



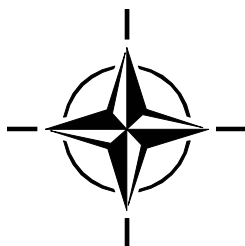
RTO MEETING PROCEEDINGS

MP-HFM-123

New Directions for Improving Audio Effectiveness

(Nouvelles orientations pour l'amélioration
des techniques audio)

Papers presented at the RTO Human Factors and Medicine Panel (HFM)
Symposium held in Amersfoort, The Netherlands, 11-13 April 2005.



Published April 2005





RTO MEETING PROCEEDINGS

MP-HFM-123

New Directions for Improving Audio Effectiveness

(Nouvelles orientations pour l'amélioration
des techniques audio)

Papers presented at the RTO Human Factors and Medicine Panel (HFM)
Symposium held in Amersfoort, The Netherlands, 11-13 April 2005.

The Research and Technology Organisation (RTO) of NATO

RTO is the single focus in NATO for Defence Research and Technology activities. Its mission is to conduct and promote co-operative research and information exchange. The objective is to support the development and effective use of national defence research and technology and to meet the military needs of the Alliance, to maintain a technological lead, and to provide advice to NATO and national decision makers. The RTO performs its mission with the support of an extensive network of national experts. It also ensures effective co-ordination with other NATO bodies involved in R&T activities.

RTO reports both to the Military Committee of NATO and to the Conference of National Armament Directors. It comprises a Research and Technology Board (RTB) as the highest level of national representation and the Research and Technology Agency (RTA), a dedicated staff with its headquarters in Neuilly, near Paris, France. In order to facilitate contacts with the military users and other NATO activities, a small part of the RTA staff is located in NATO Headquarters in Brussels. The Brussels staff also co-ordinates RTO's co-operation with nations in Middle and Eastern Europe, to which RTO attaches particular importance especially as working together in the field of research is one of the more promising areas of co-operation.

The total spectrum of R&T activities is covered by the following 7 bodies:

- AVT Applied Vehicle Technology Panel
- HFM Human Factors and Medicine Panel
- IST Information Systems Technology Panel
- NMSG NATO Modelling and Simulation Group
- SAS Studies, Analysis and Simulation Panel
- SCI Systems Concepts and Integration Panel
- SET Sensors and Electronics Technology Panel

These bodies are made up of national representatives as well as generally recognised 'world class' scientists. They also provide a communication link to military users and other NATO bodies. RTO's scientific and technological work is carried out by Technical Teams, created for specific activities and with a specific duration. Such Technical Teams can organise workshops, symposia, field trials, lecture series and training courses. An important function of these Technical Teams is to ensure the continuity of the expert networks.

RTO builds upon earlier co-operation in defence research and technology as set-up under the Advisory Group for Aerospace Research and Development (AGARD) and the Defence Research Group (DRG). AGARD and the DRG share common roots in that they were both established at the initiative of Dr Theodore von Kármán, a leading aerospace scientist, who early on recognised the importance of scientific support for the Allied Armed Forces. RTO is capitalising on these common roots in order to provide the Alliance and the NATO nations with a strong scientific and technological basis that will guarantee a solid base for the future.

The content of this publication has been reproduced
directly from material supplied by RTO or the authors.

Published April 2005

Copyright © RTO/NATO 2005
All Rights Reserved

ISBN 92-837-1147-5

Single copies of this publication or of a part of it may be made for individual use only. The approval of the RTA Information Management Systems Branch is required for more than one copy to be made or an extract included in another publication. Requests to do so should be sent to the address on the back cover.

New Directions for Improving Audio Effectiveness (RTO-MP-HFM-123)

Executive Summary

This Symposium reviewed:

- 1) The technical challenges and opportunities for more effective protection of human hearing in adverse military environments;
- 2) Efforts to optimize the portrayal of auditory information, especially via helmets and headphones; and
- 3) Research on improving the intelligibility of speech communication, especially through channels that encode and transmit multiple speech and data signals.

Participants included experts in human factors engineering, acoustics, speech & hearing sciences, military occupational health & safety, and end-users of military audio devices. The agenda included three keynote addresses and thirty-two technical papers, representing research efforts in eight nations.

This Symposium follows a 1996 AGARD meeting in Copenhagen on “Audio Effectiveness in Aviation,” and two recent HFM studies; one concerned with the effects of impulse noise, the other with technology for Active Noise Reduction (ANR).

The technical papers ranged from basic scientific studies to surveys of equipment usability in operational settings. Three papers focused on Active Noise Reduction (ANR). Four discussed bone-conducted sound and/or contact transduction for voice communication. Six papers discussed aspects of auditory perception and/or hearing mechanisms. Seven dealt with methods of presenting audio information in 3-dimensional spatial displays. Eight papers presented work on the evaluation of hearing protection devices. There were also eight on speech intelligibility.

This Technical Evaluation Report (TER) describes the background of substantial scientific and technical progress since the Copenhagen meeting, and sets forth seven primary challenges for future research efforts. The Report commends the HFM Panel and the Symposium organizers for developing a coherent and stimulating agenda, but urges planners for future meetings to seek stronger participation from the scientific community.

Although considerable progress has been made in devising hearing protection for personnel exposed to high levels of noise, such as the 150 dB sound fields near military jet aircraft, the number of persons who are disabled by noise exposure continues to mount, and it remains the No.1 military disability in this regard. One priority is to increase the use of protective equipment, through design improvements, custom fitting, education, and more rigorous implementation of policy. The goal of improved hearing protection also requires fundamental research in three specific areas:

- 1) Refined methods are needed to assess the risks from impulse noises (e.g. from small arms fire, blasts, etc.).
- 2) Research should discover how exposure to ototoxic chemicals (such as jet fuel, or solvents) exacerbates the damaging effects of noise, and evaluate possible treatments, such as anti-oxidants.

- 3) Long-term studies are needed to measure and understand the apparently wide variations among individuals in their vulnerability to noise-induced hearing loss.

This Technical Evaluation Report also recommends new efforts to improve speech intelligibility and to restore ambient spatial perception for wearers of helmets and headphones. These technical challenges cannot be met without strong investments in fundamental science aimed at a deeper understanding of human hearing. The importance of these challenges for human effectiveness cannot be overestimated. A broad, sustained effort is required in which international cooperation is vital.

Nouvelles orientations pour l'amélioration des techniques audio (RTO-MP-HFM-123)

Synthèse

Le travail du symposium a porté sur les points suivants :

- 1) défis techniques et possibilités de protéger plus efficacement l'audition du personnel dans les environnements militaires hostiles ;
- 2) efforts à déployer pour optimiser l'interprétation des informations auditives, notamment par le biais de casques et d'écouteurs ; puis,
- 3) activités de recherche sur l'amélioration de l'intelligibilité des communications verbales, notamment par le biais de canaux qui encodent et transmettent des signaux de données et vocaux multiples.

Participants au symposium : spécialistes en ergonomie, en acoustique, en sciences de la parole et de l'audition, en sécurité et santé professionnelles du personnel militaire et utilisateurs de dispositifs audio militaires. Le programme incluait trois discours et trente-deux documents techniques illustrant les efforts de recherche déployés dans huit pays.

Ce symposium fait suite à la réunion AGARD qui s'est tenue à Copenhague en 1996 et qui était consacrée à l'efficacité audio dans l'aéronautique (Audio Effectiveness in Aviation), et à deux études HFM récentes, l'une étant consacrée aux effets du bruit impulsif et l'autre aux technologies de réduction active du bruit (ANR).

Les documents techniques allaient des études scientifiques fondamentales aux recherches sur l'utilisation des équipements en configuration opérationnelle. Trois d'entre eux étaient consacrés à la réduction active du bruit (ANR). Quatre documents traitaient de la conduction du bruit par voie osseuse et/ou de la transduction par contact de la communication vocale. Six autres documents abordaient les aspects de la perception auditive et/ou de l'appareil auditif. Sept documents avaient trait aux méthodes de présentation des informations audio sur des écrans spatiaux tridimensionnels. Huit documents présentaient les travaux réalisés sur l'évaluation des dispositifs de protection de l'ouïe. Enfin, huit autres documents traitaient de l'intelligibilité de la parole.

Ce rapport d'évaluation technique (TER) décrit le contexte des avancées scientifiques et techniques substantielles réalisées depuis la réunion de Copenhague, et énumère les sept défis principaux que devront prendre en compte les efforts de recherche futurs. Le rapport félicite le Comité HFM et les organisateurs du symposium pour avoir mis sur pied un programme cohérent et stimulant, mais il recommande vivement aux planificateurs d'organiser d'autres réunions afin de renforcer la participation de la communauté scientifique.

Même si de réels progrès ont été faits pour protéger l'ouïe du personnel exposé à des niveaux de bruit élevés – personnel travaillant près des avions militaires générant des champs acoustiques de 150 dB par exemple – le nombre de personnes frappées d'incapacité en raison d'une exposition au bruit ne cesse de croître, et ce problème reste le premier facteur d'incapacité du personnel militaire. L'une des priorités

consiste à encourager davantage le port des équipements de protection en améliorant leur concept, en les adaptant aux situations, en informant le personnel et en mettant en place des politiques plus draconiennes. Optimiser la protection de l'ouïe passe également par la mise en œuvre d'activités de recherche fondamentale dans trois domaines spécifiques :

- 1) il est nécessaire d'affiner les méthodes permettant d'évaluer les risques induits par le bruit impulsionnel (p. ex. : tir d'armes à feu légères, explosions, etc.) ;
- 2) les activités de recherche doivent permettre de découvrir en quoi l'exposition à des produits chimiques ototoxiques (carburéacteur ou solvants) exacerbe les effets ravageurs du bruit, et d'évaluer les traitements potentiels, tels que le recours aux anti-oxydants ; enfin,
- 3) il est nécessaire de mettre sur pied des études à long terme afin d'évaluer et de comprendre les écarts apparemment importants qui existent entre les individus en termes de vulnérabilité à la perte d'audition due au bruit.

Ce rapport d'évaluation technique conseille par ailleurs de mettre en œuvre des moyens visant à améliorer l'intelligibilité de la parole et à restaurer la perception spatiale ambiante du personnel équipé de casques et d'écouteurs. Il n'est pas possible de faire face à ces défis techniques sans investir de façon appropriée dans les sciences fondamentales qui permettront de mieux comprendre la fonction auditive chez l'homme. L'importance de ces défis pour l'efficacité humaine ne peut pas être surévaluée. Il est par conséquent nécessaire de déployer des efforts soutenus à grande échelle pour lesquels une coopération internationale s'avère vitale.

Table of Contents

	Page
Executive Summary	iii
Synthèse	v
Human Factors and Medicine Panel	x
	Reference
Technical Evaluation Report by W.D. Larkin	T
Noise: A Limiting Factor for the Use of Modern Weapon Systems? by A. Dancer and K. Buck	KN1
A New Method for Rating Hazard from Intense Sounds: Implications for Hearing Protection, Speech Intelligibility, and Situation Awareness by G.R. Price	KN2
CONSTRAINTS and NEEDS in the Auditory and Vocal Environment of Modern Air Combat by F. Bigot	KN3
SESSION 1: NOISE	
U.S. Navy Flight Deck Hearing Protection Use Trends: Survey Results by V.S. Bjorn, C.B. Albery, R. Shilling and R.L. McKinley	1
Noise-Induced Hearing Loss in Canadian Military Personnel by S.M. Abel	2
Evaluation of Noise Effects in Auditory Function in Spanish Military Pilots by J.M. Lorente, B. Puente, B. Esteban, F. Ríos, P. Vallejo and C. Velasco	3
Distortion Products Otoacoustic Emissions as Markers of Tinnitus Persistence after Acute Acoustic Trauma by J.-B. Nottet, N. Brossard, B. Suc and A. Job	4
Auditory Tests for the Early Detection of Noise-Susceptible Individuals – A Literature Study by J. Vos	5
SESSION 2: COMMUNICATION	
Improving the Effectiveness of Communication Headsets with Active Noise Reduction: Influence of Control Structure by A.J. Brammer, D.R. Peterson, M.G. Cherniack and S. Gullapalli	6
Speech Intelligibility with Acoustic and Contact Microphones by B. Acker-Mills, A. Houtsma and W. Ahroon	7

Speech Intelligibility with a Bone Vibrator	8
by L. Pellioux, J.C. Bouy, C. Blancard and A. Guillaume	
Evaluation of Speech Synthesis Systems using the Speech Reception Threshold Methodology	9
by D.A. van Leeuwen and J. van Balken	
Extension of ITU-T Recommendation P.862 PESQ towards Measuring Speech Intelligibility with Vocoders	10
by J.G. Beerends, S. van Wijngaarden and R. van Buuren	
Military Speech Communications over Vocoders in Tandem	11
by M.A. Ericson, B.D. Simpson and R.L. McKinley	
Objective Measurement of the Speech Transmission Quality of Vocoders by Means of the Speech Transmission Index	12
by B.J.C.M. van Gils and S.J. van Wijngaarden	
SESSION 3: NOISE	
Improved Hearing Protection for Aviation Personnel	13
by R.L. McKinley, V.S. Bjorn and J.A. Hall	
Hearing Protection for Bone-Conducted Sound	14
by A.J. Dietz, B.S. May, D.A. Knaus and H.P. Greeley	
Evaluation of Acoustic Propagation Paths into the Human Head	15
by W.D. O'Brien, Jr. and Y. Liu	
Hybrid Feedforward-Feedback Active Noise Control for Hearing Protection and Communication	16
by L.E. Ray, J.A. Solbeck and R.D. Collier	
Double protection passive pour les équipages de l'hélicoptère d'attaque "Tigre" : concept et étude expérimentale	17
<i>(Double Passive Protection for "Tigre" Attack Helicopter Crews: Design and Experimental Study)</i>	
by J. Baudou, G. Reynaud, G. Poussin and A. Léger	
Experimentation to Address Appropriate Test Techniques for Measuring the Attenuation Provided by Double ANR Hearing Protectors	18
by S.E. Mercy, C. Tubb and S.H. James	
Investigating Double Hearing Protection using Human Subjects	19
by C. Tubb, S. Mercy and S. James	
A Microphone-Array-Based System for Restoring Sound Localization with Occluded Ears	20
by A.W. Bronkhorst and J.A. Verhave	

SESSION 4: DISPLAYS

Integration of Alternative Control and Display Technologies in Fighter HMI: Automatic Speech Recognition and 3DSound	21*
by A. Léger, C. Roumes, R. Gros, K. McAnally and J. Meehan	

* Paper not available at the time of publishing.

3-D Audio: Military Applications and Symbology	22
by M.A. Ericson, B.D. Simpson and R.L. McKinley	
Relative Effectiveness of Audio Tools for Fighter Pilots in Simulated Operational Flights: A Human Factors Approach	23
by S. Hourlier, J. Meehan, A. Léger and C. Roumes	
Design and Evaluation of Warning Sounds in Frigate Control Centres	24
by J.S. van Balken and S.J. van Wijngaarden	
Paper 25 withdrawn	
Spatial Audio Displays for Improving Safety and Enhancing Situation Awareness in General Aviation Environments	26
by B.D. Simpson, D.S. Brungart, R.H. Gilkey and R.L. McKinley	
SESSION 5: DISPLAYS	
Hearing Thresholds Under +GZ Acceleration in a Human Centrifuge	27*
by P. Sandor, L. Pellieux, G. Andeol and C. Dehlinger	
An Algorithm for the Accurate Localization of Sounds	28
by J.A. MacDonald	
Localization Performance of Real and Virtual Sound Sources	29
by J.A. Pedersen and T. Jørgensen	
Improving Multitalker Speech Communication with Advanced Audio Displays	30
by D.S. Brungart and B.D. Simpson	
Detection and Identification of Simultaneous Communications in a Simulated Flying Task	31
by K. McAnally, R. Martin, J. Doman, G. Eberle and S. Parker	
Comment améliorer la sélection et le traitement des messages verbaux ?	32
<i>(How to Improve the Selection and Processing of Verbal Messages)</i>	
by M. Rivenez, C. Darwin and A. Guillaume	

* Paper not available at the time of publishing.

Human Factors and Medicine Panel

CHAIRMAN

Médecin en Chef Jean-Michel CLERE
Chef de la Division
"Sciences de l'Homme et Protection"
DGA/DET/SCET/CEP/SHP – SPNuM
Tour DGA, 1626
4 bis rue de la Porte d'Issy
75509 Paris Cedex – France

VICE-CHAIRMAN

Dr. Robert FOSTER
Director, Bio Systems
Office of the Secretary of Defense
1777 N. Kent Street
Suite 9030
Rosslyn, VA 22209 – United States

PROGRAMME COMMITTEE

Chairmen

Dr. Adelbert Bronkhorst
Head of Perception Department
TNO Human Factors
3769 ZG Soesterberg – The Netherlands

Dr. Douglas Brungart
Senior Computer Engineer
Human Effectiveness Directorate, AFRL/HECB
WPAFB, OH 45433 – United States

Members

Dr. Bernd de Graaf
Manager R&D Defence
TNO Human Factors
PO Box 23
3769 ZG Soesterberg – The Netherlands

Dr. René LaRose
Director General, DRDC Toronto
1133 Sheppard Avenue West
PO Box 2000
Toronto, Ontario, M3M 3B9 – Canada

Dr. Geoff Barrett
Dstl Senior Scientific Adviser
Dstl Human Sciences
ESD Support Services, DRPS
INM, Crescent Road
Alverstoke, Hants PO12 2DL – United Kingdom

Dr. Alain Léger
THALES Avionics
Chief Scientist Human Factors
Rue Toussaint Catros
BP 91
33166 Saint Medard en Jalles Cedex – France

Maj. Dr. Ioannis Diamantopoulos
Office for Aviation Medicine
Hellenic Air Force General Staff
5 Pan. Kanellopoulo Ave
115 25 Athens – Greece

Col. Dr. Willem Tielemans
Center for Human Factors in Aviation
P.O. Box 22
3769 ZG Soesterberg – The Netherlands

Dr. Robert Foster
Director, Bio Systems
Office of the Secretary of Defense
1777 N. Kent Street
Rosslyn, VA 22209 – United States

PANEL EXECUTIVE

CDR RNLN Marten Meijer
BP 25 – 7, rue Ancelle
92201 Neuilly-sur-Seine – France
Tel: +33 1 5561 2260/62
Fax: +33 1 5561 2298
E-mail: meijerm@rta.nato.int or pelatd@rta.nato.int

Technical Evaluation Report

Willard D. Larkin

Air Force Office of Scientific Research, AFOSR/NL
875 Randolph Street, Arlington, VA 22203
USA
Tel: 703-696-7793, Fax: 703-696-8449

Willard.Larkin@afosr.af.mil

1. INTRODUCTION

This Symposium was organized to consider the challenges and opportunities offered by new audio technologies, such as active noise reduction systems and new designs for spatial audio displays. The participants included experts in human factors engineering, acoustics, speech and hearing science, audiology, military and occupational health & safety, and end users of the various devices for hearing protection and audio information portrayal. Thirty-two technical papers were scheduled, representing research and development efforts in eight nations.

The Symposium convened in Amersfoort in the conference centre, “De Eenhorn,” from Monday, 11 April to Wednesday, 13 April 2005. Open to citizens from NATO member states and NATO Partnership for Peace (PfP) states, it was capably hosted by participants from the nearby TNO Human Factors Institute, Soesterberg. Dr. Bernd de Graaf (TNO) and Ms. Marjolein Klootwijk (TNO) coordinated the local arrangements. Dr. Adelbert Bronkhorst (TNO) and Dr. Douglas Brungart (US Air Force Research Laboratory) chaired the 9-member Program Committee, assisted by Ms. Danielle Pelat (RTA, Paris). The opening session featured a welcoming address by the Mayor of Amersfoort, Mrs. Albertine van Vliet-Kuipers.

Three keynote speakers, Dr. A. Dancer (France), Dr. R. Price (USA), and Cdr. F. Bigot (France) contributed additional context for the technical papers.

The Human Factors & Medicine (HFM) Panel coordinates an array of research and development efforts beyond the scope of this Symposium, including topics such as training, selection, health, safety, and survival of military personnel. It is one of seven Panels organized under the NATO Research & Technology Board (RTB). RTB is a component of the cooperative research and information exchange undertaken by NATO's Research and Technology Organization (RTO).

2. THEME

The Symposium was organized to explore developments in three areas: noise control, speech communication, and audio displays. Several presentations reported on developments relevant to two, or all three of these inter-related topics.

The incidence of hearing loss from exposure to high levels of noise is a continuing concern, especially for military personnel whose normal work activity places them in proximity to jet aircraft engines, or exposes them routinely to impulse noise from explosives or artillery fire. Impaired hearing degrades human performance, but full protection against high noise levels may also carry a risk to human safety, because deep

attenuation also impairs the ability to localize nearby sound sources. On an aircraft carrier, for example, deck crew may face a cruel dilemma: Wear hearing protection to save the ears, or put the ears at risk to avoid bodily injury? Finding a solution to this problem has the highest possible urgency. It must be a solution that simultaneously protects human hearing and yet enables normal auditory awareness of the direction and proximity of physical danger.

Noise also impairs human safety by interfering with communication. In this case, “noise” may include a broad category of acoustic interference, such as multiple, overlapping voice communications that compete for the listener’s attention. Or it may include distortions of speech signals that stem from digital encoding or encryption. The Symposium was therefore concerned with the general question of how best to make speech intelligible, and with the closely related question of how best to measure intelligibility.

Helmets, cockpits, ship and submarine compartments have in common that they seal off the natural acoustic environment. Headphones then re-introduce electronically generated sound – speech or other audio signals – but now in mixed waveforms coinciding at the ear without distinction as to the direction or distance of these multiple sources. Recent scientific and technical developments have aimed at re-instating three-dimensional (3D) spatial perception, e.g., via head-related transfer functions (HRTFs). The Symposium addressed the technical challenges in this effort, including the assessment of 3D audio displays in the context of complex, high-workload operational tasks.

3. PURPOSE AND SCOPE

One of the Symposium’s goals was to identify the emerging technical possibilities for improving noise control, communication, and display. Another goal was to examine the problems that these new technologies may pose for adoption by end-users. This Technical Evaluation Report will describe the progress and continuing challenges brought to light in the Symposium papers and in the discussions provoked by these presentations.

The Symposium began with a session on noise, chaired by Dr. A. Bronkhorst (The Netherlands) and Dr. D. Brungart (United States). Dr. A. Dancer (France) presented a keynote address entitled, *Noise, The Limiting Factor for the Use of Modern Weapon Systems*. His address was an up-to-date review of the mechanisms of cochlear damage, the statistics of hearing disability in military occupations, the likelihood of future risks, and the need for improvements in equipment, measurement systems, and standards.

A later session on noise was chaired by Dr. Dancer and Dr. S. Abel (Canada). Thirteen technical papers were presented, overall, in the two sessions on noise.

Dr. H. Steeneken (The Netherlands) and Dr. W. Tielemans (The Netherlands) chaired a session on communication, with seven technical papers. Dr. R. Price (United States) gave the keynote address, entitled *New Considerations in Promoting Speech Intelligibility, Situational Awareness, and Hearing Protection*. His address departed considerably from this topic, and dealt mainly with an algorithm the U.S. Army Research Laboratory has developed to assess auditory hazards.

Two sessions on auditory displays concluded the Symposium. The first, chaired by Dr. A. Leger (France) and Dr. D. Brungart (United States) began with a keynote address by Cdr. F. Bigot, an experienced fighter pilot who has served in Bosnia, Kosovo, and Iraq. His address, entitled *Constraints and Needs in the Auditory and Vocal Environments of Modern Air Combat*, reflected his expertise as a pilot as well as his experience in ergonomic design. He reminded the Symposium of the importance of voice and audio in the modern cockpit,

when tactical engagements may often take place at thirty kilometre separations between the combatants, far beyond visual range.

Mr. Richard McKinley (United States) and Dr. A. Bronkhorst (The Netherlands) chaired the final session on auditory displays. Altogether, there were 12 technical papers on this topic.

The Symposium's scope can be seen in a rough enumeration of the principal areas of technical interest. These areas range from basic scientific studies to surveys of usability in operational settings. Two papers discussed audio alert symbology. Three focused on active noise reduction (ANR) techniques, or methods of testing ANR. Four discussed bone conduction and/or contact transduction. Six papers discussed aspects of auditory perception and/or hearing mechanisms. Seven papers dealt with 3D audio technology and/or its application. Eight papers presented work on the evaluation and usability of hearing protection devices. There were also eight papers on speech intelligibility.

As HFM Symposium papers and abstracts are available in published form, this Technical Evaluation Report will not review or summarize them individually. Citations to other published papers are indicated by numerals in parentheses ().

4. EVALUATION

4.1 Background & Progress

In October, 1996, NATO sponsored a conference in Copenhagen on "Audio Effectiveness in Aviation" (1). Organized by the former AGARD Aerospace Medical Panel, the Copenhagen meeting provides one point of reference for the problems and progress discussed during the Amersfoort Symposium. Two other points of reference are the recently-published HFM studies, "Assessment Methods for Personal Active Noise Reduction Validated in an International Round Robin" (2), and "Reconsideration of the Effects of Impulse Noise" (3). Prior to these, AGARD sponsored a 1981 conference on "Aural Communications in Aviation," documented in AGARD-CP-311. Held in Soesterberg, The Netherlands, this conference had 27 papers on an agenda very similar to the Amersfoort Symposium.

Consistent with the mandate and instructions given by the HFM Panel, this Technical Evaluation Report adopts a yet broader perspective, to include a general consideration of any relevant scientific and technical developments.

Hearing protection for extreme noise environments has benefited from developments of new active noise reduction (ANR) techniques and from new designs for passive attenuation by earplugs, earmuffs, and helmets. The nonlinear earplug, developed at the French-German Research Institute of Saint-Louis, (ISL) France, is a noteworthy advance (4). The ISL plug has a venturi, or orifice, such that turbulence generated at high sound levels fractures the pressure wave associated with the acoustic impulse. Thus, the effective acoustic resistance of this device increases with the peak pressure of the incident sound. For example, with an input sound level of 170 dB, the attenuation is 25 dB greater, over a broad spectrum, than it is for an input level of 110 dB. Because the ISL earplug has a lower insertion loss for low and moderate intensity sound, it preserves the possibility for speech communication, as well as for identification and localization of environmental sounds, while affording greater protection against high-level impulse noise.

NATO's interest in noise intersects with that of the International Committee on the Biological Effects of Noise (ICBEN), including concerns with noise-induced hearing loss, effects on communication and human

performance, and the development of metrics, regulations, and standards. At five-year intervals, ICBEN convenes the International Congress on Noise as a Public Health Hazard. The most recent meeting was held in Rotterdam, in 2003. The first such meeting was held in Washington, D.C., in 1968, just two years after U.S. President Lyndon Johnson's office issued a special report on jet aircraft noise. In 1972, the U.S. Congress passed the Noise Control Act. This legislation directed the U.S. Environmental Protection Agency (EPA) to develop noise exposure criteria to safeguard public health. In 1981, President Ronald Reagan terminated the EPA's noise control program, and it has never been reinstated. Thus, it falls to international organizations, such as ICBEN and NATO, to provide leadership in dealing with the problems of noise. Several participants in the HFM Symposium were represented in the recent ICBEN Congress. It is expected that several will again participate in the next Congress, scheduled for 2008, which will mark the first return of this important meeting to the United States. The Chairperson for the 2008 Congress is Dr. Jerry V. Tobias, recently retired from the U.S. Navy Submarine Medical Research Laboratory.

The 1996 Copenhagen conference inspired efforts to develop standards and metrics, especially with respect to noise exposure criteria and the evaluation of hearing protection devices. The more recent RTO-HFM study (2) reported on tests of active noise reduction (ANR) systems under conditions of use that represent adverse military deployments. The study adopted a "round-robin" testing plan, in which each participating laboratory examined several ANR headset systems in turn. To conduct these tests, the TNO Human Factors group organized groups working on military ANR in Canada, France, Germany, the Netherlands, the United Kingdom, and the United States. The study took two years to complete. Measurements among the different laboratories showed a high degree of agreement, with only small, generally non-significant differences in the test data. However, the performance of these ANR systems was found to be highly vulnerable to acoustic leakage around the headsets – i.e. to the success, or lack thereof, in fitting standard headsets comfortably to individual wearers. The problem of optimizing fits to individuals has received subsequent study, and was one topic of lively discussion in the Amersfoort Symposium.

Systems for active noise-reduction (ANR) at the ear have been developed to accommodate multiple functions, such as heart rate monitoring and communication. They also have been successfully combined with passive hearing protection. Laboratory tests of new hybrid control architectures, combining feed-forward and feedback algorithms, were reported at the Symposium (5, 6). These tests show that a hybrid system can provide improved attenuation of non-stationary noise, such as F-16 aircraft or helicopter cockpit noise, compared with the performance of either control component acting alone.

There has been significant progress toward the 50 dB attenuation goal for hearing protection in very high-noise environments, such as the 150 dBA sound fields near military jet engines. Much of this progress is very recent, and reflects success with custom-molded, deep-insertion earplugs. The process of molding the plug to individual ears has become simplified, efficient, and reliable. Anecdotal reports indicate that the plug is most comfortable only when correctly worn. Therefore it can be expected that wearers will have a comfort incentive to optimize the plug's performance. Current efforts to design custom-fitting ear cups and cranials (helmets) also show good prospects for improved hearing protection.

Despite these technical advances, noise-induced hearing impairment remains at the top of the list of military disabilities, in number of individuals affected. Yearly estimates for direct compensation for hearing loss among military personnel continue to rise at an alarming rate. In 2004, for example, the United States Veterans' Administration spent nearly \$634 million in compensation for hearing loss as a primary disability. Other countries reported similar substantial expenditures for hearing loss treatment and compensation. The urgency of this problem, obvious to all at the Symposium, calls for a vigorous and consistent research and development effort.

Another area of recent progress stems from scientific studies of binaural hearing, and the measurement of interaural differences in timing and level for acoustic waveforms reaching the two ears from various directions. These measurements provide the basis for head-related transfer functions (HRTFs), which have been exploited to enable 3D spatial perception in headphones. Although individual HRTFs differ from one person to the next, even generic versions prove useful. One important application is the use of spatial audio to localize targets and/or threats that either cannot be seen or would not be noticed if visual search were the only method of detecting them. Another important application is the imposed spatial separation of multiple audio signals or multiple speech sources, so that even if they overlap in time, they will occupy apparently different positions in the listener's acoustic space. One version of this spatial separation technology, reported in the Symposium by Dr. D. Brungart and Dr. B. Simpson, has been designed for AWAC personnel. It provides an optimal separation of speech channels, in virtual directions and distance, for up to seven simultaneous voices.

In a related Symposium paper, Dr. K. McAnally and his colleagues demonstrated the effectiveness of the spatial separation approach. They found that up to 5 communication channels, separated in azimuth, could be monitored in a simulated flying task. The 3D audio condition was better than non-spatialized conditions for speech detection and recognition, and it reduced the perceived workload.

The Symposium took place against a backdrop of greatly expanding interest and invention in the area of auditory displays. Much of this activity has exploited powerful new electronic capabilities to generate, organize, and manipulate waveforms, both for psychoacoustic study and for application in virtual audio displays. Work of this kind was under-represented at the Symposium, although its scope and variety easily could have overwhelmed a three-day program. Many ongoing developments in this area are relevant to human systems design in the military environment. A few examples will suffice to represent the scope of current efforts:

- Sonification of visual displays – e.g., to help diagnose complex turbulence patterns in the atmosphere or in ocean waves.
- Efficient methods to study perceptual classifications of sonic stimuli – e.g., multidimensional representations for very large “earcon” sets.
- Auditory navigation aids for virtual environments – e.g. to represent self-motion, platform motion, or target motion.
- Manipulation of non-traditional sonic parameters – e.g., to represent the size, shape, or composition of an auditory source.
- Control of virtual reverberation and acoustic shadowing – e.g., to assist distance and location judgments within a room or other closed areas.
- New stethoscope designs – e.g., with better impedance matching to the skin and ANR electronics for use in adverse military environments.
- Auditory alerts not based upon loudness – e.g., using temporal or spatial jitter to capture attention and better inform the listener.

Because of the growing commercial incentives in non-military applications, work in the area of auditory displays is likely to advance very rapidly. It will be increasingly important to monitor these innovations. International coordination in these areas is substantially aided by the annual International Conferences on Auditory Display (ICAD). The 11th ICAD meeting will take place in Ireland, July 6-9, 2005. The chairperson is Mikael Fernstrom, of the University of Limerick.

Technical Evaluation Report

This rapid growth of new audio display techniques, although greatly needed, confronts a long-standing human factors problem: how to evaluate the consequences of any technical innovation for human performance. Several Symposium papers brought to light the extreme care – and often the extreme expense – required in such evaluations. Given the manifold complexity of many operational tasks, how should new auditory technology be used to optimize performance?

To take just one example, consider the control interface for an uninhabited air vehicle (UAV). Considering system costs as well as ergonomics, should haptic feedback or acoustic feedback be used to convey air turbulence? Would the accident rate decrease if the navigation map were provided with combined egocentric and exocentric directional cues? How would photographic time-on-target be affected? How would fuel use be affected? Should an auditory horizon be installed, in addition to the visual one? Aside from the obvious multiplicity of such questions, one of the great impediments to human factors optimization is that trade-offs among the various options are difficult to measure in advance of actually building and testing the entire system.

A traditional response to this evaluation problem has been to conduct empirical studies of various options, usually one option at a time, in laboratory set-ups that usually represent only a small fraction of the real-world task complexity. Over the past eight years, a new approach has been developed, combining computational modeling with cognitive task analysis (7). This approach develops a Synthetic Task Environment (STE), validated against expert performance in the real-world task, along with a validated computational model of human performance (representing all levels of training, from novice to expert). Design changes in the task interface can then be implemented quickly, and the consequences of these changes can be assessed, using all the metrics relevant to system performance. (One of the first STE's was constructed in the U.S. Air Force Research Laboratory for a UAV control task. This laboratory-based STE was not designed to train UAV operators, but it proved so successful that it was later adopted for that purpose, at one-tenth the cost of the originally planned training system.)

One key element in this approach is computational modeling that not only represents human performance within a complex system, but also opens quantitative pathways to define optimal performance. There always will be a role for the traditional laboratory-bench assessments of new ideas, including new types of audio displays, but in the expensive and often protracted development of complex military systems, there can be a much greater role for Synthetic Task Environments to evaluate these ideas in a rapid prototyping mode. To succeed in this approach requires a confluence of expertise, including cognitive science and optimization theory paired with traditional ergonomics.

4.2 Scientific & Technical Challenges

This section of the Technical Evaluation Report takes up a selected group of continuing challenges, all of which were discussed during the Symposium. Although much has been accomplished in each of these areas, much remains to be done. Solutions to these problems range from near-term to far-term in accessibility. They are considered, below, in a quasi-logical order, which should not be mistaken for an endorsement of relative priorities: All of these scientific and technical challenges are important.

4.2.1 Hearing Protection: Use and Non-Use

One consensus among many at the Symposium is that military personnel do not take full advantage of the hearing protection made available to them, and that the research community lacks a complete understanding of the phenomenon. For example, in a study of 301 U.S. Navy flight deck crewpersons, Ms. V. Bjorn and her

colleagues found that nearly half never wore earplugs & muffs that would have provided 30 dB protection. Dr. S. Abel, reporting on a Canadian study of 1,057 personnel, found that many avoided hearing protection because they found it incompatible with other gear. The cumulative effects of noise exposure in these personnel are substantial, as they are for the 252 pilots Spanish pilots described in another Symposium report by Dr. J. M. Lorente and colleagues.

Continuing field studies will be needed to define the problem, the elements of which will likely differ from one operational context to another. Continuing surveillance, educational efforts, and re-engineering will be needed to reverse the dismal statistics. NATO can have an important role in promoting international exchanges regarding successful and unsuccessful practices, stimulating reviews of policy and implementation, and coordinating field trials to improve compliance. The potential benefits to human welfare give this endeavor the highest urgency.

4.2.2 Impulse Noise

Understanding the effects of exposure to impulse noise remains an important, unsolved problem. Long-term effects of this exposure have not been quantified. Nor is there a reliable way to predict the long-term hazard. The heart of the matter is that dose-effect relationships have not been developed for relevant metrics of long-term audiological outcomes, such as a permanent threshold shift (PTS). Research effort in this direction should be a high priority.

When a PTS is found, it is usually impossible to make a thorough retrospective analysis of the individual's exposure history. Impulse waveforms from small arms fire, large calibre weapons, blast noise, and the like, can vary in important details, such as duration, peak level, total energy, impulse repetition rate, and level of background noise. For example, as Dr. A. Dancer reported in his Symposium keynote address, a 120 mm mortar round can register a peak pressure of 185 dB, while a 0.50 calibre rifle registers about 175 dB at the shooter's ear. If damage risk criteria were to focus on peak pressure alone, then it is conceivable that current methods of hearing protection would not satisfy the new European Directive under consideration, which would limit residual peak pressure under earplugs to a maximum of 137 dB. On the other hand, it is likely not only that other waveform parameters matter, but also that these parameters enter a risk equation in a complex, nonlinear fashion, with an important dependence on long-term, cumulative exposures. Data on temporary threshold shifts (TTS) may be necessary, but they are not sufficient to understand this problem.

One Symposium paper, reported by Dr. J. Vos, noted that although a temporary threshold shift (TTS) can be measured under controlled (or at least well-monitored) exposure conditions, there are no data to support a fixed relationship between TTS and PTS at the individual subject level. Although a correlation may exist in aggregate epidemiological data, especially for continuous noise exposures, the problem is more complex for impulse noise. A relationship between TTS and PTS can be possible if the TTS definition were relaxed to allow a more prolonged interval between exposure to the sound and measurement of the consequent threshold shift. However, no scientific study can afford to prolong this interval without ensuring strict controls designed to actually prevent PTS.

An alternative can be to look for other psychoacoustic predictors of long-term hearing impairment. One approach discussed at the Symposium suggests that otoacoustic emissions can be useful. This research, presented by Dr. Agnes Job, asked if the time constant of tinnitus persistence after an acute acoustic exposure can be predicted from data on hearing thresholds and/or from the measurement of distortion-product otoacoustic emissions (DPOAEs). The data are encouraging. They distinguish clearly between groups of individuals whose tinnitus subsides before 72 hours, post-exposure, and individuals whose tinnitus persists

beyond 72 hours. That any such outcome difference exists is itself noteworthy, and relates to the general question of individual differences, to be considered next in this TER. Otoacoustic emissions are, of course, an incompletely understood phenomenon. Repeated measurements in the same ear do not always yield stable results. Their relationship to cochlear health and function is still an open topic for research, as they can be absent in some individuals with otherwise apparently normal hearing. Nevertheless, there is much to gain from a concerted effort to understand how the integrity of the cochlear mechanism could be assessed with this method.

A recent study (8) examined virtually all the available data on temporary threshold shifts (TTS) from impulse sounds. The analysis included results from the Blast Overpressure Project directed by the late Dan Johnson in Albuquerque, New Mexico, USA. Different results were found for short- and long-duration sounds. The data suggest strongly that frequency-dependent weighting is appropriate to reduce the contribution of low-frequency energy, but no single numerical rule emerged. The complexity of this problem demands an approach from first principles, rooted in auditory biophysics and neurophysiology. It also would be highly desirable to examine a broader array of audiological consequences, beyond the TTS and PTS, to discover, for example, how auditory masking, discrimination, and binaural processing might be affected, over the long term, by exposure to impulse noise.

4.2.3 Individual Differences

Some military recruits arrive with enough prior hearing damage, presumably from years of exposure to excessive noise, to disqualify them from many military assignments. Even those who can pass a basic audiological exam may differ greatly in hearing capabilities that can be important for human performance. The Symposium discussion repeatedly turned to this problem, as it was exposed in data from several research studies. The ability to localize acoustic sources, for example, is critical to many military tasks. To provide spatial audio information at the headphones, or inside a helmet, depends upon matching audio cues to this human ability. But tests of the ability to localize sound sources show large individual differences; apparently as variable as other differences in hearing sensitivity (e.g., pitch discrimination, absolute thresholds). It would be rather easy to develop efficient audiological tests for spatial hearing. A first priority should be to include these tests in standard audiological examinations. To understand the basis for the large differences among individuals, however, remains a major challenge for research.

A second aspect, also discussed at the Symposium, is the large variance between individuals in their apparent susceptibility to noise exposure. The word *apparent* is needed here because susceptibility is usually measured in a very limited way, such as a temporary threshold shift (TTS), and without complete data on relevant exposure histories. The notion of an “acquired resistance” to noise exposure has become popular in some quarters, based upon laboratory studies with animals. As with the now-discredited belief that people can adapt permanently to chronic sleep deprivation, this notion is likely to succumb, eventually, to more penetrating scientific inquiry. The important question is: Do some personnel require a more conservative approach to protect their hearing? Judging from the Symposium papers, the answer to this question is yes, but it is not yet possible to assess or to predict individual vulnerability with any degree of confidence. A much deeper understanding of long-term exposure is needed. To craft the next generation of damage-risk criteria, there must be a greater investment in longitudinal studies, coupled with psychoacoustic research to discover how the human auditory system adapts – or fails to adapt – to repeated sonic insults.

4.2.4 Ototoxic Exposures and Noise

It is known that exposure to high levels of noise generates reactive oxygen species (“free radicals”) in the cochlea, and that cochlear injury from noise can be exacerbated by exposures to solvents, jet fuel, antibiotics, and other chemical agents that contain reactive oxygen species (9, 10). The Symposium was not designed to include explicit consideration of this topic, but it relates to the question of hearing protection generally, as well as to the specific question of individual differences in susceptibility to noise. For individuals with hearing loss, an indeterminate prior history of exposures to various chemicals masks the effects of noise, and thereby complicates the audiologist’s task. Research on this question is beyond the scope of the Symposium, but not beyond the scope of the HFM panel. Modern research techniques, such as genomics and proteomics, could shed light on risk factors and biological mechanisms. Studies of this kind should be coupled to investigations of antioxidants, as some research has indicated they can be effective, at least in the short term, as a prophylactic against hair cell damage from noise or from chemical exposures (11, 12, 13).

4.2.5 Speech Intelligibility

Eight Symposium papers dealt directly with speech intelligibility. This topic represents a central challenge for audio effectiveness. One aspect of this challenge is to devise consistent metrics to evaluate the intelligibility of speech as it is received via electronic communication channels of one kind or another, and under circumstances in which there may be interference from ambient noise, channel noise, distortion, and overlapping non-speech signals. Algorithms for this purpose are under development in the signals intelligence community, as well as in the human factors community, as both groups attempt to keep pace with the rapid expansion in applications of electronic signal processing to speech transmission, speech separation, and speech synthesis. The era of single, dedicated transmission lines for voice communication is ending. In many cases, the same channel is used for voice, data, and video. There is an increasing use of vocoders (often in tandem, with entirely different types of voice coding at the sending and receiving stations). Signals are encrypted. Information flows in packet networks with multiple possibilities for distortion from peak-clipping, jitter, delays, packet loss, etc.

Against this backdrop, methods of intelligibility analysis that once served the need must be re-invented or replaced. The Symposium included excellent papers on this theme. In one, Dr. S. van Wijngaarden showed how to adapt the Speech Transmission Index (STI) sensibly to allow assessments with vocoders. In another, Dr. J. Beerends demonstrated that PESQ, an algorithm originally developed at TNO, Soesterberg, to assess speech quality rather than intelligibility, could be adapted to the latter task with low-bitrate encoders. That PESQ can work well in this domain deserves to be better known in the speech research community.

Efforts to improve the metrics for speech intelligibility will be continually challenged as new technologies for speech signal transmission continue to expand. This challenge equally applies to machine algorithms for speech detection, speaker identification, and speech restoration. But another, more difficult aspect of the speech intelligibility problem pleads for research attention, and offers a high potential payoff: How to achieve genuine enhancement of transmitted speech?

The term “speech enhancement,” as it is currently employed, often refers only to methods that attempt to separate speech signals from their noisy backgrounds. These methods can be categorized, roughly, either as “blind separation” techniques based upon a purely statistical analysis of the signal, such as Independent Component Analysis (ICA), or as “model-based” techniques that exploit aspects of human hearing and/or human speech production mechanisms. Both methods can improve the audibility of the speech signal. Neither approach has yet produced speech enhancement in the stronger sense of the term, in which a speech signal would be remarkably more intelligible than the original, noise-corrupted signal. If the original speech signal

can be heard at all, the human ear is a very good instrument for making sense of it. Machine techniques sometimes render speech less intelligible, because although the speech may be easier to hear, it is often truncated in ways that put a greater burden on the listener's ability to compensate for important, but missing articulatory, prosodic, and psycholinguistic features (14). This effect is the opposite of what is needed to ensure reliable voice communication. The problem is compounded, of course, when the communication is among people from diverse linguistic cultures, such as the NATO member states.

The goal of genuine speech enhancement sets a difficult course for research, but it should be embraced. It aims beyond present concerns with separating speech from noise and/or "high-fidelity" recovery of the original speech source. It is now possible to study new methods of speech modification that could sacrifice fidelity to a limited degree, yet provide a large gain in the listener's ability to understand or to identify a speaker. For example, Hideki Kawahara (15) has shown how it is possible to simulate altered vocal tract resonances and vocal tract sizes to produce exceptional or unusual voice characteristics. Roy Patterson (16) has begun to exploit this technique to study how these altered voices are perceived. Such techniques expand the palette of audible speech in ways that could provide a human factors benefit. Further benefit can be expected from research into the basic mechanisms of speech perception, especially the so-called "cognitive" operations of the human auditory system. Such mechanisms are at present almost entirely absent from speech-processing algorithms. The human factors goal should be to define the optimal characteristics of the speech signal, for intelligibility purposes, and then work to restore these features when they are missing or degraded. "Supranormal intelligibility," rather than recovery of the original speech signal, should be the goal.

4.2.6 Contact Transducers and Bone Conduction

A contact microphone, placed directly on the skin of the face or throat, has the possibility to eliminate much of the airborne background noise that could otherwise mask a speech signal. Such microphones could be extremely useful if the speaker is inside a noisy vehicle or helicopter. Of course, the microphone's placement can be problematic, because the biophysical machinery for speech production is widely distributed, from the lips to the tongue, to the back of the mouth, to the larynx: No single contact position will capture the full acoustic signal. As Dr. B. Acker-Mills reported in the Symposium, such microphones, acting as low-pass filters, may well pick up vowels, but not fricatives, resulting in poor overall speech intelligibility compared with noise-canceling air microphones. The idea is still rather new, and deserves further study. It may be possible to combine air- and contact microphones to good advantage. The restoration of missing fricatives through signal processing is another approach, though an ambitious one, in line with the speech-enhancement goals outlined above.

Contact transducers that bypass the ear canal also have potential to reduce audible noise in comparison with standard earphones. In the Symposium, Dr. L. Pellieux showed encouraging data for speech intelligibility using a vibrator in direct contact with the listener's head. As expected, the mechanical interface and the precise placement and geometry of contact on the skin surface can be critical; many such issues remain to be investigated.

In very high noise environments, adequate protection of the cochlea may require consideration of sound transmission paths that by-pass the ear canal. Helmets add some protection against non-airborne conduction, especially at frequencies above 1 kHz. However, little is known to answer very basic questions about bone conducted sound: How does acoustic energy reach the cochlea via multiple pathways through the head and torso? How dispersed and how focused is this energy? What is its masking potential? What is its contribution to long-term hearing impairment? Can it be suppressed with active noise reduction (ANR) techniques? Two Symposium papers reported new investigations of these primary biophysical and psychoacoustical problems.

Dr. A. Dietz summarized an extensive project that compared mechanical impedance calculations with measurements using a head simulator and with psychoacoustic data. Dr. W. O'Brien reported progress in the development of a model that tracks the propagation of acoustic waveforms through the human head. An ongoing French-American collaboration (A. Dancer, H. von Gierke, and R. McKinley; not reported at the Symposium) is pursuing psychoacoustic measurements related to bone-conducted loudness and masking effects. None of this research yet translates into revolutionary new techniques to control acoustic input to the cochlea, but it is an entirely new effort, it is being carried out with great care and expertise, and it deserves high priority for continued support.

4.2.7 Restoration of Ambient Spatial Perception

To the extent that hearing protectors and helmets succeed in blocking unwanted noise, they also reduce perceptual access to the ambient acoustic environment. It can be important to restore this access, especially if the environment is complex and dangerous. The human factors challenge is to devise equipment that makes headgear as "transparent" as possible to the nearby, low-level sounds of interest, yet affords maximum protection against high-level sonic intrusions. In one Symposium paper, Dr. A. Bronkhorst reported some progress toward this dual requirement. Concentrating on the need for spatial awareness of the local acoustic environment, he and Mr. J. Verhave tested a multi-microphone system that gave improved localization judgments, compared with a conventional earmuff lacking in sound pass-through. The idea is similar to that under intensive investigation for use in hearing aids, and for similar reasons: The human auditory system depends critically on spatial perception to disambiguate the acoustic world. Without this ability, human performance is disabled.

There is ample room for new research to solve this problem. One puzzle is to understand how a microphone system, such as the one reported above, should match interaural time and intensity cues to the perceptual requirements of the user. If microphones are spaced more widely than the user's two ears, the interaural timing will be inevitably mismatched to the user. Yet, preliminary data seem to indicate that people can adapt to non-optimal HRTFs. A general study of this possibility could be very informative.

5. RECOMMENDATIONS

This Symposium was well organized to consider a wide range of important technical challenges. By limiting the sessions to inter-related topics on noise, communication, and audio displays, the Symposium maintained good coherency. This plan can be recommended for future NATO meetings, even if some of the additional questions discussed in this TER are taken up.

In general, future meetings could benefit if there were a greater representation from the scientific community. This would foster stronger awareness of, and collaborations with, the community of technology users. At this Symposium, many papers demonstrated excellent connections with current scientific developments related to the technical problems under review, but some did not.

Quite aside from the value of bringing scientific perspectives to meetings such as this, it is clear that many of the more urgent technical problems require a greater investment in basic research aimed at understanding the human auditory system. As already mentioned, the need for deeper insight into speech perception is key to developing more effective speech processing technology. To cite just one further example, the hearing-protection problem should become a motivating beacon for basic scientific efforts aimed at understanding the peripheral auditory system. Most current models represent the auditory periphery as a group of feed-through

circuits, without taking into account the important role of efferent feedback. Neural efferents are widely believed to function as a neurological gain control system at the hair cell level, and they may also have a role in auditory selective attention, detection, and masking. A deeper understanding of this system could inform efforts to develop damage risk criteria and new protective technologies.

Section 4.2, above, describes a group of seven primary challenges for future research and technical development:

1. Hearing Protection: Use and Non-Use
2. Impulse Noise
3. Individual Differences
4. Ototoxic Exposures and Noise
5. Speech Intelligibility
6. Contact Transducers and Bone Conduction
7. Restoration of Ambient Spatial Perception

Aligned with these seven are many usability issues and technical questions that also deserve attention, but are included above by implication: Among them are the effects of whole-body vibration, the compatibility of hearing equipment with night vision equipment or goggles, the use of machine algorithms to separate one speech channel from another, the robustness of HRTF algorithms to noise and reverberation, and the ever-present problem of how new technologies contribute to, or reduce task workload.

The scope and importance of all these challenges for human effectiveness demands a broad and sustained effort. International coordination is vital. Although many non-governmental organizations take an interest, NATO has a leadership role in fostering the cross-disciplinary efforts that are the engines of technical progress.

6. REFERENCES

- [1] Audio Effectiveness in Aviation. AGARD Aerospace Medical Panel Symposium, Copenhagen, October, 1996. AGARD/AMP, 7 Rue Ancelle, 92200, Neuilly-sur-Seine, France.
- [2] RTO-TR-HFM-094. Assessment Methods for Personal Active Noise Reduction Validated in an International Round Robin. August, 2004. Research and Technology Organisation (NATO), BP 25, F-92201 Neuilly-sur-Seine Cedex, France. ISBN 92-837-1121-1.
- [3] RTO-TR-HFM-017. Reconsideration of the Effects of Impulse Noise. April, 2003. Research and Technology Organisation (NATO), BP 25, F-92201 Neuilly-sur-Seine Cedex, France. ISBN 92-837-1105-X.
- [4] Dancer, A., Hamery, P. (1998). A new nonlinear earplug for use in high-level noise environments. *Journal of the Acoustical Society of America*, 103, p.2878.
- [5] Cartes, D., Ray, L. R., Collier, R. D. (2002). Experimental evaluation of leaky least-mean-square algorithms for active noise reduction in communication headsets. *Journal of the Acoustical Society of America*, 111, 1758-1772.

- [6] Streeter, A., Ray, L. R., Collier, R. D. (2004). Hybrid feedforward-feedback active noise control. Proceedings of the June 2004 American Control Conference, Boston, U.S.A., 2876-2881.
- [7] Gray, W. D. (2002). Simulated task environments: The role of high-fidelity simulations, scaled worlds, synthetic environments, and mircroworlds in basic and applied cognitive research. Cognitive Science Quarterly, 2(2), 205-227.
- [8] Smoorenburg, G. F. (2003). Damage risk with impulse-noise exposure. Proceedings of the 8th International Congress on Noise as a Public Health Problem. ICBEN 2003, Schiedam, The Netherlands. Pp. 1-4. ISBN 90-807990-1-7.
- [9] Ohinata, Y., Miller, J. M., Altschuler, R. A., Schacht, J. (2000). Intense noise induces formation of vasoactive lipid peroxidation products in the cochlea. Brain Research, 878, 163-173.
- [10] Rao, D. B, Moore, D. R., Reinke, L. A., Fechter, L. D. (2001). Free radical generation in the cochlea during combined exposure to noise and carbon monoxide: an electrophysiological and an EPR study. Hearing Research, 161, 113-122.
- [11] Yamasoba, T., Nuttall, A. L. Harris, C., Raphael, Y., Miller, J. M. (1998). Role of glutathione in protection against noise-induced hearing loss. Brain Research, 784, 82-90.
- [12] Schacht, J. (1999). Antioxidant therapy attenuates aminoglycoside-induced ototoxicity. Annals of the New York Academy of Science, 884, 125-130.
- [13] Ohinata, Y., Miller, J. M., Schacht, J. (2003). Protection from noise-induced lipid peroxidation and hair cell loss in the cochlea. Brain Research, 966, 265-273.
- [14] Divenyi, P. (Ed.), (2005). Speech Separation by Humans and Machines. Dordrecht, The Netherlands: Kluwer Academic Publishers. ISBN: 1-4020-8001-8.
- [15] Kawahara, H., Irino, T. (2005). Underlying principles of a high-quality speech manipulation system STRAIGHT, and its application to speech segregation. In P. Divenyi (Ed.) Speech Separation by Humans and Machines. Dordrecht, The Netherlands: Kluwer Academic Publishers. ISBN: 1-4020-8001-8.
- [16] Smith, D.R., Patterson, R. D., Turner, R., Kawahara, H., Irino, T. (2005). The processing and perception of size information in speech sounds. Journal of the Acoustical Society of America, 117, 305-318.



Noise: A Limiting Factor for the Use of Modern Weapon Systems?

A. Dancer and K. Buck

“Acoustics and Protection of the Soldier” (APC)
French-German Research Institute of Saint-Louis (ISL)
5 rue du Gl. Cassagnou, BP 70034
F-68301 SAINT-LOUIS Cedex

dancer@isl.tm.fr

NOISE OF MODERN WEAPON SYSTEMS

Continuous and/or impulse noises produced by modern weapon systems constitute a threat for the health of the soldier and impede his operational ability. Moreover, their levels often exceed the statutory exposure limits as well for the users (unprotected and protected ears) as for the nearby community.

1.0 INTRODUCTION

Modern weapon systems produce very high levels of impulse and/or continuous noise. Exposure to intense noise induces mechanical and/or metabolic damage to the inner ear.

At the hearing threshold the amplitude of the displacements of the sensory structures of the inner ear (stereocilia) is about 10^{-12} m (1/100 the diameter of the hydrogen atom). At 120 dB this amplitude reaches 1 micrometer (corresponding to an angular deflexion of ~ 15 degrees of the stereocilia). Depending on the level of the noise, these structures may break off immediately (i.e., for large continuous and/or impulse noises) or be overpowered by fatigue failure mechanisms (figures 1 and 2).

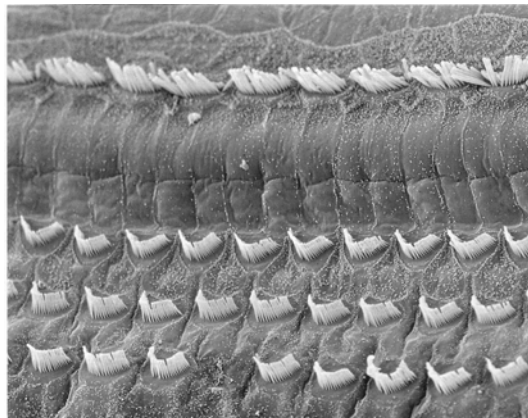


Figure 1: Intact hair cells and stereocilia

Dancer, A.; Buck, K. (2005) Noise: A Limiting Factor for the Use of Modern Weapon Systems? In *New Directions for Improving Audio Effectiveness* (pp. KN1-1 – KN1-14). Meeting Proceedings RTO-MP-HFM-123, Keynote 1. Neuilly-sur-Seine, France: RTO.
Available from: <http://www.rto.nato.int/abstracts.aps>.

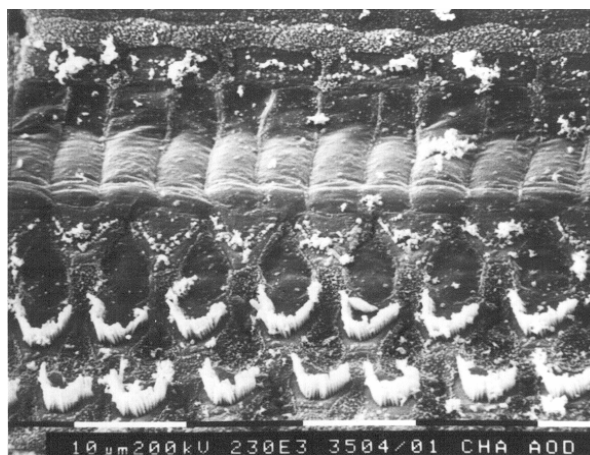


Figure 2: Damaged hair cells and stereocilia

Moreover, immediately after an exposure to a loud noise, one can observe a swelling of the afferent synapses (the interface between the sensory cells and the fibers of the auditory nerve that conveys the hearing message to the upper auditory centers) due to the glutamate excitotoxicity (figure 3). In the worst cases, the synapses burst and the nerve fibers disconnect from the sensory cells (figure 4).

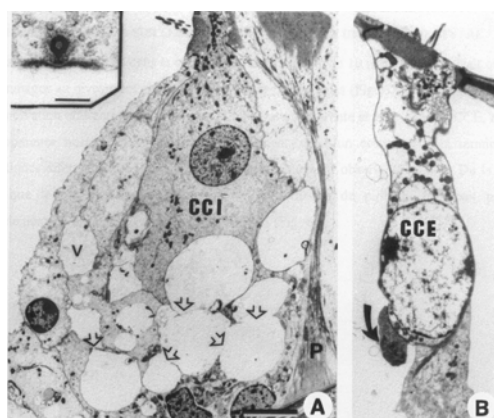


Figure 3: Swelling of the afferent synapses under the Inner Hair Cells (CCI) (CCE: Outer Hair Cells)

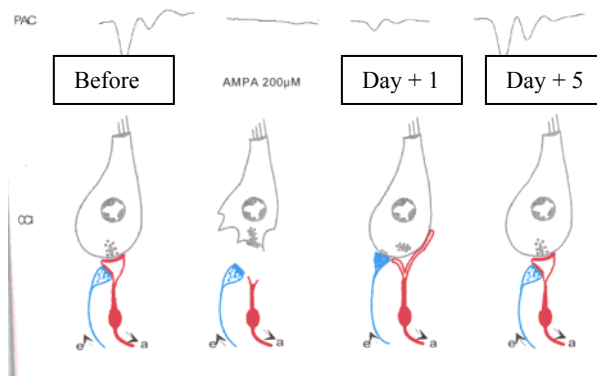


Figure 4: Schematic representation of synaptic damage due to metabolic excitotoxicity

Over-exposure to noise damages permanently the sensory cells and induce elevated threshold (Permanent Threshold Shift: PTS), impairment of frequency selectivity, recruitment and tinnitus [1].

2.0 HEARING DAMAGE CONSEQUENCES

2.1 Operational consequences

The hearing losses and the decrease in frequency selectivity induce difficulties to detect, localize and identify acoustic sources in the environment and impedes the efficiency and the security of the soldier. Moreover, the impairment of speech intelligibility (especially in noisy environments) can drastically reduce the global performance of complex and expensive weapon systems [2] (figure 5).

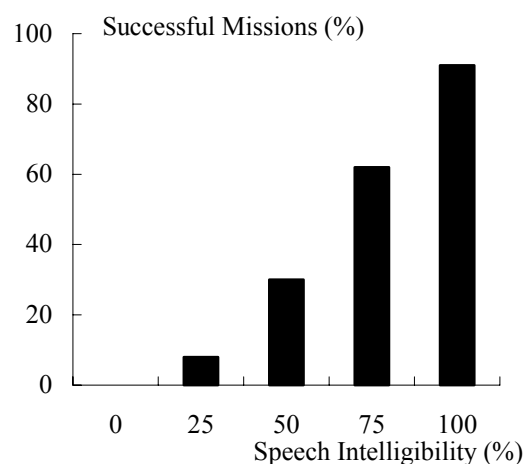


Figure 5: Tank performance: percentage of successful missions (including navigation, reporting and gunnery) as a function of speech intelligibility [2]

2.2 Financial consequences

The acoustic trauma represents the first cause of morbidity in the military during peace time !

The Noise-Induced-Hearing-Losses are responsible for many expenses.

Soldiers suffering large PTS can be definitively withdrawn from front line service. For specialized personnel high educational and training expenses may be definitively wasted. Moreover, PTS is considered as war injury and must be compensated [3].

For this cause, in 2003, 548 million dollar have been distributed to 74,363 US veterans (figure 6).

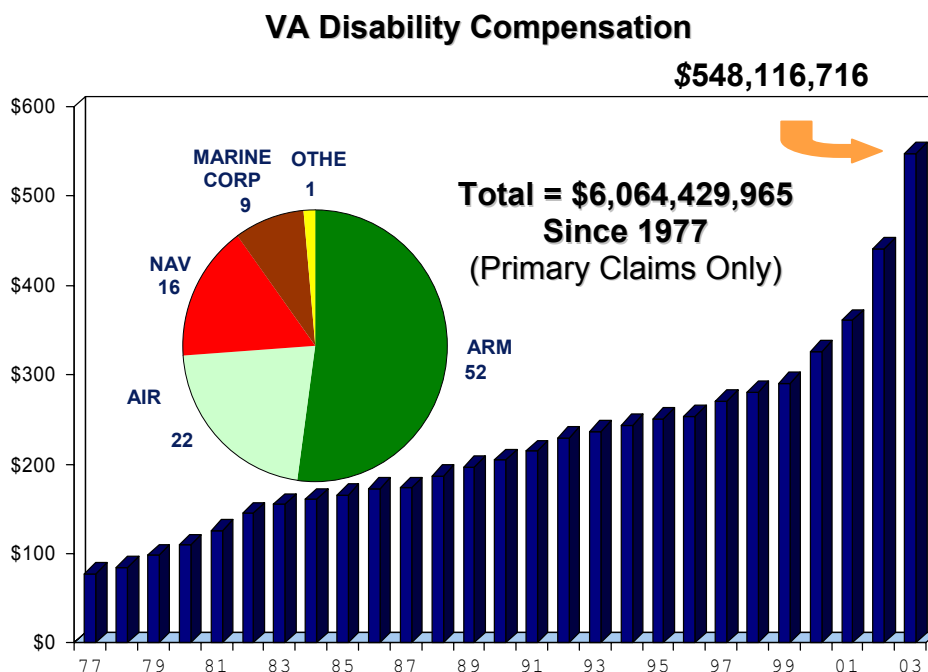


Figure 6: Hearing Loss disability costs in the USA [3,4]

In the UK, in 2002, 20% (£40m) of the litigation claims are directly related to noise and vibration and this figure is doubling every 4 years. There are also additional costs: retraining of downgraded personnel, training of new replacement recruits (~ £2m per head aircrew) [5].

In France, the annual cost of the compensations is evaluated to 50 million euros.

In Belgium, about two thirds of the 6 million euros paid yearly to the veterans for all kinds of disabilities correspond to NIHL.

Moreover, the medical treatments (of which the efficiency is questionable) are also very expensive [6].

3.0 CONTINUOUS NOISE

Heavier, faster, and more powerful weapon systems produce higher continuous noise levels. In the following, some examples will be presented for a better understanding of the problem.

3.1 Jet noise

The new fighters are probably the noisiest military weapon systems. The spectrum of the noise is generally broadband with a maximum at mid-frequencies (~ 1 kHz).

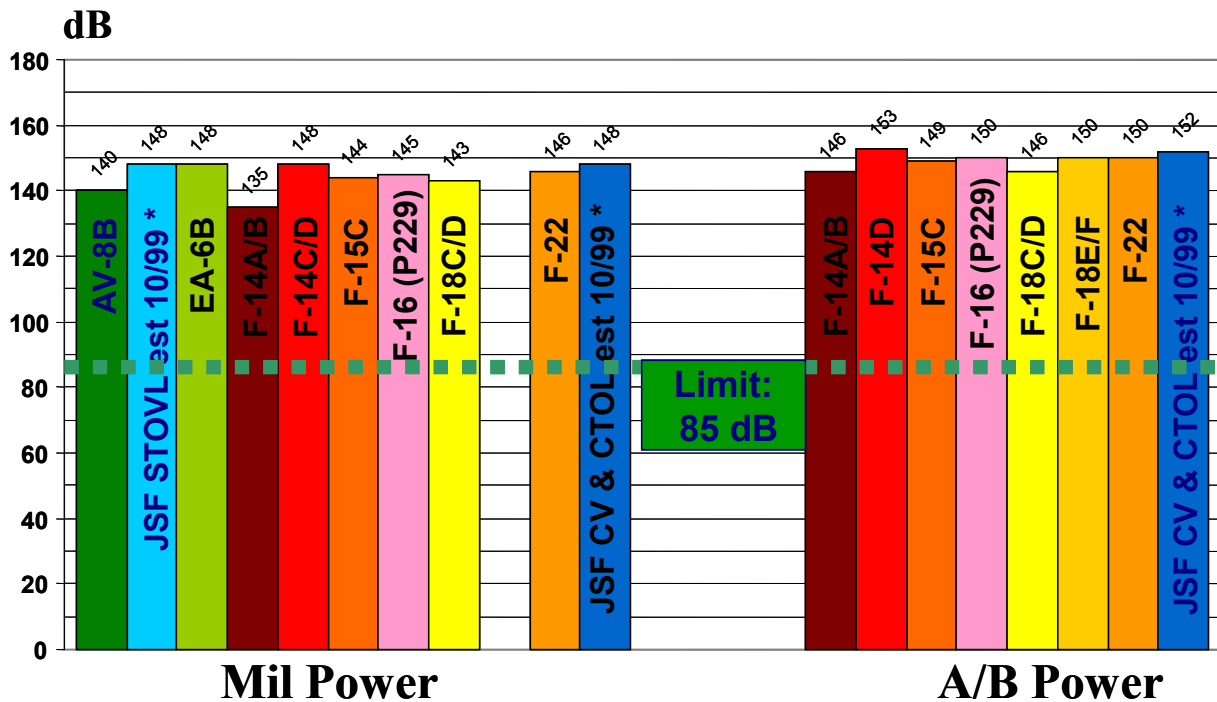


Figure 7: Worst Case Aircraft Noise Levels at 15 meters in dBA [4]

Individuals located in the immediate vicinity of these planes (ground crew, AC deck crew...) may be exposed to levels > 150 dB(A). In these conditions, the present hearing protection devices cannot afford enough attenuation [4]. Communication between the members of the crew is impossible. Because of the noise in the immediate vicinity of the plane *and* the noise inside the cockpit (≥ 120 dB in the new fighters) (see also figure 8), communication between the ground crew and the pilot may be very difficult if not impossible especially during specific takeoff and landing operations (JSF) [5].

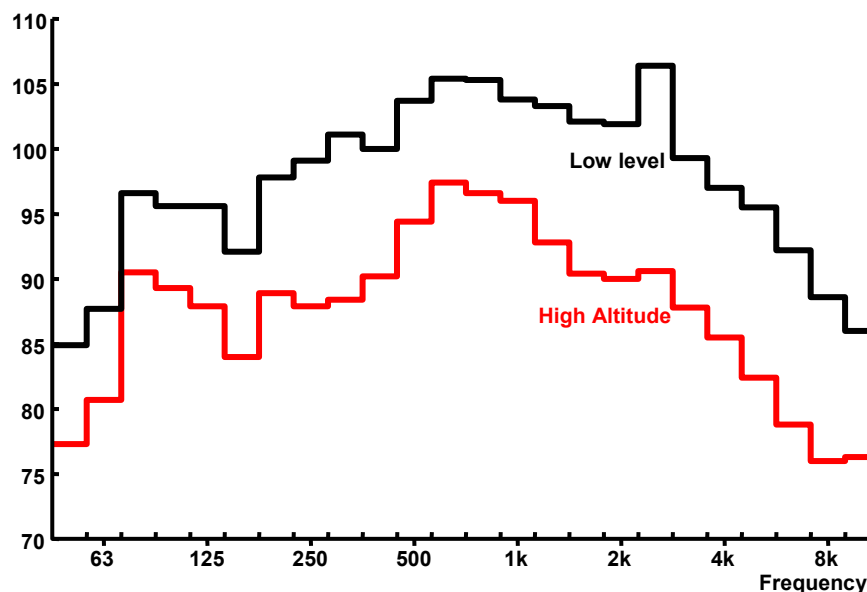


Figure 8: HARRIER cockpit noise (low level and high altitude flight) in dB SPL [5]

Noise: A Limiting Factor for the Use of Modern Weapon Systems?

Jet noise has also a large impact on the environment. The annoyance corresponding to the extended noise footprint of the new fighters could limit their normal training operations in densely populated areas (community noise) (figure 9).

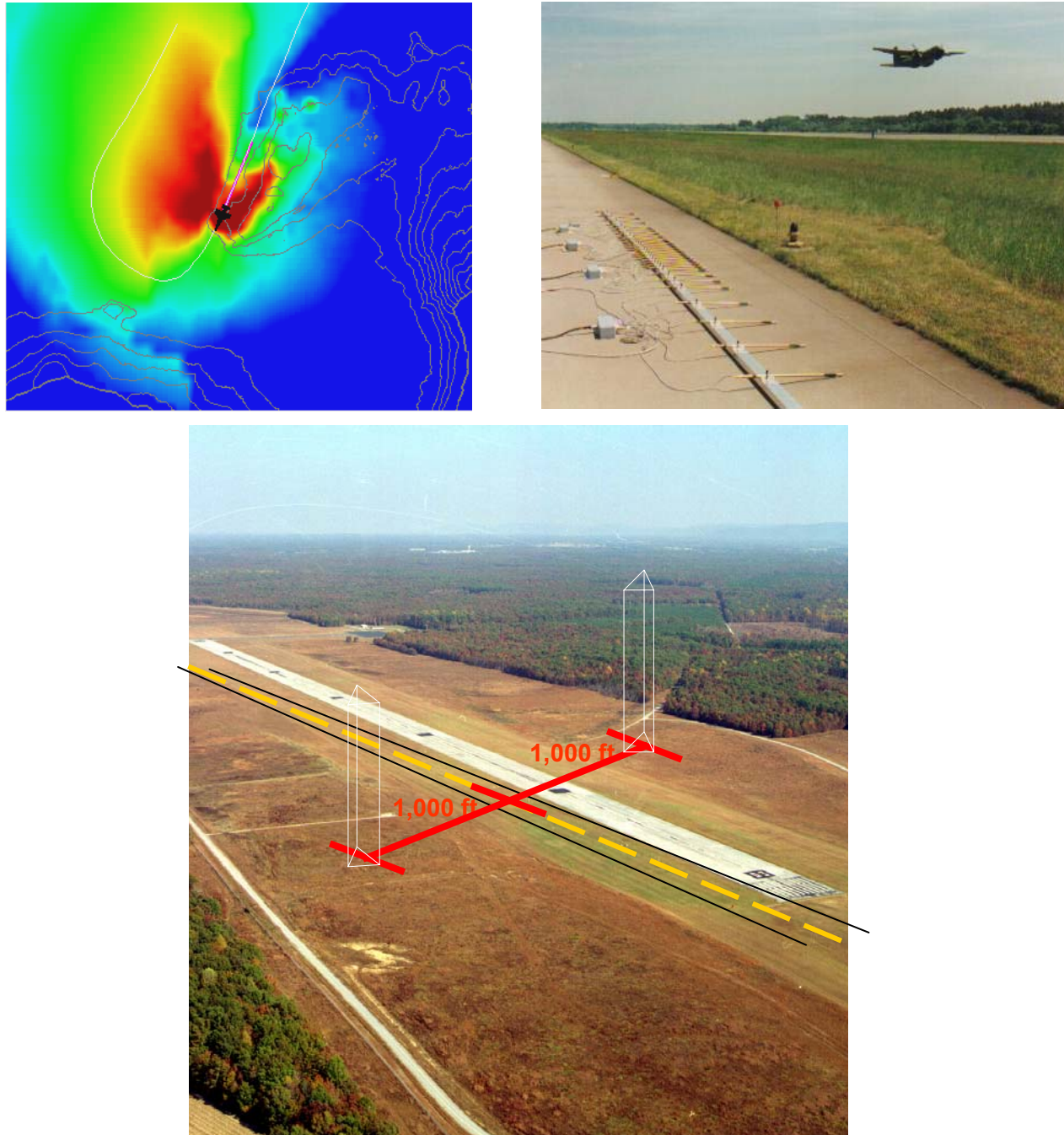


Figure 9: Noise footprint, present and future measurement installations of the US Air Force and FAA (dB Towers) [courtesy R. McKinley, AFRL/HECB Dayton]

3.2 Helicopter noise

The noise in the cabin of helicopters is made of low (rotor), medium (gearboxes) and high (jet engine) narrow band discrete tones, superimposed on a low level broadband background noise [5].

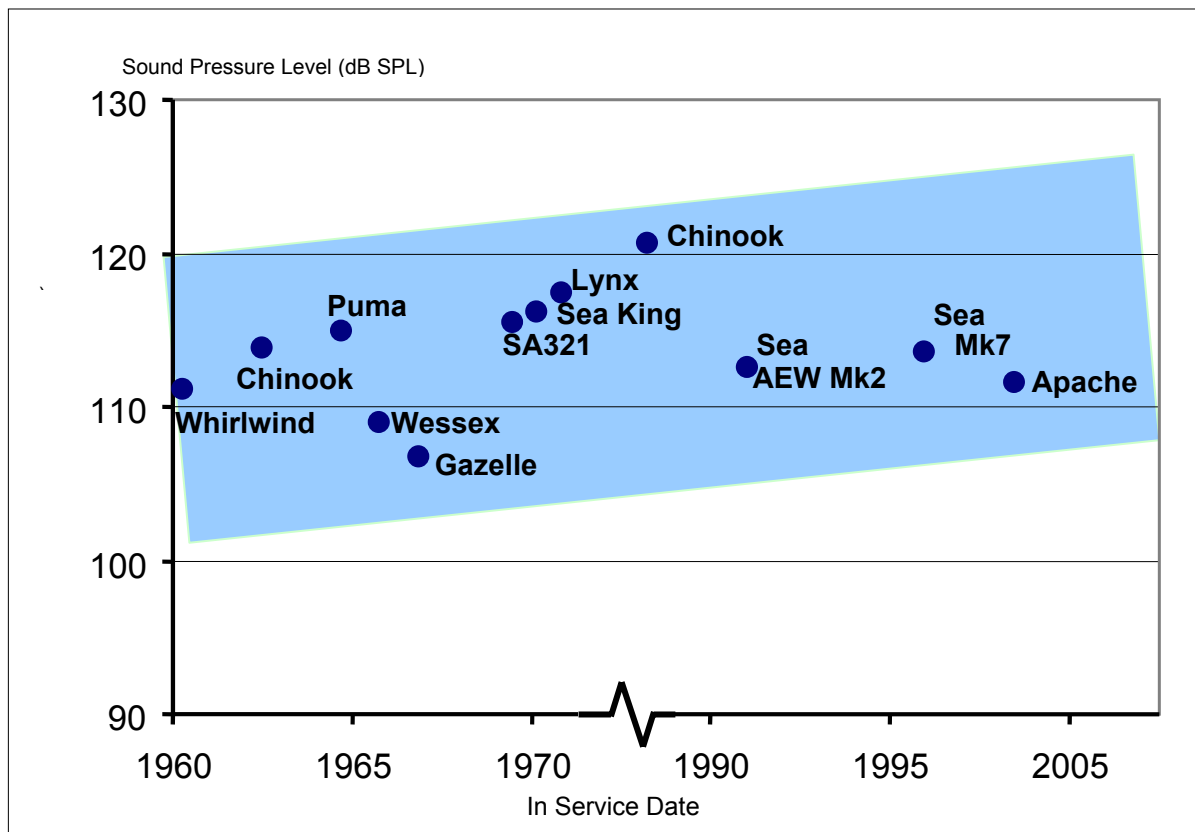


Figure 10: Noise levels in helicopters [5]

As the noise in a helicopter is rich in low frequency sound, the limited low-frequency attenuation characteristics of a helmet or headset will let through almost all of the low frequency noise [4,8]. These low frequencies will mask (nonlinear masking) the speech frequencies and impair the communication [7,8,9]. However, at the higher frequencies where the helicopter generates little noise and the helmet attenuation is at its maximum, the noise levels at the aircrews' ear will be low [5].

3.3 Land-vehicles noise

The A-weighted noise level at the positions of the crew members is reported for different land-vehicles in figure 11 (the dark colored bars span two standard deviations around the mean A-weighted sound level) [5,10]. Heavy tanks show the highest interior noise level (~120 dBA) and the spread within this category is small. The figure 12 indicates that the noise is maximum at low and very low frequencies (around 100 Hz) [5]. In these conditions and given the attenuation afforded by the passive hearing protectors there is a significant hazard for hearing and the communication is badly impaired: nonlinear masking of the speech frequencies by the very low ones [7,8,9]. Because of this masking, the crew members adjust the volume control of intercom and radio systems to very high settings (sometimes corresponding to speech-to-noise ratios in excess of 10 dB). When voice communication is frequent, a significant increase is made to the overall noise dose [10,11].

Noise: A Limiting Factor for the Use of Modern Weapon Systems?

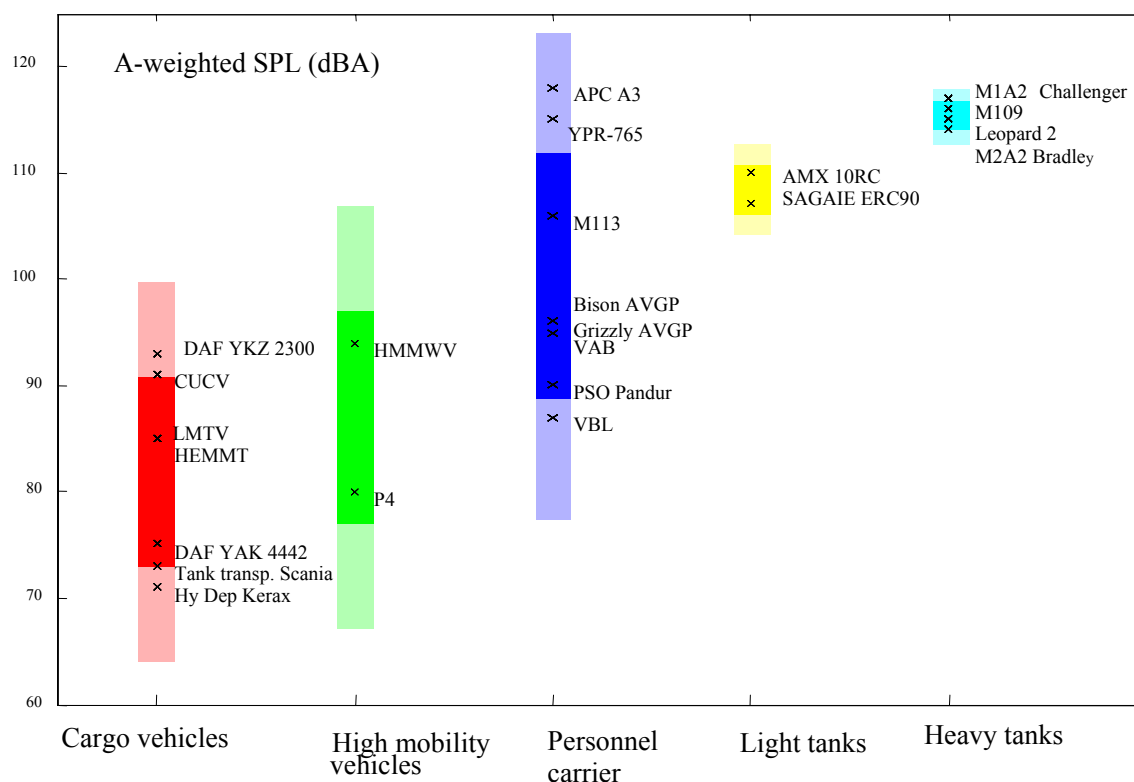


Figure 11: A-weighted interior noise level in land-vehicles [5,10]

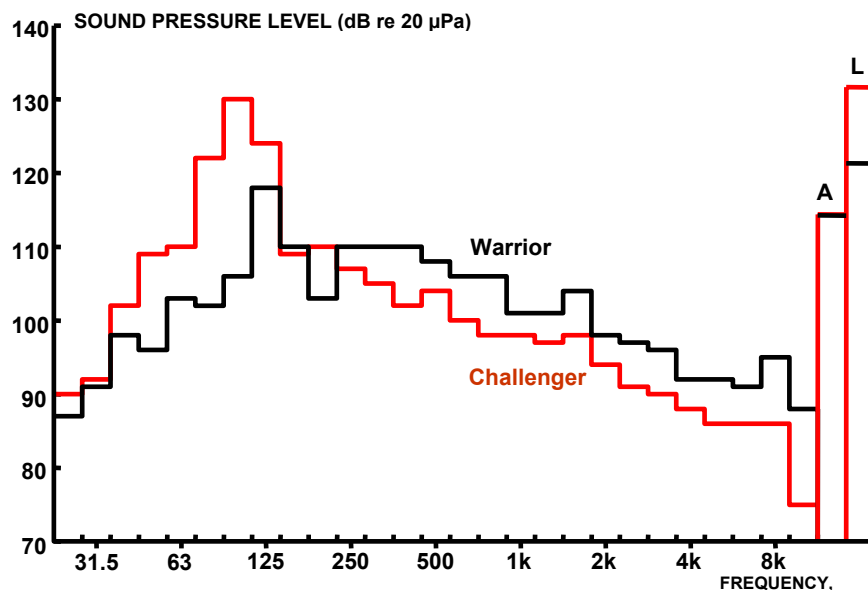


Figure 12: Noise levels inside heavy tanks [5]

3.4 Continuous noise exposure limit

The risk of hearing damage is correlated with the amount of the A-weighted acoustic energy received by the ear (isoenergy principle). An equivalent level of 85 dBA over 8 hours is generally considered as the exposure limit for unprotected ears. For every 3 dB that a LAeq exceeds the limit, the authorized exposure time must be halved: i.e., 4 hours for 88 dB, 2 hours for 91 dB..., 3 seconds for 124 dB! Therefore, in most conditions the exposure of unprotected ears to jet noise, helicopter noise and land-vehicle noise is unpractical and/or prohibited when the regulation is strictly enforced [12,13].

What should be the performance (the Insertion Loss: IL) of the hearing protection to allow a reasonable exposure duration to the noises that have been chosen as examples (§ 3.1, 3.2, 3.3)?

For the jet noise on an aircraft carrier we can estimate that a 10 minutes total exposure to 140-150 dB(A) (corresponding to about 30 launches and recoveries) is an absolute minimum requirement for the deck crew. To comply with the regulation, the hearing protection (HP) must afford an IL of 40-50 dB. No present HP (simple or double) can afford such an IL. If one refuses to break the law and if one does not accept large permanent Noise-Induced Hearing Loss to the personnel, the only solution is to develop new hearing protection devices [4,5].

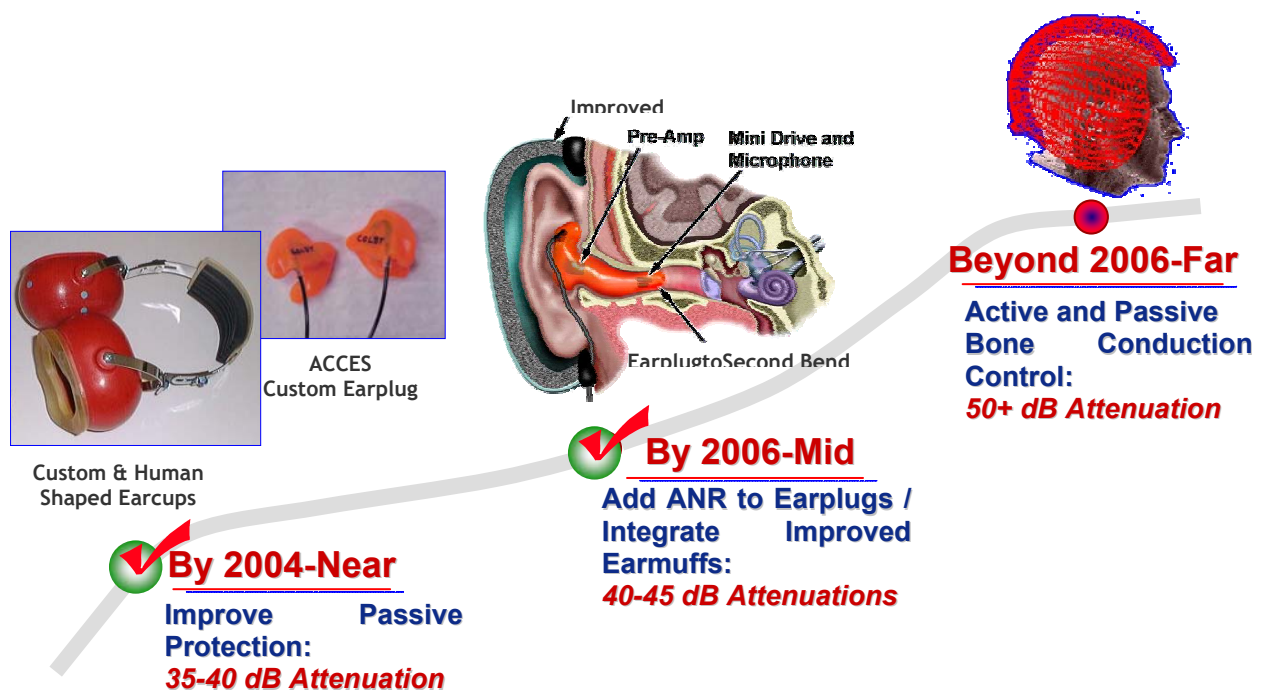


Figure 13: New Hearing Protection Technologies [courtesy: R. McKinley, AFRL/HECB Dayton]

According to [5], present hearing protection available for helicopters must be improved by about 5 dB to stay within legislative criteria. This could be achieved with the help of Active Noise Reduction techniques and/or double hearing protection (earplugs and earmuff).

In the land-vehicles (especially the heavy tanks) the present hearing protectors are generally unable to attenuate sufficiently the high-level noise (about 115 dB(A)) and to allow exposure durations in excess of a few tens of minutes. In these conditions, new (double?) hearing protectors making use of the active noise reduction techniques are necessary.

New hearing protectors will be more expensive and – probably – heavier, more cumbersome and less comfortable (especially if a double hearing protection is used). It will be necessary to ensure that the new hearing protection technologies are affordable, supportable, available and easy to use.

Last but not least, if the occupational noise exposure limit were lowered (i.e., 80 dB(A) instead of 85 dB(A) [13]) and if that new limit were enforced in the army, it is likely that in many situations no practical solution (hearing protection) could be found to comply with a new lower limit.

4.0 IMPULSE NOISE

When a round is fired a large volume of heated gas is released in the surrounding atmosphere. The rapid expansion of the gas initiates a pressure wave that takes the form of a shock wave (figure 14).

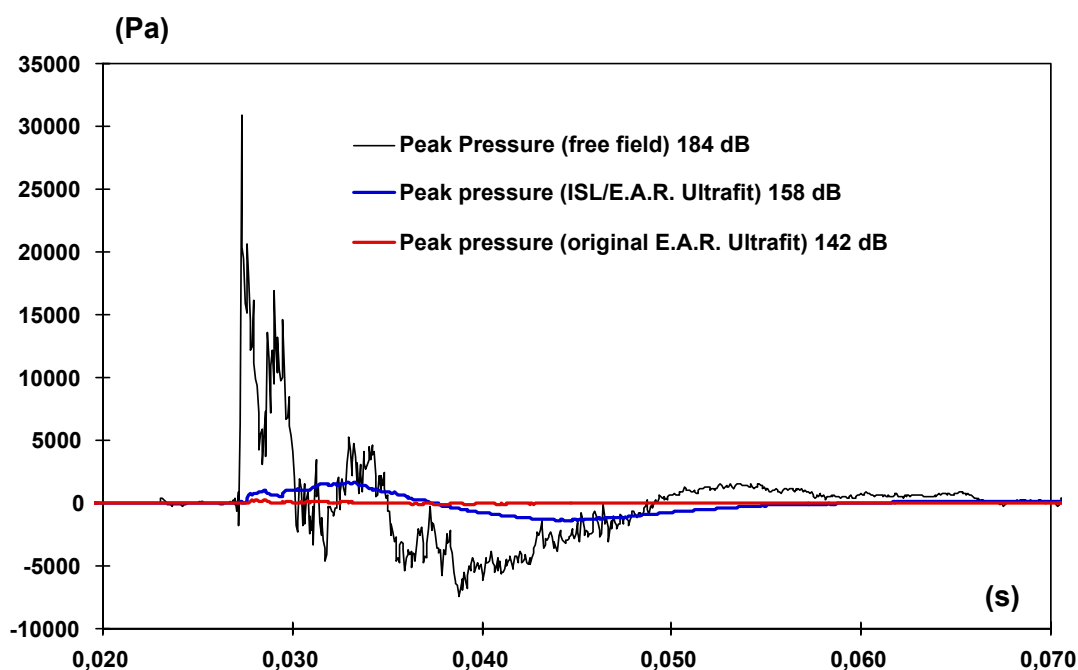


Figure 14: Pressure-time histories (120 mm mortar noise) recorded in free field and at the microphone of the ISL Artificial Test Fixture ear with a nonlinear earplug (ISL/E.A.R. Ultrafit) and a linear earplug (E.A.R. Ultrafit)

For a 120 mm mortar (top charge), the peak pressure at the loader's ear is 185 dB (figures 14 and 15). For a .50 caliber sniper's rifle, the peak pressure at the shooter's ear is 175 dB (figure 15).

4.1 Impulse noise exposure limit

To evaluate the hearing hazard due to weapon noises, a number of criteria have been proposed [14]. These criteria can be divided in three main categories:

- the first one (CHABA, 1968; Pfander criterion, 1980, 1994; MIL STD 1474B (M2), 1984; Smoorenburg criterion, 1982...) uses the peak pressure, the duration(s) (measured in the free field) and the number of the impulses, to evaluate the hazard. Among those, the criteria of Pfander and Smoorenburg (which are characterized by a line with a slope of -3dB/doubling of

either the duration and/or the number of impulses), are roughly in agreement with the iso-energy principle,

- the second one (Atherley and Martin, 1971; Martin, 1976; Dancer, 1982; DTAT, 1983) is based on the (A-weighted) iso-energy principle.
- the third one (Price and Kalb, 1991, 1992) is based on a physico-mathematical model of the auditory periphery. It aims to take into account the actual mechanics of the middle and of the inner ears (including the nonlinearities), up to the highest stimulation levels, and to calculate an index of hazard.



Figure 15: Left: 120 mm mortar, Right: .50 caliber sniper's rifle

These different criteria give different evaluations of the hearing hazard for unprotected ears (this is especially true for the noises of the large weapons). They also disagree on the predicted efficiency of the hearing protectors. No perfect Damage Risk Criterion presently exists (i.e., a DRC able to evaluate accurately the hazard in all exposure conditions: for impulse and continuous noises, for small and large weapons, for free field and reverberant exposures, for protected and unprotected ears...). However, thanks to numerous physical measurements, animal experiments and human observations performed by the members and the experts of the NATO RSG 29 [15] it can be shown that for impulse noises:

- the LAeq8 method with a limit at 85dB allows a limitation of the hearing hazard comparable to that aimed at by the other criteria,
- the LAeq8 method allows the assessment of the hazard for all kinds of weapon noises according to the well-recognized procedure used for occupational exposure (ISO 1999). It can be applied as well to impulses in free field and/or in reverberant conditions (either for small or for large caliber weapons),
- the LAeq8 method does not lead to an excessive overprotection and hence to an unjustified restriction of the use of the weapons as it is the case for most of the other criteria (especially with respect to the large weapon noises),

- the LAeq8 method allows to evaluate the hearing protection afforded by earplugs or earmuffs from classical Insertion-Loss data obtained by Real-Ear-At-Threshold or Acoustical-Test-Fixture methods in a more accurate and less conservative way than most of the other criteria.

This method has been evaluated on soldiers:

- 20 subjects equipped with AEARO foam earplugs are exposed to 20 howitzer (155 mm) rounds (175 dB peak pressure, A-duration: 8 ms, global LAeq8: 109 dB). The Insertion Loss afforded by the plugs is close to 30 dB (in these exposure conditions), therefore the subjects receive a noise dose corresponding to a LAeq8 of 79 dB. No significant TTS is observed.
- 16 subjects equipped with AEARO/ISL nonlinear earplugs are exposed to 7 mortar rounds (185 dB peak pressure, A-duration: 2.5 ms, global LAeq8: 110 dB). The Insertion Loss afforded by the nonlinear plugs is close to 30 dB (in these exposure conditions), therefore the subjects receive a noise dose corresponding to a LAeq8 of 80 dB. No significant TTS is observed (in spite of a peak pressure of 158 dB measured under the plug: figure 14).
- 14 subjects equipped with ISL nonlinear plugs are exposed to 6 shock waves (190 dB peak, A-duration: 1.5 ms, global LAeq8: 114 dB) (Albuquerque study). The Insertion Loss afforded by the nonlinear plugs is close to 30 dB (in these exposure conditions), therefore the subjects receive a noise dose corresponding to a LAeq8 of 84 dB. No significant TTS is observed in all but one subject.
- groups of 10 subjects equipped with an earmuff are exposed to 100 shock waves (187 dB peak pressure, A-duration 3 ms, 1 minute interval) (Albuquerque study). No significant TTS is observed (in spite of a peak pressure of 173 dB measured under the earmuff).

Then, we can conclude that the criterion based on the measurement of the A-weighted energy with a limit at 85 dB LAeq8 allows to assess satisfactorily the hazard corresponding to impulse noise *and* the actual efficiency of the hearing protectors. Consequently, at first sight impulse noise is not a limiting factor for the use of modern weapons.

However, the new European directive that will be enforced on February 2006 [13] prohibits the exposure to *a residual peak pressure higher than 137 dB(C) under the hearing protection*. This demand is scientifically and experimentally unfounded (see above) and originates (i) from a misreading of the actual acoustical, biomechanical and physiological phenomena related to hearing protection *versus* impulse/weapon noise and (ii) from an insufficient exchange of data between the military experts and the occupational law makers. If this new regulation were to be applied to the weapon noises, only light weapons could still be used *with a double hearing protection* (earplugs and earmuff). In all other cases (medium and heavy weapons), the residual peak pressure under the hearing protection will exceed 137 dB(C) whatever hearing protection is in use.

5.0 CONCLUSION

The noise of the modern weapon systems is a limiting factor for their use either because no present hearing protection is able to protect the ear and to avoid a large deterioration of the voice communication (continuous noise), or because unsuited regulation will make their use impossible (impulse noise).

REFERENCES

- [1] Dancer, A., "Hearing and Hearing Protection", NATO Lecture Series 244/HFM-111, 2004
- [2] Peters, L., Garinther, G., "The effects of speech intelligibility on crew performance in an M1A1 tank simulator", US Army Human Eng. Lab, Techn. Memorandum, 1990, 11
- [3] Ohlin, D., "Cost effectiveness of hearing conservation programs", NATO Lecture Series 219, RTO-EN-11, AC/323(HFM)TP/31, 2000
- [4] McKinley, R., "Passive hearing protection systems and their performance", NATO Lecture Series 244/HFM-111, 2004
- [5] James, S. and McKinley, R., "Applications overview of military noises, insertion loss, prediction of performance", NATO Lecture Series 244/HFM-111, 2004
- [6] Dancer, A., "Individual susceptibility to NIHL and new perspective in treatment of acute noise", NATO Lecture Series 219, RTO-EN-11, AC/323(HFM)TP/31, 2000
- [7] Buck, K., Wessling, T. and Dancer, A., "Adapting the STI for the use in very noisy environments", 7th Intl. Congress on Noise as a Public Health Problem, Sydney, 1998, 163
- [8] Buck, K., "Active hearing protection systems and their performance", NATO Lecture Series 244/HFM-111, 2004
- [9] Steeneken, H., "Assessment and standardization", NATO Lecture Series 244/HFM-111, 2004
- [10] van Wijngaarden, S. and James, S., "Protecting crew members against military vehicle noise", NATO Symposium, AVT-110, Prague, 4-8 October, 2004
- [11] Dancer, A., "Noise in the French Army Land Vehicles: consequences on Hearing Protection and Communication", NATO Symposium, AVT-110, Prague, 4-8 October, 2004
- [12] ISO 1999 2nd Edition (1990) "Acoustics: Determination of occupational noise exposure and estimation of noise-induced hearing impairment", Geneva, International Standards Organization
- [13] Physical Agents Directive 2003/10/EC, Official Journal of the European Communities, February 15, 2003
- [14] Dancer, A. and Franke R., "Hearing Hazard from Impulse Noise: a Comparative Study of Two Classical Criteria for Weapon Noises (Pfander Criterion and Smoorenburg Criterion) and the LAeq8 Method", Acta Acustica, 3, 539-547, 1995
- [15] NATO Research Study Group on the Effects of Impulse Noise, AC/243 (Panel 8/RSG 6) D/9, February 1987



A New Method for Rating Hazard from Intense Sounds: Implications for Hearing Protection, Speech Intelligibility, and Situation Awareness

G. Richard Price

AHAnalysis

125 Conestoga Street, POB 368

Charlestown, MD 21914

USA

AHAnalysis@Comcast.net

IMPLICATIONS OF A NEW METHOD FOR RATING HAZARD FROM INTENSE SOUNDS

The auditory hazard assessment algorithm for the human (AHAH), developed by the US Army Research Laboratory, is theoretically based and has been demonstrated to rate hazard from intense sounds much more accurately than existing methods. The AHAH model operates in a PC level computer and analyses hazard in the time domain. In addition to reproducing the conductive path from air to the cochlea accurately it includes a non-linear stapes (clips large displacements) and an active middle ear muscle system. It is being written into a new US Army MIL-STD-1474(e) and is used by the Society of Automotive Engineers for the analysis of airbag noise hazard. The model shows that low-frequency energy at high levels can act to reduce the flow of energy into the inner ear, reducing the hazard. Traditional analyses tend to overrate the hazard from large calibre weapons impulses. The model also shows that impulses with little low-frequency energy, e.g. rifles, may be underrated in hazard by traditional methods. Hearing protective devices (HPDs), in order to be effective for gunfire type impulses, were shown to need most attenuation in the mid-range and less at lower frequencies, much like the attenuation curve for the non-linear combat arms plug. At the same time, speech intelligibility with such an attenuator could be much better than for an HPD with good low-frequency attenuation. Future developments of the model will include expansion to cover a wider range of intensities and an adaptive middle ear muscle system.

1.0 INTRODUCTION

Intense noise fields around weapon systems pose a variety of critical problems in which the requirements for protection, communication and effective weapon systems intertwine in complex ways. With the continuing emphasis on high performance and light weight, the noise fields in and of themselves have become a hazard from which the ear and even the body may need to be protected. Recent statistics from the battlefield suggest that even with considerable emphasis on hearing conservation, hearing loss nevertheless ranks fourth in the list of casualty-producers. This fact makes two points. First, protecting the ear through improved equipment design and adequate hearing protection are not options, but are requirements in the modern military. And second, the fact that hearing loss is a casualty-producer demonstrates that the sense of hearing is now understood to be essential to adequate soldier performance in the military. It is needed both for the purpose of communication as well as for providing situation awareness. Thus the sense of hearing must be protected and maintained in as near a normal state of acuity as possible.

Price, G.R. (2005) A New Method for Rating Hazard from Intense Sounds: Implications for Hearing Protection, Speech Intelligibility, and Situation Awareness. In *New Directions for Improving Audio Effectiveness* (pp. KN2-1 – KN2-24). Meeting Proceedings RTO-MP-HFM-123, Keynote 2. Neuilly-sur-Seine, France: RTO. Available from: <http://www.rto.nato.int/abstracts.apis>.

However, when we protect the ear, we often inadvertently create an auditory deficit by reducing sensitivity to sound as well as by interfering with the auditory system's ability to localize noise sources. In this case the cure is remarkably similar to the disease!

If, on the other hand, we seek to reduce the noise at the source, then a second set of dilemmas present themselves. Hazard reduction at the source is the ideal answer; however, it typically involves reducing the energy in the source or providing shielding from it. The first alternative may make the system less powerful (and therefore exposes troops to greater risk) while shielding may add unacceptable weight.

Dealing effectively with these problems requires a good theoretical grasp of the parameters involved so that intelligent compromises can be made that will optimise system performance. In this talk we will focus on the new perspective brought to these issues by a new method of rating hazard from intense sounds that has been developed by the US Army Research Laboratory. The Auditory Hazard Assessment Algorithm for the Human (AHAAH) is essentially an electro-acoustic analog of the ear written as a computer program. It is presently being written into what will become MIL-STD-1474 (e) and has already been accepted by the Society of Automotive Engineers as the procedure for calculating the hazard from airbag noise exposures (SAE, 2003). The program and its supporting documents are available for download at www.arl.army.mil/ARL-Directorates/HRED/AHAAH/.

Central to dealing with the optimisation of system performance in the military is the ability to evaluate the hazard associated with intense sounds and the protection provided by HPDs. This presentation will introduce the AHAAH model, describe its validation and explore some of its implications.

In his keynote address at this meeting, Dr. Dancer has already pointed out that noise issues limit weapon design. Overestimation of hazard, often done on the pretext that it is "conservative", nevertheless leads to inferior weapon design, which in turn results in casualties and jeopardizes mission accomplishment. Present criteria tend to err in this direction. Likewise, underestimation of hazard leads to hearing loss in troops, which also creates casualties and jeopardizes mission accomplishment. The goal in predicting hazard must be accuracy – errors in either direction must be avoided.

2.0 THE AHAH MODEL

2.1 The Interpretive Challenge

Currently, the world's standards rate hazard based on some measure of pressure and duration or alternatively by measuring the A-weighted energy in the exposure. Today there is general agreement that these standards for intense noise exposures (140 dBP and above) work very poorly. In addition to being inaccurate, they lack a theoretical base, relying on correlations between various noise measures and incipient hearing loss in human subjects. Therefore, without a solid theoretical understanding, when impulses with unusual pressure histories arise, there is little certainty that the existing methods of analysis will fit. This situation has produced paradoxical data. Consider that a rifle impulse at the firer's ear contains a little over 1 J/m² of A-weighted energy. Data from soldiers firing the FNC rifle (Brinkmann, 2000) suggest that an unprotected exposure of 5 rounds fired from one's own weapon would be just on the verge of producing permanent hearing loss in the 95%ile ear. This agrees well with the French use of A-weighted energy, which would allow an 8.7 J/m² exposure (Dancer, 2000). At the same time, data from the US Army's Albuquerque studies (Johnson, 1998; 1994) have demonstrated that an exposure to 2000-3000 J/m² from a simulated cannon impulse, (also A-weighted) at the ear canal entrance, produced no hearing loss. Thus, on the basis of A-weighted energy

somewhere between 5 and 3000 rifle shots would be tolerable. It is clear that at these levels A-weighted energy rates hazard very poorly.

Or consider the interpretive problem associated with the alternative schemes used in Germany (Pfander, 1975), The Netherlands (Smootenburg, 1980), the UK (Ministry of Defence, 1982) and the US (MIL-STD-1474(d), 1997). In these criteria differing measures of peak pressure and duration are used to enter a hazard analysis diagram. The data discussed in the previous paragraph appear in Fig. 1 where we see pressure histories of the two impulses. The rifle impulse was recorded in the free field at the firer's ear location and the simulated cannon impulse was recorded at the ear canal entrance under an earmuff. In all the criteria, higher pressures and longer durations are rated as more hazardous. The problem is that 6 of the rifle impulses are hazardous and 100 of the simulated cannon impulses are in fact safe. It is difficult to see how any method of rating hazard that depends on a measure of peak pressure and duration could possibly hope to reconcile these data.

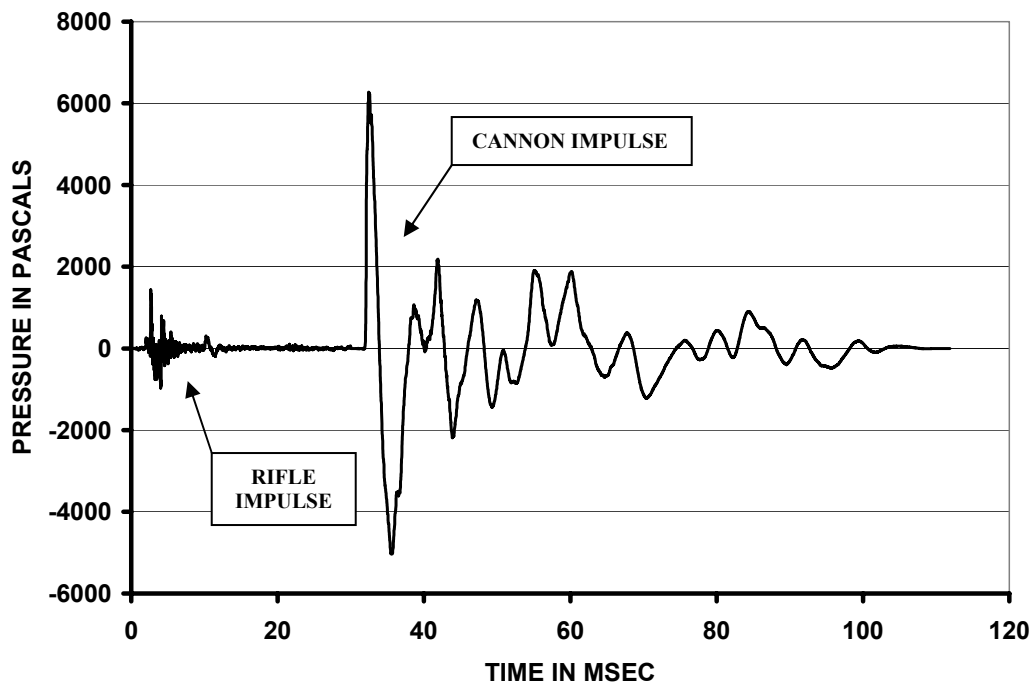


Figure 1. Pressure history of a rifle impulse followed by a cannon impulse.

2.2 Approach to Solving the Problem

Rather than seek yet another acoustic metric that might correlate with hearing loss, we approached the problem by trying to establish a theoretically based understanding of how the ear operates at high intensities. Once those processes were understood, then the appropriate metric(s) became apparent. In the end, the basis for hearing loss at high intensities can be understood initially to be mechanical disruption of the tip links within the hair cells in the organ of Corti, as illustrated in Fig. 2. If the amplitudes continue to rise, the extent

of the mechanical disruption can, of course, rise to the destruction of the organ of Corti itself. At the level of the hair cell, the basic loss process is taken to be mechanical stress modelled as a fatigue of materials, i.e. we keep track of the number of flexes of the basilar membrane and their amplitudes (in microns, upward flexes only), square them, and accumulate the sum at 23 locations within the cochlea (roughly 1/3 octave intervals). The result of this calculation is called the Auditory Risk Unit (ARU). The model does not actually calculate the hair cell movements; but takes the simpler option of calculating basilar membrane displacements, which are the forcing function for the hair cell displacements.

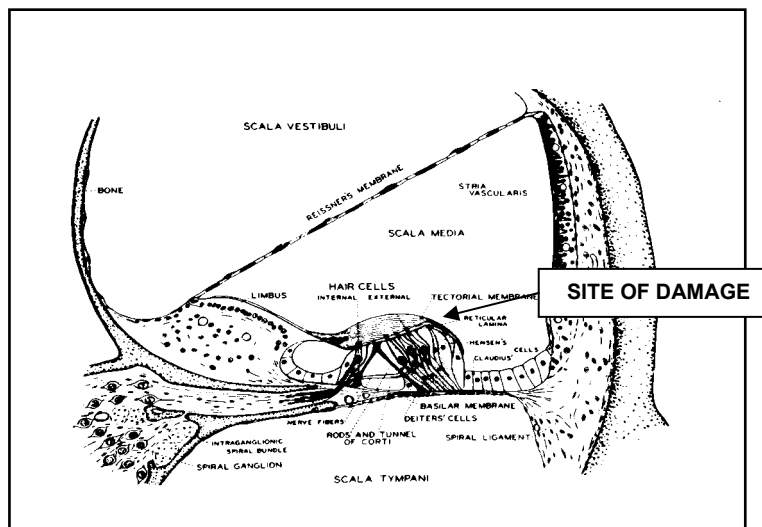


Figure 2. Cross section of the organ of Corti showing likely site of initial damage

The real complexity for predictive purposes lies in the multiple non-linearities that are part of the ear's basic physiology. It is well known that the ear is differentially sensitive to frequency, conducting sound best in the mid-range of frequencies, cutting off at the low frequencies because it is too stiff and at the high frequencies because it is too massive. The ear is also equipped with middle ear muscles that contract and stiffen the middle ear, changing its conductivity up to 20 dB or more, depending on frequency. Their behavior adds a non-linearity that changes with time (as a result of the dynamics of the muscle contraction) and differentially as a function of frequency (low frequencies are affected more than the higher frequencies). And lastly, at very high levels the middle ear itself becomes non-linear in that it is incapable of transmitting displacements larger than 20 microns or so due to the constraint imposed by the annular ligament of the stapes. This element is extremely important in explaining the ear's ability to withstand very high-energy exposures. In essence, at high levels the middle ear becomes a powerful peak-clipping device.

2.3 Creation of the AHAAH Model

Keeping track of the interactions of the various non-linearities is a task ideally suited for a mathematical model of the ear. Various elements of the ear's structure had been modelled by different researchers in the past; but the elements lacked the integrative "glue" that would enable a pressure in the free field to be propagated via head-related transfer functions, through the external and middle ears, into the cochlea and

down the basilar membrane. The non-linear dynamics of the middle ear also needed to be included in the model as well as the damage model in the inner ear. We opted to do an electro-acoustic model of the ear, maintaining conformality with the structure of the ear. The model could be simpler; but one of the functions of modelling is insight, and keeping the model's structure like that of the ear promotes engineering insight. The result was first a model specific to the cat ear that reproduced the existing data for the conductive path from the free field to the cochlea. This model was then tested with damaging stimulation in biological ears. In the end the correlation between the model's output and the average hearing loss measured for 12 differing exposures was 0.94, which meant that it was predicting very successfully. Once that model was validated, it was turned into the human version by adjusting the electro-acoustic elements to match human dimensions and values. The variables were then fine-tuned to assure that the model reproduced the transfer function data that had been reported for the human ear. Thus the AHAH model for the human ear was developed essentially without reference to human hearing loss data. At the time (1996) it was assumed that it would be challenged with data and adjusted as necessary to achieve a fit. So far no adjustment has been made. The process of the development of the model took place during the tenure of two NATO RSGs (RSG-6 and RSG-29) on impulse noise. The details of the development and the codes were shared with these groups.

2.4 The Model's Features

Operationally, the model is user-friendly. It runs in near real-time on a PC level computer and uses WINDOWS conventions. The software allows for the importation, editing, and analysis of waveforms. It also includes an analytical feature in the form of a movie that shows basilar membrane displacements during the analytic epoch and indicates what portion of the waveform produced them.

The movie has proven to be an essential element of the model. Given that the damage mechanism is mechanical stress within the cochlea, it follows that the process needs to be followed in the time domain, i.e. specific timing of individual oscillations in the pressure waveform make a difference in the transmission through the middle ear (the stapes limits large displacements) and the resulting damaging effect within the cochlea. This point is discussed in section 3.2 of this paper. The argument that the ear needs to be evaluated in the time domain is a theoretical position that differs from the arguments for the use of a frequency domain measure, e.g. A-weighted energy. Such a statistic simply ignores the pattern in which simulation arrives, taking the position that only the total energy in the analytic epoch need be accounted for.

In order to be useful as a DRC, it was necessary to find some method of including susceptibility of the ear in the calculation of hazard. In order to do this as simply as possible, we accepted the idea that a susceptible ear is like a normal ear being driven harder. If we assume that susceptibility is normally distributed, like many auditory values, with a standard deviation of 6 dB, then it follows from simple statistical considerations that to simulate the 95%ile ear, the sound pressure level should be raised 10 dB (1.64 SDs). The same logic allows the calculation of hazard for any other level of susceptibility.

2.5 Prediction of Hazard with the AHAH Model

The model calculates hazard in Auditory Risk Units (ARUs), defined earlier, which have been related to hearing loss in the cat model. The formula relating ARUs to threshold shift is:

$$CTS = 26.6 \times \ln(ARU) - 140.1$$

Where: CTS is Compound Threshold Shift measured within ½ hour

Thus an exposure totalling 500 ARUs would be expected to produce a CTS of 25 dB. One of the determinations of the Albuquerque studies (and the general agreement the NATO RSG 029, [2003]) was that a 25 dB threshold shift at (any frequency) should be considered the limit of tolerable exposure; hence, 500 ARUs could also be considered a just tolerable exposure.

The ultimate test of the model is whether or not it predicts hearing loss for the human ear. In order for an exposure to be interpretable by the model, a digitised waveform for the exposure must be available and there should be measures of hearing sensitivity before and after the exposure. With those requirements there are somewhat more than 70 experiments with interpretable human data, the largest body of which is the US Army's Albuquerque studies. In those tests, groups of 60 subjects wearing hearing protection were exposed to explosive impulses intended to mimic the free field response of Army weapons (Johnson, 1994; 1998). The model's prediction was calculated from the waveform measured in front of the ear canal entrance. To establish the validity of the model's response, the question was asked regarding whether or not the model's prediction could be rejected as inaccurate, i.e. if the model said the exposure were safe and 6 Ss showed unacceptable threshold shifts, then the model's prediction was considered inaccurate. Or if the model said the exposure was hazardous and no subject showed a 25 dB threshold shift, then the model's prediction would be rated as inaccurate for that condition. For comparison purposes, the Albuquerque data were also evaluated in the same manner with the US Army's MIL-STD-1474(D) and the A-weighted energy measure (Price, 2003).

The evaluation diagram used appears in Table 1 which shows the evaluation of the Albuquerque data based on MIL-STD-1474(D). In this diagram the cells contain codes that identify the specific exposures. Entries in the upper left and lower right quadrants represent accurate calls, while entries in the lower left and upper right quadrants represent either over – or under-estimation of hazard respectively. In the case of MIL-STD-1474(d) it is apparent that for these impulses it tends to err in the direction of over-predicting hazard, and in no case did it identify an exposure as safe that was in fact hazardous. Its prediction was correct in 19 of the 53 experiments (36% correct).

In Table 2 we see the result for the evaluation using the A-weighted energy criterion, i.e. exposure to $L_{Aeq8} > 85$ dB is considered hazardous. It is apparent that A-weighted energy, even more than MIL-STD 1474(D) tends to over-predict hazard for these impulses. It was correct on 13 of the 53 cases (25% correct).

In Table 3 we see the evaluation as done by the AHAAH model. It is apparent that it has been largely successful in rating the hazard. In no case was a hazardous impulse called safe and it erred in over-estimating the hazard in three instances. Overall, the model was correct in 50 of the 53 cases (94% accuracy).

Approximately 20 other exposures to small arms, spark gap discharges, shoulder-fired rockets, etc. could be evaluated with the model. They were gleaned from the literature of the last 40 years and while it was not possible to be as rigorous in the application of the analyses, they nevertheless provided interesting tests and an additional “reality check” on the efficacy of the model in dealing with a wide range of conditions. In all those cases the model's prediction was correct, giving it an over-all accuracy for both protected and unprotected exposures of better than 95%.

EVALUATION BY MIL STD-1474												
		OUTCOME										
		SAFE						HAZARDOUS				
PREDICTION	SAFE	01	T1	T2								
		F1	F2	F3								
HAZARDOUS		G1	G2	G3								
		R1	R2									
		O2	O3	O4	O5	O5	O7	O8	O9	OF	OH	
		T3	T4	T5	T6	T7	T8		T9	TF	TH	
		F4	F5	F6	F7	F8	F9		FF	FH		
		G4	G5	G6	G7	G8	G9	GF				
		GH	R3	R4	R5	R6	R7	R8				
		R9										

Table 1. Evaluation of the Albuquerque dataset with MIL-STD-1474(d).

EVALUATION BASED ON A-WEIGHTED ENERGY												
		OUTCOME										
		SAFE								HAZARDOUS		
PREDICTION	SAFE	G1 R1										
	HAZARDOUS	O1 T1 F1 R2 G4	O2 T2 F2 R3 G5	O3 T3 F3 R4 G6	O4 T4 F4 R5 G7	O5 T5 F5 R6 G8	O6 T6 F6 R7 G9	O7 T7 F7 R8 GF	O8 T8 F8 R9 GH	O9 T9 FF	OF TF FH	OF TH F9

Table 2. Evaluation of the Albuquerque dataset with the A-weighted energy criterion

EVALUATION BY AHAAH									
PREDICTION	SAFE	OUTCOME							
		SAFE				HAZARDOUS			
		O1	O2	O3	O4	O5	O6	O7	O8
		T1	T2	T3	T4	T5	T6	T7	T8
		F1	F2	F3	F4	F5	F6	F7	F8
		G1	G2	G3	G4	G5	G6	G7	G8
		G9							
		R1	R2	R3	R4	R5	R6	R7	R8
HAZARDOUS									

Table 3. Evaluation of the Albuquerque dataset with the AHAAH model

2.6 Acceptance and Adoption of the Model.

The AHAAH model is gaining acceptance as a method of rating hazard. It has been successfully peer reviewed by the American Institute of Biological Sciences (2001), is being used by the Society of Automotive Engineers as the basis for rating hazard from airbags (SAE, 2003), is in a draft technical report of the American National Standards Institute (ANSI, 2005), is being proposed as the basis for rating hazard in the US Army's MIL-STD-1474(e), and is being proposed to the US Army's Surgeon General as the basis for a damage-risk criterion for the Army.

3.0 IMPLICATIONS OF THE AHAAH MODEL

3.1 Range of Applicability of the AHAAH Model

The model was designed to address virtually any intense sound, regardless of its origin or its physical characteristics. It is true that the specific focus during validation studies was on gunfire level sounds (140 dBP and higher) where the loss mechanism within the cochlea is clearly mechanical stress. The human

exposure data used in validation have ranged from about 155 dBP to 185 dBP. For those levels at least, we have seen that the model fits the data very well. However, there is still a question as to how far downward in level the method will work. In that regard, it is interesting to note that at lower levels pressure specifications are typically in rms values. Waveforms with rms pressures in the 125-130 dB region might easily have peak pressures over 140 dB. Such pressures (and higher) can be found near jet engines on carrier decks or in maintenance settings where they are being test-run.

3.2 New Analytic Insights from Time-Domain Analysis.

It is often the case in modern science that technical insights are linked to the development of analytical tools that provide new ways to examine phenomena. The AHAH model provides such an opportunity. It keeps track of displacements in the middle and inner ear, arguing that in order to understand damage at the level of the hair cell, the exact pattern of instantaneous displacements must be known. In other words, it operates in the time domain. It also makes a movie of the analysis interval that shows displacements within the ear and the hazard attributable to them. Measures such as A-weighted energy lose track of instantaneous displacements and follow only the total energy in the epoch being analysed distributed across frequencies, i.e. it operates in the frequency domain. For some purposes the two methods can provide similar answers; but in other cases they do not. Consider the case of the pressure history from an airbag deployment shown in Fig. 3. The AHAH model calculates 1342 ARUs (clearly hazardous) with 94 J/m² of A-weighted energy (also considered a hazardous exposure). On the other hand, the AHAH model shows through its movie feature that most of the calculated hazard occurs at the moment pointed to by the black arrow (about 10 msec into the waveform). If that dip in the pressure history were eliminated, the hazard would drop to 435 ARUs (safe exposure); but the A-weighted energy would still be 74 J/m², which would still be rated as hazardous. The dip in pressure is a function of a bounce the bag makes as it deploys. Were some damping mechanism designed into the bag to prevent the rebounding bounce, then the deployment would be safe. The answer to this noise hazard would be an engineering change unrelated to the sound of the bag filling. Not every noise problem lends itself to this innovative kind of resolution. However, we can now observe that a new analytical tool, the AHAH model, is now available for use.

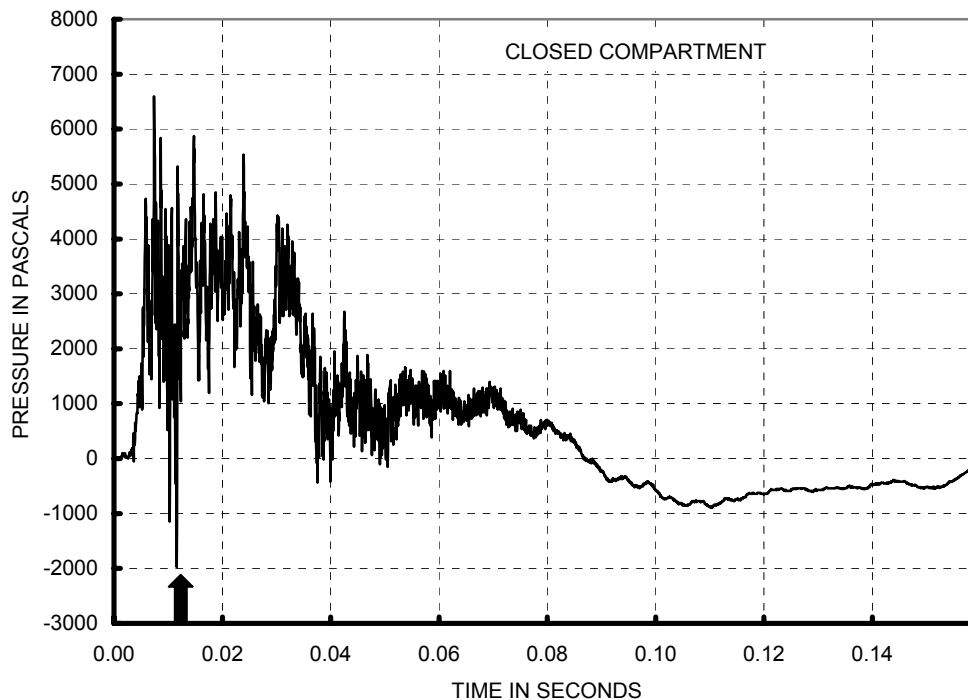


Figure 3. Pressure history of airbag waveform at the driver's left ear in a closed compartment

3.3 A New View of the Influence of Low-Frequency Energy

3.3.1 Impulses with a Lot of Low-Frequency Energy

It would be fair to say that today A-weighted energy is generally regarded in the hearing conservation world as the measure of choice for rating hazard, at least for SPLs below 160 dB. It reflects the transfer function of the ear based on the (psychological) loudness of a sound at 40 dB. It also has the additional advantage that, now that meters have been built to include its characteristics, it is easily measured. However, for many years, there has been the suspicion that for high-level noises it may include too much low-frequency energy and other weighting systems have been proposed (Buck, Dancer and Parmentier, 2003). In Fig. 4 we see the transfer function for the ear for free-field pressure to stapes volume velocity as calculated by the AHAH model. One curve depicts the transmission in a normal ear and the second curve depicts transmission with the middle ear muscles contracted. For comparative purposes the A-weighting curve has also been plotted. It is apparent that the A-weighting curve does indeed cut off slower than the transmission curve, especially if the middle ear muscles are contracted. On this basis it might be expected that the AHAH model will rate lower frequencies as less hazardous than an A-weighted measure, consistent with the transmission of the human ear.

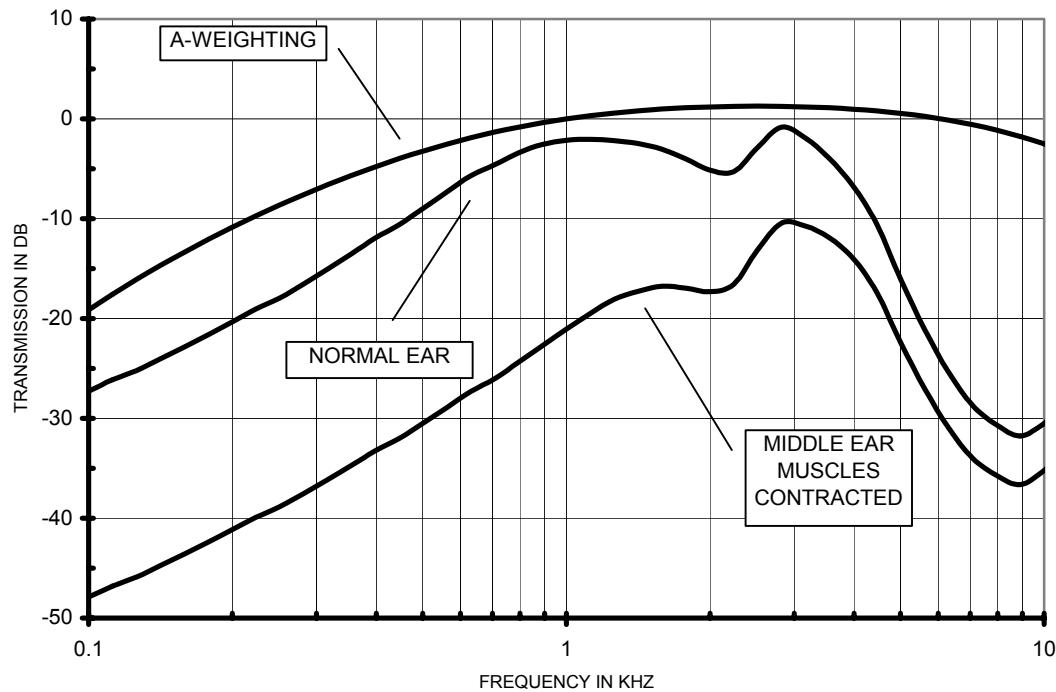


Figure 4. Transmission of the human ear with no muscle contraction and with a middle ear muscle contraction compared with the A-weighting function

If the only problem with A-weighting were that its low-frequency cut-off slope needed to be changed, then the answer to the problem would be simple. However, like many processes that take place at high intensities, the problem is that other things are happening simultaneously. Namely, the middle ear, which has an absolutely remarkable linear range (better than 120 dB), becomes non-linear at very high levels. Displacements of the stapes are constrained by the annular ligament so that it can't move more than 20 microns or so. And this means that waveforms at high levels are severely peak clipped as they enter the cochlea, which is where the basic damage takes place. This effect can be seen in Fig. 5 for an impulse from a 105 mm Howitzer. In the figure, taken from a display in the AHAH program, the upper panel is the calculated stapes displacement for the pressure history in the lower panel. In this illustration, a "warned" exposure has been assumed (middle ear muscles are contracted). It is apparent that the stapes displacement resembles the pressure history; but with an important difference. Because of the compression by the stapes, the initial peaks in the pressure history are relatively much higher than the same peaks in the stapes displacement and conversely, the smaller oscillations in the latter part of the pressure history form a much larger part of the waveform entering the cochlea. By using the movie feature of the model, we can determine that $\frac{1}{2}$ of the hazard has accrued by 17.5 msec into the waveform. That portion of the waveform represents 86% of the A-weighted energy in the impulse. It follows that 14% of the energy (the latter portion of the wave) also contains $\frac{1}{2}$ of the hazard. These effects are largely the result of the peak-clipping action of the middle ear. It can only be seen as ironic that the portion of the waveform we have traditionally focused on in rating hazard, the initial high peak pressure, contained 54% of the A-weighted energy but was responsible for only 3% of the hazard. Put another way, if the initial peak (3 msec of the waveform) were totally eliminated, the hazard in the remainder of the waveform would have

changed only from 673 to 651 ARUs for the warned ear. If the ear were unwarned we get the surprising result that the hazard would have risen 16% from 2093 to 2422 ARUs because of the delay in starting a middle ear muscle contraction.

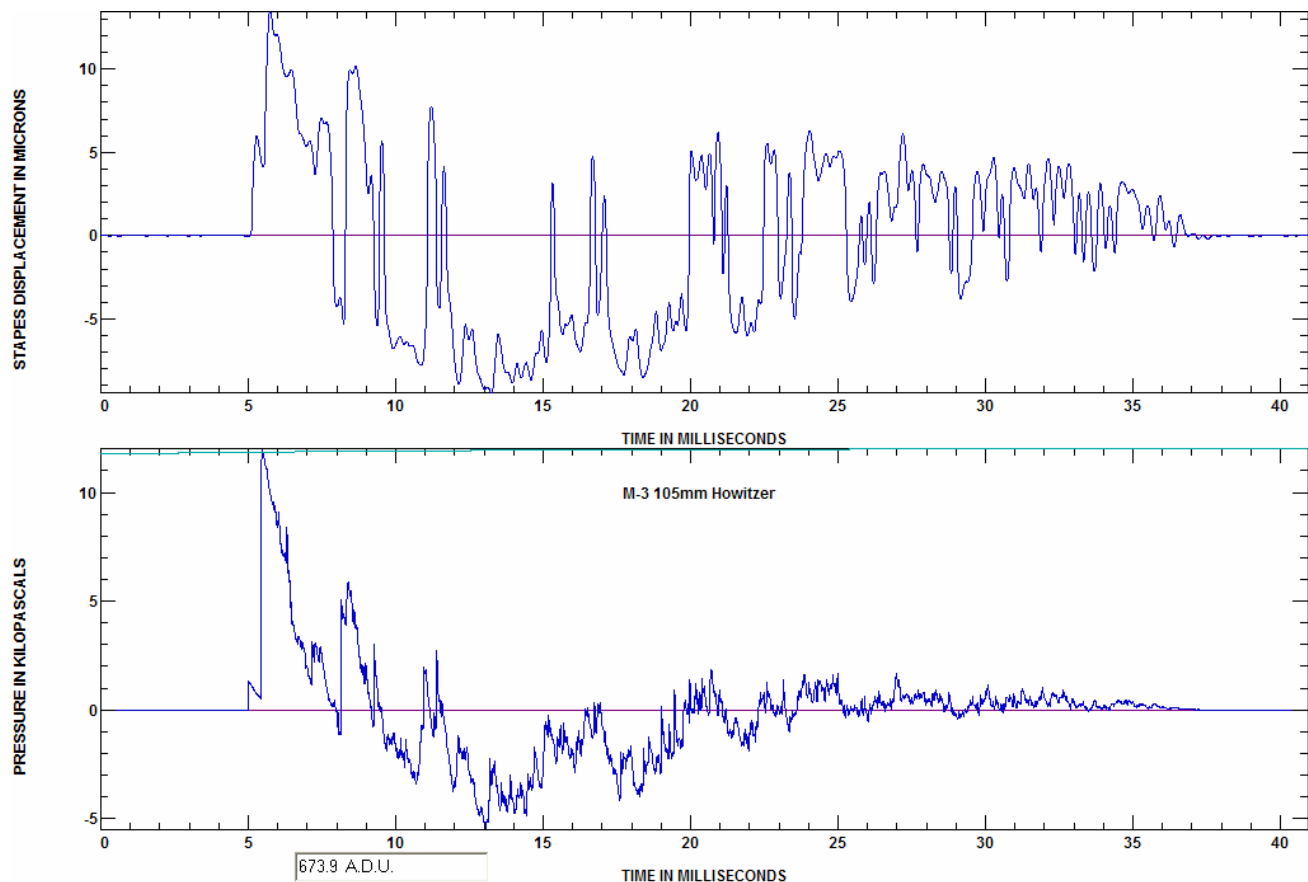


Figure 5. AHAH model's calculation of the stapes displacement (upper panel) in response to the free field pressure from a 105 mm Howitzer. Middle ear muscles are contracted during the episode.

Furthermore, the presence of such a non-linearity means that proper evaluation of its effect must be made in the time domain, rather than the frequency domain, i.e. just exactly when peaks and dips occur in the waveform matters. Oscillations that occur during periods of high displacement (during clipping) are not transmitted and oscillations that occur during moments when the stapes displacements are in their linear range are transmitted. Thus, lower frequency energy has the power to modulate the flow of higher frequency energy. This is the best explanation of why the AHAH model was successful in predicting hazard for the Albuquerque data and the other methods were not.

3.3.2 Impulse with Less Low-Frequency Energy

It might appear at this point that the AHAH model supports the idea that gunfire impulses are simply much less hazardous than we thought and that MIL-STD-1474(d) and an A-weighted energy measure are grossly over-conservative. That would be a serious misapprehension. Where there is less low-frequency energy, as in small arms fire, the picture is essentially reversed. At the levels typical for small arms, the middle ear is only moderately non-linear, in which case the assessment of hazard reverses. Consider the case for an unprotected exposure to a hypothetical rifle impulse portrayed in Fig. 6. At the highest peak pressure, 168 dB, AHAH allows one round, A-weighted energy none. However at lower pressures, as the ear becomes more linear, the two measures separate by about 10 dB with A-weighted energy allowing much more exposure. At 158 dB, where experiments with human Ss have been done (Brinkmann, 2000) these tests have been done and 3 or 6 rounds is essentially in the right region. But at lower pressures, there are simply not any data on human exposure, so this comment can only be precautionary. If the AHAH model is correct at 140 or 150 dBp, then A-weighted energy is grossly under-conservative and serious hearing loss could result from exposures predicted to be safe. If AHAH fails to predict correctly at these intensities, then the error will be one of over-conservatism. Interestingly enough, the A-weighting prediction and the AHAH prediction would be essentially identical at the lower pressures if the AHAH model were predicting hazard for the 50%ile ear.

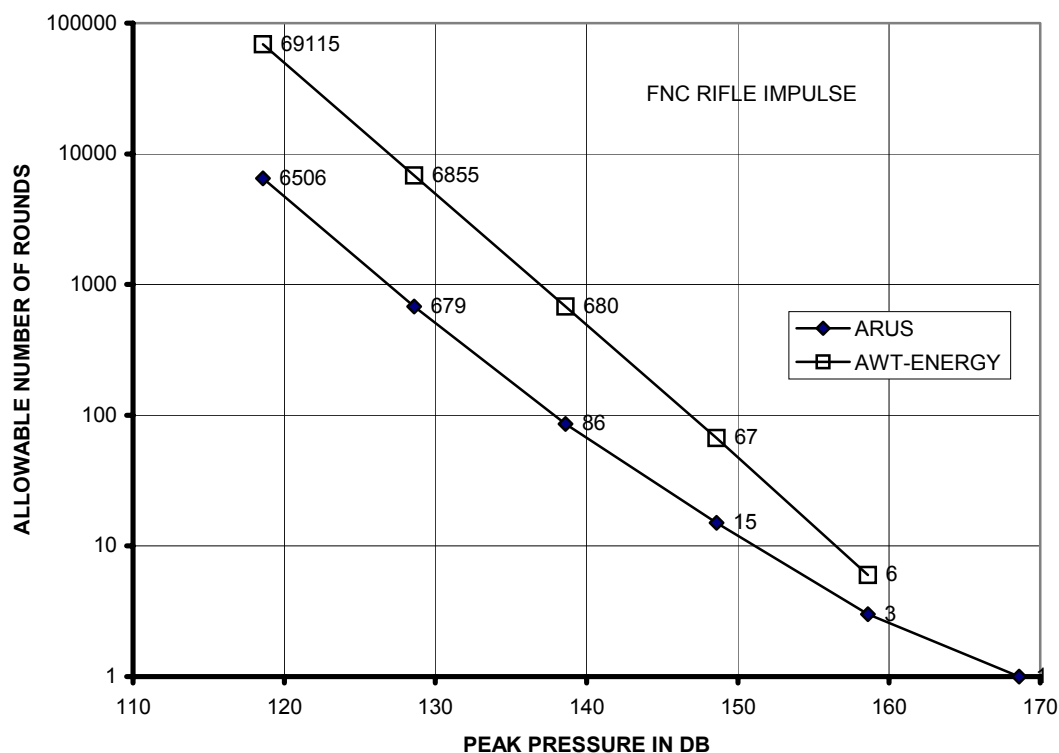


Figure 6. Allowable number of rounds for a rifle impulse, unprotected exposure as rated by the AHAH model and A-weighted energy

MIL-STD-1474(d) only allows exposures with protected hearing at levels above 140 dB. If we calculate the exposure allowed for a rifle impulse using that standard, we typically arrive at an allowable number of rounds between 100,000 and 1,000,000 for single hearing protection. If we calculate the effect of an HPD or measure the hazard using a protector on a manikin, the AHAH model estimates about 600 – 15,000 rounds for the warned ear, depending on the impulse and the protector. Again, there is an immense discrepancy between these two systems. Surely MIL-STD-1474(d) is under-conservative. It allows 15-25 dB more energy than an A-weighted energy criterion would. However, under the protector, peak pressures are in the 125-140 dB region, an area where the AHAH model is untested. More data would obviously be welcome here.

3.4 IMPLICATIONS FOR LEVEL/NUMBER TRADING RATIO CALCULATIONS

Tradition has the power to focus thinking in familiar patterns. One such persistent idea is that there is a level/number trading ratio for impulses which expresses the change in the acceptable number of impulses for a given change in level (Smootenburg, 2003). In the case of A-weighted energy, a 10-fold change in number would require a 10 dB reduction in level. The problem is that this concept presumes that the ear is essentially linear. While that is true for an immense dynamic range, it is not true at very high levels. For instance, if we take an impulse and artificially change its level over a wide range we can observe a progression of effect. In Fig. 7 we see the allowable number of impulses for the airbag impulse in Fig. 3, were its peak pressure changed in 10 dB steps from 112 to 182 dB. By way of comparison, the level/number ratio has also been plotted on the right hand ordinate. At the low intensity end, the trading ratio is essentially 10 (an energy relationship – for a 10 dB increase in peak pressure, the number of rounds allowable drops by a factor of 10).

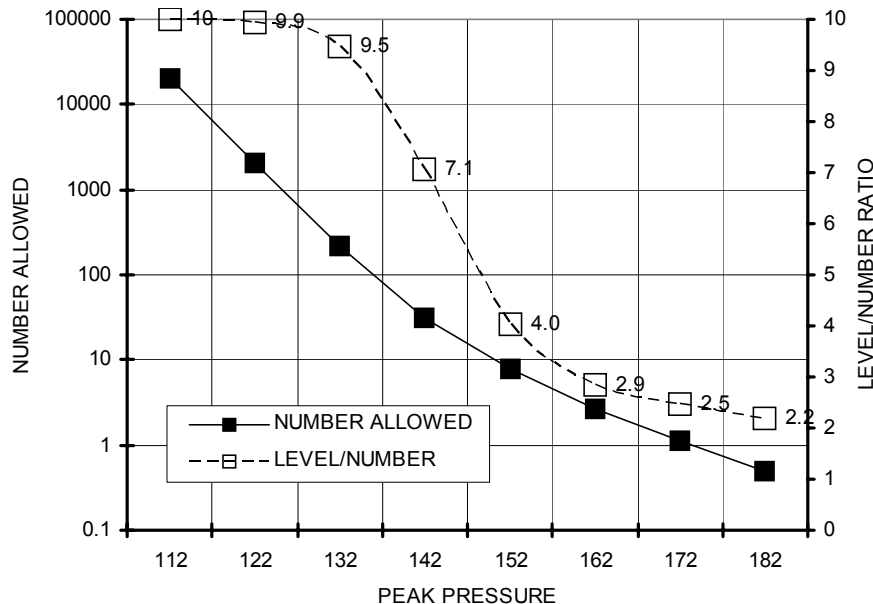


Figure 7. Allowable number of impulses for a warned ear for the waveform in Fig. 3 (left ordinate). Also plotted is the level/trading ratio for each of the levels.

However, as the pressure rises and the clipping of the stapes has its effect, the ratio changes downward to 2.2 at 182 dB. This makes sense, given the non-linear middle ear. This particular impulse is taken as illustrative of the problem in general. It is apparent that it would be irrational to seek a single level/number trading ratio to work for impulses over a wide range of pressures. Smoorenburg has nonetheless tried (2003) to find the level/trading ratio and found numbers for various impulses and experiments that fall in the range plotted in this figure. The AHAH model is thus consistent with the available data in arguing that no single level/trading ratio for impulses should be expected to exist.

We also note in passing that in his review chapter on impulse noise, Smoorenburg (2003) had credited the AHAH model as doing essentially the right thing for impulse noise. However, he faulted the model for being “over compressive” i.e. failing to find a proper trading ratio. A close examination of the text indicates that in fact Smoorenburg in one case simply misapplied the model and in a second, made a mathematical error. If these two errors were corrected, his reservations with respect to the model disappear.

3.5 IMPLICATIONS FOR WEAPON DESIGN

For large calibre weapons, which typically produce high peak pressures and contain a great deal of low frequency energy, the AHAH model represents an important change in thinking. In general, AHAH suggests, consistent with the data from the Albuquerque studies, that higher peak pressures and longer durations are in fact more tolerable than previously thought. Alternatively, it may mean that for some purposes, double hearing protection will no longer be required. Double hearing protection, without some provision of a talk-through system, has never been viewed as practical in the field. The double protected ear is essentially deaf, and that’s simply not acceptable in combat.

The move to a theoretically based hazard rating system means that a new range of design options may become available. Reducing peak pressure may not be the best way to reduce hazard and alternate processes can now be pursued. On the one hand, it may be possible to use the new tools to produce even more powerful weapons that are no more hazardous than the current ones. Alternatively, it may be possible to pursue the design of weapon systems with the capacity of the current weapons; but with exposures that are safe without hearing protection or at least no more hazardous than current systems. Given the uncertainty of HPD use in combat, it would be prudent to pursue a goal of reducing the hazard from weapons so that less protection is needed.

On the other hand, the AHAH model emphasizes the hazard from small arms, an area that has traditionally been overlooked. Many years ago, a study by Walden, Prosek and Worthington (1967) found that hearing loss in the US Army was prevalent and about equal in magnitude in the various combat arms. In terms of exposure, the artillery would have been considered the worst hazard; but it was not so. It is only a hypothesis; but it may well be that the weapon most likely to produce a hearing loss is the rifle. Surely there are more of them and exposure is more likely, especially in training. Generally, when the AHAH model is used to estimate hazard with an HPD, it allows a reasonable number of impulses, perhaps a few hundred. However this is typically much less than allowed with MIL-STD-1474(d), which might allow many thousands. Clearly small arms may presently be an under-appreciated hazard.

Currently, there is specific concern for the ability to fire shoulder-fired rockets from within a room or a bunker. Special designs have evolved; but in the end, human safety may well be the determining factor regarding their acceptability. The impulse within a room is indeed intense and prolonged with the result that the existing hazard rating methods severely restrict and/or prohibit firing. The AHAH model may provide a more accurate assessment and promote safer designs in this arena.

3.5 IMPLICATIONS FOR HPDS

One of the advantages of a model is that it allows one to try different contingencies for their effects. In one such exercise the question was asked regarding what the best possible attenuating characteristic was for an HPD when the stimuli were different types of gunfire in quantities that might be encountered in military operations, e.g. 300 rounds from a rifle or a cannon. To do the test, attenuation characteristics were created, flat across frequencies, for example or a pattern of attenuation shaped like the threshold curve. A digital filter with that characteristic was used to process waveforms from the different weapons and the waveform was run through the model. By processing a series of these trials it was possible to determine how much attenuation would be required to make a 300 round exposure acceptable. A typical outcome for a cannon impulse is plotted in Fig. 7. The rather surprising outcome of this calculation was that the low frequency attenuation

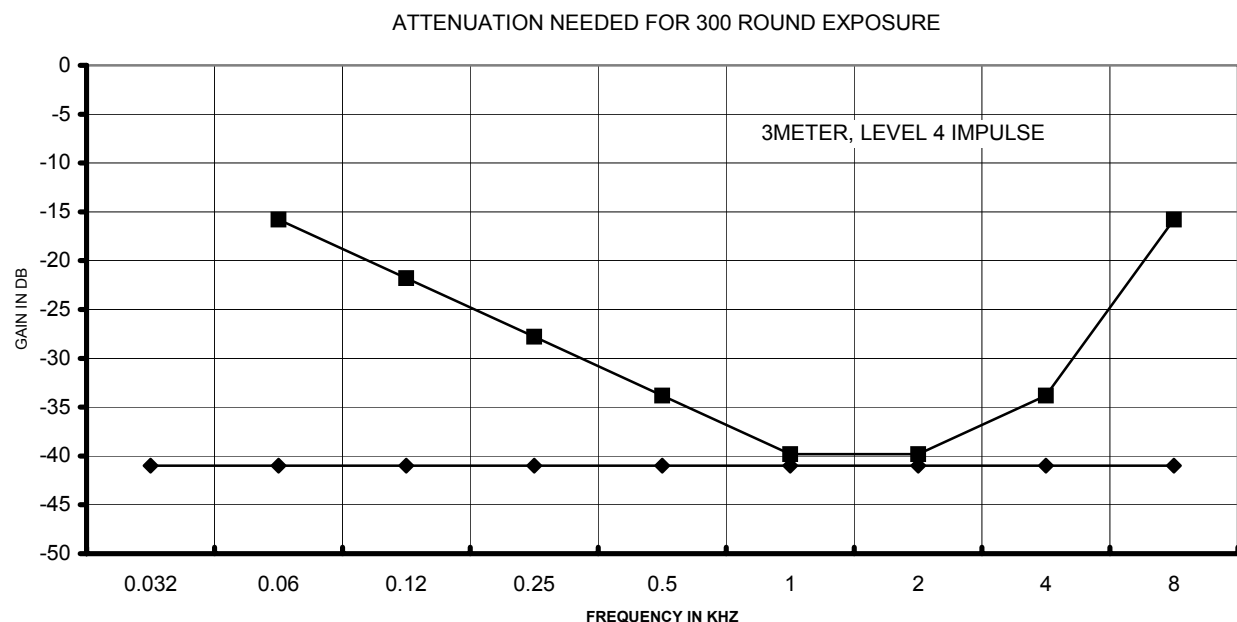


Figure 7. Calculated gain (attenuation) needed in an HPD for a 300 round cannon impulse exposure. The 3m level 4 impulse had a peak pressure of 185 dB in the free field.

did not seem to matter. Below 1 kHz the attenuation declined at 6 dB/oct for the threshold-shaped curve yet the curves reached about 40 dB in the 1-2 kHz region and they were calculated to have the same protective effect. Essentially the same pattern was observed for the other exposures (rifles, rockets, etc.) even though the absolute levels of attenuation varied with the exposure. In general, low frequency attenuation is much less critical than the attenuation in the mid-range of frequencies. Or put conversely, low frequency energy is not necessarily the enemy.

The primary reason that low frequency attenuation may not be critical at these levels is, as we have seen, that the middle ear itself has become non-linear and blocks sound transmission. This can be seen in Fig. 8 in which we have plotted the allowable number of simulated cannon impulses (measured under a protector) as a function of the A-weighted energy in the impulse or the hazard as calculated by the AHAH model. For low

SPLs where the ear is essentially linear (upper left portion of the figure) the two curves are almost parallel. But as the pressure rises, the two curves diverge markedly. The ARU curve bends to the right (allowing more rounds than A-weighting) because the stapes displacement has been limited and the energy has not been transmitted to the inner ear.

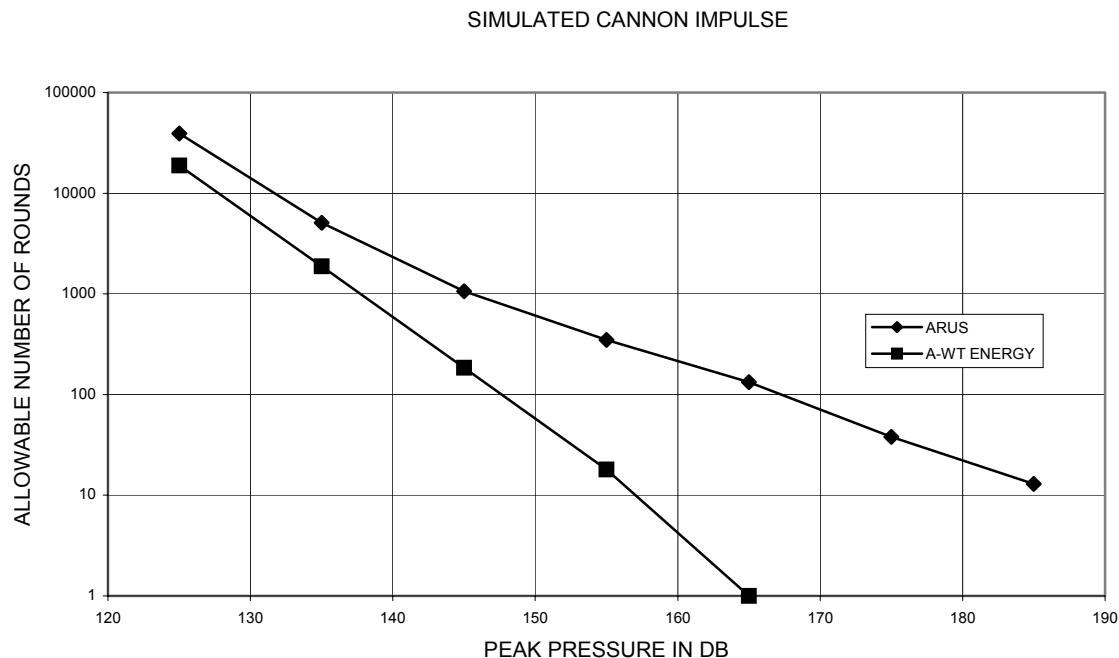


Figure 8. Allowable number of simulated cannon rounds as calculated with the AHAAH model or an A-weighted energy criterion.

So far as HPD design is concerned, then, the message is optimistic. Typically, the low frequencies are hard to attenuate and good attenuations in the mid-range are easier to achieve. Fortunately, that pattern may be acceptable. In fact, the non-linear protector now known as the “Combat Arms Plug” (CAP) in the US Army has an attenuation characteristic very much like the “threshold-shaped” attenuation curve in Fig. 7. The success of the combat arms plug is consistent with the analysis done with the AHAAH model.

There is another aspect to this as well where MIL-STD-1474(d) is concerned. In it, only two corrections are allowed for hearing protection, one for single protection and one for double protection. This is without regard to the specific attenuation of the protector or the variability of its fit in use. Thus there is no reward for better HPD design (or penalty for worse designs either). In the case of the AHAAH model (or an A-weighted energy criterion) better design could be measured, analysed and rewarded.

3.6 IMPLICATIONS FOR SITUATION AWARENESS AND SPEECH INTELLIGIBILITY

One of the strong points in favor of the non-linear CAP is that at low pressures it provides only a small amount of attenuation, which in turn permits the wearer to understand speech while wearing it and to generally stay aware of what is going on in the surroundings. The attenuation of the CAP plug for waveforms

at different peak pressures is shown in Fig. 9 along with the threshold shaped attenuation curve from Fig. 7. We see that the low frequency attenuation of the CAP and that of the threshold shaped attenuation curve are very similar. This is consistent with the success of the CAP plug in protecting the ear, in spite of its relatively poor low frequency attenuation.

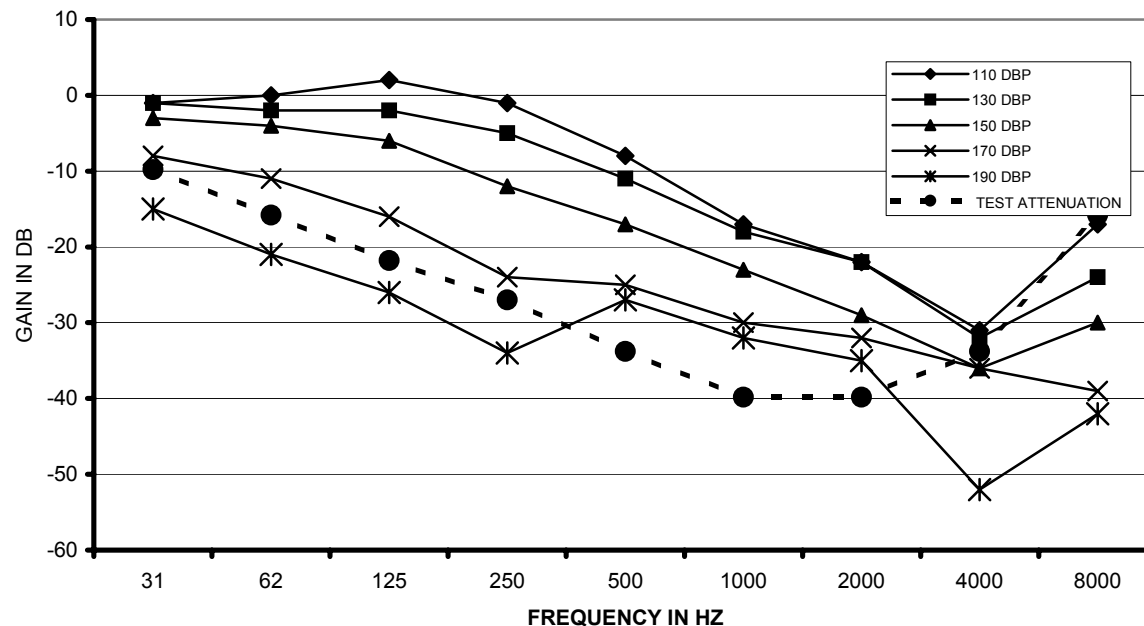


Figure 9. Gain (attenuation) of non-linear earplug as function of peak level of the incident pulse, plus the threshold attenuation curve from Fig. 7 (dashed line).

One of the purported strengths of the CAP is that it helps to maintain both situation awareness and speech intelligibility. By means of a simple calculation we can gain an appreciation of the improvement in speech intelligibility brought about by the use of an HPD with little low-frequency attenuation. The Articulation Index (AI) is an old measure that relates speech intelligibility on a variety of measures through the calculation of weighted signal-to-noise ratios in octave bands important in the perception of speech. We might, for instance ask what the difference in speech intelligibility would be if HPDs such as those in Fig. 7 were worn. If the HPD were limiting the speech signal, as it might in a quiet environment, typical of many combat situations, then we can calculate that the improvement in the AI would be about 0.2 for the plug with less low frequency attenuation. The significance of the 0.2 change in AI would depend on the specific conditions; but if intelligibility were marginal, that much change could be the difference between 20-30% intelligibility and 80-90% intelligibility, an immense difference.

3.6.1 The Effect of Speech Intelligibility on Performance

The effect of speech intelligibility on performance of combat related tasks is surprisingly strong. Peters and Garinther (1990), working at the Army Research Lab ran a series of experiments that systematically varied the speech intelligibility within a communication system and tested its effect on crew performance. The final

outcome appears in Fig. 10. In this figure, probability of mission success is plotted as a function of the percent speech intelligibility within the communications system. If the task were very simple, not requiring much communication, then the upper bound of the area fits. Speech intelligibility doesn't make too much difference. On the other hand, when the mission consisted of tank crews attacking one another in an unconstrained situation (in a simulator)(lower curve in the figure), then the effect of speech intelligibility is immense. There is an almost 1:1 correspondence between changes in speech intelligibility and mission success. There is little in the way of materiel improvement that could produce such a dramatic effect on system performance. Good hearing and good speech intelligibility are essential to successful military operations.

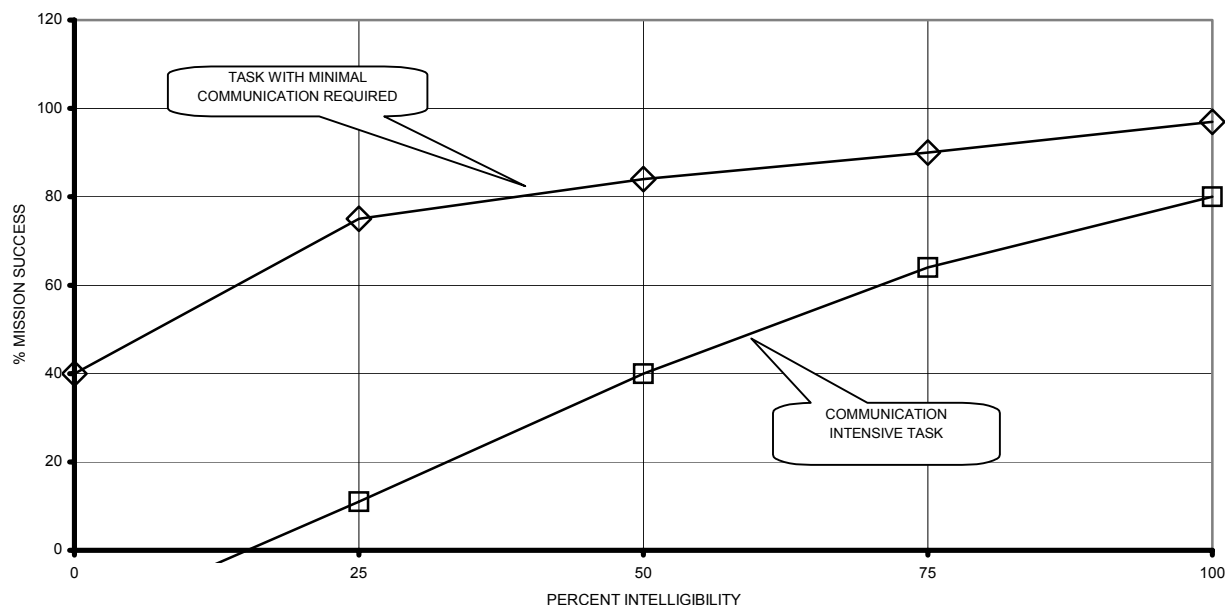


Figure. 10. Probability of mission success as a function of speech intelligibility in the crew communication system

4.0 TECHNICAL ISSUES FOR THE FUTURE

4.1 An Integrated Measure

4.1.1 The AHAH Model

As of this writing, it appears that the AHAH model is performing well for impulse noise predictions at high SPLs. However, it would clearly be desirable to have an integrated analysis system that could appropriately analyze, without compromise, all exposures from continuous noises at low levels to impulses at very high levels. As we have noted earlier, there is real question as to the range of intensities for which the AHAH model as presently configured could serve as an adequate descriptor. The prospect of creating an integrated measurement and analysis system is indeed laudable, technically challenging, and would require some theoretical insight to complete. We are presently working toward this end.

4.1.2 A-Weighted Energy

One of the justifications offered for the use of A-weighted energy as a measure is that it is an integrated measure. In this case the argument the A-weighting allows the combining of lower level continuous and intense impulsive exposures may appear to fit the situation; but it is seriously compromised from a theoretical standpoint. For example, such a measure takes no account of changes in the loss mechanism as intensity rises. An exposure to an 85 dBL_{AEQ8} with a factory noise at about 85 dB means that if the exposure is continued daily for a working lifetime, recovery will occur quickly each day and a very small loss may occur in 20 or 30 years. On the other hand, the same level of exposure to an intense impulse may have so severely stressed the cochlea that it is on the edge of permanent damage, recovery processes would be prolonged and daily exposure would probably be a bad idea. If the exposure were a few dB higher on one occasion in the factory, the consequence for that noise exposure would be negligible. But in the case of impulse noise, the result could be disaster for the ear. The fact that an 85 dBL_{AEQ8} appears to work as a descriptor across the broad range of exposures is almost certainly a happy coincidence and not a theoretical necessity. The fact that it is easy to measure should not obscure the fact that it may not truly describe stresses to the auditory system.

4.2 Adaptive Middle Ear Muscles

The middle ear muscle system may benefit the ear; but its effect is very difficult to describe. The AHAAH model has made an attempt to do so; but the method requires operator input that is relatively simple for the case of impulses in isolation. A truly excellent program would make the judgement for all exposure conditions automatically and incorporate algorithms for both attack and decay of the response in continuous noise exposures.

5.0 SUMMARY AND CONCLUSIONS

The essential message in this address is that a new analytical tool, the AHAAH model is now available for predicting hearing loss and understanding the mechanisms that operate during exposures at high SPLs. Unlike other methods, it is theoretically based and embodied in a computer program. The model will serve as the basis for noise rating in MIL-STD-1474(e) and is serving within the Society of Automotive Engineers for rating hazard from airbags.

The AHAAH has been validated with more than 70 tests with human data. These tests show that that the AHAAH model accurately predicts the onset of unacceptable changes in hearing sensitivity in over 95% of the cases. In contrast, A-weighted energy is correct about 30% of the time and MIL-STD-1474(d) is accurate about 36% of the time. Errors in prediction for MIL-STD-1474(d) and A-weighted energy tend to be in the over-prediction of hazard for impulses with a lot of low frequency energy (large calibre weapons).

On the other hand, the AHAAH model suggests that small arms impulses may be much more hazardous than previously thought. Given their pervasive presence in the military, it might be a fruitful area for the production of safer weapons.

The AHAAH model predicts hazard for any intense sound, continuous or impulsive. The unanswered question at the moment is how far down in pressure it can go and remain accurate. It is reasonable to assume it is accurate to at least 140 dBP, which could include rms levels of 125 to 130 dB.

The reduced hazard from large calibre weapons impulses can have a major impact on weapon design. Designers now have more leeway to consider more powerful weapons, lighter weight designs, or safer weapons without reduced performance.

The AHAH model includes a movie feature, which allows the analysis of the evolution of hazard within the inner ear. This feature makes it possible to devise responses that are tailored to the specific issues.

Because hazard is the result of instantaneous stresses within the inner ear, it would follow that hazard analysis needs to be done in the time domain, rather than the frequency domain.

The AHAH model points out the role of very low frequency energy in modulating the flow of energy into the cochlea. If it is intense enough to cause the stapes suspension to reach its limits, then it affects the flow of energy even though it may be lower in frequency (<20 Hz) than the normal auditory range.

For weapons impulses, HPDs need their best attenuation in the mid-range and less at low frequencies (below 1000 Hz).

With less low frequency attenuation, speech intelligibility and situation awareness are improved.

Given a non-linear middle ear, the current practice of calculating number/intensity trading ratios for exposures is irrational. Once the middle ear has become non-linear, increases in intensity result in proportionately less energy entering the cochlea. It follows that whatever trading ratio fits at one level, must not fit at another level.

1.1 REFERENCES

- [1] American Institute of Biological Sciences (2001). "Peer Review on The Human Research and Engineering Directorate (HRED) Method for Assessing the Risk of Auditory Injury for Hearing-Protected Soldiers Exposed to Impulse Noise", at www.arl.army.mil/ARL-Directorates/HRED/AHAAH/.
- [2] ANSI (2005). "Estimation of the Hazards Posed by Exposure to Impulse Noise", Draft ANSI S3.48TR-200X.
- [3] Brinkmann, H. (2000). Supplementary investigation of the German damage risk criterion with the Belgian NATO small arms rifle FNC, in Report from NATO Research Study Group 29, Panel 8, AC/243, Ch 4.
- [4] Buck, K, Dancer, A, and Parmentier G. (2003). T-weighting or A-weighting, what to use for evaluation of exposure limits. TRO Technical Report TR-017 Reconsideration of the Effects of Impulse Noise, p 25-27
- [5] Dancer, A. (2000). "Proposal for a new damage risk criterion", In Report from NATO Research Study Group RSG.29(Panel 8 – AC/243) Reconsideration of effects of impulse noise, TNO-Report TM-00-I008, pp 11-15 (second meeting).
- [6] Garinther, G. R. and Peters, L. J. (1990). "Impact of communications on armor crew performance", Army Research, Development & Acquisition Bulletin, (January-February 1990) pp 1-5.
- [7] Johnson, D. L. (1998). "Blast overpressure studies", USAARL Contract Report No. CR-98-03, U.S. Army Aeromedical Research Laboratory, P.O. Box 620577, Ft. Rucker, AL 36362-0577.
- [8] Johnson, D. (1994). "Blast overpressure studies with animals and men: A walk-up study", USAARL Contract Report No. 94-2, U.S. Army Aeromedical Research Lab, Ft. Rucker, AL 36362-0577
- [9] MIL-STD-1474(d). (1997). US Department of Defense Design Criteria Standard, Noise Limits, AMSC A7245
- [10] Ministry of Defense (1982). "Acceptable Limits for Exposure to Impulse Noise from Military Weapons, Explosives and Pyrotechnics," Interim Def Stan 00-27/1, Ministry of Defense, Directorate of Standardization, First Avenue House, London, WC1V6HE, England
- [11] NATO Research Study Group RSG.29 (Panel 8 – AC/243) (2000) "Reconsideration of effects of impulse noise", TNO-Report TM-00-I008
- [12] Pfander, F. (1975). *Das Knalltrauma*. Berlin: Springer-Verlag
- [13] Price, G. R.(2003). "Weapon noise exposure examined with the AHAAH model", in: www.arl.army.mil/ARL-Directorates/HRED/AHAAH/

- [14] Smoorenburg, G. (2003). "Risk of Hearing Loss from Exposure to Intense Sounds" Chapter 1 in TRO Technical Report TR-017 Reconsideration of the Effects of Impulse Noise, p 25-27
- [15] Smoorenburg, G. F. (1980). "Damage Risk Criteria for Impulse Noise," Institute for Perception TNO Report 1980-26, 34 pp
- [16] Society of Automotive Engineers (2003). "Impulse Noise from Automotive Inflatable Devices" SAE J2531, Nov.
- [17] Walden, B. E., Prosek, R. A. and Worthington, D. W. (1975). "The prevalence of hearing loss within selected US Army branches, (Washington, DC: US Army Medical Research and Development Command) Interagency IOA 4745.

CONSTRAINTS and NEEDS in the Auditory and Vocal Environment of Modern Air Combat

Cdt (Maj) BIGOT Frederic
EMAA/BSA
26 Bd Victor
00460 PARIS ARMEES
fred.bigot@free.fr

PREAMBLE

As far as allied operations had occurred around the world during the last 15 years, many lessons learned appeared to the fighter pilot's unity. Among them, one might be the most important to bear in mind. The faster we exchange data between aircrafts, the best we coordinate our simultaneous actions. However, human ability to understand and manage a flow of information is limited. To improve the efficiency of communications, modern technologies are to be used, to reduce pilot's workload.

I personally experienced some scientific studies about Vocal Command and 3D Sound. THALES Company and the Defence Science and Technology Organization (Australia) have formed a group to study these technologies in Modern Air Combat. In France, the simulation means were a single seat Rafale cabin integrated in a 10m diameter combat dome and managed from a development and monitoring room.

1. ENVIRONMENT CONSTRAINTS

In addition of the classical ambient noises existing in a fighter cockpit, numerous sounds are implemented into modern combat aircrafts. It allows pilots to minimize their visual checking onto the main control panel. According to the extreme cost of new generation aircrafts, any loss of aircraft in combat is to be considered as a heavy failure. Consequently, the coordination between aircrafts must be as efficient as possible, to build a common situation awareness (SA) and avoid any unexpected enemy shot.

1.1 The aeronautical environment

1.1.1 Load factor

Modern aircrafts can reach easily strong load factors up to 10 g. Their manoeuvrability is much more important than in the past. Experiencing heavy load factor during combat shows that a pilot loses his hearing capacity as long as g's are pulled. Furthermore, his analysis capacity can be reduced. There, we understand the difference between to hear and to listen to.

In a dog fight combat, a pilot is concentrated on keeping in sight his enemy, managing the energy, the trajectory and the system of his aircraft. Under load factor, vision and hearing are decreasing insidiously. Therefore, coordination between aircraft can be dangerously affected.

1.1.2 Noise due to aero dynamical rubbing

With the emergence of long range active missiles (AIM 120, MICA, AA12), modern combat is become a “beyond visual range” (BVR) combat. The sooner you shoot the sooner you can escape to the enemy shooting envelope. Escaping manoeuvres might be done at high speed and low altitude.

On the other hand, flying at high speed and low altitude may generate heavy aero dynamical noise in the cockpit. This can affect the quality of transmissions and often reduces the understanding ability of the pilot

1.1.3 Engine and air conditioning noise

Without a huge thrust, a fighter aircraft is not efficient. Most of the time, the engine is located just in the pilot’s back. It also generates vibrations, heavy noise and roaring.

Air conditioning is mandatory but not discrete The ventilation blows regularly and noisily. It doesn’t really trouble the pilot but adds yet a constant noise into the cockpit.

1.2 The tactical context

1.2.1 Number of participants

Since the beginning of the Aviation, aircrafts involved in the air warfare had always been numerous. Composite Air Operations, heavy strike deep inside the enemy lands, air combat between tens of aircrafts lead officials to solve problems of communication and coordination. Therefore, the audio environment may be fully confused and disorganized, causing misunderstandings and fratricide shots!

So far, cockpit sounds must be relevant, discrete but efficient, and give place to information about the SA. Overloading hearing pilot’s capacity with lots of specific sounds is not appropriate. The pilot must be able to catch any vital data on the radio frequency to prevent him from unexpected threats.

1.2.2 Radio jamming (TRANSEC)

During the fight, radio exchanges might be scrambled by enemy jamming stations. Generally, the noise level is pretty high and frankly disturbing for pilot’s mind..

1.2.3 Radio cryptology (COMSEC)

To avoid enemy intrusion in allied communications, a secured radio is employed (“Secure Voice”). Encrypted transmissions are not as clear as expected and may slightly trouble pilot’s understanding.

1.3 The aircraft weaponry system

1.3.1 Two radio transmitters

Two radio transmitters are commonly used onboard. The pilot may simultaneously hear to two different talks and may be confused or miss any vital information about his safety, regarding the SA, the controller orders, or any threat calls.

1.3.2 Weaponry system alerts (RWR, Missile head)

In Modern Air Combat, two main equipments are employed by the fighter pilot. Firstly, the radar gives the pilot the capacity to build his own SA, completed by the controller or the wingies. Secondly, the Radar Warning Receiver (RWR) consists in alerting pilot to enemy radar pointing over 360° around him. The alert is given visually on a screen associated with specific sounds.

In BVR combat, the visual information shown on the screen in 2D dimension is widely sufficient to build the SA.

However, in “Within Visual Range” (WVR) combat like dog fights, the aircraft attitude may be unusual, troubling the pilot’s analysis about the incoming threat. It is generally hard to locate any other threat while being upside down, head up and eyes turned to the enemy to keep visual on! At that time, 2D visual information on the RWR screen is hard to analyse. Moreover, keeping visual contact on the enemy prevents the pilot from glancing at the RWR screen regularly. Therein, audio alert may be a significant help.

1.3.3 Aircraft system alerts (failure, altitude, fuel, Auto Pilot)

Modern aircrafts are comprised of many audio alert devices like failure warning, crossing altitude warning, minimum fuel warning or Auto Pilot disconnection warning. It causes sometimes a bit of confusion when all these sounds get tangled in the pilot’s mind.

2. NEEDS FOR AN EFFECTIVE SYSTEM

2.1 Auditory needs

Because audio information can be caught at any time, head up, it is the best way to alert the pilot of a warning situation.

2.1.1 Quality of the radio transmitters

Both radio transmitters quality must be as good as possible to minimize noise interference and cluttered voice. It appears obvious that the clearer the radio transmission can be the better the coordination is achieved on a battlefield.

2.1.2 Radio hearing level presets

As two radio boxes are implemented on a fighter aircraft, one of them should be called the “Main Box”, and the other the “Secondary Box”. Usually, the Main Box’s listening level is upper than the secondary one in order to distinguish one from the other. The pilot does regularly change the noise levels when switching boxes.

A system that could preset automatically both noise levels when switching channels would be relevant. It would improve the safety and the aircraft’s ergonomics.

2.1.3 Advanced Electronic Noise Reduction System

Civilian aviation companies suggest its customers some noise reduction headsets during long haul flights. The technology is easy to implement and cheap. Reducing the noise inside the cockpit would improve quickly the pilot’s comfort and exchanges’ efficiency.

2.1.4 3D Sound System in WVR combat (RWR, IR missile lock on)

As told previously, keeping full SA during a Dog Fight is not an easy job. Analysing the RWR screen when pulling g’s and keeping visual contact with the enemy is not that easy. However, an option to visual information on a screen is available.

3D sounds technology makes all that easier during WVR combat. Two major positive points might be underscored:

① During WVR combat, 3D sounds system offers the pilot the possibility to build the SA instinctively and surely. Locating any threat direction quickly while keeping in sight his bandit, makes the pilot more efficient in his combat and more aware of the SA.

② 3D sounds system may be used as an indicator of Electronic Warfare density. Not having the capacity to locate the threats reveals the presence of multi directional threats and can be alarm.

2.1.5 Audio failure warning

Every sound should be appropriate and representative of the type of information delivered. Failure warnings are generally associated with a light button on the failure panel. The pilot must analyse this panel to take a proper action.

Following the first warning alarm, a synthetic voice announcing the type of the failure would sometimes make the analysis easier during particular phases (night, sun reflection on the panel, air refuelling process, close formation in the clouds, landing and take off).

2.2 Vocal needs

During overload flight phases, the pilot’s resources might be fully used. Studies tend to prove that the Vocal Command could free some resources in order to use them to other tasks. In fact, once the overload is present, there is a little use of vocal command.

In a modern combat aircraft, the most difficult automatism to acquire is the voice. It needs a long and tedious training to be fully operational. Talking during a fight, as a danger is imminent, requires lots of mental resources. The resources consumption is all the more important than the vocal command phraseology is long or complex.

2.2.1 Radio equipment (HQ radio and Secure Voice Radio)

To avoid enemy intrusion on the tactical frequency, it is mandatory to use HQ radio. This equipment allows the pilot to avoid radio jamming and also improves the exchanges' quality. However, efforts must be done to improve the reliability of this type of equipment which sometimes goes wrong due to problems of time reference between aircrafts.

Encrypted voice is often a bit confused and requires a minimum of attention by the pilot. It would be interesting to improve this technology to protect the pilot from difficulties to understand some exchanges.

2.2.2 Vocal Command (FQ changes)

Numerous studies have been achieved to determine the benefits of a vocal command in a fighter aircraft. This technology appeared to be very useful in some specific situations. HOTAS commands are generally the best way to improve the cockpit ergonomics. However, some functions are not always available via HOTAS command, either due to their specificity (i.e. radio control panel), or to the overloaded HOTAS commands' availability. Most of the time, pilot has to switch frequency onto the radio control panel, away from the stick and the throttle. Having the opportunity to do it vocally would be such a huge improvement for ergonomics and safety considerations.

Vocal command must be declined into three major features:

❶ It must be used quietly, without haste. Talking in combat is the most difficult automatism to get and to keep. A long training is necessary to be on top. Gestures appear much easier and faster to execute than the voice. Vocal command also requires time and mental resources to be used correctly. Consequently, the Vocal Command syntax must be short and easy to pronounce, perhaps as tactical codewords are chosen.

❷ During hot combat, pilots need a complete availability of the radio transmitter to listen to the situation awareness and to order any emergency call. No time for pronunciation efforts and hazardous results!

Vocal command is also useful when time is free, danger is out and any HOTAS command is not available (i.e. radio frequency switches, navigation waypoints management, weapon selection ...etc, in close formation or at night).

❸ Vocal command technology is efficient but system misunderstandings may regularly occur. The reliability of this function is not complete. A validation process must be added to confirm the command and to avoid wrong vocal orders.

3. CONCLUSION

Today's missions become more and more complex. It is also mandatory to improve the ergonomics of the modern combat aircraft in order to minimize the pilot's workload.

A good ergonomics must be :

- intuitive
- simple
- reliable

Modern technologies bring military aviation the opportunity to improve the building of the Situation Awareness and the fighter aircraft cockpit's ergonomics. 3D sounds and vocal command technologies appear to be relevant as the complexity of the weapon system increases regularly. As a matter of fact, they must be implemented in the modern combat aircrafts.

U.S. Navy Flight Deck Hearing Protection Use Trends: Survey Results

Valerie S. Bjorn

Naval Air Systems Command
AEDC/DOF 740 Fourth Street
Arnold AFB, TN 38389-6000, USA

valerie.bjorn@arnold.af.mil

Christopher B. Albery

Advanced Information Engineering Services - A General Dynamics Company
5200 Springfield Pike, WP-441
Dayton, OH 45431, USA

chris.albery@wpafb.af.mil

CDR Russell Shilling, Ph.D., MSC, USN

Office of Naval Research - Medical and Biological S&T Division
800 N Quincy Street
Arlington, VA 22217-5860, USA

russell_shilling@onr.navy.mil

Richard L. McKinley

Air Force Research Laboratory (AFRL/HECB)
2610 Seventh Street, Bldg 441
Wright-Patterson AFB, OH 45433-7901

richard.mckinley@wpafb.af.mil

ABSTRACT

Hearing loss claims have risen steadily in the U.S. Department of Veterans Affairs across all military services for decades. The U.S. Navy, with U.S. Air Force and industry partners, is working to improve hearing protection and speech intelligibility for aircraft carrier flight deck crews who work up to 16 hours per day in 130-150 dB tactical jet aircraft noise. Currently, flight deck crews are required to wear double hearing protection: earplugs and earmuffs (in cranial helmet). Previous studies indicated this double hearing protection provides approximately 30 dB of noise attenuation when earplugs are inserted correctly and the cranial/earmuffs are well-fit and in good condition. To assess hearing protection practices and estimate noise attenuation levels for active duty flight deck crews, Naval Air Systems Command surveyed 301 U.S. Navy Atlantic and Pacific Fleet flight deck personnel from four aircraft carriers and two amphibious assault ships. The survey included a detailed assessment of cranial helmet fit and maintenance condition (e.g., earmuff headband tension, earcup foam and cushion integrity); earplug use and insertion depth; anthropometric measures; and personal/historical data. Data analysis showed that 79% of surveyed flight deck personnel

Bjorn, V.S.; Albery, C.B.; Shilling, R.; McKinley, R.L. (2005) U.S. Navy Flight Deck Hearing Protection Use Trends: Survey Results. In *New Directions for Improving Audio Effectiveness* (pp. 1-1 – 1-20). Meeting Proceedings RTO-MP-HFM-123, Paper 1. Neuilly-sur-Seine, France: RTO. Available from: <http://www.rto.nato.int/abstracts.asp>.

U.S. Navy Flight Deck Hearing Protection Use Trends: Survey Results

ears received an estimated 0-6 dB of noise attenuation from either shallow earplug insertion depths or never wearing earplugs (47% reported never wearing earplugs). For subjects who reported they sometimes or always wore earplugs (14% reported always wearing earplugs), only 7% inserted the earplugs deeply enough in both ears to achieve the maximum expected noise attenuation of 22 dB in both ears. Worn without earplugs, the cranial helmet with earmuffs has been reported to provide approximately 21 dB of noise attenuation when correctly fit, worn, and maintained. All survey subjects reported wearing a cranial helmet with earmuffs, but 75% of subjects were issued a questionable size (most wore the largest of four sizes available), and 41% of earcup cushions and foam inserts were deteriorated, hard, creased, or missing. This survey identified numerous technological and hearing conservation policy changes to improve hearing protection for flight deck crews. Based on these findings, the U.S. Navy is improving procedural documentation for flight deck hearing protection fit, use, and maintenance, as well as developing and fielding enhanced hearing protection technology in joint efforts with the U.S. Air Force.

1.0 BACKGROUND

U.S. Department of Defense occupational safety and health instructions set 85 dBA as the safe noise exposure limit for an 8-hour time-weighted average (TWA); and for every 3 dB increase in noise level, the safe exposure time limit is cut in half.^[7] U.S. Navy instructions state that when noise levels exceed 104 dBA, double hearing protection (earplugs and earmuffs) shall be worn, and when noise exposures exceed an 8-hour TWA of 84 dBA, administrative controls like crew rotation are to be implemented, in addition to wearing double hearing protection.^[10,14,16,17]

1.1 U.S. Military Jet Aircraft Noise Levels^[20]

U.S. Navy, Marine Corps, and Air Force high-performance jet aircraft produce 130-150 dB noise. Figure 1 compares legacy military jet noise and estimated noise produced by the next-generation Joint Strike Fighter (JSF). On an aircraft carrier, each catapult launch exposes flight deck crews to approximately 20-30 seconds of aircraft noise with engines at maximum power. Launch duration is defined as the time from when the engine is first run-up past 25% of maximum power until the aircraft clears the end of the deck. When an aircraft is recovered (a cable arrested landing), pilots are required to push the throttle to maximum power again and to prepare to take off in the event they miss the arresting cables. A recovery takes approximately 3 seconds. The recovery duration is defined as the time from when the aircraft first passes the end of the deck until the engine setting is less than 25% of maximum power. In a 24-hour period, a typical busy day for a flight deck crewperson is approximately 60 launches and 60 recoveries on an aircraft carrier.

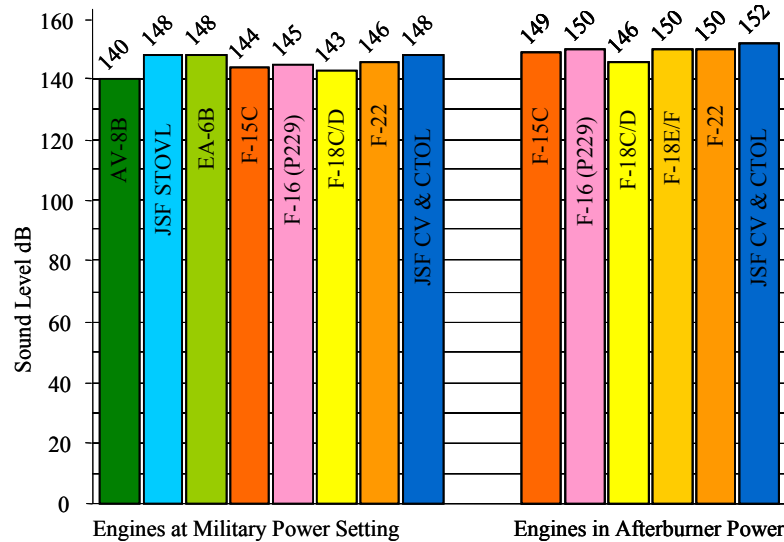


Figure 1: U.S. Military Jet Aircraft Near-Field Noise Levels
Measured Approximately 50 ft Radius and 45 deg off the Nose/Centerline^[20]

1.2 Hearing Protection on U.S. Navy Flight Decks

The hearing protection devices commonly worn on U.S. Navy flight decks are described below.

1.2.1 Earplugs Commonly Used on U.S. Navy Flight Decks

The three most prevalent earplugs used on U.S. Navy flight decks are the Aero E·A·R Classic™ foam earplug, the V-51R Single-Flange earplug, and the Triple-Flange earplug (see Figure 2). The E·A·R Classic™ is a one-size-fits-most expanding foam earplug that needs to be fully inserted in the ear canal to achieve maximum noise attenuation performance.^[8] The V-51R Single Flange earplug and the Triple-Flange earplug are available in sizes; both must be initially fit by medically trained personnel and then fit correctly again for each use by the trained wearer to achieve maximum noise attenuation.



Figure 2: Earplugs Common to U.S. Navy Flight Decks:
Aero E·A·R Classic™, V-51R Single Flange, and the Triple-Flange

U.S. Navy Flight Deck Hearing Protection Use Trends: Survey Results

1.2.2 U.S. Navy Flight Deck Crewman Helmet with Earmuffs

The U.S. Navy Flight Deck Crewman Sound Attenuating Helmet Assembly is commonly called “the cranial” (see Figure 3, left insert). The cranial is worn to protect against head injuries and high intensity noise on U.S. Navy flight decks and in some aircraft. The HGU-24/P cranial includes a sound-powered microphone and headset assembly for communication while the HGU-25(V)2/P cranial does not include communications capability.^[12] The Radio Cranial (also known as the Hydra Helmet) is another communications helmet approved for use (see Figure 3, right insert). The HGU-24/P and HGU-25/P cranials are available in four sizes (6¾, 7, 7¼, and 7½). The Radio Cranial is one size to fit all.

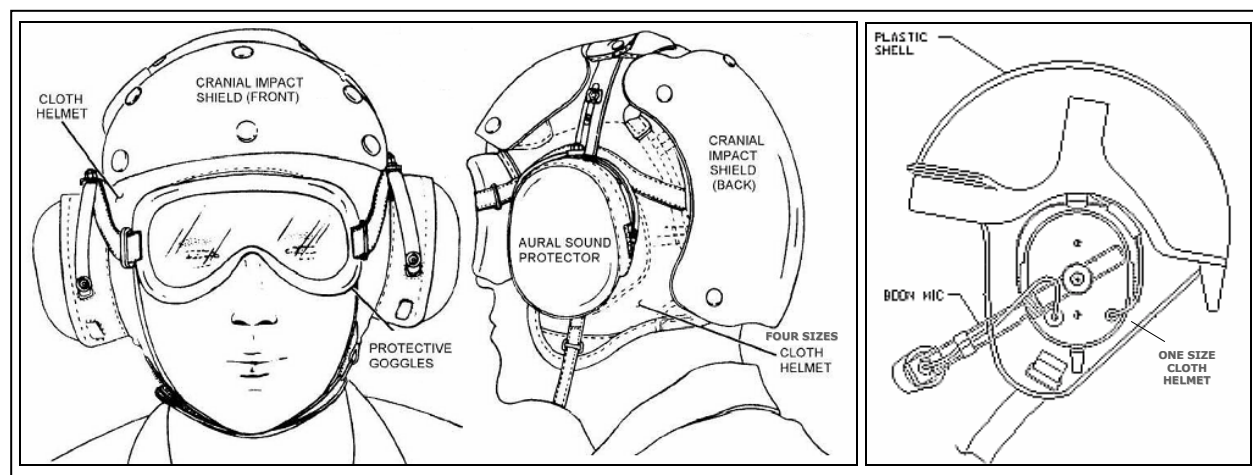


Figure 3: Flight Deck Crewman Sound Attenuating Helmets - HGU-25(V)2/P and Radio Cranial

Earmuffs (see Figure 4) are tethered in the cranial to provide noise attenuation. The left and right earcups include standard ear seals that consist of foam inside a polyurethane skin. The purpose of the ear seal is to create an acoustic seal between the earcup and the user’s head. The inside of each earcup is lined with 0.5-inch polyurethane foam to dampen noise inside the earcup.^[19]

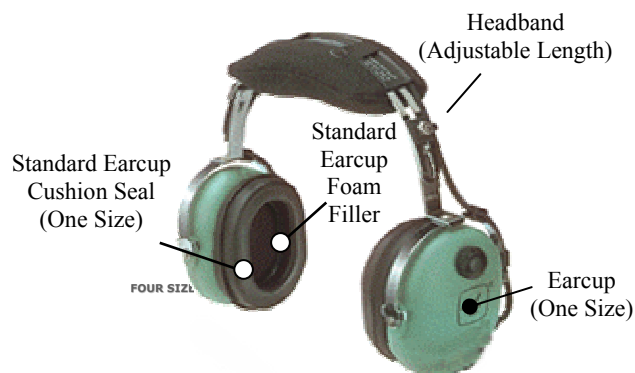


Figure 4: Sound Aural Protector (Earmuffs)

1.3 U.S. Navy Flight Deck Personnel Daily Exposures to Hazardous Noise^[20]

Figure 5 is a diagram of some flight deck personnel locations as they ready an aircraft for catapult launch. Just prior to launch, the Plane Captain (green diamond) and forward Final Checker (blue star) move to the Foul Line; however, the aft two Final Checkers (orange triangles) remain as shown. Figure 6 provides a photo of deck personnel at work around launching aircraft. Crews working at side-by-side catapults are often exposed to the noise of adjacent aircraft as well as the aircraft they are launching. Figure 7 shows noise propagation contour lines for an F-18C jet aircraft. Similar noise contours are generated by other conventional takeoff and landing aircraft. Vertical takeoff and landing aircraft like the AV-8B Harrier produce noise contours that are generally more omni-directional. It is important to note that a number of flight-deck personnel routinely work within the marked "noise hazard" area.

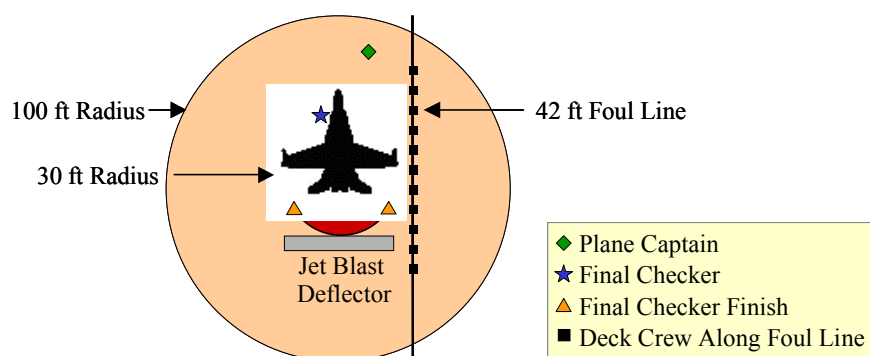


Figure 5: U.S. Navy Flight Deck Personnel Locations during Aircraft Catapult Launch



Figure 6: Photos of Deck Crews at Work in Close Proximity to Jet Noise

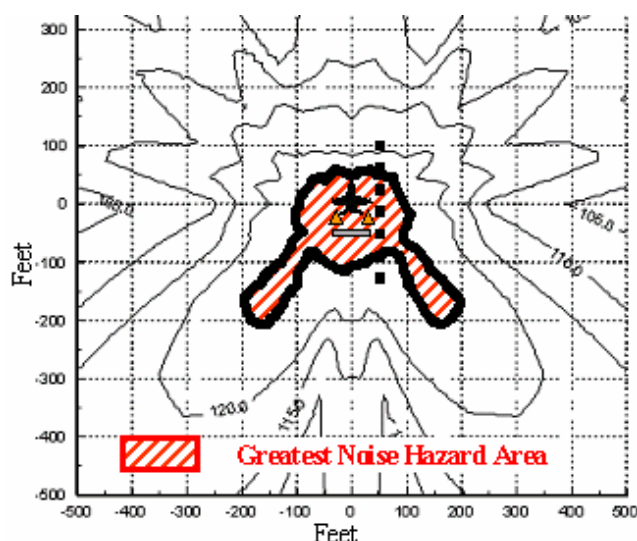


Figure 7: F-18C Noise Contours and Hazard Area

Aircraft personnel work long shifts (10-16 hours per day) in close proximity to high-level engine noise. If worn correctly, current double hearing protection of earplugs and earmuffs provides approximately 30 dB of noise attenuation protection.^[3,4] It has long been known that earplugs and earmuffs worn together offer greater protection than either item individually but less than the summation of the two devices.^[3,4,5] Double hearing protection commonly used by U.S. Navy flight deck crews provides adequate noise attenuation in jet noise environments (130-150 dB) when worn correctly and when total daily noise exposures limit crews to an 8-hour TWA of 85 dBA or less. For example, a 30 dB sound protector would allow the user less than 5 minutes total daily exposure in a 135 dB noise field.^[7] A Final Checker will exceed the safe daily noise exposure limit with just one or two high-performance jet aircraft launches. Additionally, after long flight deck duty days in jet aircraft noise, there are few, if any, quiet spaces below 84 dBA for flight deck crews' hearing to recover.^[22] The most prevalent U.S. Department of Veterans Affairs disability claim is hearing loss. For all military departments combined, hearing loss claims totalled over \$633M in 2004, over \$6.7B since 1977, and the trend is upward. These costs only include disability compensation payments and do not include the cost of treatment, audiograms, hearing aides, retraining, etc. The U.S. military total costs associated with hearing loss have been estimated at \$2-3B per year.^[22] The U.S. Navy and Marine Corps portion of these compensation costs is approximately 25-30%.^[21]

2.0 PURPOSE

The purpose of this survey was to estimate noise attenuation provided by helmets, earmuffs, and earplugs as used by U.S. Navy flight deck crews and to check the level of compliance with hearing conservation instructions. This survey was one part of a larger effort to determine both non-material (e.g., training, enforcement, crew rotation) and material (e.g., technological, pharmacological) intervention routes to improve hearing protection for U.S. Navy^[15] and U.S. Air Force aviation personnel.

3.0 METHODS

A core data collection team comprised of experienced life support equipment developers and anthropometrists collected survey data by questionnaire interview, inspection, and anthropometric measures. The survey included queries of personnel demographics (e.g., age, gender, rank, etc.), the type of hearing protection worn, and how hearing protection was selected, worn, and maintained. The survey protocol was approved by a U.S. Navy human-use review board and included gathering Informed Consent from each survey participant.^[11] The goal was to survey at least 300 personnel across six ships: two aircraft carriers per U.S. Second (Atlantic) and Third (Pacific) Fleets, and an amphibious assault ship from each fleet as well, i.e., three ships per coast.

3.1 Subject Selection

The survey was not to interfere with normal duties; therefore, subjects were not selected for discriminating variables like rank, duty station, age, gender, race, or anthropometric dimensions (e.g., head circumference). Rather, subjects volunteered or were ordered by superiors to participate (the survey included a question on reason for the participation). U.S. Navy Bureau of Naval Personnel data were used to assure the subject population was representative of actual flight deck personnel distributions for gender and military rank.

3.2 Anthropometry

Three common head dimensions were measured on each subject using spreading calipers and measuring tapes: bi-temple breadth, head breadth, and head circumference.

3.3 Earplug Use and Insertion Depth

For this survey, subjects who reported that they were earplug users were asked to insert Aero E·A·R™ expanding yellow foam earplugs in both their ears. After waiting several minutes for the earplugs to fully expand, the earplugs were marked around their circumference at the opening of the ear canal. Earplugs were removed, allowed to expand fully and then left and right earplug insertion depths were measured as the distance from the inserted earplug tip to the ink marking.

3.4 Cranial Helmet Size, Fit, and Maintenance

The following was recorded to assess overall cranial helmet fit for each subject: size of cranial worn; earcup position over left and right ears; chinstrap length; suitability of cranial helmet position on the head and relative to the brow ridge (glabella). Each cranial was also inspected, particularly for the condition of the earcup cushions and earcup foam inserts.

3.4.1 Headband Clamping Force

Each subject's cranial headband force was measured as a practical way to estimate earmuff headband condition and ability to press the earcups tightly to the head. Headband clamping force was measured using an Inspec Laboratories Ltd (Salford, UK) tension rig (see Figure 8) and following ANSI S12.6 methods (set 145 mm bitracion breadth and 130 mm head height).^[2]



Figure 8: Inspec Laboratories Ltd Tension Rig to Measure Earmuff Headband Clamping Force

3.4.2 Hair and Helmet Fit

Subject hair type was recorded (i.e., thick/thin, coarse/fine, curly/straight, close-cut/bald). Several measures were taken on females with long hair to estimate the girth added by tied-up hair under the cranial helmet: (1) head circumference under hair buns/braids, (2) head circumference including hair buns/braids at what appeared to be the greatest hair volume, and (3) distance up or down from the head circumference path, at what appeared to be the greatest hair volume.

3.4.3 Eyeglass Temples

The style of eyeglasses worn by subjects was noted. Additionally, the distance from the side of the head to the outer surface of the eyeglass temples where they passed under the earcups was estimated by subtracting bi-temple breadth from eyeglasses bi-temple breadth and dividing by two. Temple height under the earcup cushion was also measured using calipers. Figure 9 provides a diagram of eyeglass temple measurements.

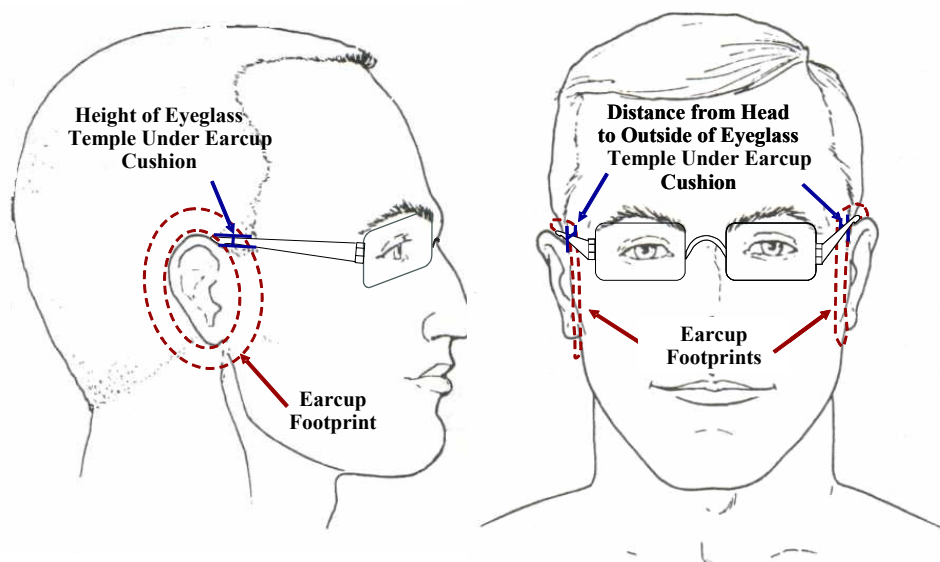


Figure 9: Eyeglass Temple Dimensions under Earcup Cushion

4.0 RESULTS AND DISCUSSION

Over 300 U.S. Navy flight deck personnel were interviewed and measured from six Atlantic and Pacific Fleet aircraft carrier and amphibious ship flight decks to assess how well hearing protection devices were used, fit, and maintained. This survey identified numerous ways to improve hearing conservation policy, policy implementation, and hearing protection designs.

4.1 Subjects

A total of 301 subjects (34 female, 267 male) were measured and interviewed in this survey. Tables 1-5 provide descriptive statistics for subject age, flight deck experience, job type / location, rank, and ship type.

U.S. Navy Flight Deck Hearing Protection Use Trends: Survey Results

Table 1: Subjects' Age and Flight Deck Experience

	Mean	Median	Std Dev	Min	Max
Age (years)	24	22	5	18	42
Flight Deck Experience (months)	34	24	37	1	204

Table 2: Subjects' Job Type (Shirt Color)

Shirt Color	N	Percent
Red	37	12.3
Blue	81	26.9
Green	47	15.6
Yellow	63	20.9
Brown	6	2.0
Purple	52	17.3
White	15	5.0

Table 3: Subjects by Job Location

Job Location	N	Percent
Checker/Shooter	11	3.7
Jet Blast Deflector	3	1.0
Catapults	28	9.3
Arresting Gear	5	1.7
Fire Crew	30	10.0
Safety	6	2.0
Chocks & Chains	61	20.3
Tractor	14	4.7
Plane captain	5	1.7
Aircraft Director	57	18.9
Fuel	52	17.3
Other	29	9.6

Table 4: Subjects by Rank

Rank	N	Percent
E1	3	1.0
E2	19	6.3
E3	152	50.5
E4	74	24.6
E5	26	8.6
E6	20	6.6
E7	4	1.3
O3	3	1.0

Table 5: Subject by Ship Type and Fleet

Ship Type	Location	N	Percent
LHA/LHD*	Atlantic	63	20.9
LHA/LHD	Pacific	53	17.6
CVN*	Atlantic	61	20.3
CVN	Pacific	53	17.6
CVN	Pacific	55	18.3
CVN	Pacific	16	5.3
*CVN – Aircraft Carrier			
*LHA/LHD – Amphibious Assault Ship			

U.S. Navy Flight Deck Hearing Protection Use Trends: Survey Results

Subjects represented the typical flight deck population distribution for rank and gender (according to unpublished data solicited for this survey from the U.S. Navy Bureau of Naval Personnel statistics). Additionally, 63% of the subjects reported spending at least 11 hours on the flight deck during a typical shift, while 29% reported durations between 6 and 10 hours per day.

4.2 Anthropometry

Head breadth and head circumference were collected on 285 subjects. These data were compared to head anthropometry data collected in 2002 on 747 U.S. Navy aircrew and flight deck crew personnel.^[13] Subjects interviewed in this survey were similar (could find no statistically significant difference) for head breadth and circumference to those measured in the referenced study. Figure 10 and Table 6 report the similarities between the two data sets.

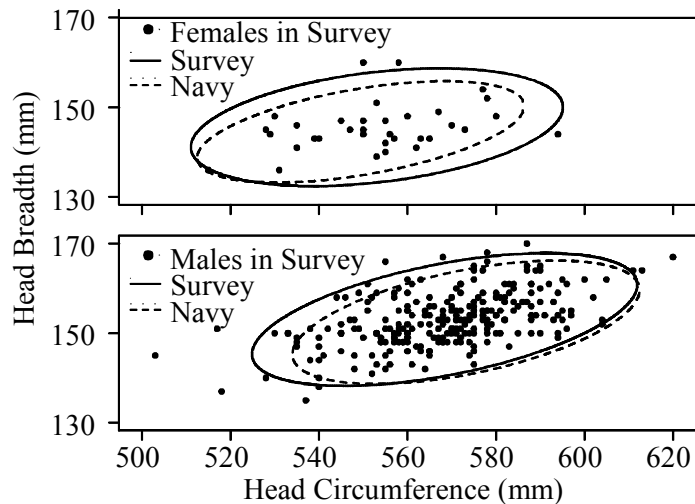


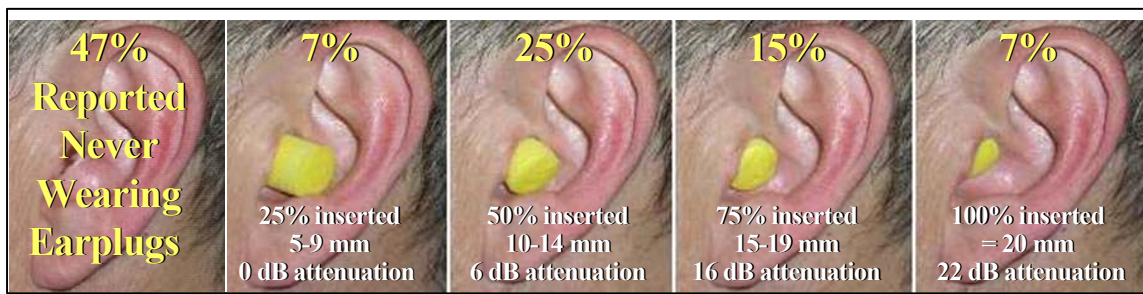
Figure 10: Estimated 95th Percentile Regions for Each Gender and Data Set

Table 6: Descriptive Statistics for Gender for Both Surveys

Gender	Data	Measure	N	Values in mm				
				Mean	Median	Min	Max	Std
Female	Navy	Breadth	220	144.5	144	132	158	4.6
		Circumference	220	549.4	548	514	593	15.1
	This Survey	Breadth	34	145.6	145	136	160	5.4
		Circumference	34	553.1	554	515	594	17.2
Male	Navy	Breadth	521	152.4	152	132	170	5.6
		Circumference	521	573.2	572	523	618	16.0
	This Survey	Breadth	251	153.1	153	135	170	6.1
		Circumference	251	568.4	570	503	620	17.8

4.3 Earplug Use and Insertion Depth

The most significant finding of this survey was that 79% of the ears of flight deck personnel interviewed received an estimated 0-6 dB of noise attenuation from either shallow earplug insertion depths or never wearing earplugs (47% self-reported never wearing earplugs). Only 14% reported always wearing earplugs beneath their cranials, i.e., 14% reported wearing the required double hearing protection. Further, of those who reported sometimes or always wearing earplugs, few inserted the earplugs deeply enough to benefit fully. Figure 11 shows how noise attenuation provided by an expanding foam earplug is directly proportional to its insertion depth.^[5] Figure 11 also shows the percentage of earplugs inserted to each depth. For example, only 7% inserted the earplugs deeply enough to achieve 22 dB noise attenuation in both ears.



Noise Reduction Ratings from Air Force Research Laboratory earplug insertion depth study using American National Standard S12.6-1997 (R2002) Methods for Measuring the Real-Ear Attenuation of Hearing Protectors, Method A (Experimenter Supervised / Verbally Coached), mean minus two standard deviations.

Figure 11: Earplug Insertion Depth, Related Noise Attenuation, Percentage of Earplugs at Each Depth
[extrapolated from 202 ears of sometimes and always earplug users]

Table 7 provides earplug insertion depth summary statistics. No significant difference could be found between left and right earplug insertion depth, suggesting handedness is not a factor, i.e., right handedness predominates yet right earplug insertion was not deeper than left. Some 85% of subjects' left and right earplugs were ≤ 2 mm different for insertion depth (8 mm was the maximum difference between left and right earplug insertion depth). No correlation could be found between having similar left / right earplug insertion depths and achieving an optimum earplug insertion depth of ≥ 20 mm. For example, one subject had a 10 mm left earplug insertion and a 12 mm right earplug insertion; the insertion depths were within 2 mm of each other, yet both were shallow and only achieved approximately 6 dB attenuation.

Table 7: Descriptive Statistics for Earplug Depth Left and Right Earplugs Accounted Separately

	Mean	Median	Std. Dev.	Min.	Max.
Earplug depth (mm)	13.6	13	3.9	7	22

Of the subjects who reported wearing earplugs at least sometimes, 75% reported wearing E·A·R Classic™ earplugs; 75% reported they replaced their earplugs daily, while 24% reported replacing them when they appeared soiled. Table 8 ranks earplug use by both job type and job location. These data indicate personnel in some of the most hazardous jet engine noise locations, such as Aircraft Directors and Jet Blast Deflector personnel, are least likely to wear earplugs.

U.S. Navy Flight Deck Hearing Protection Use Trends: Survey Results

Table 8: Earplug Usage by Job Type and Job Location

Job Type	Job Location	N	Wear Earplugs			
			Always	Sometimes	Never	Percent Never
Green	Safety	1	1	0	0	0.0
Yellow	Catapults	1	1	0	0	0.0
White	Safety	5	4	1	0	0.0
Red	Other	8	3	5	0	0.0
Green	Checker	3	1	2	0	0.0
Blue	Fire Crew	1	0	1	0	0.0
Brown	Other	1	0	1	0	0.0
Yellow	Tractor	1	0	1	0	0.0
White	Checker	8	4	3	1	12.5
Green	Other	8	1	6	1	12.5
Blue	Other	6	1	3	2	33.3
Red	Fire Crew	29	4	14	11	37.9
Brown	Plane Captain	5	1	2	2	40.0
Blue	Chocks and Chains	60	9	23	28	46.7
White	Other	2	0	1	1	50.0
Yellow	Other	4	0	2	2	50.0
Purple	Fuel	52	5	19	28	53.8
Green	Catapults	27	4	8	15	55.6
Green	Arresting Gear	5	0	2	3	60.0
Blue	Tractor	13	0	5	8	61.5
Yellow	Aircraft Director	57	4	15	38	66.7
Green	Jet Blast Deflector	3	0	1	2	66.7
Total		300*	43	115	142	47.3
* Does not total 301 because one subject did not report earplug use.						

Figure 12 shows earplug use habits across various subject groupings: ship type, ship location, gender, cranial helmet issuing, age, rank, daily time on the flight deck, overall flight deck experience, job type and location. Chi-square testing determined whether the frequency of those who always, sometimes, or never used earplugs varied significantly between these groups. No significant differences were found between subjects, fleets, or ship types for hearing protection use, fit confirmation, and maintenance. The only significant differences found ($p \leq 0.05$) between these groups were for job type and job location. For job type, White Shirts used earplugs more often than others (Appendix A provides typical jobs associated with specific shirt colors worn on U.S. Navy flight decks). For job location, Safety and Final Checker personnel used earplugs more often than others.

U.S. Navy Flight Deck Hearing Protection Use Trends: Survey Results

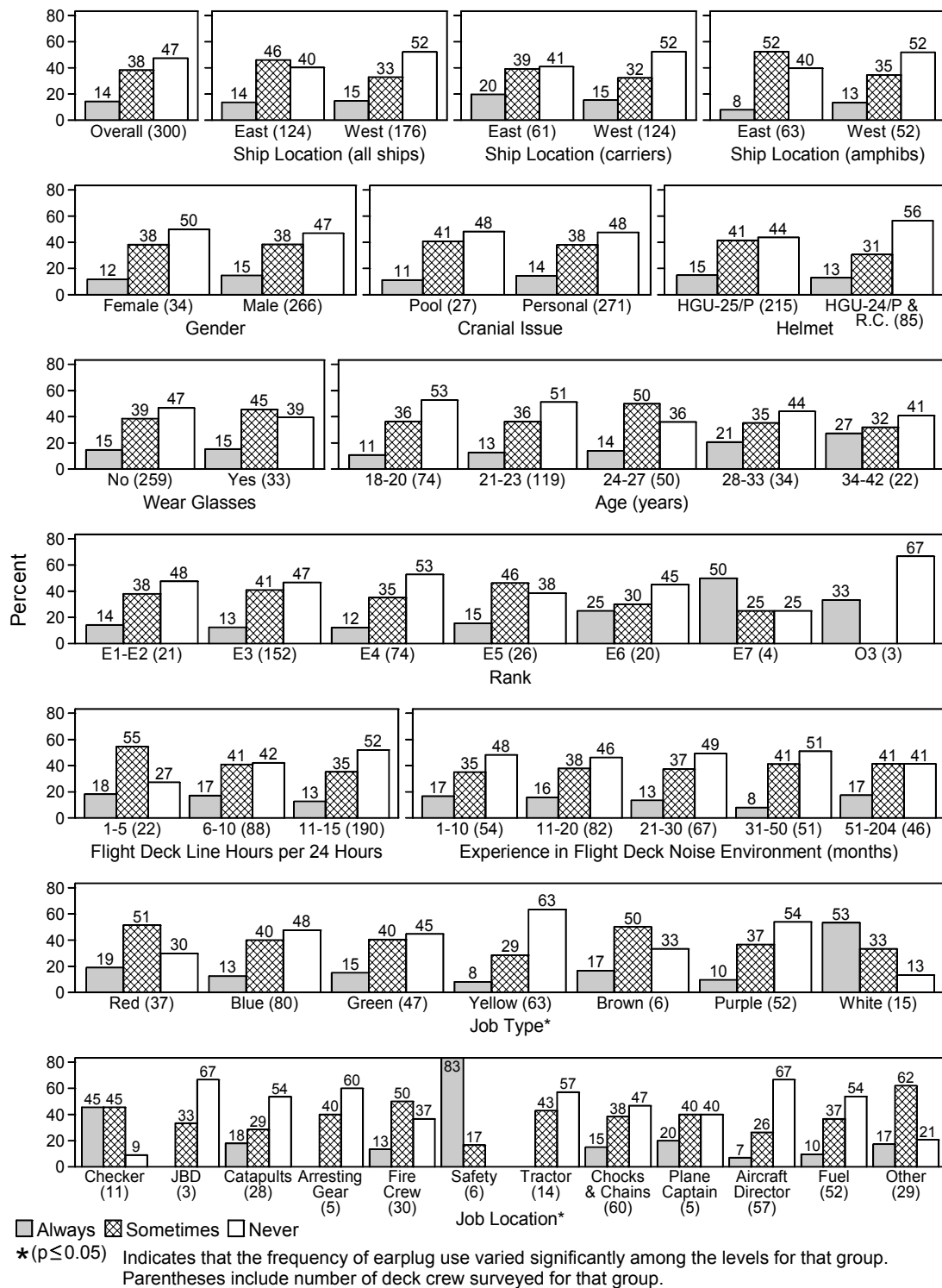


Figure 12: Percentage of Earplug Usage for Various Subject Groupings

U.S. Navy Flight Deck Hearing Protection Use Trends: Survey Results

4.4 Air (Acoustic) Leaks Under Earcups

The cranial/earmuffs, worn without earplugs, have been reported to provide approximately 21 dB of noise attenuation^[3] when correctly fit, worn, and maintained. It has been reported since the 1950s that air (acoustic) leaks between earcups and wearers' heads can reduce noise attenuation 3-15 dB across a broad range of frequencies, predominantly in the lower frequencies.^[5,23] The following criteria are important to maximize cranial/earmuff noise attenuation; these were assessed in the survey and findings are reported below.

- Sized correctly
- Adjusted to fit comfortably and to ensure earcup-to-head seal is not disrupted by items such as thick hair, eyeglasses, caps, ear warmers, etc.
- Well maintained so earcup cushions and earcup foam inserts are soft and pliable
- Headband clamping force falls in the range of 6-21 N^[1,6,18]

4.4.1 Cranial Sizes Issued and Fit Observations

Subjects wore one of three cranial helmet types: the HGU-24/P (sound powered), HGU-25(V)2/P, or the Radio Cranial (see Figure 3 above). Table 9 lists the distribution of the cranials inspected in this survey.

Table 9: Distribution of Cranials Inspected

Cranial Helmet Type	N	Percent
HGU-24/P (sound powered)	17	5.6
HGU-25(V)2/P	216	71.8
Radio Cranial	68	22.6

Of the four cranial cloth sizes, 70% of the subjects were issued the largest size (7½); 68% did not know cranials came in sizes or what size they should wear; and 67% took whatever size they were issued. Most ships stocked only the two largest sized cranials (7¼ and 7½). According to a common hat size chart (Table 10), 75% were issued a questionable size of cranial, e.g., 13% of subjects issued the largest cranial size (7½) may have been better fit in the smallest cranial size (6¾).

Table 10: Common Hat Size Chart
(Used by survey team to estimate the cranial size that should have been issued.)

Head Size		Hat Size
in.	cm	
21	53	6⅞
21½	54	6¾
21⅝	55	6⅞
22⅛	56	7
22½	57	7⅛
23	58	7¼
23⅜	59	7⅝
23¾	60	7½
24	61	7⅝
24½	62	7¾
25	63	7⅞

Despite these possible cranial size issues, the survey team reported that 90% of all gear appeared to fit well. Inspection of the earcup cushion seal around the ears indicated that 73% had both ears inside the earcup cushions, indicating a good earcup fit. The other 27% had at least one earlobe trapped under the bottom of the earcup cushion, possibly causing an acoustic leak and degraded attenuation.^[5,23] Chinstrap length was rated as good for 94% of the subjects (the chinstrap may have been too long for 4% and too short for 2%). Cranial position in the fore - aft direction was measured from the brow (glabella) to the leading edge of the cloth liner (Table 11 provides summary statistics). These cranial position data do not correspond to existing fit instructions; they provide a relative indicator of earmuff and headband fore - aft rotation on the head and around the ears. For 9% of subjects, the cranial appeared to sit too high on the head, while 1% appeared to sit too far back, forward, or low on the head.

Table 11: Distance from Brow to Cranial Cloth Front Edge

	Mean	Median	Std. Dev.	Min.	Max.
Brow to Cloth Distance (mm)	40.4	39	18.4	0	104

4.4.2 Cranial Maintenance

The survey team determined that approximately 41% of the earcup cushions and/or earcup foam inserts were in unsatisfactory condition, despite 73% reporting they inspected or looked over their cranials at least daily. Figure 13 shows a cranial with poor earcup cushions and missing earcup foam inserts. Another 15% reported they never inspected their cranials. Some 9% reported that they had to share or “hot swap” cranials throughout daily missions. These shared gear tended to be the most soiled and least maintained. Noise attenuation is likely reduced in poorly maintained cranials with deteriorated, flat, hard, and missing earcup cushions and earcup foam.^[5]



Figure 13: Hard, Creased Earcup Cushions and Missing Earcup Foam Inserts

4.4.3 Headband Clamping Force

Headband clamping force has been linked to wearer comfort and to noise attenuation. Previous research indicated that new headbands generally provide 6-21 N of clamping force but that 17 N approximates the upper limit for comfort when wearing traditional earmuff designs.^[1,5,6,18] Comfort relates to pressure (headband force per earcup seal contact area); however, higher headband clamping forces are thought to reduce acoustic leaks between the earcup seal and the wearer’s head.^[1,5,6,18,23] As headband clamping force commonly reduces with routine use, age, and active headband stretching by wearers,^[5] this survey included clamping force measurements on all cranials/earmuffs. Table 12 provides clamping force data collected on earmuffs while in the cranials. These clamping forces are consistent with previously reported headband forces and indicate that these headbands may not have been age-fatigued or excessively spread open to loosen them.

Table 12: Cranial Headband Clamping Force

	Mean	Median	Std. Dev.	Min.	Max.
Tension (Force in N)	15.3	15.5	2.9	6.5	23.2

The cranial was reported to be comfortable enough to wear at least 10 hours at a time by 30% of the subjects. However, 55% reported that after a few hours of wearing the cranial they felt major discomfort. Another 15% reported that the cranial caused severe discomfort within a few hours of donning (the subject with the maximum clamping force (23.2 N) reported immediate, severe discomfort.)

U.S. Navy Flight Deck Hearing Protection Use Trends: Survey Results

4.4.4 Effect of Hair and Other Items on Fit

Many personnel who had to share cranials wore a bandana or skullcap under their cranial as a hygiene barrier. Others reported that they wore items like bandanas to absorb sweat in hot weather and winter caps and ear warmers to keep warm in cold weather. These barriers generally passed under the earcup cushions (see Figure 14), and while attenuation was not measured in this survey, air (acoustic) leaks can reduce earcup attenuation.^[5,23]



Figure 14: Survey Subjects Wearing a Bandana, Long Hair Tied Up, Eyeglasses, and a Winter Hat under Cranial Earcups

Males are not permitted to have long hair in the U.S. Navy; however, even short hair can disrupt the earcup sealing to the head. For this reason, subject hair type was recorded (i.e., thick/thin, coarse/fine, curly/straight, close-cut/bald). The survey team did not note that any hair type correlated to a cranial fit issue. Female subjects who had long hair and wore it tied up added 2 cm on average to head circumference measures. Tied up hair, hair type, length, or style was not reported to impact cranial fit for female subjects. This may be due to the largest size cranials being issued and these sizes being large enough to accommodate the added hair mass and bulk and/or the hair buns passing inferior to the bottom edge of the cranial and fitting in the nape of the neck.

4.4.4.1 Eyeglass Temples

Eyeglass temples passing between the earcup cushion and the wearer's head creates a noise leak pathway that has been linked to a 3-7 dB reduction in noise attenuation.^[1,5,6,18,23] Eyeglasses use was reported by 11% of the subjects: 72% wore standard type frames, 21% cable-type, and 7% wore some other type. Table 13 provides the three additional eyeglass-related measurements (described in Methods). Figure 15 shows a subject with eyeglass temple worn under earcup cushion.

Table 13: Descriptive Statistics for Measures Taken on Eyeglass Wearers

At Intersection with Earcup Footprint	Mean	Median	Std. Dev.	Min.	Max.
Temple Breadth (mm)	140.8	140	6.7	127	156
Eyeglass Breadth (mm)	150.4	150	6.8	140	167
Eyeglass Temple Height (mm)	3.3	3	1.4	1.5	8

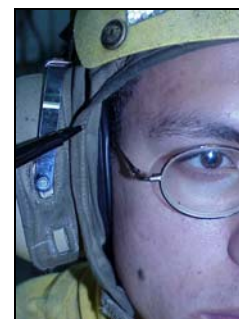


Figure 15: Eyeglass Temple Worn under Earcup Cushion

5.0 SUMMARY AND CONCLUSIONS

The most significant finding of this survey was that 79% of the ears of flight deck personnel interviewed received an estimated 0-6 dB of noise attenuation from either shallow earplug insertion depths or never wearing earplugs. Some 47% self-reported never wearing earplugs while just 14% reported always wearing earplugs with their cranials (which is the required double hearing protection). Further, of those who reported sometimes or always wearing earplugs, only 7% inserted the earplugs deeply enough to achieve the estimated 22 dB noise attenuation in both ears.

The cranial and earmuffs have been reported to provide approximately 21 dB of noise attenuation^[3] when worn without earplugs, fit correctly, and maintained. This survey found that 75% of subjects may not have had a well-fit cranial helmet and that 41% of the earmuffs needed new earcup cushions and/or foam inserts. This survey also identified numerous items worn under earcup cushions that may have led to acoustic leaks and reduced noise attenuation^[1,5,6,18,23], items such as caps, ear warmers, and eyeglasses. However, earmuff headband clamping forces were within normal expected ranges (6-23 N).

U.S. Navy flight deck noise levels (up to 150 dB) and personnel exposure durations (most over 11 hours per day, 7 days a week) are among the worst in the world. The double hearing protection available to flight deck crews can provide approximately 30 dB of noise attenuation,^[3,4,5] however, this survey identified numerous practices that likely reduce this level noise attenuation.

6.0 RECOMMENDATIONS

Analysis of the data collected in this survey identified technological and non-technological (e.g., training, enforcement) ways to improve hearing protection for U.S. Navy flight deck personnel. Based on these survey findings, the following is recommended.

- Improve helmet, earmuff, and earplug noise attenuation performance to extend the daily safe noise exposure time limit
 - Advance attenuation technology for high, repeatable performance
 - Make helmet/earmuffs personal issue equipment, i.e., do not share hearing protectors
 - Instruct users how to select the correct size and to correctly wear and maintain helmets, earcups, and earplugs
- Design earplugs that encourage the use of earplugs and correct earplug insertion
- Set, distribute, and enforce consistent hearing conservation procedures
 - Make hearing protection part of the uniform
 - Associate cost/penalty for not complying
 - Issue Surgeon General policy statement to all Ship Commanding Officers that Hearing Conservation Instructions must be enforced

Areas for additional research include the following:

- Investigate why personnel are not wearing double protection – i.e., not wearing earplugs
- Determine effective level for supervisory control for hearing protection
- Investigate the effects on attenuation of hair, caps, eyeglasses, etc., under earcups and helmets
- Measure/track effect of new hearing protection technologies and policies
- Survey U.S. Air Force flight line personnel hearing protection use

7.0 ACKNOWLEDGMENTS

The authors acknowledge the technical input from their teammates on U.S. Defense Technology Objective HS.33 Improved Aviation Personnel Hearing Protection, especially Dr. James B. Sheehy, ST (Naval Air Systems Command) and Mr. James A. Janousek (Naval Air Systems Command). Special recognition and gratitude go to Mr. James A. Rodriguez (Naval Air Systems Command) who provided invaluable fleet and carrier liaison during each survey trip. Thanks are due also to Mr. Chuck Goodyear (General Dynamics) and Maj. Thia Eades (while at Air Force Research Laboratory) for strengthening the survey approach and providing statistical analyses and data presentation advice, and to survey team members: Scot Best (Naval Air Systems Command); Scott Fleming (Air Force Research Laboratory); Steve Paquette (U.S. Army Soldier Center); Dr. William Marshak (Sytronics, Inc.) and Dr. Jeffrey Buchholz (Micro Optics).

8.0 REFERENCES

1. Acton, W.I., Lee, G.L., and Smith, D.J. (1976). "Short Communication – Effect of Head Band Forces and Pressure on Comfort of Ear Muffs," *Ann. Occup. Hyg.* 19, 357-361.
2. ANSI S12.6-1997 (R2002) American National Standard Methods for Measuring the Real-Ear Attenuation of Hearing Protectors.
3. Authorized Hearing Protection Devices, Naval Environmental Health Center Website <http://www-nehc.med.navy.mil> (search on *Hearing Protection Devices*).
4. Berger, E.H. (1983). "Laboratory Attenuation of Earmuffs and Earplugs Both Singly and in Combination," *Am. Ind. Hyg. Assoc. J.* 44(5), 321-329.
5. Berger, E.H., Royster, L.H., Royster, J.D., Driscoll, D.P., Layne, M. (Eds.). (2000). The Noise Manual (5th ed.) Am. Ind. Hyg. Assoc. Fairfax, VA.
6. Casali, J.G. , Grenell, J.F. (1989). "An Exploratory Study of Moderate Physical Activity and Selected Design Attribute Effects on Earmuff Attenuation," *Am. Ind. Hyg. Assoc. J.* 50(9), 480-485.
7. Department of Defense Instruction Number 6055.12: DoD Hearing Conservation Program
8. E·A·R Earplug Manufacturer Website <http://209.200.67.149/e-a-r.com/default2.cfm>, Aearo Company, 5457 West 79th Street, Indianapolis, IN 46268, Telephone (317) 692-6666.
9. Johnson, D.L., Nixon, C.W. Skelton, M. (1983). "The Effect of Reduced Headband Force on the Attenuation of Muff-Type Protectors," *J. Acoust. Soc. Am. Suppl.* 1, 74, S94.
10. MCO 6260.1 Marine Corps Hearing Conservation Program. (Available Online.)
11. Naval Air Warfare Center Aircraft Division Institutional Review Board Memorandum 6500, 460000A/1218 dated 20Sep02, Use of Human Subjects in Minimal Risk Protocol "Fleet Hearing Protection Survey" DoD #30116 approved.

12. NAVAIR Technical Manual 13-1-6.7-3, Aircrew Personal Protective Equipment, Chapter 7, HGU-24/P AND HGU-25(V)2/P Cloth Helmet Assemblies, May 1999 (Change 5 - 1 Sep 2002).
13. Navy Head Anthropometry Database of 747 Navy Personnel. Traditional Measurements and 3-D Scan Data at <https://pma202.navair.navy.mil/> [Join, Login, Team Sites, Science and Technology, Environmental & Hearing Protection, Documents, Open "3D_USN-USMC_HEADSCAN"], Janousek, J. (301) 342-9210.
14. NEHC TM 6260.51.99-1 Navy Medical Department Hearing Conservation Program Procedures. (Available Online.)
15. Office of Naval Research Program Element 060372N Project R2914, Improved Aviation Hearing Protection.
16. OPNAVINST 5100.19 Navy Occupational Safety and Health Program Manual, NAVOSH and Major Hazard-Specific Programs. (Available Online.)
17. OPNAVINST 5100.23 Navy Occupational Safety and Health Program Manual, Chapter 18, Hearing Conservation And Noise Abatement. (Available Online.)
18. Paakkonen, R. (1992). "Effects of Cup, Cushion, Band Force, Foam Lining and Various Design Parameters on the Attenuation of Earmuffs," Noise Control Eng. J. 38(2), 59-65.
19. SAE-AS23899 Specification, Sound Aural Protector (30 Aug 99) for use by the Department of Defense (DoD).
20. Sheehy J.B., Bjorn V.S., McKinley R.L. (2001) "Appendix C Joint Strike Fighter Noise Emissions and Hearing Protection S&T Review - Executive Summary, Office of the Deputy Under Secretary of Defense (Science and Technology) Congressional Report on Tactical and Support Aircraft Noise Reduction: Science and Technology Funding and Plans," Response to the National Defense Authorization Act of Fiscal Year 2001 (Report 106-616).
21. U.S. Army Center for Health Promotion and Preventive Medicine Hearing Conservation Program Website <http://chppm-www.apgea.army.mil/hcp/default.aspx>. [Compensation Reports].
22. Yankaskas, K.D., (2001) "Landing on the Roof," International Military Noise Conference Website http://chppm-www.apgea.army.mil/imnc/Hearing_Conversation/Yankaskas.ppt.
23. Zwislocki, J., (1957) "Ear Protectors," Handbook of Noise Control. McGraw-Hill, New York, 8 (1-27).

Appendix A

Shirt Colors on U.S. Navy Aircraft Carriers

From <http://www.chinfo.navy.mil/navpalib/ships/carriers/rainbow.html>

Blue

Plane Handlers
Aircraft Elevator Operators
Tractor Drivers
Messengers and Phone Talkers

Red

Ordnancemen
Crash and Salvage Crews
Explosive Ordnance Disposal

Brown

Air Wing Plane Captains
Air Wing Line Leading Petty Officers

Yellow

Aircraft Handling Officers
Catapult and Arresting Gear Officers
Plane Directors

White

Air Wing Quality Control Personnel
Squadron Plane Inspectors
Landing Signal Officer
Air Transfer Officers
Liquid Oxygen Crews
Safety Observers
Medical Personnel

Green

Catapult and Arresting Gear Crews
Air Wing Maintenance Personnel
Cargo-handling Personnel
Ground Support Equipment Troubleshooters
Hook Runners
Photographer's Mates
Helicopter Landing Signal Enlisted Personnel

Purple

Aviation Fuels

Noise-Induced Hearing Loss in Canadian Military Personnel

Sharon M. Abel

Defence R&D Canada - Toronto
1133 Sheppard Ave. West, PO Box 2000
Toronto, Ontario M3M 3B9 Canada

Email: sharon.abel@drdc-rddc.gc.ca

ABSTRACT

The escalating cost of claims for noise-induced hearing loss in the Canadian Forces (CF) prompted the review of current hearing conservation practices. As a first step, a prospective study was conducted to assess risk factors for the development of hearing loss in a wide range of military trades. A total of 1,057 individuals working in land, sea and air environments at five CF military bases contributed their current hearing test results and first hearing test results on record. They also completed a 56-item questionnaire relating to demographics, occupational and non occupational noise exposure history, training in and utilization of personal hearing protection, and risk factors other than noise which might affect hearing, including head injury, ear disease, medications, and solvent exposure. Military medical personnel recruited the subjects, distributed the questionnaires and assessed hearing. Apparatus and protocols for the latter conformed to current clinical practice. The results showed that the prevalence of moderate to severe hearing loss progressed with years of noise exposure, with hearing thresholds in those over 45 years ranging broadly from normal to over 70 dB HL. Unprotected exposure to solvents and leisure noise appeared to be significant determinates of adverse outcome, while the effects of head injury, history of ear disease, and the use of medications were minimal. The survey suggested that training on the hazards of noise exposure and the selection and utilization of hearing protection were inadequate. Hearing protection was reported to be incompatible with other gear, uncomfortable and an impediment to communication.

1.0 INTRODUCTION

The Canadian Forces introduced a hearing conservation program in the 1950s [1]. Components include noise measurement, the reduction of noise at the source where possible, education on the hazards of noise exposure, utilization of personal hearing protection and the regular monitoring of hearing. Statistics compiled by Veterans Affairs Canada show that, in spite of these safeguards, the cost of claims for noise-induced hearing loss, is escalating. Currently, the annual payout for a force strength of about 50,000 individuals is more than double the amount cited by the Workplace Safety and Insurance Board of Ontario [2] for a provincial labour force of about 2 million in the manufacturing, construction, agriculture and mining sectors [3]. These data support the need to review and upgrade current hearing conservation practices.

Unprotected exposure to high-level sound results in permanent hearing loss, generally referred to as either noise-induced hearing loss (NIHL) or noise-induced permanent threshold shift (NIPTS). The magnitude of the hearing threshold elevation relative to normal hearing thresholds will depend on the level of the sound to which the individual is exposed, the number of hours of exposure per day, and the use of personal hearing

Abel, S.M. (2005) Noise-Induced Hearing Loss in Canadian Military Personnel. In *New Directions for Improving Audio Effectiveness* (pp. 2-1 – 2-24). Meeting Proceedings RTO-MP-HFM-123, Paper 2. Neuilly-sur-Seine, France: RTO. Available from: <http://www.rto.nato.int/abstracts.apa>.

Noise-Induced Hearing Loss in Canadian Military Personnel

protection. For federal employees in Canada the level of continuous steady-state noise that is considered safe for hearing is 87 dBA for an 8-hour work day [4]. Based on the equal energy principal, the exposure duration should be halved for every 3 dB increment in level [5].

According to Franks et al. [6], NIHL is among the most prevalent work-related medical conditions in both the United States and Europe. Biological changes from metabolic stress and mechanical damage following noise exposure are complex and include pathology of neural, sensory, supporting and vascular cells of the peripheral end organ for hearing, i.e., the cochlea [7]. Damage to hearing caused by limited exposures may result in a temporary threshold shift (TTS) that gradually disappears over a number of hours of relative quiet. However, continued or repeated exposures will result in irreversible sensory hair cell loss and concomitant permanent sensorineural, predominately high tone, hearing loss [8]. Although the high-level sounds to which the individual is exposed may have a broad frequency spectrum, the most vulnerable frequency region for hearing loss is 4-6 kHz. This outcome reflects the natural resonance of the ear canal at 3.8 kHz [9]. Hearing loss in this region is clearly evident after 3-5 years of exposure [10]. The hearing threshold “notch” in the audiogram in this frequency region deepens with continued exposure. Hearing loss gradually spreads to lower and higher frequencies. Individual differences are large, denoting a wide range of personal susceptibility [11,12].

High-level impulsive sounds from weapons are a particularly damaging source of noise exposure in military environments [13,14]. Peak levels may be as high as 185 dB SPL. Both the nature of the injury and the pattern of recovery differ substantially from that of continuous noise. The effectiveness of hearing protectors for preventing NIHL from small firearms and large-calibre weapons was investigated by Christiansson and Wintzell [15]. The subjects were 204 infantry officers who instructed conscripts in the use of firearms. Peak exposures levels ranged between 156-185 dB SPL. For the sample as a whole, the maximum loss was observed at 6 kHz. At this frequency the average hearing threshold was 25 dB HL (hearing level, i.e., loss, relative to normal threshold) by 35 years of age, and 40 dB HL at 45 years of age. The prevalence of a moderate hearing loss by mid-life in spite of hearing protector usage suggested that hearing conservation programs developed for application in industrial settings may not be adequate for military trades.

The assessment of hearing loss from occupational noise exposure may be confounded by a number of factors. Studies reported in the literature have focused on the interaction with aging, exposure to potentially ototoxic organic solvents (e.g., toluene and styrene), and exposure to high-level sound from non occupational sources (i.e., leisure noise).

1.1 Aging

The two main causes of hearing loss are aging and noise exposure. The International Standards Organization (ISO) provides databases for males and females who are otologically normal and highly screened for factors that may affect hearing, as well as for males and females from a typical unscreened population of an industrialized country [16,17]. These may be used as baselines against which to assess the effects of noise exposure. Hearing loss is modelled as an accelerating function of age while noise-induced hearing loss is modelled as a decelerating function of the exposure duration [18]. It may be possible to determine the primary cause of the hearing loss, age or noise, from the configuration of the audiogram. The effects of aging start at the highest frequencies and gradually progress to lower frequencies. In contrast, noise-induced hearing loss is initially evident as threshold notch in the region of 4-6 kHz that deepens with exposure and gradually spreads to both lower and higher frequencies [19].

1.2 Solvent exposures

Studies conducted in both animal models and humans suggest that various aromatic hydrocarbons, particularly toluene, xylene and styrene, may be ototoxic [20-23]. Mounting evidence also indicates that exposure to organic solvents potentiates NIHL [24]. In one experiment reported in the literature, Sliwinska-Kowalska et al. [25] compared hearing loss in workers exposed to either a mixture of organic solvents at concentrations below permissible levels in paint and lacquer industries, solvents in combination with noise levels greater than 85 dBA, or neither agent. Hearing loss was defined as an elevated hearing threshold in excess of 25 dB HL at any frequency from 1-8 kHz. The prevalence of hearing loss in the control group was 36% compared with 57.5% for solvent exposure and 61.5% for solvent plus noise exposure. Mean hearing thresholds for the noise plus solvent group were about 5 dB greater than those for the solvent alone group in the range of 2-4 kHz. Research in animal models have demonstrated outer hair cell damage for both solvent and noise, supporting the hypothesis of same site of injury [26].

1.3 Leisure Noise Exposure

There has been increasing concern in recent years about the potentially damaging effect of high-level sound exposure during leisure activities. The sources are diverse and include amplified music at rock concerts and disco/dance bars, motorcycles, snowmobiles, firearms, power and chain saws, and impulse generating toys such as cap pistols. Levels can be in excess of 100 dBA [27]. This concern has been validated by hearing tests conducted as part of medical screening for military recruitment. For example, in a random sample of 500 18-year old Swedish male conscripts, Axelsson et al. [28] found a 14% prevalence of hearing loss, defined as a threshold equal to or greater than 20 dB HL at any test frequency. Most often, the loss was in the region of 4-8 kHz. Seventy-nine percent of the sample reported that they listened to pop music often or very often, and 17% had attended more than 10 rock concerts. Twenty-one percent played a musical instrument. A considerably higher prevalence of high-frequency hearing loss was reported by Borchgrevink [29] for 18-year old Norwegian male conscripts. In the latter study, prevalence increased from 15% to 25% among a sample of about 30,000 individuals over a 10-year period from the early 1980's to the early 1990's. The author argued that the observed increase in prevalence reflected the increase in exposure to leisure noise, particularly rock music.

1.4 Impact on Communication, Task Performance and Health

The impact of noise exposure on human health is wide ranging and includes both auditory and non-auditory outcomes [30]. With respect to communication, studies have shown that individuals with high-frequency hearing loss will have difficulty with speech understanding in noisy backgrounds [31,32]. Smoorenburg [32] found that the speech reception threshold (SRT - the level at which 50% of sentences can be repeated correctly) in quiet was correlated with pure-tone hearing thresholds below 1 kHz and the SRT in noise to the thresholds at 2 kHz and 4 kHz. A noticeable handicap (decrement in SRT in noise) was apparent even when the hearing loss at these frequencies was relatively mild at 30 dB HL.

The presence of high-level background sound is also detrimental to the performance of diverse non-auditory tasks, including vigilance, selective attention, sensory-motor behaviour and recall [33,34]. For example, in a study of vigilance and time-sharing reported by Hockey [35], subjects performed a tracking task (designated high priority) while detecting the presentation of lights at different spatial locations (designated low priority). In a 100 dB-SPL noise background, deficits were observed for the low priority task, particularly for signals presented from peripheral positions. The effect of the noise was a reallocation of attention. Continuous high-level sounds have also been implicated in sleep disorders [36,37], cardiovascular disease [38] and mental

Noise-Induced Hearing Loss in Canadian Military Personnel

health [39]. In a series of laboratory sleep studies Thiessen and colleagues [40,41] found that, with increases in the level of taped traffic noise, the number of awakenings increased by 40% and the probability of a decrease to a lighter sleep stage increased by 70%. Cantrell [42] reported a rise in plasma cortisol and blood cholesterol levels that persisted for at least a week after the conclusion of an experiment in which subjects were confined to a noisy dormitory for 50 days and nights and exposed to high-frequency sounds ranging from 80 to 90 dB SPL. Research relating to noise and mental health have demonstrated that psychiatric symptoms and mental hospital admissions are more prevalent in the vicinity of airports [43].

2.0 RESEARCH DESIGN

A prospective cross-sectional study was carried out at five Canadian Forces military bases located in Ontario and Nova Scotia [44]. All personnel who were scheduled for a routine hearing test within a designated 12-month period were invited to participate at the time of their clinic visit. No restrictions were placed on age, gender, rank, military trade or length of military service. The only exception was that volunteers had to have completed at least one year of employment in the Canadian Forces prior to the study. No upper limit was placed on the number of participants. Those who volunteered contributed their most recent audiogram, as well as the first audiogram in the clinic chart. The latter was not necessarily the enrolment audiogram. However, at least one year must have elapsed between the first and current assessments.

Volunteers also completed a 56-item survey relating to risk factors for acquiring a hearing loss. The survey included questions about age, gender, previous trades and work-related noise exposure history, current noise and solvent exposures, training in and utilization of hearing protection and respirators, head injury, civilian noise exposure and perceived hearing handicap. Subjects were also asked about their health, and specifically about their history of ear infections and ear symptoms, head injuries, dizziness, and medications prescribed. The responses enabled the investigators to take into account various causes of any observed hearing loss.

3.0 METHODS AND MATERIALS

The study protocol was approved by the Defence Research and Development Canada Human Research Ethics Committee. Upon arrival at the clinic subjects were given the protocol to read. If interested in participating, they provided written consent and then completed the questionnaire while waiting for the hearing assessment. For the latter, subjects were tested individually in a sound proof booth (Eckel Industries, Model AB200). The apparatus included a Beltone audiometer (Model 119) and TDH-39 matched headset. Hearing thresholds were measured by trained clinic personnel once at each of seven pure tone frequencies (0.5, 1, 2, 3, 4, 6 and 8 kHz), in each ear. Thresholds were determined using the standardized method of adjustment [45]. The sound level was systematically increased and decreased in steps of 5 dB to find the critical value (i.e., the threshold) that separated the audible from the inaudible range. Following the hearing assessment, the subject's file (the Consent Form, completed questionnaire, and copies of the current audiogram and first available audiogram) were transferred by clinic personnel to the experimenters.

4.0 RESULTS

4.1 The total sample

4.1.1 Demographics

In all, 1,057 subjects (910 males and 147 females) participated in the study. The distribution of subjects by age is given in Table 1. A comparison of these data with Canadian Forces demographics confirmed that the male/female ratio was virtually identical, as was the ratio of non commissioned members to officers (77% vs 23%). The sample underestimated the population by 10% on average in the two younger categories and overestimated the population by 10% on average in the two older categories. Because there were relatively few subjects older than 55 years, the 46-55 years and 56-65 years categories were combined for subsequent analyses.

**Table 1: The Distribution of Subjects by Age
(N=1057)**

Age (years)	n	%
16 – 25	40	4.0
26 – 35	246	23.0
36 – 45	578	55.0
46 – 55	185	18.0
56 – 65	8	1.0
Total	1057	100.0

Table 2 provides summary statistics on the current age, age at recruitment, total years of service and years of service in the current trade. For our sample, just over half the subjects (55%) were aged 36-45 years at the time of the study, the mean age at recruitment was 20 years, subjects had been employed for 19 years on average and had served in the current trade for 15 years. This information suggests that for most of the subjects, the trade (and thus likely the noise exposure profile) had not changed over the course of their military careers. Subjects' age and number of years of military service were highly correlated ($r = +0.872$, $p < 0.001$). Since number of years of service was assumed to be a valid index of number of years of noise exposure, and because of the demonstrated relationship of years of service and age, the decision was made to assess the effects of age on hearing rather than years of service in subsequent analyses.

Table 2: Age at Recruitment, Total Years of Service and Number of Years in Current Trade

	Current Age (years)	Age at Recruitment (years)	Total Years of Service (years)	Years of Service in Trade (years)
Mean	39.59	20.38	19.20	14.60
Standard Deviation	7.14	3.78	7.76	8.22
Minimum	20	15	1	0.25
Maximum	64	46	47	37

In all, 107 trades were represented in the sample. In some instances, similar trades were grouped together based on information about the work carried out. The number of subjects in each trade or trade grouping ranged from 1 to 97. Several trades known to be associated with high-level noise exposure, e.g., Infantry, Armoured Soldiers and Artillery Soldiers [46], were poorly represented. This may have been the result of overseas deployment during the period of the study. Land, Sea, Air and Purple (shared by the three environments) trades accounted for 15%, 14%, 36% and 27% of the sample, respectively. Based on a description of their duties and responses to the questionnaire, 9% of the sample were categorized as Controls, i.e., they were not routinely exposed to noise on the job.

4.1.2 Hearing thresholds

Mean current hearing thresholds (dB HL) from 0.5 kHz to 8 kHz are given for left and right ears for each of four age categories in Table 3. A repeated measures analysis of variance (ANOVA) applied to the data from those subjects for whom results were available for all seven frequencies tested (N=948) showed significant effects of age group, ear, frequency, frequency by age group, ear by frequency and ear by frequency by age group ($p < 0.001$). Post hoc pair wise comparisons using the least significant difference (LSD) test [47] indicated that, averaged across frequencies, hearing thresholds for the two younger groups were no different (8 dB HL). Thresholds for both these groups were significantly lower than those observed for each of the two older groups which were significantly different from each other (13 dB vs 20 dB). Averaged across groups, thresholds increased from 8 dB HL to 18 dB HL from 0.5 kHz to 8 kHz. The difference between ears within groups at each frequency was at most 4 dB. Although the mean hearing thresholds were within normal limits, distributions of hearing thresholds observed at 4 kHz and 6 kHz in the left and right ears were positively skewed toward high values, increasingly with age [44]. These data showed that the percentage of individuals with moderate to severe hearing loss increased with years of service. In the oldest group, approximately 17-26% had a moderate to severe hearing, a greater proportion (4-5%) in the left ear than the right at 4 kHz and 6 kHz.

Table 3: Current Hearing Thresholds (dB HL) as a Function of Age

Age (yrs)**	n	Ear	Frequency (kHz)							
			0.5	1	2	3	4	6	8 ⁺	
16 – 25	40	Left	7.4 (6.3)*	5.9 (5.8)	7.3 (8.4)	6.9 (10.2)	8.3 (9.6)	10.3 (10.2)	10.0 (13.8)	
		Right	9.8 (7.8)	7.5 (6.7)	5.6 (8.3)	6.4 (9.9)	8.0 (15.7)	9.4 (14.6)	11.1 (19.1)	
26 – 35	246	Left	5.8 (7.1)	5.6 (7.4)	5.5 (8.0)	7.9 (10.2)	9.2 (11.9)	12.5 (12.7)	11.9 (13.7)	
		Right	6.7 (7.5)	5.8 (7.0)	4.5 (7.6)	5.6 (9.6)	8.1 (11.6)	11.8 (11.5)	11.5 (10.8)	
36 – 45	578	Left	7.6 (7.9)	7.6 (8.5)	7.9 (9.3)	12.3 (12.4)	16.0 (14.8)	20.4 (16.1)	20.6 (17.5)	
		Right	8.9 (8.2)	8.7 (8.5)	7.4 (8.8)	10.3 (11.4)	13.8 (13.2)	17.7 (15.0)	18.1 (15.6)	
46 – 65	193	Left	10.6 (12.7)	11.2 (13.0)	13.5 (13.6)	21.9 (18.6)	26.4 (18.8)	30.5 (20.5)	34.2 (23.3)	
		Right	11.0 (9.1)	11.6 (9.9)	11.6 (11.3)	17.4 (14.8)	22.2 (17.4)	27.6 (18.8)	29.7 (22.1)	

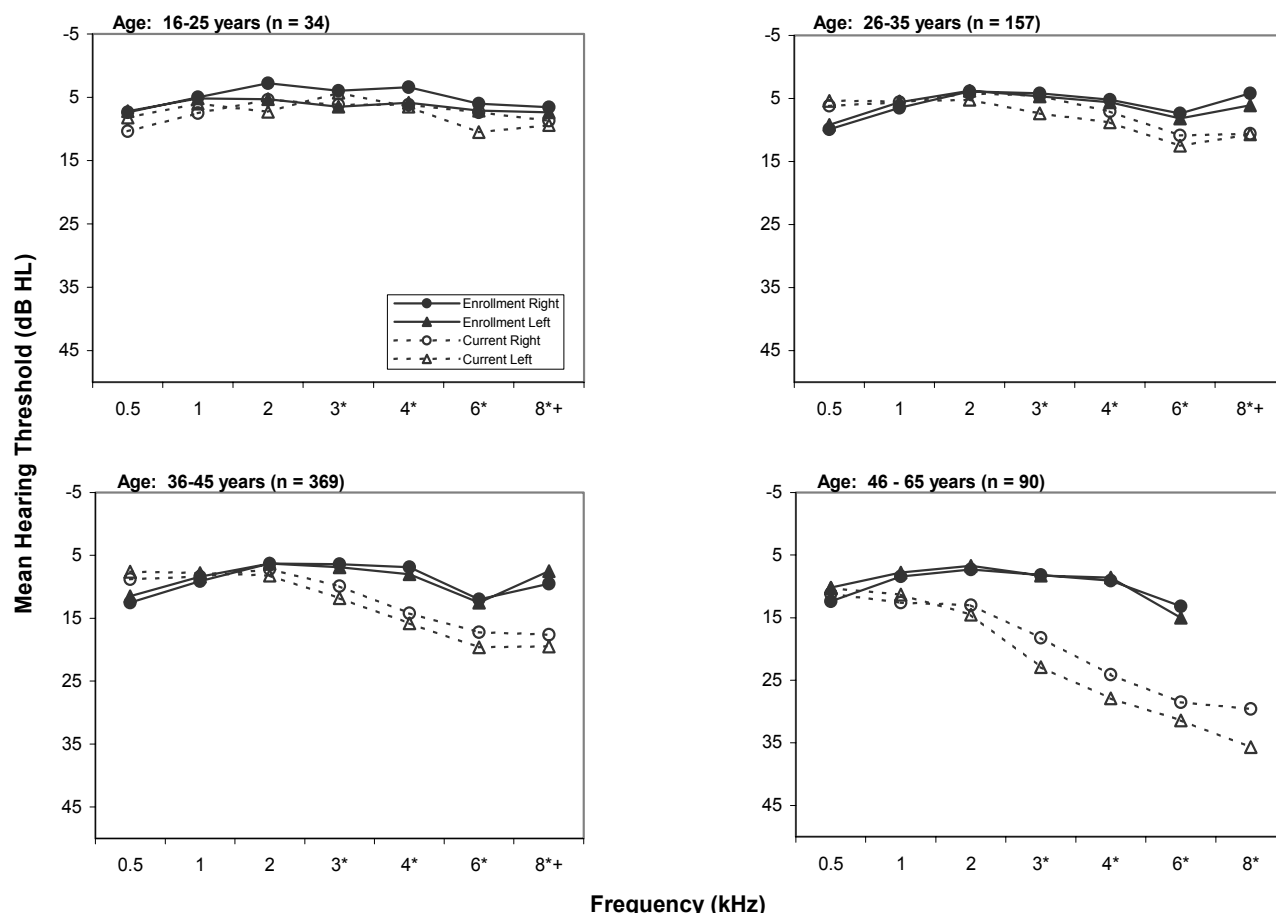
** age range = 20 – 64 years

* mean (standard deviation)

⁺ n for 8 kHz (16 – 25 yrs = 31; 26 – 35 yrs = 227; 36 – 45 yrs = 519; 46 – 65 yrs = 172)

A comparison of mean enrolment and current hearing thresholds is shown in Figure 1. Analyses were based on the data from those subjects in whom the earliest available audiogram was actually the enrolment audiogram (N=650). Results are shown separately for the four age groups. A repeated measures ANOVA applied to the data from those subjects in whom measurements were available for both ears at six frequencies from 0.5-6 kHz (N=647) showed significant effects of age group, time of measurement, ear, and frequency ($p<0.001$); time by age group, frequency by age group, time by ear, time by frequency, and ear by frequency ($p<0.001$); time by frequency by age group, time by ear by frequency ($p<0.001$); and time by ear by age group ($p<0.04$). Differences due to the time of measurement (enrolment vs current) ranged from 1-2 dB across frequency for the 18-25 year olds, from 0-4 dB for the 26-35 year olds, from 1-8 dB for the 36-45 year olds and from 0-16 dB for the 46-55 year olds. The largest differences were observed at 4 kHz or 6 kHz.

Noise-Induced Hearing Loss in Canadian Military Personnel



* n for Enrolment 3 kHz: 26-35=156; 4 kHz: 36-45=368, 46-65=89, 6 kHz: 26-35=156, 36-45=368, 46-65=89

*+ n for 8 kHz Enrolment: 16-25=22, 26-35=36, 36-45=10, 46-65=1; Current: 16-25=26; 26-35=157; 36-45=369; 46-65=81

Figure 1: Enrolment and Current Hearing Thresholds (dB HL) for Four Age Groups (n=650).

As stated above, enrolment audiograms were available for 650 subjects (61% of the sample). Table 4 provides a listing of years of service (in five year bins) for the first available audiogram together with the percentages of subjects who had hearing thresholds less than 20 dB HL (normal or slight hearing loss) and less than 30 dB HL (no more than a mild hearing loss) at 4 kHz and 6 kHz for the left ear. These are the frequencies that have been shown to be most susceptible to the damaging effects of noise. Shown for comparison are the distributions of current hearing thresholds for the same number of years of service. The table shows that for the majority of subjects (84%) the first audiogram was measured within the first five years. In these individuals, 90% had hearing thresholds less than 20 dB HL at 4 kHz and 78% had hearing thresholds less than 20 dB HL at 6 kHz. Regardless of whether the first or current audiogram was examined, the proportion with no more than a mild hearing loss of less than 30 dB HL at these two frequencies remained above 85% over the first 20 years of service, or in terms of age, until the early 40's. Steep declines were then evident.

Table 4: First and Current Hearing Thresholds as a Function of Years of Service

Frequency	Audiogram	Years of Service	n	Hearing Threshold (dB HL)	
				< 20 dB	< 30 dB
4 kHz (n = 1043*)	First	1 - 5 years	874 (84%)	784 (90%)	852 (97%)
		6 - 10 years	78 (8%)	65 (83%)	77 (99%)
		11 - 15 years	37 (4%)	28 (76%)	32 (86%)
		16 - 20 years	22 (2%)	17 (77%)	20 (91%)
		21 - 25 years	14 (1%)	7 (50%)	11 (79%)
		26 - years	18 (2%)	9 (50%)	13 (72%)
	Current	1 - 5 years	66 (6%)	52 (79%)	64 (97%)
		6 - 10 years	63 (6%)	54 (86%)	61 (97%)
		11 - 15 years	180 (17%)	149 (83%)	164 (91%)
		16 - 20 years	260 (25%)	184 (71%)	231 (89%)
		21 - 25 years	282 (27%)	156 (55%)	230 (82%)
		26 - years	192 (18%)	89 (46%)	123 (64%)

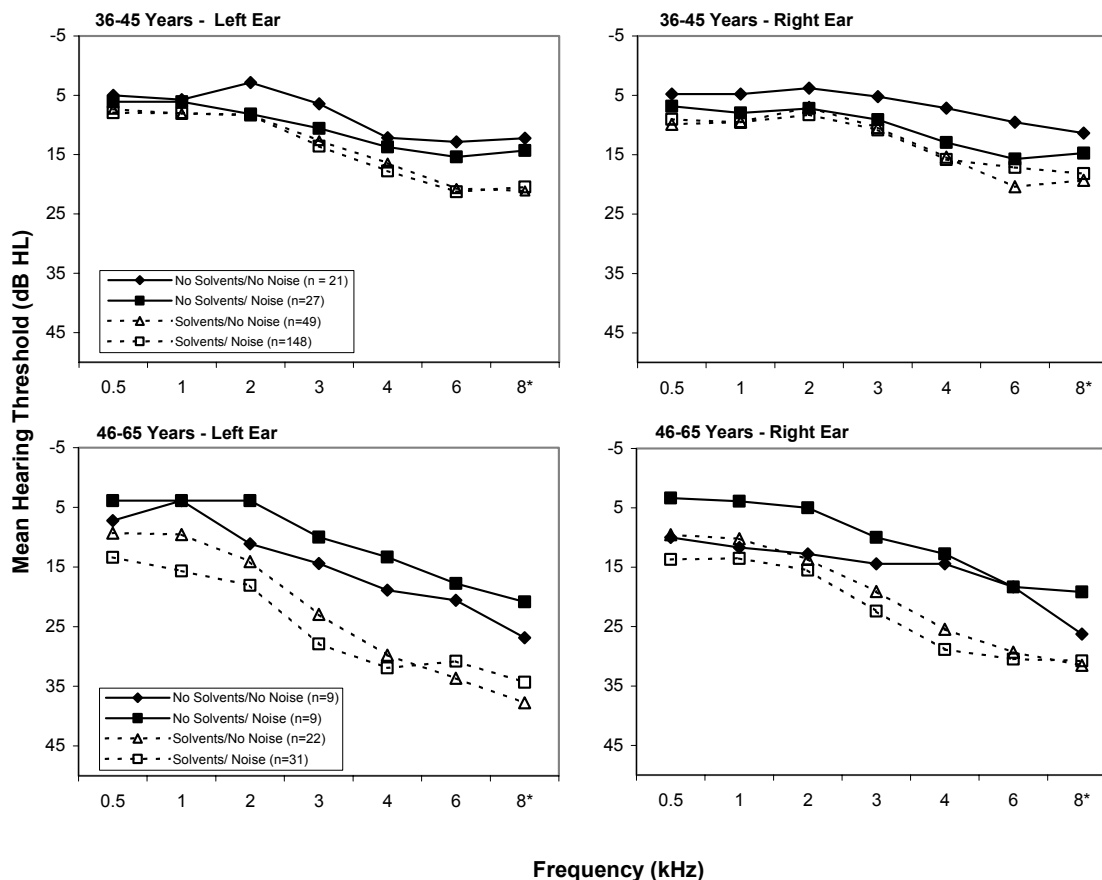
* N = 1057 – 14 participants were not included; 12 did not have a first hearing test or did not have a date for the first hearing test and 2 did not have hearing thresholds at 4 kHz

6 kHz (n = 1041*)	First	1 - 5 years	873 (84%)	682 (78%)	811 (93%)
		6 - 10 years	77 (7%)	54 (70%)	70 (91%)
		11 - 15 years	37 (4%)	26 (70%)	32 (86%)
		16 - 20 years	22 (2%)	16 (73%)	18 (82%)
		21 - 25 years	14 (1%)	5 (36%)	7 (50.0)
		26 - years	18 (2%)	9 (50%)	12 (67%)
	Current	1 - 5 years	66 (6%)	54 (82%)	61 (92%)
		6 - 10 years	63 (6%)	44 (70%)	59 (94%)
		11 - 15 years	179 (17%)	136 (76%)	161 (90%)
		16 - 20 years	259 (25%)	151 (58%)	208 (80%)
		21 - 25 years	282 (27%)	123 (44%)	200 (71%)
		26 - years	192 (18%)	64 (33%)	111 (58%)

* N = 1057 – 16 participants were not included; 12 did not have a first hearing test or did not have a date for the first hearing test and 4 did not have hearing thresholds at 6 kHz

Because a high proportion of subjects indicated that they had been exposed to solvents, a comparison was made of hearing thresholds in those with and without solvent exposure, who had or had not been exposed to noise. Subjects were categorized as solvent exposed if they had responded in the affirmative on the questionnaire and 80% or more of the individuals in their trade had been exposed. In all, 398 subjects met these criteria. Subject were defined as not exposed if they had responded negatively and if less than 20% of the those in their trade had been exposed. In all, 147 subjects met these criteria. From these groups a further selection was made of subjects who had been (1) exposed to noise “often” or “constantly” and the severity of the noise was “moderate” or “severe” and (2) those who were “never” or “occasionally” exposed to noise and the noise was “no noise” or “mild noise” based on the responses to the questionnaire. Mean hearing thresholds plotted as a function of frequency for subjects in each of the four cells are presented in Figure 2 for subjects aged 36-45 years and 46-65 years. These data show that solvents have a relatively greater negative impact than noise on hearing thresholds, particularly in the high frequencies above 3 kHz. The difference increased with age. Because of the wide difference in numbers of subjects across conditions, these trends were not evaluated statistically.

Noise-Induced Hearing Loss in Canadian Military Personnel



*n at 8 kHz: 36-45 yrs n = 18, 21, 44, and 131; 46-64 yrs n = 8, 6, 20, and 31

Figure 2: Effect of Solvent Exposures on Hearing in the Two Older Age Categories

4.1.3 Survey of risk factors

Responses to the questionnaire showed the following:

- Perception of changes in hearing and noise levels in the workplace: 27% believed that their hearing had become moderately or much worse since joining the Canadian Forces; 47% reported that their workplace was often or constantly noisy and 59% said that the noise to which they are exposed was moderate or severe.
- Training on hearing loss prevention: The greater proportion of the sample (82%) had either received no training or one hour or less on the dangers of noise exposure. Further, 89% had received either no training or one hour or less on the proper use of hearing protection devices. Finally, 50% judged their training on the dangers associated with noise exposure as either negligible or poor quality.
- Utilization of personal hearing protectors: 80% of the sample were exposed to noise at least some of the time. While relatively few (17%) of those exposed indicated that they never wore hearing protection devices in noise, only 42% wore these devices for more than half the work shift. Most

(77%) said that supervisors advised about the need for hearing protection either occasionally or not at all.

- Drawbacks of hearing protector utilization. On the whole, the participants appeared to understand the utility of wearing hearing protection. Only 19% often or definitely agreed that hearing protectors were *not* beneficial. Only 22% often or definitely agreed that they were uncomfortable but 50% believed that they would interfere with hearing. Only 10% felt that they would pose a danger at work. Concern about auditory impairment, then, seemed to be the major impediment to utilization.
- Few subjects (3%) had moderate to great difficulty listening in a quiet room. By comparison, 32% experienced moderate to great difficulty in noise. Similar outcomes were observed for listening over the telephone. Subjects who had difficulty in noise were likely those with acquired hearing loss. The percentage is similar to the percentage who believed that their hearing had become worse since joining the Canadian Forces. Both distance from the source (i.e., listening across a room) in noise without hearing protection and understanding orders in noise with hearing protection increased the percentage who experienced significant difficulty by about 20%.
- Exposure to Solvents: Over half the sample (61%) were exposed to solvents at work. Of this group, only 19% said that they wore respiratory protective equipment for more than half the work shift.
- Head injury: Relatively few subjects (9%) had a history of work-related head injury. Of these, 10% complained of a moderate or severe permanent hearing loss. In one-third of these cases, both ears were affected.
- Ear infections, non work related head injuries and the use of medications: 41% and 32% of the sample had experienced ear infections during childhood and adulthood, respectively. However, very few individuals (3% of the total sample) believed that they currently had a hearing loss from a previous ear infection. 12% had sustained a head injury outside of work. However, no subject had a permanent hearing loss as a result of the incident(s). No subject reported a permanent hearing loss as a result of medications.
- Tinnitus and dizziness: Previous research has shown that exposure to loud sounds may result in tinnitus (noise in the head or ears) as well as hearing loss. Tinnitus accompanied by both hearing loss and dizziness may be indicative of inner ear disease [48]. Half the sample (52%) reported that they had experienced some tinnitus. Of these, 34% perceived the tinnitus to be moderate or loud, 11% experienced moderate to much interference with their ability to hear, and 3% experienced sleep disturbance often or all the time. 15% of the sample had bouts of dizziness. Of these, only 24% definitely experienced the dizziness as the perception of circling. The latter would be indicative of vestibular involvement. Dizziness rarely interfered with work, and was rarely related to hearing loss.
- Exposure to leisure noise: Of the 1,057 respondents, 53% claimed that they were exposed to loud sounds during leisure activities. The most prevalent sources were power tools (58%), rock music (28%) and disco/dance bars (16%). The number of hours of exposure ranged widely across individuals from 5 minutes to 45 hours. Long durations may have resulted from subjects' inadvertently combining work and non work related exposures in their responses. Only 30% wore hearing protection more than half the time during leisure activities, with muffs and plugs used equally often.

4.2 Selected trades

4.2.1 Demographics

Seventeen trades with 15 or more subjects were selected for individual consideration (see Table 5). These trades accounted for 667 subjects or 63% of the sample of 1,057. A review of the current age, years of service and years in current trade indicated that they were comparable [44]. Mean age ranged from 35 to 43 years and mean total years of service ranged from 13 to 24 years. Mean years in the current trade was somewhat broader, ranging from 7 to 22 years. In most trades, the mean numbers of years in service and years in the current trade were virtually identical, i.e., they differed by 6 years at most. The exceptions were the flight engineers (11 years) and aircraft structures technicians (12 years). Further review indicated that the discrepancies were attributable to the fact that virtually all of the individuals in these two trades had

Table 5: Summary Statistics for Selected Trades

MOC	Job Title	n	Current Age	Total Years of Service	Years in Trade
031	Infantry Soldier	18	41.2 (6.6)*	21.7 (7.1)	22.0 (6.9)
091	Flight Engineer	33	42.9 (5.6)	23.9 (6.2)	13.1 (8.1)
215	Signal Operator	16	37.1 (5.9)	17.4 (5.9)	16.3 (6.8)
31	Air Navigator	51	35.4 (7.3)	15.2 (7.2)	12.7 (7.1)
32	Pilot	68	37.8 (7.2)	17.3 (7.6)	14.1 (7.6)
411	Vehicle Technician	47	38.7 (6.9)	19.4 (6.8)	17.3 (7.9)
514	Aviation Technician	97	38.4 (5.9)	18.2 (6.3)	11.8 (6.8)
526	Avionics Technician	43	40.4 (5.9)	20.8 (6.1)	14.5 (8.3)
565	Aircraft Structures Technician	17	38.3 (7.4)	18.9 (8.2)	7.4 (6.1)
71B	Marine Surface and Subsurface	15	40.0 (5.5)	19.8 (6.4)	17.2 (7.6)
737	Medical Technician	19	35.8 (7.9)	13.4 (8.8)	10.9 (8.3)
811	Military Police	20	35.8 (5.9)	15.5 (7.2)	12.9 (7.7)
836	Resource Management Support Clerk	64	40.8 (6.6)	18.6 (7.8)	16.2 (8.8)
861	Cook	29	39.7 (6.1)	19.5 (6.9)	17.4 (7.9)
911	Supply Technician	53	39.3 (5.3)	18.1 (6.8)	16.3 (6.8)
933	Traffic Technician	40	38.8 (8.2)	18.8 (9.2)	16.3 (8.7)
935	Mobile Support Equipment Operator	37	39.9 (6.8)	19.4 (8.4)	17.8 (8.6)

* Mean (Standard Deviation)

previously worked in an engineering technical trade (e.g., Aero Engineering Technician, Airframe Technician or Instrument Electronics Technician). Thus, on balance, it seems safe to conclude that the noise exposure had not changed

4.2.2 Hearing Thresholds

Figure 3 provides a display of the percentage of subjects in each of the selected trades with moderate to severe hearing loss exceeding 40 dB HL at 4 kHz and 6 kHz, ordered by prevalence. The data plotted indicate that the trades in which 15% or more of subjects had sustained a moderate to severe hearing loss were Flight Engineers (left ear at 4 kHz and right and left ears at 6 kHz), Signal Operators (left ear at 6 kHz) and Cooks (left ear at 6 kHz). In Figure 4, the percentages associated with hearing thresholds equal to or greater than 90 dB HL (deafness) to -10 dB HL (normal hearing) are displayed for the Flight Engineers for four frequencies, 0.5 kHz, 2 kHz, 4 kHz and 6 kHz, for all subjects in the selected trade, without respect to age. Mid-frequency (2 kHz) hearing thresholds have been shown to be a good predictor of speech understanding. Since thresholds

at this frequency were generally no more than mildly elevated, it is not expected that subjects would have difficulty with speech understanding in quiet. This outcome is in contrast to the higher frequencies, 4 kHz and 6 kHz, where moderate to severe hearing loss was more prevalent.

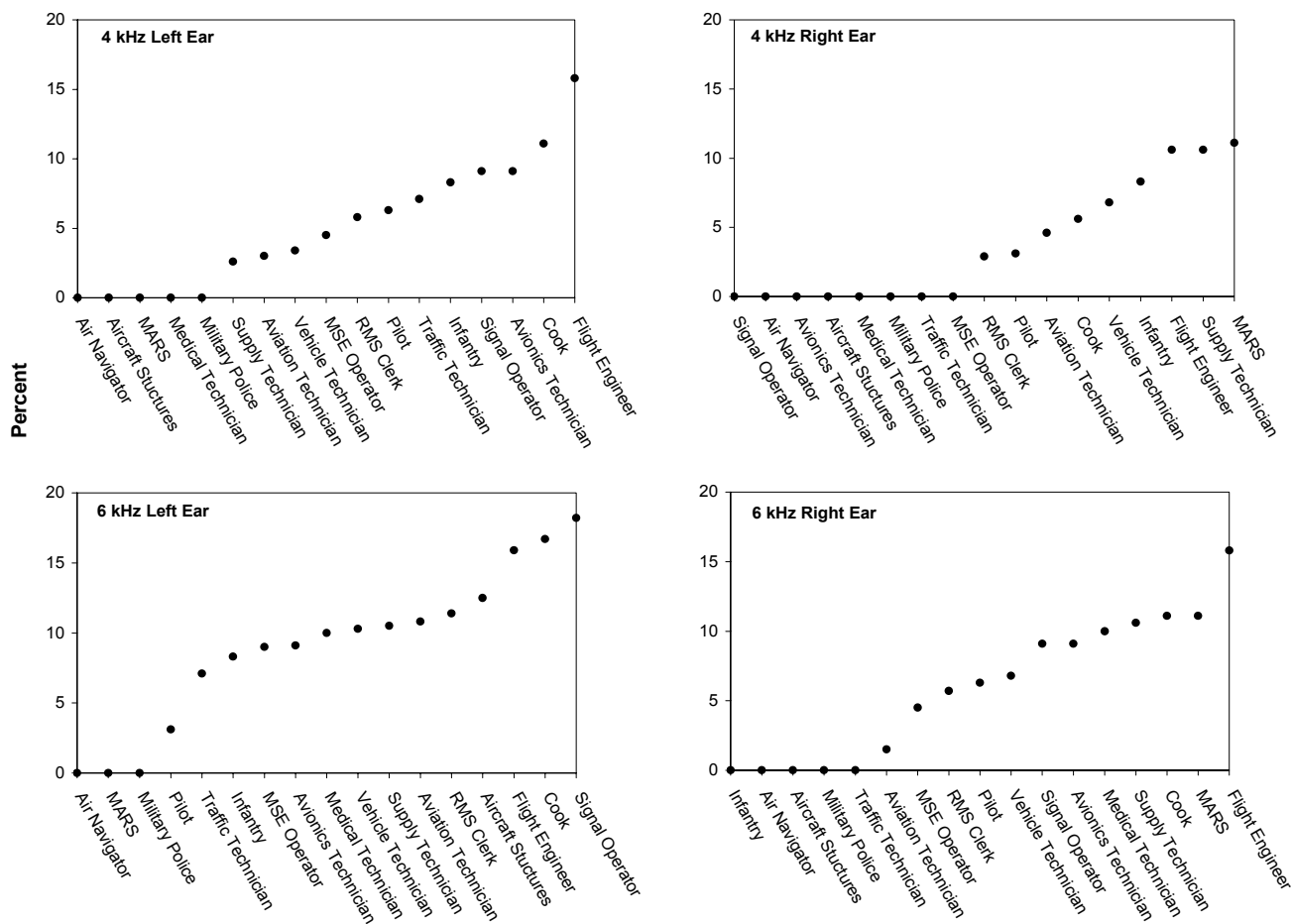


Figure 3: Percentage of 36-45 Year Olds with Moderate to Severe Hearing Loss in Selected Trades

Noise-Induced Hearing Loss in Canadian Military Personnel

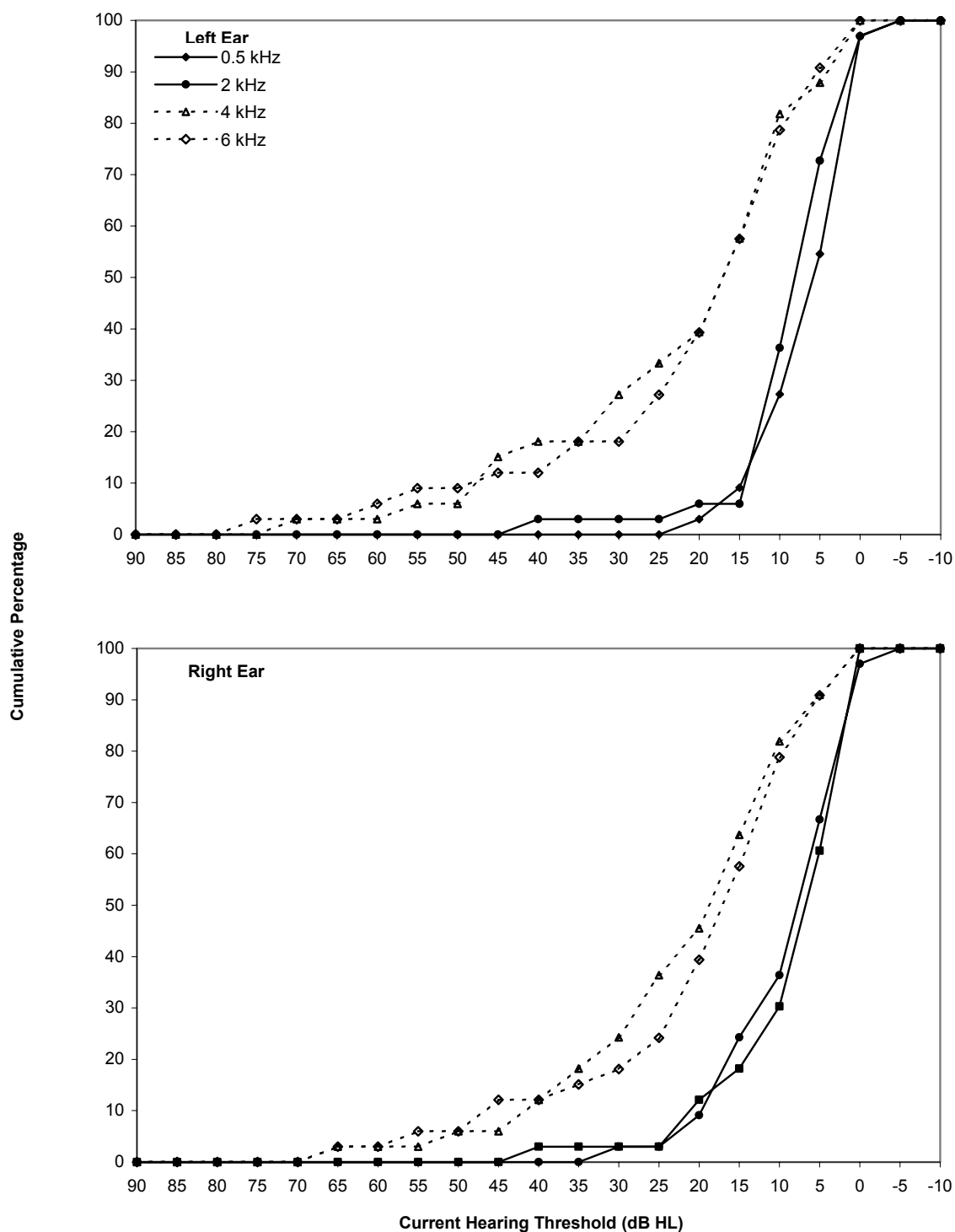


Figure 4: Hearing Loss in Flight Engineers (n=33)

4.2.3 Survey of risk factors

- Perception of changes in hearing and noise levels in the workplace: In approximately half the trades selected, 25% or more believed that their hearing was moderately or much worse since joining the Canadian Forces. Only the Medical Technicians and Military Police had fewer than 20% in these response categories. Trades groups in which more than 50% responded that their workplace was often or constantly noisy included Flight Engineers, Air Navigators, Pilots, Vehicle Technicians, Aviation and Avionics Technicians, Aircraft Structures Technicians, Cooks and Traffic Technicians. More than 40% of Flight Engineers, Air Navigators and Aircraft Structures Technicians said that the noise was severe.
- Training on hearing loss prevention: With the exception of Aircraft Structures Technicians, 68% or more of the subjects in each trade had either received no training or one hour or less on the dangers of noise exposure. Further, more than 75% in every trade had received either no training or one hour or less on the proper use of hearing protection devices. Regardless of trade, fewer than 13% judged their training on the dangers associated with noise exposure to be good.
- Utilization of personal hearing protectors: Except for Military Police, Resource Management Support Clerks and Cooks, at least 65% of individuals in each trade group were exposed to noise. Except for these trades, generally subjects said that they wore hearing protectors in noise. However, in only three trades, Flight Engineers, Air Navigators and Pilots, did more than 50% wear these devices for a full work shift. In only Infantry Soldiers and Aircraft Structures Technicians were more than 40% often or constantly reminded by their supervisors to wear hearing protection.
- Drawbacks of hearing protector utilization. Across trades, only 0-28% often or definitely agreed that hearing protection devices would not be beneficial. However, 25% or more of Flight Engineers, Pilots, Marine Surface and Subsurface personnel, Medical Technicians, Cooks, Supply Technicians and Traffic Technicians often or definitely agreed that they were uncomfortable, and 40% or more of all trades but Air Navigators and Pilots often or definitely agreed that they would interfere with hearing. Cooks (48%), Infantry Soldiers (28%) and Mobile Support Equipment Operators (24%) felt that they would often or definitely pose a danger at work.

In every trade selected, the majority of subjects (more than 70%) did not experience difficulty listening in a quiet room without protectors. The presence of noise resulted in moderate to great difficulty in more than 40% of Infantry Soldiers, Signal Operators, Vehicle Technicians and Supply Technicians. In well over half the selected trades (11 of 17), more than 40% of subjects experienced moderate or great difficulty when answering the telephone in a noisy room. As for the total sample, in every trade distance from the source (i.e., listening across a room) and the use of hearing protection increased the percentage who experienced significant difficulty.

- Exposure to Solvents: Over half the sample (60% or more) in all trades except Air Navigators, Pilots, Military Police, Resource Management Support Clerks, Cooks, Supply Technicians and Traffic Technicians said they were exposed to solvents. For the exposed subjects in each group, with the exception of the Aircraft Structures Technicians, the greater proportion (more than 70%) either did not wear respiratory equipment or did so for less than half the work shift.
- Head injury: Across the selected trades only the Infantry Soldiers, Vehicle Technicians, and Medical Technicians reported more than a 15% incidence of head injury at work. Of the 61 affected individuals in these trades, only 6 had more than a mild permanent hearing loss from the accident.

Noise-Induced Hearing Loss in Canadian Military Personnel

- Ear infections, non work related head injuries and the use of medications: At least 30% of the individuals in each trade had experienced ear infections in either childhood or adulthood. Across trades, only 7 individuals had more than a mild permanent hearing loss as a result. Head injuries outside of work occurred in more than 25% of the group for only Infantry Soldiers, Marine Surface and Subsurface personnel and Medical Technicians. There were no reported cases of permanent hearing loss. There was no evidence that any subject in any trade group was suffering from a permanent hearing loss as a result of medications.
- Tinnitus and dizziness: With the exception of Flight Engineers, less than 20% of each trade group believed that they experienced tinnitus often or constantly. In all those who experienced tinnitus in each trade, the tinnitus was perceived to cause moderate to much interference in less than 25% of the group, and often or constant sleep disturbance in less than 20% of the group. Less than 6% of any group experienced dizziness often or all the time. Of those who experienced dizziness, this symptom was definitely felt as the perception of circling in more than 40% in only Medical Technicians, Mobile Support Equipment Operators and Cooks. Dizziness rarely interfered with work or was related to hearing loss.
- Exposure to leisure noise: As in the case of the total sample, at least half the individuals in each trade had been exposed to leisure noise some of the time. The notable exception were Medical Technicians, Military Police and Cooks. In most trades the most likely source was power tools, followed by rock music. Except for Infantry Soldiers less than 50% in each trade wore hearing protection more than half the time during these exposures.
- Sideline work: Relatively few subjects in each trade worked outside the military. Those who engaged in sideline work generally did so occasionally.

5.0 DISCUSSION

A hearing conservation program was introduced by the Canadian Forces about 45 years ago [1]. Yet the cost of compensation claims adjudicated by Veterans Affairs Canada continues to escalate. The goal of the present investigation was to uncover possible oversights in the implementation of the program. Specifically, predisposing risk factors were documented in a relatively large sample of subjects that represented a wide variety of military trades and environments. A cross-sectional study design was used. Current and first available hearing test results were compared within individuals who had been scheduled for a routine hearing test and who had agreed to participate in the study. These subjects were then stratified by age and trade. Subjects were also asked to provide information on diverse factors including age, gender, number of years of military service, history of trades, training on the risks associated with noise exposure and methods of conserving hearing, as well as their history of occupational and leisure noise exposure, solvent exposure, use of protective gear, head injury, hearing health history and use of medications that might impact hearing.

It was estimated that 25% of those who would have been scheduled for a routine hearing test agreed to participate. A comparison of sample and population profiles showed that the ratios of males to females and non commissioned members to officers were closely matched. On average, the four sample age categories under and over estimated comparable population age categories by about 10%. One hundred and seven (107) out of a possible 133 current military trades were represented in the sample: 20 Land, 20 Sea, 15 Air, 23 Purple, and 29 Control (quiet). In only 17 trades, however, were there sufficient subjects for independent analysis. Of these only one (Medical Technicians) might be considered a quiet trade. In the analyses of the data undertaken, differences due to the gender and rank of subjects were not evaluated because of the relatively large disparities in the size of the subgroups that would be compared. Further, since number of

years of service and age were found to be highly correlated, the significance of only age was assessed. This approach had the advantage of allowing for the comparison of outcomes with International standards (see below).

The survey data indicated that about one-half the sample were concerned about noise exposure. Close to one-third perceived that their hearing had deteriorated over the their term of employment. Yet training on the hazards of noise exposure and methods of preventing hearing loss, as well as consistent use of personal hearing protectors was rated as poor. According to the respondents, supervisors rarely reminded personnel about the need to wear hearing protection in noise. Most subjects said that they believed that the use of hearing protection would be beneficial, and relatively few felt that such devices would pose a danger at work. Issues of concern were discomfort and the belief that the wearing of hearing protection might interfere with the performance of auditory tasks. With respect to the latter, half the sample reported moderate to great difficulty understanding orders in a noisy room while wearing hearing protection. Laboratory studies support this observation for subjects with pre-existing high-tone hearing loss [49].

Other risk factors for hearing loss include but are not limited to solvent exposure, head injury, ear infections, the use of ototoxic medications and exposure to noise during side line work or leisure. Well over half the sample (61%) said that they were exposed to solvents at work. Few (less than 20%) wore respirators, perhaps because they or their supervisors believed that the levels were within safe limits. Previous studies provide ample evidence that exposure to solvents alone or in combination with noise may result in hearing loss through injury to the inner ear [25,26]. Dose response curves for human subjects are not yet available. In contrast, the percentages of subjects who had sustained a head injury, suffered ear disease or used medications that might affect hearing were relatively small. Many reported symptoms of inner ear disease, such as dizziness (15%) and/or tinnitus (52%). However, these symptoms did not appear to be severe enough to interfere with auditory perception, the ability to work or sleep. Few subjects (8%) were engaged in sideline work but at least half participated in noisy leisure activities involving power tools, rock music and disc/dance bars. This latter finding is in line with earlier reports [46,50].

The audiometric data showed that for the sample as a whole hearing had deteriorated over time. While mean current thresholds were indicative of no more than a mild hearing loss of about 35 dB HL for the frequencies examined, the distributions of hearing thresholds across subjects were positively skewed to higher values at each frequency. At 4 kHz and 6 kHz, the most susceptible frequencies for NIHL, a hearing loss in excess of 20 dB was evident in approximately 15% of the youngest group (16-25 years). By mid-life (46 years and older) that proportion had increased to 60%, with 20% showing a moderate to severe hearing loss exceeding 40 dB HL [44]. Norms published for otologically normal persons [16] show that the change in hearing over the age range studied should be no greater than 10 dB for 50% of the population at these frequencies. In comparison, in the present study, the difference in mean hearing thresholds observed the youngest and oldest groups of subjects were 16 dB and 18 dB, respectively at 4 kHz and 6 kHz. A post hoc comparison of the effects of solvent and noise exposures either alone or in combination suggested that the solvents might be the more hazardous agent. Hearing thresholds in the oldest group at the higher frequencies, 3-8 kHz, were relatively higher for those subjects exposed to either solvents or solvents and noise than for those not exposed to solvents, with or without noise as a factor.

The 17 trades with sufficient numbers for independent consideration accounted for 61% of the total sample. They ranged widely from Cooks and Medical Technicians to Infantry Soldiers and Pilots. The profiles of subjects in these trades mirrored those of the larger group. Analysis of the survey data showed that the perception that the workplace was noisy was as likely to be reported by Cooks as Pilots or Vehicle Technicians. Generally, it was the air trades that reported that the noise was severe. Except for Aircraft

Noise-Induced Hearing Loss in Canadian Military Personnel

Structure Technicians, training on hearing loss prevention was considered poor. A relatively high proportion of only Flight Engineers, Air Navigators, Pilots and Traffic Technicians were likely to wear hearing protection for a full work shift. In most trades hearing protectors were considered to be beneficial. Cooks, Infantry Soldiers and Mobile Support Equipment Operators thought they might pose a danger for their work. A significant proportion in all 17 trades thought they would interfere with hearing. More than half the subjects in 11 of the 17 trades reported that they were exposed to solvents. In each, the use of respirators was not a common practice. As for the total sample, permanent hearing loss from head injury, ear infections and the use of medications were rare. Across the selected trades, the greatest average elevation in hearing threshold at 4 kHz in the left ear by 35-45 years of age was sustained by the Infantry Soldiers. Trades in which 15% or more subjects in this age category had a moderate to severe hearing loss of 40 dB or more in one or other ear at either 4 kHz or 6 kHz included Flight Engineers, Signal Operators, and Cooks. The prevalence of a moderate to severe hearing loss would be expected to increase as these groups aged. Unfortunately, the numbers of subjects available the oldest age category were not sufficient to extend the analysis.

While the results discussed above provide some clues as to some of the underlying factors responsible for the escalating burden of noise-induced hearing loss in the Canadian Forces, subjects' written comments help to identify some of the main problems associated with hearing conservation programs. A representative sample follows.

Aviation Technician: "I've worked on flight line on four different bases...I have made significant efforts to protect my hearing....but the loud screeching sound of turbines and turbine engines still come through..."

Mobile Support Equipment Operator: "As an MSE OP, certain vehicles are very loud inside and as an instructor I am unable to wear hearing protection."

Aviation Technician: "Engine run-ups are extremely loud. It is what causes ringing in my ears. I wear ear plugs and a communications headset, but it doesn't really matter."

Director of Music: "The noise levels during a full band rehearsal are often quite loud.....In order to hear each other, most musicians cannot wear ear defenders."

Aviation Technician: "...inside the hanger 99.8% of the personnel are not wearing hearing protection and concurrent activities...are being carried out outside the hanger which are damaging their hearing abilities."

Traffic Technician: "As it's nearly impossible to carry out a conversation in the back of the aircraft without removing one side of the headset, this practice was and is still common practice amongst in flight check personnel."

Signal Operator: "Monitoring multiple radios plus telephones makes it almost impossible to wear protective equipment. HF radio systems are particularly noisy."

Naval Communicator: "The initial 15 years of going to sea involved noise at levels higher than tolerable. Ignorance toward noise and its effects resulted in my hearing loss....."

Aviation Technician: "I have had trouble wearing earmuffs in winter with a toque."

Search and Rescue Technician: "I do not wear foam ear plugs due to comfort issues."

6.0 CONCLUSIONS

The results of this investigation lead to the following conclusions that the prevalence and escalation of noise-induced hearing loss in the Canadian Forces are due to:

- Insufficient training on the hazards of noise exposure, hearing protection selection and hearing protector utilization.
- Problems with hearing protection use including discomfort with extended usage, incompatibility with gear, and difficulty communicating in noise.
- Unprotected exposure to solvents and leisure noise.

7.0 RECOMMENDATIONS

Comments made by subjects support the creation of end-user focus groups for high-risk trades to capture the impact of hearing conservation programs and practices currently in place on individual workers. Previous studies have shown that these may be used to advantage to supplement and clarify both audiometric and structured survey data [51]. Individual users of personal hearing protection and their supervisors could use these sessions to provide insight into problems with the utilization of hearing protection.

Tailored intervention studies should be conducted to determine the type of hearing protection that would be suitable for individual users in each trade [52]. In particular, the utility of implementing newer hearing protector technologies should be assessed in field studies. Special consideration should be given to both the type of noise and the level of noise to which personnel are exposed, communication capability, comfort and compatibility with other gear [53,54].

Training on the hazards of noise exposure (both occupational and non occupational) must be mandatory for all CF personnel at recruitment and with each new posting. These lectures must include instruction on the proper fitting of the hearing protection that will be used on the job, as well as periodic monitoring to ensure maximum effectiveness.

Strategies for hearing loss prevention must include active surveillance of hearing protection practices by supervisors and flagging by clinic personnel of changes in hearing that exceed strictly defined criteria, immediate and appropriate intervention, and an evaluation of the effectiveness of the intervention [55].

Steps must be taken to control solvent exposures. Although few studies have explored dose response relationships in human subjects, it is clear from the results available that solvents alone or in combination with noise will result in cochlear pathology. The data available demonstrate clearly that even if solvent concentrations meet the criteria set out in regulatory guidelines, they may nonetheless potentiate noise-induced hearing loss [25]. Education on the risks of solvent exposure and training on the use of respiratory equipment should be mandatory for all personnel.

8.0 ACKNOWLEDGMENTS

This research was funded by Veterans Affairs Canada in partnership with Defence Research and Development Canada. The author is indebted to Ms. Stephanie Jewell for her help with database management, the statistical analysis of results and the preparation of the figures and tables.

9.0 REFERENCES

- [1] Neely K. Hearing conservation for the armed forces. Medical Services J.- Canada 1959; April: 235-247,
- [2] Vye K. WSIB battles hearing loss: The silent workplace hazard. Workplace Safety and Insurance Board News Release, May 15, 2001.
- [3] Statistics Canada. Canadian statistics - the people: Labour force, employed and unemployed, numbers and rates, 2003. Excerpted from the official website, www.statcan.ca.
- [4] MOL. Canada Labour Code, Part VII, Levels of Sound. Canada Occupational Health and Safety Regulations, Ministry of Labour, Ottawa, ON, Canada, 1991.
- [5] CSA. Hearing protection devices – performance, selection, care and use, Z94.2-02. Canadian Standards Association, Rexdale, ON, Canada, 2002.
- [6] Franks JR, Stephenson MR, Merry CJ. Preventing occupational hearing loss – a practical guide. US DHHS, PHS, CDC, NIOSH publication no. 96-110, 1996.
- [7] Henderson D, Hamernik RP. Biological bases of noise-induced hearing loss. *Occup Med* 1995; 10(3): 513-534.
- [8] Forge A. Sensory cell regeneration and functional recovery: A review. In: Axelsson A, Borchgrevink HM, Hamernik RP, Hellström P-A, Henderson D, Salvi RJ, eds. *Scientific basis of noise-induced hearing loss*. New York: Thieme, 1996, pp. 3-22.
- [9] Leikin JB, Davis A, Klodd DA, Thunder T, Kelafant GA. Selected topics related to occupational exposures. Part I. Noise exposure. *Diseases of the Month* 2000; 46: 240-255.
- [10] Bauer P, Korpert K, Neuberger M, Raber A, Schwetz F. Risk factors in hearing loss at different frequencies in a population of 47,388 noise-exposed workers. *J Acoust Soc Am* 1991; 90(6): 3086-3098.
- [11] Abel SM, Haythornthwaite C. The progression of noise-induced hearing loss. A survey of workers in selected industries in Canada. *J Otolaryngol.* 1984; 13, Suppl: 2-36.
- [12] Bohnker BK, Page JC, Rovig G, Betts LS, Muller JG, Sack DM. U.S. Navy and Marine Corps hearing conservation program 1995-1999: Mean hearing thresholds for enlisted personnel by gender and age groups. *Military Medicine* 2002; 167: 132-135.
- [13] Smoorenburg GF. Damage risk criteria for impulse noise. In: Hamernik RP, Henderson D, Salvi R, eds. *New perspectives on noise-induced hearing loss*. New York: Raven, 1982, pp. 471-490.
- [14] Dancer A, Buck K, Parmentier G, Hamery P. The specific problems of noise in military life. *Scand Audiol* 1998; 27, Suppl 48: 123-130.
- [15] Christiansson BAC, Wintzell KA. (1993). An audiological survey of officers at an infantry regiment. *Scand Audiol* 1993; 22: 147-152.

- [16] ISO. Threshold of hearing by air conduction as a function of age and sex for otologically normal persons, 7029. Int. Organization for Standardization, Geneva, Switzerland, 1984.
- [17] ISO. Determination of occupational noise exposure and estimation of noise-induced hearing impairment, 1999. Int. Organization for Standardization, Geneva, Switzerland, 1990.
- [18] Dobie RA. Separating noise-induced from age-related hearing loss. *Western J Med* 1994; 160 (6): 564-565.
- [19] Coles RRA, Lutman ME, Buffin JT. Guidelines on the diagnosis of noise-induced hearing loss for medicolegal purposes. *Clin Otolaryngol* 2000; 25: 264-273.
- [20] Pryor GT, Dickinson J, Feeney E, Rebert CS. Hearing loss in rats first exposed to toluene as weanlings or as young adults. *Neurobehav Toxicol Teratol* 1984; 6: 111-119.
- [21] Muijser H, Hoogendijk EMG, Hooisma J. The effect of occupational exposure to styrene on high-frequency thresholds. *Toxicol* 1988; 49: 331-340.
- [22] Fechter LD, Clerici WJ, Yao L, Hoeffding V. Rapid disruption of cochlear function and structure by trimethyltin in the guinea pig. *Hearing Res* 1992; 58: 166-174.
- [23] Morata TC, Dunn DE, Kretschmer LW, Lemasters GK, Keith RW. Effects of occupational exposure to organic solvents and noise on hearing. *Scand J Work Environ Health* 1993; 19: 245-254.
- [24] Barregård L, Axelsson A. Is there an ototraumatic interaction between noise and solvents? *Scand Audio* 1984; 13: 151-155.
- [25] Sliwinska-Kowalska M, Zamysłowska-Szmytko, E, Szymczak, W, Kotylo P, Fiszer M, Dudarewicz A, Wesolowski W, Pawlaczyk-Luszczynska M, Stolarek R. Hearing loss among workers exposed to moderate concentrations of solvents. *Scand J Work Environ Health* 2001; 27(5): 335-342
- [26] Fechter LD, Chen G-D, Rao D. Characterising conditions that favour potentiation of noise induced hearing loss by chemical asphyxiants. *Noise & Health* 2000; 3(9): 11-21.
- [27] Clark WW. Noise exposure from leisure activities: A review. *J Acoust Soc Am* 1991; 90(1): 175-181.
- [28] Axelsson A, Rosenhall U, Zachau G. Hearing in 18-year old Swedish males. *Scand Audiol* 1994; 23: 129-134.
- [29] Borchgrevink HM. Music-induced hearing loss >20 dB affects 30% of Norwegian 18 year old males before military service - the incidence doubled in the 80's, declining in the 90's. In: *Proc Int Congr on Noise as a Public Health Problem*, Nice, France, 1993, pp.25-28.
- [30] Berglund B, Lindvall T. Community noise. *Arch of the Ctr For Sensory Research* 1995; 2(1): 1-195.
- [31] Humes LE. Understanding the speech-understanding problems of the hearing impaired. *J Am Acad Audio* 1991; 2(2): 59-69.

Noise-Induced Hearing Loss in Canadian Military Personnel

- [32] Smoorenburg GF. Speech reception in quiet and in noisy conditions by individuals with noise-induced hearing loss in relation to their tone audiograms. *J Acoust Soc Am* 1992; 91(1): 421-437.
- [33] Davies DR, Jones DM. Noise and efficiency. In: Tempest W, ed. *The noise handbook*. London: Academic, 1985, pp. 87-141.
- [34] Kjellberg A. Subjective, behavioral and psychophysiological effects of noise. *Scand J Work Environ Health* 1990; 16, Suppl 1: 29-38.
- [35] Hockey GR. Effect of loud noise on attentional selectivity. *Quart J Exp Psychol* 1970; 22 : 28-36.
- [36] Vallet M, Gagneux J-M, Simonnet F. Effects of aircraft noise on sleep: An in situ experience. In: Tobias JV, Jansen G, Ward WD, eds. *Noise as a public health problem*. ASHA Reports 10, Rockville, Maryland, 1980, pp. 391-396.
- [37] Griefahn B, Schuemer-Kohrs A, Schuemer R, Moehler U, Mehnert P (2000). Physiological, subjective and behavioural responses during sleep to noise from rail and road traffic. *Noise & Health* 2000; 3(9): 59-71.
- [38] Kent SJ, von Gierke HE, Tolan GD. Analysis of the potential association between noise-induced hearing loss and cardiovascular disease in USAF aircrew member. *Aviat Space and Environ Med* 1986; April, 348-361.
- [39] Tarnopolski A, Watkins G, Hand DJ. Aircraft noise and mental health: I. Prevalence of individual symptoms. *Psychol Med* 1980; 10: 683-698.
- [40] Thiessen GJ. Disturbance of sleep by noise. *J Acoust Soc Am* 1978; 64: 216-222.
- [41] Thiessen GJ, Lapointe AC. Effect of intermittent truck noise on percentage of deep sleep. *J Acoust Soc Am* 1983; 64: 1078-1080.
- [42] Cantrell RW. Prolonged exposure to intermittent noise: audiometric, biochemical, motor, psychological and sleep effects. *Laryngoscope* 1974; 84, Suppl 1.
- [43] McLean EK, Tarnopolsky A. Noise, discomfort and mental health: a review of the socio-medical implications of disturbance by noise. *Psychol Med* 1977; 7: 19-62.
- [44] Abel SM. Risk factors for the development of noise-induced hearing loss in Canadian Forces personnel. External Client Report, DRDC Toronto ECR 2004-116, Toronto, ON, Canada, 2004.
- [45] Yantis RA Pure tone air-conduction testing. In: Katz J. ed. *Handbook of clinical audiology*, 3rd ed. Baltimore: Williams & Wilkins, 1985, pp. 153-169.
- [46] Pelausa EO, Abel SM, Simard J, Dempsey, I. Prevention of noise-induced hearing loss in the Canadian military. *J Otolaryngol.* 1995; 24(5): 271-280.
- [47] Daniel WW (1983). *Biostatistics: a foundation for analysis in the health sciences*, 3rd ed. New York: Wiley, 1983.

- [48] Kruger B. Basic audiologic evaluation. In: Alberti PW, Ruben RJ, eds. *Otologic medicine and surgery*, vol. 1. New York: Churchill Livingstone, 1988, pp. 365-395.
- [49] Abel SM, Armstrong NM, Giguère C. (1993). Auditory perception with level-dependent hearing protectors: The effects of age and hearing loss. *Scand Audiol* 1993; 22(2): 71-85.
- [50] Bray MB Szymanski M, Mills R. Noise induced hearing loss in dance music disc jockeys and an examination of sound levels in nightclubs. *J Laryngol Otol* 2004; 118: 123-128.
- [51] Prince MM, Colligan MJ, Stephenson CM, Bischoff BJ. The contribution of focus groups in the evaluation of hearing conservation program (HCP) effectiveness. *J Safety Res* 2004; 35: 91-106.
- [52] Lusk SL, Ronis DL, Kazanis AS, Eakin BL, Hong OS, Raymond DM. Effectiveness of a tailored intervention to increase factory workers' use of hearing protection. *Nursing Res* 2003; 52(5): 289-295.
- [53] Dancer A, Buck K, Hamery P, Parmentier G. Hearing protection in the military environment. *Noise & Health* 1999; 5: 1-15.
- [54] Abel SM, Sass-Kortsak A, Kielar A. The effect on earmuff attenuation of other safety gear worn in combination. *Noise & Health* 2002; 5: 1-13.
- [55] Helfer TM, Shields AR, Gates, K.E. Outcomes analysis for hearing conservation programs. *Am J Audiol* 2000; 9: 75-83.



Evaluation of Noise Effects in Auditory Function in Spanish Military Pilots

JM Lorente, B. Puente, B. Esteban, F. Ríos, P. Vallejo, C. Velasco

CIMA
Arturo Soria 82
28027 Madrid
SPAIN

E-mail: beapues@telefonica.net

INTRODUCTION

Military pilots are particularly exposed to a wide variety of environmental trends and one of them is the noise produce by the aircraft used for flying duties. In addition to that changes in cabin pressure and vibration can be considered as contributing factors.

Adverse effects can be produced under several situations, by self exposing to the noise environment in the flight line or during taxi, take off and cruise operations (1,2,3). Environmental stress can be partially coped by using noise reduction devices, but effectiveness is still far away from procuring a complete protection and devices are still problematic in various aspects (4).

Consequently limited effectiveness of speech communication and eventually hearing impairment or loss can be produced.

OBJECTIVES

From and epidemiological perspective we want to know the incidence of noise-induced hearing loss in a aleatory sample of aircrew members belonging to the Spanish Armed Forces in order to obtain valuable data and to proceed with eventual solutions and application of corrective measures that will lead to stop adverse effects of noise.

MATERIAL AND METHODS

Population studied in this work are aircrew members, all of them pilots of the Spanish Armed Forces. All pilots reviewed has been estuded in the Spanish Armed Forces Aeromedical Center (CIMA) (departments of Aerospace Medicine and ENT).

Samples were randomized among the pilots who currently take his annual physical check up.

Selected criteria were: to be fighter, transport or helicopter pilot and to be fit for that.

The study was anonymous, volunteer and written consent should be completed.

We rule out from the study all pilots with demonstrated ear disease or trauma detected during the physical.

Lorente, J.-M.; Puente, B.; Esteban, B.; Ríos, F.; Vallejo, P.; Velasco, C. (2005) Evaluation of Noise Effects in Auditory Function in Spanish Military Pilots. In *New Directions for Improving Audio Effectiveness* (pp. 3-1 – 3-8). Meeting Proceedings RTO-MP-HFM-123, Paper 3. Neuilly-sur-Seine, France: RTO. Available from: <http://www.rto.nato.int/abstracts.aps>.

Evaluation of Noise Effects in Auditory Function in Spanish Military Pilots

We established three phases in the design:

- Sociodemographic data: age and sex
- Aeronautical data: Service, type of aircraft, flying time, mean time of time flying period, type of hearing protection currently used (ear plugs, headsets or integrated helmet).
- ENT data:
 - ENT history, otitis, barotitis, hipoacusia, other additional data.
 - ENT Exam: general, otoscopy, impedanciometry and audiometry.

Data were included in a data base for statistical analysis by using the SPSS (Statistical Package Social Sciences) version 10.0 for windows.

We perform descriptive study for the variables and we used the Kolmogorov-Smirnov Test for normal deviation of quantitative variables.

For mean deviation we used the Kruskal-Wallis and “u” of Mann-Whitey, depending on the characteristics of the variables; and Chi2 for comparison of proportions.

We considered for statistical analysis a significance of $p < 0.05$.

RESULTS

A total of 372 pilots were studied, all of them males and mean age of 34, 74 (SD 7.97) and range between 23 and 56 years old.

A total of 276 were Air Force (AF) pilots (74.2%), 91 (24.5%) belongs to Army and 5 to the Navy (1.3%).

Table 1: Shows the Percentage of Pilots According to the Type of Aircraft and Mean Flying Time.

	Frequency	Valid Percent	Mean Flying Time
Fighter	82	22,0 %	2617,57
Transport	170	45,7 %	2785,03
Helicopters	120	32,3 %	1540,3
Total	372	100,0 %	

In order to analyse the outcome of the audiometry performed to each pilot, we have divided the sample in 4 age groups:

- 20-29 yo
- 30-39 yo
- 40-49 yo, and
- 50-59 yo

Table 2: Shows the Percentage of Pilots Distributed by Age Group.

Age Group	Frequency	Valid Percent
20-29	120	32.3 %
30-39	162	43.5 %
40-49	62	16.7 %
50-59	28	7.5 %
Total	372	100.0 %

To correlate the results of the audiometries we took into account the physiological loss by age group and in every frequency (250, 500, 1000, 2000, 4000, 8000 Hz) according to the modified T J Leisti (5) tables (Table 4), depicted in Figure 1.

Table 4: Physiological Loss According to T.J. Leisti.

Age	125	250	500	1000	2000	4000	8000
20-29	0	0	0	0	0	3	5
30-39	5	5	5	5	6	14	16
40-49	7	7	7	8	8	21	25
50-59	10	10	12	12	13	29	32
60-69	14	14	15	19	24	40	48
70-79	18	19	23	24	31	47	59
80	22	23	27	33	39	56	66

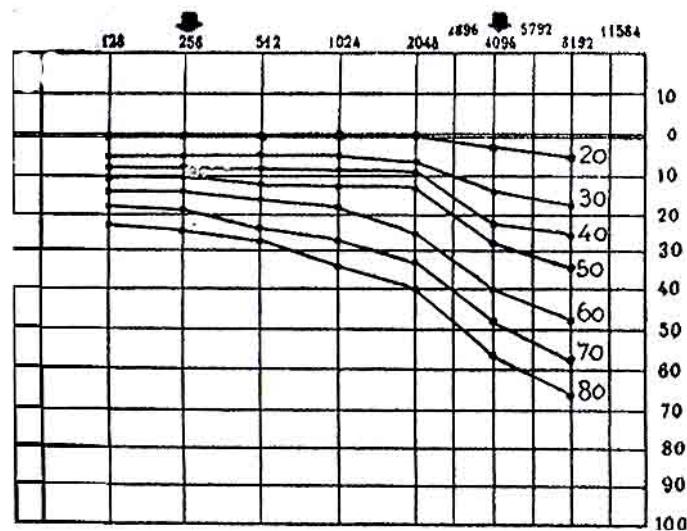


Figure 1: Shows Age Loss Depicted by Leisti.

Tables 5, 6 and 7: Shows the Findings According to Type of Aircraft.

Table 5

FREQUENCY	FIGHTER PILOTS			
	RIGHT EAR		LEFT EAR	
	NORMAL	ANORMAL	NORMAL	ANORMAL
250	45,1%	54,9%	51.2%	48.8%
500	50%	50%	59.8%	40.2%
1000	76.8%	23,2%	81.7%	18,3%
2000	76.8%	23.2%	73.2%	26.8%
4000	56.1%	43,9%	48.8%	51.2%
8000	58.5%	41,5%	56.1%	43.9%

Table 6

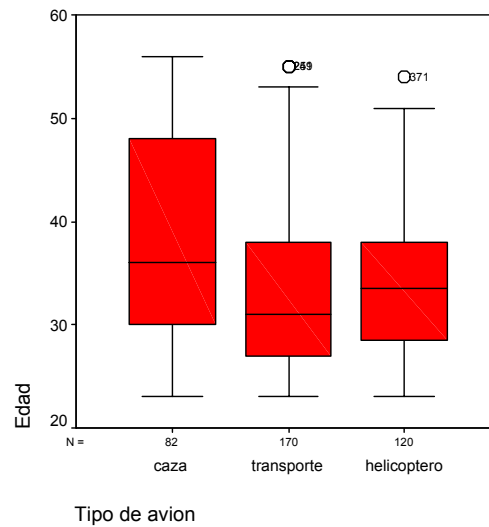
FREQUENCY	TRANSPORT PILOTS			
	RIGHT EAR		LEFT EAR	
	NORMAL	ABNORMAL	NORMAL	ABNORMAL
250	39.4%	60.6%	40.0%	60,0%
500	33.5%	66.5%	35.3%	64.7%
1000	73.5%	26.5%	68.2%	31.8%
2000	80.0%	20.0%	73.5%	26.5%
4000	58.8%	41.2%	52.9%	47.1%
8000	65.9%	34.1%	62.9%	37,1%

Table 7

FREQUENCY	HELICOPTER PILOTS			
	RIGHT EAR		LEFT EAR	
	NORMAL	ABNORMAL	NORMAL	ABNORMAL
250	45.0%	55.0%	41.7%	58.3%
500	37.5%	62.5%	41.7%	58.3%
1000	72.5%	27.5%	75.8%	24.2%
2000	74.2%	25.8%	78.2%	21.8%
4000	63.3%	36.7%	55.0%	45.0%
8000	70.0%	30.0%	69.2%	30.8%

Results coming from the comparative study of groups (Fighter vs Transport) in each frequency, shows that fighter pilots got a better audiometry with statistical significance ($p=0.012$) in the frequencies of 500 for the Right Ear. In the Left Ear $p<0.000$ at 500 and $p<0.025$ at 1000. There are no differences between transport and helicopters.

In order to evaluate if age could be considered as a causal factor we crossed age with type of aircraft and results showed that after application of Mann-Whitney U Test, the fighter pilots got an age statistical significant older than transport and helicopter pilots ($p=0.000$ and $p=0.002$), Figure 2 depicted those differences.



Regarding Ear Protection we found that 91.4% of fighter pilots used integrated helmet, but only 4.3% of transport and 57.5% of helicopters pilots currently used integrated helmet. Statistical difference were found with $p=0.000$.

DISCUSSION

(***)

Hearing loss due to exposure to aviation environment is extensively presented in the current literature. Many authors correlate noise, loss in auditory perception and exposure to noise related to aviation (2, 6, 7, 8, 9, 10, 11, 12, 13, 14). Mechanisms of ear impairment involved are: repeated barotraumas which affect the middle ear leading to low frequencies impairment, and exposure to environmental aviation noise (aircraft engine noise, etc...) which lead to impairment in frequencies of 3000 and 4000.

In this paper we have demonstrated how often we observed audio impairment in military pilots. Only 16.1% (60 pilots out of a total of 372) does not show any impairment.

Helicopter and transport pilots use to be the more noise induce impaired. In acute frequencies (500) in both ears and frequency of 1000 in the left ear, the fighter pilot got a better ear acuity.

Age is not a factor, in spite of the type of aircraft, even older pilots but better protected meant a better ear acuity. Apparently the transport pilots and helicopter pilots seems to be less protected or to receive a poor protection, and actually we demonstrated that with a good degree of statistical significance.

CONCLUSIONS

1. High incidence of audio impairment in military pilots.
2. More impairment in transport and helicopter pilots.
3. Frequencies more affected in our study: 500 Hz and 1000 Hz.
4. Age is not a factor in this study.
5. Transport and helicopter pilots are insufficiently protected.
6. Barotraumas should be consider as a contributing factor.
7. Low frequency vibrations should be consider as well as contributing hactor.

So, hearing loss in pilots should be considered as a professional disease and linked to noise production, repeated barotraumas and (low frequency) vibrations; the lack of adequate hearing protection can lead to loss, and prevention is a key factor in stopping auditory impairment and the only way to cope such problem.

Epidemiological studies must pursue a better understanding of the incidence, factors related and countermeasures establishment.

BIBLIOGRAPHY

- [1] LIENHART H., GOUTEYRON J.F., FAUGERE J.L., FILIPPINI J.F. La surdité de l'aviateur. Première partie. Méd. Aéro. Spat., XXVII, 1988, 105, 53 - 5T9.
- [2] GOMEZ CABEZAS P.: Hipoacusias del Aviador. Rev. Aero. Astro. Julio 1972, N° 380: 495-504.
- [3] GOMEZ CABEZAS P.: Barotraumatismos óticos y sinusales. Rev. Aero. Astro. Octubre 1973, N° 395: 767-782.
- [4] VON GIERKE H.E.: Exposure to combined noise and vibration environments, Proc. of 3rd Intern. Cong. of noise as a Public Health Problem, ASHA Report 10, American Speech and Hearing Association, Rockville, MD., 1980.
- [5] PORTMANN M, PORTMANN C.: Audiometría Clínica. Barcelona: Ed Toray-Masson, 1979 (3^a edición): p153 (Fig. 108).
- [6] HUSTIN A.: Troubles de láudition chez le personnel navigant civil. Communication à la Société Francaise de Physiologie et de Médecine Aéronautiques et Cosmonautiques. Paris 18 Juin 1965. Rev. Méd. Aéro. Spat., 1965, N° 13: 50-51.
- [7] GOMEZ CABEZAS P.: Hipoacusias perceptivas del aviador. El trama acústico del personal volante y de las industrias aeronáuticas. Rev. Aero. Astro. Octubre 1963, N° 275: 843-850.
- [8] BASTIEN J., GALBAN P., BLANC P., SOURDOIS G., ROBION J.: Considérations statistiques sur les hypoacusies et surdités du personnel navigant technique de láviation civile au cours des visites médicales de controle.. 21 Mai 1965. Rev. Méd. Aéro. Spat., 1965, N° 13: 12-15.
- [9] BLANC MM., BASTIEN J.: Aspect actuel de la fonction cochleaire chez le personnel navigant. Rev. Méd. Aéro. Spat., 1973, N° 45: 154-155.

- [10] RIBAK J., HORNUNG S., KARK J., FROOM P., WOLFSTEIN A., ASHKENAZI I.E.: The association of age, flying time, and aircraft type with hearing loss of aircrew in the Israeli Air Force. *Aviat. Space Environ. Med.*, 1985; 56:322-7.
- [11] BROADWELL D.K.: Letter to editor. *Aviat. Space Environ. Med.*, 1987, 82.
- [12] BASSOUMI T., FAYALA M.: Détériorations audiométriques chez le pilote militaire en Tunisie. *Méd. Aéro. Spat.*, Tome XXX, N° 119 - 1991.
- [13] LORENTE J.: El traumatismo sonoro crónico. Su incidencia en el medio aéreo. 131 Curso de Aptitud para el Ascenso a Comandante. Escuela Superior del E.A. Madrid 1991.
- [14] LORENTE J.M.: "Cambios auditivos en los pilotos de las Fuerzas Armadas. Revisión bibliográfica." *Revista Universitaria de Sanidad*. Vol. 1, Núm. 1, 1999.



Distortion Products Otoacoustic Emissions as Markers of Tinnitus Persistence after Acute Acoustic Trauma

Dr Nottet Jean-Bertrand, M.D.

Hôpital d'instruction des Armées Desgenettes
108, Bld Pinel, 69275 Lyon cedex 03
France

Email : Jean-Bertrand.Nottet@wanadoo.fr

Dr Brossard Nicolas, M.D.

Hôpital d'instruction des Armées Desgenettes
108, Bld Pinel, 69275 Lyon cedex 03
France

Dr Suc Benoît, M.D.

Hôpital d'instruction des Armées Desgenettes
108, Bld Pinel, 69275 Lyon cedex 03
France

Dr Job Agnès, Ph.D

Centre de Recherches du service de santé des
Armées « Emile Pardé » -CRSSA-
24, avenue des maquis du Grésivaudan
38702 La Tronche cedex
France

Email : ajob@crssa.net

SUMMARY

It is a common thought in the physicians military community that persistence of tinnitus after acute acoustic trauma (AAT) essentially depends on severity of hearing thresholds shifts. Nevertheless it has never been demonstrated scientifically. Here we have looked for predictive factors of tinnitus duration after AAT by using hearing thresholds and distortion product otoacoustic emissions (DPOAEs). DPOAEs are otoacoustic emissions that revealed the functional state of the active cochlear mechanical processes. Young military adults under 25 years old, without history of tinnitus and hospitalized for an AAT were followed-up for 15 days. Examination during which the tinnitus state was recorded were carried out at three periods of time : 24 ± 5 hours, 72 hours and 15 days after the trauma. Two groups were defined according to their tinnitus duration after AAT (Group 1 < 72 hours and Group 2 > 72 hours). At 24 hours after AAT, hearing levels in the high frequency range did not differ significantly ($p=0.250$) between the short-lasting tinnitus group (Group 1) and the long-lasting tinnitus group (Group 2). In contrast, 24 hours after AAT, groups differed for DPOAEs, significantly ($p=0.016$). When statistical analyses took into account the severity of the acoustic trauma (Hearing levels held constant), DPOAEs were even more significantly different between the long and short-lasting groups ($p=0.007$). This result is mainly in favour of an outer hair cell (OHC) dysfunction in persistent tinnitus. Here we show that even moderated acoustic trauma could generated persistent tinnitus and was associated with poorer hearing thresholds recovery 15 days after AAT. DPOAEs may be useful predictors at 24 hours after the trauma, but also certainly before any deleterious acoustic event. Prospective studies are in progress. They could be interesting markers for subsequent pharmacological studies.

1.0 INTRODUCTION

Tinnitus is a symptom associated with acute acoustic trauma (AAT). Literature is abundant (Temmel et al. 1999; Axelsson and Hamernik 1987; Axelsson and Barrenäs 1992; Man and Naggan 1981; etc.). It is usually the most disabling functional after effect. In the context of the study, AAT was defined as an ear trauma provoked by impulsive noise exposure with specific clinical signs (i. e., High-pitched whistling

Nottet, J.-B.; Brossard, N.; Suc, B.; Job, A. (2005) Distortion Products Otoacoustic Emissions as Markers of Tinnitus Persistence after Acute Acoustic Trauma. In *New Directions for Improving Audio Effectiveness* (pp. 4-1 – 4-12). Meeting Proceedings RTO-MP-HFM-123, Paper 4. Neuilly-sur-Seine, France: RTO. Available from: <http://www.rto.nato.int/abstracts.aps>.

(tinnitus) and a notch on audiograms corresponding to hearing losses at frequencies between 3 and 6 kHz) typically described for noise-induced hearing loss (NIHL). Human investigation on AAT are not easy for statistical comparisons because acoustic sources (impulsive noise) that came into ears are very inhomogeneous (type of arms, distance, protections, etc.).

In the military environment, permanent tinnitus associated with marked NIHL is a subject of medical concern. In spite of this, there are only few studies measuring the fluctuations of hearing levels (HL) just after impulse noise (Dancer et al. 1991; Plinkert et al. 1995; Veuillet et al. 2001). There is a lack of studies of the persistence of tinnitus over the hours and days after the AAT. In practice, subjects are usually far from hospitals and are sent to be examined by an otologist only when tinnitus persists for more than a week. Therefore, it is rarely possible to examine injured ears at an early stage. For the purpose of this preliminary study, we urged the officers of several military regiments to send any subjects complaining of tinnitus after target practice rounds to the regional hospital as soon as possible, for ear examination.

For assessment of hearing status and for identifying potential predictive factors of persistent tinnitus, we tested hearing levels and used measurements of distortion product otoacoustic emissions (DPOAEs). The relationship between NIHL and otoacoustic emissions is abundantly described in the literature showing the advantages of using OAEs as sensitive indicators of hearing loss (Bonfils et al. 1988; Bonfils and Uziel 1989; Collet et al. 1991; Kemp et al. 1990; Kim et al. 1992; Martin et al. 1990; Smurzynski et al. 1990; Wilson 1992), but it had never been proved sensitive indicators of persistence of tinnitus after AAT.

OAEs are known to represent the cochlear mechanical activity of outer hair cells (OHC) and some studies suggest that one source of tinnitus may be related to dysfunctions in cochlear mechanical activity (Ceranik et al. 1998; Chery-Croze et al. 1994; Kemp 1981; Lind 1996; Shiomi et al. 1997; Zenner and Ernst 1993). In three studies, for example, chronic tinnitus from different etiologies associated or not with hearing loss was found to be associated with low level DPOAEs (Shiomi et al. 1997)

Here, the main test was to determine if tinnitus duration after an acute acoustic trauma (AAT) due to firearm discharge was associated with hearing levels or/and DPOAEs). Thus the purpose was to approach or localize tinnitus duration mechanisms.

2.0 MATERIELS & METHODS

2.1 Subjects

Subjects were soldiers who were sent to the hospital when they complained of tinnitus after target practice rounds during regular training. Subjects were 22 ± 2.3 years old, free from pre-existing AAT or tinnitus or significant hearing loss, as assessed by hearing control at the time of enlistment in the army.

On admission to the otology department of the hospital, subjects underwent a cortico-steroid therapy (5-d, methylprednisolone: Solumedrol IV[®], 2 mg/kg/day, as usually prescribed for acoustic traumas in French military hospitals). This therapy had been shown to slightly enhance the recovery in animal models (Lamm and Arnold 1998; Gervais d'Aldin et al. 1999) and in human (Suc and Asperge 1988; Suc et al. 1994). Afterwards they were subjected to OAE measurements and a pure tone audiometry.

2.2 Instrumentations

Pure tone audiometry (Midimate 622 Madsen) was performed at 1, 2, 3, 4, 6, and 8 kHz using a 5 dB-steps method in a sound proof room of the hospital otolaryngology unit. Hearing loss were expressed in dB re ISO 389 (ISO, 1975, 1983) with reference to a standard normal ear.

DPOAEs were recorded (ILO 92, Otodynamic Analyser, Otodynamic Ltd) in the sound-proof room, with subjects sitting in a comfortable chair and instructed to remain as quiet as possible. The three port probe (one microphone and two transducers) was positioned in the subject's outer ear canal using a foam ear tip.

2f₁-f₂ DPOAE levels were measured using the DP-gram procedure, in response to pure tones of moderate level (L₁=65 dB SPL, L₂=55 dB SPL), and plotted with respect to the f₂ primary tones, ranging from 1 to 6 kHz, (8 points/octave). The f₂/f₁ frequency ratio was set at 1.22.

For each subject, the mean responses in dB SPL at frequencies centered around 1 kHz (from 875 to 1437 Hz), 2 kHz (from 1437 to 2875 Hz) and 4 kHz (from 2875 to 6025 Hz) were calculated. Average noise floor was -14 ± 5 dB SPL across subjects and frequencies.

2.3 Procedures

We used a standardized questionnaire that is completed in French military hospitals following an AAT. This questionnaire included data concerning the date and time of the accident, the damaged ear(s), tinnitus characteristics (low or high frequency, continuous or intermittent), the wearing of ear protectors, the type of ear protectors, the type of deleterious acoustic source and the distance from the patient.

According to the questionnaire, the deleterious acoustic source was less than 1 m from the ear and was due essentially to gunfire from a French manufactured automatic machine gun called FAMAS. Spectrum data of FAMAS involved frequencies from 0.06 to 16 kHz. The average sound pressure level (10 ms) was 120-125 dB SPL and the peak pressure was 158 dB SPL at the onset (0.5 ms) of the shot (DTAT 1983, Parmentier 1996).

Data collection (n=37) revealed a large variability in the time delay between AAT and the first examination at the hospital (i.e., from 2 hours to 5 days). In order to insure a homogenous delay group, we restricted the study to subjects with delays between 16 h and 32 h. This selection yielded a group of 24 young military subjects, examined at three time points after the AAT: (1) at 24 ± 5 hours (from 16 h to 32 h), (2) at 72 hours, (3) at 15 days.

Over the 15 days period, patients stay for one week in quiet at the hospital and for one week on sick leave at home with strict recommendations of no noise exposure.

During training wearing earplugs is compulsory. 54 % had no earplug at the time of the accident (accidental fall). The remaining subjects suspected a non correct insertion of their foam earplugs.

At the first examination, all patients had a continuous high frequency tone (whistling) in their ear, fluctuating in pitch and intensity. Eight subjects had persistent tinnitus less than 72 h after AAT, 16 subjects had persistent tinnitus superior to 72 hours, among them, 5 subjects still had persistent tinnitus at 15 days after the AAT. When tinnitus was bilateral (one subject), data concerning the more severely affected ear (notch on audiogram) was used for analysis.

To test for OAEs and hearing levels differences between subjects with various tinnitus duration, two classes were defined : Group 1 (n= 8) represented subjects with short-lasting tinnitus (< 72 hours) and Group 2 (n= 16) represented subjects with long-lasting tinnitus (> 72 hours). In addition, in order to contrast strongly short-lasting tinnitus with long-lasting tinnitus, we compared group 1 with a subgroup of group 2 : the five subjects with still persistent tinnitus at 15 days (group 3). Non parametric Mann-Whitney tests to compare these two groups were performed in this particular case for a better statistical robustness.

Otherwise, statistical analyses were carried out by using analysis of variance ANOVA with a between-subject factor "tinnitus duration groups" and a within-subject factor "frequencies" for comparing hearing levels and DPOAEs levels. Equality of variance were checked (Levene's test) for DPOAEs and HL.

A significant inter dependence between hearing levels and otoacoustic emissions after impulse noise exposure was previously demonstrated (Wagner and Plinkert 1999). For this reason it seemed important (in order to extract the role played by tinnitus duration itself) to analyse the relationship between HL, DPOAEs and tinnitus duration by covariance analysis COVAR, (Mc Pherson 1990) using the 24h-high frequency worse hearing level (notch) as an indicator of AAT severity (covariate).

This specific analysis was interesting because the adjustment took into account the various severity of AAT and thus allowed us to test differences between tinnitus duration groups all audiogram notches equal. The present study was approved by a national public institutional ethics committee on the use of human subjects in biomedical research (CCPRB LYON-B/06/07/2000). Written, informed consent was obtained from each subject.

3.0 RESULTS

3.1 Hearing levels and DPOAEs features

In this population, right and left ears were affected in equal proportion (Table 1). AAT was characterized by a hearing loss at high-frequency showing a notch on audiograms with a maxima focused at 4 kHz (38% of cases), 3 kHz (33%), 6 kHz (21%) or 5 kHz (8%) with a various severity comprised between 10 to 70 dB HL (Table 1).

We calculated an overall high-frequency hearing level with mean hearing losses at high frequencies

Maximal high-frequency HL at 24 hours were strongly negatively correlated with 24 h-DPOAEs (Pearson correlation coefficient, $r = -0.81$, $p < 0.001$) (Figure 1). DPOAEs decreased as hearing thresholds increased.

Subject	EAR	Loss max (dB) at 24 h	Most affected Frequency. in kHz	Tinnitus Duration
N°2	L	10	5	< 72 h
N°5	L	15	4	< 72 h
N°8	R	20	4	< 72 h
N°10	R	25	3	< 72 h
N°11	L	25	4	< 72 h
N°13	L	35	3	< 72 h
N°16	L	50	3	< 72 h
N°22	R	65	3	< 72 h
N°1	L	10	6	> 72 h
N°3	R	15	6	> 72 h
N°4	R	15	6	>72 h (*)
N°6	L	15	4	> 72 h
N°7	L	20	4	>72 h (*)
N°9	R	20	3	> 72 h
N°12	L	30	4	> 72 h (*)
N°14	L	40	3	> 72 h
N°15	L	45	3	> 72 h
N°17	R	55	5	> 72 h
N°18	R	55	3	> 72 h (*)
N°19	R	55	6	> 72 h
N°20	R	60	4	>72 h
N°21	L	60	4	> 72 h
N°23	R	70	6	> 72 h
N°24	L	70	4	> 72 h (*)

Table 1: Clinical data of 24 military subjects hospitalized 24 hours after an acute acoustic trauma (AAT) due to gunfire. R: Right, L: Left. (*) Tinnitus persistent at 15 days.

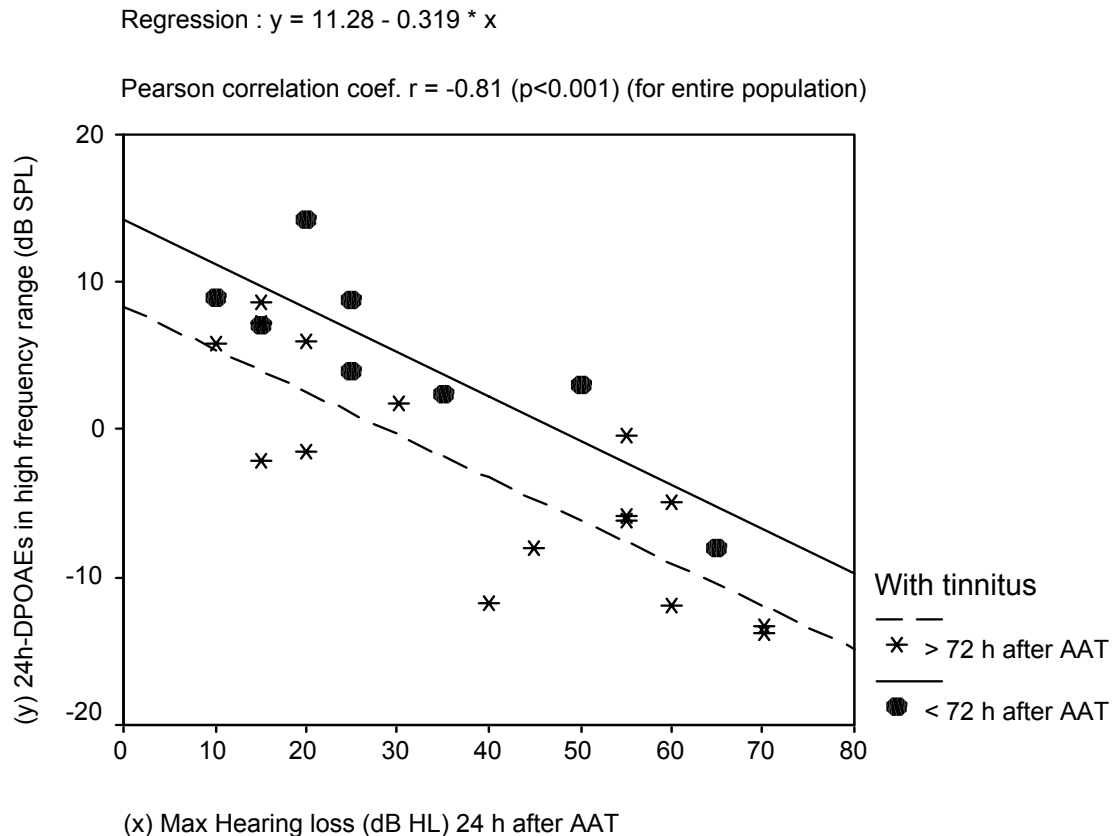


Figure 1 : Scatter plot of the association between 24 h-high-frequency-DPOAEs and 24 h-maximum hearing loss. Comparison between tinnitus duration groups. Regression slope was lower with long-lasting tinnitus.

3.2 Hearing levels and tinnitus duration after AAT

Analysis of variance with repeated measures in Table 2 showed that hearing levels were not significantly different between Group 1 and Group 2 in the 1, 2 and 3-6 kHz range at all periods of time after AAT. The same result was found with Mann-Whitney tests between Group 1 and Group 3 in Table 3.

Mean hearing levels (HL) were presented in Figure 2. Differences between groups were not significant at 24h after AAT. At 72 hours and 15 days recovery to normal threshold was poorer in the long-lasting tinnitus groups but were not statistically significant.

Between tinnitus duration groups	TIME	HL p value	DPOAE p value
<i>At 3-6 kHz</i>			
ANOVA	24 h	0.250	0.016
	72 h	0.125	0.018
	15 d	0.113	0.112
COVAR (covariate= HL)	24 h		0.007
	72 h		0.030
	15 d		0.206
Between tinnitus duration groups	TIME	HL p value	DPOAE p value
<i>At 2 kHz</i>			
ANOVA	24 h	0.345	0.677
	72 h	0.930	0.193
	15 d	0.250	0.174
COVAR (covariate= HL)	24 h		0.429
	72 h		0.160
	15 d		0.118
Between tinnitus duration groups	TIME	HL p value	DPOAE p value
<i>At 1 kHz</i>			
ANOVA	24 h	0.461	0.704
	72 h	0.663	0.160
	15 d	0.314	0.217
COVAR (covariate= HL)	24 h		0.531
	72 h		0.095
	15 d		0.138

Table 2 : Synopsis of ANOVA and Covariance analysis comparing hearing levels and DPOAEs of two tinnitus duration groups (long and short-lasting tinnitus).

FREQUENCY BANDS	TIME	HL p value	DPOAE p value
3-6 kHz	24 h	0.724	0.019
	72 h	0.435	0.030
	15 d	0.065	0.045
2 kHz	24 h	0.093	0.943
	72 h	0.622	0.354
	15 d	0.524	0.171
1 kHz	24 h	0.435	0.435
	72 h	0.943	0.045
	15 d	0.171	0.093

**Table 3 : Non parametric tests comparing more contrasted tinnitus duration groups after AAT:
group 1 (tinnitus <72 h) and group 3 (tinnitus present at 15 days).**

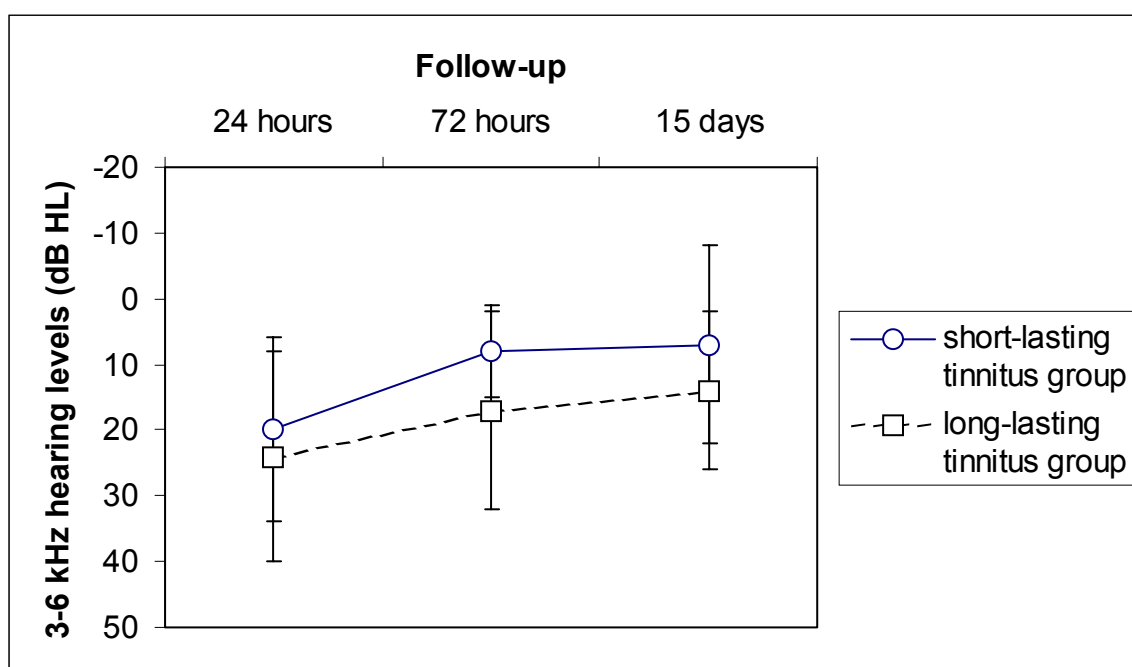


Figure 2 : Mean 3-6 kHz hearing thresholds at different periods of time in the short-lasting tinnitus group and long lasting-tinnitus group

3.3 DPOAEs and tinnitus duration after AAT

In Figure 1, the DPOAEs regression slope of the long-lasting tinnitus group was lower than the regression slope of the short-lasting tinnitus group following a marked parallelism.

This result means that independently of severity of trauma tinnitus persistence was associated with an other parameter not depending of severity of notches.

We showed that DPOAEs were significantly lower in Group 2 in the 3-6 kHz range at 24 hours and 72 hours after AAT (Table 2). When the analysis was adjusted with the covariate 24h-high frequency worse HL, significance was higher ($p=0.007$). This result clearly suggests that hearing loss after AAT was a confounding factor slightly masking the phenomenon of tinnitus persistence.

In Table 3, Mann-Whitney comparisons between the most contrasted groups (1 and 3) confirmed the previous results. Differences were also present at 1 kHz. DPOAEs were lower at 72h and 15d after trauma in the long-lasting tinnitus group.

4.0 DISCUSSION

In moderate AAT, the association between hearing levels and tinnitus duration is weak. The results of the study, clearly showed that hearing levels (at least up to 70 dB HL) are not early markers to explain tinnitus persistence after an AAT compare to DPOAEs.

Tinnitus persistence after AAT is the most disabling after effect, it generates many problems at psycho-social and operational levels.

This study shows that even in AAT, tinnitus persistence mainly depends on cochlear outer hair cells (OHC) dysfunctions rather than cochlear inner hair cells (IHC) dysfunctions.

In general, tinnitus is nearly always present at AAT onset, which is the 'normal' response to an acoustic over stimulation. Persistence of tinnitus 15 days after AAT seemed clearly to belong to an abnormal phenomenon at recovery that had never been fully explained. This study provided some new information.

Later after trauma, at 72 hours and 15 days, tinnitus was stopped in one group, we found significant differences between groups at the edge of the trauma frequencies in particular at 1 kHz.

As Avan and coll. (Avan et al. 1991) had already observed, amplitudes of OAEs seemed not only the reflection of the stimulated zone but the state of the whole cochlear partition, and thus after an AAT, several peripheral and central factors might influence cochlear response.

As soon as no reference DPOAEs or audiograms could be provided for the traumatized subjects, we might suspected latent non frequency specific sub clinical OHC dysfunctions and alterations.

First, it could possibly partly come from history of infectious disease (otitis media) or head trauma that are known to potentiate the risk of chronic tinnitus and acoustic trauma when normal hearing subjects were exposed to noise (Job et al. 1999; Job and Nottet 2002; Ceranic et al. 1998).

Second, abnormality might also be present at the retrocochlear level and in central auditory activity as suspected before some authors (Attias et al. 1993 ; Komiya and Eggermont 2000). Furthermore individual susceptibilities at emotional level were probably involved. Tinnitus has been shown to activate the limbic system which controls emotional processing (Gardner et al. 2002; Lockwood et al. 1998; Mirz et al. 2000). A dysfunction in the centrally-controlled regulation at the synaptic level of OHC via the medial olivocochlear system (MOC) possibly occurred in response to the stress provoked by the impulsive noise and acoustic trauma (Horner et al. 2001). A recent retrospective study on tinnitus persistence induced by AAT tended to strengthen the hypothesis of an involvement of psychological factors (Mrena et al., 2002).

Recently, an experimental study on clinical normal hearing subjects (Job et al., 2004) showed that tense-anxious subjects were more susceptible to develop transient tinnitus after moderate impulse noise exposure to gunfire and related to slight alterations in cochlear DPOAEs at 3 kHz. Among young subjects (22 ± 2 years old) reporting transient tinnitus we found at 2 and 1 kHz an edge effect between subjects having no history of tinnitus and subjects with previous history of tinnitus. A significant association was found between tension-anxiety and history of tinnitus.

Thus other networks besides specific sensory one may be involved in tinnitus persistence after AAT. These networks had to be elicited and taken into account. DPOAEs were sensitive to the duration of tinnitus independently of severity of audiogram notches indicating the involvement of OHC and its regulation mechanism.

Further human experimental or prospective studies on DPOAEs at larger scale should be carried out, in order to detect earlier susceptibility to tinnitus persistence in subsequent acoustic over exposure, highly probable in a military environment.

5.0 REFERENCE

- [1] Attias J, Urbach D, Gold S, Shemesh Z. Auditory event related potentials in chronic tinnitus patients with noise induced hearing loss. *Hear. Res.* 71: 106-113, 1993.
- [2] Avan P, Bonfils P, Loth D, Narcy P, Trotoux J. Quantitative assessment of human cochlear function by evoked otoacoustic emissions. *Hear. Res.* 52: 99-112, 1991.
- [3] Axelsson A, Barrenäs ML. Tinnitus in noise-induced hearing loss. In: : Noise-Induced Hearing Loss. Dancer A, Henderson D, Salvi R, Hamernik RP. Eds, Mosby-Year Book, St Louis, pp. 272-273, 1992.
- [4] Axelsson A, Hamernik RP. Acute acoustic trauma. *Acta. Otolaryngol. Stock.* 104: 225-233, 1987.
- [5] Bonfils P, Piron JP, Uziel A. A correlative study of evoked oto acoustic emission properties and audiometric thresholds. *Arch. Otorhinolaryngol.* 245: 53-56, 1988.
- [6] Bonfils P, Uziel A. Clinical applications of evoked acoustic emission : Results in normally hearing and hearing impaired subjects. *Ann. Otol. Rhinol. Laryngol.* 98: 326-331, 1989.
- [7] Ceranic BJ, Prasher DK, Raglan E, Luxon LM. Tinnitus after head injury evidence from otoacoustic emissions. *J. Neurol. Neurosurg. Psychiatry.* 65: 523-529, 1998.
- [8] Chery-Croze S, Moulin A, Collet L, Morgon A. Is the test of efferent system function a relevant investigation in tinnitus, *Br. J. Audiol.* 28 : 13-25, 1994.
- [9] Collet L, Morgon A, Veuillet E, Gartner M. Noise and medial olivocochlear system in Human. *Acta otolaryngol.* 111: 231-233, 1991.
- [10] Dancer A, Grateau P, Cabanis A, Vaillant T, Lafont D. Delayed temporary threshold shift induced by impulse noises (weapon noises) in men. *Audiology.* 30 : 345-356, 1991.
- [11] DTAT. Direction technique des armements terrestres : Recommandations relatives à l'évaluation physio acoustique du pouvoir lésionnel des bruits. Rapport AT83/27/28. ETBS, 1983.

- [12] Gardner A, Pagani M, Jacobsson H, Lindberg G, Larsson SA, Wagner A, Hallstrom T. Differences in resting state regional cerebral blood flow assessed with 99m Tc-HMPAO SPECT and brain atlas matching between depressed patients with and without tinnitus. *Nucl. Med. Commun.* 23: 429-439, 2002.
- [13] Gervais d'Aldin C, Cherny L, Devrière F, Dancer A. Treatment of acoustic trauma. In *Ototoxicity*. Henderson et al., Eds, *Annals of the new York Acad of Sc.*, 884, 328. 1999.
- [14] Horner KC, Giraudet F, Magnan J, Chays A, Cazals Y. Sympathectomy improves the ear's resistance to acoustic trauma- Could stress render the ear more sensitive ? *Eur. J. Neurosci.* 13 :405-408, 2001.
- [15] ISO 389. Acoustics-Standard Reference Zero for calibration of pure-tone audiometers. International Organization for Standardization, Geneva. 1975.
- [16] ISO 389 AD1. Addendum n°1 to ISO 389-1975. International Organization for Standardization, Geneva. 1983.
- [17] Job A, Raynal M, Rondet P. Hearing loss and use of personal stereos in young adults with antecedent of otitis media. *The Lancet.* 353: 35, 1999.
- [18] Job A, Nottet J-B. DPOAEs in young normal-hearing subjects with histories of otitis media : evidence of sub-clinical impairments. *Hear. Res.* 167: 28-32, 2002.
- [19] Job A, Cian C, Esquivié D, Leifflen D, Trousselard M, Charles C, Nottet J-B. Moderate variations of mood/emotional states related to alterations in cochlear otoacoustic emissions and tinnitus onset in young normal hearing subjects exposed to gun impulse noise. *Hear. Res.* 193:31-38, 2004.
- [20] Kemp DT. Physiologically active cochlear micromechanics-one source of tinnitus. In D. Evered G. Lawrenson, Eds., *Tinnitus*. Pitman Books, London, pp. 54-81, 1981.
- [21] Kemp DT, Ryan S, Bray P. A guide to effective use of otoacoustic emissions. *Ear. Hear.* 11: 93-105, 1990.
- [22] Kim DO, Leonard G, Smurzynski J, Jung MD. Otoacoustic Emissions and noise induced hearing loss : Human studies. In : *Noise-Induced Hearing Loss*. Dancer A, Henderson D, Salvi R, Hamernik RP. Eds, Mosby-Year Book, St Louis, pp. 98-105, 1992.
- [23] Komiya H, Eggermont JJ. Spontaneous firing activity of cortical neurons in adult cats following pure tone trauma induced at five weeks of age. *Acta. Otolaryngol.* 120, 750-756. 2000.
- [24] Lamm K, Arnold W. The effect of prednisolone and non-steroidal anti-inflammatory agents on the normal and noise damaged guinea pig inner ear. *Hear. Res.* 115, 149-161.
- [25] Lind O. Transient-evoked otoacoustic emissions and contralateral suppression in patients with unilateral tinnitus. *Scand. Audiol.* 25: 167-172, 1996.
- [26] Lockwood AH, Salvi RJ, Coad ML, Towsley ML, Wack DS, Murphy BW. The functional neuroanatomy of tinnitus: evidence for limbic system links and neural plasticity. *Neurology.* 50: 114-120, 1998.

- [27] Man A, Naggan L. Characteristics of tinnitus in acoustic trauma. *Audiology*. 20: 72-78, 1981.
- [28] Martin GK, Ohlms LA, Franklin D, Harris FP, Lonsbury-Martin BL. Distortion product emissions in humans. III. Influence of sensorineural hearing loss. *Ann. Otol. Rhinol. Laringol.* 99: 30-42, 1990.
- [29] Mc Pherson G. Statistics in scientific investigations. Its Basis, Application and Interpretation-Analysis of covariance. Springer-Verlag New York, pp. 417-418, 1990.
- [30] Mirz F, Gjedde A, Sodkilde-Jrgensen H, Pedersen CB. Functional brain imaging of tinnitus-like perception induced by aversive auditory stimuli. *Neuroreport*. 11: 633-637, 2000.
- [31] Mrena R, Savolainen S, Kuokkanen JT, Ylikoski J. Characteristics of tinnitus induced by acute acoustic trauma : a long-term follow-up. *Audiol. Neurotol.* 7: 122-130, 2002.
- [32] Parmentier G, Kronenberger G, Dancer A. Mesures des bruits par une arme d'infanterie et une arme de chasse en champs proche et lointain. ISL. Rpt RT 521/96, 1996.
- [33] Plinkert PK, Hemmert W, Zenner HP. Comparison of methods for early detection of noise vulnerability of the inner ear : amplitude reduction of otoacoustic emissions are most sensitive at submaximal noise impulse exposure. *H.N.O.* 43: 89-97, 1995.
- [34] Shiomi Y, Tsuji J, Naito Y, Fujiki N, Yamamoto N. Characteristics of DPOAE audiogram in tinnitus patients. *Hear. Res.* 108: 83-88, 1997.
- [35] Smurzynski J, Leonard G, Kim DO, Lafrenière DC, Jung MD. Distortion product otoacoustic emissions in normal and impaired adults ears. *Arch. Otolaryngol. Head. Neck. Surg.* 116: 1309-1316, 1990.
- [36] Suc B, Asperge A. Les acouphènes des traumatismes sonores par FAMAS. *Médecine et Armées*, 16 :207-210. 1988.
- [37] Suc B, Poulet M, Asperge A, Vix J, Barberot J-P, Doucet F. Evolution clinique des traumatismes sonores aigus. Bilan d'une étude de 250 cas. *Ann. Otolaryngol (Paris)*, 111 :319-324. 1994.
- [38] Temmel AF, Kierner AC, Steurer M, Riedl S, Innitzer J. Hearing loss and tinnitus in acute acoustic trauma. *Wien. Klein. Wochenschr.* 111: 891-893, 1999.
- [39] Veuillet E, Martin V, Suc B, Vesson JF, Morgon A, Collet L. Otoacoustic emissions and medial olivocochlear suppression during auditory recovery from acoustic trauma in human. *Acta. otolaryngol.* 121:278-283, 2001.
- [40] Wagner W, Plinkert PK. The relationship between auditory threshold and evoked otoacoustic emissions. *Eur. Arch. Otorhinolaryngol.* 256: 177-188, 1999.
- [41] Wilson JP. Otoacoustic Emissions and noise induced hearing loss . In : Noise-Induced Hearing Loss. Dancer, A., Henderson, D., Salvi, R., Hamernik, R.P. Eds, Mosby-Year Book, St Louis, pp. 89-97, 1992.
- [42] Zenner HP, Ernst A. Cochlear-motor, transduction and signal-transfer tinnitus : models for three types of cochlear tinnitus. *Eur. Arch. Otorhinolaryngol.* 249: 447-462, 1993.

Auditory Tests for the Early Detection of Noise-Susceptible Individuals – A Literature Study

Joos Vos

TNO Human Factors
P.O. Box 23, 3769 ZG Soesterberg
The Netherlands

vos@tm.tno.nl

TESTS FOR THE DETECTION OF NOISE-SUSCEPTIBLE INDIVIDUALS

We reviewed the literature on the availability of auditory tests that determine, prior to being affected by significant hearing loss, to what extent an individual is susceptible to noise. The predictive power of tests that determine the acoustic reflex responsivity, and both temporal integration and non-linearities of the hearing system, may be large, whereas the predictability of tests that determine otoacoustic emissions and head related transfer functions seems to be small. The validity of the indices, however, can only be determined in longitudinal studies. To our knowledge, the results of such studies have never been reported.

1.0 INTRODUCTION

As a result of noisy military systems and the limited applicability of personal hearing protection, there continues to be a high incidence of noise-induced hearing loss (NIHL) among military personnel. One of the ways to reduce the adverse effects of noise, such as a decreased hearing acuity and a noticeable reduction in the ability to understand speech in noisy conditions, might be the selection of persons who are less susceptible to noise. The Netherlands Ministry of Defense contracted TNO Human Factors to investigate whether at present reliable auditory tests are available to determine in an early stage, i.e. prior to being affected by significant hearing losses, to what extent an individual is susceptible to noise.

2.0 DEGREE OF INDIVIDUAL DIFFERENCES

To give an idea of the degree in which the susceptibility to noise varies among individuals, Figure 1 reproduces median hearing thresholds of metal workers of the drop-forging industry and office and canteen personnel (control) reported in Taylor et al. [1], as a function of the frequency of the test tone. The metal workers had been exposed to A-weighted equivalent sound levels ($T = 8$ h) varying from 102 to 112 dB. With the help of vertical lines, the 90th (top) and 10th (bottom) percentiles are given as well. The data are shown for five separate groups with ages varying between 15 and 24 years up to between 55 and 64 years. The spread in the thresholds (difference between the thresholds for the 10th and the 90th percentiles) is also given in Table 1. Averaged across frequency, the spread increases from about 20 dB for the 15-24-year-old up to about 40 dB for the 35-44-year-old. For the two higher categories, the spread stabilizes at 45 dB, which must be an artifact caused by the employees with very high hearing losses who prematurely gave up their jobs. Averaged across the age groups the spread obtained for the lower frequencies of 0.5 and 1 kHz is about 15 dB smaller than that for the higher frequencies of 2 and 4 kHz. As references, Table 1 also shows the spread (10th - 90th percentiles) in the hearing thresholds for otologically normal persons that can be revealed from the data sets given in ISO 7029 [2]. The spread is averaged across males and

Vos, J. (2005) Auditory Tests for the Early Detection of Noise-Susceptible Individuals – A Literature Study. In *New Directions for Improving Audio Effectiveness* (pp. 5-1 – 5-6). Meeting Proceedings RTO-MP-HFM-123, Paper 5. Neuilly-sur-Seine, France: RTO.
Available from: <http://www.rto.nato.int/abstracts.aps>.

females. For persons older than 34 years, the spread in the thresholds for the highly exposed at frequencies of 0.5, 1, and 2 kHz is 10 to over 30 dB higher than that for the otologically normal persons.

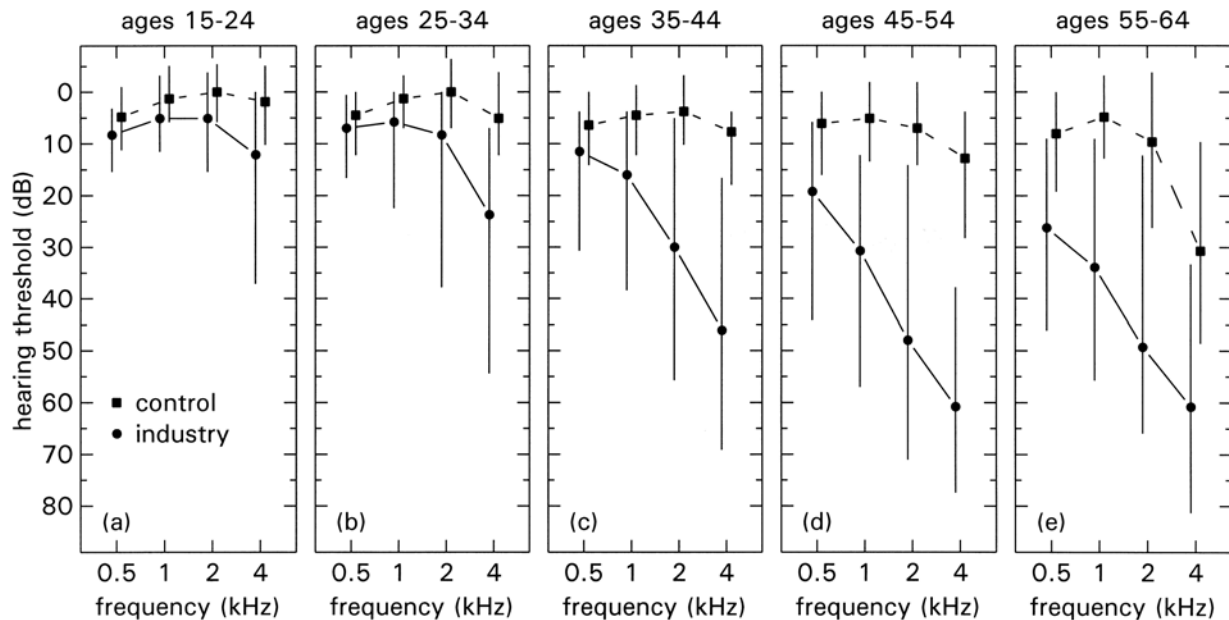


Figure 1: Median hearing thresholds of metal workers (industry) and a reference group (control) reported in Taylor et al. [1], as a function of the frequency of the test tone and for five separate age-groups. The top and bottom of the vertical lines represent the 90th and the 10th percentiles, respectively.

Table 1: Spread (in dB) in the hearing thresholds, expressed as the difference between the thresholds for the 10th and the 90th percentiles, for various frequencies and for workers in the drop-forging industry (DF) included in the study of Taylor et al. [1] and for otologically normal persons (ON) as described in ISO 7029 [2]. The difference between DF and ON is indicated by d; the mean (M) of these differences averaged across frequency or age is given in the last column and bottom row, respectively.

Age	500 Hz			1000 Hz			2000 Hz			4000 Hz			M
	DF	ON	d	DF	ON	d	DF	ON	d	DF	ON	d	
15-24	12	14	-2	15	14	1	19	16	3	37	19	18	5
25-34	16	15	1	22	15	7	38	17	21	47	20	27	14
35-44	27	16	11	35	16	19	51	20	31	53	25	28	22
45-54	38	18	20	45	18	27	57	23	34	40	32	8	22
55-64	37	21	16	47	21	26	54	28	26	48	41	7	19
M			9			16			23			18	16

3.0 AUDITORY TESTS

Changes in the hearing acuity that result from high sound exposures are often expressed as a temporary (TTS) or a permanent (PTS) shift in the hearing threshold. The effective individual sound exposure may be moderated by the acoustic reflex and anatomical properties of the ear canal and the concha. In addition, there are various psychoacoustical tests that provide insight in the way in which the hearing organ

integrates sounds in the time- and frequency domains. Since changes in these last features are observed earlier than changes as measured by means of the TTS and the PTS, much research focussed on the predictability of changes in especially the TTS from the results of tests that determine the integrating properties of the hearing system.

3.1 Acoustical Tests

Intense acoustic stimulation causes reflexive contraction of the middle-ear muscles. As long as the muscles are contracted, the transmission of sound energy is reduced. It is believed that the reduction protects the inner ear from intense sound. The literature is not quite consistent with respect to the relation between the magnitude of the *acoustic reflex* and the size of the TTS after sound exposure. Gerhardt and Hepler [3], for example, found a significant negative correlation between these two measures for tone bursts with a frequency of 1.4 kHz, but not for tone bursts with frequencies of 0.5 and 2 kHz. The interpretation about the cause-effect relationships is hampered by their observations that the magnitude of the acoustic reflex is reduced by the noise exposure itself. From a large-scale study on the relevance of various features of the acoustic reflex, comprising also measures such as the latency, rise and decay times of the reflex, Colletti and Sittoni [4] concluded that the stapedius reflex may play an important role in the protection against noise. In general however, it is felt that for a complete understanding of the cause-effect relationships, longitudinal studies are required.

As a result of individual differences in the shape of the ear canal entrance, the length of the ear canal, the cross-section area along the canal and the acoustic impedance of the tympani membrane, the *effectivity of the sound transmission* (HRTF) may differ from person to person. Hellström [5] determined the degree of TTS after exposure to noise for 36 subjects who were assigned to one of three groups depending on their mean sound transfer function for frequencies between about 3 and 6 kHz: low, middle, or high. The mean differences between the groups with a low and a high effectivity ranged between about 6 to 9 dB. The differences between the TTS as determined for either 2- or 4-kHz narrow-bandpass filtered white noise was about 5 dB at most. For the assessment of the practical relevance of these mainly anatomical differences, the test should be extended for lower frequencies of, for example, 250, 500 and 1000 Hz. Since for these lower frequencies the differences among the individual HRTFs are small, the overall effect on the TTS must be smaller as well. A related aspect, of course, is the shape of the concha and how far the ears project from the head. Ward [6] reported that there was a positive correlation between the TTS after exposure to noise and the magnitude of the ear projection.

Otoacoustic emissions (OAEs) are sounds in the ear canal that result from activity of the outer hair cells. When OAE is evoked by long series of pulse- or click trains, it is called a transient-evoked OAE (TEOAE). An OAE that occurs as a result of distortion is called a distortion product OAE (DPOAE). An OAE can also be produced spontaneously (SOAE), without deliberate sound stimulation. In several studies, the magnitude of TEOAE has been related to the hearing thresholds of groups of subjects with various degrees of hearing loss. By and large the results showed that for subjects with clearly measurable TEOAEs the hearing loss was generally small. Absence of TEOAEs, or a low reproduceability of specific components of them, however, were no reliable indicators of significant hearing losses: they could be found both in subjects with normal hearing and in subjects with deviating hearing thresholds (e.g., see Attias et al. [7]; Engdahl and Tambs [8]; Tognola et al. [9]). As a result, the TEOAE cannot be used in determining an individual's risk of hearing loss. This conclusion also holds for the DPOAE. The present author is unaware of reports about the relation between SOAE and hearing loss.

3.2 Psychoacoustical Tests

In various studies reported by Humes and colleagues (e.g., see Humes [10]; Humes and Bess [11]), the *threshold of distortion* is arbitrarily defined as the total sound level of two interfering tones at which the

beats are just no longer audible. For frequencies higher than 1 kHz, it has been found that the TTS decreases with increasing threshold of distortion. This threshold has never been related to PTS.

The *loudness discrimination* test proposed by Bienvenue et al. [12] determines the ability to detect small changes in the level of various sounds. People with a certain degree of noise induced hearing loss are able to detect smaller changes in sound level (loudness recruitment) than normal hearing subjects. Bienvenue et al. concluded that this discrimination test was an early and sensitive indicator of TTS. Without separate data on the TTS, however, this conclusion seems to be premature. In a study with naval aviators categorized as having either normal hearing or hearing threshold levels greater than 40 dB at frequencies between 4 and 8 kHz, Thomas and Williams [13] found that for frequencies of 2 and 4 kHz, the aviators with the hearing loss had deviating scores on a comparable discrimination task. Especially surprising was the result obtained for the frequency of 2 kHz: although the hearing threshold of the noise-susceptible subjects was still normal at this frequency, the result of the discrimination test already implicitly indicated the occurrence of loudness recruitment. This suggested that the test might serve as an “early warning” of noise-induced damage.

For short tone bursts the hearing threshold depends on both the frequency and the duration of the tones. For subjects with normal hearing, the trade-off between level and time is equal to –10 dB for a tenfold increase in tone duration. In a study reported by Lawton and Robinson [14] the mean slopes for a group of young and a group of older men were equal to –9.3 and –7.5 dB, respectively. For both groups the mean integration time was equal to about 150 ms. A portion of the variance in the hearing threshold at 4 kHz could be accounted for by the slope and the product of the slope and the logarithm of the integration time. For the younger and older groups the explained variance was equal to 28% and 60%, respectively. Lawton and Robinson [14] concluded that *brief-tone audiometry* might be a promising tool that should be included in future research on the detection of noise-susceptible subjects. The latter conclusion was confirmed in a subsequent study with soldiers who had experienced very heavy exposure to shooting noise whilst retaining substantially normal hearing sensitivity (Lawton and Robinson [15]).

The *threshold for octave masking* is a measure for cochlear distortion. Humes et al. [16] obtained a negative correlation between the masking threshold and the TTS: normal hearing subjects with a high threshold had a lower TTS. In the study reported by Lawton and Robinson [14] the slope of the masking threshold was significantly correlated with the hearing threshold level at 4 kHz. Especially for the group of older men the variance in the hearing thresholds accounted for by this slope was as large as 50%. After various analyses, Lawton and Robinson [15] concluded that resistance to hearing loss is associated with a steep slope of the octave masking threshold.

Bergman et al. [17] used the notched-noise procedure to determine changes in auditory *frequency selectivity* in noise-exposed industrial workers. The frequency selectivity was defined as the difference between the threshold of a test tone in broadband noise and the threshold of a test tone in noise with a spectral notch. A clear negative correlation between frequency selectivity and the hearing threshold level at 4 kHz was obtained. Since even for subjects with normal hearing there exists considerable variance in the frequency selectivity, this test might be a predictor for individual susceptibility to noise at the long term.

Speech reception thresholds and tone thresholds are clearly correlated. For the thresholds measured in 200 individuals with noise-induced hearing loss, Smoorenburg [18] found that the speech reception threshold in noise could be adequately predicted ($r = 0.72$) from the pure-tone average at 2 and 4 kHz. From the statistical association obtained in a typical cross-sectional study it cannot be inferred that hearing loss is a predictor for individual susceptibility to noise. Humes [10], however, referred to a study in which for the normal hearing subjects a significant positive correlation was found between the threshold of distortion and the discrimination scores for speech in noise at signal-to-noise ratios of 0 and 6 dB. Since a positive relation was found also between the threshold of distortion and the TTS, Humes’s hypothesis that speech reception might be a predictor for individual susceptibility to noise is plausible.

4.0 NONACOUSTIC FACTORS

The literature suggests that eye color is a predictor of individual susceptibility to noise: people with blue eyes might be more sensitive to noise than people with brown eyes (e.g., see Thomas and Williams [13]). A small effect of gender is reported also, with females being less sensitive than males. Nicotine intake might slightly enhance the prevalence of hearing loss (Henderson et al. [19]).

5.0 DISCUSSION

We reviewed the literature on indices of susceptibility to NIHL. In the literature various auditory tests are proposed that measure items such as loudness discrimination, octave masking, frequency selectivity, temporal integration, and the acoustic reflex responsivity. The main objective in the pertinent studies was to determine the correlation between the test results and the noise-induced TTS. Unfortunately, this approach has several weaknesses. Firstly, at present there are no convincing data showing that at an individual level, PTS can be predicted from TTS. Secondly, correlations between auditory test results and the TTS do not yield sufficient knowledge about real causal relationships. The difference in the effectivity of the sound transmission that results from anatomical differences among ears, is an example of a simple causal relationship. Especially the last ten years, much research has been devoted also to the relationship between various kinds of otoacoustic emissions and the hearing threshold. For subjects with measurable emissions, the hearing loss is generally small. Absence of emissions, however, is no reliable indicator of significant hearing losses.

We may conclude that the validity of the various indices of the individual susceptibility to NIHL can only be determined in longitudinal studies. In such studies, the promising auditory tests have to be administered to the participants at an early age before being affected by hearing losses. In our view the test battery should include acoustic reflex measurements, the determination of non-linearities of the hearing system (threshold of octave masking or threshold of distortion) and brief-tone audiometry. One might also consider including a test for speech reception in noise. The results obtained in the present literature study suggested that the measurement of otoacoustic emissions had a low priority. The disappointing results obtained in the cross-sectional studies, however, do not necessarily imply that individual changes in the otoacoustic emission observed in a lapse of many years cannot be used in the early detection of noise-susceptibility. This notion, and given that the measurement of otoacoustic emissions is neither time-consuming nor a burden to the subject, may justify inclusion of such a test also. After 10, 20, and possibly even after 30 years, the test results should then be related to the PTS that each participant may have suffered. In the literature, the need for such longitudinal studies has been frequently emphasized. To our knowledge, however, the results of such comprehensive studies have never been reported.

- [1] Taylor, W., Lempert, B., Pelmeier, P., Hemstock, I., & Kershaw, J. (1984). Noise levels and hearing thresholds in the drop forging industry. *J. Acoust. Soc. Am.* 76 (3), 807-819.
- [2] ISO 7029 (2000). Acoustics – Statistical distribution of hearing thresholds as a function of age. Geneva, Switzerland: International Organization for Standardization.
- [3] Gerhardt, K.J., & Hepler, E.L. (1983) Acoustic stapedius reflex activity and behavioral thresholds following exposure to noise. *J. Acoust. Soc. Am.* 74, 109-114.
- [4] Colletti, V., & Sittoni, V. (1986). Noise history, audiometric profile, and acoustic reflex responsivity. In *Basic and Applied Aspects of Noise-Induced Hearing Loss* (eds. R.J. Salvi, D. Henderson, R.P. Hamernik, and V. Colletti) Series A: Life Sciences, Vol. III (New York: Plenum Press), pp. 247-269.

- [5] Hellström, P.-A. (1993). The relationship between sound transfer functions from free sound field to the eardrum and temporary threshold shift. *J. Acoust. Soc. Am.* 94 (3), Pt. 1, 1301-1306.
- [6] Ward, W. D. (1965). The concept of susceptibility to hearing loss. *J. Occup. Med.* 7 (12), 595-607.
- [7] Attias, J., Furst, M., Furman, V., Reshef, I., Horowitz, G., & Bresloff, I. (1995). Noise-induced otoacoustic emission loss with or without hearing loss. *Ear & Hearing* 16 (6), 612-618.
- [8] Engdahl, B., & Tambs, K. (2002). Otoacoustic emissions in the general adult population of Nord-Trøndelag, Norway: II. Effects of noise, head injuries, and ear infections. *International Journal of Audiology* 41, 78-87.
- [9] Tognola, G., Grandori, F., Avan, P., Ravazzani, P., & Bonfils, P. (1999). Frequency-specific information from click evoked otoacoustic emissions in noise-induced hearing loss. *Audiology* 38, 243-250.
- [10] Humes, L.E. (1977). Review of four new indices of susceptibility to noise-induced hearing loss. *J. Occupational Medicine* 19 (2), 116-118.
- [11] Humes, L.E., & Bess, F.H. (1978). A test battery approach to the investigation of susceptibility to temporary threshold shift. *Acta Otolaryngology* 86, 385-393.
- [12] Bienvenue, G.R., Violon-Singer, J.R., & Michael P.L. (1977). Loudness discrimination index (LDI): a test for the early detection of noise-susceptible individuals. *Am. Ind. Hygiene Assoc. J.* 38, 333-337.
- [13] Thomas, G.B., & Williams, C.E. (1988). Noise susceptibility: a comparison of two naval aviator populations. In *Proceedings Noise as a Public Health Problem*, Vol. 2, 195-201.
- [14] Lawton, B.W., & Robinson, D.W. (1986). An investigation of tests of susceptibility to noise-induced hearing loss. (ISVR, Southampton, UK) ISVR Contract Report No. 86/8.
- [15] Lawton, B.W., & Robinson, D.W. (1987). Further investigation of tests for susceptibility to noise-induced hearing loss. (ISVR, Southampton, UK) ISVR Technical Report No. 149.
- [16] Humes, L.E., Schwartz, D.M., & Bess, F.H. (1977). The threshold of octave masking (TOM) test as a predictor of susceptibility to noise-induced hearing loss. *J. Auditory Research* 17, 5-12.
- [17] Bergman, M., Najenson, T., Korn, C., Harel, N., Erenthal, P., & Sachartov, E. (1992). Frequency selectivity as a potential measure of noise damage susceptibility. *British Journal of Audiology* 26, 15-22.
- [18] Smoorenburg, G.F. (1992). Speech reception in quiet and in noisy conditions by individuals with noise-induced hearing loss in relation to their tone audiogram. *J. Acoust. Soc. Am.* 91 (1), 421-437.
- [19] Henderson, D., Subramaniam, M., & Boettcher, F.A. (1993). Individual susceptibility to noise-induced hearing loss: an old topic revisited. *Ear & Hearing* 14 (3), 152-168.

Improving the Effectiveness of Communication Headsets with Active Noise Reduction: Influence of Control Structure

Anthony J. Brammer

Envir-O-Health Solutions, Box 27062, Ottawa, ON K1J 9L9, Canada, and Ergonomic Technology Center,
University of Connecticut Health Center, Farmington, CT 06030-2017

Donald R. Peterson, Martin G. Cherniack, and Subhash Gullapalli

Ergonomic Technology Center, University of Connecticut Health Center, Farmington, CT 06030-2017

anthonybrammer@hotmail.com

ABSTRACT

For communication headsets equipped with active noise reduction (ANR), the performance of the control system may influence the communication signal reaching the ear. Conversely, the communication signal may perturb the operation of the ANR system. The interaction between the communication and control signals depends primarily on the control structure, and on the bandwidths and frequency responses of the signal channels. The effects are described for two circumaural communication headsets with similar passive, and active, noise reductions, one with an analog feedback control system and the other an adaptive digital feed-forward control system. Measurements were conducted in a diffuse sound field, with the headsets mounted on a head and torso simulator. The frequency response of sound reproduced by the communication channel was measured when the ANR system was not operating, and when the control system was operating, with swept pure-tones, and broadband noise. The speech intelligibility was estimated for environmental noise shaped to represent the spectrum of speech, the noise within a tank, or the noise within an aircraft cockpit, by the Speech Transmission Index (STI). The STI and fidelity of sound reproduced by the communication channel of the device with a feed-forward control structure tended to exceed that of the more common feedback control structure. This appeared to be a consequence of the compromised frequency response of the earphone and drive electronics employed in the feedback control system to maintain stability of the feedback loop, as well as the presence of communication sounds sensed by the control microphone that were fed back into the controller. The lack of corruption of the communication signal by the feed-forward control system, together with the possibility of using electro-acoustic components with flat frequency responses, suggests that this control structure may be more consistent with the audio fidelity requirements of advanced auditory communication systems.

1.0 INTRODUCTION

A primary expectation of communication headsets is to maintain speech intelligibility and, for advanced auditory communication systems, audio fidelity, under all operational conditions. This is particularly important in circumstances in which degraded auditory communications may have life-threatening consequences, e.g., military operations, and air-traffic control. The requirement may be difficult to maintain in noisy environments and for persons with hearing loss, and also when the communication system is operated at sound levels sufficient to induce temporary threshold shift in the user.

Brammer, A.J.; Peterson, D.R.; Cherniack, M.G.; Gullapalli, S. (2005) Improving the Effectiveness of Communication Headsets with Active Noise Reduction: Influence of Control Structure. In *New Directions for Improving Audio Effectiveness* (pp. 6-1 – 6-8). Meeting Proceedings RTO-MP-HFM-123, Paper 6. Neuilly-sur-Seine, France: RTO. Available from: <http://www.rto.nato.int/abstracts.aps>.

The technology of active noise reduction (ANR) has been applied to communication headsets for more than two decades, and has achieved success in reducing environmental noise at frequencies below 1 kHz. While the contributions of ANR to improved speech intelligibility and pure-tone signal detection have been well documented [1-7], the simultaneous requirements for effective ANR, improved speech intelligibility and advanced auditory communications, such as spatialized auditory displays, have received little attention.

The purpose of this paper is to explore the interaction between the performance of the control system and sound reproduction by the communication channel in ANR headsets, with particular reference to the control structure. The intention is to identify factors that may suggest new directions for improving the audio effectiveness of future headsets, hearing protectors, helmets or earplugs equipped with ANR. The discussion is centered on the performance of two circumaural hearing protectors equipped with active noise control systems and a communication channel, one with a feedback control structure and the other with a feed-forward control structure.

2.0 APPARATUS AND METHODS

The metrics employed were the physical measurement of ANR, and the frequency response and Speech Transmission Index (STI) of the communication channel and its associated electronics. The apparatus and methods are summarized in the following sub-sections.

2.1 Test Room

A reverberation room was used for this study. The rectangular room was designed in accordance with the requirements of ISO 3741 (1975), with dimensions of 6.1 m (length) x 4 m (width) x 5 m (height). The room has an estimated diffuse-field low-frequency cut-off of 110 Hz. Four multi-element loudspeaker systems, positioned near the corners of the room at floor level, provided a broadband source of environmental noise (55 Hz to 8 kHz), and could produce an A-weighted sound level of up to about 100 dB.

2.2 Noise Reduction

The passive, and active, noise reductions of the two headsets were measured when the devices were mounted on a manikin (Bruel & Kjaer Head and Torso Simulator, HATS). The built-in microphone within the right ear simulator of HATS was used to record the sound pressures. The measurements of noise reduction were conducted using band-limited white noise with an A-weighted sound level of ~90 dB.

2.3 Frequency Response of Communication Channel

The frequency responses of the earphones in the two headsets and their associated drive electronics were obtained when the headsets were mounted on HATS with cushions sealed, first when the active control system was not operating and then when it was operating. The electronic drive signals were a swept pure tone, or broadband noise, of various amplitudes, which were fed into the input of the communication channel. The sound output of the earphone was monitored by the microphone within the ear simulator of HATS.

2.4 Speech Transmission Index

The influence of ANR on speech intelligibility was estimated using the STI, which is a figure of merit for a communication link that varies from zero (no intelligibility) to unity (perfect intelligibility) [8]. The STI

signal was fed into the input of the communication channel, and its amplitude was adjusted to produce an A-weighted sound level at the ear of 70 dB.

The environmental noise spectrum at the center-head position (i.e., in the absence of HATS) was shaped to approximate the long-term average of speech, or to correspond to the noise spectrum inside a Leopard tank, or in the cockpit of a Buccaneer aircraft. The noise spectrum of the Leopard tank is dominated by low frequency components between 100 and 500 Hz. In contrast, the noise spectrum in the cockpit of the Buccaneer aircraft is broadband, with sound pressure levels increasing with increasing frequency to above 5 kHz. For each environmental noise spectrum, the A-weighted sound level at the microphone of the ear simulator of HATS was adjusted to produce a prescribed speech signal-to-noise (S/N) ratio. In order to focus the results on the active performance of the headsets, the measurements were performed with the A-weighted sound level of the noise under the earmuff adjusted to be 67.5 dB when the ANR systems were not operating. This procedure adjusts for differences in the passive attenuation of the two headsets [9]. The "speech" S/N ratio was then 2.5 dB, and resulted in the speech-spectrum shaped environmental noise producing an A-weighted sound level of close to 90 dB at the center-head position.

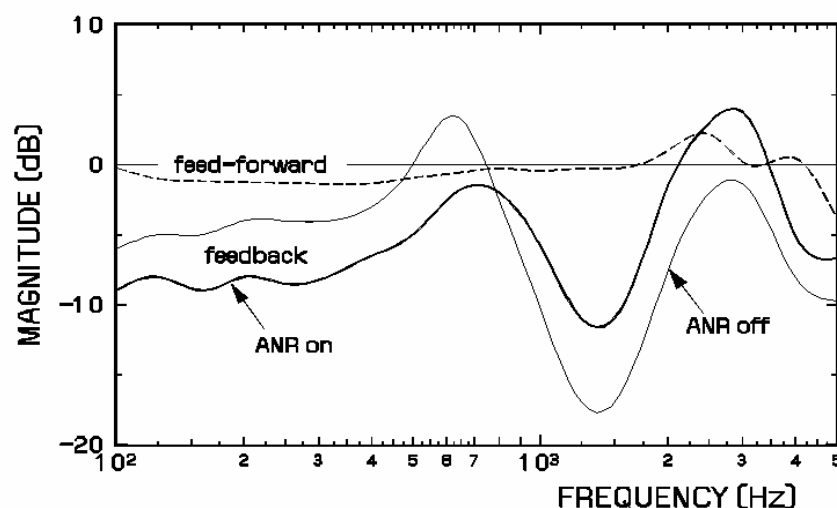


Figure 1: Frequency response of sound reproduced by communication channel of headsets with feedback ANR control system (ANR "off", and "on"), and feed-forward control system.

3.0 RESULTS

The frequency responses of the sound reproduction system in the headset with feedback control are shown in Fig. 1 when the control system was operating (curve "ANR on") and not operating ("ANR off"). As can be seen from the diagram, the frequency response of the earphone and its associated electronics in this headset displayed large frequency-dependent variations in amplitude. It should be noted that the earphone and associated electronics selected for a feedback control system is a compromise between satisfying the need for maintaining stability of the feedback loop and for communication fidelity. In contrast, the frequency response of the sound reproduction system in the headset with feed-forward control possessed little dependence on

frequency from 100 to 5000 Hz, except for a small peak (~2-3 dB) from 2 to 3 kHz (dashed line in Fig. 1). The response did not change when the control system was operating.

Table 1: STI for Headset with Adaptive Feed-forward, or Fixed Filter Feedback Control System

	Speech Spectrum	Leopard Tank	Buccaneer Cockpit
Feed-forward, ANR - ON	0.64	0.77	0.62
Feedback, ANR - ON	0.57	0.77	0.58
Feed-forward, ANR - OFF	0.64	0.77	0.62
Feedback, ANR – OFF	0.43	0.73	0.50

As already noted, the A-weighted sound level of the environmental noise at the artificial ear was set to be the same for both headsets when the control systems were not operating. This was to accommodate differences in the passive noise reductions between the headsets of up to ± 5 dB at some frequencies, and hence focus the STI measurements on the active performance of the devices. The headsets thus operated with the same "speech" S/N ratio under this condition. The ANR of the two headsets were comparable in magnitude and frequency range (~12-15 dB at frequencies below 200 Hz, falling to near zero by 300-400 Hz) [9]. The STIs of the two headsets are shown in Table 1 for the selected environmental noise spectra. It can be seen from the Table that the STI of the headset with the feed-forward control system tended to be greater than that for the headset with the feedback control system, except for the noise source dominated by low frequencies (Leopard tank). The STI recorded for the headset with feed-forward control was not affected by whether the control system was operating, or not operating, suggesting that the ANR was contributing little to the improvement of speech intelligibility in this device. This was not surprising, as the control system had been optimized to reduce low-frequency helicopter noise, including noise at the rotor fundamental blade-passage frequency (16 Hz), and not to improve speech intelligibility. The STI recorded for the headset with feedback control increased substantially when the control system was operating, but never exceeded the STI recorded by the headset with feed-forward control.

4.0 DISCUSSION

Communication signals have been introduced into the control loop of an analog fixed-filter feedback active noise control system in a number of ways [10]. An effective method has been described by Steeneken et al. [11], and is shown in the simplified block diagram of Fig. 2A. In this diagram, the control signal paths are shown by continuous lines, and the communication signal paths by dashed lines. The control filter is designed to operate in such a way as to cancel the sound sensed by the microphone E, which provides its input signal. An essential part of this process is the transformation of the electrical signal from the control filter into sound by the earphone, S, the propagation of sound from S to the microphone, E, and the transformation of sound into an electrical signal by the microphone. These transfer functions together define the error path: in this terminology, microphone E is the error microphone and its output is the error signal. The error path in many

ways governs the performance of the control system. For example, the error path transfer function will define the magnitude and frequency range of ANR: also, the error path is influenced by the presence of an air leak around the cushion sealing the earmuff to the side of the head, and an active noise control system can become unstable if this function changes sufficiently.

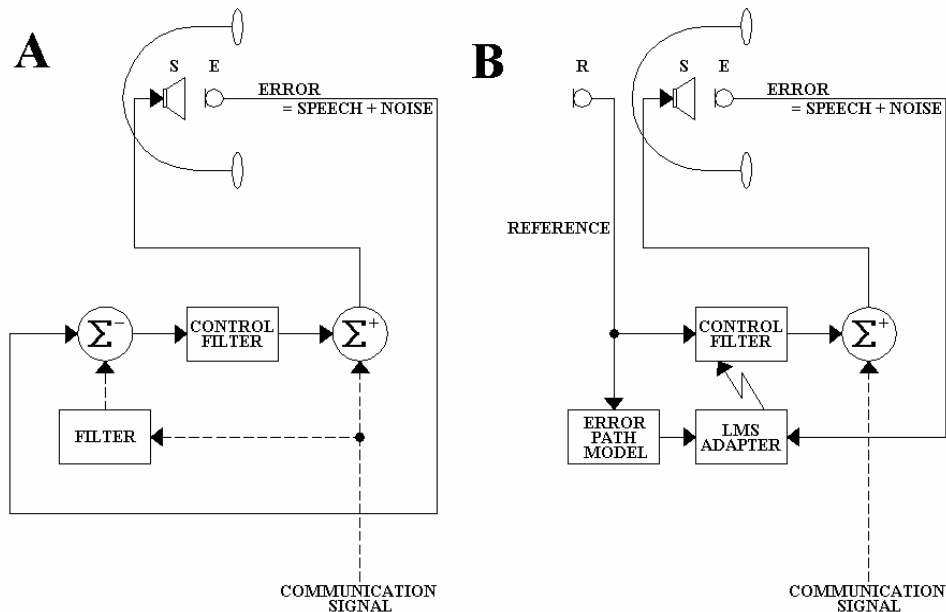


Figure 2: Communication headset with ANR and: A – feedback, or B – feed-forward control structure. For description, see text.

An important consequence of a feedback control structure is that the error signal contains both residual environmental noise and speech from the communication channel, which are fed back into the input of the control filter (Fig. 2A). As the role of the control filter is to attempt to null the error signal, this will have the effect of cancelling the communication signal as well. To mitigate this undesirable effect of the control structure, Steeneken et al. introduced an additional filter to enable the communication signal to be subtracted from the error signal prior to entering the control filter (see dashed lines in Fig. 2A). The success of this addition to the basic control structure may be judged by inspecting the results in Table 1 and Fig. 1. It is evident from Table 1 that the STI for the headset with the feedback control structure, which employed a variant of the method illustrated in Fig. 2A, is less for environmental noise spectra containing substantial components at speech frequencies above 500 Hz when the control system was not operating (e.g., results for speech-spectrum shaped noise, and Buccaneer noise). Under these conditions, the "speech" S/N is low in frequency bands contributing substantially to the STI, and no reduction in environmental noise or corruption of the speech signal by the control system is occurring. In these circumstances, the STI reflects the fidelity of the sound reproduction system in the headset with the feedback control system in comparison to that of the headset with the feed-forward control system. That the frequency response of sound reproduction by the

feedback control system is less than ideal is shown by the dashed curve in Fig. 1. As already noted, the earphone and drive electronics employed in the feedback ANR control system must satisfy the need for maintaining stability of the feedback loop, and so commonly compromise the magnitude response in order to obtain the necessary phase response and hence group delay through the system. Both the magnitude of the frequency response and the STI improve when the control system is operating (solid line, Fig.1, and Table 1, respectively), reflecting the combined effect of ANR, frequency-dependent amplification of the communication signal, and subtraction of a filtered version of the communication signal from the error signal. While the improvement in speech intelligibility anticipated from the increase in STI is encouraging, it is not clear that this control structure will produce adequate audio fidelity for advanced auditory communication systems involving, for example, spatial auditory displays. The provision of two earphones in one earmuff – one for ANR, and the other for sound reproduction – may, however, improve audio fidelity.

The basic control structure for an adaptive feed-forward active control system applied to a communication headset is shown by the continuous lines in Fig. 2B. The control structure employs an additional microphone, R, the reference microphone, that is external to the earmuff and provides the input signal to the control filter. In this control structure, the control filter must model the propagation of the environmental noise from R to E, as well as taking into account the electro-acoustic transfer function of the earphone, the propagation of sound from S to E, and the acousto-electric transfer function of microphone R. This process is implemented by successively adjusting the transfer function of the control filter to optimize the ANR, and is usually performed digitally by an adaptive filter (indicated by the curled arrow in the diagram). The adjustment involves minimizing the squared magnitude of the instantaneous error signal, and was performed by the normalized filtered-X least mean squares (LMS) algorithm in the device employed for this study [12]. Convergence to the "best" performance of the adaptive filter and the stability of the control system are assured by pre-filtering the reference signal prior to calculating the update of filter coefficients by a model of the error path. Details of the hardware and software needed to achieve broadband ANR in the small dimensions of a circumaural headset have been described elsewhere [13].

In this control structure, the communication signal is simply added to the output of the control filter (dashed lines in Fig. 2B). Note that the error signal, while still consisting of the residual noise and speech, does not become the input to the control filter and so cannot perturb the communication signal. As can be seen from Fig. 2B, the presence of the residual speech in the error signal may perturb the convergence of the filtered-X LMS algorithm resulting in less than optimum ANR, as the error signal is used to update the filter coefficients. The influence of changes in the error path, such as occur when the headset is re-fitted to the head or worn by different individuals, may be accommodated by determining person-specific error path transfer functions and in this way rendered less likely to destabilize the adaptive control system [14]. While a possible reduction in the ANR from maladjustment of the adaptive control filter cannot be excluded, inspection of the results in Table 1 indicates that the magnitude of any change in ANR is not sufficient to render the STI of the headset equipped with the feed-forward control system less than that of the feedback system in circumstances in which the ANR may be expected to contribute to the intelligibility, namely in environmental noise dominated by low-frequencies (e.g., Leopard tank). Moreover, the lack of corruption of the communication signal by the control system, together with the absence of the need for the earphone and drive electronics to possess responses tailored to maintain the stability of a feedback control loop, permits the use of electro-acoustic components with flat frequency responses. Thus, a feed-forward control structure appears to be consistent with the audio fidelity needed for advanced auditory communication systems to be built into headsets, hearing protectors, helmets, or earplugs equipped with ANR.

5.0 CONCLUSIONS

For headsets with effectively equalized passive, and similar active, noise reductions, the STI and fidelity of sound reproduced by the communication channel of a headset with a feed-forward ANR control system tended to exceed that of a headset with the more common feedback ANR control system.

The earphone and drive electronics employed by the feedback control system possessed a compromised magnitude response most probably to obtain the necessary phase response and hence group delay through the system, to satisfy the stability requirements of the feedback loop. Both the magnitude of the frequency response and the STI improved when the control system was operating, reflecting the combined effect of ANR, the subtraction of the communication signal from the error signal, and frequency-dependent amplification of the communication signal. The lack of corruption of the communication signal by the feed-forward control system, together with the absence of the need for the earphone and drive electronics to possess responses tailored to maintain the stability of a control loop, permitted the use of electro-acoustic components with flat frequency responses. A feed-forward control structure would thus appear to be more consistent with the audio fidelity needed for advanced auditory communication systems.

6.0 REFERENCES

- [1] C. W. Nixon, R. L. McKinley and J. W. Steuver, "Performance of active noise reduction headsets," in *Noise-Induced Hearing Loss*, edited by A. L. Dancer, D. Henderson, R. J. Salvi and R. P. Hamernik, Mosby, St. Louis (1992) pp. 389-400.
- [2] D. W. Gower and J. G. Casali, "Speech intelligibility and protective effectiveness of selected active noise reduction and conventional communication headsets," *Human Factors* **36**, 350-367 (1994).
- [3] R. L. McKinley, L. J. Morris and C. W. Nixon, "Special applications of active noise reduction headsets," *Audio Effectiveness in Aviation*, AGARD-CP-596, Neuilly-Sur-Seine, France (1997), pp. 21-1 - 21-4.
- [4] A. J. Brammer, R. B. Crabtree, D. R. Peterson and M. G. Cherniack, "Intelligibility in active communication headsets," *ICBEN 2003*, edited by R. de Jong, T. Houtgast, E. A. M. Franssen and W. F. Hofman, pp. 58-64 (2003).
- [5] R. B. Crabtree, S. Abel, A. L. Dancer, K. Buck, T. Wessling, H. J. M. Steeneken, J. Verhave, S. H. James, G. Rood and R. McKinley, "Assessment methods for personal active noise reduction validated in an international Round Robin," NATO Technical Report TR-HFM-094, Neuilly-Sur-Seine, France (2004).
- [6] R. B. Crabtree, "Constraints in the application of personal active noise reduction systems," *Audio Effectiveness in Aviation*, AGARD-CP-596, Neuilly-Sur-Seine, France (1997), pp. 15-1 - 15-6.
- [7] J. G. Casali, G. S. Robinson, B. C. Dabney and D. Gauger. "Effect of electronic ANR and conventional hearing protectors on vehicle backup alarm detection in noise," *Human Factors* **46**, 1-10 (2004).
- [8] IEC 60628-16 (3rd edition), "Sound system equipment, Part 16: Objective rating of speech intelligibility by speech transmission index," International Electrotechnical Commission, Geneva (2003).

- [9] A. J. Brammer, R. B. Crabtree, D. R. Peterson, M. G. Cherniack and S. Gullapalli, "Active headsets: Influence of control structure on communication signals and noise reduction," *ACTIVE 2004*, Noise Control Foundation, New York NY, pp. 1-10 (2004).
- [10] K. Buck and V. Zimpfer-Jost, "Active hearing protector systems and their performance," NATO Lecture Series HFM-111 on Personal Active Noise Reduction (2004).
- [11] H. J. M. Steeneken and J. A. Verhave, "Personal active noise reduction with integrated speech communication devices: Development and assessment," *Audio Effectiveness in Aviation*, AGARD-CP-596, Neuilly-Sur-Seine, France (1997), pp. 18-1 - 18-8.
- [12] S. M. Kuo and D. R. Morgan, *Active Noise Control Systems: Algorithms and DSP Implementations*, Wiley, New York (1996).
- [13] A. J. Brammer, G. J. Pan and R. B. Crabtree, "Adaptive feed-forward active noise reduction headset for low-frequency noise," *ACTIVE 97*, Noise Control Foundation, New York NY, pp. 365-372 (1997).
- [14] A. J. Brammer and G. J. Pan, "Application of adaptive feed-forward active noise control to communication headsets: Use of subject-specific and synthesized error path models," *ACTIVE 99*, Noise Control Foundation, New York NY, pp. 695-704 (1999).

Speech Intelligibility with Acoustic and Contact Microphones

Barbara Acker-Mills, Ph.D., Adrianus Houtsma, Ph.D., and William Ahroon, Ph.D.

U.S. Army Aeromedical Research Laboratory

P.O. Box 620577

Fort Rucker, AL 36362-0577

barbara.acker@us.army.mil

SUMMARY

Speech intelligibility of signals obtained with an acoustic microphone and three types of vibration-driven contact microphones was assessed using the Diagnostic Rhyme Test (DRT). Stimulus words were recorded digitally in a reverberant chamber with no noise and with ambient broadband noise intensity at 106 dB(A). Listeners completed the DRT task in the same settings, thus simulating typical environments of a rotary-wing aircraft. Results show that speech intelligibility is significantly worse for the contact microphones than for the acoustic microphone, particularly in noisy environments, and some consonant types are affected more than others. Therefore, contact microphones are not recommended for use in any situation where fast and accurate speech intelligibility is essential.

INTRODUCTION

The ability to communicate effectively is of paramount importance in the rotary-wing aircraft environment. Effective communication leads to increased aircrew safety and performance and contributes to successful mission completion. As communication is degraded, mission capability is reduced and the safety of the aircrew is compromised. A communication system involves at least one “talker” (sender), one “listener” (receiver), and any equipment used to augment or transmit information. Most research focuses on the listener and on devices that can increase the speech-to-noise ratio and reduce noise-induced hearing loss.

While the standard U.S. Army aviator helmet, the HGU-56/P Aircrew Integrated Helmet System, is designed primarily for impact protection, it also includes a set of headphones mounted in sound-attenuating earcups, a noise-canceling acoustic microphone (the boom microphone), and an optional Communications Earplug (CEP). The boom microphone, positioned close to the speaker’s mouth, consists of two transducer elements that are faced in opposite directions and are wired out of phase. Thus, a diffuse noise field yields a small residual output signal while directed speech sound yields a much larger differential output. Thus, by limiting the impact of ambient noise, the helmet improves the speech-to-noise ratio (for the listener) and also protects the crew from noise-induced hearing loss (for the listener and talker). A large corpus of information exists on the listener component of the communication system, but very little research assesses problems at the talker level.

Although the noise-cancelling boom microphone in the HGU-56/P works well if positioned and used properly, improper microphone use and noise conditions may impair performance. Additionally, an open microphone, in contrast to the usual “keyed” microphone, often is necessary in situations that require use of both hands (e.g., a crew chief operating a hoist may need to use both hands on the hoist control and cable). Indeed, there

Acker-Mills, B.; Houtsma, A.; Ahroon, W. (2005) Speech Intelligibility with Acoustic and Contact Microphones. In *New Directions for Improving Audio Effectiveness* (pp. 7-1 – 7-14). Meeting Proceedings RTO-MP-HFM-123, Paper 7. Neuilly-sur-Seine, France: RTO.
Available from: <http://www.rto.nato.int/abstracts.aps>.

Speech Intelligibility with Acoustic and Contact Microphones

are some aircraft environments in which an open microphone is the normal operating condition (e.g., in the British Army). Furthermore, noise conditions may be encountered (e.g., air movement from an open window or door) for which the normal noise-canceling boom microphone was not designed to minimize. These problems surrounding boom microphones are not new. During World War II, acoustic microphones faced similar noise issues, signals were very noisy, and microphones were impractical for manual operations or for tasks requiring excessive head motion (6). Throat microphones were developed in response to these problems and used during the later years of WWII.

Throat microphones are designed to pick up (transduce) the vibrations of the vocal apparatus at the throat instead of the vibrations of air molecules at the mouth. Since microphones convert sound signals (spoken words) into electrical signals to be transmitted into the communication system, boom microphones must necessarily also transmit any ambient noise around the talker's mouth. Removing this ambient noise from the communication signal should improve the signal-to-noise ratio for the listener. Thus, by virtue of being relatively insensitive to airborne sound, throat microphones, as compared to boom microphones, should produce a signal that contains less noise.

There are a variety of bone-conduction communication systems, but all operate on the same principle. The microphone is placed somewhere on the skull and is sensitive to internal vibrations created by the production of speech waves that travel through the facial and skull bones to the microphone. The current study used two different communication systems that included both a microphone and a loudspeaker. The head-gear system consisted of a microphone that was in contact with the top of the skull, and bone-conduction speakers that were in contact with both sides of the forehead and with both upper jaw bones. The other system contained a bone-conduction microphone in an earpiece, but the speaker was a regular acoustic speaker. The whole unit fit in the ear canal and was worn under a helmet.

During the development of throat and other contact microphones, there has been little systematic evaluation of speech intelligibility in noise using these devices, and the few existing studies contain conflicting results. Snidecor, Rehman, and Washburn (10) explored vowel intelligibility with contact microphones located on different areas of the head and neck. Stimuli were recorded in quiet and presented over headphones to listeners who also were in a quiet environment. One group of listeners rated intelligibility and another group of listeners gave quality judgments of the vowels. Contact microphone locations at the forehead, mastoid, and larynx were highest in intelligibility and quality ratings. While somewhat informative, this study does not use objective speech intelligibility measures, and with all recording and listening completed in a quiet environment, does not assess how contact microphones might function in noise.

Oyer (9) evaluated intelligibility of words recorded simultaneously with an acoustic microphone placed at the mouth and another acoustic microphone placed in the ear canal. The microphone placed in the ear canal was intended to pick up vibrations created by speech signals and transmitted through the skull to the ear canal. Air traffic control words in carrier phrases were recorded in quiet and mixed with 74 dB white noise prior to being presented to subjects over standard headphones. The signal-to-noise ratio was manipulated by attenuating the speech signal (-12, -15, and -18 dB). Results revealed a microphone $\times$ signal-to-noise ratio interaction, where speech intelligibility decreased for both microphones as the signal-to-noise ratio decreased, but the decrement was less for the ear microphone. Although not part of the formal study, it was reported anecdotally that simultaneous presentation of the acoustic and ear microphone stimuli resulted in very good speech intelligibility.

A study by Moser and Dreher (7) is most relevant to the current research project. They used a noise-canceling acoustic microphone, an ear microphone, and a bone conduction microphone placed on the forehead to record

the Phonetically Balanced (PB) word lists (developed by Egan (3) and later standardized by the American National Standards Institute (ANSI) S3.2-1989 (1)). The words were recorded by pilots in two different transport aircraft. Ambient noise in the KC-97 aircraft was measured at 97 dB(C) and was measured at 106 dB(C) in the C-124 aircraft. Listeners were in a quiet environment and stimuli were presented at 77 dB(C) over PDR-8 acoustic headphones. Listeners became familiar with the two different microphone stimuli by listening to a paragraph followed by operational instructions. They then completed the PB task which consisted of writing the word that was presented through headphones. All microphone transmissions were judged as “acceptable” during the familiarization phase, but PB results were better with the acoustic microphone than either the bone or ear microphone in both aircraft environments. Moser and Dreher (7) also conducted an informal evaluation of bone and ear receivers (not microphones) in an aircraft and found ear receivers to be rated as “excellent.” The bone conduction receivers placed on the mastoid, however, were considered “not acceptable” if the ears were not shielded from the aircraft noise.

In view of the sparse experimental research concerning the original contact microphones, it is surprising that very few studies address speech intelligibility using modern contact microphones. In fact, a thorough literature search found only one peer-reviewed paper that mentioned speech intelligibility secondary to a study of temporary threshold shifts (4). A few recent conference presentations have discussed bone-conduction communication systems (5), but these studies have not yet been published, nor did the listening conditions approximate the noisy environment of rotary-wing aircraft.

Even in the absence of speech intelligibility data, contact microphones are marketed to law enforcement agencies (e.g., Los Angeles Police Department SWAT), fire departments, and to a variety of users for applications that require special environmental controls (respirators, hazardous material suits, etc.) or extremely rugged construction (waterproof, dustproof). Several segments in the DoD are strong advocates of these devices, but systematic research evaluating the intelligibility of speech transmitted with the devices should be completed before recommendations can be made for use in military environments.

A direct comparison of the boom and contact microphones is important since each has its own unique strengths and weaknesses as related to transmitting a speech signal. Although contact microphones may increase the overall signal-to-noise ratio, they also may produce a speech signal that is less intelligible than that produced by boom microphones because of the lack of encoding some important speech components. Specifically, consonant sounds are produced with the articulators (the tongue, lips, jaw position, etc.). Because throat microphones pick up information before the level of the articulators, they should not be very effective in transmitting consonant sounds. However, throat microphones should effectively transmit vowel sounds (which are produced by the vocal cords). Boom microphones, on the other hand, are located close to the articulators and thus should efficiently transmit consonant sounds, but with the trade-off of also transmitting ambient noise, resulting in a lower signal-to-noise ratio.

Throat microphones make contact with the soft tissue of the throat and record information before the effects of the articulators (hard/soft palate, tongue, lips, and teeth) have been added. It is possible that bone-conduction microphones will be more effective, because the bones vibrate in response to vibrations of the vocal cords and the articulators. Thus, less ambient noise will be recorded than with an acoustic microphone, but more consonant and vowel information will be available than with the throat microphone.

The current study provides an objective, experimental evaluation of speech intelligibility for stimuli recorded using the HGU-56/P acoustic microphone, and commercially-available throat and bone-conduction microphones. The experimental conditions include realistic noise conditions and thus address the feasibility of use of these microphone options in rotary-wing aircraft.

EXPERIMENT 1: THROAT MICROPHONE

Method

Speech intelligibility was measured with the Diagnostic Rhyme Test (DRT) using procedures specified by ANSI S3.2-1989 (1). The test stimuli consists of six categories of consonants, with each category containing 16 word pairs that differ only in the initial consonant. The six consonant categories are voicing, nasality, sustention, sibilation, graveness, and compactness. These categories are based on acoustical properties of the consonants, not on place and manner of articulation. Table 1 describes the categories (8).

Table 1: Diagnostic Rhyme Test consonant category descriptions.

Category	Description	Example
Voicing (V)	Voiced (vocal cords vibrate) and voiceless consonants (no vibration).	bean (voiced) vs. peen (voiceless)
Nasality (N)	The initial nasal stops (voiced) are paired with their bilabial stop counterparts.	meat (nasal) vs. beat (voiced, bilabial stop)
Sustention (Sust)	No movement compared with movement of the articulators during production. Sustained speech instead of interrupted as with a stop consonant.	vee (sustained) vs. bee (interrupted)
Sibilation (Sib)	Fricatives accompanied by a hissing sound; produces aperiodicity in the high frequencies. Paired with non-hiss fricatives.	zee (hiss) vs. thee (no hiss)
Graveness (G)	Labial consonants (produced at lips) with energy focused in the lower frequencies. Paired with consonants produced further back in the mouth (alveolars, palatals, etc.).	weed (labial) vs. reed (alveolar)
Compactness (C)	Consonants produced with a concentration of energy in a narrow, central area of the spectrum. Paired with more spectrally diffuse consonants.	key (narrow) vs. tea (diffuse)

Stimuli

The 96 DRT stimuli were recorded in a reverberant chamber without background noise and with a background of spectrally-shaped broad-band noise intensity of 106 db(A). The latter simulates a UH-60A Black Hawk helicopter in straight-and-level flight at 120 knots indicated airspeed. A single-transducer LASH II throat microphone (based on the Thales Acoustics RA440 throat microphone) was used in conjunction with the HGU-56/P noise-canceling acoustic boom microphone to record the stimuli. The LASH II throat microphone is a small lightweight device with medium sensitivity (-47 dB re 1V/Pa) specifically designed for use in very high noise environments such as rotary-wing aircraft. The microphone has a frequency response of about 150 to 5000 Hz which is an improvement over throat microphones used in the past.

The male talker wore a throat microphone and an HGU-56/P helmet with the standard noise-canceling boom microphone (frequency sensitivity from about 200 Hz to 6000 Hz). The talker fastened the throat microphone at a comfortable position and pressure, which was measured at about 200 grams of force. (Thales Acoustics

does not provide specific directions for use of the throat microphone, except that it should fit comfortably without undue pressure.) The sound-attenuating earcups of the HGU-56/P and use of the CEP protected the talker from noise during the recording session.

Two separate analog-to-digital channels (16-bit, 40kHz sampling rate) were used to record stimuli simultaneously from the two separate microphones. Two separate sound files, one from the boom microphone and one from the throat microphone, were created for each stimulus. Stimuli were post-processed to ensure equivalent overall root mean square levels within each microphone type.

Several stimuli were selected randomly for spectral analysis. Because each of the target words contained different types of consonants that differ in intensity during natural speech, none of the signal-to-noise ratios were exactly the same. However, the signal-to-noise ratios for stimuli recorded from the throat microphone were always higher (approximately 10 dB) than for stimuli from the boom microphone.

Participants

Participants were Soldiers at Fort Rucker, Alabama, awaiting the Army Warrant Officer Course and flight training school. Eight males (mean age = 25) volunteered for the study. All but one volunteer had normal hearing as confirmed by recent physical exams or by audiograms performed at the U.S. Army Aeromedical Research Laboratory (USAARL) acoustics laboratory. One male (age 43) reported having tinnitus in his right ear. The study protocol was approved in advance by the USAARL Human Use Committee, and each participant provided written informed consent before participating.

Procedure

All testing took place in the USAARL-Acoustics Laboratory reverberant chamber, and stimulus presentation/response collection was coordinated using the Avaaz Experiment Generator and Controller software (2). The purpose of the study and experimental conditions were explained to the participant and then the participant was fitted with an HGU-56/P. The headphones in this helmet have a frequency range of about 200 Hz to 5000 Hz.

The DRT word-pairs were visually displayed on a computer monitor, followed by the target word presented in the earphones. Participants had 3 seconds in which to use a mouse to select the target word that was spoken. There was a 500 ms delay between the word-pairs appearing on the screen and the auditory stimulus presentation. The correct choice was displayed 500 ms after a response and lasted for 750 ms. If a participant did not respond in the allocated 3000 ms, an incorrect response was recorded. Trials were separated by 750 ms.

A $2 \times 2 \times 6$ repeated-measures factorial design was used, with microphone type (boom, throat), noise (none, 106 dB(A)), and consonant category (see Table 1) as the independent variables. Each block contained 96 word-pair visual displays, with microphone and noise type held constant. Word-pairs, the order of words within a pair, and the word presented over headphones were randomized within and across blocks to avoid learning effects that would occur if there was any consistency across conditions. In addition, this procedure resulted in the consonant categories being completely randomized within blocks of trials so participants would not focus on any specific types of consonants (e.g., voicing, nasality, etc). Presentation order of the four blocks (each microphone and noise combination) was counterbalanced across the eight participants, and stimulus presentation levels were at least 10 dB above masked threshold.

Results and Discussion

Hit and false alarm rates were used to compute a sensitivity index (d') for each subject in each condition. d' is monotonically related to proportion correct, but is not influenced by response bias. Table 2 lists percent correct values and approximate equivalent d' values. Because d' could not be calculated if there were either no hits (performance at floor) or no misses (performance at ceiling), if a participant had all hits and no false alarms in a particular condition (floor performance never occurred), the number of hits was reduced by one-half and the number of false alarms was increased by one-half before converting to proportion correct scores. This procedure resulted in a maximum d' score of 3.74 and allowed d' to be calculated for all conditions.

Table 2: Proportion correct values and equivalent d' values

Proportion correct	d' value
.97	3.74
.90	2.56
.80	1.68
.70	1.05
.60	.51
.50	0.00

Results were analyzed using a $2 \times 2 \times 6$ repeated-measures ANOVA with microphone, noise, and consonant category serving as independent variables. Main effects were observed for all three factors (microphone type, noise type, and consonant type). DRT performance was significantly better with the boom microphone stimuli ($M = 2.41$) than with the throat microphone stimuli ($M = 1.43$), $F(1, 9) = 225.99$, $p < .05$. As expected, performance was best in the no-noise condition ($M = 2.73$) compared to the 106 dB noise condition ($M = 1.11$), $F(2, 18) = 127.65$, $p < .05$. Inspection of the significant main effect of consonant type, $F(5, 45) = 15.17$, $p < .05$, showed that the voicing, sibilant, and compactness categories were perceived better than the other three consonant types.

Beyond the simple main effects, the significant three-way interaction ($F(5,45) = 9.40$, $p < .05$) demonstrated that speech intelligibility was influenced by microphone type, noise, and consonant category. These results are summarized in Figure 1, and the consonant category abbreviations can be found in Table 1.

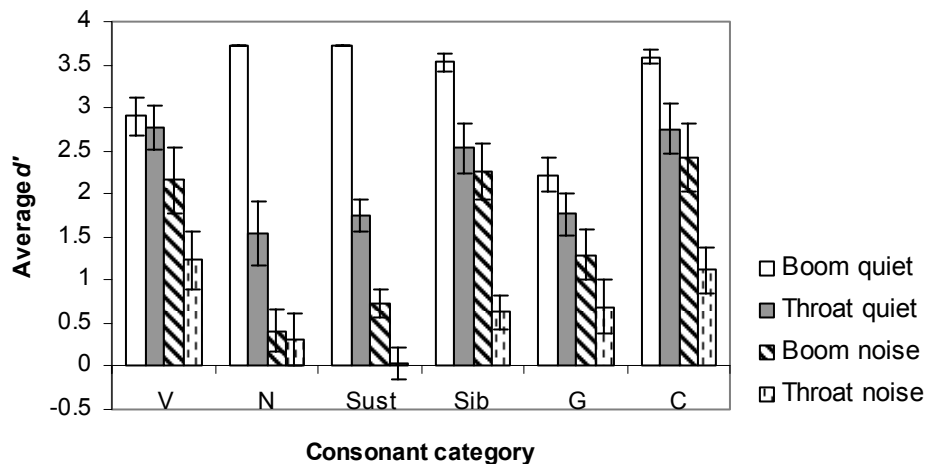


Figure 1: DRT performance as a function of microphone type, noise, and consonant category

The main effects are evident in Figure 1; the boom microphone almost always performs better than the throat microphone, and performance always declines significantly in noise (regardless of microphone type). Exceptions based on consonant category do occur. For example, performance between the boom and throat microphones was similar for the voicing category in quiet. This result is expected, as the throat microphone is in an ideal location to encode vibration, or lack thereof, of the vocal cords. Thus, information after the speech signal passes through the articulators is not essential for perception of voicing. Performance also was similar in the graveness category (in quiet and noise), which is dependent on frequency (see Table 1), but the distinctions involve lower frequencies that are transmitted by both microphones. In contrast, the absence of the effects of the articulators is rather evident for consonants with broadband frequency characteristics, such as /z/ of the sibilant category. The Appendix contains spectrographs of the words THEE and ZEE recorded in 106 dB(A) noise for the boom and throat microphones. Note the lack of distinction between the spectrograms of the throat microphone. This same principle holds for the compactness category, where distinctiveness of the consonants is dependent on perception of a broad or narrow frequency spectrum produced by the articulators, and the throat microphone fails to encode this information, particularly in noise. After inspection of the spectrograms, it is unclear why performance in noise dropped so low for both microphones in noise for the nasality and sustention categories.

The current results clearly demonstrate that while the throat microphone enhances the signal-to-noise ratio, the insensitivity to the effects of the articulators on the speech signal degrades speech intelligibility compared to the acoustic boom microphone. The DRT results indicated that some consonants are affected more than others by noise and/or use of a throat microphone. Spectrograms (see examples in the Appendix) revealed which acoustic features of the consonants were affected by microphone type and noise.

Because significant speech intelligibility differences occurred between acoustic and throat microphones, two other contact microphone systems were examined. It is possible that these bone-conduction systems that pick up vibrations from the entire vocal tract will perform better than the throat microphone which encodes information before the level of the articulators. In addition, signals transmitted through bone (a hard surface) may contain more information than those transmitted from the throat (soft tissue).

EXPERIMENT 2: BONE-CONDUCTION SYSTEMS

This Experiment used the same procedures as Experiment 1, except that the stimuli were recorded from head and ear bone-conduction communication systems. The head equipment consisted of a bone-conduction microphone on top of the skull (frequency response of approximately 200 Hz to 5000 Hz) and four bone-conduction speakers (frequency response of approximately 300 Hz to 3000 Hz) that contact the upper jaw bone and sides of the forehead on each side of the head (Temco Communications, Inc. model HG17). The ear equipment contained a bone-conduction microphone and an acoustic speaker contained in an earpiece that is placed in the ear canal and worn under a helmet (Temco Communications, Inc. model EM-P2). This microphone has a frequency response of approximately 200 Hz to 5000 Hz and the receiver 100 Hz to 3500 Hz. This particular model is designed to be used in noisy environments above 95 dB. Unfortunately, the person who recorded the boom and throat microphone stimuli was no longer available, so a different male speaker recorded the head and ear microphone stimuli. Whereas this situation is not ideal, it is highly unlikely that differences in the boom/throat and head/ear microphones are due to the recordings and not the microphones themselves. To support this notion, DRT data from an earlier study with a different speaker using the boom microphone were compared to the data of Experiment 1. Speech intelligibility was similar as for the speaker used in Experiment 1 ($d' = 2.48$ in the earlier study vs. $d' = 2.41$ in Experiment 1).

Thresholds for the two different speaker systems were measured using stimuli from the DRT task, and testing presentation levels were 10 dB above masked threshold. The head equipment prevented the helmet earcups from forming a tight seal, and therefore, participants had to wear earplugs in the noise condition in order to be protected from the 106 dB noise. The ear microphone/speaker system was worn in the right ear and one earplug was placed in the left ear during the noise condition. Eight participants, different from those in Experiment 1, completed the DRT task.

Results and Discussion

Results were compiled in the same manner as in Experiment 1, where hits and false alarms were used to calculate d' . Results were analyzed using a $2 \times 2 \times 6$ repeated-measures ANOVA with microphone, noise, and consonant category serving as independent variables. Significant main effects occurred for all three independent variables. The head microphone and speaker combination ($\underline{M} = 1.648$) performed better than the earpiece system ($\underline{M} = .976$), $F(1, 7) = 13.13$, $p < .05$, and performance was better in quiet ($\underline{M} = 1.74$) than in noise ($\underline{M} = .879$), $F(1, 7) = 27.59$, $p < .05$. Evaluation of the main effect of consonant category ($F(5, 35) = 3.54$, $p < .05$) revealed that intelligibility for the sibilant and sustention categories was less than for the other four categories.

Speech intelligibility is explained more completely by the microphone $\times$ consonant category interaction ($F(5, 35) = 9.24$, $p < .05$). As can be seen in Figure 2, performance was similar for the two systems in the nasality and sustention categories, but otherwise, intelligibility was better with the head microphone/speaker system. The spectrograms do not readily reveal why performance is similar in these two categories.

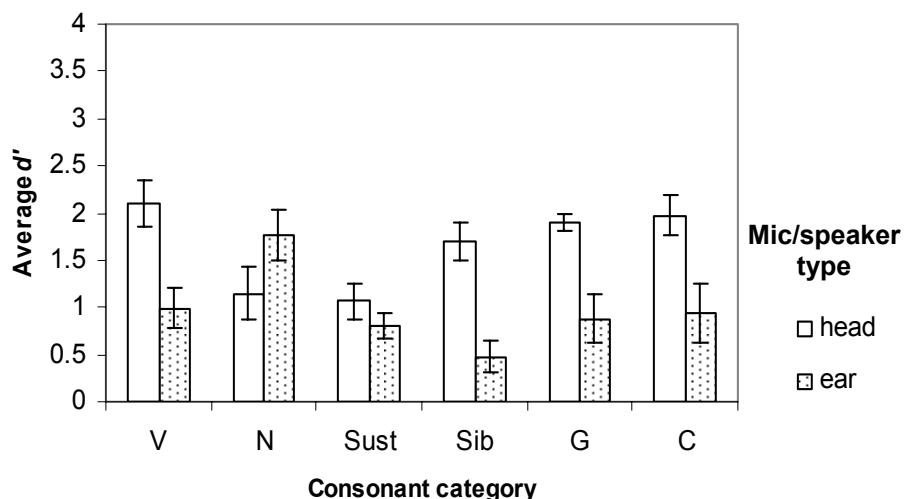


Figure 2: DRT performance as a function of microphone type and consonant category.

One advantage of d' as a sensitivity measure is that results between experiments legitimately can be compared. Figure 3 represents data combined over consonant category from Experiments 1 and 2, and it is evident that the boom microphone and HGU-56/P helmet earphones produce the best speech intelligibility, even in noise. The difference between the quiet and noise condition for the boom microphone may be artificially large because few errors were made, and even a few errors affect d' when performance is so good. The more important microphone comparison is in noise; the only system that had somewhat acceptable performance was the boom microphone at approximately 78 percent correct (performance probably would be even better with the Communication Earplug). Performance is at an acceptable level in quiet for the throat and head microphones, but is not acceptable in noise. Finally, the ear microphone/speaker results in the lowest performance even in quiet conditions (approximately 76 percent correct compared to 95 percent correct for the boom microphone in quiet). Whereas these effects are representative of microphone performance in general, keep in mind that both Experiments also exhibited interactions, where performance was affected by consonant category (see Figures 1 and 2).

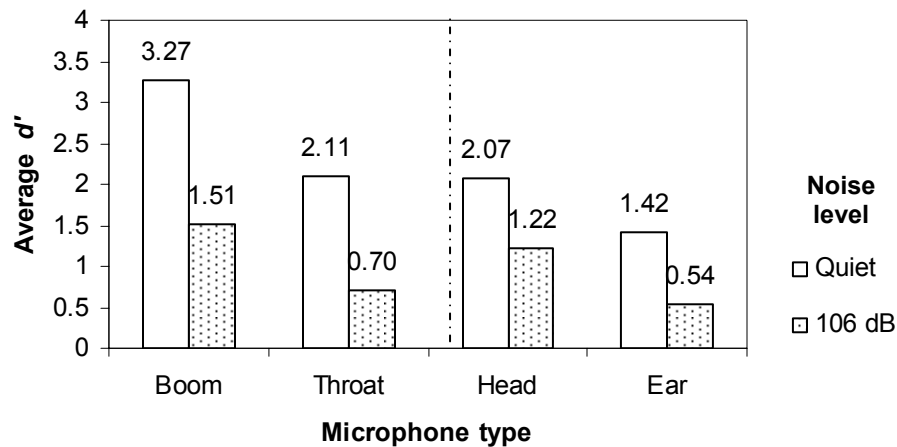


Figure 3: DRT performance as a function of microphone type and noise. Experiment 1 data are in the left panel and Experiment 2 data are in the right panel.

GENERAL DISCUSSION

The problem of noise and its detrimental effects on communication and hearing loss typically focuses on the “listener” (receiver). Whereas devices such as helmet earmuffs and the Communication Earplug can be useful (especially for hearing protection), speech intelligibility is still dependent on the quality of the original signal produced by the “talker” (sender). As noted above, the effectiveness of the noise-canceling boom microphone is reduced under various flying conditions that create unpredictable and highly variable noise. If this unpredictable ambient noise could be eliminated in the transmitted speech signal, the signal-to-noise ratio would be enhanced, and speech intelligibility also might be improved. Contact microphones greatly reduce or eliminate ambient noise because the microphone has a higher impedance that is matched only by vibrations on a surface (e.g., the skull) and not by the vibration of molecules in the air. Thus, the signal-to-noise ratio in a noisy environment is better than that of an acoustic microphone.

Even though the contact microphones have a better speech-to-noise ratio than the boom microphone, speech intelligibility was adversely affected by use of these microphones, particularly in noise. The most probable cause is that none of the contact microphones effectively encode information from the articulators, and this information is essential for differentiating consonants (e.g., the broadband noise in /z/ produced by the tongue). The results are troubling in that the DRT task represents a closed set of words and should be conditions where intelligibility is best (11). Whereas normal flight procedures also have standard communication phrases (basically a closed set), nonstandard speech will most likely occur in emergency or high-intensity combat situations. These are precisely the situations in which good speech intelligibility is critical, and the current results show that intelligibility using contact microphones is poorer than with the use of boom microphones even under the most benign of situations. Thus, it is recommended that contact microphones and speakers not be used in noisy environments where fast and accurate speech perception is critical.

ACKNOWLEDGEMENTS AND DISCLAIMER

We thank Mr. Christopher Madsen for his assistance supporting the software used to conduct this research, CPT Martin Robinette and SGT Kent Loo for their assistance in conducting this study, Ms. Janet Mauldin for her usual expert secretarial support, and our editor, Ms. Linda Burt. This research was funded in part by the U.S. Army Medical Research and Materiel Command In-house Laboratory Independent Review initiative.

The opinions, interpretations, conclusion, and recommendations are those of the authors and are not necessarily endorsed by the U.S. Army and/or the Department of Defense.

REFERENCES

- [1] American National Standards Institute. Method for measuring the intelligibility of speech over communication systems. 1989; Melville, NY: Standards Secretariat, Acoustical Society of America. ANSI S3.2-1989 (R1999).
- [2] Avaaz Innovations, Inc. ECoS Version 2™ Experiment generator and controller. North London, Ontario. 2002.
- [3] Egan, J.P. Articulation testing methods. *Laryngoscope* 1948; 58(9): 955 – 991.
- [4] Horie, S. Improvement of occupational noise-induced temporary threshold shift by active noise control earmuff and bone conduction microphone. *J Occup Health* 2002; 44(6): 414 – 420.
- [5] Letowski, T. Bone conduction communication: Applications and limitations 2003. Paper presented at the National Hearing Conservation Association 28th Annual Conference, February 20-22, at Dallas, TX.
- [6] Martin, D.W. Magnetic throat microphones of high sensitivity. *J Acoust Soc Am* 1947; 19(1): 40 – 43.
- [7] Moser, H.M. and Dreher, J.J. Operational tests of miniature microphones and receivers. Columbus, OH: The Ohio State University Research Foundation 1956; Technical Report No. 36 for Contract No. AF 19(604)-1577.
- [8] Nicolosi, L., Harryman, E., and Kresheck, J. Terminology of Communication Disorders in Speech-Language-Hearing (3rd Edition) 1989; Baltimore: Williams and Wilkins.
- [9] Oyer, H.J. Relative intelligibility of speech recorded simultaneously at the ear and mouth. *J Acoust Soc Am* 1955; 27(6): 1207 – 1212.
- [10] Snidecor, J.C., Rehman, I., and Washburn, D.D. Speech pickup by contact microphone at head and neck positions. *J Speech Hear Res* 1959; 2(3): 277 – 281.
- [11] Suter, A.H. The effects of noise on speech and warning signals. Aberdeen Proving Ground, MD: U.S. Army Human Engineering Laboratory 1989; AD-A212 520.

APPENDIX

Spectrograms for stimuli recorded in 106 dB(A) noise with a throat and an acoustic microphone.

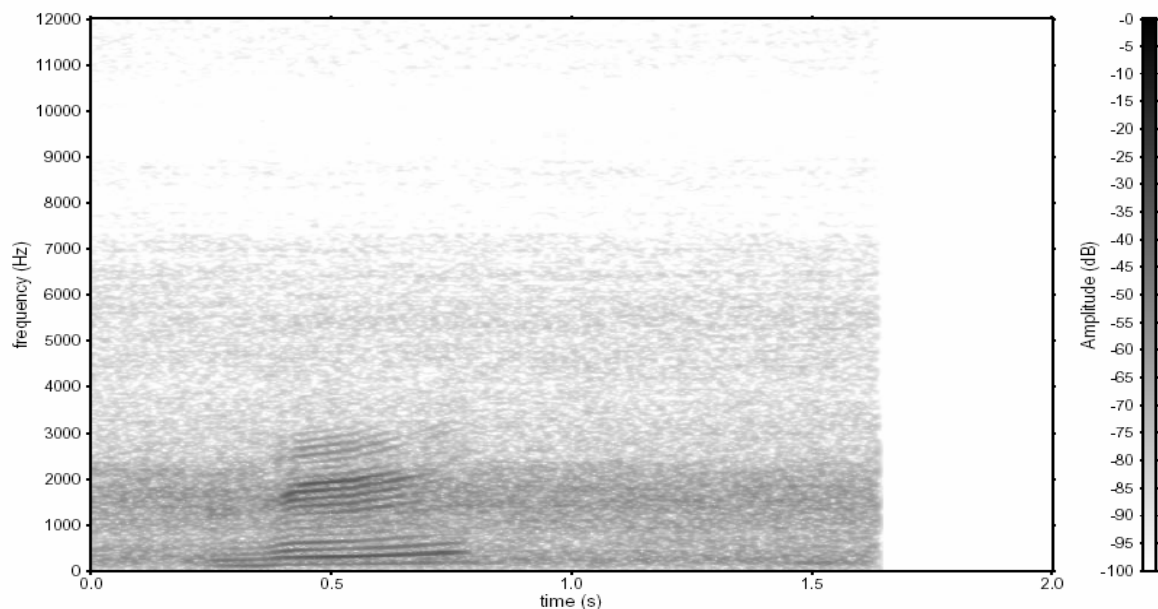


Figure A1. Spectrogram for THEE recorded with a throat microphone.

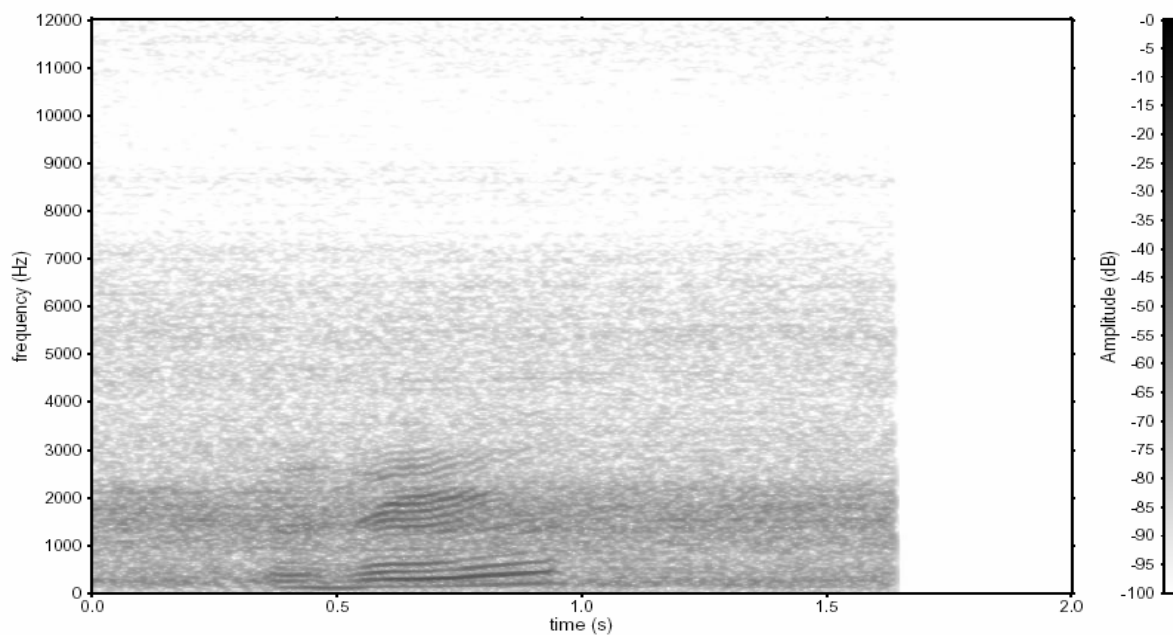


Figure A2. Spectrogram for ZEE recorded with a throat microphone.

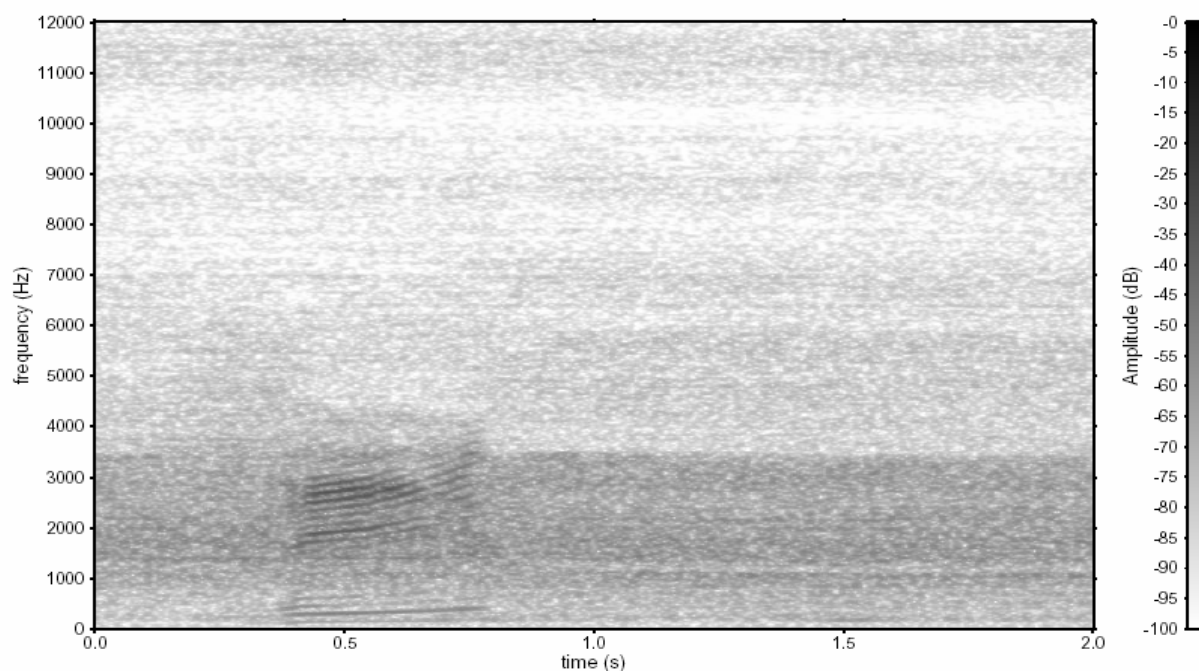


Figure A3. Spectrogram for THEE recorded with a boom microphone.

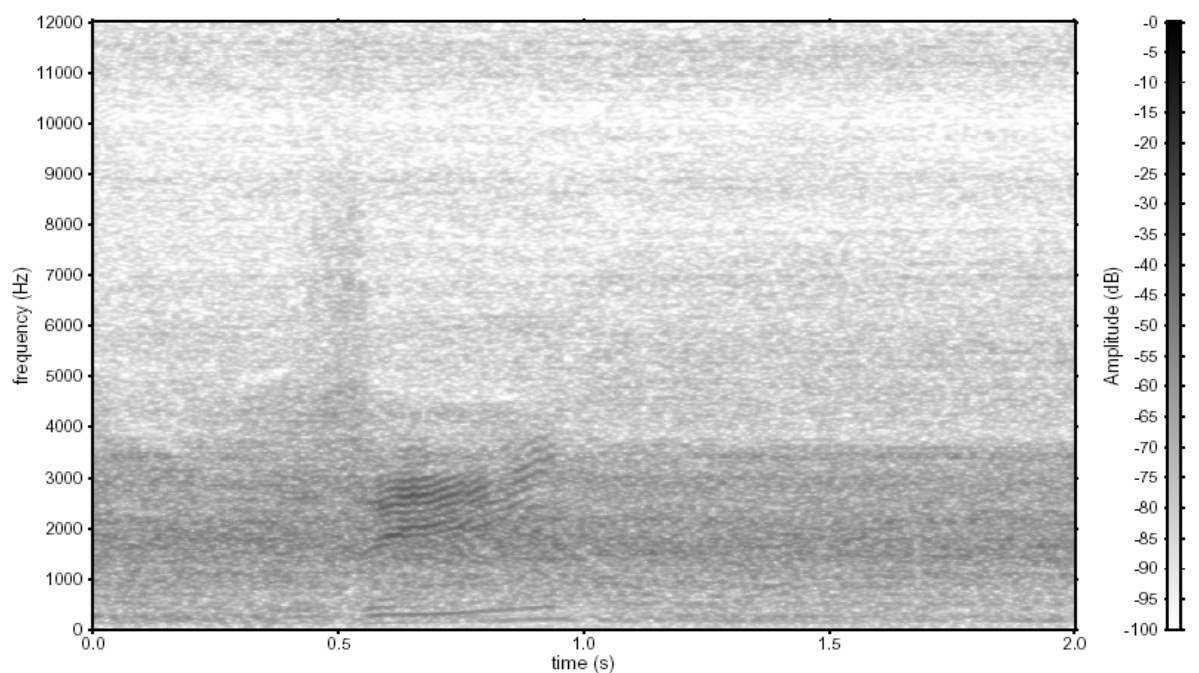


Figure A4. Spectrogram for ZEE recorded with a boom microphone.

Speech Intelligibility with a Bone Vibrator

Pellieux L., Bouy J.C., Blancard C., Guillaume A.
IMASSA

Département de sciences cognitives
BP 73, 91223 Brétigny sur Orge, France

lpellieux@imassa.fr

ABSTRACT

The FELIN project (Foot soldier with Integrated Equipment and Connectivity) of the French DGA (Délégation Générale à l'Armement – French Military Procurement Agency) aims at gearing tomorrow's foot soldier with a "flexible and manoeuvring" equipment. Among requirements, the foot soldier will have to be fitted with a communications headband operating through bone conduction. The main advantage of listening through bone conduction is that it allows for the transmission of information to the soldier, while leaving the soldier's ears free to perceive the surrounding environment and use sound cues for orientation in space. Furthermore, this device is light, not bulky and quite comfortable.

The laboratory assessment performed aimed at comparing voice intelligibility scores of a prototype bone conduction device with the scores obtained by Peltor's® militarized COMTAC-type headset. This latter device is a hearing device, i.e. fitted with two microphones, left and right, to reproduce a spatial hearing capability. It is thus possible to compare these two technologies, which both allow for orientation in space using acoustic cues.

Tests were performed in silence and in operationally realistic noise conditions (reproducing the noisy environment inside an armoured vehicle). Voice material is made of nonsense CVC words (Consonant – Vowel – Consonant), spoken by two speakers of different genders.

In silence, the prototype bone vibrator headband and the COMTAC headset obtain the same intelligibility score. In noise, the performance of the prototype headband is slightly lower than that of the COMTAC headset. A number of reasons can explain this: 1) in a noisy environment, finding the right position for the vibrator to get optimal hearing conditions is hard to find. 2) the voice transmission level is too low. It seems that the transmission level is limited because of parasite vibrations which start occurring at high levels.

As a conclusion, bone conduction technology (limited here to "listening") provides a significant positive potential, based on the performance obtained in silence. However, some design improvements must be made to reach levels allowing for intelligibility in noisy environments (notably for use in armoured vehicles).

INTRODUCTION

Project FELIN (Fantassin à Equipement et Liaisons Intégrées – Foot soldier with Integrated Equipment and Connectivity) of the French DGA (Délégation Générale de l'Armement – French Military Procurement Office) is designed to provide tomorrow's foot soldier with a "flexible and manoeuvring" equipment. Among requirements, the foot soldier should be fitted with a communications headband, operating thanks to bone conduction.

Pellieux, L.; Bouy, J.C.; Blancard, C.; Guillaume, A. (2005) Speech Intelligibility with a Bone Vibrator. In *New Directions for Improving Audio Effectiveness* (pp. 8-1 – 8-20). Meeting Proceedings RTO-MP-HFM-123, Paper 8. Neuilly-sur-Seine, France: RTO. Available from: <http://www.rto.nato.int/abstracts.aps>.

Speech Intelligibility with a Bone Vibrator

The bone conduction headset (see figure 1) offers three advantages, coming from the fact that ears are left free.

- The soldier's hearing is not jeopardized: weak acoustic signals, like a branch squeaking under an invisible enemy's foot, can be detected; sound sources can be located, and orientation in space remains possible.
- The device is not bulky, and does not prevent from wearing various masks.
- Psychologically, it makes contact with local populations easier.

This study only addresses the "listening" function of the communications headband. Its goal is to validate the technological choice of the communications headband, using intelligibility tests performed in a laboratory. This study is necessary because selecting a technology based on transmission by bone conduction is new for operational applications involving a great number of soldiers.

Testing principles are presented, along with a justification of choices made. After an overview of intelligibility tests, results and answers to questionnaires are then provided and discussed.

1. TESTING PRINCIPLES

1.1 Background

This study aims at comparing the performance of the bone conduction headband with that of PELTOR ®'s COMTAC headset. Both devices make it possible to locate sound sources and hear low intensity sounds.

1.2 Communications Headband (listening function)

Photo 1. Hearing headband in place.



The communications headband uses bone conduction [1] to transmit sounds. A vibrator, made of a electromagnet, controls and moves a small mass placed in contact with the skin. Mechanical vibrations are transmitted through the skin, towards skull bones. Parts of the vibrations are channeled through bone and reach the two cochlea. Distinct signals cannot be applied to each ear, given the high bone conduction speed and the great number of wave propagation channels in skull bones. Another part of the vibrations makes the air in the ear canal resonate, notably if the ear canal is plugged off. Note that vibrators also transmit an acoustic signal, which a priori is not desired, and which is later perceived through air conduction.

The communications headband is made of a metallic arch which makes sure the two electro mechanic vibrators are correctly applied to the zygomatic apophyses of the left and right temporal bones. The arch

Speech Intelligibility with a Bone Vibrator

can be adjusted in length, thanks to a telescopic adjustor. A length-adjustable elastic headband is set on top of the head, to maintain the device in place. The vibrator pads are connected to a small molded casing, placed on the arch, in charge of signal processing.

1.3 Hearing Headset

Photo 2. PELTOR's COMTAC Headset



Photo 3. Close-up of PELTOR's COMTAC headset, showing the right microphone coated with anti-wind foam, and the two push buttons used to adjust the sound level for surrounding noise.



The COMTAC hearing headset (photos 2 and 3) is made by PELTOR, and has two miniature microphones, one per earcup. As sound is reproduced and sent to each ear, spatial hearing cues are preserved, allowing to locate the source of ambient sounds. Furthermore, this headset amplifies low intensity surrounding sounds, thus theoretically making the detection of slight noises possible. However, if the sound level is too high, the “hearing” headset becomes “deaf”, this headset having a level limiter. When ambient noise increases, the gain of the transmission chain is reduced, to make sure the noise level never exceeds 85dBA. Beyond that limit, there is no longer any gain, and hearing protection becomes identical to that of a traditional passive headset.

Two limitations must be mentioned:

Like in any electronic system, this headset generates a background noise, and the signal to noise (S/N) ratio is often unfavorable for low intensity noises. This headset's performance is below that of natural hearing.

Speech Intelligibility with a Bone Vibrator

This headset provides poor protection in low frequencies, because of the reduced size of its flat earcups.

This headset is a militarized model, and is standard equipment for French GIPN and RAID forces, for example.

2. BASICS OF INTELLIGIBILITY TESTS [2]

Two kinds of methods can be used to assess intelligibility: objective and subjective methods.

2.1 Objective methods

The best known objective method to measure speech intelligibility is the Speech Transmission Index (STI) [3]. It works by analyzing a test signal obtained at the receiving end of the device under assessment. The STI index is then calculated, according to the deformation and degradation of the test signal in various frequency bands. The STI was validated thanks to a procedure using iteratively physical and subjective measures. A simplified version of the STI, called RASTI (Room Acoustical Speech Transmission Index) has been standardized (IEC268).

This objective assessment method has two specific strong points:

- Measurement “standardization”, making it easy to compare two systems. The STI index can be considered as part of a system’s characteristics, along with bandwidth, distortion rate, etc.
- Speed. A single measurement is sufficient to qualify the system’s intelligibility, which saves a lot of time compared to subjective assessments.

However, this method cannot be used for the bone conduction headband, because the output signal of the device under test must be available. The objective method can be used to assess a headphone, by placing a microphone in the hearer’s hearing canal. But there is no output signal available with the hearing headband. It would have to be captured in the cochlea, which is impossible. Some physical measurements of the transducer’s properties can be obtained, by installing it on a mechanical mock-up simulating a human head (artificial mastoid), fitted with an accelerometer measuring vibration amplitude. But these measurements do not take into account transmission from the cochlea to the hearer, and only provide data the transducer’s manufacturer can use to characterize its device.

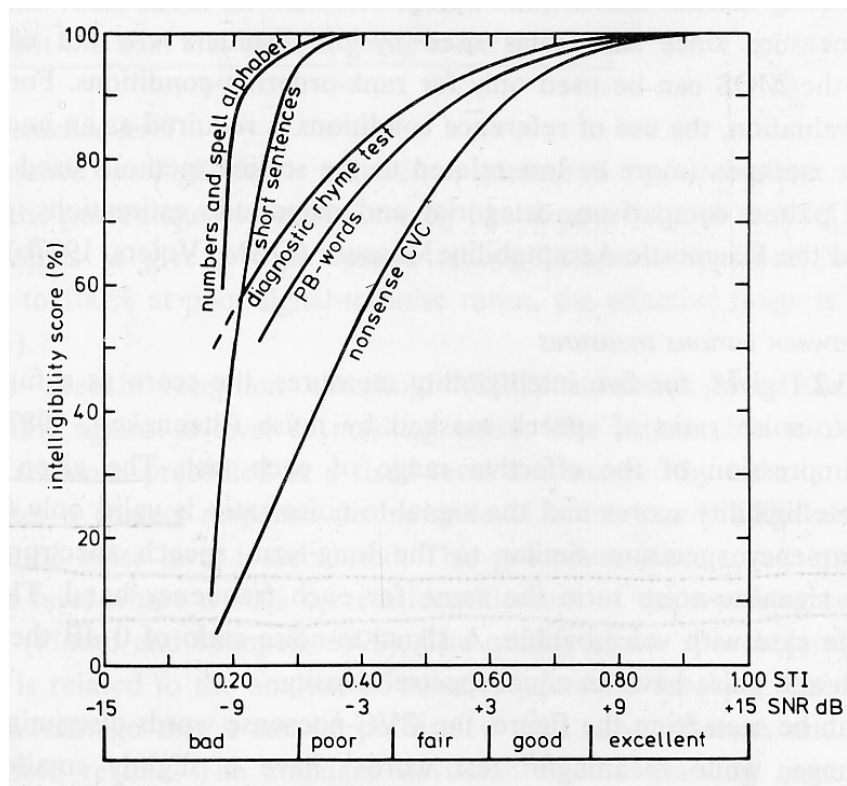
Assessing the intelligibility of an osteophone (bone vibrator) is thus mainly based on subjective tests.

2.2 Subjective methods

During subjective tests, a subject must identify speech transmitted through the device requiring testing. The word “speech” is used deliberately, to be as wide as possible. Intelligibility tests differ according to the nature of the speech transmitted: it can be alphabet letters, figures, words, sentences, various associations of consonants and vowels, or of elementary components of language such as phonemes.

Figure 1 [3] gives intelligibility scores for five kinds of tests, according to the signal to noise ratio of noise-polluted speech. This makes it possible to appreciate the effective dynamic range of each test. On Figure 1, nonsense CVC words (Consonant, Vowel, Consonant) allow for differentiation in a wide dynamic range, whereas PB words (Phonetically Balanced: statistical distribution of phonemes representative of language) offer a narrower dynamic range. Tests with figures and letters show saturation at a -5dB signal to noise ratio, because there are few test words, and identifying them relies mainly on vowel recognition.

Figure 1. Comparison of subjective intelligibility tests (for sounds having a frequency spectrum similar to that of speech - Steeneken, 1992)



Because of its broader dynamic range, the nonsense CVC test is the best suited test to compare the bone vibrator headband with PELTOR's COMTAC.

3. CVC TEST AT IMASSA

The subjective test used provides for open responses, and for a significant discrimination on a wide range of signal to noise ratios. The listener must recognize a Consonant Vowel Consonant word inserted in a carrier sentence. CVC words are made of initial consonants, vowels and final consonants appearing in spoken language. 19 initial and final consonants were chosen, as well as 11 vowels. CVC words are taken at random, and placed in 57 different carrier sentences. These sentences and the CVCs are recorded and played back using a computerized device. Voice material (3,000 CVC words and 57 carrier sentences per speaker) were recorded by a male speaker and a female speaker during a past experiment ("Study of the architecture and design of a new generation audio system, Contract STTE/AVI n°94/86002).

In practice, the subject, in front of a desktop, listens to recorded speech and used the mouse to click on any recognized CVC phoneme. The visual interface can be seen on Figure 2. This program was developed with Matlab®. The sound is transmitted by a TDT RP2.1 processor, connected to the PC through a USB interface. Excel® is used to process resulting data.

Figure 2 - CVC test visual interface

TEST

C

V

C

b
d
f
g
j
k
l
m
n
p
r
s
t
v
z
ch
gn
gz
ks
?

a
e
é
i
o
u
an
on
ou
un
?
réponse:
1 / 57
valider

b
d
f
g
j
k
l
m
n
p
r
s
t
v
z
ch
gn
gz
ks
?

4. TEST CONDITIONS

4.1 Sound environments

Photo 4. Reverberation chamber



- Measurements in silence are carried out either in the anechoic chamber (below 16 dBA), or in the reverberation chamber (below 27 dBA, with power amplifiers in maximum gain position).
- Measurements in noisy environments are carried out in the reverberation chamber, simulating the ambient sound of an armoured troop carrier vehicle, driving off-road (97dBA).

4.2 Telephone filter

The manufacturer supplies headbands equipped with receiver/transmitter devices. However, our goal is to evaluate the transducers and not the quality of radio-electric transmissions. Consequently, cabling was modified by the supplier.

However, it is normal to test the communication devices in the telephone bandwidth. The two devices were thus tested by inserting a digital telephone filter, integrated into the processor transmitting speech signals. This handicaps the COMTAC headset, which has a wide bandwidth.

4.3 Subjects

Nine subjects performed the test. Tests with a tenth subject had to be stopped, because this person complained of headaches and loss of balance right after the first test with the bone vibrator headband.

Speech Intelligibility with a Bone Vibrator

Audiogram

Frequencies tested are as follows:

125 – 250 – 500 – 1,000 – 2,000 – 3,000 – 4,000 – 6,000 and 8,000 Hz.

For six subjects, losses are below 15dB, whatever frequency or ear.

For two others, losses reach 20dB for one ear and one frequency.

For the last subject, the right ear is normal (<15dB). For the left ear, losses are 25dB at 3 kHz, 20dB at 4 kHz and 20dB at 6 kHz.

4.4 Selecting the “listening” level

4.4.1 General

Choosing the listening level is critical when listening in a noisy environment. The louder the speech, the better the intelligibility. It is thus important to assess the sound level delivered by each transducer and compare them. To this end, an equivalent sound level must be assessed for each technology. Furthermore, the speech signal is an extremely variable signal, with a transmission level hard to measure. The calibration signal is taken to be the 60 first seconds of the sequence of all CVCs starting with the “p” consonant and spoken by the male speaker.

We must also be careful not to expose our subjects to hearing levels which could jeopardize their hearing, notably during exposure to loud noise. When subjects test the bone vibrator headband in noise, they must wear hearing protection devices, i.e. ear plugs.

4.4.2 Listening level with COMTAC

The listening level with the COMTAC headset is adjusted on an artificial head [4] at 84dBA in noise and at 72 dBA in silence. The equivalent level (free front field) is obtained by withdrawing the FTOE (Fonction de Transfert de l'Oreille Externe – External Ear Transfer Function), i.e. 78 dBA in noise and 66 dBA in silence. Please note that these levels cannot harm hearing.

4.4.3 Listening level with the bone vibrator headband

a. Selecting the electric emission level

At high levels, the headband tested is subject to distortion and vibration. Selecting the emission electric level requires finding a compromise between the listening level, which must be high enough to perceive correctly in noise, and distortion, which must remain low for intelligibility to be preserved.

The maximum emission level we selected for the bone vibrator communications headband was approved by the manufacturer.

b. Problem with measuring the vibration level.

We need to determine the level used by each subject to listen with the bone vibrator headband. This level depends greatly on the pressure applied, on the nature of bone, on the presence of hair, fat, etc. It is thus necessary to evaluate an equivalent acoustic level.

c. Equivalent acoustic level

Given these conditions, we developed a subjective method, making a loudness comparison with a sound transmitted by a loudspeaker placed in front of the subject. This method was first developed by Bekesy in 1949. In silence, the equivalent listening level (in front free-field) ranges from 46 dBA to 56 dBA, depending on subjects, with an average of 51,3 dBA and a standard deviation of 3,6 dB.

In noise, the electric signal is increased by 10 dB. Furthermore, subjects are fitted with earplugs, and the occlusion effect occurs [1]. Sound level is then assessed to be 75 dBA.

4.4.4 Summarized listening levels

Sound level equivalent to front free-field (dBA)	COMTAC	OSTEO (bone vibrator)
Silence	66	estimate: 51
Noise	78	estimate: 75

Table 1: Equivalent sound level in front free-field with COMTAC and bone vibrator

4.5 Protecting subjects from noise

French Law requires limiting exposure to noise. Noise must not exceed 85dBA for a daily exposure of 8 hours, 5 days a week. Beyond 85 dBA, each time the sound level increases by 3 dB, admissible exposure time is cut in half.

In a 97dBA noisy environment, subjects can only be exposed for half-an-hour with the headband, and bare ears. Of course, intelligibility would be zero. For this reason, during tests in noise, subjects are fitted with classic foam Ear® plugs (7/9), or Quies® earplugs (2/9). Under these conditions, our subjects are well protected from noise, and can be exposed to noise up to 4 hours a day.

5. INTELLIGIBILITY MEASUREMENTS

5.1 Test conditions

A training or test session is made up of 57 sentences, each including a CVC word.

Each subject first undergoes five training sessions. The first four are conducted in silence, two with COMTAC, two with the bone vibrator headband. The last session is conducted in noise, with the bone vibrator headband.

Each subject then undergoes eight test sequences:

2 environments (silence, noise)* 2 transducers (Bone vibrator, COMTAC) *2 speakers (male, female). The order of sequences is randomly drawn.

For COMTAC, in silence, the “hearing” function is activated with minimal gain. In noise (moving armoured vehicle), the “hearing” function is cut off.

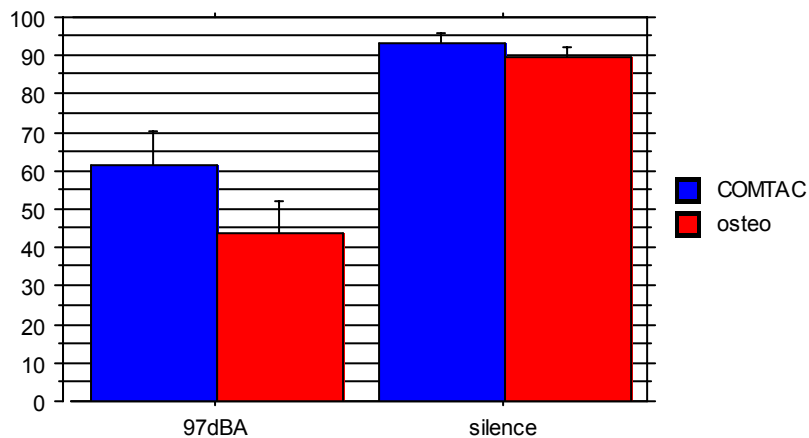
5.2 Results

5.2.1 Overall performance

In silence (figure 3), results are close: $93 \pm 5\%$ words are perceived correctly with COMTAC, and $90 \pm 5\%$ with the bone vibrator (significant difference, $p < 0,05$). When checking the performance for each C-V-C answer (initial consonant, vowel, final consonant), there is no significant difference between the two technologies.

In noise, the bone vibrator performs more poorly than the COMTAC: $44 \pm 17\%$ versus $62 \pm 17\%$ (significant difference, $p < 0,004$). Difference in performance comes from the final consonant. There is no significant difference between the two technologies for the initial consonant and the vowel.

Figure 3. Percentage of words correctly perceived, effect of environment*headset



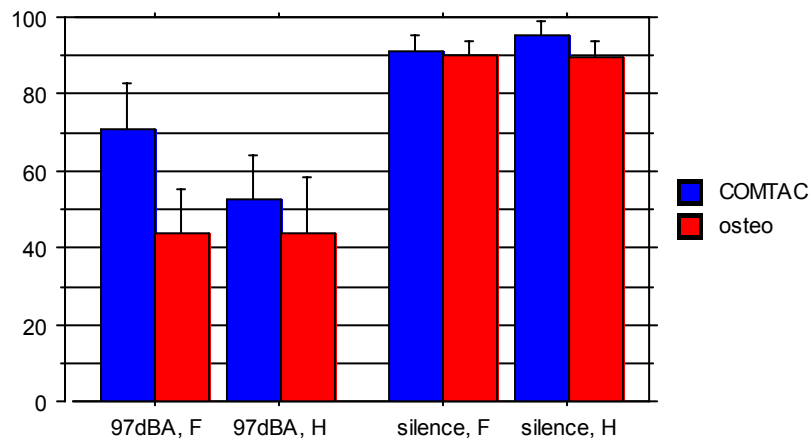
Error bars correspond to a 95% confidence interval

5.2.2 Speaker effect

There is no speaker effect with the bone vibrator (figure 4). In other words, male and female voices are perceived identically.

COMTAC in noise performs better with a female speaker.

Figure 4. Percentage of words correctly perceived, effect of environment*speaker (M/F)*headset



Error bars correspond to a 95% confidence interval

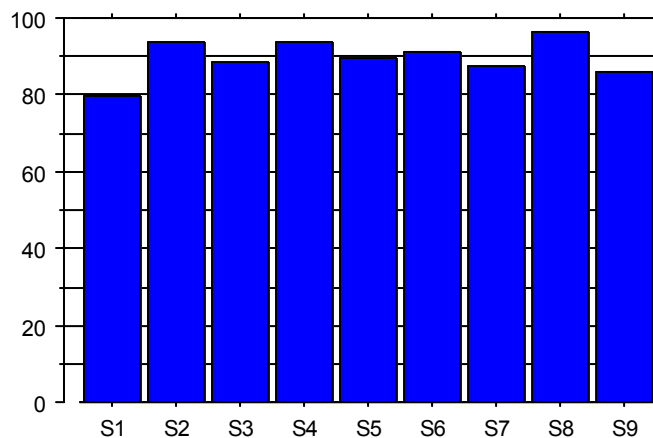
5.2.3 Subject effect

Does performance with the bone vibrator depend on the subject?

Subject effect in silence

There is a statistically significant subject effect (see figure 5), but performance is still above 80%.

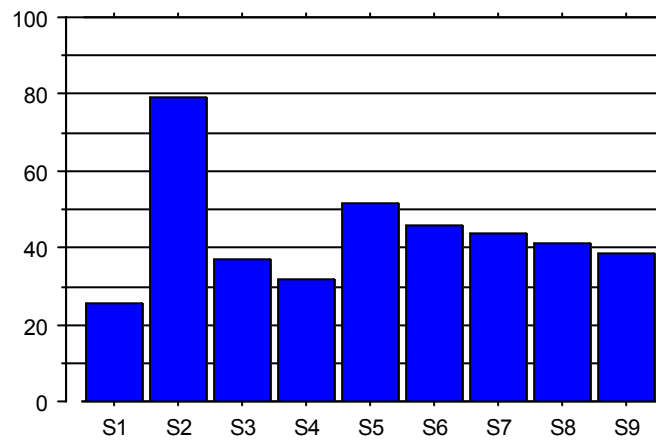
Figure 5. Percentage of words correctly perceived, bone vibrator in silence, subject effect



Subject effect in noise

Subject effect is much greater in noise. Subjects do not all perform identically in noise (see figure 6)

Figure 6. Percentage of words correctly perceived, bone vibrator in noise, subject effect



5.2.4 Summarized results

In silence, both bone vibrator and COMTAC obtain identical intelligibility scores.

In noise, the bone vibrator is not as good as the COMTAC.

With the bone vibrator headband:

- there is no speaker effect
- subjects do not all perform identically, notably in noise.

With the COMTAC,

- there is a speaker effect in noise,
- all subjects perform identically, in silence and in noise.

6. ANSWERS TO THE QUESTIONNAIRE FILLED BY SUBJECTS AFTER TESTS

6.1 COMTAC “hearing” headset

1. Regarding the voice transmission level, would you say it is:

	Too low	Low	Adequate	Strong	Too strong
In silence	0	0	6	3	0
In noise	0	2	7	0	0

The majority of subjects find the transmission level adequate. However, 3/9 subjects find the volume loud in silence.

2. Regarding outside noise perceived, would you say it is:

Not much nuisance	A little nuisance	Nuisance	Great nuisance
1	4	4	0

Overall, noise is considered as being a nuisance. However, this nuisance is moderate.

3. Regarding finding a comfortable listening position, would you say it is:

Very easy	Easy	Adequate	Difficult	Very difficult
4	3	2	0	0

Fitting the COMTAC headset is considered easy, or very easy.

4. Regarding voice quality, could you say it is:

Very good	Good	Average	Poor	Very poor
1	7	1	0	0

A majority of subjects considers voice quality to be good.

Speech Intelligibility with a Bone Vibrator

5. During or after the test with this headset, did you get a headache or did you feel dizzy ?

No	Yes
9	0

Among the nine subjects tested, none felt any side effects with the COMTAC headset.

6.2 Bone vibrator with earplugs

6. Regarding the voice transmission level, would you say it is:

	Too low	Low	Adequate	Strong	Too strong
In silence	0	0	9	0	0
In noise	2	4	3	0	0

In silence, the transmission level is considered as adequate. However, in noise, the transmission level is considered to be insufficient.

7. Regarding outside noise perceived, would you say it is:

No nuisance	Slight nuisance	Nuisance	Great nuisance
1	4	3	1

Overall, noise is considered as a nuisance. However, the nuisance remains moderate, in general. Noise is only considered to be a great nuisance once.

8. Regarding voice quality, would you say it is:

Very good	Good	Average	Poor	Very poor
0	4	4	1	0

Voice quality is not considered as good as with the COMTAC headset, but it remains good or quite good on average.

9. Regarding finding a comfortable listening position, would you say it is:

Very easy	Easy	Adequate	Difficult	Very difficult
0	1	3	4	1

For a small majority, finding a comfortable listening position is considered difficult. It seems to greatly depend on subjects' morphology.

10. Regarding the possibilities of adjustment provided on the metallic headband, would you say it is:

sufficient	Adequate	Not sufficient	I don't know
2	4	2	1

The metallic headband's adjustment possibilities were considered as insufficient twice. This number is high, given the number of subjects in test. This judgment is highly dependent on the difficulty the subject had in fitting the headband to his/her morphology.

11. Regarding fitting the elastic headband tightly on the head, would you say it is:

Not tight enough	Adequate	Too tight	I don't know
5	4	0	0

A majority of subjects considered the headband not tight enough.

12. In loud voice levels, have you felt vibrations in the headset ?

	Never	Sometimes	Often	Very often
In silence	4	4	1	0
In noise	2	7	0	0

A majority of subjects felt vibrations in silence, and mainly in noise.

13. If you felt vibrations, would you say they were:

	Not a great nuisance	A nuisance	A great nuisance
In silence	4	2	0
In noise	4	3	0

Vibrations felt were considered not a great nuisance, or a nuisance.

14. During loud voice levels, did you feel twitching?

	never	sometimes	often	Very often
In silence	7	1	1	0
In noise	6	3	0	0

A number of subjects felt twitching, especially in noise.

Speech Intelligibility with a Bone Vibrator

15. If you felt twitching, would you say it was:

	Not a great nuisance	A nuisance	A great nuisance
In silence	1	1	0
In noise	1	1	1

Twitching seemed to be quite a nuisance.

16. During or after the test with the hearing headset, did you get a headache or did you feel dizzy?

No	Yes
9	0

None of the nine subjects felt any headaches or dizziness after the tests.

6.3 . Miscellaneous

17. Regarding the attention required for the test, would you say it is:

	Greater with the hearing headset (COMTAC)	Greater with the headband (Bone Vibrator)	Identical for the two devices
In silence	0	5	4
In noise	0	8	1

In silence, a small majority considered the test with the bone vibrator to require more attention than with the other device.

In noise, almost all subjects considered the test with bone vibrator to require more attention. The subject who considered that the two devices required the same attention level is the one which scored very high when testing the bone vibrator in noise.

7. DISCUSSION ON INTELLIGIBILITY RESULTS AND ON ANSWERS TO THE QUESTIONNAIRE

7.1 Intelligibility in silence

Intelligibility is good in silence for both technologies (90% with the bone vibrator, 93% with COMTAC). In silence, the bone vibrator performs almost as well as the COMTAC. The difference in listening level (51dBA for bone vibrator vs. 66 dBA for COMTAC, see §4.4.4) probably helps making the COMTAC more intelligible. 3/9 subjects find the COMTAC's level "strong", and 9/9 find the bone vibrator's level "adequate".

As a technology, bone conduction seems to be a suitable choice.

7.2 Intelligibility in noise

In noise, intelligibility is more degraded with the bone vibrator than with the COMTAC (44% intelligibility with bone vibrator vs. 62% for COMTAC).

The lower scores obtained with the bone vibrator have various explanations:

- Our psychological and physical assessment shows a 3dB difference between listening levels in noise (78 DBA for COMTAC vs. 75 dBA for the bone vibrator, see §4.4.4). This is confirmed by subjects who declare that the bone vibrator's listening level is low (4/9) or too low (2/9). Using Steeneken's intelligibility curves (figure 2) [3], it seems that intelligibility drops down to a mere 40% after a -3 dB drop in the S/N ratio occurring around the operating point corresponding to 60% CVC intelligibility. Our results therefore seem consistent; the intelligibility loss observed may come from the listening level being too low. 7/9 subjects found the listening level with the bone vibrator low or too low. Electrically, it is possible to send more energy into the transducer. Unfortunately, this transducer then sends out parasite vibrations. The level used in the experiment, which is a compromise, was agreed upon together with the manufacturer.
- Fitting the device into the right position was difficult or very difficult for 5/9 subjects.

During the learning phase, the subject is taught how to position the headband, while listening to a continuous flow of words. The last training session is done in noise, with the headband. Results were always satisfactory. During the actual test session, CVCs are randomly selected, and the procedure aiming at adjusting the headband position was limited to the first five sentences of the session. These sentences, intertwined with silence, are not well suited to an adjustment phase. In these conditions, two subjects performed much worse in noise than during the learning phase. These sessions were not taken into account, and were repeated after the subjects were able to properly adjust the headband with a continuous flow of words.

Finding the right position can only be done by ear, and only if a continuous flow of words can be listened to. These conditions might be hard to meet in operations.

More adjustment possibilities should be available. 5/9 subjects find the headband too loose. However, even though it was not asked in the questionnaire, subjects clearly only want the headband to be tighter to improve its effectiveness in noisy environments.

It is important to note, that for at least one subject, placing the vibrating surface in contact with the zygomatic apophyses was extremely difficult.

7.3 Drawbacks of the Communications Headband (Bone vibrator)

* Listening with the bone vibrator requires more attention than with the COMTAC (8/9 subjects). This can be linked to the fact that 8/9 subjects find voice quality good or very good with COMTAC, whereas only 4/9 consider it satisfactory with the bone vibrator. The bone vibrator's reduced bandwidth and distortion are probably to blame. The listening level, which is too low, probably also negatively influences subjects.

* Even in test conditions when vibrations are moderate, they were felt by all subjects. However, only 3 considered them to be a nuisance.

In today's development phase of the mock-up device, 2 subjects felt twitching.

Speech Intelligibility with a Bone Vibrator

At loud level, test subjects felt a muscular contraction, called fasciculation. It is impossible to say whether fasciculation, which is a nuisance, comes from the amplitude of useful vibrations or from parasite vibrations.

* Out of the 9 people who performed the entire test, none experienced any headache or dizziness. The tenth subject felt bad after his first sequence with headband n°2 (the one with a difference in level between the right and the left channel). This person usually suffers from motion sickness. Tests were not continued.

It will be necessary to make sure that the proportion of people suffering from this kind of ailment is low.

CONCLUSION

Our conclusions only address the acoustic and technical aspects of the “listening” function.

In silence, intelligibility is identical with the bone vibrator as with the control system (COMTAC).

However, in noise, intelligibility is more seriously degraded with the bone vibrator than with the COMTAC

A number of reasons can explain this:

- the voice level is too low, because at high level, the transducer starts vibrating.
- it is difficult to properly position the bone vibrator to obtain optimal perception conditions.

In brief, bone conduction technology (reduced here to the mere listening function) has an appreciable initial potential, based on the performance obtained in silence. However, design improvements will be needed to reach the voice levels required for intelligibility in loud environments (notably for use in armoured vehicles).

BIBLIOGRAPHY

- [1] L'audition par conduction osseuse: intérêt pour le "combattant du futur" , A. DANCER, J. A. VERHAVE, T. HOUTGAST, Rapport I.S.L. n° SR115/2001, 30 p 2001.
- [2] Evaluation de casques à réduction active de bruit: protection auditive et intelligibilité, L. PELLIEUX, D. SARAFIAN, G. REYNAUD, AGARD Conference proceedings 596 (Audio effectiveness in aviation), 16.1-16.19, 1996.
- [3] On measuring and predicting speech intelligibility - H.J.M. STEENEKEN, Thèse de doctorat - 1992.
- [4] Artificial head for evaluation of hearing protectors, G. PARMENTIER., A. DANCER., K. BUCK, G. KRONENBERGER, C. BECK., Acustica Vol 86 pp 847-852, 2000.

Evaluation of Speech Synthesis Systems using the Speech Reception Threshold Methodology

David A. van Leeuwen and Johan van Balken

TNO Human Factors
Postbus 23
3769 ZG Soesterberg
The Netherlands

{vanLeeuwen,vanBalken}@tm.tno.nl

ABSTRACT

The intelligibility of speech synthesis systems that are available nowadays is usually high enough to enable comparisons between different synthesis systems based on the speech quality. However, in some situations, like a civil aircraft cockpit, the acoustic environment may be such that intelligibility is a discriminating factor between systems. In this paper we propose a methodology for comparing speech synthesis systems based on the Speech Reception Threshold (SRT). With this method the signal-to-noise ratio is found at which 50% intelligibility of redundant sentences is reached. A system with a lower SRT value is said to be more robust against masking noise. We have compared 5 commercial speech synthesis systems (4 male voices, 5 female voices) in an SRT experiment using a masking noise that was spectrally equivalent to cockpit noise. SRT values range from -4.1dB to 1.1dB. An ANOVA revealed that two of the nine systems had a significantly lower SRT than the rest. There was also an effect of the test subject, which is remarkable because the SRT has usually small variability over listeners.

INTRODUCTION

The EU 6th framework programme project SAFESOUND involves research to implement modern speech and audio technologies in the cockpit of a civil airline aircraft, in order to enhance the safety of flights. One of the subsystems implements Direct Voice Output (DVO) in order to communicate system messages to the pilots. For the DVO system a commercial-off-the-shelf (COTS) speech synthesis system is envisaged. The main selection criterium for the system is maximum intelligibility of the messages in the noisy civil aircraft cockpit.

The quality of commercially available speech synthesis systems is considered to be high enough that for most applications the *intelligibility* of the systems is not an issue, but rather other performance measures such as speech quality or listening comfort. However, for applications that should increase safety in a noisy cockpit environment, we argue that intelligibility *is* of highest concern in the choice of a synthesis system, since the mis-interpretation of a spoken system message can have negative consequences to the safety of the aircraft. Several methodologies for assessing the intelligibility can be found in literature, e.g., diagnostic rhyme test (DRT, see [Voiers, 1977](#)), or the Diphone Test ([Pols et al., 1987](#); [van Bezooijen and Pols, 1987](#)). For a good overview of these methodologies, see [van Bezooijen and van Heuven \(1997\)](#). These methodologies were developed at a time when the technology of speech synthesis systems was not as advanced as it is nowadays. Then, typical systems were formant-based or diphone-based, while modern systems are based on unit selection. In these methodologies ambient noise is not the key factor for determining the intelligibility, but rather the specific technology that is used. We therefore investigated new methodologies and came up with extending the use of the Speech Reception Threshold (SRT) to speech synthesis systems.

The SRT methodology uses short redundant sentences as speech stimulus material that subjective listeners are exposed to. The sentences are linguistically meaningful, and consist of 8-9 syllables. This type of

sentence is better suited to the modern unit selection based synthesis system than more diagnostic linguistic material, such as Semantically Unpredictable Sentences (SUS) or nonsense consonant-vowel-consonant (CVC) words or variants thereof. The SRT determines the signal to noise ratio (SNR) at which 50% of the test sentences are perceived correctly by test subjects. The measure can be determined accurately using a relatively low number of listeners, because the variance within (normal hearing native) listeners is small. Because the psychometric curve that describes the relation between fraction of correctly recognized sentences and SNR is rather steep, the SRT value is a good measure for the minimum SNR needed for proper intelligibility of messages. Therefore, the SRT value for a speech synthesis system in a noisy environment is directly related to the sound level at which synthesized messages need to be played in order to be perceived correctly. In a civil aircraft cockpit, with many different sound sources and a non-negligible background noise level, it is desirable to keep the sound level of these synthesized messages low, implying the need for a low SRT value.

EXPERIMENTAL SETUP

In an SRT experiment ([Plomp and Mimpen, 1979](#)) a test subject listens to a list of 13 recorded sentences one by one. The speech signal of each sentence is masked by noise, and the level of the noise (or signal to noise ratio) varies within the list. The subject is requested to orally repeat the sentence. The first sentence in a list is presented at a very low SNR, and repeated with increasing SNR until it is repeated correctly. In the next steps a new sentence is presented irrespective of the listeners response, until all 13 sentences have been played. If the sentence is repeated correctly, the noise level is increased, and if one or more words is not correct the noise level is decreased. The SRT-level is defined as the average signal-to-noise ratio over the last 10 sentences of the list. If the standard deviation of the last 10 SNR values exceeds 3dB, the SRT value is considered invalid and the data point is not used in subsequent analysis. The speech level is determined using the 'speech level meter' (SLM) algorithm ([van Velden, 1991](#)), which accounts the root-mean-square level of the upper 14dB of the speech level histogram.

We used American English SRT sentences applied previously in non-native speech research ([van Wijngaarden, 2001](#)). The sentences were transformed to speech by 5 different COTS systems, with both male and female voices resulting in 9 voice conditions (one supplier had only a female voice available). The systems voices were also of American English accent. The systems were used 'out of the box', and no adaptation of the sentences to particular systems was applied. An exception was made for the word 'aeroplane' used in one of the sentences, which several synthesis systems could not pronounce properly. This word was changed to the less British 'airplane.'

10 × 10

A panel of 10 Dutch subjects participated in the SRT test in a Latin square design. In total we used 10 different lists of 13 sentences, so that for each synthesized voice 130 sentences were produced. Because there was one voice condition less than the number of subjects or sentence lists, the Latin square design skipped one list of sentences for each subject. The masking noise used for the SRT was spectrally equivalent to noise in a Airbus A320 cockpit recorded at the location of the flying pilot during climbing of the aircraft. This stage of the flight has a higher noise level in the cockpit than during other stages, so it was chosen to represent the most critical condition with respect to intelligibility.

RESULTS

In Figure 1 the distributions of SRT over listeners are depicted for the various voices (systems). The measured average SRT values for the nine systems lie in the range -4.2 dB to $+1.1$ dB. These results are in the same range as normal SRT values of -3 dB for undistorted wideband speech of human

speakers with native listeners, and to SRTs around $+1$ dB for Dutch subjects listening to native English speakers (van Wijngaarden, 2001). The better systems appear to score as well as human speakers in the non-native condition for the SRT, but we must remark here that in those SRT experiments the masking noise is shaped according to the long-term average spectrum of the speaker, thus providing optimal masking. The spectrum of the noise used in this experiment has most energy in the lower frequencies, rolling off with an approximate slope of 3dB/octave. Both speech signal and noise spectrum have been A-weighted for determining the SNR.

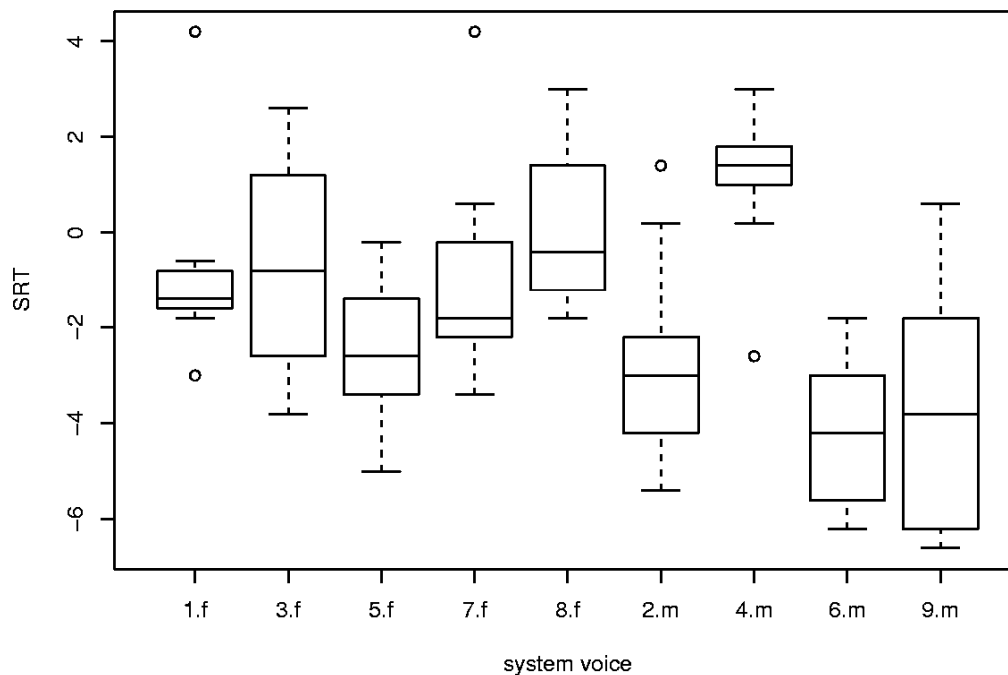


Figure 1: Boxplots of the measured SRT values averaged over synthesis system. A box indicates the first to third quartile of the distribution of SRT value among subjects. The horizontal line represents the median. The 'whiskers' indicate the range of remaining data points, except for extremes which are plotted individually. The gender of the system's voice is indicated in the system label as '.m' or '.f'.

Of the 90 SRT data points (9 systems $\times$ 10 subjects), 78 were valid. The results of the experiments have been analyzed in an analysis of variance (ANOVA). The results for the main effects ('system' and 'subject') have been tabulated in Table 1.

Table 1: SRT Anova results of main effects

Factor	<i>F</i>	<i>p</i>
Speech synthesis system	12.6	2×10^{-10}
Subject	8.4	5×10^{-8}

There are clear main effects for the factors 'system' and 'subject'. A pairwise t -test shows that only systems 4, 6 and 9 have SRT values significantly different ($p < 0.05$) from most other systems. It is interesting that all these systems use a male voice. Despite the wide range of SRT values for systems with male voice, a t -test between male voice and female voice data reveals a significant difference in SRT means of 1.3dB ($p = 0.024$).

As can be seen from Figure 1 the sizes of the boxes, indicating the 25th and 75th percentile, are generally large. This suggests a large spread in SRT values over the different subjects. In order to test if this is systematically due to the subjects, the means of SRT values over systems for the different test subjects are

plotted in Figure 2. The means over system range from -3.2 dB to 0 dB with an outlier at 1.6 dB for subject 10. This is more variability than expected, because the SRT method normally yields results with small standard deviations within subjects (Plomp and Mimpen, 1979). The observed variance within listeners, $\sigma = 2$ dB might be due to the inherent non-native character of the experiment. Van Wijngaarden et al. (2002) report a $\sigma = 1$ dB. In this case, subject 10 might have been less familiar with listening to the English language than the others. Although the subjects were asked if they felt they would be able to understand the English well enough, their English proficiency was not explicitly measured.

CONCLUSION

The experiment shows that the methodology of using the SRT with speech synthesis systems can be useful in an applied scenario such as an aircraft cockpit, where ambient noise is an important operational factor. A noteworthy difference with traditional SRT measurements is that the masking noise is not spectrally shaped like the long-term spectrum of speech, but according to the actual noise at the application domain. This leads to SRT values which are not directly comparable to other values reported in literature, but we argue they *are* meaningful to the specific application.

The power of the current experiment seems to be a little too low to make a well-funded choice for a particular speech synthesis system. With this first experiment we can estimate the power needed for subsequent experiments. We argue that more investigation of the non-native factor is necessary given the observed high listener variability and the international use of English-centered aviation equipment.

It is interesting that synthesized male voices have lower SRT values (i.e., higher intelligibility in a noisy environment) than female voices. There are several ways of reasoning about this difference. The traditional explanation for this is that with a higher fundamental frequency the female voices lead to less filled formants, and therefore less discriminable vowels, than their male counterparts. For the redundant sentences used in the SRT experiments the intelligibility in noise might be based mainly on the recognition of these vowels with only timing information from the consonants, which themselves are masked by the noise. Another cause may be sought in the relative position of the male and female speech spectra with respect to the masking noise, and the A-weighting of the different octave bands for applying a particular SNR. For example, the 125 Hz octave band is known to contribute to the intelligibility, but does hardly count in the A-weighting. The difference between female and male speech in this band is significant.

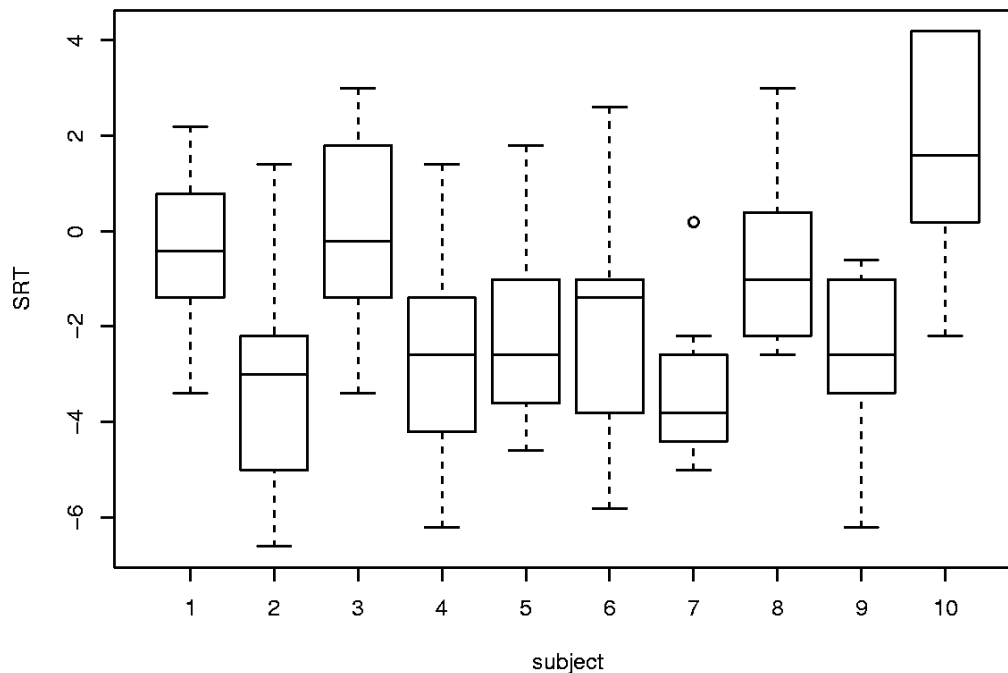


Figure 2: Boxplots of the SRT values averaged over different listeners.

Given the large spread in measured SRT values among systems, it appears that the intelligibility in noise is particularly dependent on the voice of the system. For instance, systems 8 and 9 are the female/male counterparts of the same manufacturer, and therefore share the same synthesis technology. Yet they show a large difference in SRT value, about 3.6dB.

Perhaps, referring to the well-known Turing test, we might claim that the quality of speech synthesis systems has really matured when they consistently score the same SRT values that humans do...

BIBLIOGRAPHY

- [1] R. Plomp and A. M. Mimpen. Improving the reliability of testing the speech reception threshold for sentences. *Audiology*, 8:-52, 1979.
- [2] L. C. W. Pols, J.-P. Lefèvre, G. Boxelaar, and N. van Son. Word intelligibility of a rule synthesis system for french. In *Proc. Eurospeech*, volume 1, pages 179-182, 1987.
- [3] R. van Bezooijen and L. C. W. Pols. Evaluation of two synthesis-by-rule systems for dutch. In *Proc. Eurospeech*, volume 1, pages 183-186, 1987.
- [4] R. van Bezooijen and V. van Heuven. *Handbook of Standards and Resources for Spoken Language Systems*, chapter Assessment of synthesis systems, pages 481-563. Mouton de Gruyter, 1997.
- [5] J. G. van Velden. Speech Level Meter, vesion 2, SAM_SLM. Technical report, TNO Human Factors, Soesterberg, 1991. SAM-TNO-043.

- [6] S. J. van Wijngaarden. Intelligibility of native and non-native Dutch speech. *Speech communication*, 35:-113, 2001.
- [7] S. J. van Wijngaarden, H. J. M. Steeneken, and T. Houtgast. Quantifying the intelligibility of speech using non-native listeners. *J. Acoust. Soc. Am.*, 111:-1916, 2002.
- [8] W. D. Voiers. Diagnostic evaluation of speech intelligibility. In M. E. Hawley, editor, *Benchmark papers in Acoustics*, number 2, pages 374-384. Dowden, Hutchinson, and Ross, Stroudsburg, 1977.

Extension of ITU-T Recommendation P.862 PESQ towards Measuring Speech Intelligibility with Vocoders

John G. Beerends

TNO Telecom, P.O. Box 5050, 2600 GB Delft, The Netherlands
Telephone: +316 51612563

E-mail: j.g.beerends@telecom.tno.nl

Sander van Wijngaarden, Ronald van Buuren

TNO Human Factors, P.O. Box 23, 3769 ZG Soesterberg, The Netherlands
Telephone: +31 346 356 330

E-mail: vanwijngaarden@tm.tno.nl, vanbuuren@tm.tno.nl

ABSTRACT

ITU-T recommendation P.862 PESQ was developed for assessing speech quality. The basic idea in PESQ is to compare a reference speech signal with the degraded signal through a psycho-acoustic model and a model of human quality comparison (cognitive model). Within NATO, testing of low bit rate speech codecs is more focused on speech intelligibility than on speech quality. Although the cognitive model of PESQ was designed to represent the quality judging process, it was already found that, for specific applications, PESQ can also predict intelligibility. In this paper PESQ is validated for assessing speech intelligibility with low bit rate vocoders. The results show that improvements in PESQ are necessary in order to obtain high correlations between objective and subjective intelligibility scores.

1 INTRODUCTION

ITU-T recommendation P.862 PESQ [1], [2], [3] was developed for assessing speech quality and was never validated in terms of speech intelligibility. Although there is a relation between speech quality and speech intelligibility, it is not clear that PESQ can be used to predict intelligibility. One should be aware of the fact that one can improve speech quality while decreasing intelligibility and reversely increase intelligibility while decreasing the speech quality. An example is the trend to improve the end-to-end perceived speech quality by using noise suppression systems that tend to improve the quality regarding the background noise but that tend to decrease the speech intelligibility.

This paper gives the results of a validation and extension of ITU-T Rec. P.862 on speech intelligibility. The database that was used results from a NATO speech intelligibility test on vocoders/noise suppressors and uses CVC (Consonant Vowel Consonant) intelligibility score. This data base consisted of long speech files (about 3 minutes long) containing 50 CVC (Consonant Vowel Consonant) words embedded in a carrier sentence. The intelligibility score is defined as the percentage of words for which both consonants and the vowel were identified correctly. This score is calculated per vocoder/noise condition and was measured as an average over four speech files (2 male and 2 female talkers). Twelve different noise conditions were used to assess the quality of 9 vocoders/noise suppressors. Bit rates of the codecs were between 1 and 5 kbit/s. For each condition 8 different sentences were used to assess the overall speech intelligibility score. Signals were presented diotically to listeners seated in a sound-treated room over wideband headphones.

2 PESQ P.862 VALIDATION RESULTS

Applying ITU-T Recommendation P.862 to the input and output speech samples of the database produces the results as given in Figure 1. Although the correlation between the objective PESQ MOS score and the CVC intelligibility score is quite high, $r = 0.86$, it is not up to a level that reliable intelligibility prediction can be made.

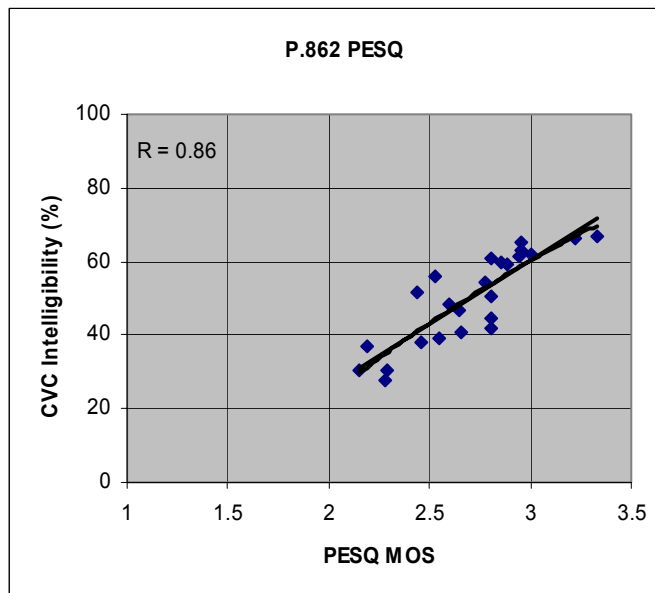


Figure 1: Relation between CVC Intelligibility and the P.862 PESQ Mean Opinion Score for the NATO speech intelligibility database.

3 EXTENDING PESQ P.862 TOWARDS MEASURING INTELLIGIBILITY

PESQ P.862 was developed for telephone band filtered speech signals [4] while the processing of the NATO speech signals involved no telephone band filtering in either the input to the system under test nor in the listening of the speech files in the subjective evaluation. The input files thus contained frequencies up to about 4000 Hz, half the sampling rate used in the experiments, and down to about 70-100 Hz, the natural lower limit of the human voice. The spectral balance tends to show that a microphone has been used that suffers from a low frequency boost to the proximity effect [5]. The output files in general showed almost no filtering effects although it would have been beneficial to filter the incoming speech below 300 Hz in order to get an optimized bit allocation in the speech encoder. Figure 2 gives a typical example of an input output frequency spectrum over a time span of about 2 seconds. It also shows significant amounts of distortion in the lower part of the frequency spectrum, between 50 and 300 Hz, a region that in normal telephone connections is filtered severely in order to optimize the signal to noise ratio (headroom), the perceived timbre and to minimize self masking as occurs when speech is played on a louder than natural level. Note that in a normal telephone connections the play back level is up to about 20 dB louder than a live conversation [6], making the self masking effect significant.

This analyses shows that a first straight forward idea for improving the correlation between PESQ scores and intelligibility scores is to extend the processed frequency range from the currently used IRS receive telephone band filtering (300-3400 Hz) to a wide band filtering (50-7000 Hz). However one should realise that PESQ was never developed to correctly model the loudness perception in the frequency bands

between 50 and 300 Hz. A simple removal of the input IRS filtering as used in PESQ only showed a marginally improved correlation and clearly other improvements need to be implemented in order to get higher correlations. Two ideas related to the spectral balance problem that should improve correlation are the introduction of upper slope frequency domain masking and the loudness growth in the lower frequency region (below 300 Hz). From these ideas only the loudness growth gave a significant improvement and was used in the PESQ intelligibility extension.

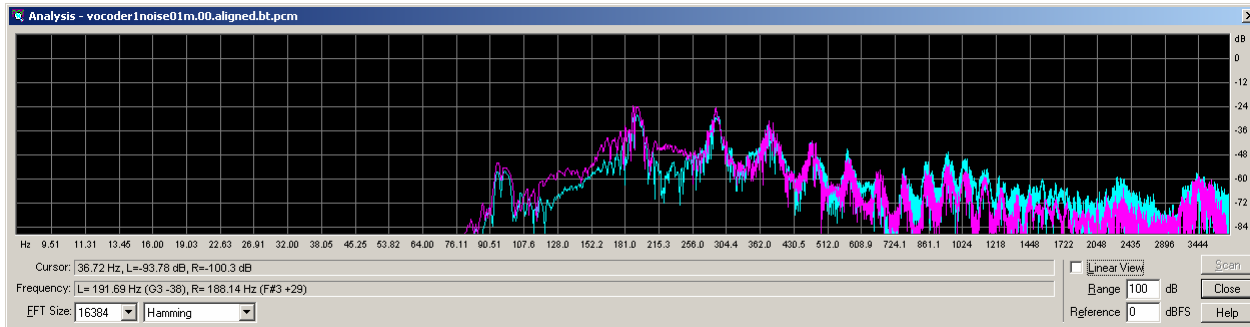


Figure 2: Typical example of the input and output power spectral density of a 2 second fragment as used in the NATO speech intelligibility database. In the lower part of the spectrum (below 300 Hz) the distorted vocoder output is higher than the input, while for the upper part of the spectrum (above 800 Hz) the reverse is true. From the view point of both quality and intelligibility this is a non optimal frequency distribution.

In order to further improve correlation one can take inspiration from the standard manner of measuring intelligibility using either the Articulation Index [7], [8] or the Speech Transmission Index [9], [10]. The AI is more or less a classical signal to noise approach, while the STI uses a modulation degradation calculation. Both of these effects are taken into account in the calculation of the PESQ MOS which is based on a frame by frame (32 ms length) bark power spectrum difference between the time aligned input and output speech signal [3]. Noise will result straight forward in a frame by frame bark power spectrum difference. Modulation differences between input and output will predominantly be seen as a global simultaneous change in all bark spectral frequency components. The effective frame length is in the order of 20 ms thus allowing to see modulation differences up to about 25 Hz. A major difference between the STI approach and the PESQ approach is that these global spectral changes only provide a global modulation difference and do not allow for a modulation frequency decomposition. An advantage is that the impact of each frequency component on the modulation is automatically weighted with its effective loudness. Because extreme low modulation frequencies (<0.1 Hz) are only perceived as a temporal change in volume these modulations are compensated for in PESQ by an adaptive frame by frame power scaling with a time constant of about four frames.

The time constant with which volume changes (i.e. low frequency modulations) are processed has a significant impact on the way that modulation differences are taken into account in PESQ and it was one of the further parameters that was changed in order to get a better correlation with the speech intelligibility.

Further significant improvements could be made by implementing the following ideas (in order of importance):

- Re-optimizing the integration over the time frequency distortion plane.
- Re-optimized to take into account the difference in impact of noise in the silent parts when comparing speech on intelligibility with speech quality. The impact of additive (background)

noise is taken into account in P.862 in an early stage of the perceptual mapping by applying a local scaling that reduces the noise level.

- Re-optimizing the linear frequency compensation. In P.862 PESQ linear frequency distortion are simply disregarded below a level of 20 dB, and distortions above 20 dB are reduced by 20 dB. This simple processing is replaced by a more advanced compensation that partly compensates all distortions, including distortions below 20 dB.

For the NATO database these improvements increased the correlation from 0.86 to 0.95. The results are given in Figure 3 where the PESQ MOS scale is replaced by an intelligibility score that is based on a compromise intelligibility scale with values that lie between a percentage CVC intelligibility and a sentence intelligibility score [11].

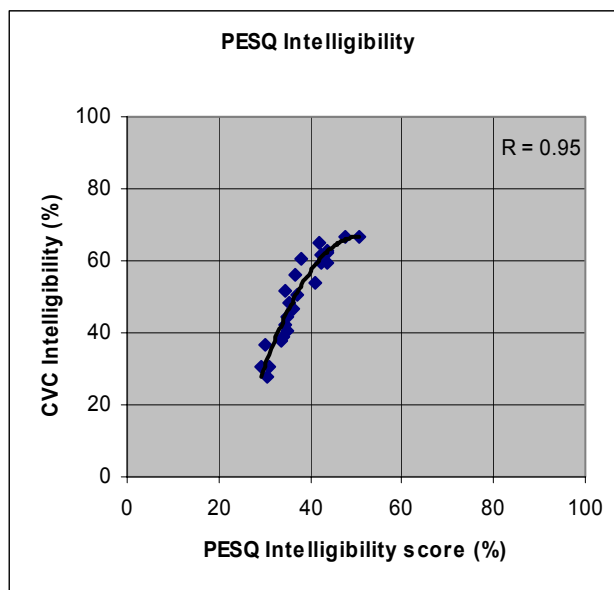


Figure 3: Relation between CVC Intelligibility and the P.862 PESQ Intelligibility score for the NATO speech intelligibility database.

4 CONCLUSIONS

The results show that PESQ P.862 [1], [2], [3] provides acceptable results when used to predict intelligibility. Significant improvements can be formulated that increase the correlation between objective and subjective intelligibility scores from 0.86 for PESQ up to 0.95. Further validations are necessary in order to see if the improvements that are implemented can cope with a wide range of distortions.

5 REFERENCES

- [1] ITU-T Rec. P.862, "Perceptual Evaluation Of Speech Quality (PESQ), An Objective Method for End-to-end Speech Quality Assessment of Narrowband Telephone Networks and Speech Codecs," International Telecommunication Union, Geneva, Switzerland (2001 Feb.).
- [2] A. W. Rix, M. P. Hollier, A. P. Hekstra and J. G. Beerends, "*PESQ, the new ITU standard for objective measurement of perceived speech quality, Part 1 - Time alignment*," J. Audio Eng. Soc., vol. 50, pp. 755-764 (2002 Oct.).

- [3] J. G. Beerends, A. P. Hekstra, A. W. Rix and M. P. Hollier, "*PESQ, the new ITU standard for objective measurement of perceived speech quality, Part II - Perceptual model*," J. Audio Eng. Soc., vol. 50, pp. 765-778 (2002 Oct.).
- [4] ITU-T Rec. P.48, "Specification for an Intermediate Reference System," International Telecommunication Union, Geneva, Switzerland (1989).
- [5] D. Josephson, "A Brief Tutorial on proximity Effect", contribution to the AES 107th Convention," preprint 5058, Sep. 1999.
- [6] ITU-T, "Handbook on Telephonometry," International Telecommunication Union, Geneva, Switzerland (1992).
- [7] N. R. French and J. C. Steinberg, "Factors governing the intelligibility of speech sounds," *J. Acoust. Soc. Am.*, vol. 19, pp. 90-118 (1947 Jan.).
- [8] K. D. Kryter, "Methods for the calculation and use of the Articulation Index," *J. Acoust. Soc. Am.*, vol. 34, pp. 1689-1697 (1962 Nov.).
- [9] H.J.M. Steeneken and T. Houtgast, "A physical method for measuring speech-transmission quality," *J. Acoust. Soc. Am.*, vol. 67, pp. 318-326 (1980 Jan.).
- [10] H.J.M. Steeneken, "On measuring and predicting speech intelligibility," PhD University of Amsterdam (1992).
- [11] J. G. Beerends, E. Larsen, N. Lyster, J. M. van Vugt, "*Measurement of speech intelligibility based on the PESQ approach*," Proceedings of the Workshop Measurement of Speech and Audio Quality in Networks (MESAQIN), Prague, Czech Republic, June 2004 (equivalent to TNO Telecom publication 33366).

ACKNOWLEDGEMENT

The authors wish to thank NATO NC3A for making available the speech intelligibility database used in this paper.



Military Speech Communications over Vocoders in Tandem

Mark A. Ericson, Brian D. Simpson, and Richard L. McKinley

Air Force Research Laboratory (AFRL/HECB)

2610 Seventh Street, Bldg 441

Wright-Patterson, AFB, OH 45433-7901

mark.ericson@wpafb.af.mil

VOCODERS IN TANDEM

Speech intelligibility of two types of vocoders was measured using the modified rhyme test. One type of vocoder, a continuous variable slope delta (CVSD), was a waveform encoder. The other type, an advanced multi-band excitation (AMBE), was a parametric encoder. In the first experiment, clear speech was processed through the vocoders. Intelligibility was measured in a control condition, i.e. without vocoding, with each type alone and with two vocoders in tandem. AMBE and CVSD performed similarly, 92.6 and 90.4%, respectively. CVSD-to-AMBE had little effect on intelligibility, measured at 89.2%. However, AMBE-to-CVSD had a large degrading effect on intelligibility. The AMBE-to-CVSD direction scored about 81.7% intelligibility with clear, unaltered speech signals. The asymmetry between waveform-to-parametric and parametric-to-waveform encoders underscores the non-linear nature of tandem vocoders on intelligibility. When vocoders of the same type were in tandem, there was no additional effect on intelligibility. The double CVSD condition yielded 92.2% intelligibility and the double AMBE condition yielded 91%. The deleterious effects of speech clipping were measured in a second experiment, as these are ubiquitous in military radio transmission systems. The AMBE parametric vocoder performed at the 88% level in isolation and at 84% when tandemed with the CVSD waveform vocoder. Alternative methods of encoding speech signals are being explored to improve speech intelligibility performance in military communication systems.

1.0 INTRODUCTION

There are two general classes of vocoders used in military communication systems today; these are parametric and waveform vocoders. Waveform vocoding techniques, such as continuous variable slope delta (CVSD), are highly resistant to noise and bit error effects [1, 2]. Parametric encoders, such as advanced multi-band excitation (AMBE) [3], greatly reduce signal bandwidth, which is helpful in reducing encryption processing requirements and the cost of transmitting a wide band signal. Both waveform and parametric vocoders can provide good speech intelligibility alone at adequate bandwidths [4]. However, a “staging” or “tandem” problem occurs when waveform encoders and parametric encoders are placed in sequence in a given communication system [5]. The distortion of the speech waveform produced by the first vocoder causes the second vocoder to severely distort the speech waveform, thereby reducing the overall intelligibility.

Vocoder algorithms have typically been developed to reduce bandwidth for long distance or secure communications [6, 7]. These devices are not necessarily designed to be compatible in conjunction with other types of vocoders. Military communication systems are likely to have legacy equipment, which will include parametric [7] and waveform types of vocoders [8]. In future military operations, speech communications are likely to occur among more operators from multiple points in the chain of command.

Military Speech Communications over Vocoders in Tandem

The tandem problem will potentially increase as remote-controlled air vehicles become more numerous in military operations. Operators of remote-control reconnaissance and attack air vehicles, such as the Global Hawk and uninhabited combat aerial vehicles (UCAVS), must communicate with civilian air traffic controllers and military command and control personnel. Ground troops equipped with satellite phones will need to communicate with other military operators via multiple communication links. Achieving good speech intelligibility over such multiple-link communication systems will be critical for safely and efficiently accomplishing military missions.

The test objective was to measure the effects of continuous variable slope delta (CVSD) and advanced multi-band excitation (AMBE) vocoding algorithms on speech intelligibility. These components are considered the critical links in the air traffic controller (ATC) to UAV ground control station communication path. A typical communication path is depicted in Figure 1. Note that the communication can go in both directions from the air traffic controller and the UAV ground control station. The direction of the path determines the order of the vocoders in tandem.

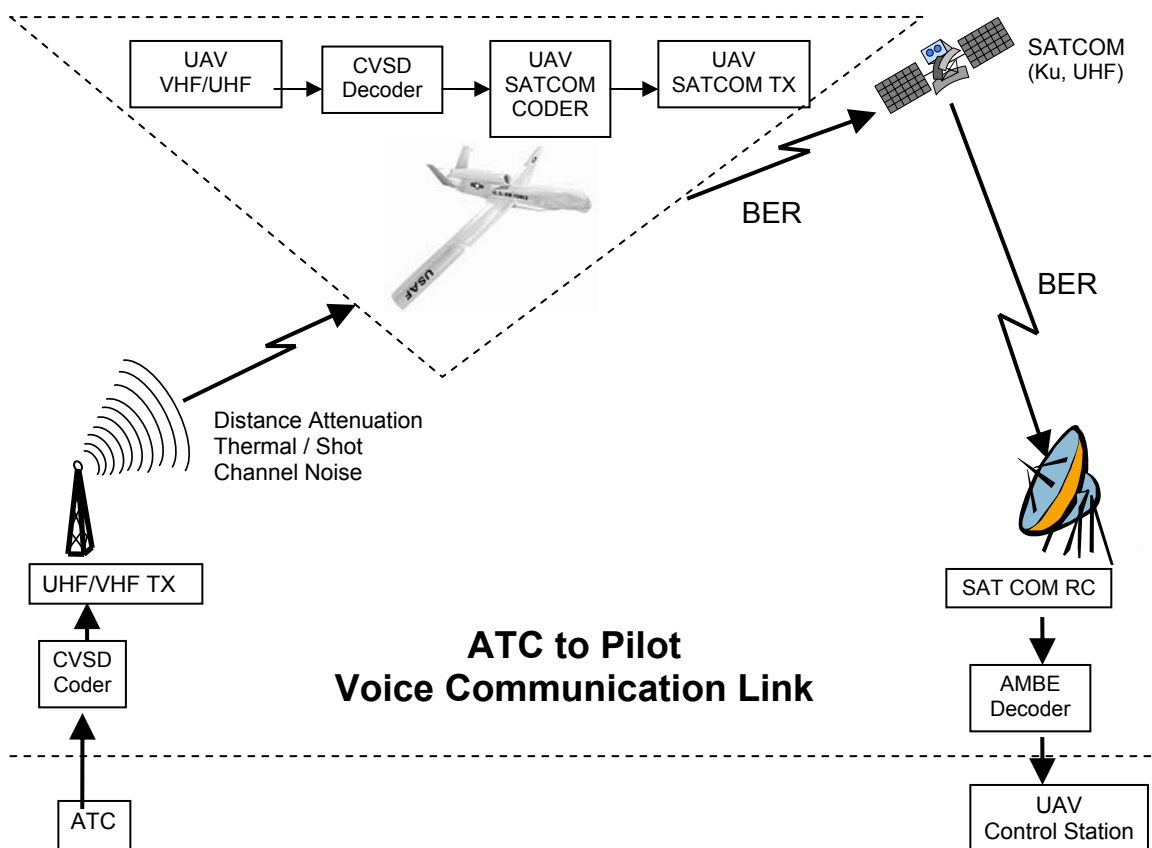


Figure 1: Depiction of communication paths from air traffic controllers to a UAV command and control station

2.0 METHODS

2.1 Equipment

The Air Force Research Laboratory's Battlespace Acoustics Branch (AFRL/HECB) operates and maintains unique facilities for researching and developing voice communication systems for military operating environments. The Voice Communication Research and Evaluation System (VOCRES) [9] is capable of simulating all parts of the military communication path from the talker, via the medium, to the receiver. The six modified rhyme test (MRT) [10] lists of 50 words each were read by three male and three female talkers and recorded in 16 bit format. The speech stimuli were pre-processed off-line using a Windows 98 PC, a 4.8 kbps AMBE processor, and a 16 kbps CVSD algorithm. The processed speech files were played back to the panel of professional listeners. The two types of vocoders were staged together and in both directions of the communication path. Listeners wore H-157A headsets at the response desks of VOCRES.

2.2 Subjects

Five listeners participated in the vocoder studies. The paid volunteer subjects ranged in age from 18 to 51 years with a mean age of 27 years. Each volunteer subject had normal hearing threshold levels and consented to participate in the speech intelligibility experiments.

2.3 Procedures

The degrading effects of the vocoding by single and tandem systems on speech intelligibility were measured using the MRT. The MRT is one of three standardized procedures for measuring the intelligibility of speech over communication systems. The speech utterances were recorded with the test word imbedded in a carrier phrase to reduce the deleterious effects of the attack portion of the automatic gain control circuitry on the intelligibility of the initial consonant. In each session, six pre-recorded talkers processed with the vocoders were played to listeners in a quiet environment. Responses were automatically collected and scored by the computers in the VOCRES facility. Baseline conditions were tested in which each talker's voice was processed with each class of vocoder in isolation and in tandem with each other.

3.0 RESULTS

3.1 Unaltered Speech Vocoded in Tandem

The first experiment determined the performance of the vocoders with unaltered (clear) speech signals. Raw data scores were corrected for guessing by using the formula of 2.4 times the raw number correct out of 50 minus 20. The mean and standard deviations for percent correct intelligibility are plotted in Figure 2. The percent correct values of speech intelligibility were measured to be 96.3% for the control condition of no vocoders, 92.6% for AMBE-alone, 90.4% for CVSD-alone, 89.2% for CVSD-AMBE, and 81.7% for AMBE-CVSD. In the like-vocoder condition, speech intelligibility was found to be 92.2% for CVSD-to-CVSD and 92.6% for AMBE-to-AMBE.

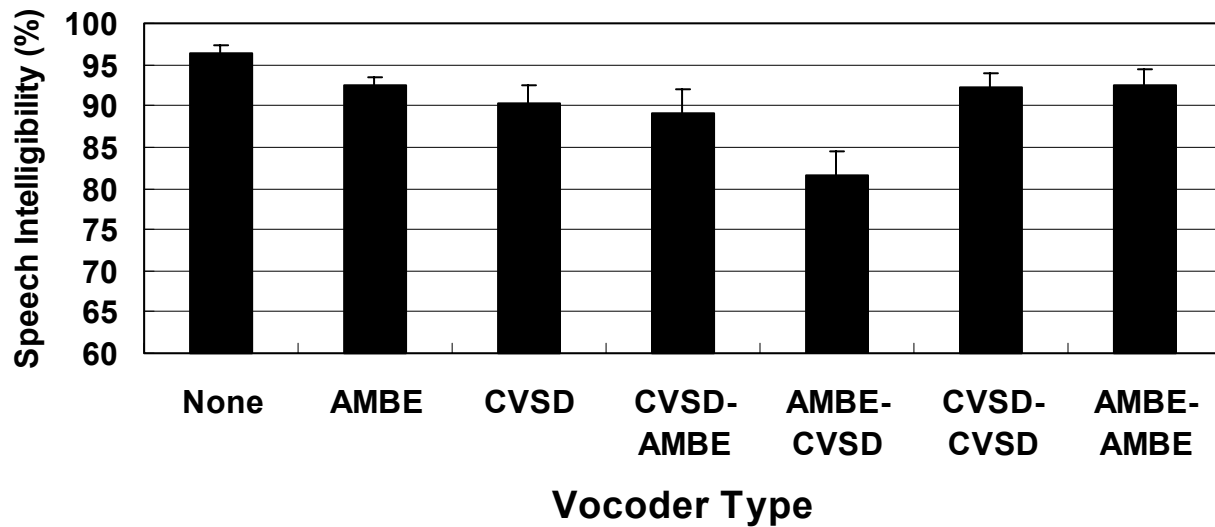


Figure 2: Speech intelligibility versus vocoder type with unaltered speech

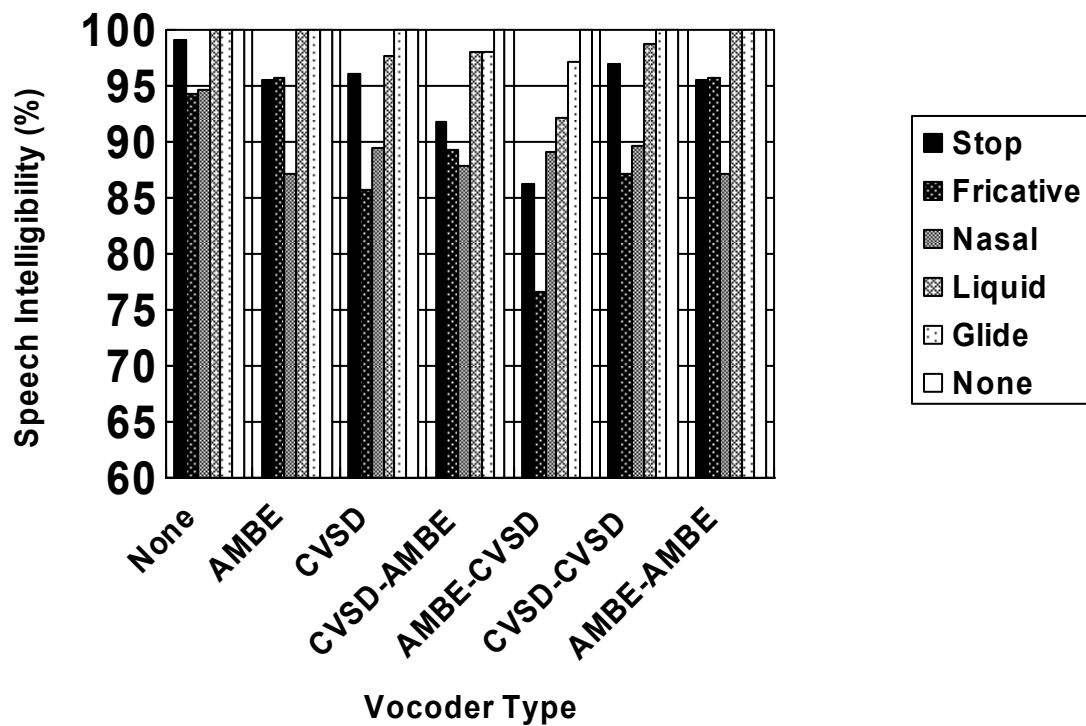


Figure 3: Speech intelligibility versus vocoder type and speech attribute with unaltered speech

A one-way analysis of variance (ANOVA) was performed on the speech intelligibility data. Percent correct scores were subjected to an arc-sine transformation before the ANOVA was performed [11]. The ANOVA revealed a main effect for vocoder type ($F=2.75(6, 6)$, $p=.042$). A post hoc least significant difference (LSD) analysis was performed on the main effects to look for differences among vocoders. The control condition of no vocoding and the AMBE-to-CVSD tandem condition were found to be different than other vocoder conditions.

The data were further analyzed by the effects of vocoding on the manner of articulation. The speech attributes included stop, fricative, nasal, liquid, glide, and phonemic absence. The mean values are shown in Figure 3 for each of the seven vocoding conditions. A two-way ANOVA again revealed a main effect for vocoder type ($F=12.75(6, 6)$, $p=.006$) and a main effect for attribute ($F=14.53(6, 6)$, $p=.002$). A post hoc LSD analysis was performed on the speech attribute data. Stops, fricatives, and nasals were found not to be significantly different from each other. Liquid, glide, and absent attributes were also found not to be different from each other.

3.2 Clipped Speech Vocoded in Tandem

The second experiment was designed to measure the performance of the vocoders with hard clipped speech signals at 10 dB down from the peak. The mean and standard deviations for percent correct intelligibility are plotted in Figure 4. The percent correct values of speech intelligibility were measured to be 96.3% for the control condition of no vocoders, 92% for AMBE-alone, 87% for CVSD-alone, 84% for CVSD-AMBE, and 75% for AMBE-CVSD. In the like-vocoder condition, speech intelligibility was found to be 90.2% for CVSD-to-CVSD and 82.6% for AMBE-to-AMBE.

A one-way analysis of variance was performed on the speech intelligibility data. Percent correct scores were subjected to an arc-sine transformation before the ANOVA was performed [11]. The ANOVA revealed a main effect for vocoder type ($F=2.45(6, 6)$, $p=.048$ MP-HFM-123-11). A post hoc LSD analysis was performed on the data to look for differences among vocoders. The control condition of no vocoding, AMBE, and CVSD were found to be not different from each other. The four tandem conditions, AMBE-to-CVSD, CVSD-to-AMBE, AMBE-to-AMBE, and CVSD-to-CVSD, were found to be not different from each other.

The data were further analyzed by the effects of vocoding on the manner of articulation. The speech attributes included stop, fricative, nasal, liquid, glide, and phonemic absence. The mean values are shown in Figure 5 for each of the seven vocoding conditions. A two-way ANOVA again revealed a main effect for vocoder type ($F=8.45(6, 6)$, $p=.012$) and a main effect for attribute ($F=3.53(6, 6)$, $p=.042$). A post hoc LSD analysis was performed on the speech attribute data. Stops, fricatives, and nasals were not found to be significantly different from each other. Liquid, glide, and absent attributes were also found not to be different from each other.

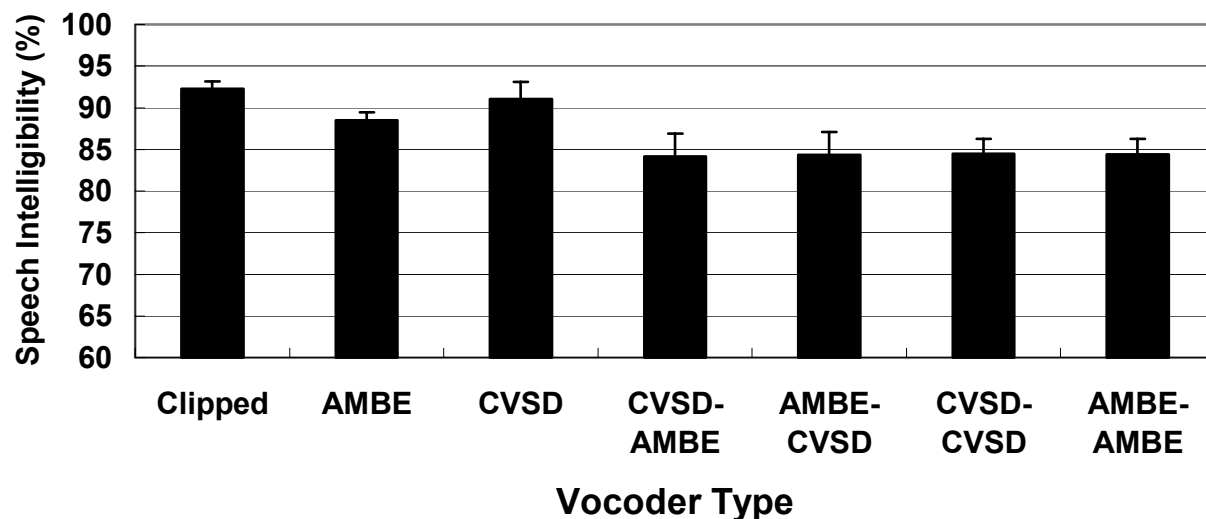


Figure 4: Speech intelligibility versus vocoder type with input speech signal hard clipped by 10dB

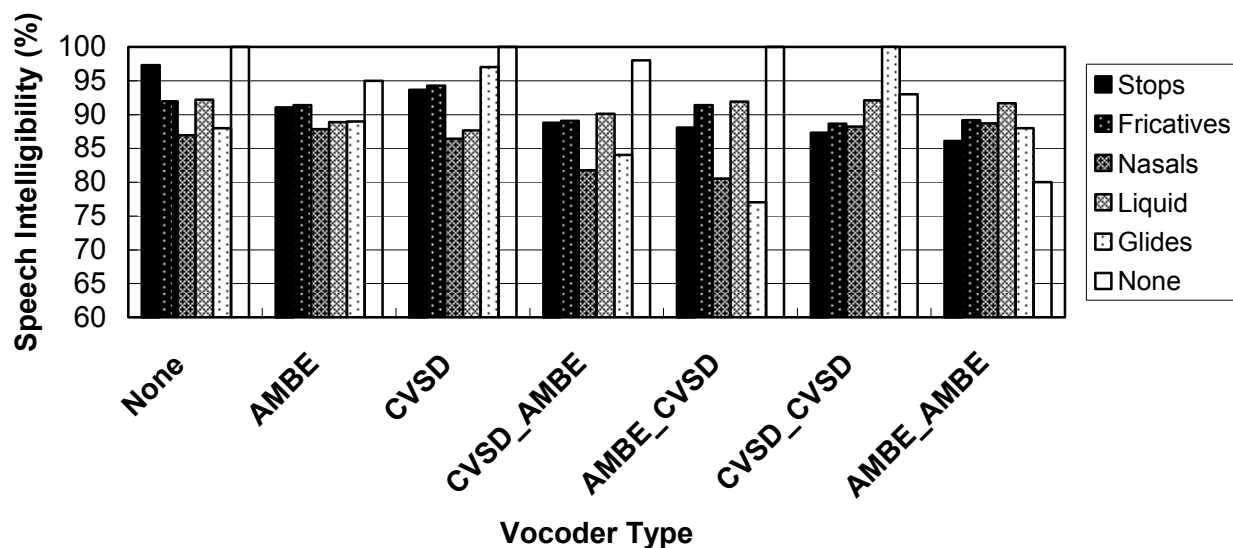


Figure 5: Speech intelligibility versus vocoder type with the speech signal hard clipped by 10dB

4.0 DISCUSSION

Both vocoders performed well with clean speech inputs. However, the AMBE parametric type of encoder performed much worse with the clipped speech than did the CVSD encoder. The CVSD vocoder was not able to process the output of the AMBE vocoder well. Conversely, the AMBE vocoder was able to process the output of the CVSD vocoder when unaltered speech was used. The CVSD vocoder performed well despite its simple algorithm and lack of knowledge of speech characteristics.

The effects of tandem vocoding and speech clipping were investigated in the current study. Many natural environmental factors and hostile jamming devices can potentially degrade the end-to-end speech intelligibility of the communication system. Deleterious effects include radio frequency channel noise, signal loss over long transmission distances, ambient acoustic noise at the talker or listener locations, encryption errors, bandwidth limitations, channel bit errors in the digital link, and burst disturbances. All of these factors and possibly more can have a degrading effect on speech intelligibility.

An implicit question in all speech intelligibility measurements is how much intelligibility is good enough for a given application. MIL-STD-1472F [12], the Department of Defense Design Criteria Standard for Human Engineering, suggests 91% intelligibility performance should be achieved on the MRT for operational military communication equipment and 97% performance for critical communications. Most military operators consider a voice communication system with 80% or better performance on the MRT to be acceptable, those between 70 and 80% to be marginal, and those below 70% to be unacceptable [4]. A further consideration is that air traffic controllers who are non-native speakers of English need a voice communication system with higher than normally acceptable speech intelligibility levels, i.e. greater than 80% with the MRT.

Since CVSD was developed in the 1970's, other waveform coding schemes have been created. Two such methods are adaptive differential pulse code modulation (ADPCM) and voice over internet protocol (VOIP). Each of these methods can out-perform CVSD in unperturbed, laboratory environments. However, all three react differently to disturbances in real-world environments. ADPCM is slightly more tolerant of bit error rates than CVSD, but requires more bandwidth to accomplish that level of performance [2]. VOIP issues deal with delay, jitter, and packet loss during transmission over digital communication links [6]. Further research is needed to understand the trade-offs between disturbances on speech intelligibility with VOIP techniques. The interaction of new and legacy vocoders when coupled in tandem should be investigated before such new systems are introduced into military communication systems.

5.0 CONCLUSIONS

Speech intelligibility performance of two types of vocoders was measured with clear, recorded speech and altered speech, hard clipped at 10 dB. The waveform vocoder was a continuous variable slope detector (CVSD) and the parametric encoder was the advanced multi-band excitation (AMBE) encoder. Intelligibility was measured in a control condition, with each type, in tandem with the opposite type, and in tandem with the same type. The AMBE and CVSD methods performed well in isolation, measured at 92.6 and 90.4% intelligibility, respectively. The CVSD-to-AMBE direction had little effect on intelligibility, measured at 89.2%. However, the AMBE-to-CVSD direction had a large degrading effect on intelligibility, 81.7%. The addition of a second vocoder of the same type had little effect on intelligibility. Intelligibility with the clipped speech input signals was generally lower than with the unaltered speech signals. Intelligibility levels in the tandem conditions with clipped speech were lower than with the unaltered speech, except for CVSD-AMBE.

A finer analysis of the effects of clipping and vocoding on speech intelligibility was evidenced in the phonemic analysis. Stops, fricatives, and nasals were adversely affected by the processing through a single vocoder and especially with dissimilar vocoders in tandem. Clipping had a large effect on all phoneme identifications. Liquids and glides were unaffected by vocoding in the clear speech condition, but were reduced in intelligibility by vocoding in the clipped speech condition.

The combination of two vocoders in tandem generally reduced performance in a non-linear way, especially in the clipped speech conditions. The asymmetry between waveform-to-parametric and parametric-to-waveform

encoders underscores the non-linear nature of tandem vocoders on intelligibility. Consideration should be given to the application of new vocoding techniques when embedded in military environments with legacy vocoders, such as the CVSD algorithm. Alternative methods of encoding speech signals are being explored to improve speech intelligibility performance in military communication systems.

6.0 REFERENCES

- [1] McKinley, R. L. and Moore, T. J. (1985). *The effect of audio bandwidth on selected digital speech coding algorithms*, 1985 IEEE Military Communications Conference MILCOM'85, Boston, MA, October 20-23, 1985161-165, 9.1.1-9.1.5.
- [2] McKinley, R. L. and Kurtz, R. V. (1984). *Subjective Performance of Selected Speech Coders in the Presence of Channel Errors* 1984 IEEE Military Communications Conference MILCOM'85, Los Angeles, California, October 21-24, 1985161-165, 9.1.1-9.1.5.
- [3] <http://www.dvsinc.com/papers/mbe.htm>
- [4] Moore, Thomas J. (1989). Speech Technology in the Cockpit, *Proceedings of Aviation Psych.*, 50-65.
- [5] <http://www.arcon.com/ddvpc/tandem.htm>
- [6] Atal, Bishnu S. (2005). 50 Years of Progress in Speech Waveform Coding, *Echoes The Newsletter of the Acoustical Society of America*, Volume 15(1), pp 1 and 4.
- [7] Street, Michael D. (2002). Host Laboratory Role in the Selection of the Future NATO Narrow Band Voice Coder, IEEE Workshop on Speech Coding, L3.4, Tsukuba, Ibaraki, Japan, Oct 6-9, 2002.
- [8] Military Communications, RTO MP-065, IST-023, February, 2002.
- [9] McKinley, R. L. (1980). Voice Communication Research and Evaluation, AMRL-TR-80-25.
- [10] ANSI Standard S3.2-1989 (rev. 1999) Methods for Measuring the Intelligibility of Speech Over Communication Systems.
- [11] Studebaker, G. A. (1985). A "Rationalized" arcsine transformation. *Journal of Speech and Hearing Research* 28:455-462.
- [12] Mil-Std-1472F (23 August 1999) Department of Defence Design Criteria Standard Human Engineering.

Objective Measurement of the Speech Transmission Quality of Vocoders by Means of the Speech Transmission Index

Mr. Bastiaan J.C.M. van Gils

TNO Human Factors P.O. Box 23, 3769 ZG Soesterberg, The Netherlands
Telephone +31 346 356 218, Telefax +31 346 353 977

vanGils@tm.tno.nl

Dr. Sander J. van Wijngaarden

TNO Human Factors P.O. Box 23, 3769 ZG Soesterberg, The Netherlands
Telephone +31 346 356 230, Telefax +31 346 353 977

vanWijngaarden@tm.tno.nl

ABSTRACT

Nearly all types of military speech communication involve the use of so-called (narrow band) voice coders or vocoders. Usually the Speech Transmission Index (STI) uses artificial test signals, which can not be reproduced by vocoders with the usual fidelity. Therefore the STI is not able to evaluate vocoders at this time. Although it is theoretically feasible to measure the Speech Transmission Index with natural speech instead of the usual artificial test signals, each of the various speech-based STI measurement methods proposed in the literature has its own shortcomings and inaccuracies. A new procedure is proposed for estimating a speech based modulation transfer function (MTF), on which the STI is based, that approaches the accuracy of conventional STI implementations based on artificial signals. The new method enables evaluation of vocoders by means of the STI. Applying the method to a voice coder database shows promising results, giving an average squared correlation coefficient R^2 of 0.87 between the subjective CVC scores and the calculated STI for male speech.

INTRODUCTION

Nearly all types of military speech communication involve the use of so-called (narrow band) voice coders, or vocoders. These vocoders efficiently digitize human speech before transmission, making use of knowledge of the characteristics of the human voice and hearing. Speech communication equipment based on such vocoders allow more efficient use of the electromagnetic spectrum available for radio transmissions, while at the same time offering greatly increased possibilities for encryption and end-to-end secure transmission. Unfortunately, the Speech Transmission Index method, perhaps the most commonly used method for objective measurement of speech intelligibility, can not be applied when vocoders are involved. This method normally makes use of artificial test signals, which can not be reproduced by vocoders (which are optimized for natural speech) with the usual fidelity. As a consequence, speech intelligibility verifications of vocoders, as frequently required by NATO, depend on expensive and cumbersome subjective procedures involving listening panels. It is theoretically feasible to measure the Speech Transmission Index (STI) with natural speech instead of the usual artificial test signals. This has already been shown shortly after the introduction of the STI [1]. The speech-based approach would eliminate the problems as outlined above. However, each of the various speech-based STI measurement methods proposed in the literature [2] - [4] has its own shortcomings and inaccuracies. A new procedure is proposed for estimating the modulation transfer function (MTF), on which the STI is based, that approaches the accuracy of conventional STI implementations based on artificial signals.

van Gils, B.J.C.M.; van Wijngaarden, S.J. (2005) Objective Measurement of the Speech Transmission Quality of Vocoders by Means of the Speech Transmission Index. In *New Directions for Improving Audio Effectiveness* (pp. 12-1 – 12-6). Meeting Proceedings RTO-MP-HFM-123, Paper 12. Neuilly-sur-Seine, France: RTO. Available from: <http://www.rto.nato.int/abstracts.aps>.

METHODS AND DATA

A preliminary comparison of speech-based MTF (sMTF) estimation methods showed that the methods proposed by Steeneken en Houtgast [1] and Ludvigsen [2] do not perform as well as those by Drullman [3] and Payton [4]. Therefore we focused on the latter two methods. Mathematically the latter two methods are alike:

$$MTF_{drullman} \sim \frac{\Re(\text{crossspectrum})}{\|\text{autospectrum}\|}$$
$$MTF_{payton} \sim \frac{\|\text{crossspectrum}\|}{\|\text{autospectrum}\|}$$

While the Drullman method uses the real part of the cross-correlation spectrum to estimate modulation transfer, the Payton method uses the cross-correlation spectrum magnitude. Difference between the methods is the weighting of out-of-phase transfer of modulations. Comparison of measurement results according to the Drullman and Payton methods indicates that the accuracy of the STI method may be improved by weighting out-of-phase modulations. Although the (almost implicit) phase-weighting approach adopted by Drullman is effective, its method and degree of phase-weighting is rather arbitrary.

This observation led us to propose a new speech-based STI phase-weighting method. This method is basically a generalization of the methods by Payton and Drullman, using an explicit and adjustable phase-weighting of the cross-correlation spectrum.

$$MTF_{phaseweight} \sim \frac{\|\text{crossspectrum}\| \cdot f(\angle(\text{crossspectrum}))}{\|\text{autospectrum}\|}$$

To test the phase-weighting method we first compared the results to a theoretical STI, calculated for simulated noise and reverberation conditions [5]. From the results we selected the best performing variant of the method.

This variant was applied to material of a narrow-band voice coder (NBVC) database [8] – this database was also used in Beerends & van Wijngaarden, this volume, consisting of speech signals, as processed through a set of 9 different military vocoders (including legacy systems as well as the current state-of-the-art, at bit rates ranging from 1200 – 4800 bps) in 12 channel conditions. Each of the 108 combinations was tested using 8 talkers (4 male – 4 female). Speech intelligibility was measured using standard subjective assessment methods (CVC and SRT). The database also entails the audio of the original CVC word lists and the input and output signals. We extracted the mean CVC score in 216 different vocoder – condition – gender combinations and calculated the mean STI of each combination. The results were compared to the well established relation between STI and CVC [6].

RESULTS

We simulated 14 speech intelligibility reducing conditions (Noise: 7 SNR ratios / Reverberation: 7 reverberation times) and calculated the theoretical STI value. The simulated speech files were processed. Since time delay between input and output signals influences the results of the phase-weighting method, we aligned the input and output signals in time, based on maximum correlation of the input and output speech envelope.

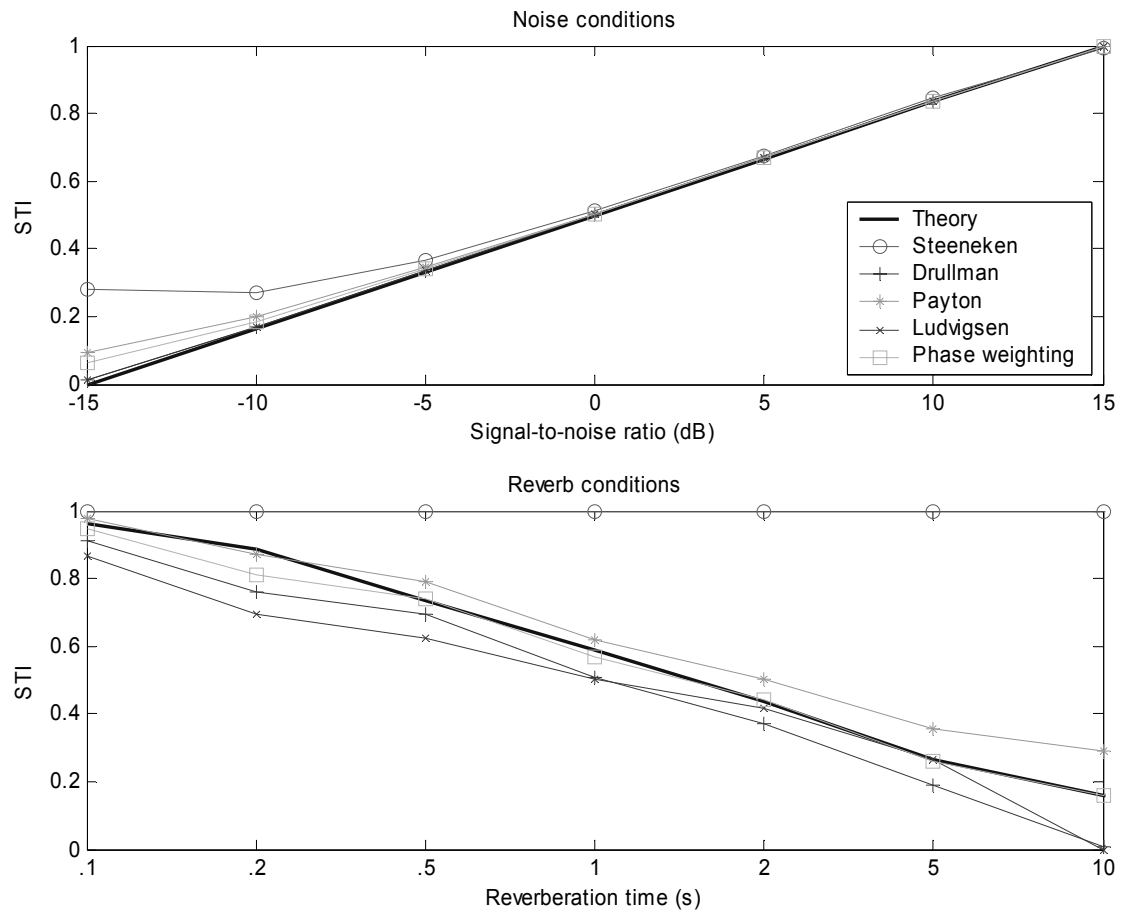


Figure 1. Comparison of speech-based STI methods with theoretically computed STI in conditions with added noise and reverberation.

The resulting STI from the existing sMTF methods and the phase-weighting method were compared with the theoretical value. Figure 1 shows the results of the existing sMTF methods, including the best variant of the phase-weighting method. In noise conditions (top) the methods perform equally well in SNR ratios higher than -5dB. In reverberation conditions (bottom) however, the differences are larger and the phase-weighting method performs the best.

In a second step we used the phase-weighting method to process the NBVC vocoder database. Figure 2 shows that for male speech, although there is an offset, the correlation between the reference curve and the calculated data is high for the older vocoder methods (CVSD, LPC-10 and CELP). The high correlation for the CVSD data can be explained by the fact that CVSD is a general waveform coder instead of a vocoder, and does not affect the temporal fine structure of the signal as much.

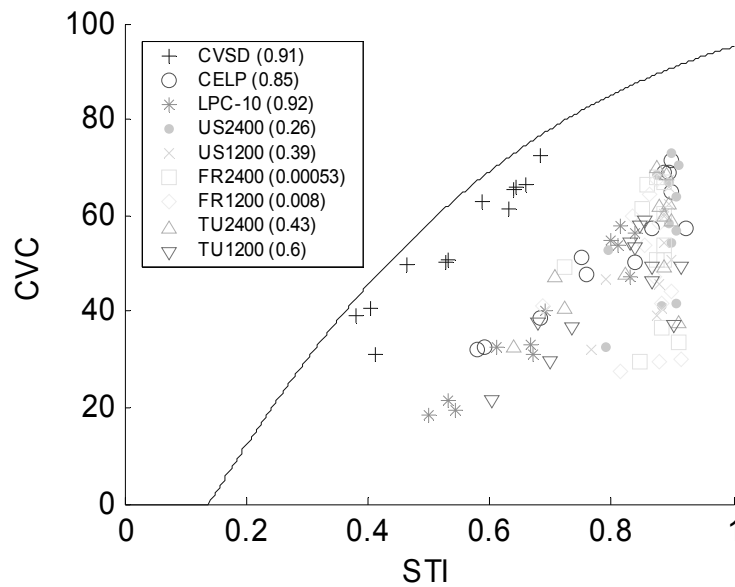


Figure 2. Scatter plot of speech-based STI with corresponding CVC scores. Solid line: STI-CVC reference curve. Numerical values: squared correlation coefficient of each data set.

The results for state-of-the-art vocoders (indicated here by their nation of origin: US, FR and TU) are not as good. Analysis of the results showed that deviation from the reference curve was highly dependent on the SNR at the vocoder input. This result can be explained by the noise suppression accomplished by the vocoders.

Vocoders function as, or explicitly implement, a noise suppression algorithm. It has been known that noise suppression does, at best, provide a modest improvement in speech intelligibility. A reference experiment however, showed that the STI does increase considerably after noise suppression. The noise reduction restores the signal envelope, while the intelligibility remains approximately the same. We conclude that the STI is blind to reduction of intelligibility due to noise addition and suppression.

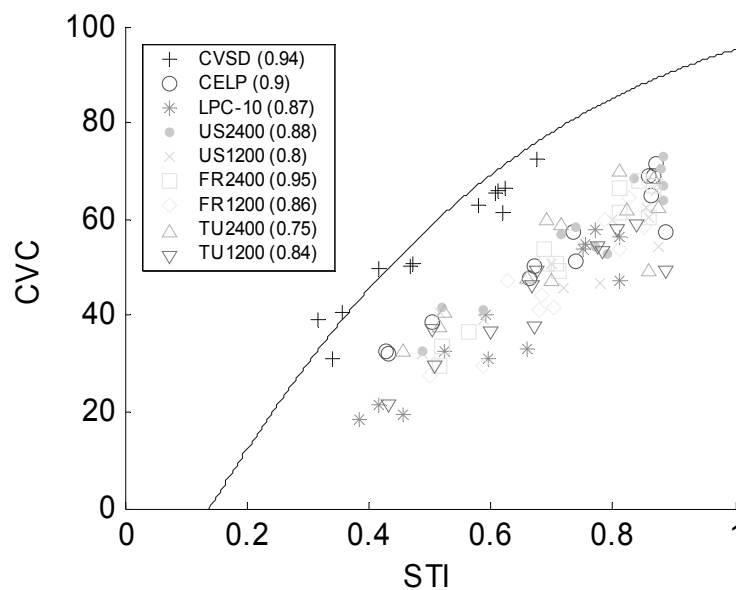


Figure 3. Scatter plot of speech-based STI using noise suppression correction with corresponding CVC scores. Solid line and numerical values as in Figure 2.

The blindness to noise suppression can be brought into the results externally. Figure 3 shows the results for male speech after combining the sMTF between the CVC source and vocoder output and the sMTF of the transfer between CVC source and vocoder input [7]. This procedure improves the correlation of all vocoders greatly as shown in Table 1.

The remaining offset however, seems to be constant among the tested vocoder types, having a value of about 0.3 STI. It may be attributed to the fact that vocoders erode the spectral content of speech fine structure, while the signal envelope remains unaffected. Again speech intelligibility is reduced without any effect noticed in the STI, which is only observing changes in signal envelope.

Not shown in this paper, the same procedure was used to estimate the STI for female speech. These results are less convincing; the squared correlation coefficient varies from .63 to .93. One reason could be that vocoder algorithms are mostly optimized for male speech.

Coder	Squared correlation coefficient R^2	Coder	Squared correlation coefficient R^2
CVSD	0.94	FR 2400	0.86
CELP	0.90	FR 1200	0.95
LPC-10	0.87	TU 2400	0.75
US 2400	0.88	TU 1200	0.84
US 1200	0.80	Average	0.87

Table 1. Squared correlation coefficient R^2 of the data sets and the STI-CVC reference curve.

CONCLUSIONS AND DISCUSSION

A new method was proposed to measure the STI with real speech instead of an artificial test signal. On conventional channels (without vocoders), this procedure yields more accurate results than existing methods to measure the STI with real speech. Initial results obtained with a database of speech processed through vocoders show that the speech-based STI can indeed be used to predict speech intelligibility of vocoders, using a correction for noise suppression effects, but that additional work must be done to improve accuracy.

- [1] Steeneken, H.J.M., Houtgast, T. (1983). The temporal envelope spectrum of speech and its significance in room acoustics. Proceedings 11th ICA Congress, Paris 1983. Vol. 7, 85-88.
- [2] Ludvigsen, C., Elberling, C., Keidser, G. and Poulsen, T. (1990). Prediction of intelligibility of non-linearly processed speech. *Acta Otolaryngol Suppl.*, 469, 190-5.
- [3] Drullman, R., Festen, J.M. and Plomp, R. (1994) Effect of reducing slow temporal modulations on speech reception. *J. Acoust. Soc. Am.*, 95, 2670-80.
- [4] Payton, K.L., Braida, L.D., Chen, S. Rosengard, P. and Goldsworthy, R. (2002) Computing the STI using speech as a probe stimulus. In *Past, Present and Future of the Speech Transmission Index.*, (ed. van Wijngaarden), TNO Human Factors: Soesterberg, the Netherlands, 125-37.
- [5] Houtgast, T., Steeneken, H.J.M. and Plomp, R. (1980) Predicting speech intelligibility in rooms from the modulation transfer function. I. General room acoustics. *Acustica* 46, 59-72.
- [6] Steeneken, H.J.M. (1992) On measuring and predicting speech intelligibility. Soesterberg: TNO Institute for Perception, ISBN 90-6743-209-1.
- [7] van Wijngaarden, S.J. and Verhave, J. Prediction of speech intelligibility for public address systems in traffic tunnels, submitted.
- [8] Tardelli, J.D., van Wijngaarden, S. J., Hassanein, H. and Collura, J.S. (2002) A Precision-weighted Rank-ordering Procedure for the Combination of Voice Coder Evaluation Results. *IEEE Speech Coding Workshop*, Tsukuba, Ibaraki, Japan, October 6-9 2002, pp 99-101

Improved Hearing Protection for Aviation Personnel

Richard L. McKinley¹, Valerie S. Bjorn², and John Allen Hall³

Air Force Research Laboratory^{1,3}
AFRL/HECB, Bldg 441
2610 Seventh Street
Wright-Patterson AFB, OH 45433-7901, USA
Naval Air Systems Command²
AEDC/DOF
740 Fourth Street
Arnold AFB, TN 38389-6000, USA

richard.mckinley@wpafb.af.mil

ABSTRACT

Hearing loss has long been associated with the operation of aircraft. Some of the first hearing protectors were developed for use around military aircraft. Today's high performance military aircraft generate noises which typically range from 110 dB to 150 dB. Normally, the source of the noise cannot be quieted without loss in performance. Therefore hearing protection is the primary tool to mitigate aviation personnel noise exposures during operations of aircraft. This paper describes a joint U.S. Air Force and U.S. Navy approach to improve hearing protection and reduce hearing loss risk. The approach included research and development to improve hearing protection as well as technologies to allow personnel to be moved from high noise work areas; recommendations for administrative controls; and investigation of hearing protective pharmaceuticals. The development of improved passive and active hearing protection technologies employed a three phased approach with attenuation performance goals for near-term (35-40 dB), mid-term (40-45 dB), and long-term (45-50+ dB) solutions. The technologies which have been developed to achieve the first two hearing protection goals will be described along with their attenuation performance characteristics. Ongoing research to achieve the long term (45-50+ dB) goal will be described with considerations of bone conducted noise pathways.

1.0 INTRODUCTION

For over 100 years, aircraft have generated levels of noise sufficient to cause hearing loss in personnel operating and maintaining those aircraft. Pilots, aircraft mechanics, and flight line personnel have long understood the risk of hearing damage when flying and working around aircraft. Early attempts at hearing protection included stuffing cotton or chewing gum into the external ear canal. The first earmuffs were basically jelly jars dipped in rubber and put on a headband. The introduction of jet aircraft brought broadband noise spectra and higher overall noise levels into the personnel noise environment. High performance military aircraft frequently generate levels of noise not normally experienced in the general civilian or industrial workplace. These levels typically range from 140 dB to over 150 dB at the worst case maintenance personnel locations (see Figure 1) and from 110 dB to 120 dB in the cockpit (9) at the pilot location.

McKinley, R.L.; Bjorn, V.S.; Hall, J.A. (2005) Improved Hearing Protection for Aviation Personnel. In *New Directions for Improving Audio Effectiveness* (pp. 13-1 – 13-12). Meeting Proceedings RTO-MP-HFM-123, Paper 13. Neuilly-sur-Seine, France: RTO. Available from: <http://www.rto.nato.int/abstracts.aps>.

Improved Hearing Protection for Aviation Personnel

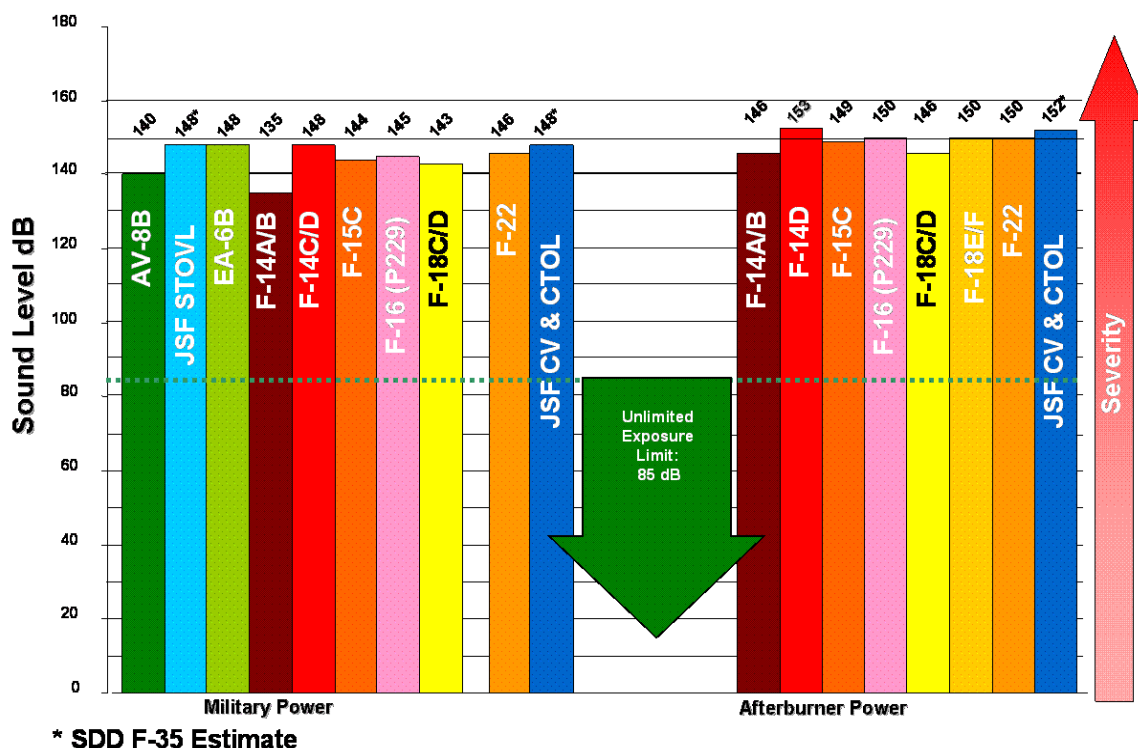


Figure 1: Worst case noise levels at personnel locations of legacy high performance military aircraft

Double hearing protection consisting of earmuffs and earplugs is required to work in these higher near-field personnel noise levels. However, passive hearing protectors available today, even when worn as double hearing protection, do not reduce the highest jet noise enough to prevent noise induced hearing loss. The concept of Active Noise Reduction (ANR), or electronically sensing and transmitting anti-noise to reduce noise in headset earcups, was demonstrated in the late 1950's by Meeker *et al* (12) working for the Radio Corporation of America (RCA) at Wright-Patterson Air Force Base. By the 1980s, the first practical ANR headsets (5) were available. Typically, ANR headsets improve attenuation 10-15 dB in the noise frequencies below 800 Hz and in sound pressure levels below 135 dB. While providing a significant gain in hearing protection, ANR headsets fall short for protecting the hearing of personnel working in noise above 135 dB.

The design of a next generation high performance aircraft usually begins with a new engine. These engines are generally more powerful, more efficient, and unfortunately, frequently produce high noise levels. Concerns over the increasing prevalence of military personnel hearing loss and related disability compensation (see Figure 2) prompted the U.S. Navy and U.S. Air Force to jointly pursue improved hearing protection for aviation personnel – pilots and maintainers – with a focus on the highest noise level environments associated with military aircraft and engine maintenance operations.

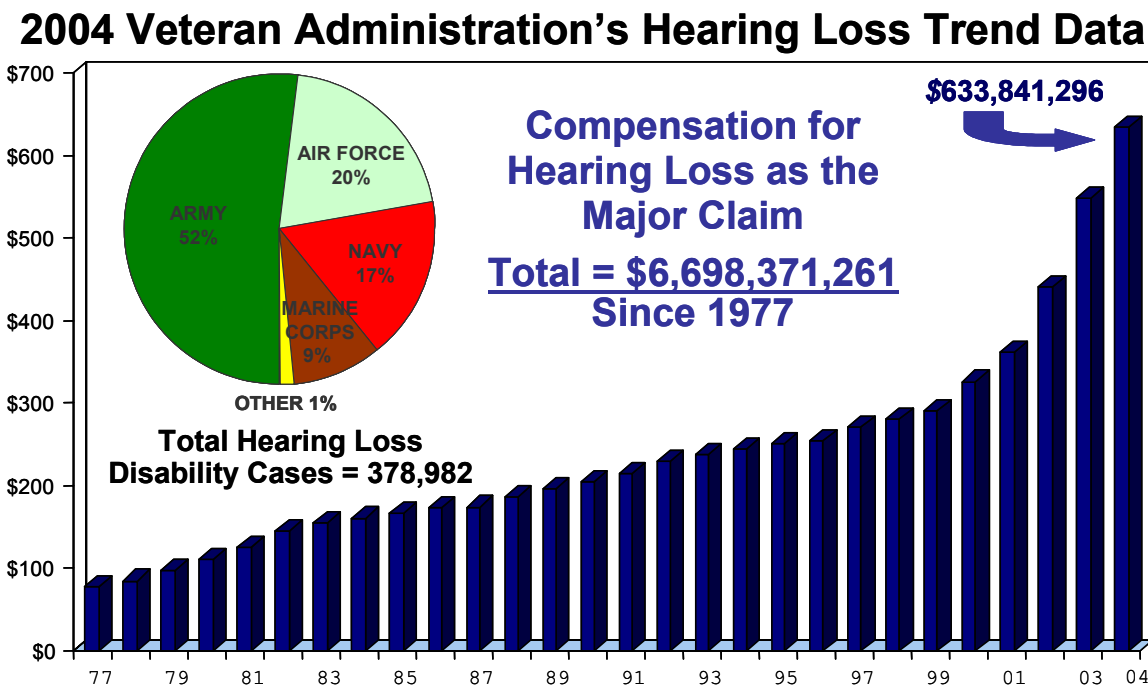


Figure 2: United States Veterans Affairs Compensation Payments for Hearing Loss Disability

2.0 RATIONALE

In 2001, the U.S. Air Force and U.S. Navy developed a joint service 7 year plan to reduce hearing loss in military aviation personnel. In addition to the U.S. Navy and U.S. Air Force, the total effort involved researchers and developers from universities and commercial technology developers as well. Figure 3 provides a schematic of the three-part plan implemented. The first part, improved hearing protection, is the focus of this paper. The goal was to develop a hearing protection system that would protect personnel working in 150 dB environments for up to 15 minutes per day, i.e. a nominal 50 dB hearing protector. Figure 4 shows how 50 dB hearing protection increases safe allowable exposure times. To meet the 50 dB hearing protection goal, near-, mid-, and long-term noise attenuation goals were set: 2004 goal was 35-40 dB; 2006 goal is 40-45 dB; 2008 goal is 45-50+ dB. The second part of the plan, administrative controls, focused on stiffening rules and regulations concerning personnel noise exposure and hearing protector use. The third part, enabling technologies, broadened the approach to include aircraft and ship technologies that could allow personnel to be moved out of noise hazards and to include pharmaceuticals that had a potential to protect against permanent hearing loss. In summary, the 50 dB hearing protector concept would protect most personnel, and the remaining hearing hazards could be addressed by administrative controls and enabling technologies.

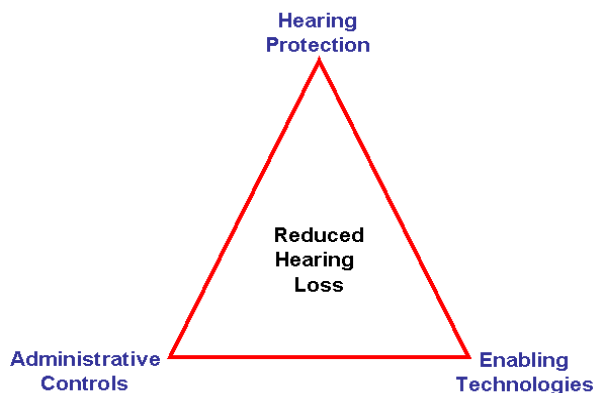


Figure 3: Three Part Plan to Reduce Hearing Loss

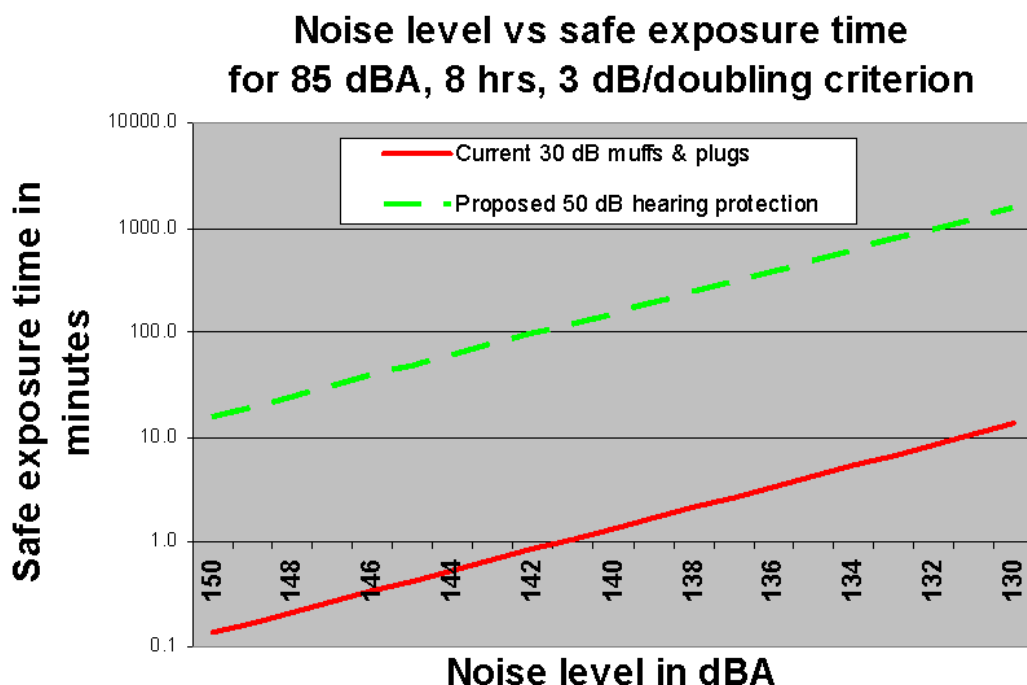


Figure 4: allowable exposure vs noise level with 30 dB and 50 dB hearing protection

Figure 5 shows the hearing protection performance goals, technologies, and time lines for improved aviation personnel hearing protection.

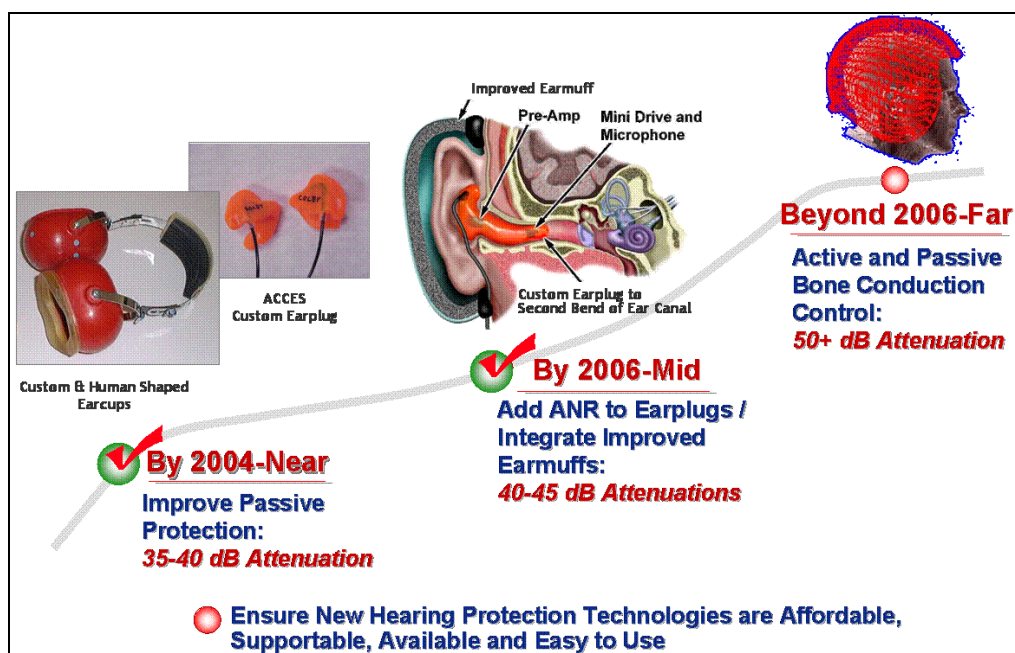


Figure 5: Hearing protection attenuation performance goals

3.0 METHODS

The joint U.S. Air Force – U.S. Navy plan included developing a combination of passive attenuation and active attenuation systems and testing to verify their attenuation performance. They included passive and active noise reduction custom fit earplugs, custom fit earmuffs, custom fit earcushions, cranial helmets, new materials, and new fabrication methods. The technologies were initially applied to existing designs (4, 6, 8, 11, 14, 15, 16, 19, 20, 21) to look at the potential advantages of the new technologies. Then the new technologies were combined in attempt to make more significant attenuation improvements. In all cases, the goal was to minimize the total A-weighted noise exposure of the personnel.

The noise fields at personnel locations were measured and analyzed along with the operational times at each condition. These individual exposures were summed, as shown in equation 1, to arrive at a total daily noise exposure, or TDE, for personnel across a flight line/deck. Typically, it was the highest level noise conditions which were the primary factors in the TDE calculation. Noise contours around an aircraft were plotted, then exposure time computed, and the resulting TDEs were mapped as seen in Figure 6. The black line is the TDE = 1 condition. Therefore, locations outside the black line are TDE < 1 and locations inside the black line are TDE > 1, indicating a daily over-exposure to noise. Personnel locations and TDE calculations were used to determine the amount of hearing protection attenuation needed to protect an individual at a particular location and to meet the allowable noise exposure criterion.

Equation 1.
$$TDE = \sum_{i=1}^n (t_{li} \div 480) \times 2^{(L_{Ali} - 85) / 3}$$

where $i \equiv$ exposure segment

where $n \equiv$ total number exposure segments

where $t_{li} \equiv$ the duration in minutes of the i^{th} exposure segment

and $L_{Ali} \equiv$ A-weighted sound pressure level (dBA) at the ear of the i^{th} exposure segment

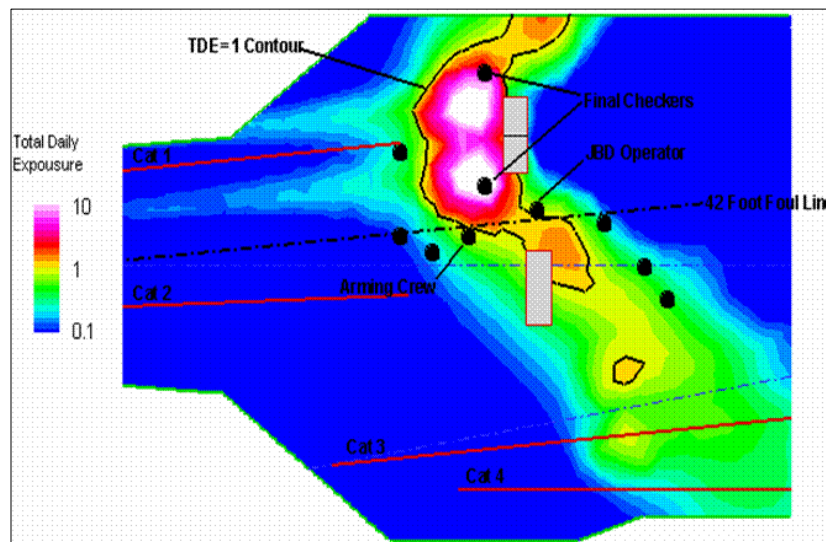


Figure 6: Notional Total Daily noise Exposure (TDE) contours for an aircraft carrier flight deck

Improved Hearing Protection for Aviation Personnel

Additionally, consideration was given to the statistical distribution of hearing protector performance. The mean or average attenuation adequately protects 50% of the personnel. If hearing protector attenuation performance was normally distributed (which it sometimes is) then subtracting 2 times the standard deviation of the attenuation would protect approximately 98% of the population. However, hearing protector performance can be a bimodal distribution, or some other non-normal distribution. To increase the accuracy of an attenuation rating of a hearing protector for a given population (like U.S. military flight line/deck personnel) attenuation performance was measured on either an actual sample from that population or a subject population selected to represent that population for age, gender, and head size.

Hearing protector attenuation testing (1, 2, 10, 17, 18) followed national ANSI and/or international ISO consensus standards. Passive attenuation hearing protectors were tested on subjects using subjective and objective test methods in a 115 dB pink sound field, i.e., 115 dB at all frequencies. The subjective attenuation test method was the Real Ear Attenuation at Threshold, or ANSI S12.6 REAT method. In this method, the subject determined his threshold of hearing at several frequencies with and without the hearing protector. The difference in the two thresholds was the attenuation of the device. This method was used to test passive attenuating earplugs and earmuffs (REAT is not recommended for active noise reduction devices). The objective attenuation test method used was the Miniature Microphone in Real Ear, or ANSI S12.42 MIRE method. In this method, the subject was configured with two microphones, one at the entrance of each ear canal. The subject was placed in the sound field. MIRE data were collected with the subject wearing and then not wearing the hearing protector over the microphones. The difference between the two data sets was the attenuation of the device. The MIRE method was used to test earmuff attenuation. MIRE can not be used to test passive attenuating earplugs but is used to characterize the active attenuation performance of ANR headsets. ANR earplug attenuation test methods are evolving to use the microphone that is integral to the ANR earplug system, since it is not practical to add an independent microphone alongside an ANR earplug, deep inside the ear canal.

The data in this report include REAT data for earmuffs and earplugs, and both REAT and MIRE data for custom earmuff testing. The ANR earplug total attenuation was estimated using REAT data for the passive earmuff and earplug attenuation, and MIRE data from the ANR earplug microphone to estimate the active attenuation. ANR earplug MIRE data presented in this report were measured in the trapped volume between the tympanic membrane and inserted end of earplug tip. This trapped volume was always less than 1 cc and in most cases less than 0.5 cc; for this small volume, a phase shift from the ANR earplug microphone to the tympanic membrane was not considered to be an error factor. This conjecture was supported by subject reports of perceived decreases in the noise level when the ANR earplug system was powered.

4.0 RESULTS AND DISCUSSION

The near term and mid term goals to develop 40 dB and 45 dB overall attenuation devices were met and exceeded. Custom shaped earcups provided a 4.4 dB mean protective attenuation gain over standard, flat earcups (47 dB attenuation at approx. 1 kHz). Passive custom earplugs worn under standard earmuffs resulted in 42 dB mean overall attenuation. Active noise cancelling custom earplugs worn with standard earmuffs resulted in 47 dB mean overall attenuation.

The near term effort focused on improving passive attenuation both in earplugs and in earmuffs. Custom fitting procedures developed and advanced in this program resulted in improved fit and comfort. Custom earplugs and earcups minimized the size of acoustic leaks by improving fit. Figure 7 shows mean attenuation data that compares custom earplugs and earmuffs to expanding foam earplugs and standard flat earcups, both commonly used by U.S. Navy and U.S. Air Force aviation personnel.

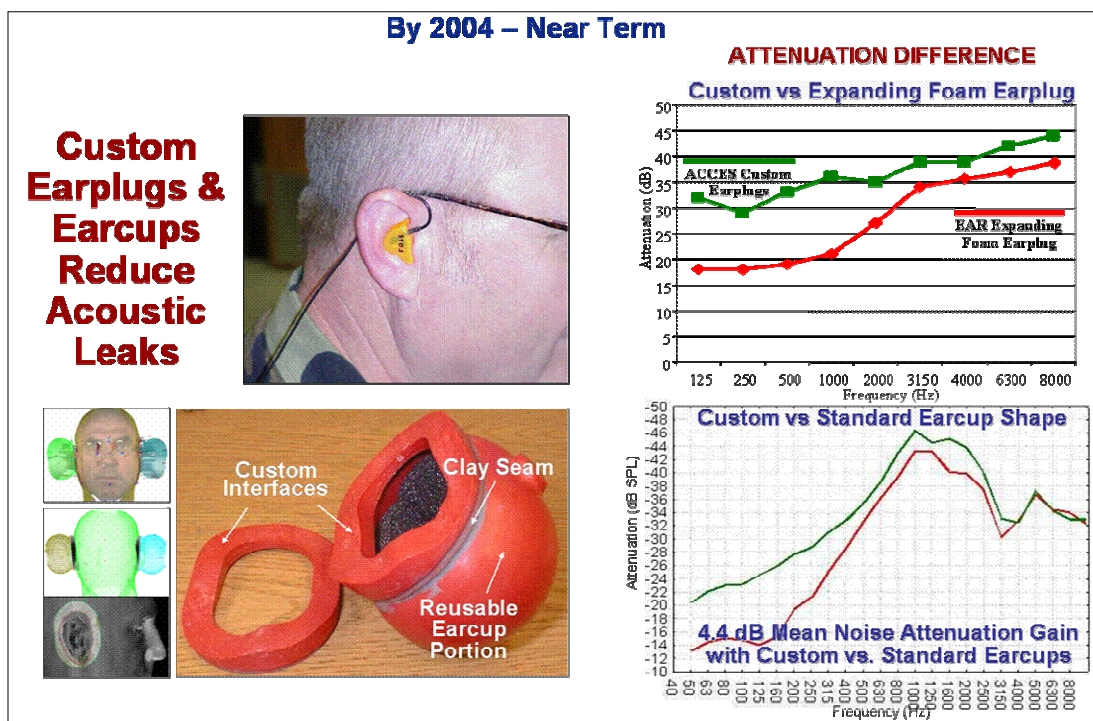


Figure 7: Near term hearing protection attenuation performance and technologies

Customizing earplugs and earcups improved the reliability of fit as evidenced by a reduced standard deviation in attenuation scores compared to currently used earcups (see Figure 8).

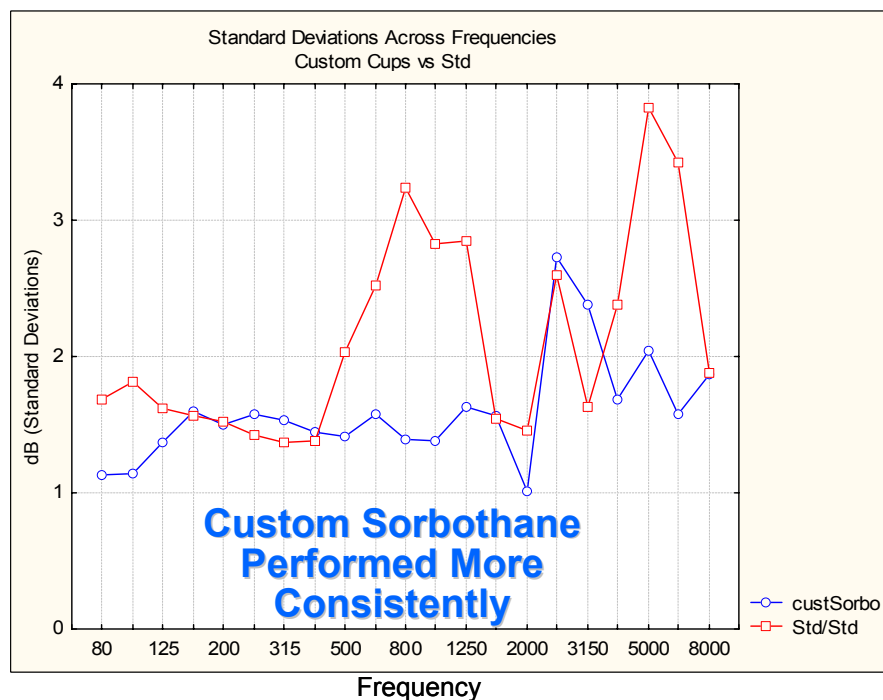


Figure 8: Custom versus Standard Earcup Attenuation Standard Deviations

Improved Hearing Protection for Aviation Personnel

A hearing protector that was comfortable and provides consistent attenuation within and between subjects was desirable for ensuring field use and ensuring field attenuation performance. The current hearing protection consists of a combination of earplugs and earmuffs, typically foam earplugs and a large volume earmuff. The reliability and consistency of fit in the field was an issue(3).

In a recent survey of 301 aviation personnel, Bjorn *et al* (3) found that while 73% reported inspecting their cranial helmets at least daily, 41% of the earcup cushions and/or earcup foam inserts were in poor condition (deteriorated, flat, hard, or missing). Only 14% reported always wearing earplugs beneath their cranials (double hearing protection), while 47% percent reported never wearing earplugs. Of those who used earplugs, only 7% inserted earplugs deeply enough in both left and right ear canals to benefit fully (full attenuation). For subjects who reported wearing earplugs sometimes or always, Figure 9 shows earplug insertion depths and the percentage who achieved each depth. In total, 79% of the ears of flight deck personnel interviewed received an estimated 0-6 dB of noise attenuation from either shallow earplug insertion depths or never wearing earplugs.

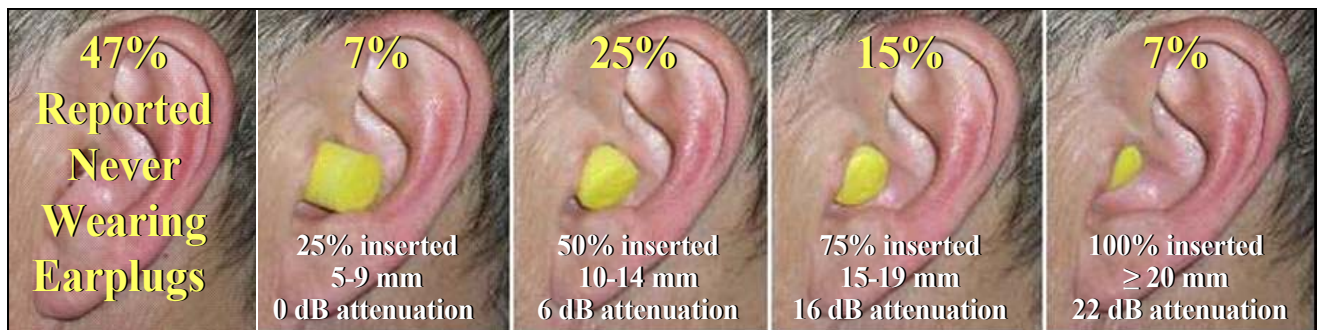


Figure 9: Left and right earplug insertion depths percentage of subjects at each depth

In comparison, Figure 10 shows the results of a limited study of custom earplug fit attenuation reliability and consistency of fit over a 4-6 week period. Once trained, subjects were able to reinsert the custom earplugs without assistance and achieve consistently high attenuation. One factor working in favor of deep insert custom earplugs was that incorrect earplug insertion was very uncomfortable. There was no viable partial insertion option for the custom earplugs; this is a significant difference when compared to almost every other available earplug. This factor, it is believed, will result in more consistent and higher overall attenuation when in actual use by personnel on flight lines and flight decks.

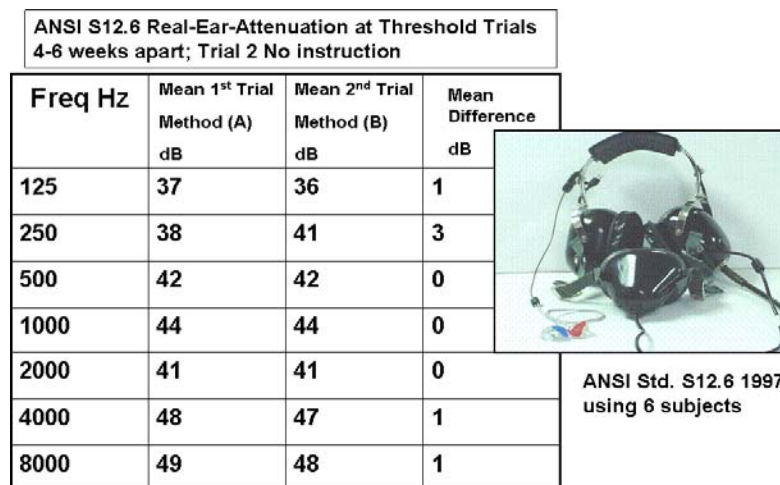


Figure 10: Custom deep insert earplug consistency and reliability of fit

With custom earplugs and earcups, double protection is still required. Either or both of these near-term technologies could be immediately fielded. The custom earmuff would result in an approximate 4-5 dB improvement in attenuation performance while the custom deep insert earplug would result in an approximate 10 dB improvement in attenuation performance.

Active noise reduction (ANR), i.e. active noise cancellation in a small volume, has been available in headsets and earmuffs since the mid 1980s. In this current U.S. Air Force and U.S. Navy program, the mid-term effort to improve hearing protection for aviation personnel was to add ANR to deep insert custom earplugs. The best available ANR headsets were able to actively cancel noises up to approximately 135 dB depending on the noise spectrum. With any ANR system, a limit is determined by the maximum excursion of the driver generating the canceling noise. The amplitude of the canceling noise needs to be equal and nearly 180 degrees out-of-phase to the insulting noise. Simply, there were no drivers available to generate the necessary 135 dB or greater canceling noise in a large volume earcup. However, the earplug approach offered a number of advantages. First, the driver in the earplug had to move a much smaller volume of air, typically less than 0.5cc. Second, the amplitude of the insulting noise to be cancelled had been reduced by the passive attenuation of both the earmuff and earplug, again reducing the excursion requirements on the driver. Finally, due to the small volume and associated small time delay, the maximum frequency of active attenuation would increase from the typical 800 Hz for an ANR headset to almost 3 kHz in the system shown in Figure 11.

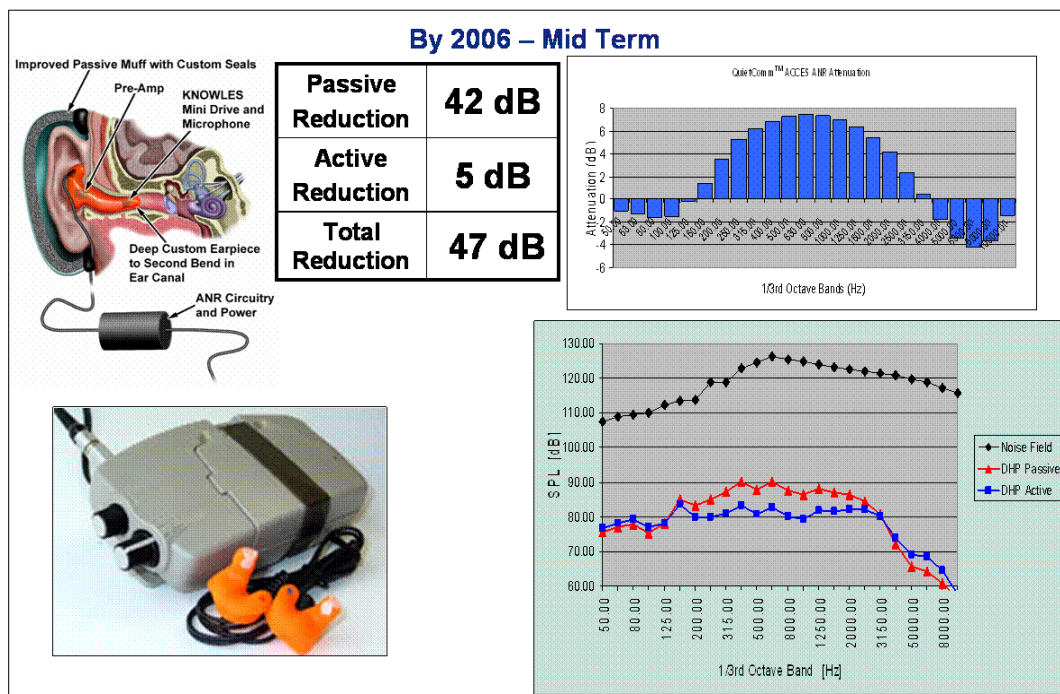


Figure 11: Active noise reduction earplug passive and active attenuation performance

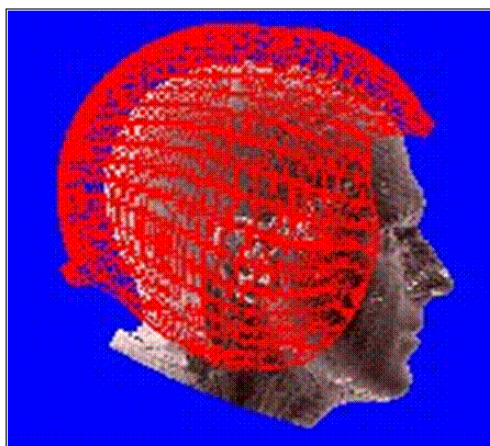
An important point should be considered when designing any ANR system. The overall goal should be to reduce the A-weighted noise level at the tympanic membrane in the noise spectra in which the device is expected to be used. This may mean giving up large, low frequency active attenuation performance to gain modest mid-frequency active attenuation performance that more directly translates to reductions of the overall A-weighted sound pressure level at the ear and correlates to more accurate speech communications.

Improved Hearing Protection for Aviation Personnel

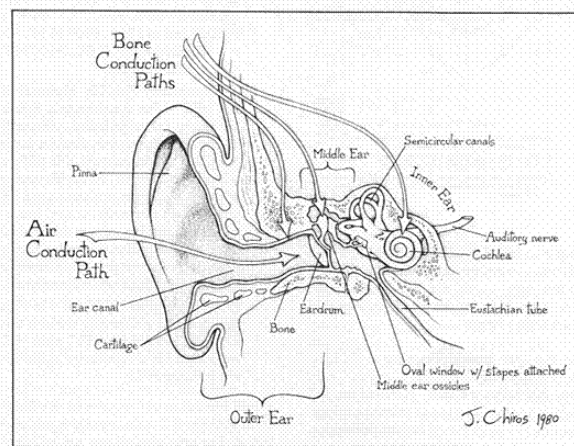
Once noise has been sufficiently attenuated via the ear canal, other noise pathways that flank the ear canal to arrive at the cochlea become the next technology challenge for hearing protection developers. The bone conduction limits have been described by many authors including Nixon, von Gierke, and Berger (7, 13). The consensus of these articles was that once 43-45 dB attenuation is achieved, bone and tissue noise pathways predominate at 2 kHz. Bone and tissue conduction were a factor at other frequencies but at greater attenuations.

Clearly in 150 dB noise environments, where 50 dB or more attenuation is required, bone conduction becomes an important issue. Two concepts have been conceived in an attempt to deal with the flanking paths provided by bone/tissue conduction. The first approach was a brute force approach with multiple sensors and drivers as depicted in Figure 12, left insert. The goal was to minimize the bone/tissue vibrations mechanically. The second approach (Figure 12, right insert) was to use an air conducted signal to cancel the bone/tissue conducted noise. This assumed the cochlea could not discriminate the source of vibrations reaching it; therefore, the cancellation signal for the bone/tissue conducted noise could be presented via an earphone. Some initial progress toward this goal of actively cancelling bone conducted noise has been made.

Active Vibration Control Cranial with ANR Earplug



Active Control/Cancellation of Bone Conducted/Flanking Path Noise



Basic Anatomy of the Ear with Illustration of the Air Conduction and Bone Conduction Sound Paths.

Figure 12: Active reduction of bone conducted noise concept

Measurements of noise levels present in the ear canal were made in high (>130 dB) noise fields. Also, noise measurements were made in the ear canal of noise reradiated in the ear canal from a bone conduction driver on the head of the subject. In both cases the noise levels present in the ear canal were sufficient to be potentially useful in an active noise reduction/cancellation system. Preliminary studies have shown that pure tone noise generated by a bone conduction driver can be actively cancelled by an air conducted cancellation signal. In this study, the phase and amplitude of the cancellation signal was controlled by the subject and the frequency was constant.

In theory, active cancellation of bone conducted noise would allow development of hearing protectors delivering virtually as much attenuation as required. To achieve this goal, an improved understanding of physical acoustics of the head and of the psychoacoustics of bone/tissue conducted noise is required.

5.0 CONCLUSIONS

Hearing loss has been and will continue to be a major disability associated with the operation of high performance aircraft. The joint U.S. Air Force – U.S. Navy program to reduce aviation personnel hearing loss has fostered new research and development in both passive and active noise attenuation resulting in significant improvements in attenuation performance with a 47 dB attenuation demonstration. The first two goals of the joint U.S. Navy – U.S. Air Force program have been met and work continues toward the development of a 50+ dB hearing protector. At 50+ dB, bone conduction is the predominant noise pathway at some frequencies. However, significant research and development in mitigating the bone conducted noise is being accomplished. The near-term technologies developed under this research and development effort are moving to operational use and should have a significant impact on reducing the incidence and severity of noise-induced hearing loss in aviation personnel.



Figure 12: Part of the joint U.S. Air Force – U.S. Navy improved hearing protection team

6.0 BIBLIOGRAPHY

1. Berger, E. H., Review and Tutorial - *Methods of Measuring the Attenuation of Hearing Protection Devices*, J. Acoust. Soc. Am. 79(6), pages 1655-1687, 1986
2. Berger, E. H., *International Activities in the Use, Standardization, and Regulation of Hearing Protection*, J. Acoust. Soc. Am. 99(4), Pt.2, page 2463, 1996
3. Bjorn, V. S., Albery, C. B., Shilling, R. D., *U.S. Navy Flight Deck Hearing Protection Use Trends: Survey Results*, NATO RTO HFM-123 Symposium Proceedings, "Directions for Improving Audio Effectiveness", Amersfoort, The Netherlands, April 2005.
4. Blackstock, D.T. and von Gierke, H.E., *Development of an Extra Small and Extra Large Size for the V-51 R Earplug*, Wright Air Development Center Technical Report, WADC 56-142, Wright-Patterson Air Force Base, Ohio, April 1956
5. Carter, J., *Active noise reduction*, AFAMRL-TR-84-008, 1984
6. Dancer, A., Buck, K., Hamery, P., and Parmentier, G., *Hearing Protection in the Military Environment*, Noise & Health 5, 1-15, 1999

Improved Hearing Protection for Aviation Personnel

7. Franke, E.K., von Gierke, H.E., and von Wittern, W.W., *The Jaw Motions Relative to the Skull and Their Influence on Hearing by Bone Conduction*, Aero Medical Laboratory Technical Report number 6466, Wright Air Development Center, Wright-Patterson Air Force Base, Ohio, May 1951
8. Guild, E., Nixon, C.W., and Baker, D.J., *User Test of Flents Anti-Noise Ear Stopples*, Internal Memo, Wright Air Development Center, Wright-Patterson Air Force Base, Ohio, May 1959
9. James, S., Hancock, M., Hazel, A., and Rood, G. M., *Primary Contributors to Hearing Damage Risk in the Military Aircraft Cockpit*, Intl. Military Noise Conf., Baltimore, MD, 2001
10. Johnson, D.L., and Nixon, C.W., *Simplified Methods for Estimating Hearing Protector Performance*, Journal of Sound and Vibration, June 1974
11. Johnson, D. L., Nixon, C. W., and Skelton, M., *The Effect of Reduced Headband Force on the Attenuation of Muff-Type Protectors*, J. Acoust. Soc. Am. Suppl. 1, 74, S94, 1983
12. Meeker, W., *Active ear defender systems: component considerations and theory*, WADC TR 57-368, 1958
13. Nixon, C. W. and von Gierke, H. E., *Experiments on the Bone-Conduction Threshold in a Free Sound Field*, J. Acoust. Soc. Am. 31(8), pages 1121-1125, 1959
14. Nixon, C. W. and Knoblach, W. C., *Hearing Protection of Earmuffs Worn Over Eyeglasses*, Aerospace Medical Research Laboratory, Report No. AMRL-TR-74-61, Wright-Patterson AFB, OH, 1974
15. Nixon, C.W., Chapter 5 *Hearing Protector Standards*, in *Personal Hearing Protection In Industry*, edited by Alberti, P.W., Raven Press, New York, 1982
16. Nixon, C.W., Chapter 12 *Hearing Protective Devices: Ear Protectors*, in *Handbook of Noise Control*, edited by Harris, C., McGraw-Hill, 1979
17. Parmentier, G.; Buck, K.; Kronenberger, G., and Beck, C., *Artificial Head (ATF) for Evaluation of Hearing Protectors*, Institut Franco-Allemand De Recherches De St. Louis, St Louis, France, 1999
18. Rood, G. M., *The In-Situ Measurement of the Attenuation of Hearing Protectors by the Use of Miniature Microphones*, in *Personal Hearing Protection in Industry*, edited by P. W. Alberti, Raven Press, New York, NY, pages 175-197, 1982
19. Shaw, E.A.G. and Thiessen, G.J., *Improved Cushion for Ear Defenders*, J. Acoust. Soc. Am., Vol 30, Num 1, pages 24-36, Jan 1958
20. Shaw, E.A.G., *Hearing Protector Design Concepts and Performance Limitations*, International Symposium on Personal Hearing Protection, University of Toronto, Canada, May 1980
21. von Gierke, H. E., *Protection of the Ear from Noise: Limiting Factors*, J. Acoust. Soc. Am. 95(5), Pt. 2, page 2914, 1994

Hearing Protection for Bone-Conducted Sound

Anthony J. Dietz, Ph.D.

B. Scott May, Ph.D., Darin A. Knaus, Ph.D., Harold P. Greeley

Creare Incorporated

PO Box 71, Hanover, NH 03755, USA

ajd@creare.com

ABSTRACT

The high-noise environments in and around modern military jet aircraft can impair voice communications and cause permanent damage to the hearing of pilots and maintenance crews if adequate hearing protection is not worn. With external noise levels of some aircraft approaching 150 dBA, adequate hearing protection must provide better than 50 dB of attenuation. In order to reach such a high level of attenuation, sound-transmission routes that bypass the ear canal must be attenuated in addition to the ear canal transmission routes. These bone- and tissue-conduction paths limit the maximum attenuation that can be achieved, even if all air-conducted sound in the ear canal were eliminated.

An investigation of bypass sound transmission mechanisms is presented. The aim of this investigation was to characterise the transmission levels of bypass sound to the cochlea and to determine the effectiveness of different hearing protection components in attenuating this sound. Measurements were made with a head simulator as well as with human subjects.

In order to make detailed measurements of the vibration of skull bones in an intense sound field, we developed a head simulator. The simulator was built from a human skull, with silicon gel used to model internal organs and silicon and latex used to model the skin. Accelerometers attached to the skull bones were used to measure the skull vibration response. Mechanical point impedance measurements on the simulator were compared to results reported in the literature for humans, cadavers, and skulls. Reasonable agreement with these results served to validate the simulation. The relationship between skull vibration measured on the skull simulator and bone-conducted sound at the cochlea of a human subject was determined by comparing mechanical point impedance measurements recorded from the skull simulator and from human subjects, with the latter including psycho-acoustic responses. The effect of different hearing protection components on skull vibration and on the attenuation perceived by human subjects was also measured. Finally, the attenuation limits for bypass sound routes through the head and through the torso were determined.

The response of the head simulator in a sound field peaked in the 1 to 3 kHz frequency range, which is where the bone-conducted limit to hearing protection attenuation is at a minimum. The bypass sound levels measured on human subjects also increased in this frequency range. The performance of different hearing protection components including ear muffs, helmets, goggles, and face shields was measured on the head simulator and on human subjects, and the results are reported. General guidance concerning the design of devices for providing protection against bypass sound transmission is given. Finally, the need for additional protection of the torso is discussed.

Dietz, A.J.; May, B.S.; Knaus, D.A.; Greeley, H.P. (2005) Hearing Protection for Bone-Conducted Sound. In *New Directions for Improving Audio Effectiveness* (pp. 14-1 – 14-18). Meeting Proceedings RTO-MP-HFM-123, Paper 14. Neuilly-sur-Seine, France: RTO.
Available from: <http://www.rto.nato.int/abstracts.apc>.

1.0 INTRODUCTION

In situations where personnel are required to work in extreme noise environments, the protection provided by hearing protection devices (HPDs) is often insufficient. Even if the sound transmitted to the ear canal was completely attenuated, the sound reaching the cochlea via bone conduction transmission paths that bypass the HPD can still be loud enough to cause hearing damage. Improved understanding of these bypass sound transmission paths will lead to improved hearing protection, allowing safe exposure to high sound levels.

1.1 Extreme Noise Environments

Military jet aircraft operating at full afterburner can produce sound levels in the vicinity of 150 dBA. Crews working in close proximity to these aircraft need hearing protection that allows them to complete a work shift without exceeding the allowable exposure limit. With a hearing conservation limit of 85 dBA for eight hours with a 3 dB exchange rate and assuming sixty, thirty-second takeoff or landing events in a work shift, the allowable noise exposure during each event is 97 dBA. For events approaching sound levels of 150 dBA, hearing protection providing attenuation in excess of 50 dB is required.

1.2 Hearing Protection Limits

Bone conduction transmission paths limit the maximum attenuation that can be achieved with double hearing protection (earplugs and earmuffs). For example, in the 2 kHz octave band, Real Ear Attenuation at Threshold (REAT) measurements by Berger et al. (2003) gave a mean attenuation of 37 dB for deeply inserted foam earplugs and 36 dB for earmuffs. However, the attenuation measured with both earmuffs and earplugs in place was only 41 dB, far less than would be expected from the combined component attenuation values. The attenuation was limited by transmission paths that bypassed the hearing protection devices. When a third level of protection was introduced in the form of a helmet, the total attenuation increased to 49 dB.

1.3 Bone Conduction Transmission Paths

The various transmission paths that bypass the attenuation provided by double hearing protection devices are collectively known as bone conduction transmission paths. Four primary bone conduction transmission paths have been identified in the numerous research studies on this phenomenon (e.g., Tonndorf et al. 1966; Sohmer et al. 2000). These are: (1) inertial movement of the ossicular bones relative to the skull at low frequencies, (2) distortion of the temporal bone and cochlear shell at high frequencies, (3) osseotympanic transmission of sound radiated from the walls of an occluded ear canal, and (4) sound conduction via fluid pathways connecting the cochlea to the brain cerebrospinal fluid. The resultant sound level at the cochlea is a frequency-dependent vector sum of the contributions from each of these transmission mechanisms.

A graphical representation of the various transmission paths is shown in Figure 1. The ability of hearing protection devices to attenuate the sound level at the cochlea will depend on their effectiveness against each of these transmission paths. Earplugs are effective against air-conducted sound transmission and against bone-conducted sound transmitted to the ear canal. Earmuffs are effective against air-conducted sound and provide limited shielding of the head. A helmet further shields the head, attenuating both air- and bone-conducted sound. Finally, even if the head is completely shielded, a body conduction path remains. In order to design effective hearing protection for extreme noise environments, the relative limits between the air, head, and body conduction transmission paths must be known.

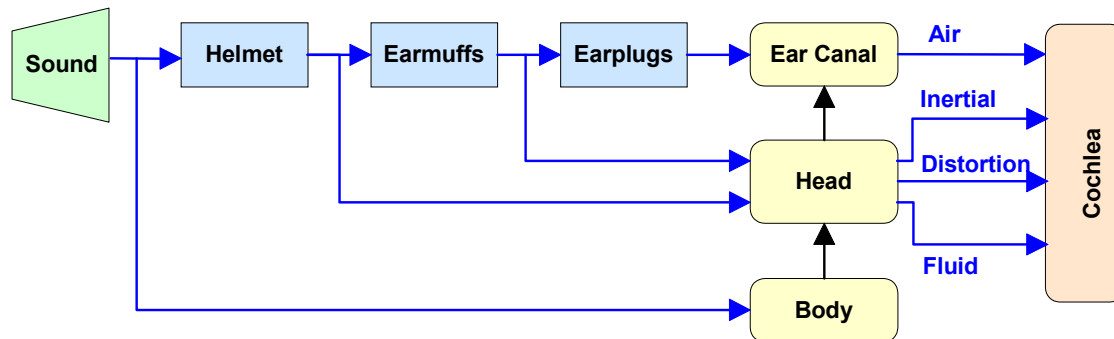


Figure 1. Transmission of Sound to the Cochlea.

1.4 Previous Studies

Previous studies have focused on the vibration response of the head to mechanical driving point forcing and on the psycho-acoustic perception of the bone-conducted sound generated by such forcing. Studies of the vibration response of the head have primarily reported impedance measurements. Smith and Suggs (1976) measured the mechanical impedance to point forcing of human heads. Hakansson et al. (1986) measured the mechanical impedance at the mastoid with and without skin penetration, and Stenfelt and Hakansson (1999) measured the impedance at the teeth. Impedance measurements have also been made on monkeys and cadavers (Stalnaker et al. 1971) and on dry skulls (Stenfelt et al. 2000). There are significant differences in both the magnitude and frequency dependence of the impedance data reported. No studies of the vibration response of a head in a sound field were located.

A number of studies on bone-conducted sound have been concerned with identifying the relative thresholds or limits of bypass sound transmission through the head and body. Berger et al. (2003) compiled an extensive review of studies on the limits of bone conduction through the head and made further measurements of the attenuation provided by deeply inserted earplugs and by earmuffs and a helmet. Ravicz and Melcher (2001) made additional estimates of the body conduction threshold over a limited frequency range.

The aim of the work reported here was to build on the current understanding of the bone conduction transmission paths and to evaluate the effectiveness of different hearing protection components in providing protection against bone-conducted sound.

2.0 METHODS

2.2 Head Simulator

2.2.1 Need for the Simulator

Measurements of the bone conduction transmission paths in human subjects are limited by the need to make non-invasive measurements and by the difficulty in separating psycho-acoustic responses to air- and bone-conducted sound. Studies on animal models and cadavers have a limited test window that hinders detailed long-term investigations of component performance. In order to overcome these limitations, we developed an instrumented head simulator that reproduced the dynamics of a human head. We validated the simulator

Hearing Protection for Bone-Conducted Sound

against the human and cadaver mechanical impedance data reported in the literature. We then measured the response of the head to a sound field. Once validated, the simulator provided a valuable test platform for repeat measurements of component effectiveness.

2.2.2 Simulator Construction

The head simulator was built up from a dry human skull. Stages of the build processes are shown in Figure 2. A triaxial accelerometer was bolted onto the right mastoid bone, and uni-axial accelerometers were bolted onto the left mastoid and also onto the left and right interior temporal bones. The mastoid accelerometers were aligned in the transverse direction, and the temporal bone accelerometers were aligned in the axial direction. The skull bones were knitted together using a low-viscosity epoxy (Gougeon WEST System #105 Resin & #206 Hardener). The brain, eyes, nose, and tongue organs and some jaw and neck muscles were simulated by balloons filled with Silicone Dielectric Gel (Dow Corning 3-4150), which set with a consistency and density similar to that of the organs being simulated. The silicon replaced gelatine, which was used in earlier models but began to decompose over time. The jaw was tied into place with rubber bands, and a foam pad was placed between the teeth. The skull was covered with 5-mm thick silicone skin tissue, formulated according to ANSI S12.42-1995. The skin was held in place and tensioned with elastic cloth tape. The external skin layer was simulated by two layers of balloons, stretched tightly over the head. The skull was then inverted and filled with silicon oil (Dow Corning 510) to simulate the cerebrospinal fluid (CSF). Earlier versions of the simulator had used water, and then a dilute solution of hydrogen peroxide, to simulate the CSF, but these solutions caused the bone and other materials to decompose over time. Before making this substitution, we determined the difference between the dynamics of water-soaked bone samples and those of bone samples soaked in silicon oil. Our results showed that the water softened the bone, lowering its natural frequency by 10% compared to dry bone. This result agrees with that reported by Khalil and Viano (1979). The natural frequency of the silicon-soaked bone was 5% lower than that of dry bone, a shift that can be attributed to the increased mass of the soaked sample. As these differences are within the skull-to-skull variability that might be expected, we chose to use the silicon fluid because of its superior preservative and dielectric properties, which significantly increased the shelf life of the skull simulator.



Figure 2. Creare Head Simulator. Left: dry skull with accelerometers mounted on the mastoid bones and on the interior temporal bones. Center: jaw added and organs simulated by latex balloons filled with silicone. Right: tissue simulated by silicone, muscle tension by elastic tape, skin simulated by balloons, remaining cavities filled with silicone oil to simulate cerebrospinal fluid.

2.2.3 Driving Point Impedance Tests

Driving point impedance tests were conducted using a B&K Type 4190 shaker with an impedance head attached (Type 8000). The skull was held against the impedance head with a 2 to 6 lb. force provided by a latex band as shown in Figure 3. Signals from the impedance head (force and acceleration) and from the skull-mounted accelerometers were recorded with the skull driven at the forehead and at the mastoid. Impedance was expressed as velocity divided by force, with velocity determined from the measured accelerations assuming a sinusoidal response. The application force had a minor effect (less than 10%) on the measured driving point impedance and a negligible effect on the response of the bone-anchored accelerometers.

2.2.4 Tests in the Sound Field

The response of the head in a sound field was measured in Creare's acoustic test facility, a 9' by 7' by 7' hemi-anechoic chamber. The head simulator was inverted and suspended in front of a single speaker as shown in Figure 4. The sound field in the head mounting location was carefully mapped and found to be uniform to within 2 dB over a frequency range from 250 Hz to 10 kHz. The response was expressed as acceleration divided by sound pressure level. The latter was measured by a reference microphone positioned adjacent to the head. The head simulator was then fitted with various hearing protection components, and the vibration attenuation was determined by comparing the measured response to that recorded for the bare head.



Figure 3. Driving Point Impedance Test Setup. The head simulator is held against a shaker using latex bands.



Figure 4. Head Simulator Mounted in the Acoustic Test Facility. Simulator is inverted and suspended approximately 1 m from the sound source.

2.3 Human Subject Tests

Human subject tests were carried out with volunteer subjects. All test protocols were reviewed by the local Dartmouth-Hitchcock Medical Center Institutional Review Board. Subjects were healthy males and females between 18 and 50 years of age with normal hearing (less than 15 dBA hearing loss for the frequency range 500 Hz to 6000 Hz) and a normal otoscopic exam.

2.3.1 Human Subject Driving Point Impedance Tests

Driving point impedance tests similar to those conducted on the head simulator were also conducted with human subjects using a B-70 bone oscillator in series with the B&K Type 8000 impedance head. Two subjects were tested and similar responses were obtained from each.

2.3.2 Human Subject Tests in the Sound Field

HPD component performance was measured using a modified Real Ear Attenuation at Threshold (REAT) technique. The technique is defined by ANSI Standard S12.6-1997. HPD attenuation is determined from hearing threshold measurements with and without the hearing protection in place. Pulses of third-octave noise are presented to the subject in a diffuse sound field. The modified technique employed in this study used a directional hemi-anechoic sound field instead of a diffuse field. This was primarily due to the facility available, but it did allow the directional performance of the HPD to be determined. Subjects were fitted with deeply inserted “Grande™” earplugs from E.A.R for the first level of protection. The helmet included Howard Leight Industries’ “Thunder 29” earmuffs for the second level of protection. The helmet shell, fitted with an edge seal, inner liner, and face shield, provided the third level of protection.

2.3.3 Head Isolation Tests

The goal of the head isolation test apparatus was to expose a subject’s body to sound, while isolating their head and ears from this sound. This was done by fitting the subject with triple hearing protection (earplugs, earmuffs, and a sound-attenuating helmet) and then placing them in a sound-attenuating box with their head protruding, as shown in Figure 5. A speaker capable of producing 130 dB SPL was mounted inside the box. Before fitting the subject in the box, the attenuation provided by the triple hearing protection was measured in the sound room using the modified REAT technique described above. The attenuation provided by the box was measured directly by subtracting the average sound level from a sound field survey inside the box from the sound level measured by a reference microphone adjacent to the subject’s head outside the box. The perceived attenuation of the sound presented inside the box was then measured by a threshold test. A valid measurement of the body-conducted sound threshold is obtained if the attenuation of sound perceived by the subject was less than the sum of that provided by the box and the helmet together. If this is the case, the perceived sound must be coming via a body-conducted sound route, as the attenuation is less than the measured bypass route through the box and the HPD.

3.0 RESULTS

3.1 Head Simulator Validation

The fidelity of the head simulator was validated by comparing driving point impedance measurements from the forehead and mastoid to similar measurements on human subjects and to results reported in the literature. The results, shown in Figure 6, are in very good agreement, indicating that the simulator responds in a similar manner to a human head. The forehead impedance measurements were made using a B-70 bone oscillator, with the same apparatus being used for both the head simulator and for human subjects. Other measurements on the head simulator used a B&K shaker, which provided a larger excitation. The results compare well with the non-skin penetrating data of Hakansson et al. (1986), who made measurements on the mastoid of human subjects and with the impedance specified in the standard for artificial mastoids (IEC 373:1990). The data are similar in form but slightly lower than those of Smith and Suggs (1976), who obtained non-skin penetrating impedance measurements for both frontal and occipital locations.

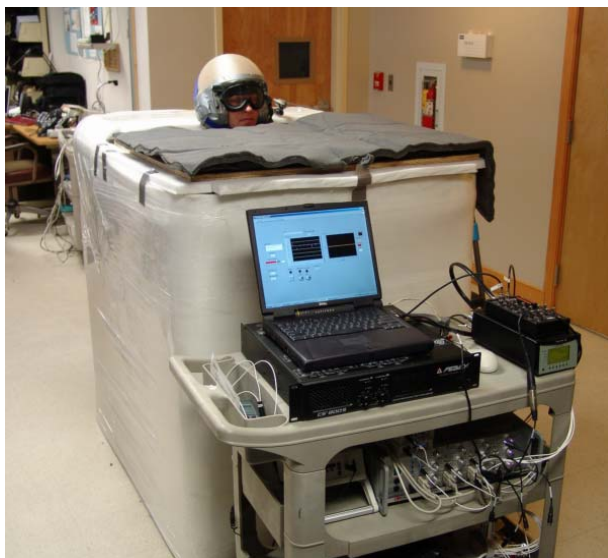


Figure 5. Head Isolation Experiment Setup.

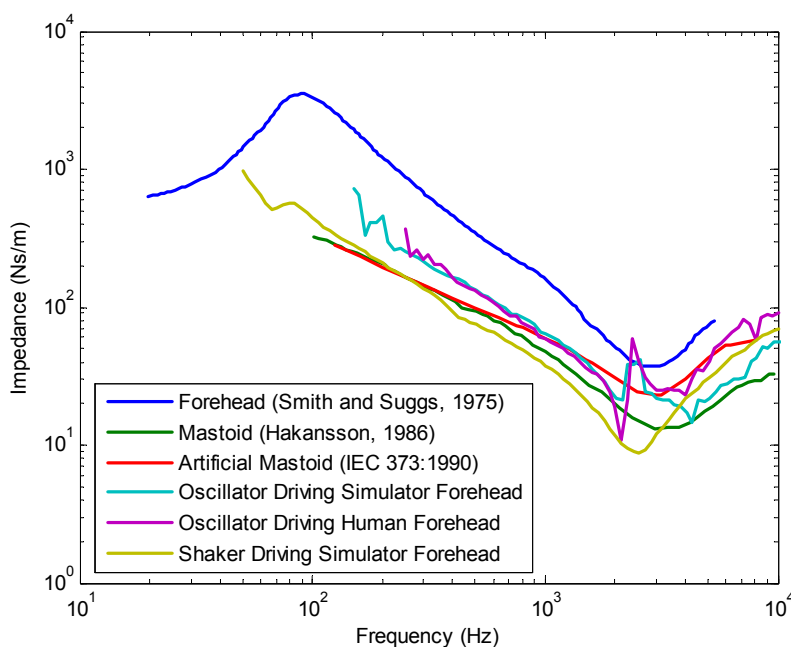


Figure 6. Driving Point Impedance Data from the Head Simulator Compared with Data from the Literature. Force and acceleration are measured at the driving point.

Impedance results from the bone-anchored accelerometers mounted in the mastoid of the head simulator are shown in Figure 7. The measurements by Stalnaker et al. (1971) from a bone-anchored accelerometer mounted in a cadaver mastoid are also included in the figure, as are measurements by Hakansson et al. (1986) from an accelerometer fixed to a titanium implant for a bone-anchored hearing aid. The impedance measured at the bone-anchored sites is higher than that measured at the skin, and there is significant variability between

studies and between measurement sites within a study. Both the location and the magnitude of the anti-resonance and the resonance were dependent on the forcing and measurement locations. However, the general agreement between the form and magnitude of the response from the head simulator compared with that presented in the literature provides confidence that the simulator has a similar response to that of a human head.

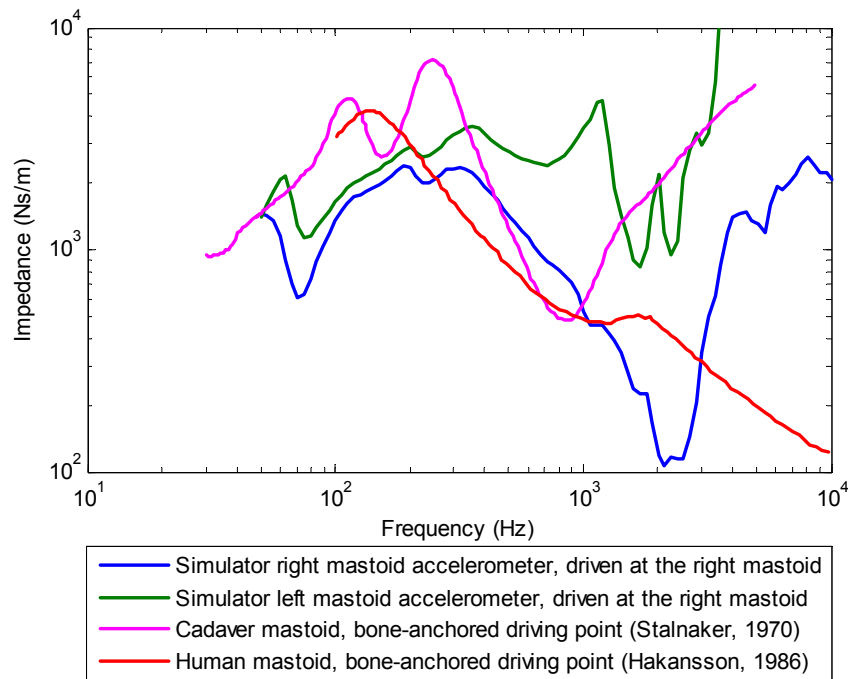


Figure 7. Driving Point Impedance Data from the Head Simulator Compared with Data from the Literature. Driven at the mastoid and acceleration measured with bone-anchored accelerometers.

3.2 Driving Point Response

The modal behaviour of the head can be understood by considering the magnitude and phase of the driving point impedance measurements. The response of the dry skull and of the reconstituted head simulator when driven at the right mastoid is shown in Figure 8. The mastoid bone accelerometers were mounted facing outwards, opposing each other. Below the first resonant frequency, the skull moves as a rigid body, and the phase difference between the z-axis of the right mastoid accelerometer (tri-axial) and the signal from the left mastoid accelerometer (uni-axial) is approximately 180 degrees. At the frequency of the first anti-resonance of the dry skull (800 Hz), the phase difference moves to zero, indicating the bones are vibrating in opposite directions, and the skull is entering the first compression mode. The dry skull resonance occurs at 1000 Hz and is followed by more modes and complicated phase transitions at higher frequencies. The largest response (impedance minimum) occurs at 2 kHz. The increased damping present in the head simulator smoothes out the sharp phase transitions seen in the dry skull data, but the same general trends are apparent. The low-frequency response of the skull and head are similar to that of two-degree-of-freedom systems, which exhibit a 180-degree phase shift at the first anti-resonance. Above the first resonant frequency (approximately 1000 Hz for the dry skull), the phase relationship changes continually, and the skull exhibits many vibration modes.

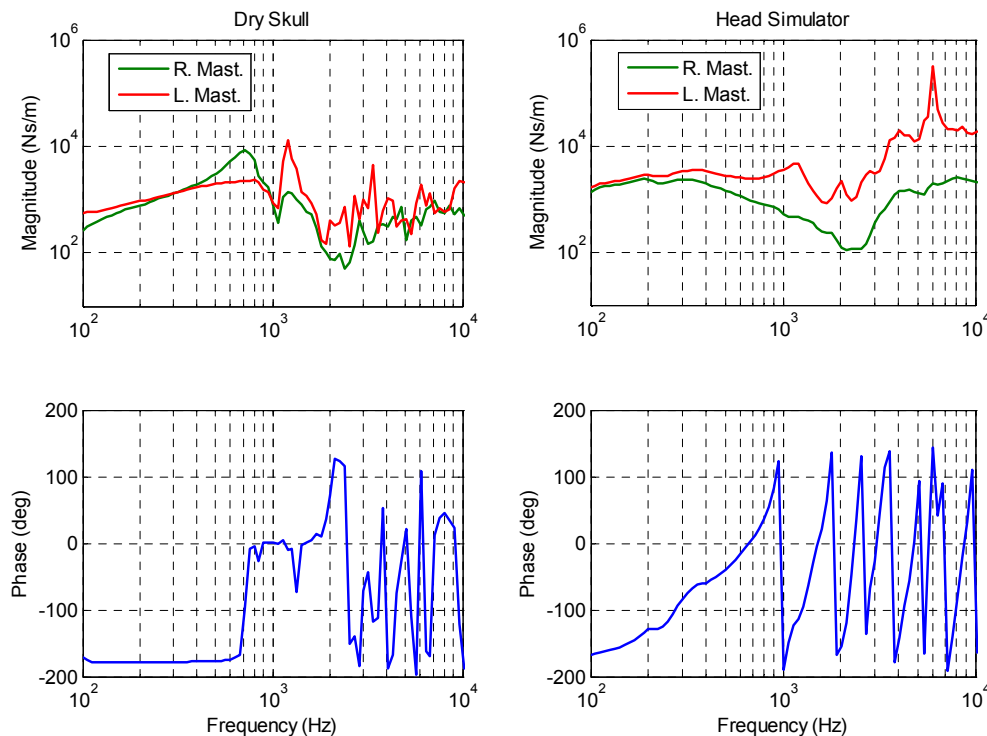


Figure 8. Driving Point Response. Dry skull response shown on the left, head simulator response shown on the right. Driven at the right mastoid. Results for the bone-anchored accelerometers mounted on the right and left mastoid are shown in the upper plots. The relative phase between the two accelerometers is shown in the lower plots.

3.3 Sound Field Excitation

The response of the head simulator to forcing from a sound field is shown in Figure 9 and Figure 10, with the head positioned facing the sound source and facing to the left of the sound source respectively. Data for each of the mastoid and temporal bone accelerometers are plotted in the figures. The acceleration values have been normalized by the forcing sound pressure level measured by a reference microphone adjacent to the head. Results for four separate tests are plotted to show test repeatability. Both the mastoid and the temporal bone accelerometers exhibit a rising response, which reaches a maximum at approximately 2 kHz. A complex response with many modes is seen at higher frequencies. Note that some directional effects can be seen below 500 Hz, but the responses are largely independent of orientation to the sound field above this frequency. For instance, the response of the uni-axial left mastoid accelerometer is higher when the helmet is oriented sideways, aligning the accelerometer axis with the direction of the sound forcing. The temporal bone accelerometers have a higher response when the head is facing the sound field, aligning these accelerometers with the direction of the sound forcing. The right mastoid acceleration is the resultant acceleration measured by the triaxial accelerometer, and as such, it shows little dependence on orientation. These data support the results from the driving point tests, which indicated that the head was vibrating as a rigid body for frequencies below 1 kHz and transitions to distortional modes of vibration at higher frequencies.

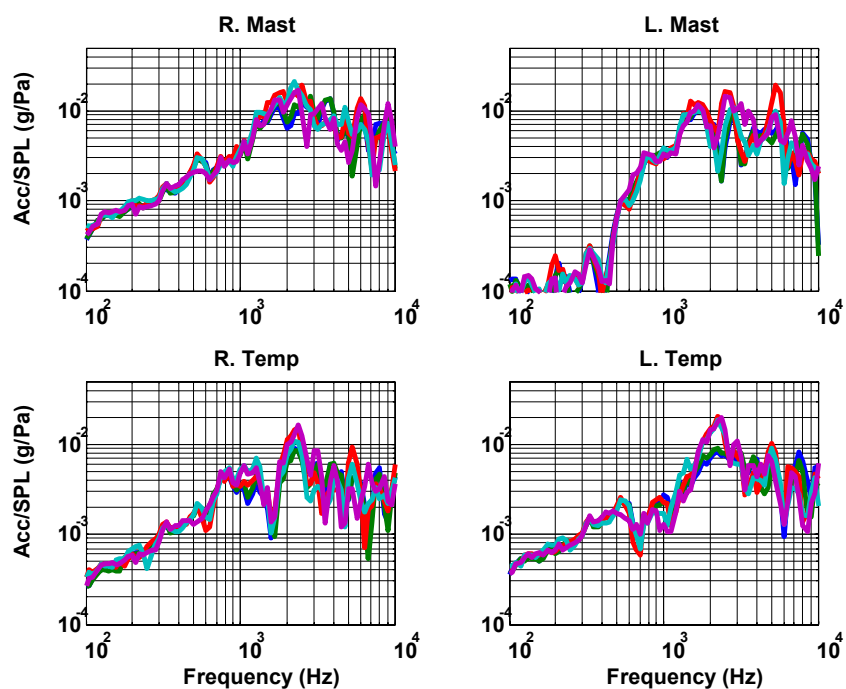


Figure 9. Head Response to Forcing from a Sound Field. Head simulator was placed facing the speaker in a hemi-anechoic chamber.

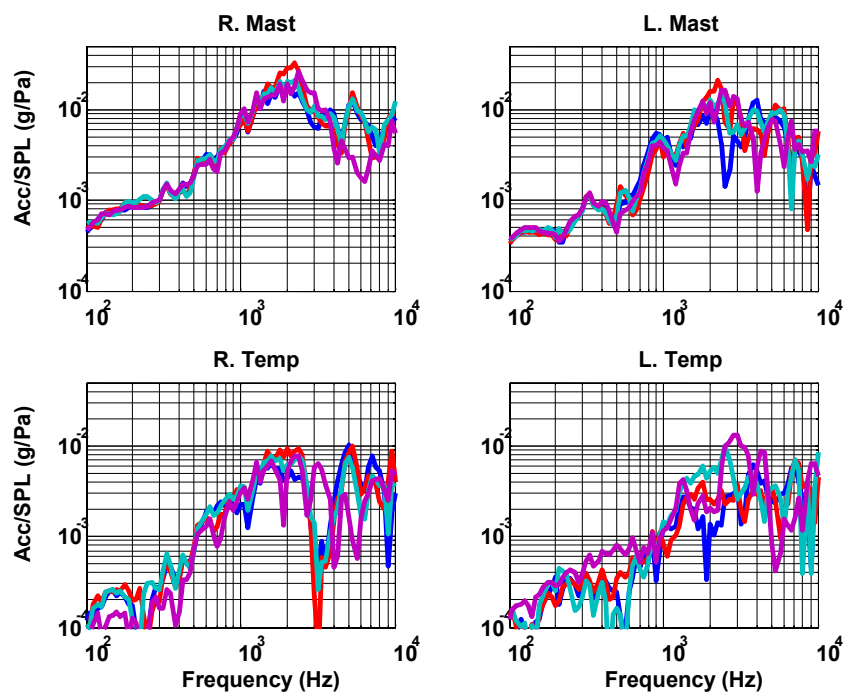


Figure 10. Head Response to Forcing from a Sound Field. Head simulator was placed facing to the left of the speaker in a hemi-anechoic chamber.

3.4 HPD Performance on Head Simulator

The complex nature of the high-frequency vibration would make it difficult to tune out or cancel the vibrations. The results of limited experiments with added damping and with active vibration cancellation supported this conclusion. Therefore, our focus was on attenuating the bone conduction transmission paths by shielding the head from the sound field. Shielding is achieved by wearing a helmet as a third layer of hearing protection, in addition to earplugs and earmuffs. A helmet was fabricated using rapid prototyping techniques for this study. The helmet, similar to that shown in Figure 11, comprised an outer shell, a foam liner, a sealed foam edge seal, and a face shield. The vibration attenuation provided by the helmet components was measured by fitting the helmet onto the head simulator and measuring the bone vibration with the head positioned in different orientations to the sound field. Representative results are plotted in Figure 12 and in Figure 13. The addition of earmuffs has a minimal effect on the head vibration, with the attenuation remaining below 5 dB for all frequencies. The same was also true for the foam liner, although these results are not included in the plot. The addition of a hard shell sealed to the head provided an approximately 10-dB increase in attenuation in the critical bone conduction frequency range around 2 kHz for sound directed from the back, while a face shield was required to give equivalent attenuation of sound directed from the front. At higher frequencies, the attenuation approached 20 dB. However, at frequencies below the first resonance where the head is vibrating as a rigid body, the helmet actually amplified the skull vibrations.



Figure 11. Hearing Protection Helmet. Helmet includes a hard shell, a foam liner, a sealed foam edge roll, earmuffs, and face shield.

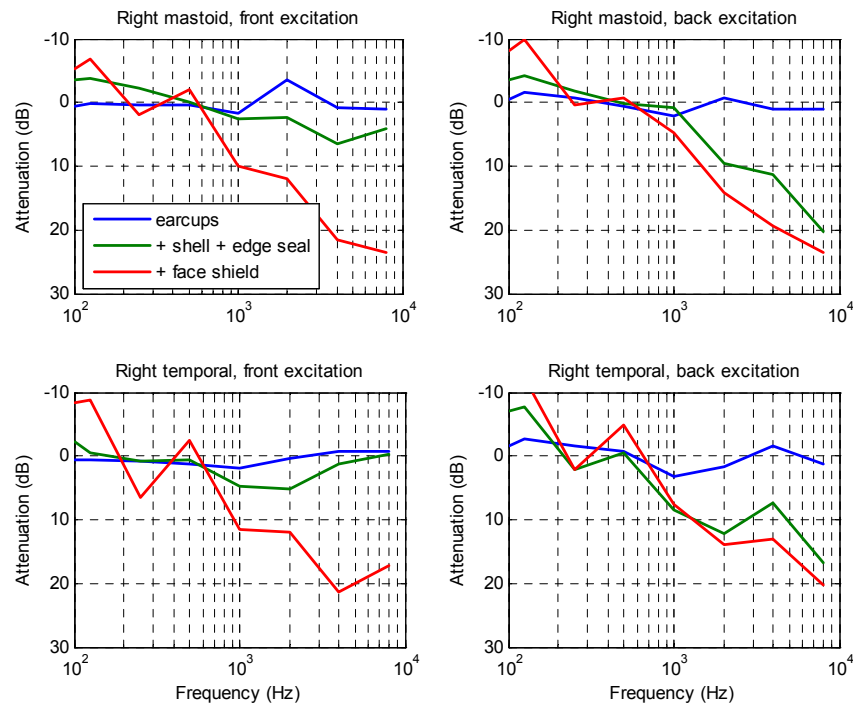


Figure 12. Vibration Attenuation Provided by Helmet Components on the Head Simulator in a Directional Sound Field. Head is oriented towards the sound field in the plots on the left, and away from the sound field in the plots on the right.

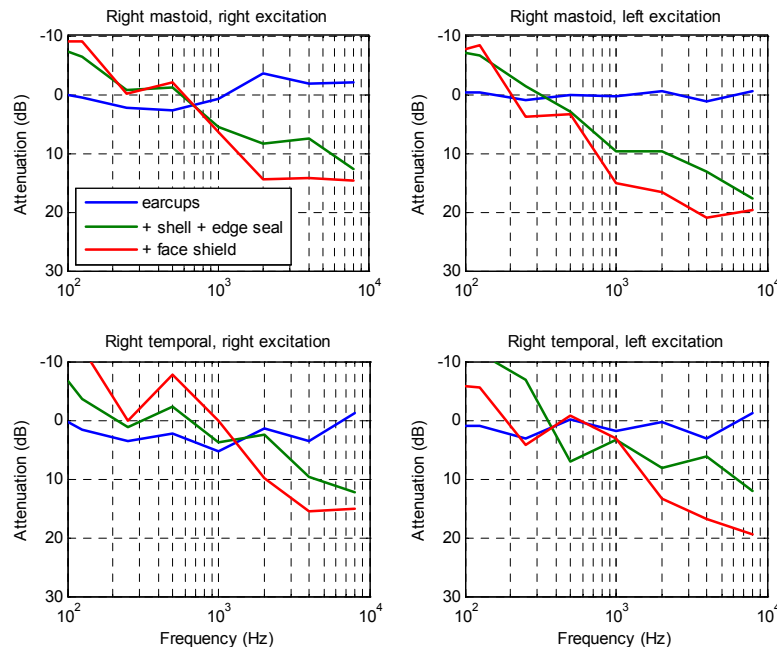


Figure 13. Vibration Attenuation Provided by Helmet Components on the Head Simulator in a Directional Sound Field. Head is oriented towards the sound field in the plots on the left, and away from the sound field in the plots on the right.

3.5 HPD Performance on Human Subjects

Various hearing protection configurations were also tested on human subjects to confirm the results from the simulator tests. Measured attenuation for plugs alone (single protection), plugs and muffs (double protection), and plugs, earmuffs, and helmet (third level) are shown in Figure 14. The helmet provides a significant increase in attenuation at frequencies above 1 kHz. At the lower frequencies, triple protection is no better than double protection, and at some frequencies may be worse. At higher frequencies, triple protection provides significantly greater attenuation. At 2 kHz, where the bone conduction response to sound is the greatest, the bone conduction limit is 40 dB, as represented by the limited double hearing protection attenuation. The triple protection provided by a helmet increases the mean attenuation to 55 dB. Therefore, the helmet is providing a 15-dB increase in attenuation. If the double protection is assumed to attenuate the air-conducted sound beyond the bone conduction limit, then the attenuation of bone-conducted sound is represented by the difference between the double protection attenuation and the triple protection attenuation (i.e., between plugs and muffs and plugs, muffs, and shell). A comparison of Figure 14 with Figure 12 shows that the frequency-dependent magnitude of the bone conduction attenuation provided by the helmet on human subjects is similar to the vibration attenuation provided by the helmet fitted to the head simulator. This result further confirms the value of the head simulator as a tool for hearing protection design and suggests that bone vibration is an adequate indicator of bone-conducted sound transmission.

The incremental increases in attenuation provided by partial coverage goggles or a full coverage face shield are shown in Figure 15. These data further support the results from the head simulator, which showed a similar increase in attenuation when the face shield was introduced.

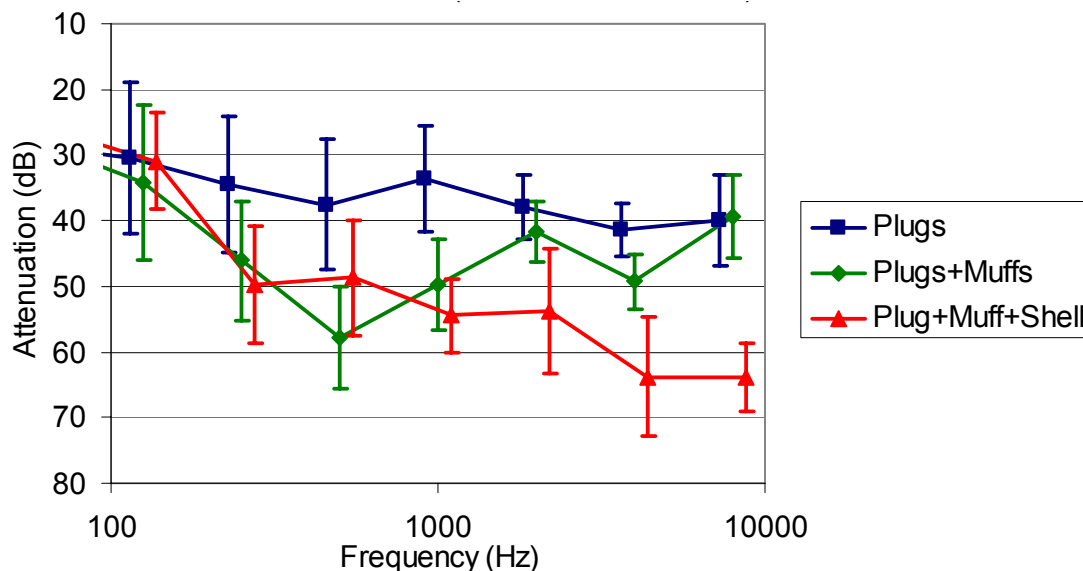


Figure 14. Attenuation Provided by Hearing Protection Components on Human Subjects. Results are the ensemble average of 8 subjects. Modified REAT protocol used (subjects facing towards directional sound field, pure tones, hemi-anechoic chamber). Error bars show ± 1 standard deviation (note points have been offset to distinguish the error bars).

Hearing Protection for Bone-Conducted Sound

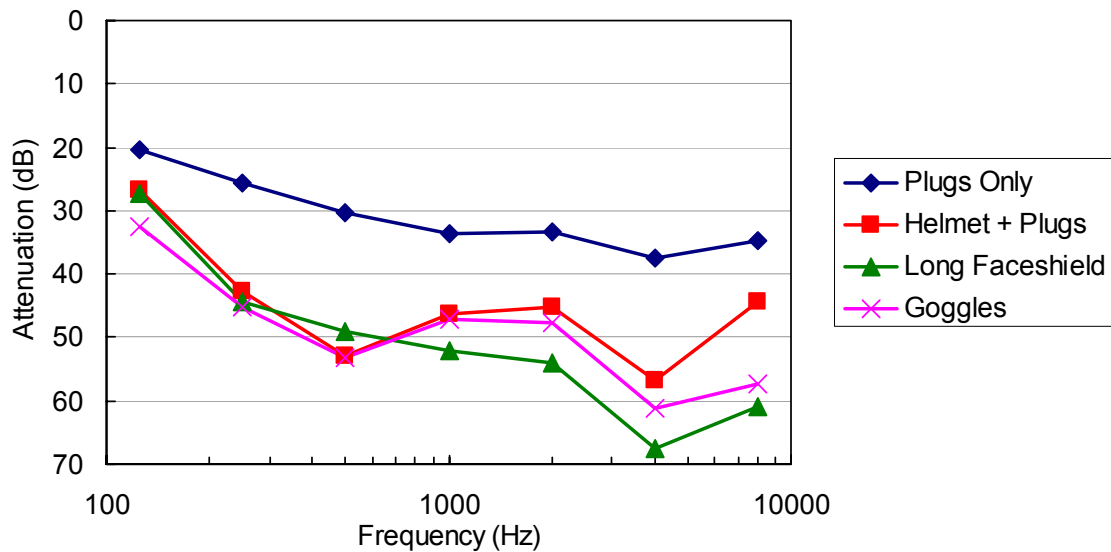


Figure 15. Attenuation Provided by Hearing protection Components on Human Subjects. Results are the ensemble average of 7 subjects. Modified REAT protocol used (subjects facing towards directional sound field, pure tones, hemi-anechoic chamber).

3.6 Head Isolation Tests

The body conduction thresholds measured for six test subjects are shown in Figure 16. The threshold levels increase with frequency from approximately 40 dB at 63 Hz to 80 dB at 8 kHz. There is reasonable consistency between the test subjects. These results are in general agreement with those of Ravicz and Melcher (2001) who measured a body conduction threshold of approximately 60 dB at 1 kHz.

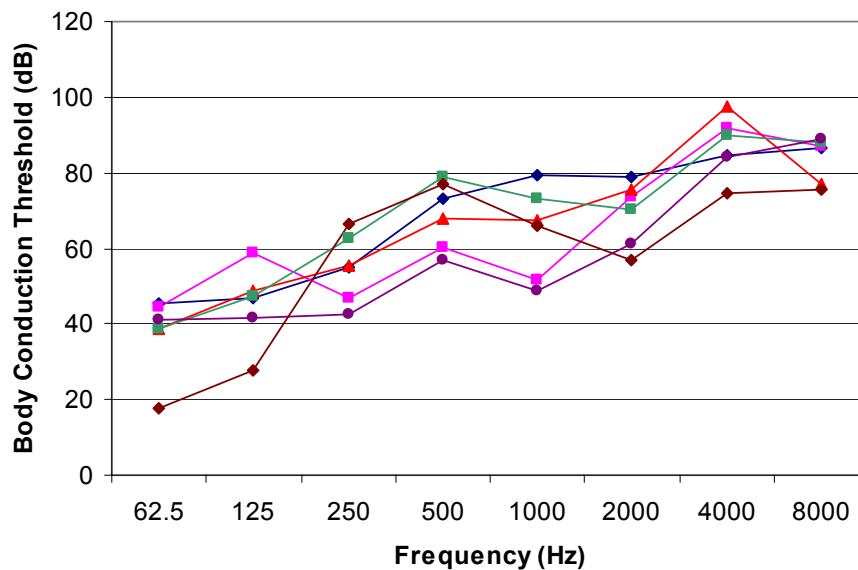


Figure 16. Body Conduction Thresholds. Results for six subjects shown.

The validity of the results may be checked by testing whether the difference between the measured body conduction attenuation is significantly different to that of the bypass path (hearing protection and box attenuation). A statistical t test showed a highly significant ($P = 0.01$) difference in five of the eight frequencies tested (250 Hz, 1 kHz, 2 kHz, 4 kHz, and 8 kHz). This significance is apparent in the summary plot shown in Figure 17, where the error bars correspond to a standard error with a 95% confidence interval. The helmet attenuation is also included in this plot.

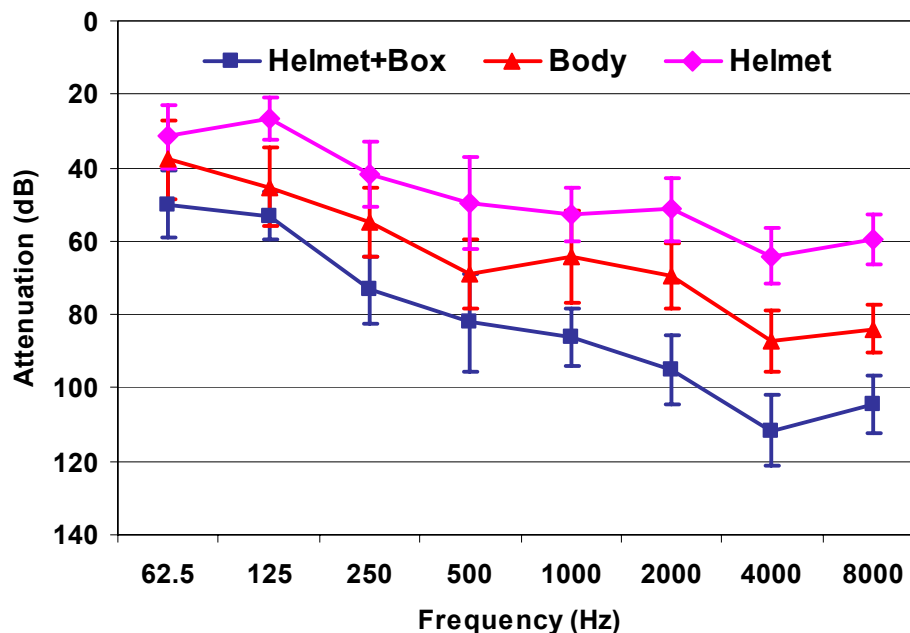


Figure 17. Summary Plot. Mean data for all subjects is shown with standard errors representing 95% confidence intervals.

5.0 SUMMARY AND CONCLUSIONS

An investigation into bone-conducted transmission paths and the effectiveness of hearing protection devices against these paths was conducted using human subjects and a head simulator. The response of the head simulator was validated by comparing its driving point impedance with results from tests on human subjects and with results reported in the literature. The vibration response of the head simulator in a sound field was then measured. The response was greatest above the first resonance in the frequency range between 2 kHz and 3 kHz, which is the frequency range where bone conduction transmission limits the attenuation of hearing protection devices. The modal behavior of the skull is complicated and bone vibration is most effectively reduced by shielding the head from the sound field with a helmet and a face shield. Measurements with the head simulator showed that shielding can provide between 10 and 20 dB of attenuation of bone vibration above 1 kHz. Measurements on human subjects showed a similar increase in sound attenuation beyond the head conduction limit represented by double protection (earplugs and earmuffs). If the head were completely shielded from the sound field, the attenuation would be limited by body conduction. Measurements of the body conduction limit found that the limit increases from 40 dB at 63 Hz to 80 dB at 8 kHz. The relative head and body conduction limits measured in this study are summarized in Figure 18.

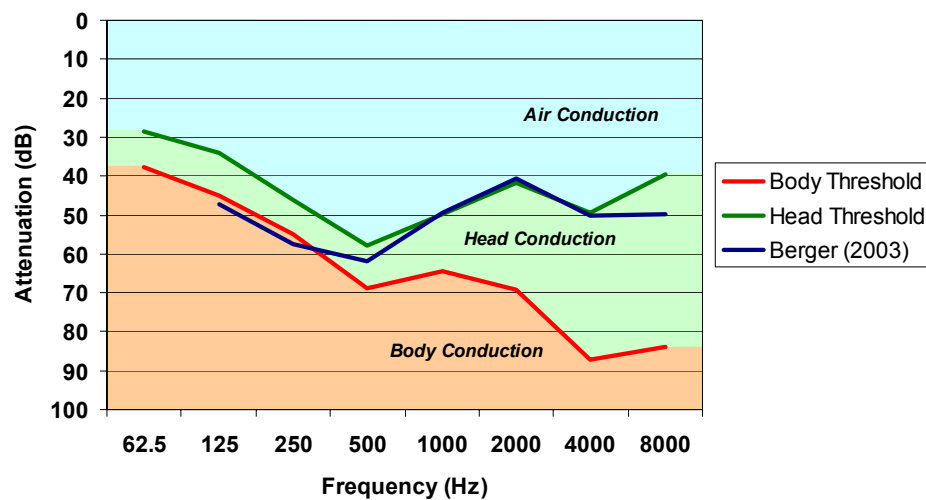


Figure 18. Head and Body Thresholds for Sound Conduction.

6.0 REFERENCES

- [1] Berger, E.H. and Kieper, J.E., "Hearing Protection: Surpassing the Limits to Attenuation Imposed by the Bone-Conduction Pathways," *J. Acoust. Soc. Am.*, Vol. 114, No. 4, Pt 1, Oct 2003, pp. 1955–1967.
- [2] Berger, E.H. and Kerivan, J.E., "Influence of Physiological Noise and the Occlusion Effect on the Measurement of Real-Ear Attenuation at Threshold," *J. Acoust. Soc. Am.*, Vol. 74, No. 1, July 1983, pp. 81–94.
- [3] Hakansson, B. and Carlsson, P., "The Mechanical Impedance of the Human Head with and without Skin Penetration," *J. Acoust. Soc. Am.*, Vol. 80, No. 4, 1986, pp. 1065–1075.
- [4] Khalil, T.B., and Viano, D.C., "Experimental Analysis of the Vibrational Characteristics of the Human Skull," *J. Sound and Vibration*, Vol. 63, No. 3, 1979, pp. 351–376.
- [5] Ravicz M.E., and Melcher, J.R., "Isolating the Auditory System from Acoustic Noise During Functional Magnetic Resonance Imaging: Examination of Noise Conduction through the Ear Canal, Head, and Body," *J. Acoust. Soc. Am.*, Vol. 109, No. 1, Jan 2001, pp. 216–241.
- [6] Smith, J.B. and Suggs, C.W., "Dynamic Properties of the Human Head," *J. Sound and Vibration*, Vol. 48, No. 1, 1976, pp. 35–43.
- [7] Sohmer, H., Freeman, S., Geal-Dor, M., Adelman, C. and Savion, I., "Bone Conduction Experiments in Humans – a Fluid Pathway from Bone to Ear," *Hearing Research*, Vol. 146, No. 1-2, 2000 pp. 81–88.
- [8] Stalnaker, R.L., Fogler, J. L. and McElhaney, J. H., "Driving Point Impedance Characteristics of the Head," *J. Biomechanics*, Vol. 4, 1971, pp. 127–139.
- [9] Tonndorf, J., "Bone Conduction. Studies in Experimental Animals," *Acta Oto-Laryngol.*, Suppl. 213, 1966, pp. 1–132.

ACKNOWLEDGEMENT OF SUPPORT AND DISCLAIMER (MAY 1995)

This material is based upon work supported by the U.S. Department of the Navy under Contract No. N00421-02-C-0266 and by the U.S. Air Force under Contract No. F49620-03-C-0008. The support of the Navy technical monitor, Mr. Jim Wilt, and of the Air Force technical monitor, Mr. Willard Larkin, is gratefully acknowledged. Our collaboration with the Dartmouth Engineering School including Professor Laura Ray, Professor Bob Collier, and Mr. Matt Maher is also gratefully acknowledged.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the U.S. Department of the Navy or the U.S. Air Force.

Hearing Protection for Bone-Conducted Sound



Evaluation of Acoustic Propagation Paths into the Human Head

Prof. Dr. William D. O'Brien, Jr. and Ms. Yuhui Liu

Bioacoustics Research Laboratory
Department of Electrical and Computer Engineering
University of Illinois
405 N. Mathews
Urbana, IL 61801
United States of America

Email: wdo@uiuc.edu / yuhuiliu@uiuc.edu

SUMMARY

The overall goal has been to develop an acoustic wave propagation model using well-understood and documented computational techniques that track and quantify an air-borne incident acoustic wave propagated around, into and in the human head. This model serves as a computational tool to elucidate the acoustic wave propagation around, into and in the human head. Specifically, the model determines two features: (1) alternate acoustic propagation paths to the cochlear shell that exist besides the normal air-borne acoustic propagation path (eardrum-ossical path) through the auditory canal and (2) sound pressure amplitude in the cochlear shell (relative to the air-borne sound pressure amplitude) via the alternate propagation paths. A 3D finite-element solid mesh was constructed using a digital image database of an adult male head. Coupled acoustic-mechanical finite-element analysis (FEA) was used to model the wave propagation through the fluid-solid-fluid media. The pressure field in fluid media and the displacement field in solid structures were computed at each time step. Instantaneous acoustic pressure waveforms were recorded at various positions inside and outside of the head model, and propagation trajectories (ray paths) were constructed and evaluated from wavefront normals as a function of frequency and incidence angle. The acoustic loss across the skull was estimated to be approximately 33 dB, consistent with theoretical estimates. The computational ray-path results and the theoretical solutions calculated using Snell's law gave a 0.7° difference for low-angle incidence; 10° difference was obtained for larger angle incidence.

1.0 INTRODUCTION

The overall objective of the proposed program is the development of an acoustic wave propagation model using well-understood and documented computational techniques that track an air-borne incident acoustic wave propagated around, into and in the human head. This model will serve as a computational tool to elucidate the acoustic wave propagation around, into and in the human head to evaluate alternate acoustic propagation paths that exist besides the normal air-borne acoustic propagation path (eardrum-ossical path) through the auditory canal to the cochlea, a cavity in the temporal bone. Specifically, the computational model will evaluate bone conduction that is produced from a sound field in air, and track the bone-conduction acoustic propagation paths into the cochlear shell.

The validation testing of the computational model will involve human subject testing. The interplay between human subject (experimental) testing and theoretical modeling has always benefited the advancement of science. The theoretical model provides a known fundamental basis, and the experiment provides the

O'Brien, Jr., W.D.; Liu, Y. (2005) Evaluation of Acoustic Propagation Paths into the Human Head. In *New Directions for Improving Audio Effectiveness* (pp. 15-1 – 15-24). Meeting Proceedings RTO-MP-HFM-123, Paper 15. Neuilly-sur-Seine, France: RTO.
Available from: <http://www.rto.nato.int/abstracts.aps>.

Evaluation of Acoustic Propagation Paths into the Human Head

evidence to evaluate the model. When the experimental and modeling results essentially agree, then a basis is provided to understand at a fundamental level why the experiment performs in the way it does. Further, and most important, with the model validated against experimental observations, the model then becomes a valuable tool to test other hypotheses without the need to conduct expensive and time consuming experiments.

2.0 MOTIVATION AND OBJECTIVE

Noise-induced hearing loss (NIHL) has been an important issue for many years. Human beings that are exposed to excessively high levels of noise can result in temporary hearing loss or permanent hearing damage, *e.g.*, crews that must work near military aircraft are subject to severe noise environments, with Sound Pressure Levels (SPL) reaching 145 to 150 dB. Even a single impulse exposure per day to such intense sounds exceeds permissible noise exposure rules [OSHA 1910.95]. Different hearing protection devices (HPDs) have been designed to prevent NIHL, such as earplugs, earmuffs and helmets. However, there is a limit to how much attenuation can be achieved by protecting the ear canals with muffs and/or plugs, because they do not attenuate sounds that are conducted to the inner ear by the hard and soft tissues of the head and body.

Bone-conduction pathways have been hypothesized for nearly the past three-quarters of a century [Guild 1936; Barany 1938; Wever 1954] and are still largely unknown [Sohmer 2000; Sohmer 2004]. However, some bone-conduction pathways have been proposed, but most of the work to deduce these pathways has been *via* direct contact methods with some type of vibrating device [Wever 1954; Freeman 2000; Sohmer 2000; Stefan 2004]. Human beings sense air-borne sound transmitted by bone conduction when the sound pressure of the external field is about 60 dB above the threshold required for hearing *via* the eardrum-ossical path [Békésy 1948; Zwislocki 1957]. This suggests that noise above 60 dB SPL cannot be fully attenuated by covering or blocking the auditory canal.

When individuals are exposed to severe noise environments, such as those generated by aircraft engines and military weapons that approach and even exceed 150 dB SPL, even if hearing protection equipment is worn, they may be subject to hearing damage. Furthermore, NIHL at low frequencies (*e.g.*, 250 Hz and less) are even more challenging. Thus, for very high noise environments and within certain frequency ranges, conventional HPDs do not provide sufficient protection.

Thus, the goal of the proposed program is to deduce bone-conduction pathways into the cochlear shell. From such knowledge, the design of advanced HPDs can then be based on well-founded scientific fundamentals; it should be noted, however, that the program's goal does not include the development of advanced HPDs.

The normal hearing process termed "air transmission" or "eardrum-ossical path" refers to air-borne sound entering the auditory canal and transduced by the organ of Corti. However, other propagation path processes are considered for purposes of this study. The process, termed "bone conduction," refers to sound entering at any location other than the auditory canal and transduced by the organ of Corti, that is, other than the eardrum-ossical path. The term bone conduction does not necessary imply that the propagation path is entirely bone. The distinction between these two processes (eardrum-ossical path and bone conduction) is necessary because the acoustic propagation paths are evaluated *via* finite-element analysis (FEA) procedures, and the eardrum-ossical path is not included in the evaluation.

For purposes of this study, sound levels will be computationally evaluated to and in the region referred to as the cochlear shell, a cavity that is located in the highly porous temporal bone that contains the organ of Corti. The cochlear response is a vectorial integration of the air-conduction pathway, different bone-conduction pathways, and any other potential alternative pathways. It is possible that each of these pathways is more

effective at different frequencies. Improving the current hearing protection devices thus focuses the need to better understand the sound propagation around, into and inside of the human head, to evaluate different propagation pathways and the paths that are taken to the cochlea.

3.0 FUNDAMENTALS FOR VALIDATION STUDIES

3.1 Wave Equation

The necessary assumptions used in ANSYS acoustic analysis include:

- The fluid is compressible, but only relatively small pressure changes with respect to the mean pressure are allowed.
- The fluid is inviscid (no viscous dissipation).
- There is no mean flow of the fluid.
- The mean density and mean equilibrium pressure are uniform throughout the fluid.

The acoustic wave equation is given by [Pierce 1989]

$$\frac{\partial^2 P}{\partial t^2} - c^2 \nabla^2 P = 0, \quad (1)$$

where P is the acoustic pressure, c is the sound speed in the fluid medium, t is time and ∇ is the Laplacian operator.

The finite-element statement of the wave equation is given by

$$[M_e^p] \{\ddot{P}_e\} + [K_e^p] \{\dot{P}_e\} + \rho_0 [R_e]^T \{\ddot{u}_e\} = \{0\}, \quad (2)$$

where

$$[M_e^p] = \frac{1}{c^2} \int_{vol} \{N\} \{N\}^T d(vol) \text{ is the fluid mass matrix} \quad (2a)$$

$$[K_e^p] = \int_{vol} [B]^T [B] d(vol) \text{ is the fluid stiffness matrix} \quad (2b)$$

$$\rho_0 [R_e] = \rho_0 \int_S \{N\} \{n\}^T \{N'\}^T d(S) \text{ is the fluid-structure coupling mass matrix} \quad (2c)$$

$\{N\}$ is the element shape function for pressure

$\{N'\}$ is the element shape function for displacements

$\{P_e\}$ is the nodal pressure vector

$\{u_e\} = \{u_{xe}\}, \{u_{ye}\}, \{u_{ze}\}$ is the nodal displacement component vectors

$\{n\}$ is the unit normal at the fluid boundary

vol is the volume of domain

S is the surface where the derivative of pressure normal to the surface is applied

Well-characterized analytical solutions [Morse 1968] are compared with the finite-element simulation results to validate the analysis.

3.2 Scattering by Rigid Cylinder (2D Rigid Cylinder Case)

The acoustic scattering from a rigid cylinder of radius a due to a plane wave propagating in a direction perpendicular to the cylinder's axis is given by

$$\begin{aligned}
 p_{\text{inc}} &= P_0 e^{ik(r \cos \phi - ct)} = P_0 [J_0(kr) + 2 \sum_{m=1}^{\infty} i^m \cos(m\phi) J_m(kr)] e^{-i\omega t} \\
 p_{\text{sca}} &= \sum_{m=0}^{\infty} A_m \cos(m\phi) [J_m(kr) + iN_m(kr)] e^{-i\omega t} \\
 A_m &= -\varepsilon_m P_0 i^{m+1} e^{-i\gamma_m} \sin \gamma_m \\
 \tan \gamma_0 &= -\frac{J_1(ka)}{N_1(ka)}, \quad \tan \gamma_m = \frac{J_{m-1}(ka) - J_{m+1}(ka)}{N_{m+1}(ka) - N_{m-1}(ka)},
 \end{aligned} \tag{3}$$

where

- p_{inc} is the incident plane wave
- p_{sca} is the scattered wave
- a is the cylinder radius
- r is the radial distance from the center of the cylinder
- ϕ is the axial angle for cylindrical coordinates
- c is the speed of sound in the medium surrounding the cylinder
- k is the $\omega/c = 2\pi/\lambda$, wavenumber
- γ_m is the phase shift
- $\varepsilon_m = 1$ when $m = 0$; $\varepsilon_m = 2$ when $m > 0$

The total acoustic pressure at the surface of the cylinder at an angle ϕ relative to the x axis is given by

$$p_a = p_{\text{inc}} + p_{\text{sca}} = \frac{4P_0}{\pi ka} e^{-i\omega t} \sum_{m=0}^{\infty} \frac{\cos(m\phi)}{E_m} e^{i[-\gamma_m + (\pi m/2)]}, \tag{4}$$

where E_m , the radiation amplitude for a cylinder, is

$$E_0 \approx \sqrt{\frac{8}{2\pi ka}}, \quad E_{m>0} \approx \sqrt{\frac{2}{\pi ka}}, \tag{4a}$$

when $ka \gg m + 1/2$, and

$$E_0 \approx \frac{4}{\pi ka}, \quad E_{m>0} \approx \frac{m!}{2\pi} \left(\frac{2}{ka}\right)^{m+1}, \tag{4b}$$

when $ka \ll m + 1/2$.

3.3 Scattering by Rigid Sphere (3D Rigid Sphere Case)

The acoustic scattering from a rigid sphere of radius a due to a plane wave propagating to the right along the polar axis is given by

$$p_{inc} = P_0 e^{ik(r \cos \vartheta - ct)} = P_0 \sum_{m=0}^{\infty} (2m+1) i^m P_m(\cos \vartheta) j_m(kr) e^{-i\omega t}$$

$$p_{sca} = -P_0 \sum_{m=0}^{\infty} (2m+1) i^m e^{-i\delta_m} \sin \delta_m P_m(\cos \vartheta) [j_m(kr) + i n_m(kr)] e^{-i\omega t} \quad (5)$$

where p_{sca} is the wave scattered from the sphere whose center is the polar origin. The total acoustic pressure at a point on the sphere an angle ϑ relative to the polar axis (note that the point $\vartheta = 0^\circ$ is the location farthest away from the source of the sound) is given by

$$p_a = (p_{inc} + p_{sca})_{r=a} = P_0 (ka)^{-2} \sum_{m=0}^{\infty} \frac{(2m+1)}{B_m} P_m(\cos \vartheta) e^{-i(\delta_m - \frac{1}{2}\pi m - \omega t)}$$

$$\approx (1 + \frac{3}{2} ika \cos \vartheta) P_0 e^{-i\omega t}, \quad ka \ll 1 \quad (6)$$

3.4 Scattering by Nonrigid Sphere (3D Elastic Sphere Case)

The scattering from a sphere that is not perfectly rigid is more complicated. The total acoustic pressure at a point on the sphere an angle ϑ relative to the polar axis (note that the point $\vartheta = 0^\circ$ is the point farthest away from the source of the sound) is given by

$$p_a = P_0 \sum_{m=0}^{\infty} (2m+1) i^m P_m(\cos \vartheta) [j_m(kr) - \frac{1}{2}(1 + R_m) h_m(kr)], \quad (7)$$

where

$$R_m, \text{ the reflection coefficient, is } 1 + R_m = 2 \frac{j'_m(ka) + i\beta_m j_m(ka)}{h'_m(ka) + i\beta_m h_m(ka)} \quad (7a)$$

$$\beta_m, \text{ the effective admittance, is } \beta_m = i \frac{\rho c}{\rho_e c_e} \left[\frac{j'_m(k_e a)}{j_m(k_e a)} \right] \quad (7b)$$

$$k_e = \omega/c_e$$

ρ, c are the density and speed of the medium surrounding the sphere

ρ_e, c_e are density and speed of the sphere material

4.0 VALIDATION STUDIES

The ANSYS (ANSYS, Inc., Canonsburg, PA) finite-element analysis (FEA) general processing code for both harmonic and transient solutions has been validated using well-understood two-dimensional (2D) and three-dimensional (3D) models. Harmonic (continuous-wave) validation studies were conducted for (1) 2D rigid cylinder, (2) 3D rigid sphere and (3) 3D elastic sphere. Transient (one-cycle sine wave) validation studies were conducted for (1) homogeneous air, (2) 2D elastic cylinder shell, and (3) 3D elastic (water) sphere. For all of the validation studies, either a harmonic or transient acoustic plane wave was initiated in air. The air-borne acoustic wave was incident on the geometric model that was either rigid, elastic or water. Water is an ideal fluid for this study because it has acoustic propagation properties similar to those of brain and other soft tissues. In all cases, the computational solutions of acoustic pressure distribution agreed well with the analytic solutions of acoustic pressure distribution.

To be familiar with different types of acoustic analysis in ANSYS and ensure building our FEA model in a proper way, several case studies were carried out on different geometric models. For these validation studies, the simulation results are compared with the theoretical solutions to evaluate the model. Herewith are the symbols used:

f = center frequency of the incident wave (kHz)
 c = sound speed in the medium (m/s)
 λ = wavelength in the medium (m)
 ρ = material density (kg/m^3)
 a = cylinder radius / shell outer radius / sphere radius (3D) (m)
 d = shell thickness (m)
 r = radial distance from the center of the cylinder
BOUND = FEA absorbing boundary radius (m)
 ϕ_{inc} = incident angle, the angle between incident wave and +x axis (degree)
 X_{inc} = incident wave position (m)
 P_{inc} = incident wave pressure amplitude (Pa)
 T = period of the incident wave
DPW = number of mesh divisions per wavelength
ITS = integration time step size

4.1 2D Rigid Cylinder (Harmonic Analysis Case)

Figure 1 shows the FEA model used in ANSYS for the 2D rigid cylinder case (§3.2). The air-borne incident harmonic plane wave propagates in the +x-axis direction. When the propagating wave meets the rigid cylinder target, acoustic pressure is distributed around the cylinder surface.

Parameters:

$$f = 3 \text{ kHz}$$

$$c_{\text{air}} = 340 \text{ m/s}$$

$$\lambda_{\text{air}} = 340 / 3000 = 0.113 \text{ m}$$

$$\rho_{\text{air}} = 1.2 \text{ kg/m}^3$$

$$a = 0.4 \lambda_{\text{air}} = 0.0452 \text{ m}$$

$$\text{BOUND} = 1.4 \lambda_{\text{air}} = 0.158 \text{ m}$$

$$\phi_{\text{inc}} = 0^\circ$$

$$X_{\text{inc}} = -0.625 \lambda_{\text{air}} = 0.0706 \text{ m}$$

$$P_{\text{inc}} = 1 \text{ Pa}$$

$$\text{DPW} = 20$$

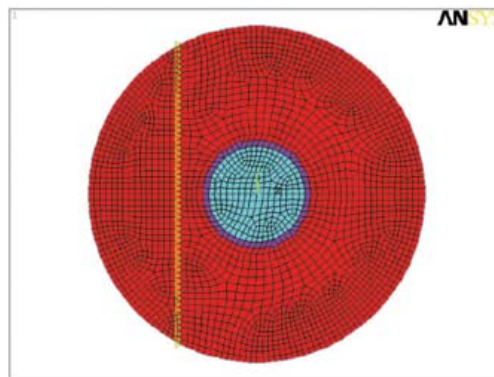


Fig 1: The 2D rigid cylinder FEA model generated by ANSYS. Red: air; Light blue: rigid cylinder; Dark blue: connecting elements; Yellow: location of plane wave initiation.

Figure 2 shows the ANSYS simulation results (circles) of the pressure distribution on the cylinder surface compared with the analytical solutions (line), and along the radial component. Good agreement is obtained.

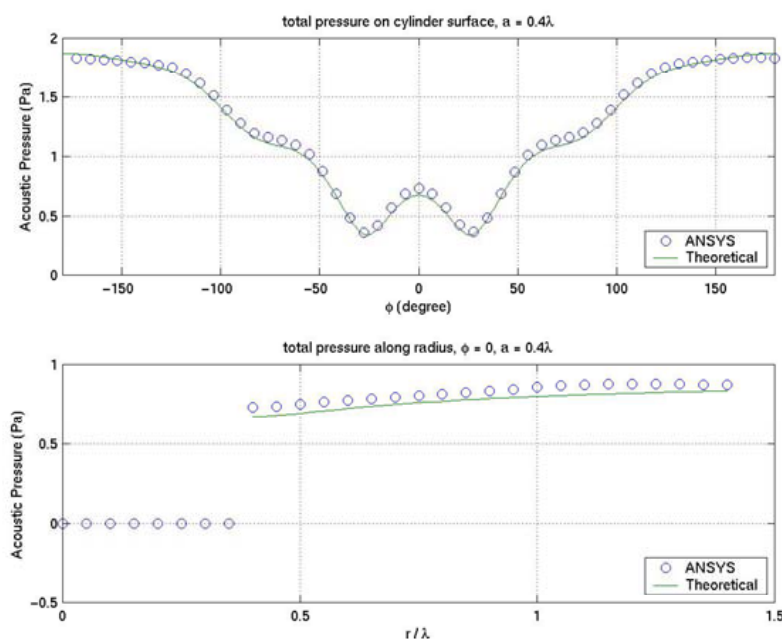


Fig 2: 2D rigid cylinder simulation results (circles) vs. analytical solutions (lines). Top plot: total pressure on cylinder surface (note that the point $\phi = 0^\circ$ is the point farthest away from the source of the sound). Bottom plot: total pressure along radius, observation angle $\phi = 0^\circ$ (note that r = radial distance from the center of the cylinder).

4.2 3D Rigid Sphere (Harmonic Analysis Case)

Figure 3 shows the FEM model used in ANSYS for the 3D rigid sphere case (§3.3). The air-borne incident harmonic plane wave propagates in the +x-axis direction. This case is of considerable practical importance because many scattering objects are more or less spherical. When the propagating wave meets the rigid spherical target, acoustic pressure is distributed around the spherical surface.

Evaluation of Acoustic Propagation Paths into the Human Head

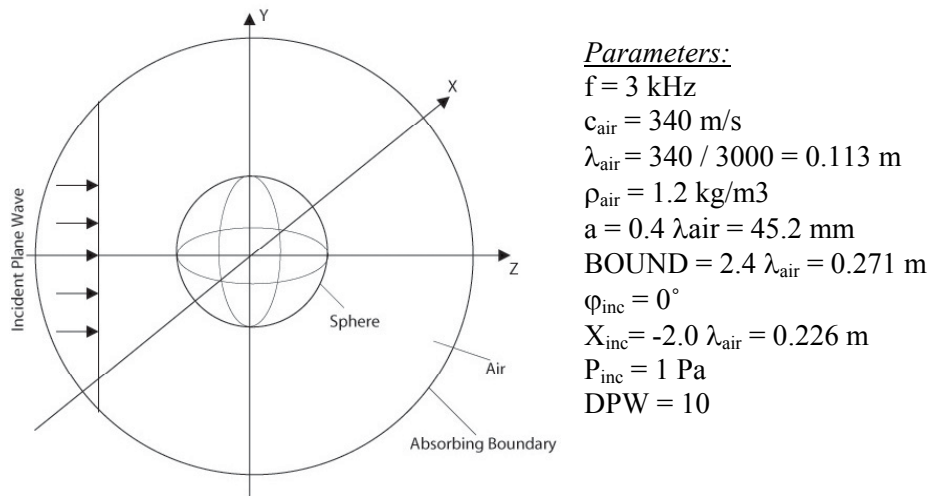


Fig 3: The 3D rigid and nonrigid sphere FEA model.

Figure 4 shows the ANSYS simulation results (circles) of the pressure distribution on the spherical surface compared with the analytical solutions (line), and along the radial component. Good agreement is obtained.

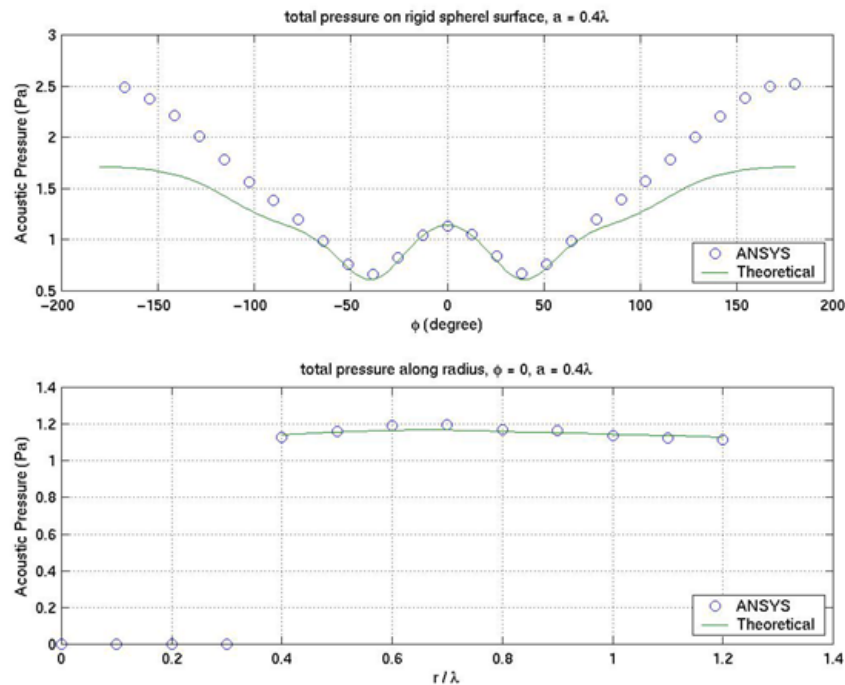


Fig 4: 3D rigid sphere simulation results (circles) vs. analytical solutions (lines). Top plot: total pressure on the surface of the center sphere cross section (note that the point $\phi = 0^\circ$ is the point farthest away from the source of the sound). Bottom plot: total pressure along +z axis, observation angle $\phi = 0^\circ$ (note that r = radial distance from the center of the sphere).

4.3 3D Elastic Sphere (Harmonic Analysis Case)

For the perfectly rigid cases (§4.1 and §4.2) the propagated wave does not enter the object. However, for real objects, the propagated acoustic wave will enter the object. Figure 3 shows the FEA model used in ANSYS for the 3D nonrigid sphere case (§3.4). Here the nonrigid sphere's density and speed are assumed to be twice those of air (parameters same as in Fig 3 except for $c_{\text{sphere}} = 680 \text{ m/s}$ and $\rho_{\text{sphere}} = 2.4 \text{ kg/m}^3$). The air-borne incident harmonic plane wave propagates in the +x-axis direction. This case is of considerable practical importance because many scattering objects are more or less spherical. When the propagating wave meets the nonrigid spherical target, acoustic pressure is distributed around the spherical surface, and also enters into the sphere.

Figure 5 shows the ANSYS simulation results (circles) of the pressure distribution on the elastic spherical surface compared with the analytical solutions (line), and along the radial component. Fair agreement is obtained.

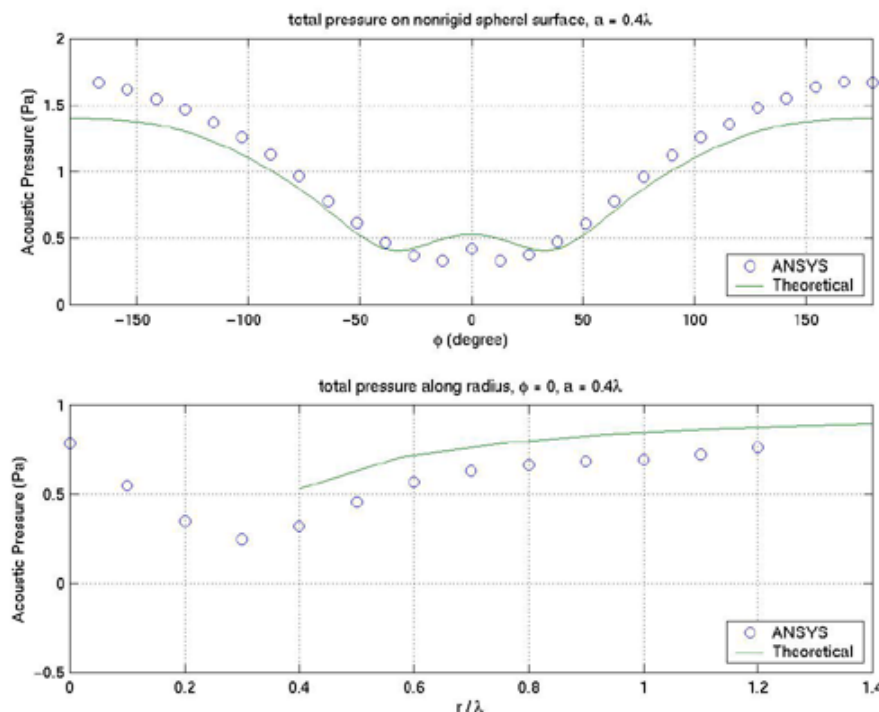


Fig 5: 3D elastic sphere simulation results (circles) vs. analytical solutions (lines). Top plot: total pressure on the surface of the center sphere cross section (note that the point $\phi = 0^\circ$ is the point farthest away from the source of the sound). Bottom plot: total pressure along +z axis, observation angle point $\phi = 0^\circ$ (note that r = radial distance from the center of the sphere).

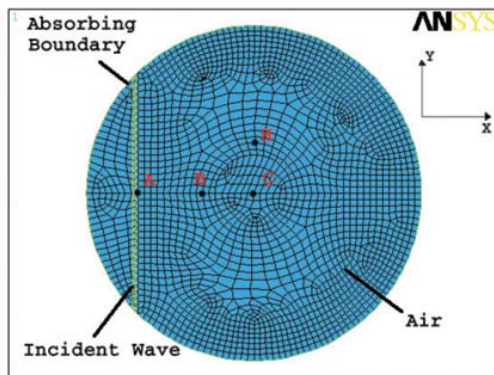
4.4 2D Analysis in Homogenous Media (Transient Analysis Case)

Transient analysis is a more realistic approach than harmonic analysis for modeling the incident acoustic pressure waveforms onto and in the human head. However, transient analysis is more involved because it requires added computer resources and more of *our* resources, in terms of the “engineering” time involved.

Evaluation of Acoustic Propagation Paths into the Human Head

To save a significant amount of these resources, preliminary studies were conducted to understand further the physics of the problem for validation purposes, that is, analyzing a simpler model provides better insight into the problem at minimal cost.

The 2D transient analysis (air is treated as the object) is evaluated to determine whether the computation domain space introduces artifacts. Here, a pulse wave (gated sinusoid wave) initiated in homogenous air medium is simulated. A simple geometry model (Fig 6) is used that is similar to the 2D cylinder harmonic-analysis model except that the 2D cylinder is air. The purpose of using an air-filled cylinder submerged in air is to consider possible artifacts that might be caused by different meshes. The pulse is applied in air and propagates in the +x direction. The locations A $(-0.9\lambda, 0)$, B $(-0.65\lambda, 0)$, C $(0.4\lambda, 0)$, H $(0, 0.4\lambda)$ (in Fig 6) denote where the acoustic pressure waveforms are plotted in Fig 7 to evaluate mainly DPW and ITS.



Parameters:

$$f = 3 \text{ kHz}$$

$$\text{Air } (c_{\text{air}} = 340 \text{ m/s}, \lambda_{\text{air}} = 0.113 \text{ m}, \rho_{\text{air}} = 1.2 \text{ kg/m}^3)$$

$$\phi_{\text{inc}} = 0^\circ$$

$$P_{\text{inc}} = 1 \text{ Pa}$$

$$\text{DPW} = \text{variable}$$

$$\text{ITS} = \text{variable}$$

$$a = 0.4 \lambda_{\text{air}} = 45.2 \text{ mm}$$

$$\text{BOUND} = a + 0.9 \lambda_{\text{air}} \text{ or } a + 1.4 \lambda_{\text{air}}$$

Fig 6: The wave propagation through air only. Locations A, B, C and H are denoted as red letters where A, B and C are on the x axis and H is on the y axis.

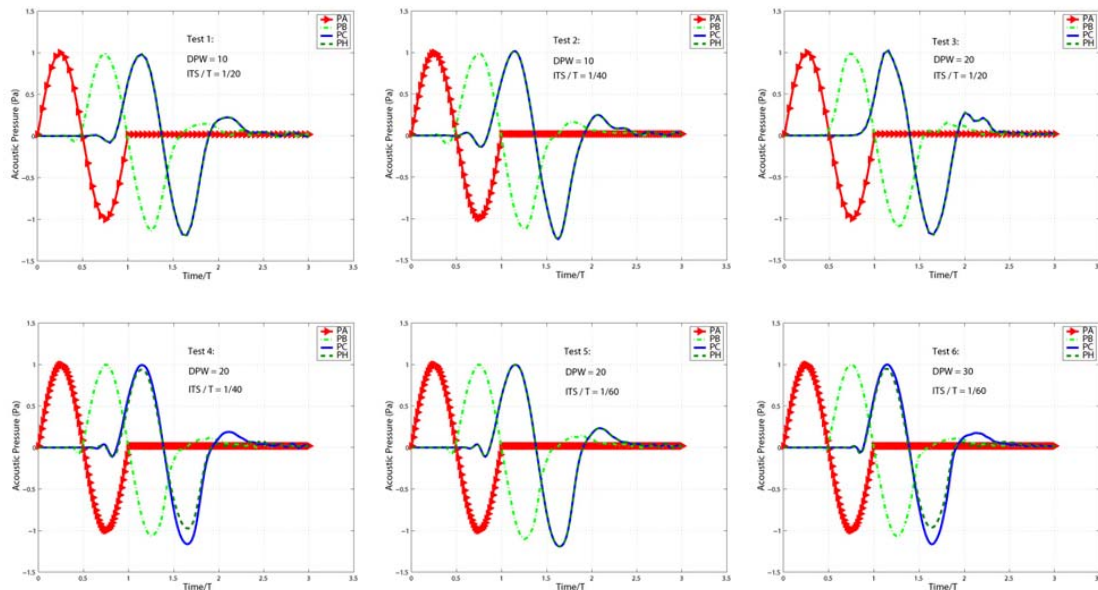


Fig 7: Pressure waveforms at four locations in homogeneous air medium to test DPW and ITS.

DPW (number of mesh divisions per wavelength, affects spatial resolution) and ITS (integration time step size, affects temporal resolution) are two key transient-analysis parameters, and BOUND (FEA absorbing boundary) determines the degree to which the computational domain can be assumed infinite in extent. Note: the incidence plane is located left of the center plane in the computation domain and BOUND affects the degree of the incident wave's distortion along the circular boundary of the computation domain. In general, the mesh must be fine enough to resolve the largest dominant frequency. A general guideline is to have at least 20 elements per wavelength along the propagation direction, that is, $DPW \geq 20$ [ANSYS 6.1 Documentation: Modeling and Meshing Guide]. For transient analysis, using at least twenty points per cycle at the center frequency yields a reasonably accurate solution. ITS is given by $ITS \leq T/DPW$, where $T = 1/f$, f is the center frequency of the incident pulse wave [ANSYS 6.1 Documentation: Transient Dynamic Analysis]. The acoustic absorbing boundary should be located at least 0.2λ from the target object [ANSYS 6.1 Documentation: Coupled-Field Analysis Guide]. Six combinations of DPW, ITS, and BOUND were evaluated (Fig 7). It was observed that, in general, larger BOUND enhanced the computational accuracy (Tests 5 and 6 vs. Tests 1, 2, 3, and 4) while larger DPW and smaller ITS yield insignificant improvements (Test 2 vs. Test 1, Test 3 vs. Test 2). For future transient-analysis cases, $DPW = 20$ and $ITS = T/20$ will be considered to be sufficient and BOUND will be further evaluated in next section.

4.5 2D Elastic Cylinder Shell (Transient Analysis Case)

Table 1 lists the properties for most components of human head [Goss 1978; Goss 1980; Duck 1990]. Among these properties, sound speed in air is the lowest and thus the wavelength in air is the shortest. Thus, for the FEA model, the smallest elements used are based on propagation in air. In the other words, to achieve good computation resolution for the mixed-property model, the optimal parameters for simulating propagation in air is determined.

Table 1: Material Properties in the Human Head

Material	Speed of Sound (m/s)	Densities (kg/m ³)
Air	340	1.2
Water	1500	1000
Soft Tissues	1520-1580	980-1010
Lipid-based tissues	1400-1490	920-940
Collagen-based tissues	1600-1700	1020-1100
Aqueous humor	1002-1006	1500
Vitreous humor	1090	1530
Blood	1580	1040-1090
Brain – grey	1532-1550	1039
Brain – white		1043
Skull - compact inner and outer tables	2600-3100	1900
Skull - Spongy diploe	2200-2500	1000

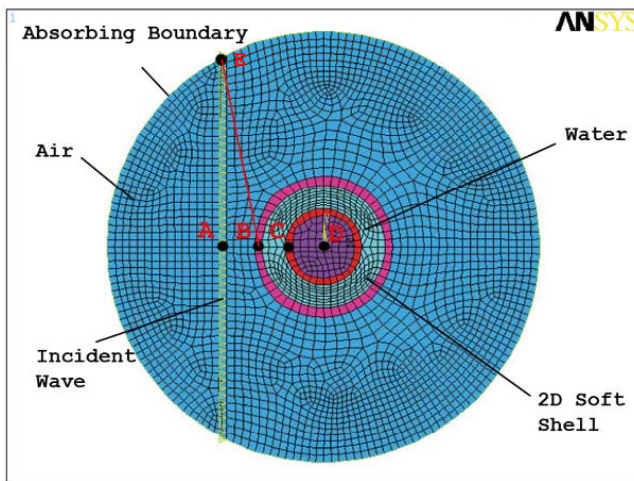
A transient analysis on a 2D soft (elastic) cylinder shell is evaluated that includes property values of typical tissue (Table 1). Figure 8 shows the 2D soft shell geometry used in the simulations. In this case study, the effect of BOUND is further evaluated in object spaces of different sizes.

Two 2D cylinder shell tests were carried out with different FEA absorbing boundary on the same target object (Table 2).

Evaluation of Acoustic Propagation Paths into the Human Head

Table 2: Small 2D Shell Transient Analysis

Parameters	Test 1	Test 2
Shell inner radius (a)	$0.1 \lambda_{\text{air}}$	$0.1 \lambda_{\text{air}}$
Shell thickness (d)	$0.025 \lambda_{\text{air}}$	$0.025 \lambda_{\text{air}}$
BOUND	$a + d + 0.5 \lambda_{\text{air}}$	$a + d + 2.5 \lambda_{\text{air}}$
Incident wave position	$-(a + d + 0.2) \lambda_{\text{air}}$	$-(a + d + 1.2) \lambda_{\text{air}}$
Distance between B and E (Fig 8)	$0.57 \lambda_{\text{air}}$	$2.56 \lambda_{\text{air}}$



Parameters:

$f = 3 \text{ kHz}$

Outer medium: air ($c_{\text{air}} = 340 \text{ m/s}$,

$\lambda_{\text{air}} = 0.113 \text{ m}$, $\rho_{\text{air}} = 1.2 \text{ kg/m}^3$)

Inner medium: water ($c_{\text{water}} = 1500 \text{ m/s}$,

$\rho_{\text{water}} = 1000 \text{ kg/m}^3$)

Soft Shell: ($c_{\text{shell}} = 2800 \text{ m/s}$,

$\rho_{\text{shell}} = 1900 \text{ kg/m}^3$)

$\phi_{\text{inc}} = 0^\circ$

$P_{\text{inc}} = 1 \text{ Pa}$

DPW = 30

ITS = $T/30$

Computational length = $5 T$

BOUND = variable

Fig 8: The soft shell 2D FEA model generated in ANSYS.

Figure 9 shows x-axis acoustic pressure waveforms at locations A, B, C and D (Fig 8). The circular shape of the absorbing boundary causes distortion of the propagated wave along the boundary, and thus induces additional artificial incident waves along non +x-axis direction. When the absorbing boundary is positioned at a greater distance from the object, less undesirable interference reaches the target object in the same computational length scale. The distance between B and E (E is at the absorbing boundary; Fig 8) in Table 2 provides a rough estimate of the shortest possible time required for the interference wave to reach B from the absorbing boundary. Comparing the first period of the pressure waveform at B, the waveform in Test 2 is more symmetric than in Test 1. A reasonable explanation is that it requires about $0.5T$ for the interference wave at E to reach B in Test 1 whereas it requires about $2.5T$ in Test 2. It is also suspected that the significant variation of the pressure waveform shape at C in Test 1 at about $1T$ after the incident wave reaches C is due to the interference wave from the absorbing boundary reaching the target shell cylinder by that time. In Test 2, this interference is smaller, but still present, because it takes longer for the interfering waves at a further absorbing boundary location to reach the shell cylinder.

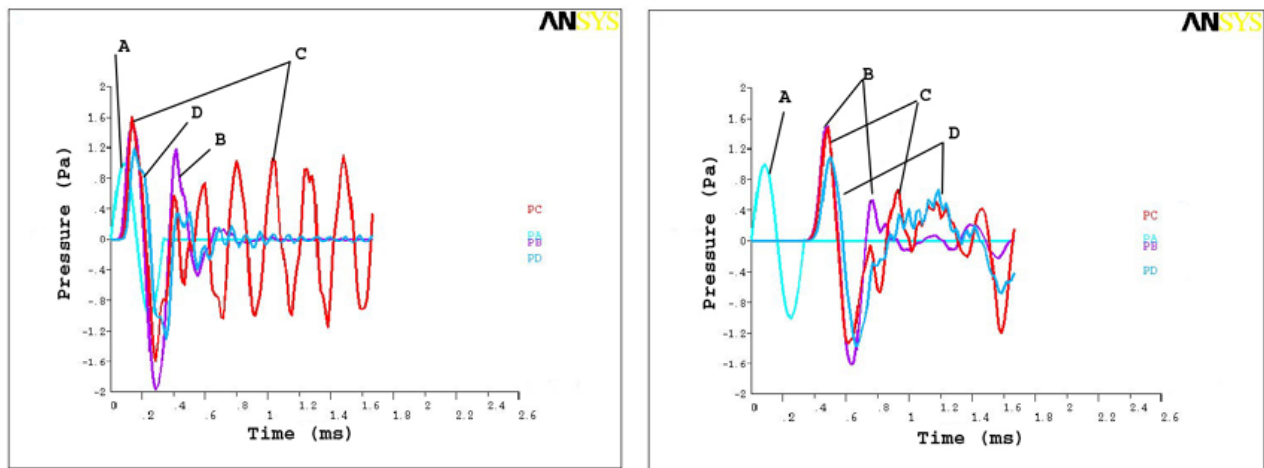


Fig 9: Acoustic pressure waveforms in Test 1 (a, left) and Test 2 (b, right)

To further demonstrate this observation, two large 2D shell cylinder tests were carried out with different FEA absorbing boundary (Table 3). In Test 3, the absorbing boundary is located the same distance from the shell cylinder as in Test 1. And in Test 4, the absorbing boundary is moved further away than in Test 3 but not as far as in Test 2.

Table 3: Large 2D Shell Transient Analysis

Parameters	Test 3	Test 4
Shell inner radius (a)	$0.58 \lambda_{\text{air}}$	$0.58 \lambda_{\text{air}}$
Shell thickness (d)	$0.09 \lambda_{\text{air}}$	$0.09 \lambda_{\text{air}}$
BOUND	$(a + d + 0.5) \lambda_{\text{air}}$	$(a + d + 1) \lambda_{\text{air}}$
Incident wave position	$-(a + d + 0.2) \lambda_{\text{air}}$	$-(a + d + 0.4) \lambda_{\text{air}}$
Distance between B and E (Fig 8)	$0.81 \lambda_{\text{air}}$	$1.34 \lambda_{\text{air}}$

Figure 10 shows x-axis acoustic pressure waveforms at locations A, B, C and D (Fig 8). In Test 4, an abnormal peak pressure amplitude occurs at C about 1T after the incident wave reaches this position, which is not observed in the other tests. Again it is very likely due to the interfering waves coming from the absorbing boundary. Similar to the previous small shell cylinder case, more symmetric pressure waveforms for the first period are found in Test 4. In both Tests 3 and 4, significant variation of the pressure waveform shape at B, C and D are observed about 1T after the incident wave reaches that position, which implies the absorbing boundary is not sufficiently further away from the shell cylinder. Based on the observations (§4.2 and §4.3), it is concluded that FEA absorbing boundary has a significant impact on the computation accuracy and larger BOUND is preferred to decrease the interference due to the circular absorbing boundary and to better simulate the incident plane wave.

Evaluation of Acoustic Propagation Paths into the Human Head

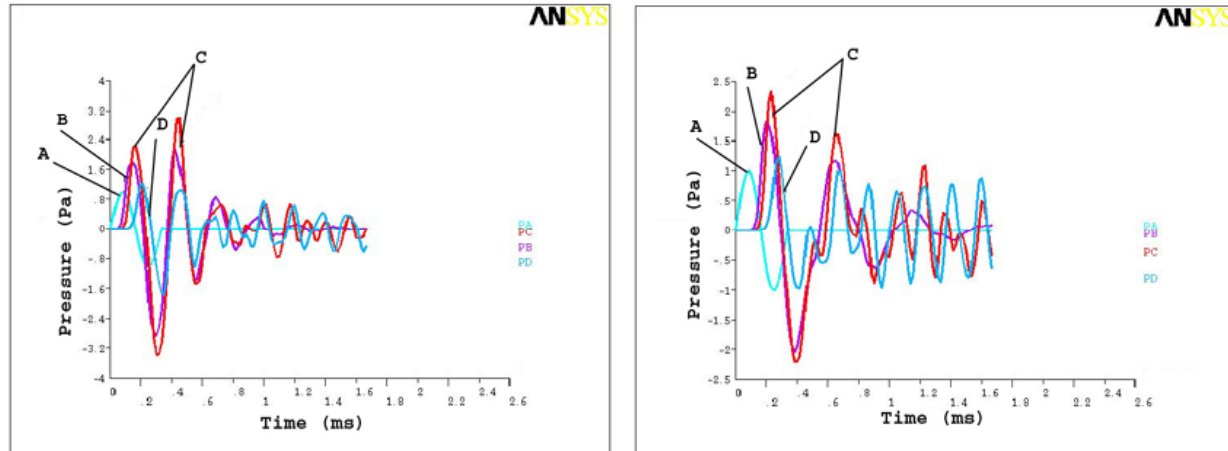
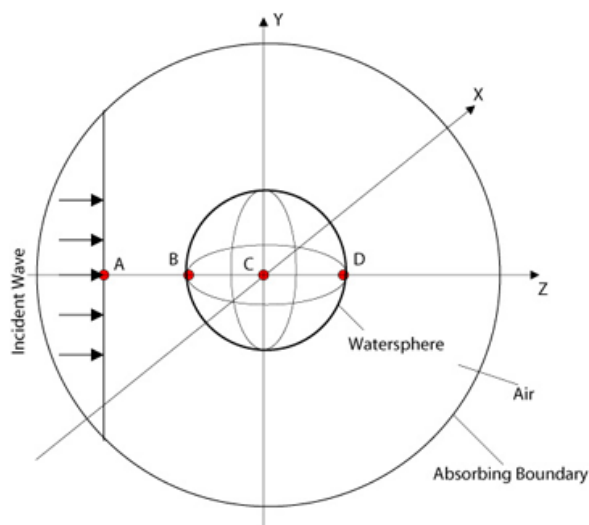


Fig 10: Acoustic pressure waveforms in Test 3 (a, left) and Test 4 (b, right)

4.6 3D Elastic (Water) Sphere (Transient Analysis Case)

In this case study, transient analysis is conducted on a 3D sphere of water that is submerged in air (§3.4), as shown in Fig 11. The acoustic pressure waveforms are shown at different locations (Fig 11) in Fig 12. The locations are A (0, 0, -0.9 λ), B (0, 0, -0.4 λ), C (0, 0, 0), D (0, 0, 0.4 λ). From the pressure waveforms it is determined that about 0.5 T is required for the incident wave to travel from A to B whereas the travel time is around 0.1 T from B to C. These travel time estimates agree with theory and thus validates the temporal accuracy of the FEA.



Parameters:

$$f = 3 \text{ kHz}$$

$$\text{Air: } c_{\text{air}} = 340 \text{ m/s, } \lambda_{\text{air}} = 0.113 \text{ m, } \rho_{\text{air}} = 1.2 \text{ kg/m}^3$$

$$\text{Water sphere: } c_{\text{water}} = 1500 \text{ m/s, } \rho_{\text{water}} = 1000 \text{ kg/m}^3$$

$$a = 0.4 \lambda_{\text{air}}$$

$$\text{BOUND} = 1.3 \lambda_{\text{air}}$$

$$X_{\text{inc}} = 0.9 \lambda_{\text{air}}$$

$$\phi_{\text{inc}} = 0^\circ$$

$$P_{\text{inc}} = 1 \text{ Pa}$$

$$\text{DPW} = 20$$

$$\text{ITS} = T/20$$

Fig 11: The 3D water sphere FEA model.

Furthermore, the instantaneous acoustic intensity was calculated based on the FEA acoustic pressure data from

$$I_A = \frac{p_A(t)^2}{\rho_{\text{air}} c_{\text{air}}}, \quad I_C = \frac{p_B(t)^2}{\rho_{\text{water}} c_{\text{water}}}, \quad (8)$$

where $p(t)$ are the peak values of the acoustic pressure at their respective locations. The instantaneous acoustic intensity insertion loss (IL) from air to water was estimated from

$$IL_{\text{dB}} = 10 \log_{10} \left(\frac{I_{\text{peak,C}}}{I_{\text{peak,A}}} \right). \quad (9)$$

From the FEA data (Fig 12), the intensity loss from air to water was estimated to be -34 dB, a reasonable value relative to a theoretical loss estimate (-33 dB) for normal incidence at a planar air-water boundary.

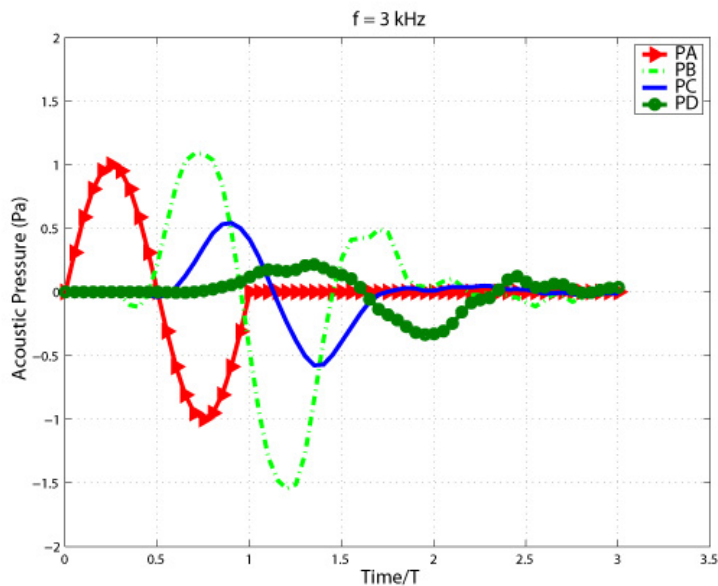


Fig 12: Acoustic pressure waveforms for the water sphere case at locations A, B, C, and D (Fig 11).

5.0 HUMAN-HEAD RELATED STUDIES

5.1 NIH Human Head

Transient FEA procedures of the 2D NIH human head (National Library of Medicine's Visible Human Project, National Institutes of Health, Bethesda, MD) have been constructed, simplified and verified. The male anatomic dataset was used. The anatomic dataset consists of 0.33-mm-wide transverse sections of the head, with each section 2048 pixels by 1216 pixels and each pixel 8-bit RGB scale. A simplified 2D human head analysis was conducted using one of the 0.33-mm-wide transverse sections. The first challenge (computational, not scientific) was to convert the NIH human head digital raster image dataset into a vector format file for import to ANSYS that uses IGES-format vector data. Then, the outer surface of the head was segmented to yield only the head contour. Skull was modeled as a 1-cm-thick layer immediately inside the human head contour. The interior region was modeled as water (Fig 13).

Evaluation of Acoustic Propagation Paths into the Human Head

Transient (one-cycle sine wave) analyses were performed at 4 frequencies (0.125, 1, 3 and 10 kHz) and two incident angles (0° : towards the right side, and 45° : approximately towards the right cheek) (Fig 14).

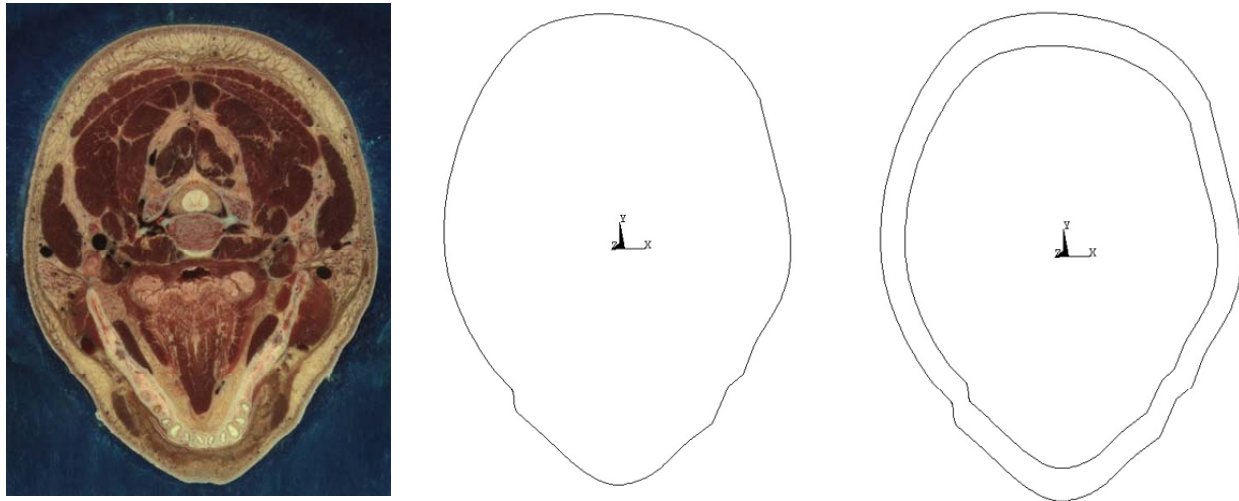
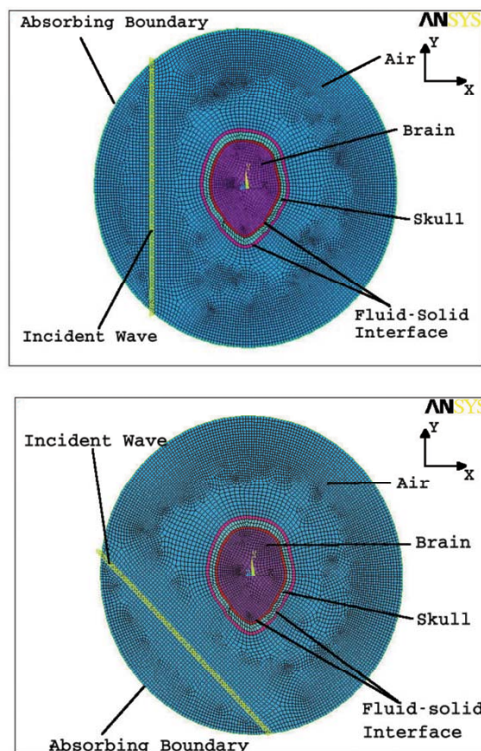


Fig 13: Development of two-dimensional geometric model of human head based on anatomic image.

The instantaneous acoustic pressure waveforms were recorded at 4 locations, A, C, F and H (Fig 15), which are at approximately the left and right ear locations near each side of the skull, and shown in Fig 16. Validation of the process was accomplished by estimating the acoustic intensity loss across the skull from the instantaneous acoustic pressure waveforms on each side of the skull (one location in air and the other location in water). The acoustic loss (Eqns 8 & 9; §4.5) across the skull was estimated from the FEA data to be approximately 26 dB, reasonably consistent with theoretical estimates (33 dB) considering this is a 2D analysis. The instantaneous acoustic intensities are shown in Fig 17 for two tests with same incident frequency (3 kHz) and the two incident angles (0° and 45°).

5.2 Analyze Human Head

The transient FEA procedure of the 3D Analyze (Mayo Clinic, Rochester, MN) human head model has also been demonstrated. The human head model was derived from the 3D Analyze MRI dataset. The Analyze human head outer surface was segmented slice-by-slice as shown in Fig 18.



Parameters:

Outer medium: air ($c_{\text{air}} = 340 \text{ m/s}$, $\lambda_{\text{air}} = 0.113 \text{ m}$,
 $\rho_{\text{air}} = 1.2 \text{ kg/m}^3$)

Inner medium: water ($c_{\text{water}} = 1500 \text{ m/s}$, $\rho_{\text{water}} = 1000 \text{ kg/m}^3$)
 Human skull ($\rho = 1412 \text{ kg/m}^3$, Young's modulus = 6.5 GPa,
 Poisson's ratio = 0.22) [Sauren 1993]

Human head: 15 mm x 20 mm

Skull thickness: 10 mm

BOUND = 7.5 mm + 2.0 λ_{air}

$X_{\text{inc}} = \text{BOUND} - 0.4 \lambda_{\text{air}}$

$\phi_{\text{inc}} = 0^\circ$ or 45° (see Fig 14)

$P_{\text{inc}} = 1 \text{ Pa}$

DPW = 20

ITS = T/20

Fig 14: (a, top) Incident angle $\phi_{\text{inc}} = 0^\circ$. (b, bottom) $\phi_{\text{inc}} = 45^\circ$.

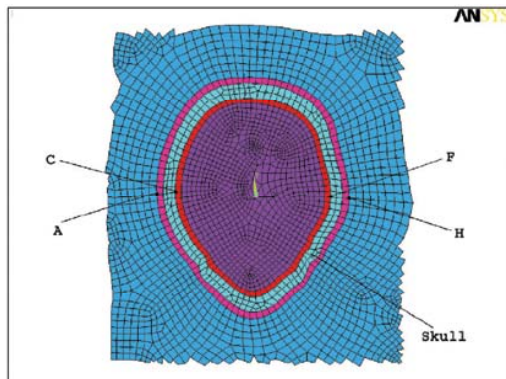


Fig 15: Four locations (A, C, F, H) along inner and outer skull surface which are at approximately the left and right ear locations near each side of the skull.

Evaluation of Acoustic Propagation Paths into the Human Head

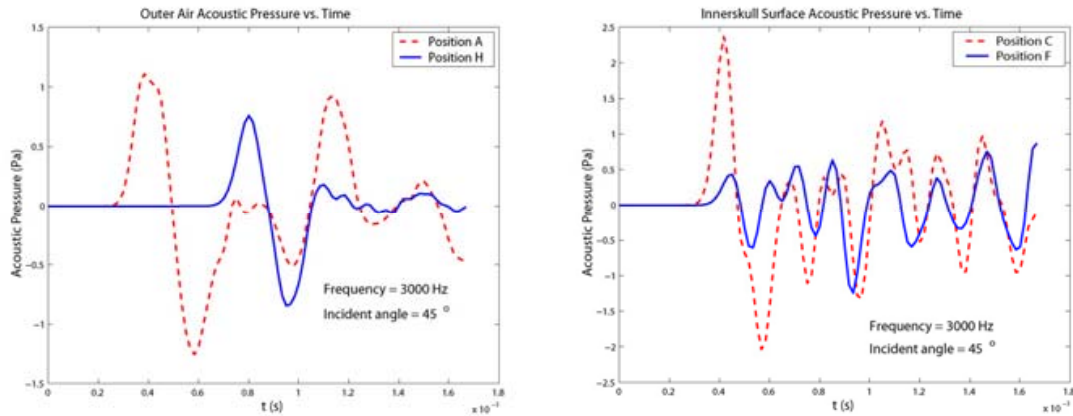


Fig 16: Instantaneous acoustic pressure waveforms recorded at four locations (see Fig 15).

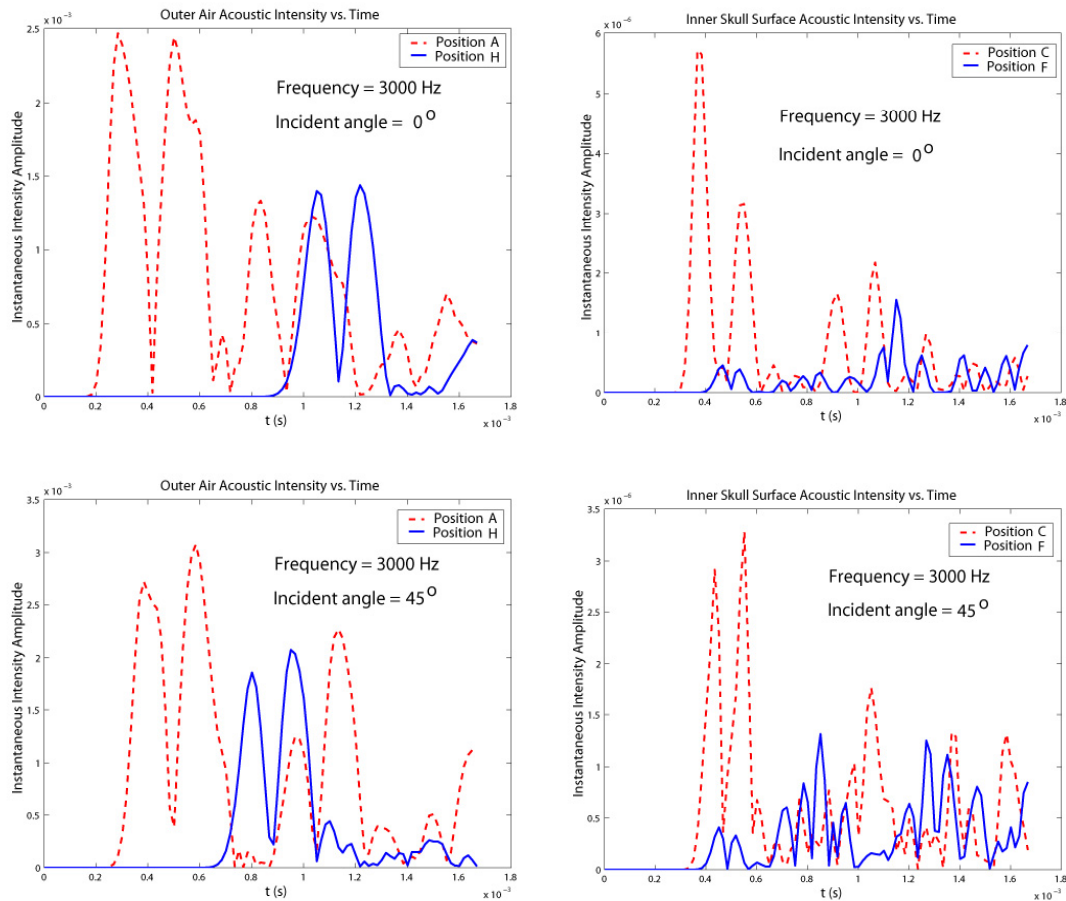


Fig 17: Instantaneous acoustic intensity distributions at a frequency of 3 kHz at four locations (see Fig 15). (top panels) $\phi_{inc} = 0^\circ$ and (bottom panels) $\phi_{inc} = 45^\circ$.

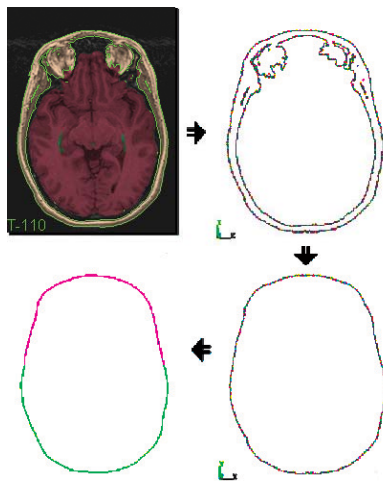


Fig 18: The Analyze human head MRI image slice is segmented to derive the outer head contour.

The Analyze human head model contains a large number of points (order: 10^5). Therefore, this model was further simplified (order: 10) while maintaining the original outer geometry. The material internal to the outer 3D head surface was constructed of skull material properties to demonstrate that it was feasible to quantify the sound pressure amplitude at a location within the head relative to the air-borne sound pressure amplitude. The 3D FEA model is shown in Fig 19.

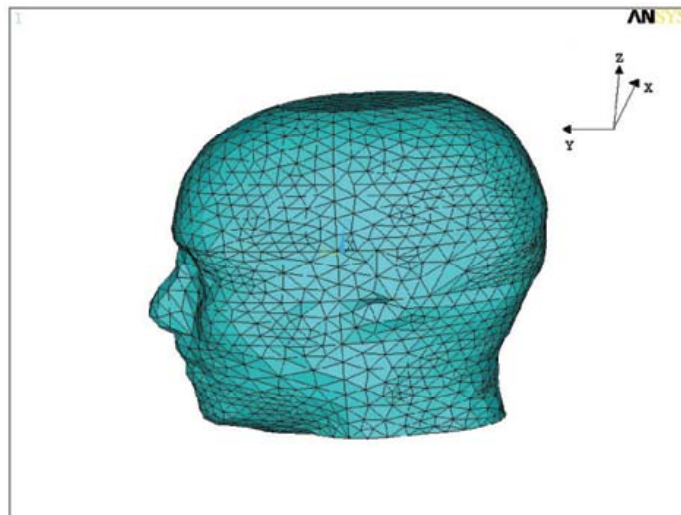


Fig 19: The 3D FEA human head model.

Parameters:

Outer medium: air ($c_{\text{air}} = 340 \text{ m/s}$,
 $\lambda_{\text{air}} = 0.113 \text{ m}$, $\rho_{\text{air}} = 1.2 \text{ kg/m}^3$)
 Inner medium: water ($c_{\text{water}} = 1500 \text{ m/s}$,
 $\rho_{\text{water}} = 1000 \text{ kg/m}^3$)
 Human skull ($\rho = 1412 \text{ kg/m}^3$,
 Young's modulus = 6.5 GPa,
 Poisson's ratio = 0.22)
 Human diameter: $\sim 0.18 \text{ m}$
 $\text{BOUND} = 0.09 \text{ m} + 0.9 \lambda_{\text{air}} = 0.192 \text{ m}$
 $X_{\text{inc}} = \text{BOUND} - 0.4 \lambda_{\text{air}} = -0.1467 \text{ m}$
 $P_{\text{inc}} = 1 \text{ Pa}$
 $\text{ITS} = T/20$

Admittedly this is an unrealistic realization of the head but one that can be evaluated in three dimensions and allows for the evaluation quantitatively of propagation into skull material. A transient 3-kHz one-cycle sine wave was incident from air onto the simple human head model and acoustic pressure waveforms were obtained at three locations (Fig 20). The instantaneous acoustic intensity peak (Eqn 8; §4.5) in air was determined to be about 2.7 mW/m^2 and the acoustic intensity in the skull bone was about $1.24 \text{ } \mu\text{W/m}^2$. The acoustic loss (Eqn 9; §4.5) across the 3D skull model surface was estimated to be approximately 32 dB, quite consistent with theoretical estimates (33 dB).

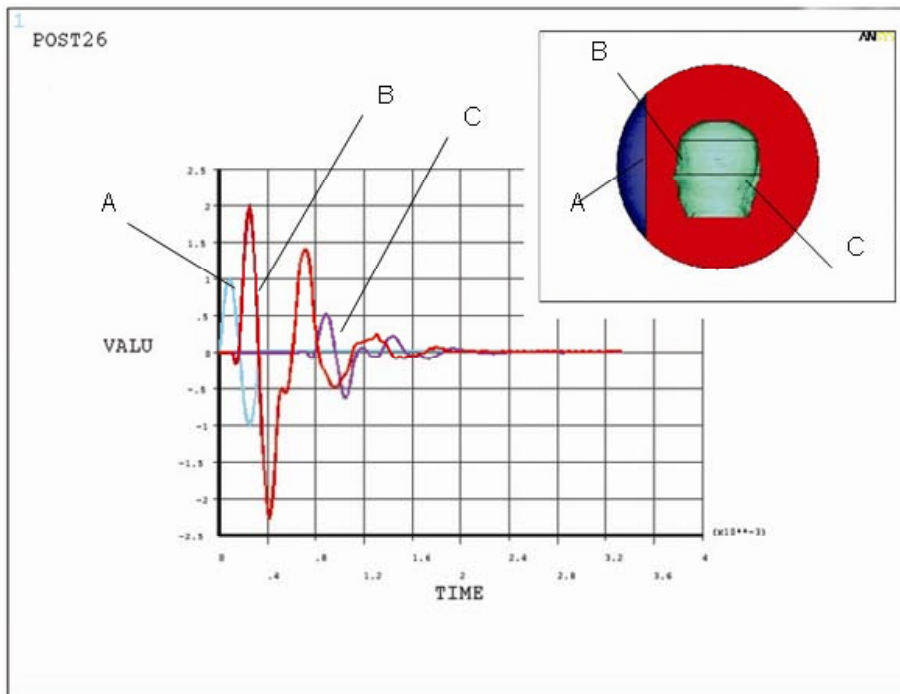


Fig 20: Acoustic pressure waveforms at three selected locations for the 3D FEA human head model.

These human head studies identified a problem that we knew we were going to have: how to represent graphically and quantitatively a 3D process? This work with the simpler Analyze human head model provided needed insight to represent a 3D process. Therefore, while the work continues to develop a 3D human head from the more complete NIH dataset, a procedure was initiated to represent graphically and quantitatively a 3D transient process.

5.3 Ray Tracing

A ray tracing approach to represent graphically and quantitatively a 3D transient process has been developed. Ray tracing is a common procedure by which wave propagation is displayed. For example, in the study of optics, rays are used to depict the path or paths taken as a light wave travels through a lens. However, in optics, the eikonal equation can be solved because the wavelength is assumed to be zero so that propagation laws can be formulated in terms of geometry. This is also the case for geometric acoustics and is often used to solve acoustic propagation problems in the ocean. However, for the case of an acoustic wave that has a wavelength comparable to the object onto which it is incident, the full wave equation must be solved because diffraction needs to be included as part of the analysis. Therefore, ray paths need to be deduced from the propagated acoustic wavefront. The transient ANSYS FEA procedure yields data from which the propagated acoustic wavefront has been deduced. From the propagated acoustic wavefront, rays have been calculated by taking the normal components of the wavefront surface as a function of time. The transient FEA procedure yields data as a function of time steps, thus allowing the wavefront to be animated as a function of time. Also, from these data, as long as the time steps are sufficiently small, ray paths can be calculated. Further, after the ray paths have been calculated, the acoustic energy can be determined from the density (number of ray paths that intersect a unit volume) of ray paths in a particular volume.

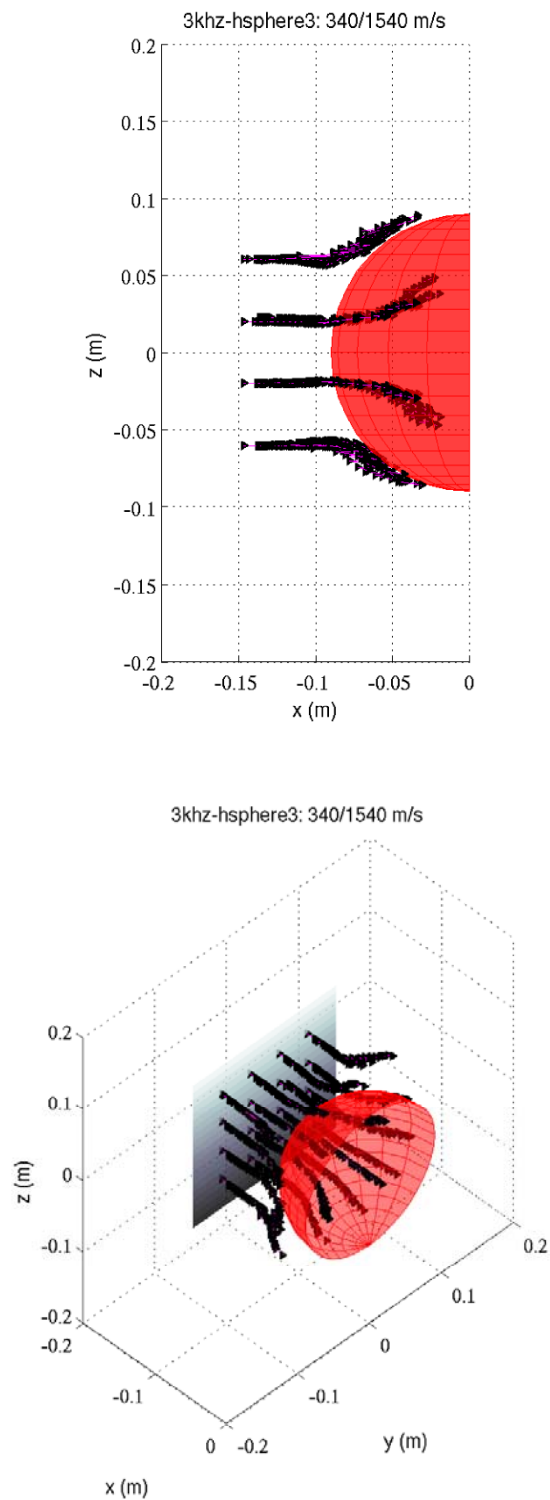


Fig 21: Ray paths calculated based on FEA data.

Evaluation of Acoustic Propagation Paths into the Human Head

Transient (3-kHz one-cycle sine wave) ANSYS FEA procedures were performed using a 9-cm-radius hemisphere of known propagation speeds. The plane wave was incident in air (speed = 340 m/s) onto the spherical surface of the hemisphere that had propagation speeds consistent with water (speed = 1540 m/s). Wavefronts were deduced from a time-domain correlation technique. From the wavefronts, ray paths were calculated (Fig 21 depicts ray paths). The ray paths originated in air and propagated into the hemisphere. For low-angle incidence, the ray path directions into the hemisphere were consistent with those calculated from Snell's Law. However, as the angle of incidence onto the hemisphere increased, the ray path directions into the hemisphere became progressively greater from those expected from a Snell's Law calculation. We are currently investigating the cause of this phenomenon. We suspect that the ray path directions are correct because diffraction phenomenon could be causing this problem. For this example, the wavelength in air is 11.3 cm (radius of the hemisphere is 9 cm) and ka is 5.0 (all based on a center frequency of 3 kHz but the frequency content of the one-cycle wave is much broader). To test our hypothesis, we conducted the transient ANSYS FEA analysis at 10 kHz using the 9-cm-radius hemisphere. However, at the high frequency of 10 kHz, we hit the current ANSYS version limits on elements/nodes number and a coarse meshing is used in the FEA model that compromised the computation accuracy. Comparing the computational results and the theoretical solutions calculated using Snell's law, up to 0.7° difference is found for low-angle incidence while up to 10° difference is found for larger-angle incidence (Fig 22). The next step is to conduct the FEA analysis using planar surfaces as a function of incident angle. A well-documented and verified ray tracing procedure will provide an easy to understand (and visualize) dynamic process. Also, the budget requests funds to support partially the unlimited ANSYS version.

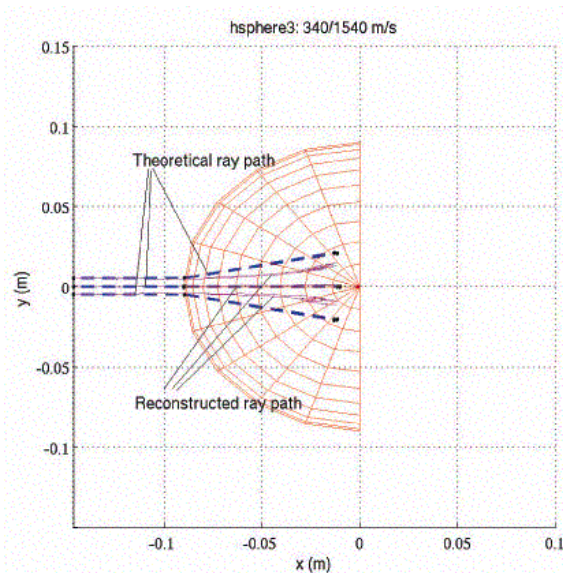


Fig 22: Ray path for the hemisphere model.

Figure 23 shows the wavefront at three time steps as the wavefront propagates through the hemisphere. If the shown arrows were connected at each wavefront step, the resultant lines would depict ray paths.

In summary, we have (1) validated finite-element analysis (FEA) general processing code for both harmonic and transient solutions, (2) constructed, simplified and verified transient FEA analyses of the 2D NIH human head, (3) demonstrated FEA analysis of the 3D Analyze human head, and (4) developed a ray tracing approach to graphically and quantitatively represent a 3D transient process.

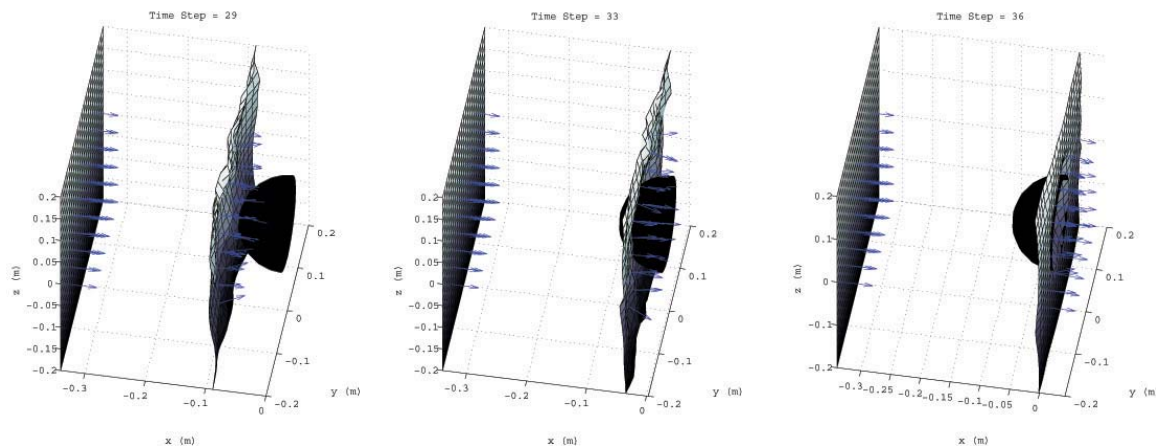


Fig 23. Three time steps as the wavefront propagates through the hemisphere.

6.0 ACKNOWLEDGMENT

Support for this work from the US Air Force Office of Scientific Research through US/AFOSR Award # F 49620-03-1-0188 is acknowledged.

7.0 REFERENCES

- [Békésy, 1948] G von Békésy, "Vibration of the head in a sound field and its role in hearing by bone conduction," *Journal of the Acoustical Society of America*, 20, 1948, 749-760.
- [Barany 1938] E. A. Barany, "A contribution to the physiology of bone conduction," *Acta Otolaryngol Suppl*, 26, 1938, 1-223.
- [Duck 1990] F. A. Duck, *Physical properties of tissues: A comprehensive reference book*, New York, NY: Academic Press, 1990.
- [Freeman 2000] S. Freeman, J.-Y. Sichel, H. Sohmer, "Bone conduction experiments in animals -- evidence for a non-osseous mechanism," *Hearing Research*, 146, 2000, 72-80.
- [Goss 1978] S. A. Goss, R. L. Johnston, F Dunn, "Comprehensive compilation of empirical ultrasonic properties of mammalian tissues," *Journal of the Acoustical Society of America*, 64, 1978, 423-457.
- [Goss 1980] S. A. Goss, R. L. Johnston, F Dunn, "Compilation of empirical ultrasonic properties of mammalian tissues - II," *Journal of the Acoustical Society of America*, 68, 1980, 93-108.
- [Guild 1936] S.R. Guild, "Hearing by bone conduction: the pathways of transmission of sound," *Ann Otol Rhinol Laryngol*, 45, 1936, 736-754.
- [Morse 1968] P. M. Morse and K. U. Ingard, *Theoretical acoustics*, New York, NY: McGraw Hill, 1968.
- [OSHA 1910.95]. Occupational noise exposure. Standard number 1910.95, Occupational Safety & Health Administration, U.S. Department of Labor, Washington, DC

Evaluation of Acoustic Propagation Paths into the Human Head

[Pierce 1989] A. D. Pierce, *Acoustics: An introduction to its physical principles and applications*, Woodbury, NY: Acoustical Society of America, 1989.

[Sauren 1993], A. H. J. Sauren and M. H. A. Classens, “Finite element modeling of head impact: The second decade”, *International IRCOBI Conference on Biomechanics of Impacts* Eindhoven, 241-254, September 8-10, 1993, Eindhoven, the Netherlands.

[Sohmer 2000] H. Sohmer, S. Freeman, M. Geal-Dor, C. Adelman, I. Savion, “Bone conduction experiments in humans – a fluid pathway from bone to ear,” *Hearing Research*, 146, 2000, 81-88.

[Sohmer 2004] H. Sohmer, S. Freeman, “Further evidence for a fluid pathway during bone conduction auditory stimulation,” *Hearing Research*, 193, 2004, 105-110.

[Stefan 2004] S. Stefan, N. Hato, R. L. Goode, “Fluid volume displacement at the oval and round windows with air and bone conduction stimulation,” *Journal of the Acoustical Society of America*, 115, 2004, 797-812.

[Wever 1954] E. G. Wever and M. Lawrence, *Physiological acoustics*, Princeton, NJ: Princeton University Press, 1954.

[Zwislocki 1957] J. Zwislocki, “In search of the bone-conduction threshold in a free sound field,” *Journal of the Acoustical Society of America*, 29, 1957, 795-804.

Hybrid Feedforward-Feedback Active Noise Control for Hearing Protection and Communication

Laura E. Ray and Jason A. Solbeck
Thayer School of Engineering, Dartmouth College
8000 Cummings Hall
Hanover NH 03755

Robert D. Collier
Sound Innovations, Incorporated
Lebanon NH

Laura.Ray@Dartmouth.edu

ABSTRACT

Research over the past decade has demonstrated substantial increases in noise reduction performance for circumaural hearing protectors through feedforward active noise reduction (ANR) based on least mean square (LMS) methods. However, commercialization of feedforward ANR hearing protection devices has yet to occur. This paper explores issues related to robust realization of feedforward ANR for hearing protection. Specifically, the dynamic range of noise sources, the potential for leakage around circumaural earcups, vibration conditions of high noise environments, and the need for communication requires ANR algorithms that are robust to large variations in the acoustical dynamics of hearing protectors. To meet this need, a hybrid ANR architecture that exhibits excellent stability margins and performance for both stationary and non-stationary noise sources is presented. The hybrid system is comprised of a Lyapunov-tuned leaky LMS feedforward component and a broadband digital feedback ANR system. The contribution of each component to ANR performance is adjusted by individual feedforward and feedback gain factors, and stability margins are defined by the net increase in these gains that can be accommodated while maintaining system stability. Stability and noise reduction performance of the hybrid system are validated experimentally using an earcup from a commercial circumaural hearing protector and using a prototype active earplug. Experiments are performed through flat plate or manikin testing using a DSP development system. Results demonstrate that stability margins are increased, in some cases, by several orders of magnitude.

1.0 INTRODUCTION

The need for improved hearing protection in the high noise fields present in military applications motivates the use of active noise reduction (ANR). Commercial ANR systems employ feedback control, in which an error microphone serves as the feedback signal from which a cancellation signal is produced. Traditional stability-performance tradeoffs that pose limitations on feedback control of noise are evident within commercial circumaural headsets; in feedback ANR, the cavity resonant behaviour forces low feedback gains and thus lower levels of active attenuation. Moreover, feedback circuitry often adds to the noise in the mid-frequency speech communication bands. Practical limitations in ANR performance and in improving speech intelligibility through feedback ANR have generated research in feedforward ANR using least-mean-square (LMS) adaptive filters [1-9]. These computationally simple filters are used widely for stationary noise cancellation. However, LMS filters have stability and performance tradeoffs caused by non-stationary, impulsive noise, finite precision, and measurement noise. Noise fields found in

Ray, L.E.; Solbeck, J.A.; Collier, R.D. (2005) Hybrid Feedforward-Feedback Active Noise Control for Hearing Protection and Communication. In *New Directions for Improving Audio Effectiveness* (pp. 16-1 – 16-10). Meeting Proceedings RTO-MP-HFM-123, Paper 16. Neuilly-sur-Seine, France: RTO. Available from: <http://www.rto.nato.int/abstracts.apc>.

military settings – tanks, fixed and rotary winged aircraft, and environments served by ground crew –vary with operating conditions, posing significant noise reduction challenges.

References 7-8 introduced the Lyapunov-tuned LMS filter (LyLMS), developed in response to environments in which noise sources exhibit temporal variations over a large dynamic range. The Lyapunov-tuning method provides a *time-varying* leakage factor and adaptive step-size combination that optimizes LMS filter stability and noise reduction performance in response to time-varying SNR. The algorithm incurs minimal additional computation over a traditional leaky LMS filter, and it has the major advantage in eliminating empirical tuning. The Lyapunov-tuned leaky LMS filter can be further augmented with a filtered-X implementation, in which a non-constant cancellation path gain is accommodated [6].

Recent studies consider hybrid ANR for hearing protection [9-11]. In [10], a complex filtered-X LMS filter is combined with a commercial analog feedback controller to attenuate infrasonic noise due to the 17.7 Hz fundamental blade passage frequency in a helicopter and its harmonics. While the feedback system, combined with passive attenuation provided by the earcup, provides 20 dB broadband performance, reduction of the fundamental blade frequency is minimal. Addition of feedforward ANR reduces the fundamental blade passage frequency by an additional 20 dB. Reference [11] evaluates a hybrid system in both a reverberant sound field and a directional sound field, and explores the relation between ANR performance and forward path delay. The results show improved performance through hybrid ANR in a reverberant field, as compared with analog feedback alone, while a directional field, which affects the acoustic delay, degrades performance when the noise source does not face the reference microphone directly. In [9], two additional aspects of hybrid ANR are evaluated. A hybrid feedforward-feedback control architecture is shown to (1) improve noise reduction performance for nonstationary noise sources, and (2) increase stability margins. A digital feedback system provides low level, broadband performance, independent of the noise source. The feedforward system acts on the resulting error signal to further increase noise attenuation. Unlike previous studies, where feedforward ANR is hybridized with a commercial analog controller, [9] develops a broadband digital feedback compensator designed to work with the feedforward system. The presence of this feedback compensator increases the feedforward gain stability margin substantially and reduces sensitivity of overall performance to the temporal characteristics of the noise source.

These references show the benefit of hybrid ANR in retaining the positive characteristics of feedback control of providing good broadband performance irrespective of the noise source as well as the positive characteristics of feedforward control of achieving good tonal noise attenuation. Hybridization allows the use of lower feedback gain without sacrificing overall performance, providing positive benefits on stability margins. In this paper, the concept of a general hybrid ANR architecture is applied to two hearing protection devices, including circumaural earcup and an earplug.

2.0 HYBRID FEEDFORWARD-FEEDBACK ANR

Figure 1 shows the hybrid ANR system block diagram. The incoming noise $X(t)$ is measured by a microphone embedded on the exterior of the hearing protector. The past L samples of $X(t)$ constitute the reference input X_k , where L is the filter length. Electronic and quantization noise enters as Q_{xk} . $X(t)$ passes through passive materials, an unknown acoustic process, to become noise signal $d(t)$. The LMS filter finds weight vector $W(z)$ modelling the unknown passive response and produces a cancellation signal $-y_k = W^T X_k$. An error microphone inside the hearing

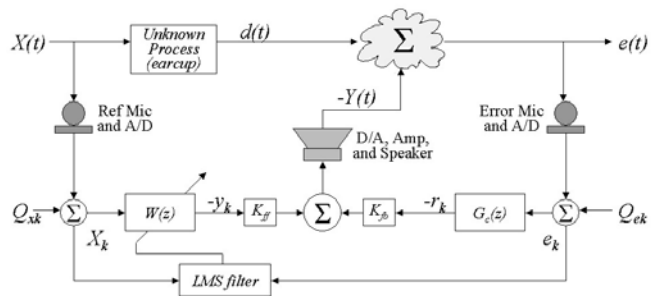
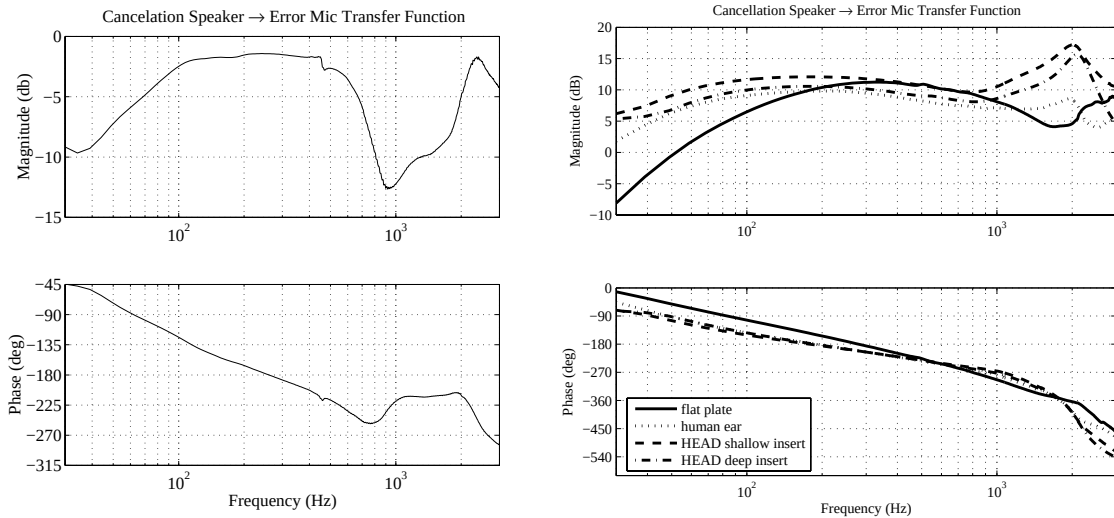


Figure 1: Hybrid ANR topology



(a) commercial circumaural eardrum

(b) earplug

Figure 2: Open loop response from input to cancellation speaker driver to output of error microphone amplifier for (a) eardrum mounted on flat plate and (b) ANR earplug in flat plate, human ear, and manikin.

protector registers the residual noise e_k , subject to measurement noise Q_{ek} . e_k adjusts the LMS filter and also passes through digital feedback compensator $G_c(z)$ creating a feedback cancellation signal $-r_k$. The two cancellation signals are scaled by gains K_{fb} and K_{ff} , summed, and passed to a D/A converter and amplifier to drive the cancellation speaker inside the eardrum as $-Y(t)$. The residual noise, $e(t) = d(t) - Y(t)$, is what is “heard” by the error microphone. In an eardrum, the error microphone is necessarily about one inch from the concha, while in an earplug, the error microphone is generally within the external auditory ear canal.

Figure 2 shows cancellation path responses from an internally generated cancellation signal to the error microphone for a commercial circumaural hearing protector and for several configurations of a commercial earphone modified for hybrid ANR. Both devices have good passive attenuation. Earplug data were collected within a flat plate (enclosing a volume comparable to that enclosed within human ear), on a manikin at two insertion depths, and in a human ear, showing different responses that can be elicited using common test configurations. The general characteristics of the cancellation path for these devices are similar, indicating that a common ANR architecture is viable. However, the variability in the cancellation path necessitates a system with good stability margins, which poses a design challenge for feedback and feedforward ANR individually. The non-constant cancellation path gain can be accommodated using a filtered-X implementation of the Lyapunov-tuned leaky LMS filter, in which either a finite-impulse response or infinite-impulse response filter shapes the reference input prior to the LMS filter update based on the cancellation path response [12]; however, to the extent that this path varies from user to user, the shaping filter needs either to be adaptive or robust to variations in the cancellation path. Similarly, the feedback system must also be robust to variations in the cancellation path response.

The hybrid architecture provides a means to retain performance while building in adequate stability margins in the face of these variations. For feedback ANR, a low-order digital feedback compensator $G_c(z)$ is designed to increase the frequency range in which the phase response is near -180° , such that the internally generated cancellation signal is 180° out of phase with the error signal e_k for a broad frequency range. Unlike commercial feedback ANR, which provides up to 20 dB of narrowband performance, this compensator provides lower level (5-10 dB), broadband attenuation. Since phase characteristics of various hearing protectors are similar, a common compensator structure is selected for feedback ANR. Feedforward ANR is based on the Lyapunov-tuned LMS (LyLMS) feedforward algorithm from [7,8], which is summarized here. $X_k \in \mathbf{R}^n$ is the reference input at time t_k , and is subject to noise Q_{xk} .

The unknown process produces output d_k . The LMS filter seeks a weight vector $W_k \in \mathbf{R}^n$ to minimize the mean-square-error. The optimum or Wiener solution is $W_0 = E[X_k X_k^T]^{-1} E[X_k d_k]$, where $E[\cdot]$ is the expected value and $e_k = d_k - W_k^T X_k$. A recursive update equation $W_{k+1} = W_k + \mu e_k X_k$ comprises the basic LMS filter. Convergence speed and stability depend on the step size μ , and a normalized step size μ_k is introduced to improve convergence. To improve stability in the face of measurement noise, the weight vector update equation is altered through a leakage factor, providing the leaky, normalized LMS (LNLMS) filter $W_{k+1} = \lambda W_k + \mu_k e_k X_k$, where $0 \leq \lambda \leq 1$ for stability. The LNLMS filter forces a stability-performance tradeoff: for high performance λ should be close to unity, whereas stability requires leakage (lower λ). The optimum λ changes with signal strength and spectral composition. An adaptive leakage factor provides a solution to this tradeoff. The Lyapunov tuning method provides an adaptive leakage factor and step size combination to optimize both stability of the weight update equation and noise reduction performance. Tuning laws are detailed in [7,8].

3.0 EXPERIMENTAL EVALUATION

Experimental evaluation of the hybrid feedforward-feedback architecture was conducted in a Low Frequency Acoustic Test Cell (LFATC), described in [13], for the circumaural earcup, and for the earplug, a combination of test cell and manikin testing was performed. The LFATC, shown in Figure 3 has a flat acoustic frequency response from 10 to 200 Hz. Digital equalization extends this range to approximately 1000 Hz. A single earcup is secured over the base plate of the test cell. A 15.2 cm diameter 100 W speaker mounted in the top plate of the cell provides the noise signal and two Brüel & Kjær 4190 Type I microphones mounted through the sidewall and mounted axially in the base plate under the earcup provide source level and error signal, respectively. The microphone under the earcup represents the location of the external opening to the ear canal. Noise floors of these precision microphones average 50 dB in the measurement range 40-1250 Hz.



Figure 3: Low Frequency Acoustic Test Cell



Figure 4: HEAD Measuring System

The HEAD Acoustics Artificial Head Measurement System, shown in Fig. 4, reproduces head related transfer function characteristics in order to further evaluate the hearing protection system objectively at realistic source levels. This system provides an intermediate facility between that of the LFATC and human subject testing for evaluation of active hearing protection. In a quiet room with negligible building vibration and mechanical equipment, the noise floor of the HEAD microphones averages 55 dB.

The earcup was developed in-house from commercial components. Modification of a commercial feedback earcup for feedforward ANR is described in [1]. The passive attenuation of the earcup is 5 dB at 50 Hz, increasing to 30 dB at 800 Hz [9]. The earplug consists of a commercial earphone for military applications modified for ANR. Knowles FG-3329 microphones are introduced for reference and error sensing. A foam tip provides good passive attenuation. Since earphone performance depends on the trapped volume of air, the cancellation path transfer function of the earplug is measured in the test cell (with an appropriate adaptor to insert the earphone into the flat plate above the B&K microphone), on the manikin, and on a human subject. While neither flat plate testing nor manikin testing represent the characteristics of a human ear, to first order, the cancellation path transfer function shown in Figure 2b is qualitatively similar to that of the earplug when inserted into a human ear.

A variety of noise sources were selected for the performance evaluation, depending on the device tested: (1) Individual tones at 1/3-octave band center frequencies, (2) a sum-of-tones signal, (3) F-16 cockpit

noise, (4) Huey helicopter noise, and (5) UH-60A Blackhawk noise. Within the LFATC, signals are band limited 50-800 Hz. These noise sources can be viewed as increasingly less ideal operating conditions. F-16 aircraft noise is similar to band-limited pink noise. However, the two minute noise source recording used in experiments exhibits significant temporal variation as described in [8]. Huey helicopter noise also resembles pink noise, but with the addition of 55 Hz tonal components and associated harmonics and impulsive staccato components in the time domain from the rotor blades passing 10.7 times per second. Blackhawk noise has a similar broadband spectrum, but with high frequency tones due to gear whine and its harmonics. For earcup testing within the LFATC, sources are set to levels between 105 and 110 dB to avoid distortion in the cancellation speaker and saturation of the reference microphone. For earplug testing on a manikin, source levels are set at approximately 100 dB to permit a human to work in test area. All noise levels are reported in dB relative to a 20 μ Pa reference pressure.

The hybrid controller is implemented on a dSPACE DS1103 controller board. The controller board is based on a PowerPC 604e microprocessor running at 400 MHz. The dSPACE board provides on-board 16-bit A/D and 14-bit D/A converters. The hybrid system operates at an update frequency of 10 kHz for the earcup and 15 kHz for the earplug. Anti-aliasing filters are present on all microphone channels, and anti-imaging filters on all speaker driver channels. In the LFATC, performance data are given as the difference between the precision microphone outside of the earcup and the one inside in the base of the test cell, which accounts for the separation path between the noise source and the wearer's ear, or as the difference between the reference and error ANR microphones, which are calibrated to the precision B&K microphones. For earplug testing, performance is measured as the insertion loss at the manikin ear microphone, or as the difference between the reference and error ANR microphones, which are calibrated to B&K microphones.

4.0 EXPERIMENTAL RESULTS

4.1 Commercial Circumaural Earcup

The results for the earcup are summarized from [9]. Figure 5 shows the active attenuation of the sum-of-tones 50-800 Hz, F-16, and Huey helicopter noise sources, as measured by the B&K precision microphone located inside the earcup. *Active Attenuation* is the difference between the noise level with passive attenuation only and the noise level with both passive and active attenuation. For all three noise sources, the feedback system has low level (5-10 dB) but high bandwidth noise reduction capabilities. For tonal noise, the feedforward system performs exceptionally well in the range 80-400 Hz, with diminished performance above and below that range. Whereas both systems have only moderate to good attenuation below 100 Hz, the combined system is able to provide approximately 30 dB of active attenuation at these frequencies. Combining the two independent systems has resulted in performance that is greater than the sum of its parts, with total attenuation of the sum-of-tones of 36 to 51 dB in the 40-800 Hz band.

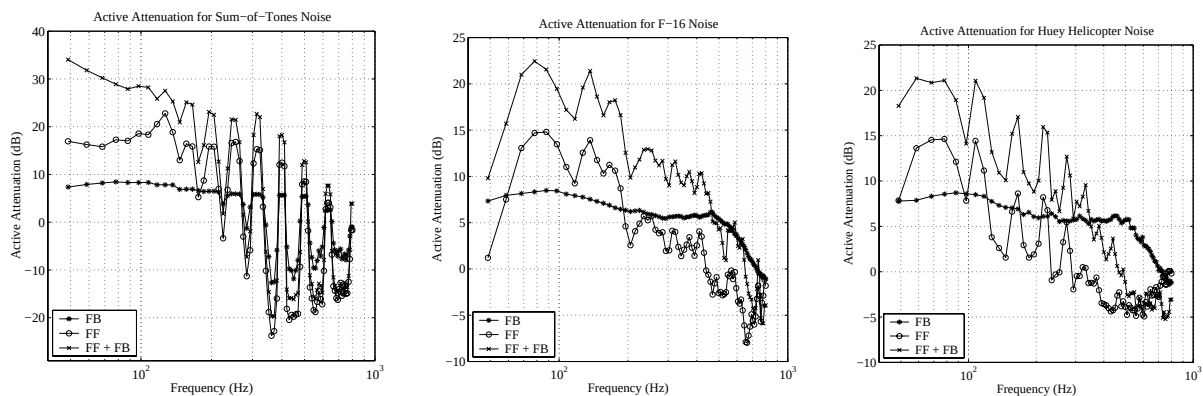


Figure 5: Active Attenuation performance of the hybrid ANR for a commercial circumaural earcup (a) Sum-of-Tones noise at 110 dB, (b) F-16 cockpit noise at 110 dB, and (c) Huey helicopter noise at 105 dB.

Table 1: Summary of performance for commercial earcup modified for hybrid ANR

Noise Source	Average Noise Level (dB)					Total Attenuation (dB)			Active Attenuation (dB)		
	Source	Passive	Feedb.	Feedf.	Hybrid	Feedb.	Feedf.	Hybrid	Feedb.	Feedf.	Hybrid
Sum-of-tones 50 - 800 Hz	110	98	90	81	71	20	29	40	8	17	27
F-16 Cockpit 50 - 800 Hz	110	95	88	86	78	23	25	32	8	10	17
Huey Cockpit 50 - 800 Hz	105	94	86	84	76	19	22	30	8	11	18

Total noise reduction, which include passive attenuation, causes the error microphone signal to approach its noise floor, thus the noise reduction performance approaches its physical limits. In this case, the feedforward and feedback gains assume one value for all frequencies, and the filtered-X algorithm is not implemented. Nevertheless, the average source level of 110 dB is successfully reduced to 71 dB.

The F-16 cockpit noise most closely resembles band-limited pink noise in that it contains no purely tonal content and, over long time periods, has a fairly uniform spectral component. However, during short periods of time its spectral content shifts considerably, which presents problems for traditional LMS filters. Despite this, the hybrid system provides an average active attenuation of 17 dB (32 dB total attenuation), reducing the 110 dB source level to 78 dB. Once again, the hybrid system has substantially greater performance than either of the independent systems acting alone, particularly for frequencies less than 200 Hz. Additionally, whereas the feedforward system alone *added* noise for frequencies above 500 Hz, the hybrid system largely avoided adding noise in the 50-800 Hz band.

Huey helicopter noise contains broadband nonstationary components like the F-16 noise, but also has a tonal component following a 55 Hz fundamental attributed to the tail rotor, and a temporal component due to blade passage that resembles a periodic broadband impulse. In order to keep the periodic impulse from forcing the ANR systems to over-drive the cancellation speaker, the source level is reduced to 105 dB. Active attenuation results are shown in Figure 5c. Once again, the addition of the feedback system to the feedforward system significantly improved the low-frequency attenuation, in this case by 5-10 dB. The tonal component is eliminated by both the feedforward and hybrid system, but is largely untouched by feedback ANR. The feedback system is unsuccessful in removing the temporal component of the helicopter noise; the feedforward system cannot completely remove it either. In contrast, the combined hybrid system is able to almost completely remove the periodic *thwt*, leaving behind a broadband background noise at an average level of 77 dB. Table 1 summarizes these performance results, showing average source, passive, active, and total noise reduction performance for each noise source for this earcup.

The hybrid system not only improves active performance, but it also improves the gain margin of the individual systems. In the feedback system, increasing the path gain, K_{fb} , generally increases the feedback attenuation. However, the gain that provides maximum noise attenuation is close to the threshold of instability, forcing a common stability-performance tradeoff. In a similar way, without a filtered-X LMS, there is a maximum feedforward gain, K_{ff} , above which the weight vector grows without bound, or overexcites certain frequencies. When the two systems are combined, both gains can be increased to levels that otherwise would cause instability, and the increased gain allows for higher overall active attenuation. Alternatively, the hybrid system can provide the same attenuation levels as individual feedforward and feedback components with larger stability margins. Larger stability margins provide a margin against uncertainty in the earcup transfer function or speaker performance. For the feedback system, adding the feedforward system allows K_{fb} to be increased by approximately 20% before instability reoccurs. However, the increased stability is

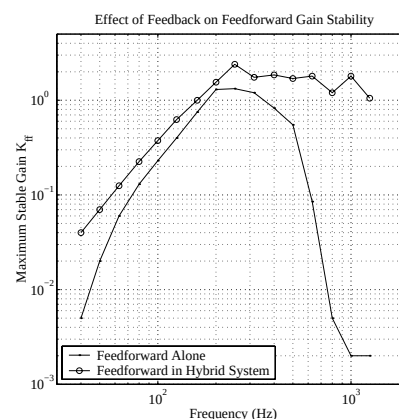


Figure 6: Maximum stable gain (K_{ff}) of the feedforward system for earcup

most notable in the feedforward system. Figure 6 shows the *maximum stable* (not necessarily optimal) feedforward gain K_{ff} , as determined experimentally, as a function of frequency. Figure 6 shows that by augmenting the feedforward system with feedback, the maximum stable K_{ff} can be increased by orders of magnitude at some frequencies.

4.2 Prototype ANR Earplug

Figure 7 shows the ANR earplug developed for evaluation of the hybrid architecture, with the earplug inserted in the left ear of the HEAD manikin. An unmodified communication earplug is inserted in the right ear. Manikin test results are reported for tonal noise, sum-of-tones at 1/3 octave band center frequencies 80-1250 Hz, and UH-60 helicopter cockpit noise. Passive and total ANR performance is measured as the insertion loss between the unoccluded ear and the occluded ear. In addition, total performance is provided as measured by the difference between the ANR reference microphone and the ANR error microphone, both of which are calibrated to the B&K microphones within the LFATC.



Figure 7: ANR earplug on manikin

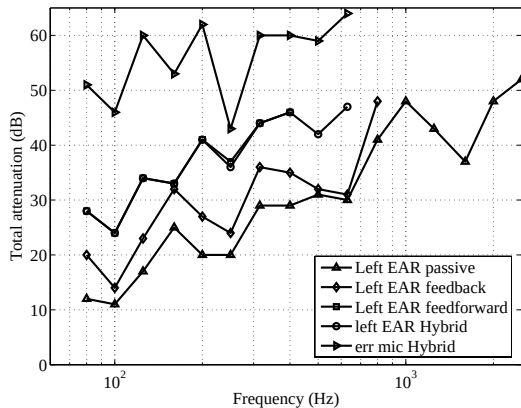


Figure 8: Pure tone performance of ANR modified earplug as measured by insertion loss at HEAD in-ear microphone and between ANR microphones.

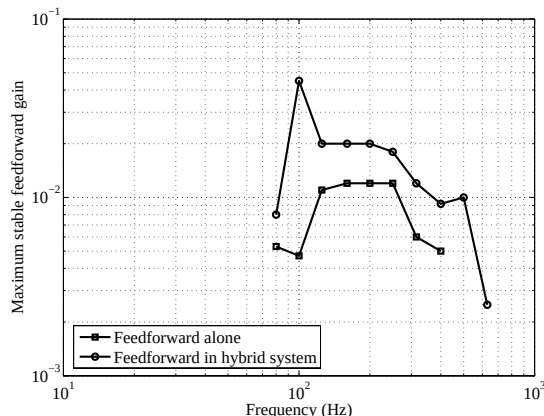


Figure 9: Maximum stable gain (K_{ff}) of the feedforward system for ANR earplug

Figure 8 shows passive performance of the ANR modified earplug, measured for pure tones, as well as the total (active+passive) performance for feedback, feedforward, and hybrid ANR measured at the manikin ear. Hybrid performance is also provided at the error microphone. Passive attenuation is reduced in the mid-frequency range by earplug modification, but high frequency passive attenuation is similar to the unmodified earplug. The feedback compensator gain is fixed at a value that provides a 6 dB gain margin and the feedforward compensator gain is tuned to maximize performance at each frequency. The feedback compensator provides a modest (2 to 6 dB) attenuation over the frequency range measured, while the feedforward compensator alone or hybridized with the feedback compensator provides an additional 8 to 15 dB from 80 to 630 Hz as measured at the in-ear microphone. However, for both the feedforward system alone and the hybrid system, the sound level at the ANR error microphone is reduced to its noise floor for all tones except 80 Hz and 160 Hz, thus at source levels tested, ANR performance is maximized at the error microphone. Total performance for tonal noise measured at the error microphone is 45 to 60 dB. For pure tone ANR experiments, after passive attenuation, the in ear microphones range from 0 (at 800 Hz) to 23 dB (at 200 Hz) above the 55 dB noise floor, thus at some frequencies there is insufficient signal-to-noise ratio to demonstrate ANR performance at the in-ear microphones.

Although both feedforward and hybrid ANR can drive the error signal to its noise floor for single tones, as with the earcup, there is a stability benefit of hybrid ANR. Figure 9 shows the maximum stable

feedforward gain for the feedforward system alone and for the hybrid system. The maximum stable feedforward gain changes sign above 400 Hz for the feedforward system alone. For the hybrid system, the crossover increases to 630 Hz. Figure 11 shows that the maximum stable feedforward gain increases by a factor of two to ten over the 80 to 630 Hz range through hybridization.

The variation in cancellation path gain requires the implementation of a filtered-X LyLMS filter for composite noise. The filter models the acoustic cancellation path of the manikin with a deep inserted ANR earplug. With the filtered-X LyLMS, a single feedforward gain is chosen to accommodate all frequencies of interest. Figure 10 shows passive and total ANR performance for sum-of-tone noise 80-1250 Hz for the ANR earplug and the unmodified earplug, and Figure 11 shows passive and total ANR performance for UH-60 helicopter noise. For sum-of-tone noise, the hybrid system provides a modest increase in performance of the ANR earplug over its passive performance of 0 dB (at 1250 Hz) to 19 dB (at 200 Hz) as measured at the ear microphone, with no spillover (negative attenuation). The ANR performance for the modified earplug is 0 to 25 dB at the error microphone for sum-of-tone noise and 0 to 15 dB for UH-60 noise.

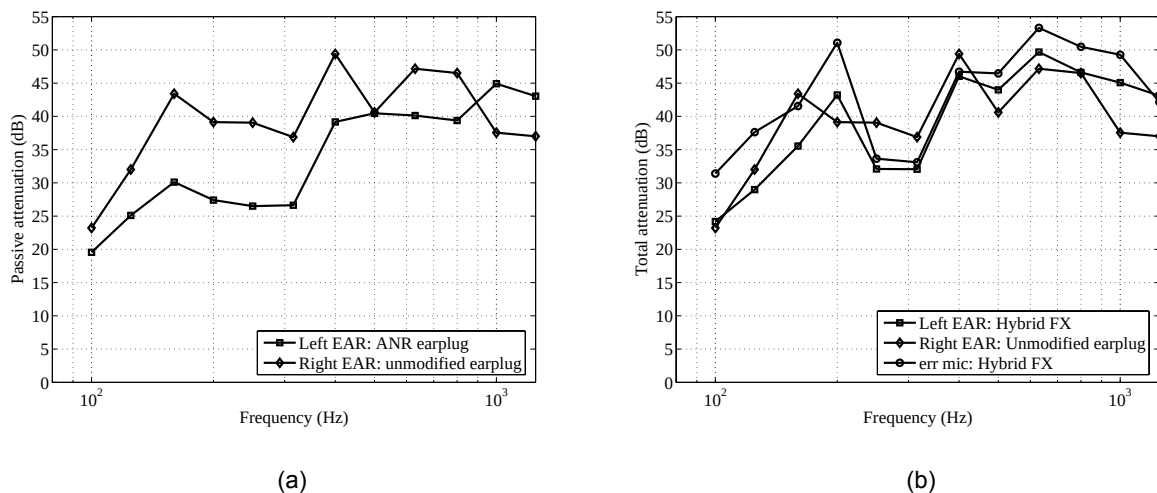


Figure 10: Performance of ANR earplug for sum-of-tones at 1/3 octave band center frequencies 80-1250 Hz, source level 102 dB, (a) passive attenuation of modified and unmodified earplug, (b) Total attenuation of ANR earplug.

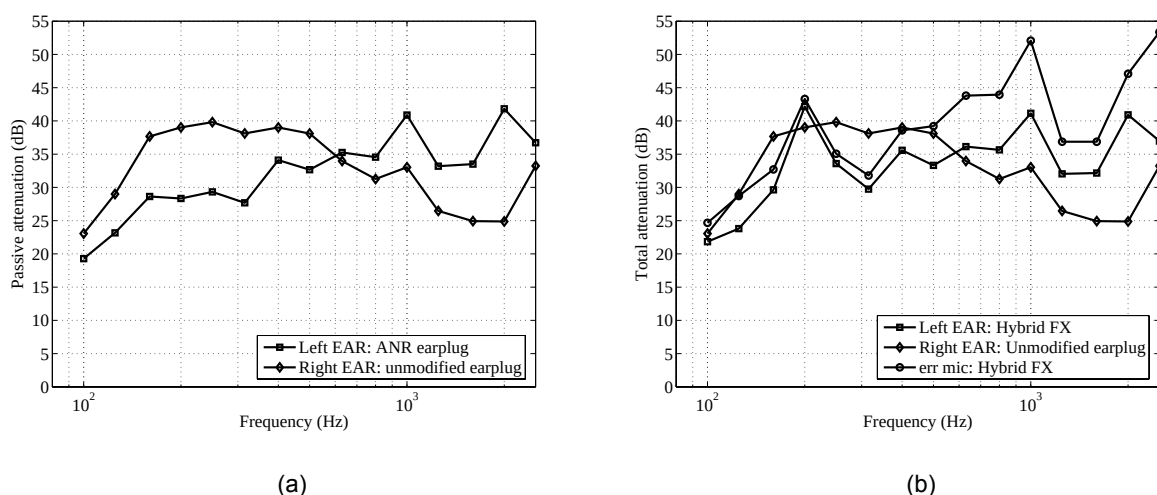


Figure 11: Attenuation performance of ANR earplug for UH-60 helicopter noise, source level 97 dB (a) passive attenuation of modified and unmodified earplug, (b) Total attenuation of ANR earplug.

Table 2: Summary of performance for communication earplug modified for hybrid ANR

Noise Source	Average Noise Level (dB)					Total attenuation (dB)		Active attenuation (dB)	
	Source	Passive unmodified	Passive ANR	Feedb.	Hybrid	Feedb.	Hybrid	Feedb.	Hybrid
Sum-of-tones 80 - 1250 Hz	102	69	73	70	66	32	36	3	7
UH-60 cockpit noise	97	63	66	62	60	35	37	4	6

Table 2 summarizes manikin test results for the earplug showing that, on average, the ANR modified earplug loses 3-4 dB of passive attenuation through modification and gains 6-7 dB through active noise reduction at source levels tested. While the total attenuation of the ANR earplug remains below the passive attenuation of the unmodified earplug at some frequencies, a production earplug, in which the error microphone is integrated in the earplug body could retain good passive attenuation; hence, the projected total attenuation achievable by the active earplug is the sum of the passive performance of the unmodified earplug and the active performance of the ANR earplug. These results suggest that an ANR earplug that retains the passive performance of the unmodified earplug can achieve an average performance improvement of 9-11 dB using the present hybrid ANR algorithm at source levels tested. Unlike the earcup experiments within the LFATC, where source levels allowed operation significantly above the noise floor, source levels used for earplug testing do not permit operation at more than 5-18 dB above the noise floor of the manikin in-ear microphones, thus performance can be expected to improve as source level increases.

CONCLUSION

A hybrid ANR architecture has been evaluated for two hearing protection devices – a commercial circumaural earcup and a commercial communication earplug, both modified for feedforward ANR. The basic hybrid architecture comprised of a broadband feedback controller and a Lyapunov-tuned leaky LMS feedforward component provides increased stability margins over each control component acting alone. Experimental evaluation of the earcup through flat plate testing shows total performance ranging from 30 dB for Huey helicopter noise to 40 dB for sum-of-tone noise. For the ANR earplug, total performance is 36 to 37 dB for sum-of-tone noise and UH-60 helicopter noise. For both systems, stability margins increase through hybrid ANR, without a loss in ANR performance.

ACKNOWLEDGEMENTS

This research is supported through an AFOSR Defense University Research Instrumentation Program award F49620-03-1-0248, the David Clark Company, Inc., and by the U.S. Army Medical Research and Materiel Command contract No. W81XWH-05-C-0031.

REFERENCES

- [1] R.D. Collier, K.H. Kaliski, and L.R. Ray, Experimental Techniques for Evaluation of Active Noise Reduction Communication Headsets with the Thayer Low Frequency Acoustic Test Cell, Proc. NOISECON03, on CD, Cleveland OH, June 2003.
- [2] G.J. Pan, A.J. Brammer, and R.B. Crabtree, Adaptive Feedforward Active Noise Reduction Headset for Low-Frequency Noise, Proc. Symposium on Active Control of Sound and Vibration, Ed. Elliott and Horvath, 1997.

- [3] G.J. Pan, A.J. Brammer, R. Goubran, J.G. Ryan, and J. Zera, Broad-Band Active Noise Reduction in Communication Headsets, *Canadian Acoustics*, 22(3), 113-114, March 1994.
- [4] G.J. Pan and A.J. Brammer, Adaptive Noise Reduction Headset for Helicopter Aircrew, AGARD Symposium on Audio Effectiveness in Aviation, AGARD, Nevilly-sur_seine, pp. 19-1 to 19-6, 1996.
- [5] A.J. Brammer and G.J. Pan, Opportunities for Active Noise Control in Communication Headsets, *Canadian Acoustics* 26(3), 32-33, September 1998.
- [6] L.R. Ray, R. D. Collier, and K.H. Kaliski, Optimization of Stability and Performance of LMS Filters for Feedforward Active Noise Reduction in Communication Headsets, *Proc. ACTIVE02*, 705-715, July 2002.
- [7] D. Cartes, L.R. Ray, and R.D. Collier, Lyapunov Tuning of the Leaky LMS Algorithm for Single-Source, Single-Point Noise Cancellation. *Mechanical Systems and Signal Processing*. 17(5), 925-944, 2003.
- [8] D. Cartes, L.R. Ray, and R.D. Collier, Experimental Evaluation of Leaky Least-Mean-Square Algorithms for Active Noise Reduction in Communication Headsets, *J. Acoustical Soc. of America*, 111, 1758-1772, April 2002.
- [9] A. Streeter, L.R. Ray, and R.D. Collier, Hybrid Feedforward-Feedback Active Noise Control, *Proceedings of the 2004 American Control Conference*, Boston MA, 2876-2881, June 2004.
- [10] M. Winberg et al. A new passive/active hybrid headset for a helicopter application, *Int. J. Acoustics and Vibration*, 4(2), 51-58.
- [11] B. Rafaely and M. Jones, Combined Feedback-Feedforward active Noise-Reducing Headset – The Effect of the Acoustics on Broadband Performance, *J. Acoust. Soc. Amer.*, 112(3), 981-989, 2002.
- [12] S.M. Kuo and D.R. Morgan, *Active Noise Control Systems: Algorithms and DSP Implementations*, John Wiley & Sons, Inc., 1996.
- [13] D. Cartes, L.R. Ray, and R.D. Collier, Low Frequency Acoustic Test Cell for the Evaluation of Circumaural Headsets and Hearing Protection, *Canadian Acoustics*, 30(1), 13-20, 2002.

Double protection passive pour les équipages de l'hélicoptère d'attaque "Tigre" : concept et étude expérimentale

J. Baudou ; joel.baudou@fr.thalesgroup.com

G. Reynaud ; gerard.reynaud@fr.thalesgroup.com

G. Poussin ; gilles.poussin@fr.thalesgroup.com

A. Leger ; alain.leger@fr.thalesgroup.com

THALES Aerospace
Rue Toussaint Catros
33187 Le Haillan
FRANCE

RESUME

Les coques des écouteurs du casque Topowl ont déjà fait l'objet d'une étude visant à optimiser leur protection auditive dans les stricts budgets de masse et de volume impartis. La présente démarche consistait donc à sélectionner puis évaluer des solutions permettant d'améliorer sensiblement cette protection face aux niveaux de bruit plus élevés que prévu relevés en cockpit. Par sa simplicité de mise en œuvre, sans impact sur le casque, son faible coût et son efficacité d'atténuation propre, la protection complémentaire offerte par les bouchons d'oreille passifs s'est naturellement imposée comme axe de travail.

Préalablement évalués de façon empirique, des bouchons de type " mousses Quies " ont été sélectionnés pour leur port confortable et leur forte atténuation dès les basses fréquences. Les mesures, réalisées par test audiométrique (méthode REAT), ont effectivement démontré une atténuation globale très importante sur l'ensemble du spectre. Malheureusement les essais en vol effectués ont immédiatement révélé de multiples défauts dont certains étaient rédhibitoires d'un point de vue opérationnel : dégradation de l'intelligibilité des communications, non perception d'alarmes critiques, sensation d'occlusion. des bouchons personnalisés ont alors été évalués en laboratoire mais leur atténuation dans la bande audiophonique était encore trop importante pour tenter de nouveaux essais en vol avec des perspectives de succès.

En revanche, des bouchons personnalisés à atténuation linéaire – i.e. approximativement constante sur l'ensemble du spectre – ont fourni des résultats remarquables. Les mesures montrent un couplage très satisfaisant avec le casque puisque la protection globale est significativement améliorée alors que la bande audiophonique est modérément atténuée. Suite aux essais en vol effectués avec différents modèles de casques et sur divers types d'hélicoptères, les impressions des neuf pilotes évaluateurs, y compris chez les plus réticents au port des bouchons, sont très favorables à l'utilisation de ce type de protection : préservation de la sécurité du vol par la réduction de la fatigue liée au bruit, amélioration de la qualité d'écoute et de l'intelligibilité globale par un meilleur filtrage des bruits parasites, ergonomie satisfaisante grâce à la personnalisation.

Sous réserve d'une sélection adéquate des bouchons d'oreille et de leur évaluation poussée, en couplage avec le casque, tant en laboratoire qu'en situations réelles, cette démarche montre donc l'intérêt de la double protection passive comme solution simple et efficace dans le cas où la protection d'origine ne peut être modifiée.

Baudou, J.; Reynaud, G.; Poussin, G.; Leger, A. (2005) Double protection passive pour les équipages de l'hélicoptère d'attaque "Tigre" : concept et étude expérimentale. Dans *Nouvelles orientations pour l'amélioration des techniques audio* (p. 17-1 – 17-14). Compte rendu de réunion RTO-MP-HFM-123, Communication 17. Neuilly-sur-Seine, France : RTO. Disponible sur le site : <http://www.rto.nato.int/abstracts.aps>.

1 DESCRIPTION DU VISEUR DE CASQUE

Un viseur de casque n'est acceptable du point de vue opérationnel que si les fonctions ajoutées sur la tête n'altèrent pas les niveaux de protection du casque ni la sécurité du pilote. C'est pourquoi un viseur de casque performant est conçu comme un équipement de tête et non comme un accessoire du casque. En effet, c'est en intégrant les fonctions de visualisation au casque de protection qu'on pourra atteindre en particulier les critères biomécaniques de masse et de centre de gravité, tout en respectant les niveaux de protection passive.

Il a été montré [2] que pour un pilote d'hélicoptère, la masse supportée par la tête est le paramètre dimensionnant vis à vis des risques de lésions du rachis cervical. Elle ne doit pas dépasser 2,3 kg pour un niveau de crash normalisé.

L'autre paramètre dimensionnant est le centre de gravité de l'équipement de tête qui doit être aussi proche que possible de celui de la tête nue afin de respecter la mobilité naturelle de la tête et réduire les risques de cervicalgies et de fatigue.

Le respect de ces critères amène naturellement le concepteur à alléger les éléments de protection passive du casque au profit des fonctions actives du visuel.

Le viseur de casque TopOwl qui équipe l'équipage de l'hélicoptère d'attaque TIGRE a été conçu selon cette démarche. Il utilise la visière pour présenter des images d'aide au pilotage, conduite de tir et navigation, il intègre un système de Détection de Posture qui le relie au calculateur du système de mission et comprend des modules de vision nocturne à intensificateurs de lumière.



Figure 1 : Viseur de casque Topowl.

Le tableau ci après présente une décomposition des fonctions des équipements de tête Topowl et SPH 5 en comparant leurs masses respectives. On voit que le casque qui assure principalement la protection passive a été allégé pour compenser la masse ajoutée par le dispositif de visualisation.

Tableau 1 : Comparaison des masses par fonction entre un viseur de casque et un casque conventionnel.

Equipement de tête	TopOwl	SPH 5
Protection faciale	100 g	2 x 90 g
Protection aux impacts	475 g	700 g
Protection acoustique	75 g	200 g
Communication	80 g	120 g
Liaisons mécaniques	120 g	220 g

Dans le viseur de casque Topowl, une partie de l'énergie d'impact est absorbée par les structures du dispositif de visualisation afin d'alléger d'autant le casque et atteindre au global un niveau de protection équivalent à celui d'un casque conventionnel.

Afin d'obtenir un centre de gravité acceptable, les dispositifs de vision nocturne ont été placés sur les cotés du casque et à hauteur des yeux pour ne pas dégrader le champ de vision.

Cette disposition a entraîné une réduction du volume disponible pour la protection acoustique dont la conception est présentée dans cet article.

2 SPÉCIFICATION

La protection acoustique à obtenir est en général définie par l'avionneur pour l'équipementier sous forme d'une atténuation en fonction de la fréquence. Les valeurs d'atténuation spécifiées devraient être définies en fonction du spectre de bruit ambiant et de la dose maximale admissible pour l'oreille humaine. La spécification d'atténuation est aussi souvent définie en référence à ce qui se fait de mieux dans le domaine. La spécification pour le casque du Tigre est la suivante :

Tableau 2 : Spécification d'atténuation en fonction de la fréquence.

Fréquence	< 250 Hz	250 Hz à 500 Hz	500 Hz à 1 kHz	1 kHz à 3 kHz	3 kHz à 4 kHz
Atténuation	> 10dB	> 20 dB	> 25 dB	> 35 dB	> 40 dB

La méthode de mesure associée à cette spécification n'est pas précisée.

A cette spécification d'atténuation, s'ajoutent les contraintes de conception suivantes pour les protecteurs acoustiques :

Les protecteurs incluent les transducteurs audio.

Les dimensions des protecteurs assurent une couverture anthropométrique du 5^{ème} percentile au 95^{ème} percentile (MIL STD 1472 C).

Les protecteurs doivent être conçus pour un port prolongé de plus de 4 heures en continu, 8 heures par jour.

Les protections acoustiques ne doivent pas engager la sécurité de l'utilisateur en cas d'impact latéral.

La masse de chaque protecteur complet ne doit pas dépasser 50g.

3 CONCEPTION

Les protecteurs acoustiques des casques d'écoute, casques anti-bruit ou des casques conventionnels pour pilotes d'hélicoptères sont constitués d'une cavité comportant une mousse interne absorbante, un coussinet d'étanchéité et un dispositif de maintien et d'ajustement de la pression de contact autour de l'oreille.

On sait [3] que la masse et le volume de la cavité, la souplesse du coussinet, l'effort de maintien et surtout l'étanchéité sont les facteurs clé pour une bonne atténuation.

Une première étude théorique a été menée avec l'aide du laboratoire spécialisé du Centre de Transfert de Technologie de l'université du Mans afin de d'appréhender l'influence de chaque composant et préparer la conception des cavités acoustiques à volume réduit. Cette étude a montré que pour le volume et la masse donnés :

- le coussinet et la liaison du protecteur au casque sont les paramètres principaux pour l'atténuation des basses fréquences avec une préférence pour un coussinet de faible périmètre afin de faciliter l'étanchéité circumaurale et une épaisseur aussi faible que possible afin de limiter la transmission de vibrations,
- la coque et la mousse absorbante n'influencent que les moyennes et hautes fréquences avec une préférence pour une coque en matériau intrinsèquement amortissant ou très rigide afin de repousser ses vibrations aux hautes fréquences où elles seront atténuées par le matériau absorbant.

Un calcul des fréquences de résonance par éléments finis a été réalisé afin de vérifier cette hypothèse avec le dessin de la cavité finale.

Ces résultats ont permis de réaliser un prototype utilisé pour la sélection des matériaux et composants avec le concours de l'Institut de Médecine Aéronautique du Service de Santé des Armées (IMASSA). Différentes configurations de protecteurs ont été testées sur une tête artificielle de l'Institut Saint Louis (ISL), les protecteurs étant maintenus par un arceau à serrage contrôlé.

Le matériau de la coque a d'abord été sélectionné parmi des composites carbone, verre et aramide avec matrice époxy, tissus et nombre de plis variables. Le composite carbone époxy en 3 plis a été retenu car il présente le meilleur compromis masse / atténuation.

A partir de cette coque, différents coussinets issus de protecteurs anti-bruits du commerce ont été testés. Parmi plus de 20 références de produits comprenant également des coussinets avec gel de silicone pour le confort thermique et l'étanchéité, le coussinet BILSOM 2742 est celui qui présente le meilleur rapport masse / atténuation.

Le matériaux absorbant interne retenu est une mousse mélamine car elle présente une atténuation supérieure de 1 dB dans les basses et moyennes fréquences et jusqu'à 8dB sur certaines fréquences par rapport aux autres mousses testées. L'ajout de matériau viscoélastique type Butyl et polychloroprène à l'intérieur de la cavité n'a pas apporté d'amélioration.

Enfin, le mode de liaison au casque a été conçu de manière à limiter la transmission des vibrations à la cavité et assurer son étanchéité. Ainsi, la coque est reliée au casque par l'intermédiaire de cales en tissu auto agrippant (Velcro) et le casque possède un dispositif d'ajustement de la position du protecteur permettant de loger l'oreille entière à l'intérieur du coussinet et d'ajuster la répartition de la pression autour de l'oreille.

Le casque qui intègre les cavités acoustiques est lui même fermé par un bourrelet de contour disponible en plusieurs tailles afin d'assurer le contact avec la tête et participer à l'isolation acoustique.



Figure 2 : Conception des protecteurs.

4 MESURE DE L'ATTÉNUATION DE L'ÉQUIPEMENT DE TÊTE

Pour mesurer l'atténuation du viseur de casque ainsi défini, c'est l'équipement de tête complet qui a été testé afin de tenir compte des éventuels effets de transmission et de couplage. La mesure a été réalisée dans le laboratoire de l'IMASSA selon la méthode MIRE définie par la norme ANSI S12-42-1995 sur 4 personnes et non plus sur tête artificielle, pour plus de représentativité.

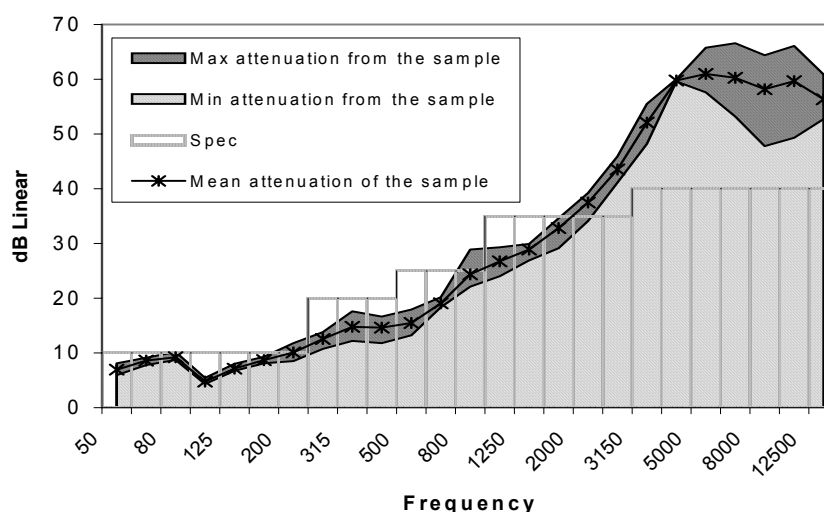


Figure 3 : Atténuation de l'équipement de tête sur 4 sujets, méthode MIRE à l'IMASSA.

Ces mesures étant faites sur 4 sujets seulement, on conserve les valeurs extrêmes et la moyenne plutôt que des écarts types. On constate que les écarts entre les différents sujets sont faibles, ce qui démontre l'efficacité du dispositif de maintien qui assure un effort de serrage reproductible et une bonne étanchéité du protecteur. L'atténuation globale est donc insuffisante au regard de la spécification. On sait que la méthode MIRE donne des résultats d'atténuation inférieurs à basse fréquence comparativement à une méthode subjective de type Real Ear Attenuation at Threshold (REAT) et on peut estimer être proche de l'objectif en dessous de 250 Hz mais entre 250Hz et 1500 Hz, il manque jusqu'à 10 dB pour atteindre la spécification.

5 ANALYSE

Il apparaît donc impossible de respecter la spécification initiale avec des protecteurs dont le volume est réduit et la masse limitée. On remarque que cette spécification n'est pas tenue non plus avec un casque conventionnel comme le casque MK4 britannique [4] pourtant réputé performant sur le plan de l'atténuation acoustique.

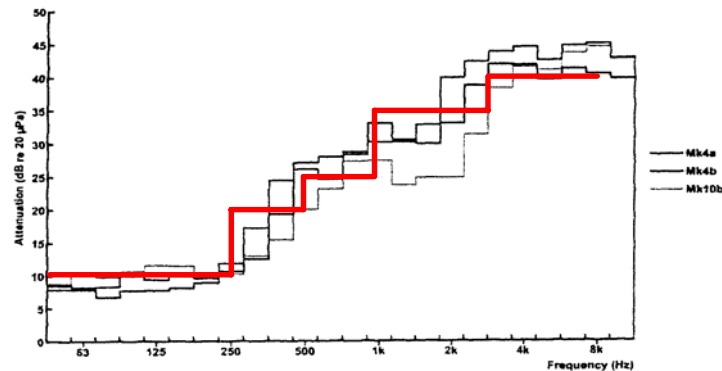


Figure 4 : Atténuation de casques Mk4 et spécification Tigre [4].

Législation

La législation du travail française applique la directive européenne 2003/10/CE basée sur la norme ISO 1999 :1990. Cette législation ne s'adresse pas aux pilotes militaires mais elle nous sert de référence pour l'analyse du risque de dommage auditif.

Cette directive fixe la durée d'exposition en fonction de la dose de bruit continu au delà de laquelle l'usage de protection est nécessaire. Pour un niveau global de 85 dBA, la durée d'exposition quotidienne ne peut dépasser 8 heures. En travaillant à énergie constante, cette durée d'exposition peut être divisée par deux chaque fois que le bruit augmente de 3 dBA, ainsi pour un bruit ambiant de 100 dBA, la durée du vol ne doit pas dépasser 15 min. En pic, le niveau de bruit ne doit pas dépasser 135 dBLin.

Il est rare que l'équipage d'un hélicoptère soit en vol 8 heures par jour et il est improbable que cet équipage vole 5 jours par semaines et 8 heures par jour. S. James [3] propose donc un facteur de correction du seuil en fonction du nombre d'heures de vol rapporté sur une année. Ainsi pour 350 heures de vol annuelles, le seuil devient 92,4 dBA.

Le dépassement de ces doses peut provoquer une réduction temporaire de l'acuité auditive et aboutir à terme à une baisse définitive de l'audition. Le temps de récupération d'une perte temporaire d'audition dépend du niveau et de la dose auxquels on a été exposé, en général une audition normale est recouvrée au bout de 3 à 5 jours.

La perte d'audition est détectée lors d'un audiogramme lorsque le seuil d'audition est supérieur à 20 dB. En général, l'acuité auditive est d'abord plus faible pour les fréquences situées autour de 6 kHz.

D'autre part, un niveau de bruit ambiant trop élevé cause également des risques opérationnels tels que fatigue d'où diminution de l'efficacité et dégradation de l'intelligibilité de la communication électro-acoustique. Buck et al. [1] reporte que le succès d'une mission est en rapport direct avec l'intelligibilité de la communication, il est donc nécessaire d'atténuer les bruits de basses fréquences pour améliorer l'intelligibilité des plus hautes fréquences de la parole.

On retient donc le critère de 85 dBL_{Aeq8} pour remplacer la spécification initiale mais également pour réduire les risques de perte temporaire d'audition et la fatigue auditive, avec le temps de repos qui pourrait s'en suivre. On comprend que c'est l'atténuation des basses fréquences qu'il faudra privilégier pour une bonne intelligibilité des communications.

Bruit hélicoptère

Le bruit d'un hélicoptère est un bruit continu, il présente en général un niveau élevé pour les basses fréquences et un spectre très étendu de 15 Hz à plus de 6 kHz. Chaque hélicoptère possède un bruit caractéristique qui dépend entre autres du nombre de turbines, de la charge embarquée, de la configuration des rotors, de l'aérodynamique et donc de la vitesse, des vibrations des structures mécaniques etc. ; dans le cockpit on trouve également des sources supplémentaires de bruit comme le conditionnement d'air et l'avionique.

Dans un hélicoptère de transport ou de liaison, le bruit en cockpit est plus faible que dans un cockpit d'hélicoptère de combat comme le Tigre où les postes sont en tandem et dans lequel le co-pilote, en place arrière, se situe très près de la boîte de transmission et des turbines.

Des mesures de bruit on donc été faites en cockpit arrière dans différentes configurations de vol, du vol stationnaire à la vitesse maximale.

Ces mesures ont été réalisées avec un micro placé dans l'oreille et un micro placé à l'extérieur du casque, comme dans une mesure MIRE. Ces mesures ont également permis d'évaluer le volume de la radio communication.

Atténuation réelle de l'équipement de tête

A partir des atténuations du casque mesurées aux mêmes fréquences que dans on obtient un niveau sous le casque de 84 dBA. Il s'agit d'une valeur moyenne couvrant 50% de la population. Pour couvrir une plus grande part de la population, on tient compte des dispersions individuelles et donc de la valeur de l'écart type sur la moyenne des atténuations globales individuelles : 0,6 dBA. La dose de bruit sous casque devient ainsi 84,6 dBA pour 84% de la population et 85,6 dBA pour 95% de la population. Ces valeurs sont indicatives et seraient à affiner avec un plus grand nombre de mesures : 10 suivant normes ANSI.

Communication radio

S. James [4] rapporte que la pression acoustique due à la communication radio est de 6 à 10 dB. Compte tenu de l'occurrence et de la durée des messages, on évalue le niveau acoustique pondéré de la communication à 3 dBA sur tout le vol.

Bruit d'armes

Lors d'un tir canon, le bruit impulsionnel est très élevé mais reste inférieur au critère de 160 dBSPL retenu par la France [5] pour l'exposition aux bruits d'armes. La pression acoustique au niveau de l'oreille étant encore diminuée par le casque, il n'y a donc pas de risque de dommage auditif. Dans le cas des bruits d'armes, la durée admissible d'exposition dépend du nombre de coups tirés durant le vol, voire 24 heures. L'utilisation d'armes est naturellement limitée dans le temps, même pour une mission de combat. On définit donc l'exposition admissible au bruit impulsionnel par l'intermédiaire d'une réduction de la durée d'exposition au bruit continu. Il a été déterminé que la durée d'exposition pouvait être réduite jusqu'à 30% pour une utilisation intensive du canon.

Niveau de bruit réel

Compte tenu du bruit réel de l'hélicoptère sous le casque et du volume de communication, la dose de bruit perçue par le co-pilote atteint 88,6 dBA pour un taux de couverture de 95%.

Le risque de perte d'audition avec la protection actuelle est donc assez faible mais il faut respecter les durées d'exposition définies par iso-énergie si on veut être à l'abri.

Dans ce cas, la durée de vol doit être inférieure à 3 h 30 min pour respecter le critère de 85 dBL_{Aeq8}, ce qui paraît trop faible en temps de paix, notamment pour des pilotes instructeurs.

En cas de tirs intenses, cette durée est réduite à 2 h 30 min ce qui n'est pas suffisant en temps de guerre.

6 AMÉLIORATION DE L'ATTÉNUATION

Il nous faut donc améliorer l'atténuation globale de l'équipement de tête afin de respecter le critère de 85 dBA. Les améliorations possibles sont listées dans le tableau suivant :

Tableau 3 : Comparaison des solutions d'amélioration d'atténuation.

	Performance	Impact sur casque	Impact sur masse
Protecteurs passifs plus efficaces	+5 dB jusqu'à 2 kHz	Augmentation volume et déplacement des intensificateurs de lumière	+ 110 g
Atténuation Active de Bruit	+5 dB à +30 dB jusqu'à 500 Hz	Augmenter le volume des cavités et ajouter l'alimentation	+ 50 g
Bouchons d'oreilles actifs Type CEP	> 20 dBA	Ajout d'un connecteur sur le casque	+ 25 g
Bouchons d'oreille passifs Type mousse ou moulés	> 20 dBA	Aucun	< 3 g

Amélioration des protecteurs

Pour rendre les protecteurs actuels plus efficaces en basses fréquences, il faudrait augmenter leur volume et leur masse si on se réfère aux casques anti-bruit du commerce ou ceux des casques d'hélicoptère type SPH 4, HGU 56 ou Alpha MK 4. Une augmentation de masse supérieure à 100 g remettrait en cause la tenue du critère de lésion au crash, en particulier pour les femmes. L'augmentation du volume pour atteindre celui des protecteurs connus nécessiterait de revoir la disposition des intensificateurs de lumière avec pour conséquence la dégradation du centre de gravité de l'équipement de tête et de l'ergonomie visuelle.

Atténuation active de bruit

Parmi les différents systèmes ANR [1], le plus simple à intégrer est certainement l'ANR feedback analogique (mesure du son à l'intérieur du protecteur) car il suffit d'intégrer le module ANR dans la cavité à la place du transducteur et d'ajouter les fils d'alimentation. Une adaptation du dessin du protecteur est quand même à prévoir pour un fonctionnement optimum dans tout le spectre, le surcroît de masse de 50g environ pourrait être admissible pour une population de pilotes supérieure au 25ème percentile.

Le système feed-forward (mesure du son à l'extérieur du casque) paraît théoriquement mieux adapté au bruit d'un hélicoptère peu fluctuant mais sa mise en œuvre en pratique est délicate car il nécessite une parfaite reproductibilité du fonctionnement du protecteur sur chaque pilote, ce qui est difficile compte tenu des variations anatomiques naturelles. Cette contrainte pourra être levée avec un filtrage digital à la place du filtrage analogique actuel.

L'efficacité optimale de l'ANR est centrée autour de 300 Hz avec 15 dB d'atténuation en fonction du filtrage adopté. La contribution de l'ANR sur l'atténuation du bruit se situe entre 50 Hz et 500 Hz, au delà l'atténuation est nulle et il peut même apparaître une amplification du signal autour de 800 Hz.

Cette atténuation du bruit en basses fréquences diminue l'effet de masque des fréquences plus hautes de la communication et facilite leur audition sans nécessiter un fort volume. Cette performance est donc intéressante vis à vis du bruit hélicoptère pour abaisser le niveau global d'exposition mais n'est peut être pas suffisante pour combler le déficit d'atténuation du protecteur actuel dans la bande 500 Hz à 1500 Hz compte tenu notamment du risque d'amplification du bruit par l'ANR dans cette même bande.

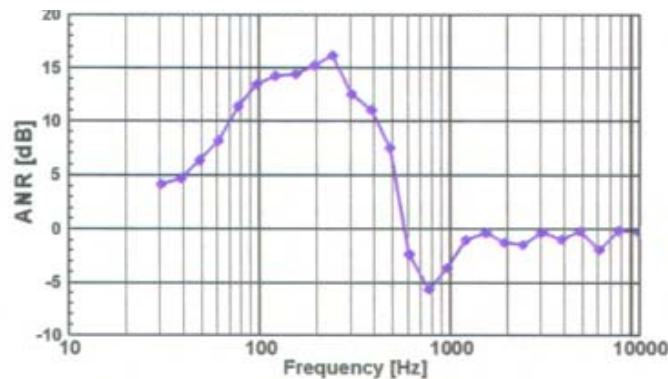


Figure 5 : Contribution de l'ANR sur l'atténuation [1].

Bouchons d'oreilles

Les bouchons d'oreilles sont largement utilisés dans l'industrie sous diverses formes, en mousses avec différentes formes, bouchons souples en élastomère ou personnalisés par moulage en acrylique ou silicone.

Le système présentant le meilleur rapport efficacité/simplicité est certainement le bouchon en mousse qui se forme facilement à la taille du conduit auditif et qui procure une atténuation supérieure à 20 dB lorsqu'il est mis en place correctement [3].

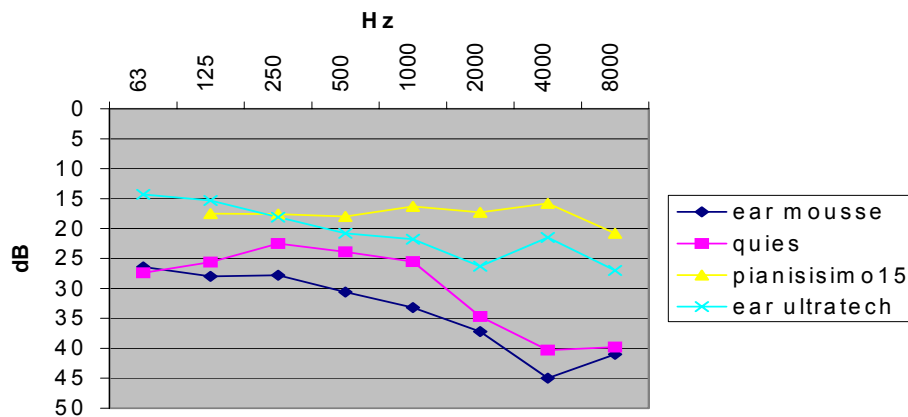


Figure 6 : Atténuation de différents bouchons (IMASSA).

L'utilisation de bouchons d'oreille en complément de la protection apportée par l'équipement de tête actuel paraît donc être la voie à la fois la plus simple et la sûre d'améliorer l'atténuation.

Des premiers tests en laboratoire effectués selon la méthode REAT, ont confirmés que l'ajout de bouchons permet d'atteindre une très bonne protection de l'ordre de 35 dBA.

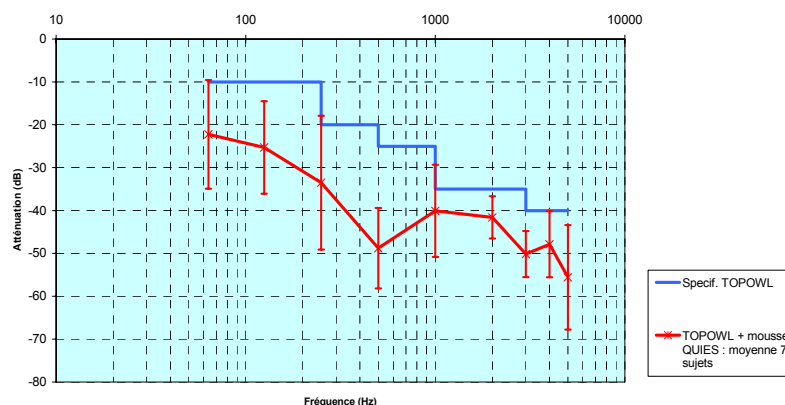


Figure 7 : Atténuation de la double protection avec bouchons mousse.

L'apport des bouchons est très important en basses fréquences et surtout autour de 500 Hz, justement où on recherche une amélioration de la performance de l'atténuation du protecteur. A hautes fréquences, l'atténuation est inférieure à la somme des atténuations du casque et des bouchons. Les écarts types sont très importants à cause d'une part des mesures selon une méthode subjective mais surtout à cause de l'efficacité des bouchons qui dépend fortement de leur insertion dans le canal auditif. En effet, l'atténuation peut être nulle si le bouchon n'est inséré que de 25% de sa longueur dans le conduit auditif [3] et peut atteindre 22 dB lorsque le bouchon est complètement inséré (presque plus visible). Cette insertion dépend de la taille du conduit auditif mais aussi de l'effet psychologique du risque de ne pas pouvoir ressortir le bouchon.

Avec cette double protection, l'atténuation est telle que l'audition des communications est affaiblie, ce qui peut être compensé par une augmentation du volume au niveau du casque, mais surtout celle des alarmes cockpit devient insuffisante, ce qui est beaucoup plus critique du fait de l'impossibilité d'ajuster leur volume individuellement. Enfin, une sensation d'occlusion et d'isolation de l'environnement perçue lors des essais en vol a définitivement rendu réducteur cette double protection.

Bouchons actifs

L'utilisation de bouchons actifs type CEP (Communication Ear Plug) paraît être une bonne solution pour régler à la fois le problème de l'atténuation du bruit et de la conservation de l'audition des communications.

Un modèle de ce type de bouchon comprend une partie en mousse qui obstrue le conduit auditif et un transducteur miniature à l'extérieur, le son est transmis via un conduit qui traverse la partie en mousse. La encore, l'efficacité de la protection dépend de l'insertion correcte du bouchon. Dans quelques cas, des problèmes de confort liés à la taille du transducteur externe ou à la présence du fil d'alimentation contre l'oreille ont également été notés. Le problème de l'audition des alarmes en cockpit demeure le même qu'avec des bouchons mousse.

Bouchons personnalisés

Les bouchons personnalisés règlent le problème de l'insertion et de la taille du conduit auditif par prise d'empreinte et reproduction de la forme du conduit de l'oreille du sujet. Les bouchons testés sont insérées dans la parties externe du conduit auditif, c'est à dire la partie malléable autour du cartilage. Il existe des techniques qui améliorent encore plus l'étanchéité du bouchon en prolongeant la partie insérée jusqu'au niveau de la partie du canal comprise dans la partie osseuse rigide de l'oreille [3]. Compte tenu de la délicatesse du procédé de réalisation de ces inserts profonds, seuls le premier type de bouchons a été évalué.

La transmission du son à l'oreille est assurée par un orifice percé à posteriori dans le bouchon moulé, un filtre peut être ajouté à l'entrée ou dans l'orifice de manière à obtenir une atténuation adaptée au bruit à atténuer. Ainsi l'atténuation peut être linéaire, c'est à dire pratiquement constante sur l'ensemble du spectre et présente une très faible distorsion dans les fréquences vocales.

Un premier modèle avec filtre intégré à l'orifice a été testé en laboratoire mais l'atténuation dans la bande audio était encore trop importante pour tenter de nouveaux essais en vol avec des perspectives de succès. En revanche, un modèle avec filtre externe utilisé par les musiciens en particulier et facilement disponible chez les audioprothésistes, a donné des résultats remarquables à plusieurs points de vue. En effet, les mesures en laboratoire montrent un couplage très satisfaisant avec le casque puisque la protection globale est significativement améliorée alors que la bande audio est modérément atténuée.

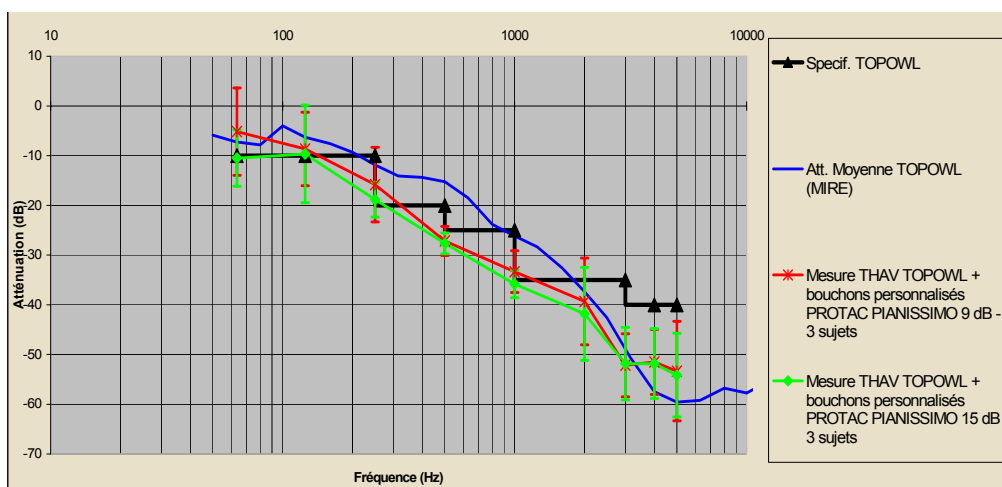


Figure 8 : Atténuation comparée de double protections avec bouchons linéaires 9 dB et 15 dB.

L'atténuation moyenne tangente même la spécification initiale avec des filtres 15 dB. Elle n'est toutefois pas atteinte si on tient compte des écarts types encore importants car liés à la méthode REAT et à cause d'un petit nombre de mesures (3 sujets). L'atténuation à hautes fréquences paraît plus faible avec la double protection qu'avec le casque seul mais les résultats ne sont pas parfaitement comparables car obtenus avec 2 méthodes MIRE et REAT et dans 2 laboratoires différents. Selon ces mesures, la double protection ainsi réalisée atteint 80,4 dBA au global avec les communications soit un apport de 3,6 dBA supplémentaires, ce qui paraît faible mais le nombre d'expérimentation est trop faible pour conclure sur la valeur réelle. Des mesures complémentaires sont donc à prévoir, le bruit en cockpit et l'atténuation devront être déterminés de façon cohérente et aux mêmes fréquences. On a toutefois constaté que la double protection a apporté un gain d'atténuation à chaque fréquence pour chaque expérimentateur, ce qui confirme l'intérêt de la solution et permet d'envisager son application.

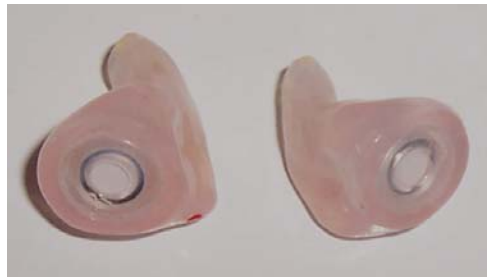


Figure 9 : Bouchons linéaires personnalisés avec filtre 25 dB.

7 ESSAIS EN VOL

La validation de la double protection est passée par une première phase d'essais en vol. Les trois filtres disponibles : 9dB, 15 dB et 25 dB ont été testés par trois pilotes d'essais.

Les rapports d'essais indiquent que :

- l'ergonomie est jugée satisfaisante,
- la sécurité de vol est améliorée du fait de la réduction de fatigue liée au bruit, et qu'aucun problème de décompression n'est relevé,
- la qualité d'écoute est améliorée par un effet de réduction de bruits parasites,
- les bouchons de type 15 dB présentent le meilleur compromis atténuation/qualité d'écoute,
- pour éviter leur perte, les bouchons doivent être équipés d'une cordelette en caoutchouc, suffisamment souple pour s'étirer ou casser et d'éviter de blesser l'oreille en cas de traction involontaire.

A la suite de ces indications, une deuxième phase d'essais avec des pilotes opérationnels a été menée dans différentes machines pour des missions réelles comprenant des vols de navigation, tactique et convoyage, de jour et de nuit, avec tirs canon et roquettes.

Les rapports des pilotes confirment les résultats précédents :

- Qualité de l'écoute : les bouchons utilisés améliorent sensiblement la qualité de l'écoute en filtrant effectivement les bruits parasites. Toutefois, l'utilisation d'un filtre change l'environnement sonore et demande à l'utilisateur une accoutumance lui permettant de se créer de nouveaux repères sonores.

- Audition des alarmes : l'audition des différentes alarmes rencontrées ne pose pas de problème particulier.
- Ergonomie : les bouchons utilisés ne causent pas de gêne particulière pendant le vol. Cette remarque reste valable en tirs canon réalisés au casque. Une gêne peut être perçue si le positionnement de l'écouteur n'est pas parfait. L'utilisateur devra donc avoir une attention particulière lors de la mise en place de son casque et notamment sur le positionnement effectif des écouteurs. Les bouchons eux-mêmes ne provoquent pas de gêne particulière ni d'irritation.
- Tirs canon et roquette : le port des filtres n'apporte pas de modification notable du bruit, généré par le tir, perçu par le tireur. Ceci semble être cohérent si l'on prend en compte que ces bruits sont "sourds" (basse fréquence).
- Effet secondaire : l'utilisation des bouchons lors des vols Tigre a permis de mettre en évidence l'absence de bourdonnements observés lors de vols similaires réalisés sans bouchons. D'où un confort acoustique après le vol qui n'est pas du tout négligeable.

Suite aux essais en vol réalisés à ce jour, les bouchons linéaires 15 dB ont été jugés acceptables par les pilotes.

8 CONCLUSION

Le bruit d'un hélicoptère est surtout riche en basses fréquences, ce qui gêne les communications et entraîne une fatigue auditive avec nécessité de récupération. Le respect du critère d'exposition à 85 dBLAeq8 permet de contenir le risque auditif, que ce soit en bruit continu ou lors de tirs d'armes. L'atténuation active de bruit est actuellement la solution la plus efficace pour réduire le bruit à basses fréquences jusqu'à 500 Hz, elle doit donc être couplée à une protection passive performante pour le reste du spectre. Les futurs développements avec filtrage digital ou intégration dans des bouchons d'oreille permettront certainement d'étendre l'efficacité de l'atténuation active.

Le viseur de casque est un équipement incontournable des hélicoptères de combat actuels. Le respect des critères de lésion au crash et de fatigue limite la masse supportée par la tête et nécessite donc l'allègement des fonctions de protection passive sans en réduire le niveau.

La double protection passive présente un intérêt pour les viseurs de casques car elle permet de réduire la spécification d'atténuation passive du casque et donc de l'alléger. Il s'en suit une plus grande liberté de conception pour les protecteurs et la possibilité d'implanter les capteurs de vision de nuit au niveau des oreilles pour un meilleur équilibrage de l'équipement de tête.

Les bouchons d'oreilles personnalisés à atténuation linéaire représentent un bon complément à l'atténuation du casque car ils réduisent le bruit parasite dans les basses fréquences sans trop atténuer les fréquences de la communication. Ce qui améliore l'intelligibilité des communications tout en réduisant le volume sonore global perçu par l'oreille. Ainsi la fatigue liée au bruit et aux communications est réduite et l'efficacité de la mission est améliorée. L'atténuation peut être ajustée par le choix du filtre, ce qui permet d'éviter l'effet d'occlusion et d'isolement afin de conserver la perception spatiale et de l'environnement.

La double protection est également intéressante dans le cas de casques conventionnels pour améliorer l'atténuation globale face à des machines de plus en plus bruyantes car elle ne nécessite aucune modification matérielle.

Il reste à convaincre les opérationnels de l'utiliser car c'est une opération supplémentaire au départ de la mission et l'effet de la protection supplémentaire n'est pas encore sensible à ce moment là. Cela passe par une sensibilisation et une formation du personnel aux risques d'exposition au bruit.

La logistique reste également à mettre en place dans les forces afin d'équiper les pilotes et maintenir les protections à disposition.

RÉFÉRENCES

- [1] K. Buck and V. Zimpfer-Jost - Active Hearing Protection Systems and their Performance ; Lecture Series HFM-111 on " Personal Active Noise Reduction ", October 2004.
- [2] A. Leger, L. Portier, J. Baudou, X. Trosseille – Crash Survivability And Operational Comfort Issues Of Helmet Mounted Display In Helicopters: Simulation Approach And Flight Tests Results; PROCEEDING "The Design and Integration of Helmet Systems"; Additional paper; December 2-5 1997.
- [3] R. McKinley and V. Bjorn – Passive Hearing Protection Systems and their Performance.
- [4] S. James – Defining the cockpit noise hazard, aircrew hearing damage risk and the benefits Active Noise Reduction can provide.
- [5] A. Dancer et R. Franke – Hearing Hazard from Impulse Noise : a Comparative Study of Two Classical Criteria for Weapon Noises (Pfander Criterion and Smoorenburg Criterion) and Laeq8 Method ; acta acoustica 3 (1995) 539-547.
- [6] ANSI S12.42-1995 – Microphone In Real Ear and Acoustic test Fixture method for the Measurement of Insertion Loss of Circumaural hearing Protection Devices.
- [7] ANSI S12.6-1984 – Method for the Measurement of Real Ear Attenuation of Hearing Protectors.
- [8] ISO 1999:1990 – Détermination de l'exposition au bruit en milieu professionnel et estimation du dommage auditif induit par le bruit.

Experimentation to Address Appropriate Test Techniques for Measuring the Attenuation Provided by Double ANR Hearing Protectors

Susan E. Mercy, Christopher Tubb and Soo H. James

QinetiQ

Cody Technology Park, Ively Road
Farnborough, Hants, GU14 0LX
UK

semercy@qinetiq.com

ABSTRACT

Increasingly the military environment can involve exposure to high levels of noise; levels of up to 150dB are anticipated (in some quarters) for deck-crew and ground-crew working with the Joint Combat Aircraft. In order to protect the hearing of personnel in high levels of noise, the use of double protection (earplugs worn in conjunction with earmuffs) is also increasing. Active Noise Reduction (ANR) systems are now widely incorporated into headsets and have also been incorporated into personally moulded earplugs. Therefore, the combination of an ANR Headset and ANR earplugs is the next obvious step in reducing high levels of noise at ear. Difficulties arise in determining the appropriate technique for establishing levels of sound attenuation achieved by the use of this Double ANR system. The Real Ear At Threshold (REAT) method has the advantage that it is a measure of the noise perceived by the subject. However, it can result in masking errors at low frequencies due to physiological noise. REAT cannot be used with ANR devices since the electronic noise of the ANR circuitry at approximately 1-2kHz will raise perceived threshold at this frequency range. The Microphone In Real Ear (MIRE) method has the advantage that it measures the absolute sound pressure levels in the ear canal and can be used with ANR devices. However, this technique does not necessarily measure what the subject hears, as it cannot measure the sound transmitted to the inner ear via routes other than the primary path of the ear canal.

The personally moulded earplugs used by QinetiQ have been designed to incorporate microphones at the tip to measure the sound pressure level within the ear canal, and the sense microphone of the ANR earplugs can also be used to measure the sound pressure level within the device. Experiments using ANR headsets in conjunction with passive personally moulded earplugs were used to compare REAT and MIRE methods and to show the differences between the techniques. The experimental results show that the overall attenuation provided by combinations of ANR headsets and passive earplugs can be found by choosing the appropriate technique in any given one third-octave band to produce a combined REAT/MIRE assessment of the attenuation.

This work was funded by the Human Sciences Domain of the UK Ministry of Defence Scientific Research Programme.

1.0 INTRODUCTION

Noise induced hearing loss is, in theory, entirely preventable and legislation [1, 2] exists to minimise and hopefully eliminate noise hazards and working practices that result in this type of hearing damage. The legislation provides limits on daily personal noise exposure and since the daily personal noise exposure is defined as a combination of noise level and exposure duration, the level can be traded against the duration to ensure that employees' personal exposure stays within the limits.

There are environments where the reduction of noise exposure by decreasing the noise level at source or by shortening the duration of exposure in order to comply with the legislative limits is impractical or costly to implement. These environments tend to be military, where the high levels of noise encountered by personnel are difficult to limit at source. In particular, the noise levels encountered by aircraft-carrier deck-crew, and potentially by ground-crew, are predicted in some quarters to be as high as 150 dB(A) when aircraft such as the Joint Combat Aircraft are brought into service. With noise levels this high, where a reduction in exposure duration by alteration in shift-patterns would result in unwieldy complements of personnel, the only remaining option in the reduction of noise exposure is to improve the hearing protection provided and thereby reduce the noise at the ear.

Standard hearing protectors, although improving in design and the levels of sound attenuation that they can achieve, do not sufficiently reduce the noise at the ear when working in these high noise environments. Hearing protection devices such as earplugs and earmuffs can be worn in combination to attempt to further reduce the noise travelling along the normal air-conductive pathways, i.e., down the ear canal, through the middle ear to the inner ear. However, the use of this type of double hearing protection does not increase the level of sound attenuation achieved by the sum of the attenuation provided by each device separately, as might be expected. The combined performance of the two devices is not simple to predict and although various empirical techniques have been used in the past to attempt to calculate the total attenuation achieved by double protection [3, 4] the most accurate method is still that of direct measurement.

One explanation for the failure of double protection to provide a level of sound attenuation equivalent to a summation of the separate attenuation levels is that at frequencies at and above 2kHz the limits of sealing the air conductive pathway have been reached. This means that at these frequencies any further addition of protection to the outer ear canal provides no extra benefit in reducing the noise dose received. Sound can also reach the cochlea via a number of routes through the head and torso, known as body or bone conduction, that were determined by Tonndorf [5] and have been investigated further by other researchers [6, 7, 8]. At low frequencies, vibration of the head causes the ear canal walls to vibrate and generate additional noise within the ear canal; this additional noise is transmitted to the cochlea via normal air-conductive routes. At higher frequencies the vibration of the bones and fluids of the head cause direct stimulation of the cochlea.

A number of researchers have investigated the bone conduction threshold i.e., levels of sound attenuation that are required before bone conduction pathways dominate the air conductive pathways. The highest bone conduction threshold was measured by Zwislocki [9], who achieved the high levels of attenuation in his experiment using metal rods and wax earplugs in combination with heavy earmuffs incorporating Helmholtz resonators. Watson and Gales [10] also established bone conduction thresholds using laboratory devices and Berger [11] showed that with double protection the insertion depth of a conventional earplug influenced the level of overall sound attenuation regardless of the weight and type of earmuff. These researchers all established different bone conduction thresholds.

The combined use of standard earplugs and earmuffs produces levels of hearing protection that reach bone conduction limits at the higher frequencies, however, at lower frequencies the bone conduction limits are not met. Following on from Berger's work with deeply inserted foam earplugs, some researchers are developing personally moulded earplugs that extend down into the bony meatus, thereby reducing the vibration of the ear canal walls at low frequencies. An alternative approach is to increase the protection provided by the earplug and earmuff, by using high quality passive devices incorporating Active Noise Reduction (ANR) systems. ANR provides extra active attenuation below 1kHz and the extra active attenuation in the ANR earmuff and the ANR earplug may enable the noise at the ear to be reduced further, increasing levels of attenuation until the bone conduction thresholds are met at all frequencies.

The Royal Aerospace Establishment developed a universal-fit ANR earplug [12] in which a t-shaped device was fitted through a shortened foam earplug. This ANR earplug contained a microphone that sensed the sound pressure level in the ear canal and a transducer that emitted the ‘cancelling’ waveform. The earplug was connected via a thin cable to an external box that contained the ANR circuit. Although the ANR earplug worked well it was difficult to wear comfortably under earmuffs, as the top section did not fit well into the subjects’ conchas.

Further development of the ANR earplug by QinetiQ has produced a unique-fit device that is personally moulded to the subject’s own ears. In this form, the earplug is easy to fit but can still contain the ANR sense microphone and the transducer to emit the cancelling noise. As before, a thin cable connects the earplug to the external ANR system. This device is still a prototype and materials used in the construction have not been optimised for either their sound attenuating properties or comfort. However, since the earplug has been moulded to the subject’s ears, the occurrence of friction or pressure hot-spots is greatly reduced compared to the universal-fit model.

In order to facilitate the research into the combination of the ANR earplug and an ANR earmuff, a number of personally moulded earplugs have been manufactured that contain an additional miniature microphone positioned at the tip of the earplug, which will be positioned in the ear canal itself. The electronics have also been modified so that the ANR sense microphone, used to measure sound pressure level for the ANR system, can be read by the experimental operator. This means that the personally moulded ANR earplugs have two microphones that can be used for observation of the sound pressure level within the earplug and within the ear canal.

The headset chosen for use during this research is a Peltor Optime III device that has high levels of passive sound attenuation. One of these Peltor headsets was fitted with a high quality ANR system that has been proven in other types of earshell and that is currently being flown by the UK Royal Navy.

A full range of experiments investigating the sound attenuation achieved by a personally moulded passive earplug, the personally moulded ANR earplug described above, a Peltor Optime III headset and the modified Peltor ANR headset are described in a companion paper [13]. The following sections of this paper examine the validity of two common measurement techniques used to determine the sound attenuation afforded by double ANR protection.

2.0 MEASUREMENT TECHNIQUES

Two different techniques are widely used to determine the sound attenuation of different types of hearing protector. The Real Ear at Threshold (REAT) technique uses a human subject to detect their threshold of hearing when no hearing protection is worn (unoccluded) and with the hearing protector in place (occluded). This method can be used with all types of passive hearing protectors, including earmuffs, headsets and earplugs since it is reliant on the subject’s perceived change in threshold and not on any instrumentation. The second technique, known as Microphone in Real Ear (MIRE), uses miniature microphones that are mounted either into the earmuff of a hearing protector using clips or by fixing them at the entrance to the ear canal of a subject [14]. A noise field is generated and the noise level at the microphone is measured both with and without the hearing protector in place. The subject in this case is purely used as an accurate ‘dummy’ head and the inter-subject variation in the measured sound attenuation will be due to the difference in the fit of the hearing protector on the individual. The location of the microphone (in the earmuff or in the ear canal) can provide different sound attenuation figures for the same hearing protector; however, if the same location is used throughout an experiment, the results are highly repeatable across subjects and provide lower standard deviations than exhibited using the REAT technique. Microphones can also be placed on the occluded side of an earplug to perform MIRE as long as sufficiently small microphones and very narrow wires are used to connect the microphone to the instrumentation.

REAT has the advantage over MIRE in that the sound attenuation determined by this technique is what the subject hears, regardless of how the sensation of hearing is generated. However, at low frequencies (particularly at 63 Hz and to a lesser extent at 125 Hz) physiological noise due to the noise of blood flow and the pulsing of the blood at heartbeat frequencies produces masking effects. This masking noise is noticeable when the ear is occluded and is known as the occlusion effect. Since the masking does not occur when the ear is unoccluded, at low frequencies REAT will over-estimate the sound attenuation of a device [15]. Additionally, REAT cannot be used with ANR devices since the residual noise generated by the active noise system will also cause masking at some frequencies up to 2kHz and will therefore raise thresholds and again over-estimate the sound attenuation achieved. REAT can, of course, be used with standard passive earplugs and with ANR devices that are not used in their active mode, i.e. with the active system turned off. REAT will also measure the sound that reaches the cochlea via bone conduction pathways. However, since REAT is performed at threshold, whereas MIRE is measured in a sound field at high noise levels, REAT assumes that the hearing protector is linear with increased noise level, in order that the sound attenuation measured at threshold will still be valid at higher noise levels.

MIRE measures the sound pressure level in the ear canal or within the earmuff, not at the cochlea. Therefore the effects of blood flow and bone conduction are ignored. It has the advantage over REAT that it can be used with ANR, since MIRE measurements are made in high ambient noise levels that greatly exceed the masking produced by the residual noise of the active system. The main disadvantage of MIRE over REAT is that it does not necessarily measure what the subject hears, since it misses the contribution from the bone conduction pathways. Hence, at higher frequencies MIRE will over-estimate the sound attenuation.

3.0 EXPERIMENTAL PROCEDURE

3.1 Measurement Facility

QinetiQ have a high noise facility that has been used to measure sound attenuation by MIRE for many years. The facility and techniques used to measure sound attenuation by this method follow ANSI S12.42 1995 [16]. Recently, the facility has been upgraded so that measurements can be taken by the REAT technique in accordance with ANSI S12.6 1997 [17]. The capability has also been extended so that as well as recording threshold levels at the frequencies set out in the standard REAT method, thresholds can be measured across the whole third-octave spectrum at centre frequencies from 50Hz to 12kHz. This enables REAT results to be compared with MIRE results in all one third-octave bands, or to pursue detailed investigation of REAT at frequencies of interest.

An extensive range of measurements using combinations of ANR earmuffs and passive and ANR earplugs have been undertaken, and a full description of these measurements and results are provided in the companion paper [13]. In order to examine the effect of the measurement technique, only a small number of these measurements are examined in this paper.

3.2 Measurement of ANR Residual Noise

In order to establish the level of residual noise generated by the personally moulded earplugs, a measurement was taken of each earplug fitted in turn to a Brüel and Kjær (B&K) artificial ear type 4153. The artificial ear was located within an Anechoic Chamber and the earplugs were sealed to the artificial ear using a re-useable adhesive putty, taking care to prevent leakage and any associated feed-back within the ANR system. This provided the frequency range over which the residual noise was generated by each system, and thus indicates the frequencies at which REAT measurements should not be attempted with an ANR system in active mode. The sound pressure levels measured by this method will not correspond to

those encountered on the ear, since the volume that the system is working into on the artificial ear will be different from that on the human ear.

3.3 MIRE measurements

The MIRE measurements in these experiments were conducted on each of the subjects seated in a diffuse noise field of 120 dB(A) and wearing a combination of the passive and active earplugs and earmuffs. The sound pressure level at several points along the sound transmission path into the ear canal was measured with a series of microphones, as shown in Figure 1. The microphones either side of the earmuff, labelled microphones 1 and 2, are not discussed in this paper. The measurement positions of interest are microphone 3, the ANR sense microphone, and microphone 4, referred to as the probe microphone. The probe microphone is not positioned deeply in the ear canal and does not necessarily record the sound pressure level achieved at the tympanic membrane. MIRE measurements were obtained on Brüel and Kjær 2133 Dual Channel Frequency Analysers, taking a 16 second average when the signal in the room had become stationary, over the frequency range 31.5 Hz to 10 kHz. One measurement was taken at each microphone for each combination of hearing protectors.

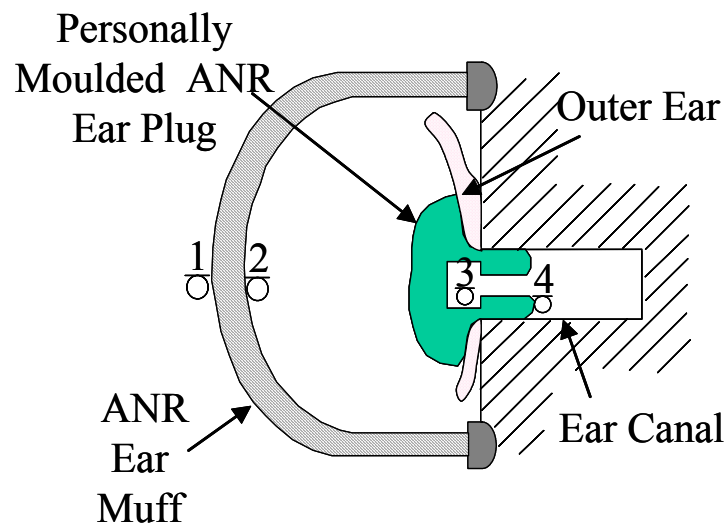


Figure 1: Measurement Microphone positions for MIRE with double ANR hearing protection. Microphones measure the SPL at: 1) the unoccluded field, 2) within the earmuff, 3) at the ANR sense microphone and 4) within the ear canal.

Since the ANR sense microphone and the probe microphone are embedded within an earplug, a method of achieving an unoccluded measurement had to be determined. The microphones from each earplug could obviously not be removed and placed in their respective positions at the entrance to the subject's ear canal and within it without destroying the earplug. One approach was to measure the sound pressure level with the earplugs suspended in the sound field at head height, disregarding any influence of the ear and head on the unoccluded sound field. A second method was to locate the earplug in an artificial ear canal, made of a 30mm length of aluminium tubing, 10 mm in diameter with walls 1 mm thick, and suspend this combination within the unoccluded sound field. However, this approach best approximates microphones that are positioned at the tympanic membrane, not at the open entrance to the ear canal.

The provision of the two microphones, the ANR sense microphone and the probe microphone located in the ear canal, enabled a comparison of the MIRE attenuation at each position. The probe microphone on these earplugs is easily fitted when the earplug is inserted and is located in a repeatable position. The wire leading from the earplug was taken over the top of ear, passing behind the ear and following the

depression between the jaw and the mastoid bone. This reduced the breakage of the acoustic seal of the headset to a minimum. The double ANR protection is shown in Figure 2 on a B&K Head and Torso Simulator for clarity.



Figure 2 Double ANR protection, cut-away shows personally moulded earplug and position of cable, minimising acoustic leakage.

The sound attenuation was measured for the personally moulded ANR earplug alone and in combination with the ANR headset. The ANR system of either device could be activated or the system left in its passive mode and the combinations discussed in this paper are:

- a) ANR headset passive, ANR earplug passive
- b) ANR headset passive, ANR earplug active

The ability to measure the systems in their passive mode is important for comparison with REAT measurements. The results of the attenuation measured by MIRE at the sense and the probe microphones for the ANR earplug alone and in combination with the ANR headset passive are reported here.

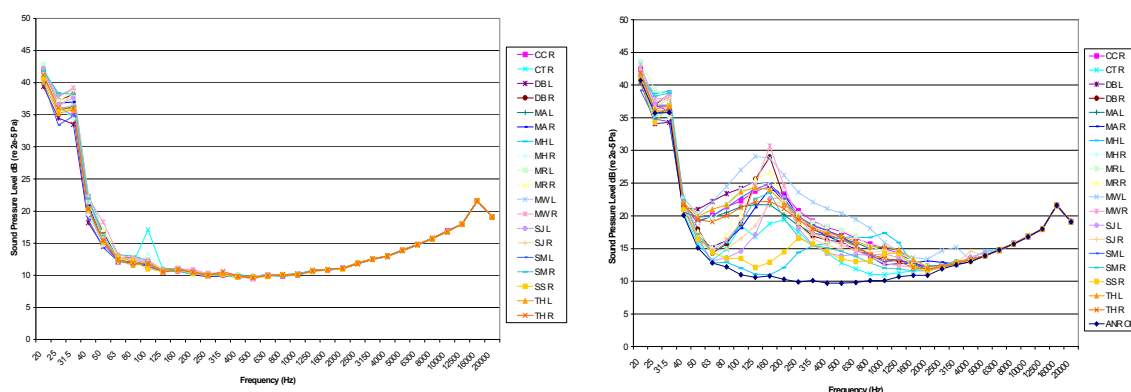
3.4 Comparison of MIRE and REAT measurements

REAT measurements were also undertaken for the same combinations of passive and active earmuffs and earplugs as for the MIRE measurements. Although REAT was not measured for the ANR earplug in its active mode, the ANR earmuffs could be used when active, since the earplug prevented the residual noise being heard by the subject. Therefore, the combination (a) above was also measured using REAT over an extended number of third-octave bands. As we were particularly interested in whether the limits of bone conduction had been reached, REAT was performed at the centre frequencies required by the standard test method (125 Hz, 250 Hz, 500 Hz, 1 kHz, 2 kHz, 4 kHz and 8 kHz) with the addition of the third-octave bands centred on 63 Hz and at 2.5 kHz, 3.15 kHz, 5 kHz, 6.3 kHz, and 10 kHz. An experimenter-supervised fit was used throughout and one REAT measurement performed for each combination of hearing protectors.

4.0 RESULTS AND ANALYSIS

4.1 ANR Residual Noise

The measurements of the residual noise produced by the personally moulded ANR earplugs when positioned on the artificial ear are shown in Figure 3. The measurements indicate that about half of the earplugs generate an ‘electronic hiss’ over the frequency range 50 Hz to 2 kHz, with a similar number generating noise over the slightly narrower band of 80 Hz to 2 kHz. Two earplugs only generate the noise from 160 Hz to 2 kHz and therefore do not exhibit a peak level at the same frequency as the other earplugs. For the majority of the earplugs, the maximum noise is generated at 160 Hz. The range of frequencies of the residual noise corresponds to the low frequencies that the ANR system is designed to reduce, centred on the particular frequency of 160 Hz, and extends up into the mid-frequency range into those regions where ANR systems reinforce or enhance the noise.



Figures 3a and 3b Sound pressure level produced by personally moulded earplugs mounted on an artificial ear. a) ANR system off and b) ANR system on.

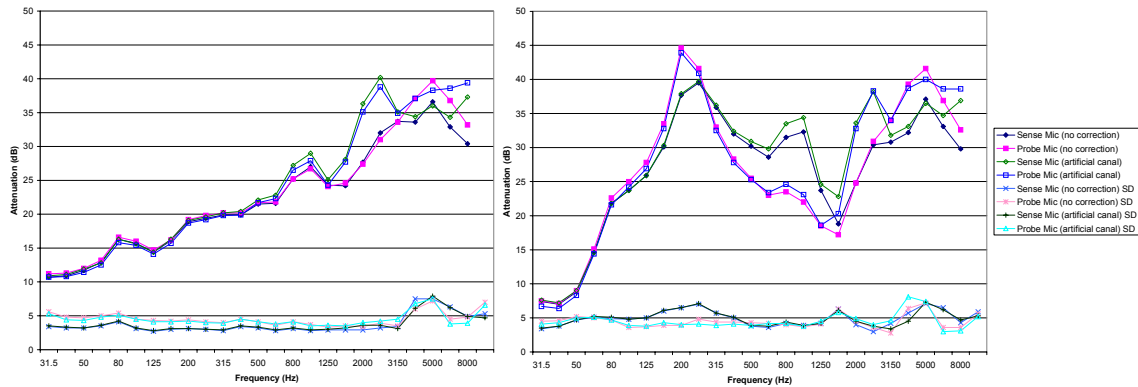
Hence, the REAT technique cannot be used with these personally moulded ANR earplugs at frequencies below 2 kHz, as the residual noise of the system will produce masking and therefore over-estimate the sound attenuation provided by the earplug at these frequencies.

4.2 MIRE measurement microphone

The levels of attenuation achieved by the ANR earplug were measured using the MIRE technique for the earplug in its passive and its active modes of operation. In order to calculate the attenuation, the sound pressure level measured in the occluded ear should be subtracted from the sound pressure levels measured with the ear unoccluded.

Figure 4 illustrates the differences in attenuation achieved using the different methods of measuring the unoccluded sound field for the ANR earplugs in their passive and active modes. The curves are the mean of the attenuation calculated for various numbers of subjects; with ANR off, sense microphone (18 ears out of a maximum 20 ears, i.e. 10 pairs of ears), probe microphone (17 ears), with ANR on, sense microphone (16 ears) and probe microphone (14 ears). Measurements were not taken for those subjects where the microphone failed to work and for those for whom the ANR produced feed-back when switched on. These were caused by failures in the manufacture of the earplugs and the feedback was caused by an incomplete acoustic seal of the earplug in the ear canal. Several attempts were made to re-set those earplugs generating feed-back, however, it was not always possible to prevent feed-back from occurring, suggesting errors in the moulding of the earplug.

Figure 4a shows a distinct peak in the response at 2500 Hz for the passive attenuation data corrected by the artificial ear canal. This peak corresponds to the resonance of the artificial ear canal. A similar peak can be seen in Figure 4b for the attenuation with the ANR system on. There are differences between the sense and the probe microphone at frequencies above 3.15 kHz but these occur for both the ANR off and on, with and without the use of the artificial ear canal and are likely to be due to the difference in the position of the sense and probe microphones.

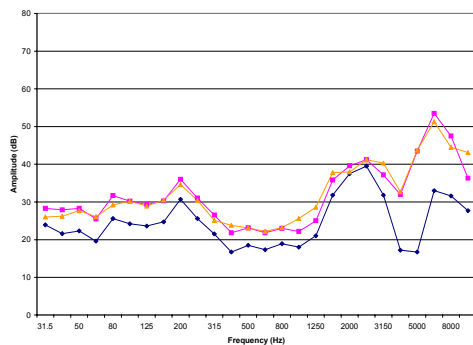


Figures 4a ANR off
Figure 4b ANR on
Mean attenuation and SD for ANR earplug measured at Sense and Probe microphone positions. Unoccluded measurement in free field corrected by use of an artificial ear canal and not corrected.

When the ANR system is activated, the differences between the position of the ANR sense microphone and the probe microphone become more apparent. The peak in the ANR response is at 200 Hz for both microphones, but the sense microphone shows an additional peak between 800 Hz and 1 kHz. The probe microphone has a minor increase in the attenuation in this region, as do both microphones when measured with the ANR system off. The slight peak in the region of 800 Hz is not particularly noticeable on the sound pressure level data used to obtain the unoccluded condition and so may be an occlusion effect within the ear canal. However, when the ANR system is off the difference between the sense and probe microphone at this point is negligible; with the ANR system active, the peak at the sense microphone becomes much larger than the peak for the probe microphone. The difference between the sound pressure levels at the two microphones is primarily a function of the positions of the microphones and the operation of the ANR system. The ANR has been optimised at the sense microphone position, so the best measure of the sound attenuation achieved by the system will be at the sense microphone. The probe microphone measures the sound attenuation in the ear canal and the difference between the measurement at this position and at the sense microphone provides an indication of the effectiveness of the ANR beyond the earplug. The constriction of the ear canal portion of the earplug also acts as a low-pass filter and this will also reduce the effectiveness of the ANR within the ear canal. Since the probe microphone is only located part-way into the ear canal, it is not measuring the sound pressure level at the eardrum and so the acoustics of the ear canal itself may be influencing the levels of attenuation measured.

An attempt was made to further investigate the differences between the sense microphone and the probe microphone. Personally moulded earplugs had also been manufactured for a B&K Head and Torso Simulator (B&K HATS); of these, the left earplug did not appear to function correctly. Figures 5a and 5b show the attenuation achieved on the B&K HATS with the Right personally moulded ANR earplug. The attenuation determined at each microphone should provide an indication of what is causing the differences. The curves for the probe microphone and for the microphone located within the B&K HATS at the 'eardrum' have an almost identical response. This is due to the short length of the ear canal within the B&K HATS, which results in the probe microphone being positioned very close to the 'eardrum'

microphone. The sense microphone curve is approximately 5 dB below that of the probe microphone curve up to 2 kHz, as shown in Figure 5a for the ANR in passive mode. This off-set does not occur on the human subjects, where in the passive mode the earplug sense and probe microphones provide very similar results (Figure 4a).



Figures 5a ANR off
Attenuation achieved measured by MIRE on B&K HATS using Personally moulded ANR earplug

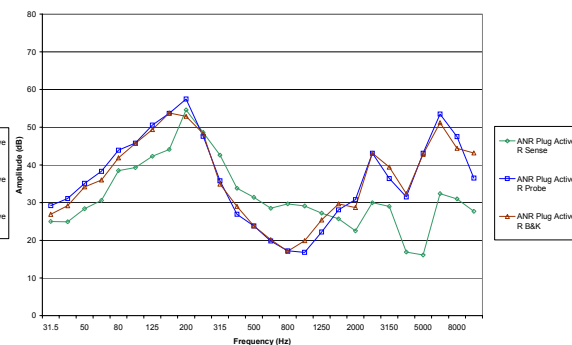
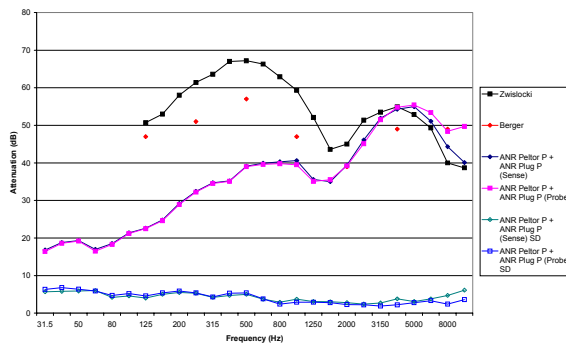


Figure 5b ANR on

The peak in the attenuation for all three microphones on the B&K HATS is at 200 Hz and this is a function of the electronics in the ANR system. The difference in the frequency range of the attenuation with the ANR system active for the HATS and for the human subjects may be due to the differences in the ear canal of the human subjects compared to the simulated canal of the HATS. The ANR sense microphone again provides higher levels of attenuation at frequencies greater than the ANR centre frequency compared to the probe microphone, which is indicative of the low-pass filter effect caused by the constriction in the earplug through the ear canal portion, which is particularly noticeable when the ANR is active.

The attenuation measured by MIRE for the combinations of the passive ANR Peltor headset worn together with the personally moulded ANR earplugs in Passive and Active mode are shown in Figures 6 a) and b). The MIRE attenuation was again measured at both the sense and the probe microphones, but in the graphs below only the measurements with the uncorrected ear canal data have been shown. The attenuation data when the artificial ear canal was used for the unoccluded measurement provided a higher level of attenuation around 2.5 kHz and so the uncorrected version gives lower attenuation, even though this may be an underestimate of the total attenuation provided by the double protection. The curves for the bone conduction thresholds of Zwislocki [9], achieved by uncomfortable laboratory techniques, and for Berger [11], achieved by the use of deeply inserted E.A.R. foam earplugs together with a headset, are given for comparison.



Figures 6a ANR Headset Passive and Active with ANR earplug Passive

Attenuation of Double ANR system using MIRE. Measured at probe and sense microphones with no correction applied to unoccluded measurement.

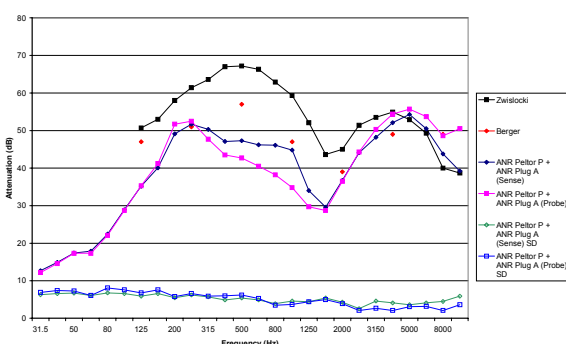


Figure 6b ANR Headset Passive and Active with ANR earplug Active

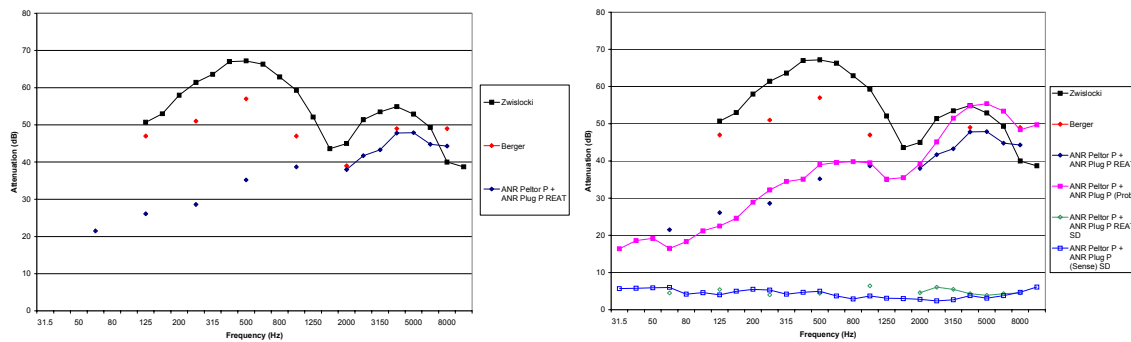
The passive ANR Peltor headset plus passive ANR earplug is a mean of 18 measurements (i.e. 18 different ears) for the sense microphone and 16 measurements for the probe microphone. The passive ANR Peltor headset plus active ANR earplug is a mean of 15 measurements for the sense microphone and 14 measurements for the probe microphone. Again, measurements were excluded for those earplugs where the microphone did not work and where activation of the ANR system resulted in feedback in the ear canal.

The differences between the sense and the probe microphones are again most noticeable around 1 kHz when the ANR earplug is in its active mode, whereas the differences at 3.15 kHz are much less when using double protection than when the earplug is worn alone. The combination of the ANR headset and active ANR earplug reaches the limits of bone conduction set out by Berger at 200 to 250 Hz for both the probe and the sense microphone. At 1 kHz the sense microphone also shows levels of sound attenuation that reach Berger's bone conduction limits. The ANR system has been optimised at the sense microphone, which suggests that if the ANR system could be optimised at the probe microphone rather than at the sense microphone the bone conduction thresholds could be achieved over a broader frequency band.

For both cases, with the ANR earplug passive and active, the MIRE measurement shows levels of attenuation greater than the bone conduction limits of Zwislocki and of Berger at frequencies above 2 kHz.

4.3 Comparison of MIRE and REAT measurements

For double ANR protection, the MIRE measurements show that the bone conduction limits at mid to high frequencies have been exceeded. However, the MIRE technique cannot measure the sound reaching the cochlea by bone conduction routes at frequencies above 2 kHz. Therefore, a similar set of measurements was performed using REAT and Figure 7a shows the attenuation achieved by REAT under the ANR Peltor headset and the personally moulded ANR earplug both operating in their passive modes. This is the mean of the results for 10 subjects. A comparison of the REAT measurement and the MIRE probe microphone measurement for this condition is shown in Figure 7b.



Figures 7a REAT with ANR Headset Passive and Active and ANR earplug passive **Figure 7b comparison of REAT with MIRE**
Attenuation of Double ANR system

The graph of attenuation measured by REAT reaches the limits of bone conduction as set out by Berger between 2 kHz and 5 kHz. The comparison with the MIRE technique shows that this method is over-estimating the attenuation because it cannot measure the sound transmitted via bone conduction. A t-test performed on the results of these two techniques shows that there is a significant difference between the methods above 2.5 kHz ($p=0.0009$, $\alpha=0.05$). Below 2 kHz, MIRE produces slightly higher levels of sound attenuation between 200 Hz and 2 kHz than REAT, although these differences are only 4 dB. The t-test shows a significant difference between the methods at 250 Hz and 500 Hz ($p=0.045$, $p=0.025$, $\alpha=0.05$), but no difference at 1 kHz, 2 kHz and 2.5 kHz. At 63 Hz and at 125 Hz REAT produces higher levels of sound attenuation than MIRE, which is due to the physiological noise produced in the ear canal when occluded and is a function of this technique. However, the t-test analysis of the results shows that there is no significant difference between the two methods at these two frequencies ($p=0.083$, $p=0.091$, $\alpha=0.05$).

5.0 MEASUREMENT OF ATTENUATION: REAT VS MIRE

The introduction of active noise reduction into headsets and earplugs has resulted in difficulties in choosing the correct technique to measure the levels of sound attenuation actually experienced by a human subject. The two techniques open to the investigator have their own peculiarities and present different difficulties in interpretation. The influence of bone conduction on the levels of attenuation achieved as experienced by the subject suggests that REAT is the better technique, however, the problems of masking presented by the residual noise of the ANR system lead to the choice of the MIRE technique.

It has been suggested that REAT should be used to determine the level of attenuation achieved when the ANR system is in its passive mode (i.e. the ANR system is turned off), and that the effect of the ANR system as measured by the MIRE technique should be added to this to present the total attenuation at the ear. This results in the device under test being measured three times per subject, once for REAT with the ANR system off and twice for MIRE with the ANR system off and on. It is the difference between the two MIRE measurements that produces the purely 'active' attenuation that is added to the passive REAT data. REAT is usually measured at the centre third-octave of an octave band, whereas MIRE can be measured at third-octave bands covering the whole frequency spectrum. MIRE is a faster technique than REAT, requiring only tens of seconds to complete a single measurement, however, it seems excessive to acquire this data only to ignore two thirds of it. With current hearing protection devices used singly, the limits of bone conduction are not met and so it would seem that MIRE alone would be sufficient to provide the attenuation due to an ANR device.

The attenuation provided by double hearing protection, e.g. a headset and earplug in combination, will result in the bone conduction limits being met at frequencies above 2 kHz. The residual noise of the ANR

system used in the personally moulded earplugs extended up to 2 kHz, which corresponds to those frequencies where the ANR system reduces the sound pressure level in the ear canal and those frequencies where the ANR system reinforces and thereby increases the sound pressure level in the ear canal. Although MIRE provides the best measure of attenuation below 2 kHz, above 2 kHz it is probable that REAT should be used. If the decision to use two techniques has already been taken, as mentioned above, it may perhaps be an improvement to use different techniques over different parts of the frequency spectrum. Therefore, it would seem appropriate to use MIRE to measure the sound attenuation at frequencies up to 2 kHz and for REAT to be used at frequencies above 2 kHz.

However, the difficulty of choosing the position of the microphone with which to perform the MIRE measurements now becomes significant. If the ANR system under test is an earplug, the sense microphone could be used for the measurement if the electronics allows the microphone signal to be tapped, as shown in these experiments. This may over-estimate the levels of attenuation achieved by the device, as it will measure the sound pressure level when optimally reduced. If the ANR sense microphone cannot be used, then a microphone must be positioned within the ear canal with the signal fed back underneath the earplug by a means that does not compromise the acoustic seal. As has been shown already in this paper, the results produced by the sense microphone and the probe microphone seated in the ear canal can be significantly different over parts of the frequency spectrum. If the ANR sense microphone does not provide a measure of the sound pressure level at the eardrum, then MIRE will not give a correct indication of the attenuation experienced by the subject. If the probe microphone, situated in the ear canal, is located at a position where the ANR system is not reducing the sound pressure level sufficiently, this microphone will not give a true indication of the attenuation experienced by the subject.

6.0 CONCLUSIONS

The measurement of the sound attenuation provided by ANR devices worn in combination leads to conflicting decisions when choosing the measurement technique. The use of REAT is not possible at those frequencies when the ANR system produces masking noise. On the other hand, the likelihood of reaching the limits of bone conduction above 2 kHz precludes the use of MIRE. As a result of experiments on human subjects together with a unique personally moulded ANR earplug incorporating two measurement microphones, it is suggested that MIRE should be used to obtain the attenuation below 2 kHz and that REAT should be used above 2 kHz. Further investigation should be undertaken into the location of the microphone for MIRE.

REFERENCES

- [1] UK Noise at Work Regulations: 1989
- [2] EU Physical Agents (Noise) Directive 2003/10/EC
- [3] A. Behar *Sound Attenuation from Combinations of Earplugs and Earmuffs* Applied Acoustics Vol 32 1991 149-158
- [4] S. M. Abel, N. M. Armstrong *The Combined Sound Attenuation of Earplugs and Earmuffs* Applied Acoustics Vol 36 1992 19-30
- [5] J. Tonndorf. *Bone Conduction*, In *Foundations of Modern Auditory Theory Vol II*. Academic Press London, 1972

- [6] S. M. Khanna, J. Tonndorf., J. E. Queller *Mechanical parameters of hearing by bone conduction* J. Acoust. Soc. Am. Vol 60 (1) July 1976 139-154
- [7] S. Stenfelt, T. Wild, N. Hato, R. L. Goode *Factors contributing to bone conduction: The Outer Ear* J. Acoust. Soc. Am. Vol 113 (2) Feb 2003 902-913
- [8] S. Stenfelt, N. Hato., R. L. Goode *Factors contributing to bone conduction: The Middle Ear* J. Acoust. Soc. Am. Vol 111 (2) Feb 2002 947-959
- [9] J. Zwislocki *In Search of the Bone-Conduction Threshold in a Free Sound Field* J. Acoust. Soc. Am. Vol. 29 (7) July 1957 795-804
- [10] N. A. Watson, R. S. Gales *Bone-Conduction Threshold Measurements: Effects of Occlusion, Enclosures and Masking Devices* J. Acoust. Soc. Am. Vol. 14 (4) April 1943 207-215
- [11] E. Berger *Laboratory Attenuation of Earmuffs and Earplugs Both Singly and in Combination* Am. Ind. Hyg. Assoc. J. 44 (5) 1983 321-329.
- [12] E. A. Goodfellow *A Prototype Active Noise Reduction In-Ear Hearing Protector* Applied Acoustics Vol 42 1994 299-312
- [13] C. Tubb, S. E. Mercy, S. H. James experimental paper
- [14] G. Rood *The In-Situ Measurement of the Attenuation of Hearing Protectors by the use of miniature microphones* In *Personal Hearing Protection in Industry*, Raven Press, New York, 1982
- [15] E. Berger, J. Kerivan *Influence of Physiological Noise and the Occlusion Effect on the Measurement of Real-Ear Attenuation at Threshold* J. Acoust. Soc. Am. 74 (1) July 1983
- [16] *Microphone in Real Ear and Acoustic Test Fixture Methods for the Measurement of Insertion Loss of Circumaural Hearing Protection Devices* American National Standard ANSI S12.42 1995 (R 2004)
- [17] *Methods for Measuring the Real Ear Attenuation of Hearing Protectors* American National Standard ANSI S12.6 1997



Investigating Double Hearing Protection using Human Subjects

Christopher Tubb, Susan Mercy and Soo James

QinetiQ

Human Factors in Acoustics and Vibration

Cody Technology Park, Ively Road, Farnborough

Hampshire, GU52 OLX, UK

Tel: +44 (0) 1252 392668, Fax: +44 (0) 1252 397034

E-mail: ctubb@qinetiq.com / semercy@qinetiq.com / shjames@qinetiq.com

ABSTRACT

In the increasingly noisy military environment and with the growing need to protect the hearing of military personnel, the use of double protection, earplugs and earmuffs used in combination, has become more and more prevalent. Furthermore Active Noise Reduction (ANR) headsets and earplugs appear to offer a means of increasing the attenuation of these double protection systems. However, it has been shown that the attenuation afforded by double protection is not a simple additive process, the combined attenuation tends to be less than the sum of the individual earplug and earmuff attenuation figures. Also, at the levels of attenuation provided by double protection, bone conduction pathways start to play an important part in the sound level received at the ear. A fuller understanding of these processes is required if the full benefit of double protection is to be achieved in future devices.

Most previous assessments of the attenuation the afforded by double protection and bone conduction limits have been carried out using the REAT (Real Ear At Threshold) test technique. However, this technique is limited for predicting the attenuation from active systems due to the masking effect of the residual electronic noise. Other work has used ATF (Acoustical Test Fixtures) for predicting the attenuation provided by ANR double protection. However, this can only take into account bone conducted noise by the use of mathematical models. The study presented here employed both REAT and MIRE (Microphone In Real Ear) test techniques to investigate the attenuation given by both passive and active noise protection devices when used in combination. The MIRE experimentation was carried out in noise fields up to 120dBA and the sound pressure level, in the earshells and ear canals, was monitored via miniature microphones and probe microphones fitted in personally moulded earplugs respectively. The study included testing on ten subjects, where all the hearing protectors were fitted as they would be used in the field. The MIRE and REAT experiments have both shown that headsets with varying performance of attenuation appear not to have a major effect on the overall attenuation when worn over earplugs and it is the performance of the earplugs that has a more direct effect on the attenuation of the double protection system. However, from direct measurements of the Sound Pressure Level (SPL) under the earmuffs during the MIRE procedure it can be concluded that the attenuation of the earmuff remains constant, implying that there is an interaction between earmuff and earplug that is leading to the degradation of the attenuation afforded by the double protection system.

This work was funded by the Human Sciences Domain of the UK Ministry of Defence Scientific Research Programme.

1.0 INTRODUCTION

The noise environment that military personnel are subjected to is becoming increasingly harsh, with noise levels to which aircrew are exposed rising from 110 dB to a maximum of 120 dB in the last thirty years. Fortunately, the optimisation of passive hearing protectors and the introduction of active noise control systems have meant that the ability of personal hearing protection to reduce noise at the ear has kept pace with the increased noise risk.

In future aircraft, such as the Joint Combat Aircraft (JCA) and Euro Fighter (EF) Typhoon, the noise levels in the cockpit are expected to increase only marginally. However, deck-crews and possibly ground-crews will be exposed to the environmental noise generated by these aircraft and the noise levels to which they will be exposed are predicted in some quarters to be as high as 150 dB. At these noise levels single protection earmuffs or earplugs worn alone will not provide sufficient attenuation to bring noise dose within legislative criteria. Double protection systems, earplugs and earmuffs worn in combination, are the obvious answer to increase the attenuation afforded. However, double protection is a complex process and the overall attenuation is not the addition of the two single protector attenuations. Various empirical techniques have been used to attempt to calculate the total attenuation achieved by double protection [1][2], but the most accurate method is still that of direct measurement.

One of the factors that has been found to limit double protection systems is bone conduction (BC), sound reaching the cochlea by pathways other than the standard air conduction (AC) pathway via the ear canal. The limiting effect occurs when the BC influence becomes greater than that of the AC pathway, i.e. when the air conducted pathway is sufficiently occluded. It has been found that above 2kHz earplugs and earmuffs worn in combination are sufficient for bone conduction limits to be reached [3][4][5]. This explains, at these frequencies, why the simple addition of single hearing protector attenuations does not result in the overall attenuation measured. At the lower frequencies it has been conjectured that the limitation of attenuation is the mechanical coupling between the earplug and earmuff [3]. More over it has been found that the individual attenuation of a headset worn over the earplug does not effect the overall attenuation [4][6][7], i.e. a poor attenuating headset or a good attenuating headset worn over the same earplug gives the same overall performance. Behar and Kunov [6], in their work, found for poor attenuating earplugs there was no advantage given by better hearing protector attenuation but for good attenuating earplugs, the attenuation of the headset did benefit the overall attenuation. However, this work used an Acoustical Test Fixture (ATF) that does not monitor bone conduction and no corrections were made to take account of it.

Therefore to further increase hearing protection performance, bone conduction has to be considered. Tonndorf identified three specific bone conduction pathways [8]:

- External ear component of bone conduction: the external ear canal walls vibrating cause pressure perturbations in the ear canal, which then excites the eardrum in the standard manner.
- Middle ear component of bone conduction: the inertial vibrations of the ossicles (middle ear bones).
- Inner ear component of bone conduction: mechanical distortion of the cochlear walls.

The external ear component has been identified as the dominant pathway below 2kHz, when the ear canal has been occluded at it's entrance [9][10]. This effect has been termed the occlusion effect, as when the ear is unoccluded the ear canal acts as a high pass filter, which reduces the external ear BC pathway at low frequencies [8]. Stenfelt et al. identified this component of bone conduction to be specifically the oscillation of the soft ear canal tissue as opposed to the bony portion of the ear canal. Deeply inserted earplugs (inserting into the bony portion of the ear canal) circumvents this effect [3][4]. Above 2kHz and when the ear is unoccluded, the middle ear and inner ear pathways are dominant. However, the relative effect of the two pathways is hard to attain.

The main aim of the current study was to investigate the effect of Active Noise Reduction (ANR) systems when used in double protection systems. Berger et al. [7], used an ANR headset in active mode over deeply inserted earplugs and found a statistically significant increase in attenuation as compared to the ANR off mode, but questioned practical benefit of the increase. For the current study the ANR headset system chosen for use in the testing of double protection systems was a Peltor Optime III, modified with a production ANR system that is currently in use in flight helmets worn by UK Navy Aircrew.

In previous work the highest low frequency BC limits were reached using deep inserted earplugs, because as mentioned these circumvent the external ear BC pathway. ANR earplugs provide a possible method for attaining higher levels of attenuation, close to previously found BC limits, without the impracticality of deep inserted earplugs. The Royal Aerospace Establishment (RAE) developed a ‘top-hat’ style ANR capsule that was fitted into the ear through a shortened standard foam earplug [11]. Although this device worked well, the earplug did not sit well under a headset and was therefore uncomfortable when used in double protection. QinetiQ have now developed the design so that now a personally moulded earplug is used to house the ANR components and these fit more comfortably under earshells.

The work reported in this paper presents the attenuation data attained for double protection systems from both the objective Microphone In Real Ear (MIRE) method [12] and the subjective Real Ear At Threshold (REAT) method [13]. MIRE was used because REAT cannot measure “direct” ANR performance due to the residual electronic noise of these systems. REAT was used because MIRE does not take into account the middle and external ear BC pathways because it only measures the sound pressure level in the ear canal. A full discussion of the measurement techniques is presented in a companion paper [14].

2.0 EXPERIMENTATION

In order to investigate the protection provided by double protection a number of experiments were devised to measure the levels of sound attenuation afforded by these devices. These experiments involved four individual hearing protection devices, a passive and an ANR headset and a set of passive and ANR earplugs (see Table 1). Two independent experiments were carried out to calculate the attenuation of the hearing protectors worn both singly and in combination (earplugs + headset). The first experiment was based on an objective microphone method and the second experiment based on a subjective method. The microphone method uses the MIRE (Microphone In Real Ear) standard ANSI S12.42 1995 [12] and the subjective method uses the REAT (Real Ear At Threshold) standard ANSI S12.6 1997 [13]. The REAT method has been extended to include all of the third-octave bands from 50Hz to 12.5kHz, to allow direct comparison with MIRE results.

Hearing protector	Type	Graph Key
Peltor Optime III	Passive Headset	Passive HS
Modified Peltor Optime III	ANR Headset	ANR HS (off/on)
Personally moulded passive earplug	Passive Earplug	Passive EP
Personally moulded ANR earplug	ANR Earplug	ANR EP (off/on)

Table 1: Hearing protectors used

The signal from the ANR sense microphone in the ANR earplugs could be accessed by the experimenter and the ANR earplugs were also designed to incorporate an additional microphone located at the ‘tip’ of the earplug, positioned within the ear canal (see Figure 1). This additional microphone is known as the ‘probe’ microphone as it measures the actual sound pressure level experienced in the ear canal. The signals from both microphones were fed along a thin cable, routed over the subject’s ear and down through the depression between the subject’s jaw and mastoid bone. Therefore, when the earplug was worn underneath a headset, the cable did not break the acoustic seal of the headset.

Passive versions of the personally moulded earplugs were also made for each subject. These contained a 'probe' microphone in the same location as for the ANR earplug, but did not contain any ANR components. The cable from the probe microphone was routed around the subjects' ears as described above.

2.1 MIRE Experimental Procedure

The laboratories at QinetiQ have a custom built reverberation chamber that allows generation of a diffuse sound field up to 120dB(A). To assess the attenuation of the hearing protection over a range of noise levels measurements were made at five noise levels, from 80dB(A) to 120dB(A) in 10dB(A) steps. A diffuse pink noise field was used, which has a flat 1/3 octave band spectrum.

Measurement No.	Measurement Condition
0	Noise field calibration, with BK4134 mic. at the centre of subject head position.
1	Passive Earplug: Probe mic. unoccluded in free field, at the five noise field levels
2	Passive Earplug: Probe mic. unoccluded in "ear canal simulator", at the five noise field levels
3	ANR Earplug: Probe, Sense mics. unoccluded in free field, at the five noise field levels
4	ANR Earplug: Probe, Sense mics. unoccluded in "ear canal simulator", at the five noise field levels
5	Passive Headset
6	Passive Earplug
7	Passive Earplug + Passive Headset
8	Passive Earplug + ANR Headset (off)
9	Passive Earplug + ANR Headset (on)
10	ANR Headset (off)
11	ANR Headset (on)
12	ANR Earplug (off)
13	ANR Earplug (on)
14	ANR Earplug (off) + ANR Headset (on)
15	ANR Earplug (off) + ANR Headset (off)
16	ANR Earplug (on) + ANR Headset (off)
17	ANR Earplug (on) + ANR Headset (on)
18	ANR Earplug (off) + Passive Headset
19	ANR Earplug (on) + Passive Headset

Table 2: MIRE experimental procedure for one subject run

The MIRE experimentation involved assessing the attenuation of all four hearing protectors worn singly and in combination; these devices are listed in Table 1. The full list of combinations tested is given in Table 2 and constitutes one complete set of measurements for a given subject. This set of measurements is referred to as a 'subject run'. The first four measurements in Table 2 were for calibration and setup of the reverberation chamber and subjects were only present from measurements 5 to 19. One subject run took, on average, four hours to complete. Subjects were allowed breaks from the testing but care was taken not to affect the consistency of the fit of the hearing protectors. For each hearing protector fit, subjects were encouraged to attain the best fit possible and were allowed to alter the fit whilst in the 80dB(A) noise field. This allowed the best possible attenuation to be attained, which is desirable for achieving the levels of attenuation necessary to investigate bone conducted noise.

Microphones were positioned along the air conductive sound transmission path and the positioning of the microphones during the experimentation is shown in Figure 1. The first pair of microphones were used to measure the attenuation of the earshell; microphone 1 was fitted to the outside of the headset, thereby measuring the unoccluded sound field, and microphone 2 was inside the earshell, allowing occluded

measurement at this point. The second pair of microphones were fitted in the personally moulded earplugs and allowed measurement of the attenuation of the earplugs themselves and of the total attenuation afforded by the double protection systems. These two were the probe and sense microphones. The probe microphones were fitted to both passive and ANR earplugs, but since the sense microphone was the microphone of the ANR system it was only present in the ANR earplugs. The output of the microphones was analysed on two Brüel & Kjær (B&K) 2133 third octave band analysers.

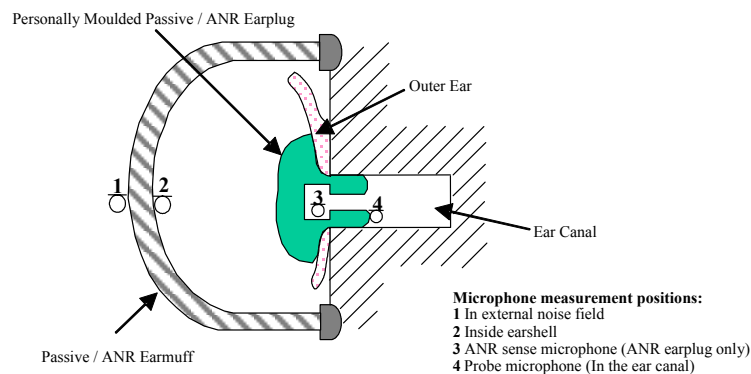


Figure 1: Microphone positions for the MIRE experiment

Measurements were made at the five noise levels, from 80dB(A) to 120dB(A) in 10dB(A) steps. In addition to the five noise levels, the noise floor of the microphones for each condition was also measured. The noise floor is the noise level measured at the microphones without any noise field present and includes the electrical response of the system. To get valid noise level readings from the microphones it is important that the noise level is 10dB higher than the noise floor of that microphone.

The MIRE test standard fully specifies the case of measuring circum-aural hearing protectors, however a method for measurement of MIRE for earplugs is not yet defined. Ideally, a probe microphone would be fitted in the ear canal for both the unoccluded and occluded case, but there are then problems fitting the earplug properly with a microphone and associated cables already in the ear canal. It is for this reason that personally moulded earplugs with probe microphones fitted were used to measure the sound pressure level in the ear canal. In order to obtain the attenuation, this still required the measurement of the unoccluded sound pressure level. The best measurement of the unoccluded sound field would be to remove the sense and probe microphones from the earplugs and locate them at their respective positions at the entrance to the subject's ear canal and within it. This was not practical as it would have required the destruction of all of the earplugs. Therefore, two techniques were adopted, the first involved placing the earplug microphones at the end of a metal tube (dimensions length: 0.03m, inside diameter: 0.007m) and measuring the SPL at the microphones for all 5 noise fields, measurements 03 and 05 from Table 2, for the passive earplug and ANR earplug respectively. This simulated the effect of having the microphones in an open ear canal, as these are the approximate dimensions of an average human ear canal. The second technique ignored the effect of the ear canal resonance and measured the diffuse sound field directly, by simply suspending the earplugs in the diffuse noise field, measurements 02 and 04 from Table 2, for the passive and ANR earplug.

2.2 REAT Experimental Procedure

The combinations of earplugs and headsets that were tested using the REAT method are given in Table 3, and this is a subset of the combinations used with the MIRE procedure. The reasons for this were firstly that the MIRE experimentation had highlighted some key areas of interest, in particular that the double protection systems were reaching the bone conduction limits at frequencies over 2kHz. Secondly, that REAT cannot be used with ANR systems when they are in their active mode of operation, because of the

masking effect of the residual ANR noise. Therefore, no ANR ‘on’ conditions were tested, apart from when the ANR headset was worn over earplugs. Finally REAT depends heavily on the subject’s ability to concentrate, hence the number of systems that an individual subject can reliably test in one sitting is limited. Subjects were required to complete the testing without removing their earplugs in order for the same ‘fit’ to be used for measurements when the headset was changed or removed.

REAT Measurement No.	Measurement Condition	Equivalent MIRE Measurement No.
1	Unoccluded	00
2	Passive Earplug	06
3	Passive Earplug + Passive Headset	07
4	ANR Headset (off)	10
5	ANR Earplug (off)	12
6	ANR Earplug (off) + ANR Headset (off)	14
7	ANR Earplug (off) + ANR Headset (on)	15
8	ANR Earplug (off) + Passive Headset	18

Table 3: REAT experimental procedure for one subject run

The goal was to identify whether bone conduction limits were being reached and, as mentioned above, the previously collected MIRE data indicated that bone conduction limits were being attained above 2kHz, but not at lower frequencies. Therefore, the REAT measurement was made at all the 1/3 octave band frequencies in the range 2kHz to 8kHz and at the standard REAT frequencies (125Hz, 250Hz, 500Hz, 1kHz, 2kHz, 3.15kHz, 4kHz, 6.3kHz and 8kHz). An additional measurement was also made at the 63Hz band. This enabled the region of bone conduction to be investigated in greater detail than the standard REAT method would normally allow.

The REAT method employed 1/3 octave band-limited white noise for the stimulus signal. The experiment was carried out in the reverberation chamber, where the suitability of the room and amplifier/speaker system has been subjected to thorough inspection to make sure it meets the requirements of the ANSI REAT specification.

3.0 RESULTS AND DISCUSSION

The companion paper [14], compares in more detail the review of the measurement techniques used. For the results presented here, all the MIRE earplug / double protection attenuation data has been attained from the probe microphone using the unoccluded freefield attenuation calculation, i.e. without the ear canal simulator. However, it has to be noted, that this will under estimate the MIRE measured attenuation, i.e. attenuation of the sound pressure level in the ear canal. The underestimate will be most noticeable around 2kHz – 3kHz bands, because the unoccluded measurement in the freefield does not measure the open ear canal resonance and the ear canal resonance shifts from around 2.7kHz to around 5.5kHz when the ear canal is occluded [15].

The REAT data presented is mean data for 10 subjects. During the MIRE experimentation the attenuation was shown to be linear over all the noise levels. Hence the MIRE earshell data is the mean over 100 measurements, 10 subjects, left/right ears, 5 noise levels. The earplug and double protection data was the mean over 80 measurements, as it was found that at the 80dB(A) noise level the occluded SPL at the earplug microphones was under the noise floor.

3.1 Single Protection

The mean MIRE attenuation measurements for the single hearing protector devices used during the trial are shown in Figure 2, where the Zwischlocki [3] and Berger 1983 [4] BC limits have also been included to provide a comparison. Figure 2a) shows the attenuation for the two headsets used, with the ANR headset showing performance with both ANR on and off. Fitting the ANR system to the Peltor Optime III has reduced the passive performance of the headset. The ANR recoups this “lost” attenuation and over the frequency range 80Hz to 630Hz the attenuation outperforms the unmodified headset. The ANR headset is only a prototype device and with more development in fitting the ANR system it is conceivable that the passive attenuation can be brought to match that of the unmodified device. These devices worn singly only approach the Berger bone conduction limit protection around 2kHz.

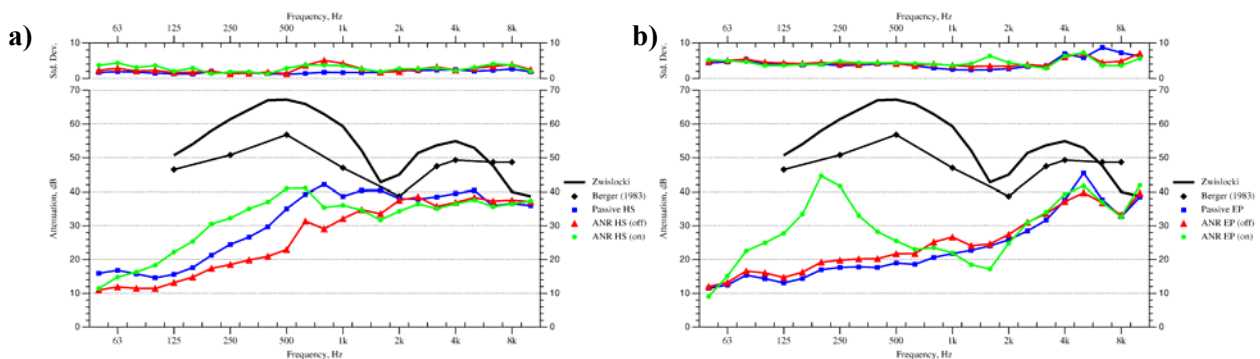


Figure 2: Mean and std. dev. of the attenuations (MIRE data) for the hearing protectors worn singly a) headsets b) earplugs

Figure 2b) shows the individual attenuations for the earplugs. The passive performance of the two earplug types is comparable and, in general, the ANR earplug has better attenuation except in the 1/3 octave band centred at 5kHz. It should be noted that the earplugs were prototypes and specifically designed for the fitting of an ANR system. The passive performance has not been optimised and provides a lower passive attenuation performance than conventional earplugs. The ANR performance provides an increase in attenuation at the low frequencies, with a maximum attenuation increase of 27dB in the 1/3 octave band centred at 200Hz. This brings the attenuation close to the Berger 1983 BC limit for this 1/3 octave band. Figure 2b) shows that the ANR earplug produces noise enhancement in the range of 800Hz to 2kHz (i.e. attenuation performance is worse than when the ANR is switched off). This effect is characteristic of ANR systems and due to a trade off between the levels of active performance at the lower frequencies and the bandwidth of performance. It can be seen that the same enhancement effect, to a lesser degree, is shown by the ANR headset (Figure 2a).

3.2 ANR and Double Protection

As already stated, prior research has indicated that double protection is not a purely additive process. At the higher frequencies this limitation may be attributed to bone conduction limits being reached. However at lower frequencies it has been conjectured that there is some mechanical coupling between the earplug and earshell that limits the performance. The effect of the ANR headset worn over an earplug was considered by Berger et al. [7], where it was found that there was a statistical increase in attenuation in the 125Hz frequency band with the ANR headset switched on. Although, Berger did comment that the relative effect of the 3dB increase on the overall attenuation is negligible. Figure 3 shows the MIRE measured mean attenuation for the ANR headset switched on and switched off worn over the passive earplug and the ANR earplug turned off. When worn over the passive earplug the difference in attenuation with ANR on and off is negligible and a t-test showed there to be no statistical significance ($p < 0.05$) in the difference

Investigating Double Hearing Protection using Human Subjects

between the attenuations. When worn over the ANR earplug there is a noticeable increase in attenuation between ANR on and off in the range 80 to 160 Hz, which was found to be statistically significant ($p < 0.05$). However at 125Hz there is only a 3dB increase in the double protection as opposed to an increase of 11dB for the headset worn singly. This minimal advantage provided by the headset ANR system is also shown in the REAT measurements, see figure 5b). At the higher frequencies (above 2kHz) there is no change in overall performance between ANR on and off for the device worn alone, therefore no change at these frequencies is not entirely unexpected. At the lower frequencies, particularly 250Hz and 500Hz, the additional active attenuation of the headset is not passed onto the overall double protection. At these frequencies the attenuation levels fall short of the bone conduction limits previously identified and therefore the limitation cannot be due to bone conduction limits being reached. Zwislocki [3] and Berger [4] also indicated this effect for headsets with different attenuating properties not changing the overall attenuation when headsets were worn over earplugs.

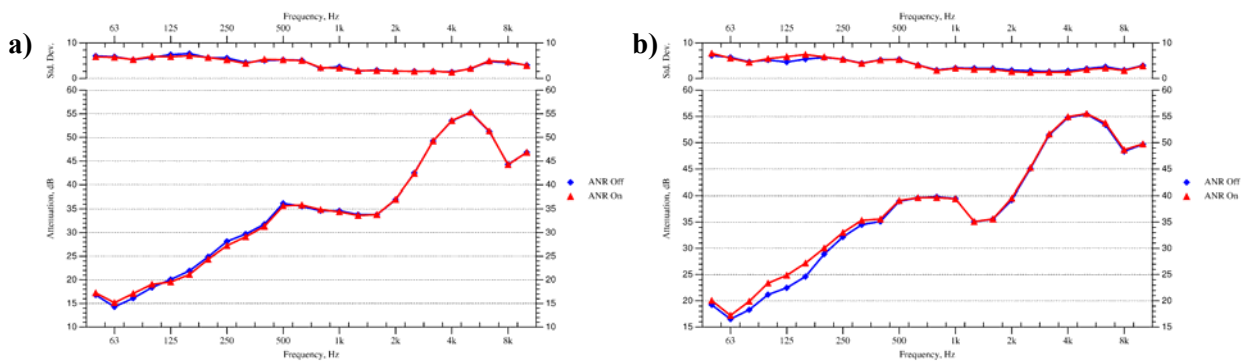


Figure 3: Mean and std. dev. of the attenuations (MIRE data) for ANR headset (on/off) worn in combination with a) Passive earplug and b) ANR earplug (off)

Earplug attenuation has been found to have a direct impact on the overall attenuation of double protection systems. The ANR of the earplug (see Figure 4a) evidently has an effect on the sound pressure level in the ear canal. Below 2kHz, when the ear canal is occluded at its entrance (i.e. standard fitted earplugs), the dominant BC pathway is via the SPL in the ear canal. Therefore at these frequencies for standard fit earplugs, the MIRE measured attenuation will give an accurate prediction of the attenuation seen by the user. Hence, the active attenuation of ANR earplugs can give a significant increase of attenuation when used in a double protection system. Figure 4b compares the attenuation data between the passive earplug and ANR earplug turned off when both were worn in conjunction with the passive headset. The results are similar to Figure 4a) in that the change in earplug attenuation changes the overall double protection attenuation.

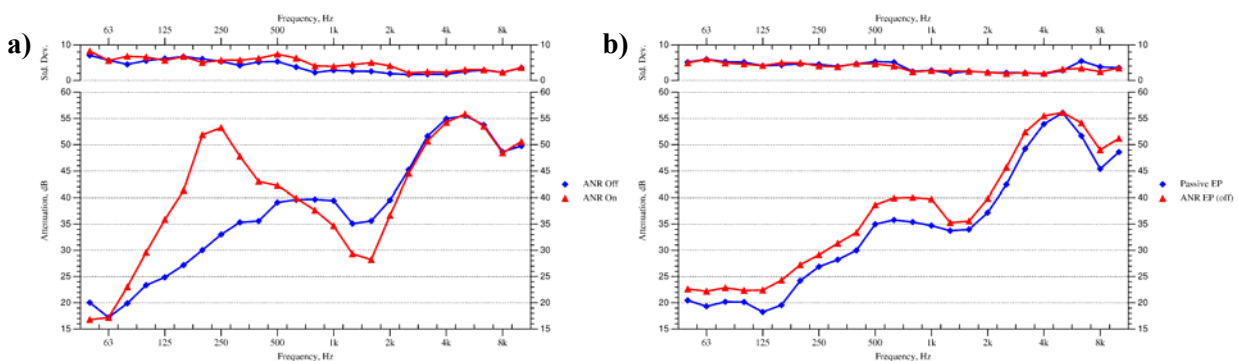


Figure 4: Mean and std. dev. of the attenuations (MIRE data) for a) ANR earplug (on/off) worn in combination with the ANR headset (on) and b) Passive earplug or ANR earplug (off) worn in combination with the passive headset

Figure 5a) shows the effect of changing the earplug attenuation when worn under a headset when using the REAT technique. At frequencies above 2kHz there is strong correspondence between the two curves, and as we are considering REAT data it is likely that this is due to bone conduction limits, noting also that this is not the case for the MIRE data in Figure 4b). However, below 2kHz there is marked effect on the attenuation and here the attenuation falls short of the bone conduction limits. This indicates that for double protection it is the attenuation of the earplug that is most important in defining the overall attenuation.

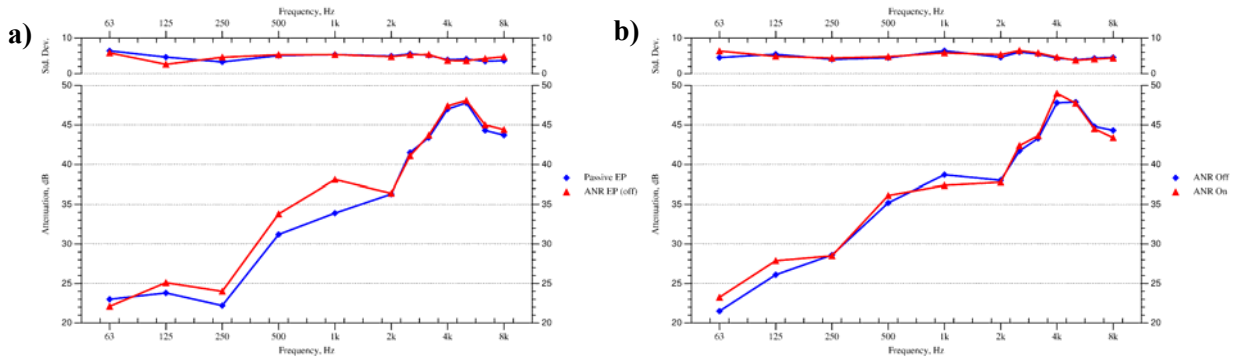


Figure 5: Mean and std. dev. of the attenuations (REAT data) for a) Passive earplug or ANR earplug (off) worn in combination with the passive headset and b) ANR headset (on/off) worn in combination with ANR earplug (off)

3.3 Earshell Attenuation

During the MIRE experimentation, the sound pressure level was measured under the earshells, when the headset was worn over the earplugs as well as when worn singly (see figure 2a). This was to facilitate further investigation into why the addition of double protection is not purely additive. Figure 6a) shows the mean attenuation provided by the passive headset when worn alone and with all combinations of earplug. Figure 6b) shows the mean attenuation provided by the ANR headset (on) when worn alone and with all combinations of earplug.

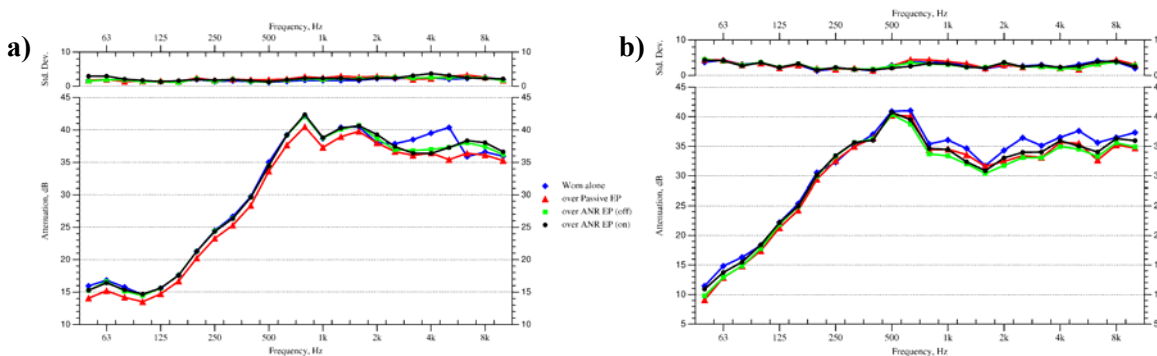


Figure 6: Individual attenuation of hearing protectors when worn as double protection, for a) passive headset mean attenuation over all combinations of earplug and b) ANR headset (on) mean attenuation over all combinations of earplug

The attenuation under both earshells is consistent apart from some small discrepancies, which could conceivably be put down to earshell fit, the position of the microphone within the ear shell and the reduced volume due to the earplugs being present. These graphs indicate that not only is the passive attenuation consistent but the active attenuation of the ANR headset is also working consistently when worn in combination with the earplugs. Exactly the same results were found for the ANR headset with the ANR

turned off. This means there is no degradation of earshell attenuation, implying that there is mechanical coupling between the earplug and earshell that reduces the overall attenuation.

3.4 Bone Conduction Limits

Figure 7 compares the MIRE and REAT measured mean attenuation. Figure 7a) shows the attenuation for ANR earplug turned off worn in conjunction with the ANR headset turned on and Figure 7b), the passive earplug worn in conjunction with the passive headset. The dominant bone conduction pathway above 2kHz is via the middle ear and inner ear, hence the sound pressure level in the ear canal does not factor in predicting the attenuation once the BC limit has been reached. As discussed in the companion paper [14], this implies that MIRE will predict higher attenuations, when at BC limits, than occur in reality above 2kHz. REAT on the other hand includes the inner and middle ear BC components and so measures the BC limited attenuation. In Figure 7a) and b) the MIRE data over predicts the attenuation above 3.15kHz, and implies that at these frequencies the BC limit has indeed been reached. Also, as mentioned, the MIRE technique used for calculating earplug attenuation (i.e. freefield – unoccluded) will tend to under estimate the attenuation, particularly in the 2-3 kHz bands. Implying that the MIRE measured attenuation would be higher in the 2kHz and 2.5kHz bands, indicating bone conduction being reached at these frequencies also.

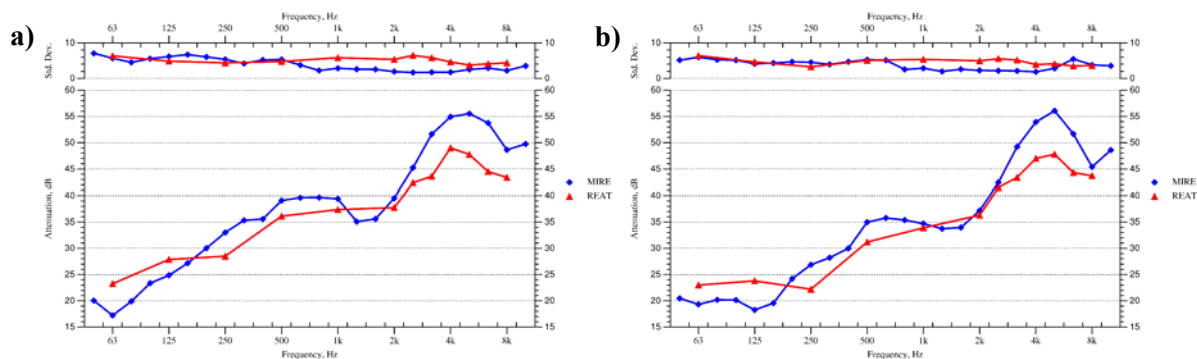


Figure 7: Comparing MIRE and REAT for double protection systems, for a) ANR earplug (off) worn in combination with the ANR headset (on) and b) Passive earplug worn in combination with the passive headset

Figure 8 shows the BC limits that can be inferred from this study. The curves in Figure 8a) show the attenuation of the ANR Earplug turned off when worn under the ANR headset turned on. This is considered the “passive” BC limit as there is no direct ANR effect on the attenuation. The REAT curve is simply the REAT attained attenuation. The second (“hybrid”) curve is a combination, as suggested in [14], of the REAT measured data above 2kHz and the MIRE measured data below 2kHz. As with Figure 2 the Zwislocki and Berger 1983 bone conduction limits are included for comparison. It is clear that there is a shortfall of attenuation at the low frequencies. However, in order to attain the levels of attenuation at low frequencies both Zwislocki and Berger used deeply inserted earplugs (into the bony part of the ear canal), which circumvents the BC pathway of the soft tissue of the ear canals oscillating. It is unsurprising that the earplugs in the current study, with a poorer passive performance and not deeply inserted, don’t reach the Berger and Zwislocki BC limits.

At the frequencies above 2kHz the attenuation is comparable to the previous BC data either close to the Berger 1983 curve or Zwislocki’s curve, where at 3.15kHz there is 5dB drop from the Berger 1983 curve. This implies that, further to the above discussion the work has reached bone conduction limits above 2kHz.

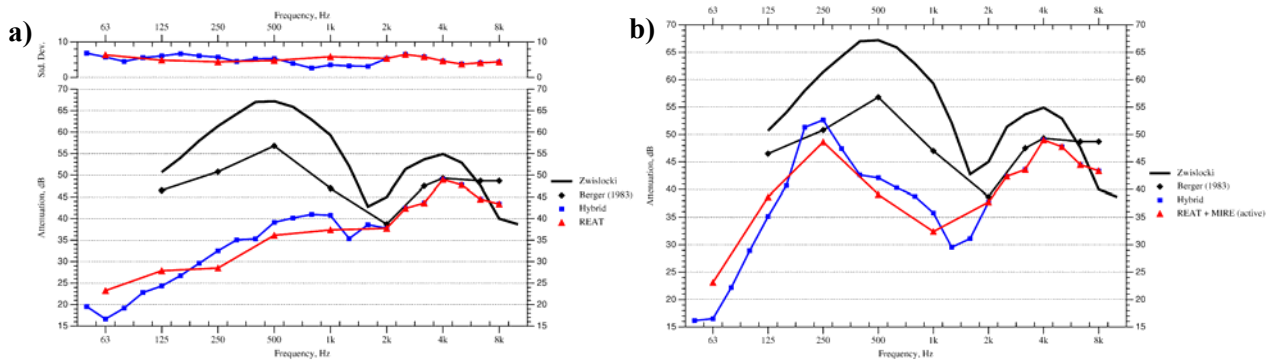


Figure 8: Predicted bone conduction limits for the current study, for a) “passive” BC and b) “active” BC

Figure 8b) shows the “active” BC limit that can be inferred from the current work. The “hybrid” curve is constructed from the mean REAT attenuation of the ANR Earplug (off) under the ANR Headset (on) above 2kHz and, below 2kHz, from the mean MIRE attenuation of the ANR Earplug (on) under the ANR Headset (on). The REAT plus MIRE (active) curve, is the same as the “hybrid” curve above 2kHz and, below 2kHz, it is the REAT attenuation with ANR Earplug (off) under the ANR Headset (on) plus the MIRE measured active attenuation only for the ANR earplug (on). The justification of not including the ANR earplug effect over 2kHz is that as the ANR only effects the SPL in the ear canal and as this is not part of the dominant BC pathway it will give a negligible effect on the overall attenuation. However the importance of this curve is the attenuation given at the lower frequencies. This shows that ANR fitted in an earplug allows a shallow fit of earplug (in a double protection system) to reach the bone conduction limits achieved by deeply inserted plugs (used in a double protection system), even if only over a limited range of frequencies. If the ANR earplugs were designed to provide better passive attenuation, the active attenuation seen here could provide similar levels of low frequency attenuation as deeply inserted earplugs.

5.0 CONCLUSION

The work has shown that the active attenuation of an ANR headset appears to offer little advantage when used over earplugs in a double protection system. However ANR earplugs have been shown to increase the low frequency attenuation when worn in combination with either active or passive headsets. Even with the poor passive performance of the moulded earplugs used in this study the active attenuation of ANR earplugs used in a double protection system has reached BC limits attained by deeply inserted plugs. This indicates that standard fit ANR earplugs are a viable alternative to deep inserted earplugs, as they are more practical and easier to fit.

The work has also shown that although the attenuation afforded by the double protection is not a simple additive process, the headset suffers no degradation of performance when worn over earplugs. This implies that there is likely to be some mechanical coupling between the headset and earplug that reduces the overall attenuation attained.

REFERENCES

- [1] A. Behar. Sound attenuation from combinations of earplugs and earmuffs. *Applied Acoustics*, 32(2):149–158, 1991.
- [2] S. M. Abel, N. M. Armstrong. The combined sound attenuation of earplugs and earmuffs. *Applied Acoustics*, 36:19-30, 1992.
- [3] J. Zwislocki. In search of the bone-conduction threshold in a free sound field. *Journal of the Acoustical Society of America*, 29(7):795–804, July 1957.
- [4] E. H. Berger. Laboratory attenuation of earmuffs and earplugs both singly and in combination. *American Industrial Hygiene Association Journal*, 44(5):321–329, 1983.
- [5] C. W. Nixon and H. E. Gierke. Experiments on the bone-conduction threshold in a free sound field. *Journal of the Acoustical Society of America*, 31(8):1121–1125, August 1959.
- [6] A. Behar and H. Kunov. Insertion loss from using double protection. *Applied Acoustics*, 57(4):375–385, August 1999.
- [7] E. H. Berger, R. W. Kieper, and D. Gauger. Hearing protection: Surpassing the limits to attenuation imposed by bone-conduction pathways. *Journal of the Acoustical Society of America*, 114(4):1955–1967, 2003.
- [8] J. Tonndorf. Bone Conduction in *Foundations of Modern Auditory Theory Vol II*. Academic Press London, 1972.
- [9] S. M. Kanna, J. Tonndorf, and J. E. Queller. Mechanical parameters of hearing by bone conduction. *Journal of the Acoustical Society of America*, 60(1):139, July 1976.
- [10] S. Stenfelt, T. Wild, N. Hato, and R. L. Goode. Factors contributing to bone conduction: the outer ear. *Journal of the Acoustical Society of America*, 113(2):902–913, February 2003.
- [11] E. A. Goodfellow. A prototype active noise reduction in-ear hearing protector. *Applied Acoustics*, 42: 299-312, 1994.
- [12] ANSI S12.42-1995: Microphone-in-Real-Ear and Acoustic Test Fixture methods for the measurement of insertion loss of circumaural hearing protection devices. *American National Standard*, 1995.
- [13] ANSI S12.6-1997: Methods for the measuring the Real-Ear Attenuation of hearing protectors. *American National Standard*, 1997.
- [14] Susan E. Mercy, Christopher Tubb and Soo H. James. Experimentation to address appropriate test techniques for measuring the attenuation provided by double ANR Hearing Protectors, *In Proceedings of the HFM Symposium on New Directions for Improving Audio Effectiveness*, April 2005.
- [15] S. Stenfelt, N. Hato, and R. L. Goode. Factors contributing to bone conduction: the middle ear. *Journal of the Acoustical Society of America*, 111(2):947–959, February 2002.

A Microphone-Array-Based System for Restoring Sound Localization with Occluded Ears

Adelbert W. Bronkhorst
TNO Human Factors
P.O. Box 23, 3769 ZG Soesterberg
The Netherlands
adelbert.bronkhorst@tno.nl

Jan A. Verhave
TNO Human Factors
P.O. Box 23, 3769 ZG Soesterberg
The Netherlands
jan.verhave@tno.nl

ABSTRACT

Current helmets and hearing protectors interfere with the sound transmission to the ears and therefore affect the perception and localization of speech and other useful sounds. This can be a serious drawback especially when the person wearing the protection has to operate in complex, unpredictable environments. A novel electro-acoustic system for sound pass-through was developed that can make hearing protection acoustically 'transparent'. By using external microphone arrays tuned to have a directional sensitivity similar to that of the open ears, the system can not only improve audibility of low-level sounds but also restore normal sound localization. The tuning was done by selecting specific microphone positions and by designing digital filters through which the individual microphone signals are passed. The system was evaluated in a sound localization experiment. Two versions were tested: one with individualized digital filters and one with universal (generic) filters. A comparison was made with a system with single external microphones, and with an earmuff with no sound pass-through. An open-ear condition was included as reference. Results show that, across all occluded-ear conditions, localization performance is best for the microphone-array system with individualized digital filters. Compared to listening through passive earmuffs, the percentage of confusions (quadrant errors) is nearly halved. However, localization performance is still not as good as with open ears.

1 INTRODUCTION

An important disadvantage of almost any type of hearing protection is that it suppresses all sounds, not only the noise, so that it interferes with the perception of speech and other useful sounds. Especially in critical situations (e.g. a soldier fighting in an urban environment, with a threat of snipers) this can be an important drawback. Several hearing protectors that are currently on the market address this problem, by including a controlled acoustic leak (pass-through) that restores the audibility of low-level sounds, while blocking the harmful high-level sounds. The pass-through can be realized either with a non-linear acoustic filter, or with an electro-acoustic system, consisting of a microphone, amplifier, limiter and telephone. However, these systems only solve part of the problem. Hearing protectors do not only decrease the level of the incoming sound – resulting in a reduced audibility – but they also distort the sound field, which has a severe effect on the capability to localize sounds (Smootenburg & Geurtsen, 1991; Abel, 1996; Vause &

Bronkhorst, A.W.; Verhave, J.A. (2005) A Microphone-Array-Based System for Restoring Sound Localization with Occluded Ears. In *New Directions for Improving Audio Effectiveness* (pp. 20-1 – 20-8). Meeting Proceedings RTO-MP-HFM-123, Paper 20. Neuilly-sur-Seine, France: RTO. Available from: <http://www.rto.nato.int/abstracts.aps>.

Grantham, 1999; Bolia et al., 2000; Bolia & McKinley, 2000). The latter effect is not compensated for by existing systems with a pass-through.

In order to address this issue a system was developed that uses microphone arrays instead of single microphones (Bronkhorst, 2002). This system can restore both audibility and directional hearing because the microphone array not only picks up the external sounds – just as a single microphone does – but is also able to simulate the direction- and frequency-dependent acoustic properties of the open ear. The simulation is realized by optimizing the placement of the microphones and by passing the output signals of each microphone through a specially designed digital filter, before they are added.

The system developed by Bronkhorst (2002) has only 2 microphones per array and is therefore not capable of simulating directional dependencies outside the plane passing through the microphones. In the current study, an improved version was developed that has arrays with 3 microphones placed on the corners of an equilateral triangle. The system was evaluated in a sound localization experiment. Localization performance was measured for passive hearing protection and for hearing protection equipped with either single microphones or microphone arrays, using open-ear performance as reference. The system was also subjected to an acoustical validation.



Figure 1: Setup for measuring head-related transfer functions

2 SYSTEM DESCRIPTION

The fundamental idea underlying the design of the microphone array system is that directional hearing can be restored when the acoustical effects of the head, ears and body on incoming sounds are simulated. The main effects are the differences in level and arrival time between the two ears and the spectral features (i.e. the frequency-dependent amplification and attenuation of the sound) present at each ear (e.g. Kuhn, 1987). It is thought that the interaural differences mainly code the left-right dimension, whereas the spectral features code both the up-down and front-back dimensions. The acoustical effects can be represented mathematically by the complex-valued transfer functions from a sound source to both eardrums – the so-called head-related transfer functions (HRTFs; see e.g. Wightman and Kistler, 1989). The HRTFs used in the current study were measured with small Sennheiser KE 4-211-2 microphones placed in ear plugs completely sealing the ear canals. Subjects were seated in the center of a hoop on which a trolley with a loudspeaker was mounted (see Fig. 1). Movements of the hoop and loudspeaker were controlled by a PC.

Measurements were done for about 1000 source positions, which cover the sphere around the subject's head almost completely (except for positions with elevations more than 60° below the horizontal plane) with a resolution of $5\text{-}6^\circ$.

In order to simulate the HRTFs with microphone arrays it is first necessary to measure the transfer functions of the microphones themselves. This was also done in the facility shown in Fig. 1. The hearing protector with the microphone arrays was placed on a Head Acoustics HMS II artificial head which was positioned in the center of the hoop. The microphone transfer functions are quite different from HRTFs because they contain less spectral features. In order to make the microphone transfer functions similar to the HRTFs, the microphone signals are passed through digital filters, added, and then presented through telephones mounted inside the hearing protector, as indicated schematically in Fig. 2. The system used in this study consists of two arrays of 3 microphones placed on the outside of a Peltor H7A earmuff, custom-built pre-amplifiers, a PC with an RME Hammerfall audio device and two Sony MDR CD 999 telephones mounted inside the earmuffs. The audio device does the A/D and D/A conversion and has a telephone amplifier. The digital filtering is performed by a custom-made Delphi program running on the PC.

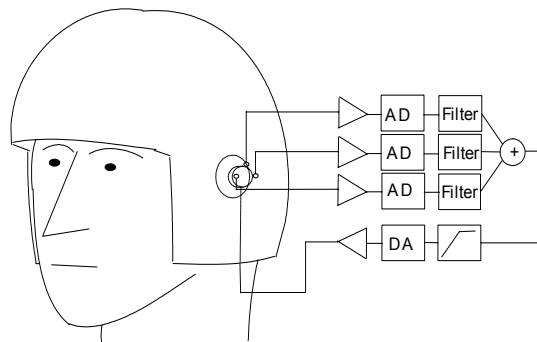


Figure 2: Schematic diagram of the microphone array system

The digital filters were calculated using software that minimizes the differences between the HRTFs and the transfer functions of the microphone array. Because digital filters are used, the filter characteristics only need to be determined at discrete signal frequencies. This means that six coefficients (two per filter) should be determined for each frequency. The calculation was done by iteratively minimizing a difference measure, which was the weighted sum of log-amplitude differences and group delay differences, averaged over angle of incidence. A set of 500 angles of incidence was used that covered the horizontal band with elevations between -22.5° and $+22.5^\circ$. In order to compensate for the characteristics of the telephones in the earmuffs, the calculated filter was multiplied by the inverse of the telephone transfer function. This was done separately for each ear.

3 EVALUATION

3.1 Design of the listening experiment

The effects of the presence or absence of hearing protection, and the type of pass-through were tested in the following five conditions:

1. open ears
2. earmuffs with the electronics switched off
3. earmuffs with a single microphone operating on each side

4. earmuffs with the microphone arrays and individualized filters
5. earmuffs with the microphone arrays and generic filters

Condition 1 was always presented first; the order of the other conditions was balanced across subjects. In these conditions, subjects wore the Peltor earmuffs equipped with the microphone arrays. The arrays were only switched on conditions 4 and 5; in condition 3, only the front microphone of the array was used and no digital filtering was performed. The individualized filters were calculated by fitting the transfer function of the array to the HRTFs of the listener, and by subsequently multiplying the filters by the inverse of the telephone transfer functions that had also been measured individually. The generic filters were based on HRTFs of one particular subject (not taking part in this study) which have proved to be suitable for many listeners. The inverse telephone transfer function was in this case the average of all individually measured transfer functions.

Table 1: Source positions used in the sound localization experiment

Azimuth Elevation	-180	-135	-120	-90	-45	-30	0	45	60	90	135	150
-45					X						X	
-30	X						X					
-20				X								
0	X		X			X	X		X	X		X
30	X			X			X					
45		X						X				
60	X						X			X		
90							X					

The experiment was conducted using the same setup as used for the HRTF measurements. Eight normal-hearing listeners were used as subjects. They were blindfolded, seated on an adjustable chair, and positioned so that their head was in the center of the hoop. Listeners indicated the perceived sound direction using a pointer (a small stick that could be rotated and tilted). In order to prevent confounding of perceived direction and distance, they were instructed not to set the pointer in the direction of the source, but to align it to the imaginary line between their head and the perceived source position.

Twenty-one loudspeaker positions were used (see Table 1); 9 positions were located in the median plane (the symmetry plane of the head), 6 in the right hemisphere, and 6 in the left hemisphere. Stimuli were 500-ms burst of pink noise with level of 75 dBA, measured at the position of the center of the listener's head. When the sound pass-through was switched off (condition 2), the level was increased by 10 dB. In each condition, 42 stimuli were presented (two per position); the order of the positions was randomized.

3.2 Results of the experiment

From the raw data, three different measures were determined: horizontal error, vertical error and percentage of confusions. All measures were derived from the directions of vectors pointing to the target and response (denoted as target and response vector, respectively). In order to calculate the horizontal error, the target vector was first rotated so that it had the same elevation as the response vector. The error was then taken as the angle between the response vector and the rotated target vector. The vertical error was taken simply as the difference in elevation between target and response vectors. A response was counted as a confusion when the angle between target and response vector was reduced by at least 30° after the response vector was mirrored in either the horizontal or the vertical plane passing through both ears (this corresponds to up/down and front/back confusions, respectively). Because confusions were resolved before the horizontal and vertical errors were calculated, there is a trade-off between confusions

and directional errors: any response that is not counted as confusion will cause an increase of one of the directional errors. This means that the choice of criterion for counting confusions is, to a degree, arbitrary, because it does not affect the sensitivity of the overall analysis.

Mean results and standard errors for the three measures are listed in Table 2. It can be seen that performance with open ears is clearly better than that with occluded ears, irrespective of the presence and type of sound pass-through. In order to verify this conclusion, we conducted, for each of the three measures, a single-factor repeated-measures ANOVA. A highly significant effect of condition was found in all cases, and Tukey HSD post-hoc analysis showed that the results of the open-ear condition were indeed significantly better than those for the other conditions ($p < 0.0001$). In the further analysis, we focused on the conditions with hearing protection, and we performed similar ANOVA's, excluding the data for the open-ear condition. In this analysis, we only found significant effects of condition on vertical error ($p = 0.025$) and percentage of confusions ($p = 0.008$) but not on horizontal error ($p = 0.24$). Tukey HSD post-hoc analysis revealed that the effects were due to the better performance when using microphone arrays with individualized filters (condition 4). There was a trend that the use of arrays with non-individualized filters (condition 5) also improved performance, but this effect did not reach the 5% significance limit.

Table 2: Results of the sound localization experiment. For each condition, the horizontal error, the vertical error, and the percentage of confusions are listed. Both mean values and standard errors are given.

Condition	Horizontal error (°)	S.e. (°)	Vertical error (°)	S.e. (°)	% confusions	S.e.
1 (open ears)	9.3	0.9	15.2	1.2	6.3	1.3
2 (passive muff)	18.6	1.5	26.7	0.6	35.1	4.2
3 (1 microphone)	16.1	1.4	26.6	0.8	32.1	2.5
4 (individualized)	15.8	0.9	23.3	1.4	20.2	2.4
5 (generic)	17.6	1.7	24.0	0.7	26.8	3.2

3.3 Acoustical measurements

Acoustical measurements were carried out to check the similarity between the transfer functions of the microphone array system and different sets of HRTFs. The earmuff with the microphone arrays was placed on the Head Acoustics artificial head, which was positioned in the center of the loudspeaker hoop. The microphone signals were amplified and passed through the digital filters calculated for one of the subjects; these were, in this case, not multiplied by the inverse telephone transfer function. Transfer function measurements were performed for the same set of directions as used in the earlier measurements: about 500 angles of incidence, with elevations between -22.5° and $+22.5^\circ$.

In the analysis, the results of the measurements were compared with the HRTFs on which the filters were based. Separate comparisons were carried out for amplitude and phase. The amplitudes were converted to dB and differences were averaged over frequency and angle of incidence. Only the first 124 bins were included in the averaging (corresponding approximately to the frequency range 100 Hz – 12 kHz). The phase of each measurement or HRTF was unwrapped, converted to time and then averaged, resulting in an average group delay. Because high-frequency delays are less relevant and possibly distorted by inaccurate unwrapping, this calculation was only based on bins 1-30 (roughly 100 Hz – 3 kHz). The comparison was repeated for all sets of HRTFs (also the 'generic' set), in order to see how effective the adaptation of the filters to specific HRTFs is. It appears that the match between the transfer functions of the microphone array system and the HRTFs is, indeed, optimal for the HRTFs on which the filter calculation was based. The average difference in amplitude is in that case 4.1 dB, whereas values between 4.8 and 5.6 dB are obtained for the other HRTF sets. Although the group delay differences depend less on HRTF set than the

amplitude differences, the difference was also smallest (0.06 ms) for the HRTF set on which the filters were based. Differences for the other sets ranged from 0.07 to 0.11 ms.

4. CONCLUSION

This study describes the development and evaluation of a microphone-array-based system that can be used to improve sound perception and localization of persons wearing hearing protection. The system uses arrays of 3 microphones placed on the shells of standard earmuffs, applies digital filtering to the microphone signals and presents the resultant signals through telephones mounted inside the shells. A listening experiment demonstrates that the system can significantly improve sound localization, compared to performance with passive earmuffs. The most notable effect is a reduction of the number of quadrant errors (confusions) from 35% to 20%. However, localization performance is still worse than that achieved with open ears (6% confusions). This means that we must look at how the system can be further improved, for example by increasing the number of microphones and by optimizing the placement of the microphones.

Interestingly, the subjective differences between listening with open ears and through the microphone array system are smaller than the localization experiment suggests. Other factors play a role as well, in particular the reproduction of background noise, which sounds much more natural through the microphone array than through single microphones or passive earmuffs. This can be explained by the fact that single microphones are not directional while the microphone arrays (like open ears) are most sensitive to sounds coming from frontal-lateral directions. The difference in global directional sensitivity will particularly have effects on sound perception in noisy environments and should be tested in separate (e.g. speech intelligibility) experiments.

5. ACKNOWLEDGEMENT

This research was supported by the European Research Office of the US Army, grant N62558-02-M-6380

6. REFERENCES

- [1] Abel, S.M. (1996). Combined effects of hearing loss and hearing protection on sound localization: Implications for worker safety. In: Scientific Basis of Noise-Induced Hearing Loss. Axelsson, A., Borchgrevink, H., Hamernik, R.P., Hellstrom, P.-A., Henderson, D. & Salvi, R.J. (Eds.). Thieme, New York. Pp. 286-295.
- [2] Bolia, R.S., & McKinley, R.L. (2000). The effects of hearing protectors on auditory localization: evidence from audio-visual target acquisition. *International Journal of Occupational Safety and Ergonomics* 6, 309-319.
- [3] Bolia, R.S., D'Angelo, W.R., Mishler, P.J. & Morris, L.J. (2000). Effects of hearing protectors on auditory localization in azimuth and elevation. *Human Factors* 43, 122-128.
- [4] Bronkhorst, A.W. (2002) Richtinghoren met hoofdbescherming.. Report TM - 02 - A014, TNO Human Factors, Soesterberg, The Netherlands (in Dutch).
- [5] Kuhn, G.F. (1986). Physical acoustics and measurements pertaining to directional hearing. In "Directional Hearing" (Eds.: William A. Yost & George Gourevich), Springer-Verlag, New York, pp. 3-25.

- [6] Langendijk, E.H.A. & Bronkhorst, A.W. (2002) Contribution of spectral cues to human sound localization. *Journal of the Acoustical Society of America* 112, 1583-1596
- [7] Smoorenburg, G.F. & Geurtsen F.W.M. (1991) De invloed van het dragen van gehoorbeschermers en van gehoorverliezen ten gevolge van lawaai op het richtinghoren. Rapport 1991 C17. TNO Technische Menskunde, Soesterberg.
- [8] Vause, N.L. & Grantham, D.W. (1999). Effects of earplugs and protective headgear on auditory localization ability in the horizontal plane. *Human Factors* 41, 282-294.
- [9] Wightman, F.L., & Kistler, D.J. (1989). Headphone simulation of free-field listening: I. Stimulus synthesis. *Journal of the Acoustical Society of America* 85, 858-867.



3-D Audio: Military Applications and Symbology

Mark A. Ericson, Brian D. Simpson, and Richard L. McKinley

Air Force Research Laboratory (AFRL/HECB)

2610 Seventh Street, Bldg 441

Wright-Patterson, AFB, OH 45433-7901

mark.ericson@wpafb.af.mil

3-D AUDIO SYMBOLOGY

Five experiments were conducted to study the acoustic attributes that enable the accurate identification and localization of rudimentary spatial warning sounds. In each experiment, two sounds were played simultaneously over loudspeakers at various azimuths and elevations. The stimuli consisted of pure tone complexes with a 13 kHz bandwidth. The fundamental frequencies, the amplitude modulation rate of the complex, the harmonicity of the carrier and modulation frequencies, and the coherence of the carrier and modulation phase were varied in the experiments. The combination of all the cues provided the best localization and identification performance. When and only when all cues were used, the subjects were able to accurately localize and identify the target sound.

1.0 INTRODUCTION

When warning sounds are played simultaneously over a single loudspeaker, they tend to combine into a single, jumbled warning clamour. In addition, the cacophony often tends to distract and annoy rather than inform an operator [1]. Alarm systems tend to be turned off by the operator, or never turned on, to avoid distractions under high-stress, high-workload conditions. The situation will potentially become more intractable if the current warning sounds, designed for single channel or monaural displays, are presented via spatial auditory displays.

Coding techniques have typically been employed to facilitate the identification of multiple warning sounds during monaural presentation over a headphone or over a single loudspeaker. Repetition rate has been found to be the most salient feature of warning sounds for purposes of identification [2] and for encoding urgency [3]. When the repetition rates of two sounds are similar, frequency coding and modulation can be used to segregate multiple warning sounds. However, such encoding techniques require the user to remember the association among a repetition rate, a frequency, a modulation and the airborne event that triggered the warning. Most pilots do not recall these associations accurately when asked to identify pre-recorded warning tones [4], [2], [5]. Presumably, identification performance would decrease further under high-workload conditions [6].

Naish [7] was one of the first to report the potential benefits of combining aircraft warning sounds with auditory lateralization cues. Using interaural phase and level differences, reaction times to left and right signals were found to be significantly faster when the perceived locations of the warning sounds correlated with the verbal direction of the sounds. Over the past sixteen years, many improvements have been made in 3-D auditory display technology. The accurate measurement and implementation of head-related transfer

Ericson, M.A.; Simpson, B.D.; McKinley, R.L. (2005) 3-D Audio: Military Applications and Symbology. In *New Directions for Improving Audio Effectiveness* (pp. 22-1 – 22-10). Meeting Proceedings RTO-MP-HFM-123, Paper 22. Neuilly-sur-Seine, France: RTO.
Available from: <http://www.rto.nato.int/abstracts.aps>.

3-D Audio: Military Applications and Symbology

functions provides many opportunities for integrating warning sounds into spatial aircraft display systems. Of particular importance to the military is the rapid and accurate display of threat warning information to the pilot.

Radar warning receivers typically display missile threats at one of four priority levels: 1) search, 2) scan, 3) alert, and 4) launch. Warnings are commonly coded by one or two tones. Search tones are usually a single pure tone with a relatively low frequency. Scan tones often sweep up and/or down in frequency. Alert tones, which signify a radar lock, alternate in pairs. Missile launch tones are represented by a single, high-frequency pure tone. Although radar warning receiver sounds are different for all aircraft, they can be roughly categorized by the preceding descriptions [8].

Experience with the integrated helmet auditory visual system (IHAVS) symbology demonstrated that pilots could effectively identify threat and priority levels for only one monaurally presented threat [5]. Two to four simultaneous radar warning receiver (RWR) sounds were not reliably identified. Spatial separation and the addition of uncorrelated noises enabled pilots to localize and identify up to four simultaneous warning sounds.

These spatial warning sounds were used in flight tests at Edwards Air Force Base to measure in-flight performance [9]. Although pilots could determine the location of four sounds in the laboratory, they reported being able to utilize only two sounds in flight because of the high workload. Daniels, Ericson, and French [10] described other airborne applications of combined 3-D audio visual displays.

Other techniques have been explored for extending the frequency range of current warning sounds for application in spatial auditory displays. Martin, Parker, McNally, and Oldfield [11] measured localization performance with rotary- and fixed-wing warning sounds. The sounds with the greatest number of pulses were more easily localized than the wide bandwidth sounds with fewer pulses. Therefore, the temporal structure of sounds is generally more important than signal bandwidth for high localization accuracy.

Patterson and Datta [12] extended existing warning sounds for use in spatial auditory displays. Three aspects of sounds were extracted and applied to wider bandwidth representations of similar sounds. These techniques included envelope filtering, Nyquist whistling, and fine structure doubling. Extending the frequency range of current warnings to 12 kHz was found to significantly improve localization accuracy.

In the current study, the amplitude modulation envelope was varied to impose a type of repetition rate onto the signals. The values chosen represent four regions: steady-state (0 Hz), slow temporal envelope (3 Hz), trill (10 Hz), and roughness (100 Hz). There are four parameters to vary in the amplitude modulation formula. These are the carrier frequency and phase and the modulation frequency and phase. This has the equivalent effect of adding noise to the spatial frequency, affecting the low-frequency resolution more than the high-frequency resolution. Unlike adding random noise, this technique maintains optimum signal strength for use in high noise environments.

2.0 METHODS

2.1 Equipment

The auditory localization facility at Wright-Patterson Air Force Base was used to conduct the experiments. The auditory localization facility was used for the presentation of free-field stimuli for multiple sound source localization and identification. The large set of loudspeaker locations in this facility reduced the possibility of location choice biases due to set size.

The God's-Eye Localization Procedure (GELP) [13] was used for collection of localization responses. In this technique, the subject responded after each stimulus presentation by positioning the tip of an electro-magnetic stylus at a point on the surface of a 20-cm plastic spherical model of auditory space to indicate the perceived direction of the auditory image. The subject's chin was restrained by a chin rest to reduce head motion cues.

The stimuli consisted of pure tone complexes with a 13 kHz bandwidth. The fundamental frequencies were in the range of 500 to 1000 Hz and are described for each experiment. The harmonicity of the carrier and modulation frequencies was either purely harmonic or randomized over a plus-or-minus five percent range. The coherence of the carrier and modulation phase was either completely in-phase or uniformly randomized over a two Pi range. All sounds were generated with Tucker-Davis-Technology System II equipment and a personal computer with Pentium II processor and a Windows 98 operating system.

2.2 Subjects

Ten naïve subjects were recruited from the general population. All had normal hearing and normal or corrected normal vision. The paid volunteers participated in all experiment conditions of the first and second experiments. Each volunteer subject had a normal hearing threshold levels and consented to participate in various listening experiments.

2.3 Procedures

The subjects were instructed to localize the center of the "auditory image" using the GELP technique. Two sequential presentations of the threat warning sounds were played before each pair of simultaneous sounds. Subjects were instructed to respond by indicating the locations in the order in which the sequential sounds were played immediately before the simultaneous sounds.

3.0 RESULTS

3.1 Steady State Pure Tone Complexes

In the first experiment, the two complexes differed in their fundamental frequencies by no difference, a musical fifth, a random non-musical interval, and a musical octave. The components were either regularly spaced (harmonic) or irregularly spaced (inharmonic). Data for this experiment are shown in Figure 1 below.

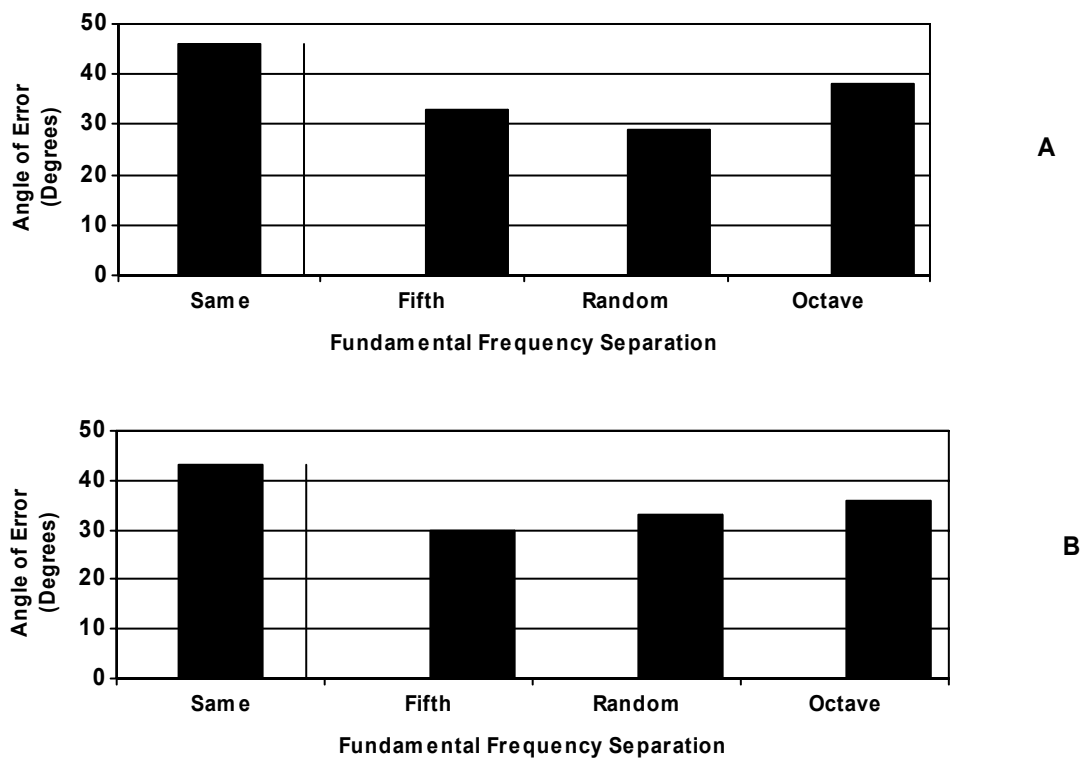


Figure 1. Angle of error versus fundamental frequency separation for (A) harmonic spacing and (B) inharmonic spacing of frequency components

3.2 Amplitude Modulation Rate

In the second experiment, the two complexes differed in their rate of amplitude modulation. The amplitude modulation rates included 0 (no modulation), 3, 17, and 100 Hz. Localization performance was relatively poor for these signals. Performance was worst when the target and masker were of the same amplitude modulation frequency. The 3 Hz stimulus was generally the worst target and the worst masker. The low modulation rate caused an apparent motion effect between the target and masker, which contributed to poor localization performance. Data for this experiment are shown in Figure 2 below.

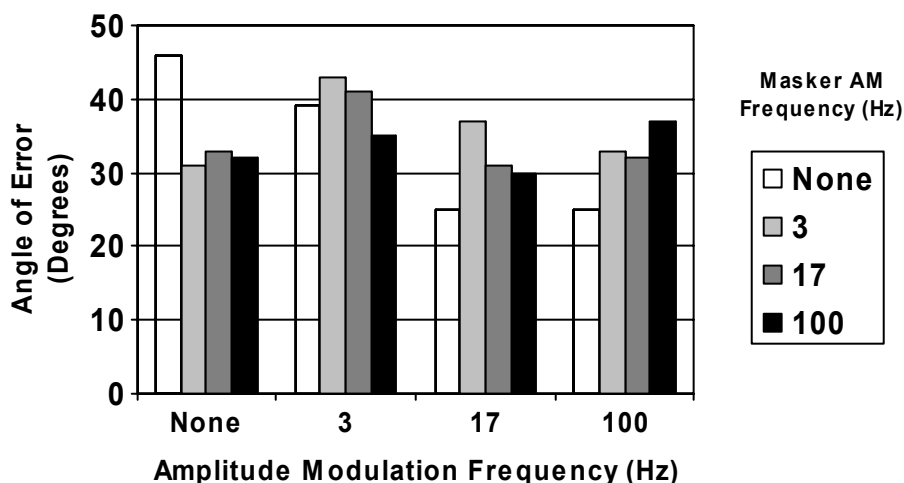


Figure 2. Angle of error versus amplitude modulation frequency of target sound

3.3 Amplitude Modulation Rate with Modulation Rate Randomized +/- 5%

In the third experiment, the two complexes differed in their rate of amplitude modulation as in experiment 2. However, the rate was randomized for each frequency component over a +/- 5% range. Localization performance was better than in experiment 2. The 3 Hz stimulus was still the worst target and the worst masker. The addition of the randomized modulation frequency greatly reduced the apparent motion effect between the target and masker. Data for this experiment are shown in Figure 3 below.

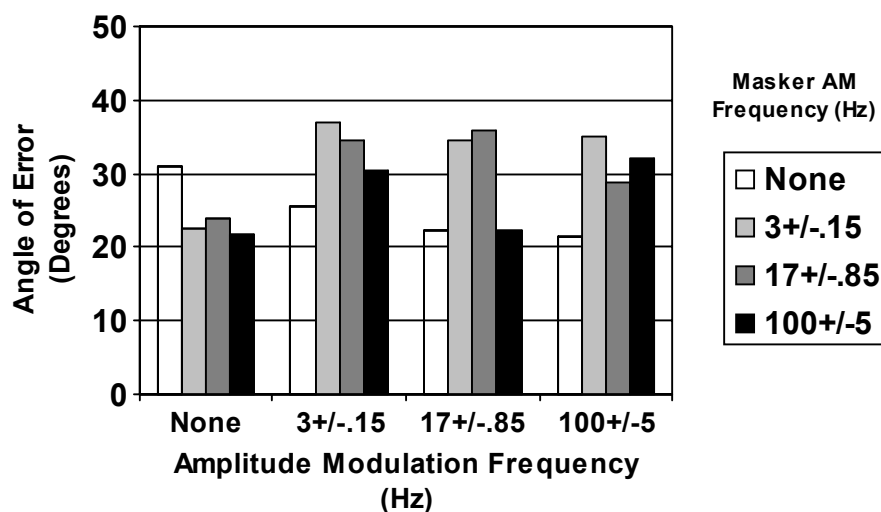


Figure 3. Angle of error versus amplitude modulation frequency of target sound

3.4 Amplitude Modulation Rate with Randomized Modulation Rate and Carrier Frequency

In the fourth experiment, the two complexes differed in their rate of amplitude modulation and had randomized modulation rates as in experiment 3. In addition, the carrier frequencies were also randomized over a plus-or-minus five percent range. Localization performance improved over those measured in experiments 2 and 3. The addition of the randomized carrier frequency to the randomized modulation frequency nearly negated the apparent motion effect between the target and masker. Data for this experiment are shown in Figure 4 below.

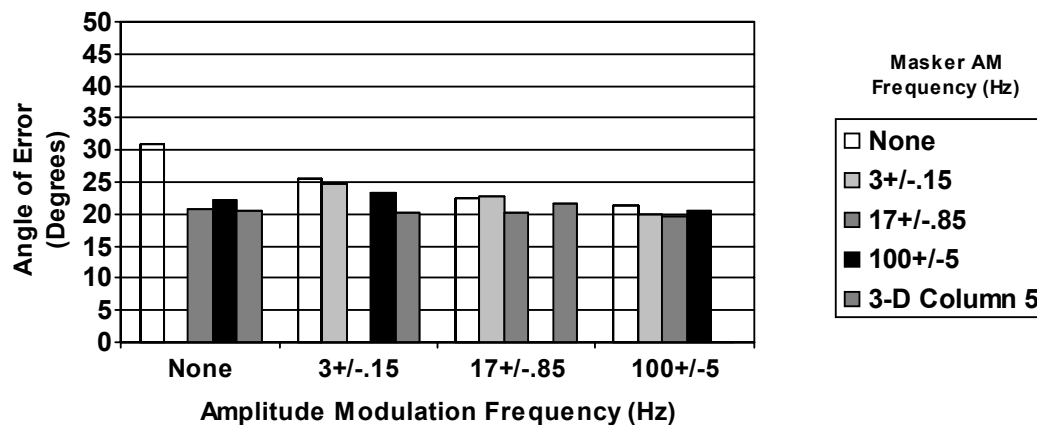


Figure 4. Angle of error versus amplitude modulation frequency of target sound

3.5 Amplitude Modulation Rate with Randomized Frequencies and Phases

In the fifth experiment, the two complexes differed in fundamental frequency (500 Hz and 571 Hz) and all four possible ways of creating amplitude modulated signals. The carrier frequencies and modulation frequencies were randomized over a plus-or-minus five percent range. The phases of the carrier and modulation frequencies were randomized over a 2 Pi range. Data were collected on the most difficult amplitude modulated signals to localize, the 3 and 17 Hz cases. Localization performance was close to localization acuity without a masker. Acuity averaged about fifteen degrees and correct identification performance about 86%. Some of the identification errors were due to loss of attention over the course of the twenty-minute sessions. Data for this experiment are shown in Figure 5 below.

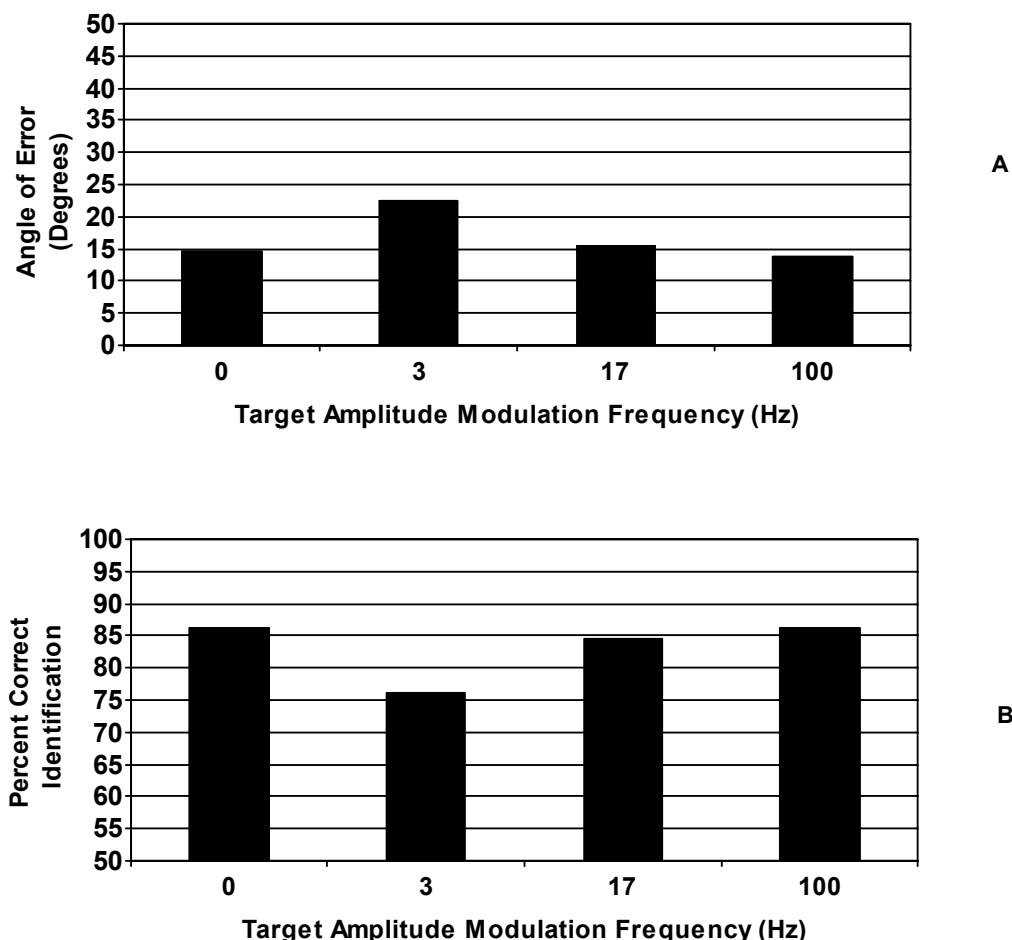


Figure 5. (A) Angle of error versus amplitude modulation frequency of target sound with a 3 Hz masking sound and (B) percent correct identification versus target sound with a 3 Hz masker

4.0 DISCUSSION

There are special challenges for designing warning sounds for military applications. One is that multiple threats of the same type may need to be presented at the same time. Another challenge is the harsh acoustic environment in which the sounds are typically presented, usually over headphones. One advantage of the current approach to warning sound design is that it maximizes the signal-to-noise-ratio without loss of localization acuity or identification performance. Perturbations to the component frequencies and relative phases do affect the noisiness of signals, which affects urgency but only a small amount. Repetition rate and loudness are the most salient cues for perceived urgency; harmonicity is third most important.

The introduction of spatial warning sounds can potentially improve military and non-military aviation safety. Spatial auditory displays for collision avoidance would reduce reaction times of pilots to take evasive manoeuvres and provide a more intuitive display of impending trouble. In military aircraft, integration of

3-D Audio: Military Applications and Symbology

spatial warning sounds with radar warning receivers should greatly improve their utility by allowing the pilot to hear the location of the threat and keep his head and eyes free to attend to other tasks. Pilots would improve their situation awareness by hearing the location of their team members' voices and threat locations. Onboard aircraft warning sounds could be displayed from the location of the failure. In command and control centers, the locations of friendly and enemy forces could be simultaneously monitored. The laboratory research on spatial warning sounds will continue to provide a framework for their integration into military auditory displays.

5.0 CONCLUSIONS

Several experiments were conducted on the ability of trained subjects to localize and identify pairs of sounds. All sounds were presented over loudspeakers in anechoic space. Presentation of multiple sounds over headphones in virtual auditory space can potentially degrade performance further if the virtual auditory cues are not properly implemented. The major experimental findings are summarized in the statements below.

- 1) Pure tone, in-phase, harmonic complexes were poorly localized. Auditory images tended to fuse into a single large source, especially when located close to each other and when their fundamental frequencies were equal or at octave multiples of each other.
- 2) Any single cue manipulated in the experiments, e.g. fundamental frequency, harmonicity of the carrier and modulation frequencies, phase of the carrier and modulation frequencies, and amplitude modulation rate, was relatively ineffective in improving localization acuity.
- 3) The combination of all the cues provided the best localization and identification performance.
- 4) When applying these findings to the design of threat warning sounds, consideration should be given to their effects on the perceived urgency of the warning sounds.

6.0 REFERENCES

- [1] Edworthy, J. and Haas, E. (1995). Toward a safer, more efficient auditory warning signal, *J. Acoust. Soc. Am. Newsletter*, 5(2).
- [2] Patterson, R. D. (1982). Guidelines for auditory warning systems on Civil Aircraft, *CAA Paper 82017*.
- [3] Edworthy, J. (1994). Urgency mappings in auditory warning signals, *Human Factors and Alarm Design*, Ed. Neville Stanton, 15-30.
- [4] Duross, S. H. (1978). Civil aircraft warning, systems – A survey of pilot opinion with British Airways. *Tech. Report 78056*, Royal Aircraft Establishment, Farnborough.
- [5] Gerth, J. M., Folds, D. J., Fain, W. B., McKinley, R. L., and D'Angelo, W. R. (1995). Development and evaluation of auditory symbology for representation of simultaneous threats, *NAECON 95*, Dayton, Ohio, 8.

- [6] Tun, P. and Wingfield, A. (1994). Speech recall under heavy load conditions: Age, predictability and limits on dual-task interference. *Aging, Neuroscience, and Cognition*. **1**(1) 29-44.
- [7] Naish, P.L.N. (1988). The addition of localisation cues to auditory warnings. *Technical Memorandum MM2*, Royal Aerospace Establishment.
- [8] Folds, D. J. (1990). Advanced audio displays in aerospace systems: Technology requirements and expected benefits, *NAECON*, Dayton, Ohio, 739-743.
- [9] McKinley, R. L. and Ericson, M. A. (1997). *Flight demonstration of a 3-D auditory display*, in Binaural and Spatial Hearing in Real and Virtual Environments, Gilkey, R. and Anderson, T., Eds. Lawrence Erlbaum Associates, Publishers, Mahwah, New Jersey.
- [10] Daniels, Reginald, Ericson, Mark A., and French, Guy A. (2002). Improved performance from integrated audio video displays, *Proceedings of SPIE* **4712**, 113-119.
- [11] Martin, R., Parker, S., McNally, K., and Oldfield, S. (1996). The abilities of listeners to localize defense research agency auditory warnings, *AGARD-CP-596*, 9-1 - 9-7.
- [12] Patterson, R. D. and Datta, A. J. (1996), Extending the frequency range of existing auditory warnings, *AGARD-CP-596*, 8-1 - 8-8.
- [13] Gilkey, R. H. Goode, M. D., Ericson, M. A., Brinkman, J., and Stewart, J. M. (1995). A pointing technique for rapidly collecting localization responses in auditory research, *Behavioural Research Methods, Instruments, and Computers* **27**, 1-11.



Relative Effectiveness of Audio Tools for Fighter Pilots in Simulated Operational Flights: A Human Factors Approach

Sylvain Hourlier[#], James Meehan^{*}, Alain Léger[°], Corinne Roumes

[#]Département de Sciences cognitives – IMASSA – BP 73 – 91223 Brétigny-sur-orge Cedex – France
Tel. 33 1 69 23 78 74 – Fax. 33 1 69 23 70 02

shourlier@imassa.fr

^{*} DSTO (Melbourne / Australia)

[°] Thales Avionics (Le Haillan / France)

ABSTRACT

The French-Australian Collaboration on Emerging Technologies (FACET) investigated appropriate means of delivering situation-awareness into the cockpit of fighter aircraft under simulated operational conditions. Increasing use of audio has been suggested as a means to reduce visual workload, to enhance situation awareness, and mitigate the manual and cognitive demands of HOTAS and existing command-and-display concepts. An open design for the pilot interface should incorporate redundant information in the auditory and other sensory modes, while enabling commands to be delivered through voice or manual control interfaces. However, sensory and cognitive resource competition may still limit delivery of the implicit benefits of such a design. The objective of FACET was to investigate this proposition.

Eight military pilots from France and Australia flew four full-mission scenarios in a simulated air combat environment. Auditory signals comprised radio messages, 3-D sounds and alarms, while the pilot's oral responses were verbal responses or direct voice inputs via automatic speech recognition. Some of the signals were drawn from those already in use, but all were selected for their potential to support situation awareness and support visual information, all delivered via a helmet-mounted display.

The four scenarios had specific, embedded combinations of events requiring activation of multiple resources. All events were videotaped. Both planned and unexpected events that were the outcome of human-interface interactions occurred. The pilots' behaviour (e.g., tool selection, prioritization, errors) as well as their efficiency in processing inputs and output control were analysed in order to describe their cognitive resource management. Resource competition imposed by the new technologies was analysed. Interviews were also performed after each run. Following the simulation, Repertory Grids were constructed to elicit each pilot's cognitive representation of the control-and-display concepts.

Pilot performance appeared to be independent of the time course of resource use during an event. Resource management was therefore not based on sequential allocation of resources, but rather on a cognitive prioritization of the task at hand. New audio technologies proved to be easy to use in conjunction with a broader use of the helmet-mounted display. Direct voice input required strict initial training to ensure reliability. Following such training, it was attractive to the pilots, even under high workload. The efficiency of 3-D sound was pilot dependent; one pilot in particular was able to fully use it immediately yielding a significant performance increase with a low cost in cognitive resources. In general, pilots' use of 3-D sound was dependent upon their level of training as well as their trust in, and willingness to experiment with the technology. These results may have implications for the training of pilots in the use of these new technologies.

Hourlier, S.; Meehan, J.; Léger, A.; Roumes, C. (2005) Relative Effectiveness of Audio Tools for Fighter Pilots in Simulated Operational Flights: A Human Factors Approach. In *New Directions for Improving Audio Effectiveness* (pp. 23-1 – 23-8). Meeting Proceedings RTO-MP-HFM-123, Paper 23. Neuilly-sur-Seine, France: RTO. Available from: <http://www.rto.nato.int/abstracts.aps>.

Besides the remarkable effectiveness of Franco Australian collaboration and workforce sharing that needed to be noted, this project highlighted that operational pilots had a relative easy appropriation of technology they had never used before in realistic simulated air combats.

Their strategy of use was workload dependant, yet still efficient at the highest workload levels even for DVI, revealing the true nature of cognitive resource management in flight. Pilots relied on an in/output management mastery where "tradeoff guided" resource allocation is the key to success. Interestingly, new and resource hungry interface systems were not especially discarded at highest levels of workload because pilots thought their use brought a fair tradeoff. Beyond the investigation of "tradeoff guided" resource management we should try to establish how to build systems that blends into this high level resource management strategy.

1.0 INTRODUCTION

In this paper we describe an evaluation in simulated flight of new control and display technologies for future combat aircraft. The technologies involved are helmet-mounted display (HMD) delivered visual and auditory information, and direct-voice commands. This encompasses visor-projected imagery (symbology and sensor images), advanced auditory displays (3-D sound, auditory icons, synthetic speech) and direct voice recognition by on-board systems. The essence of the problem is that these technologies are approaching a level of maturity where plans for their incorporation into new and existing aircraft are becoming firm, and we do not yet fully understand the likely impact on the pilot. The specific aim of the study reported here is to understand the behavioral and cognitive outcomes of addressing information to the different sensory modes and enabling inanimate on-board systems to respond to voice commands of the pilot. The general objective is the evolution of some principles for design and implementation of such systems in the next generation combat aircraft. This paper is focusing on the use of a set of audio tools and the pilot's cognitive resource management that ensued.

2.0 PURPOSE OF THE STUDY

As a scientific background we used Wickens Multiple Resource Theory [1]. It adds a qualitative aspect to resources that explains low performance on certain concurrent task because they draw on the same qualitative pool at the same time. For example, it is most difficult to speak while listening.

Resources are thus defined according to three dimensions; stage defined (early *versus* late processing), modality defined (auditory *versus* visual encoding) and processing-code defined (spatial *versus* verbal encoding). For the evaluation of a system, that uses both visual and auditory cues simultaneously with manual and verbal commands, such qualitative discrimination in resources will reveal incompatibilities in the system usage. The most expected one being issuing a DVI (direct voice input) command while listening to audio inputs. In such a frame of reference, the evaluation of the "alternate technology suite" workload depends on the number of concurrent uses, the respective type of resource they feed on and the level of ongoing demands due to flight and mission. The idea is that the more an individual feels his resources are invested, the highest his perceived workload will be. A bench model of resource competition was designed to represent the most probable resource conflicts to be expected given the HMI at hand (figure1).

Inspired by Flanagan's [2] critical incidents, critical events were embedded in four operational scenarios designed by the French Flight test centre (in Istres). Each critical event drew on specific allocation of resources (auditory, visual, verbal, etc.) predicted by the model to be of a problem for pilots, and were scheduled along increasing workload phases. Due to the complexity of the missions at hand, some events combinations were expected to happen during high workload phases but could not be predicted before hand, so an in-house event videotaping and labeling system (comparable to the "Experimental Video Annotator" [3] was designed accommodate the appearance of any "unscheduled" event. Each recorded

event is composed of a series of activities (basic interactions with the HMI, either the "standard" ones or the "alternate" ones) in their original order of appearance.

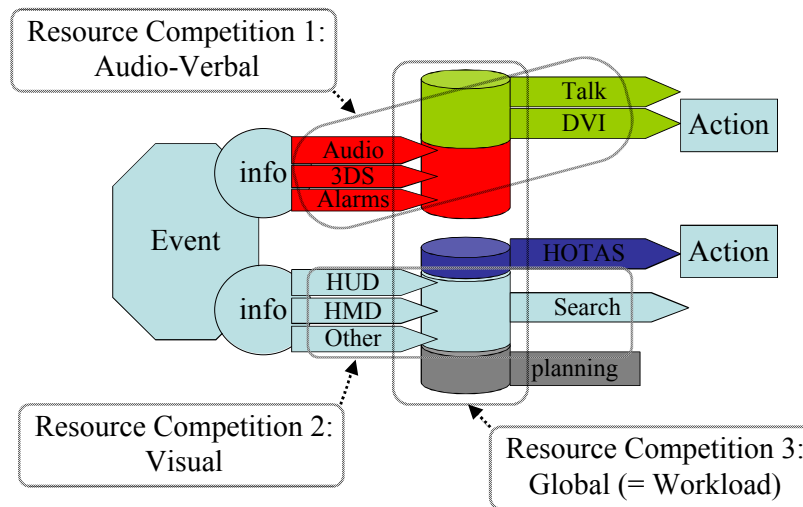


Figure 1: Resource competition model outlining audio-verbal resources competition, visual resources competition and global resource competition (i.e. Workload)

The pilots' behavior (e.g., tool selection, prioritization, errors) as well as their efficiency in processing inputs and output control was analyzed in order to describe their cognitive resource management. Resource competition imposed by the new technologies was analyzed. Interviews were also performed after each run. Following the simulation, Repertory Grids were constructed to elicit each pilot's cognitive representation of the control-and-display concepts.

3.0 METHODS

3.1 Simulation

The Rafale simulation at Istres flight test centre was used for this experiment (full mission sphere).

3.1.1 Alternate technology suite

Thales's alternate HMI suite consisted of a 3DS generating system integrated in the Topsight Rafale helmet (HMD) with a direct voice inputs (DVI) possibility. The standard Rafale system was always available, but pilots were encouraged to use the alternate system whenever possible. Auditory signals comprised radio messages, 3-D sounds and alarms, while the pilot's oral responses were verbal responses or direct voice inputs via automatic speech recognition. Some of the signals were drawn from those already in use, but all were selected for their potential to support situation awareness and support visual information, all delivered via a helmet-mounted display.

3.1.2 Scenarios

Four full-mission scenarios were designed in a simulated air combat environment.

3.2 Population

Seven military pilots from France and Australia (4 French, 3 Australians) participated in the study, though they had various background, all were flying combat aircraft at the time of the experiment.

3.3 Event analysis

Each event (expected and unexpected) was named after its activity composition; for example if an incoming communication (A) appeared when the pilot was issuing a DVI call (DVI) that happened on a preexistent cognitive load (G), the resulting event was labeled "G-DVI-A".

Events were analyzed with regards to their composition (number and order of activities), each event and activity being assessed in terms of pilot realization (good or bad performance for the event and its activities, as well as for the impact of an activity over a preexisting event).

3.4 Controlled debriefings

The "on the spot" debriefings enabled understanding of event performance as well as insights on acceptance and usability, both with regards to the systems (HMD, DVI, 3DS) used.

3.5 Collected Data

The data collected consist of:

- 28 debriefed runs,
- 57 types of events by combination of 6 types of activities: A (alarms, com-in), 3DS, DVI, HMD, Ve (com-out), G (cognitive workload), Vi (visual search).
- 271 recorded events (135 predictable and 136 unpredictable) over 5 levels of workload

4.0 SELECTED RESULTS AND DISCUSSION

4.1 Resource management strategy

One could think that the arrangement of activities over time that create an event can influence the performance of the included activities or on the performance of the subsequent event. To address this, the performance of every activity was tested with regards to its position in an event. Only 3DS, DVI and HMD had sufficient occurrences to be statistically analyzed with an ANOVA. In fact, the evaluation of event performance with regards to the position of activities shows no significant impact whether the activity position (last or first, or in the middle): DVI ($F^3_{79}=0.26$; $p=0.853$; NS), 3DS ($F^3_{186}=0.59$; $p=0.619$; NS) and HMD ($F^2_{128}=0.38$; $p=0.686$; NS). One could have thought that interface management information had to be "cumulative". Prioritization strategy dominates thus sequentiality of activities. The prioritization is governed by task and objective at hand and activities are evaluated in terms of "resource investment demand" versus "performance gain" and integrated to or discarded from the ongoing event, whatever their timing position. (i.e. an incoming communication can lead to the abandon of another ongoing activity if the pilot thinks the tradeoff is worth it). Pilot resource management seems to be based on a cognitive prioritization of the task at hand.

4.2 Effect of 3DS and DVI use on performance

To test the effect of DVI or 3DS use on the performance of an ongoing event, the relative performances of events associated with their respective use (failed or successful) were compared. It appears that when a DVI or 3DS command is issued, its performance does not influence the performance of the ongoing event.

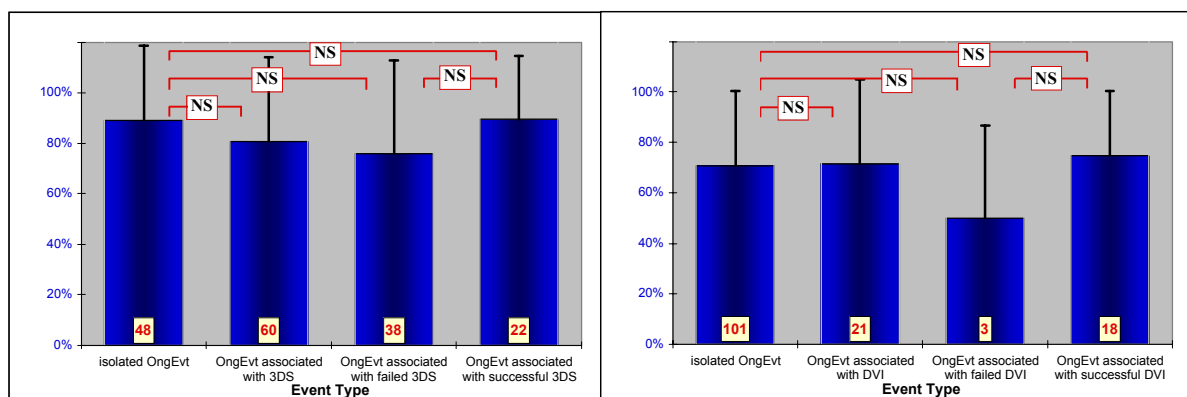


Figure 2: effect of a DVI (left) or 3DS (right) activity on an ongoing event (performance is in blue, number of events in red, black bars show standard deviation), ANOVA tests are shown in red brackets.

The relative high performances (80 to 90%) of the ongoing events presented on figure 2 (left) show that the complementary activities in an event including 3DS are highly successful whatever the success of the 3DS. A 3DS does not alter the prioritization of ongoing events. 3DS is generally ignored and the performance of the ongoing event is not altered because pilots (but one) did not delegate resources to address the 3DS information.

As for DVI use, when a command is issued, its performance does not influence the performance of the ongoing event (figure 2 right). This could be interpreted as a transfer of resources from one standard previous activity (HOTAS for example) to DVI. The ongoing performance variation is not identifiable but gain comes through ease of use as pilots reported.

4.3 Effect of training with DVI and 3DS over runs

The DVI learning curve happens during all 3 first scenarios and the effect is significant ($F^3_{79}=6.34$; $p=0.0007$). The effect is though not significant for 3DS. The only activity of the HMI suite that has a strong potential for early learning is the DVI (figure 3 left). As for the 3DS it seems that the progress was not consistent enough to be noticeable. There has been no significant progress of the 3DS activity over the four scenarios ($F^3_{186}=1.57$; $p=0.19$; NS): It is possible that the learning curve was uneven throughout pilots (figure 3 right). Though, the less than 45% overall success with 3DS is noteworthy.

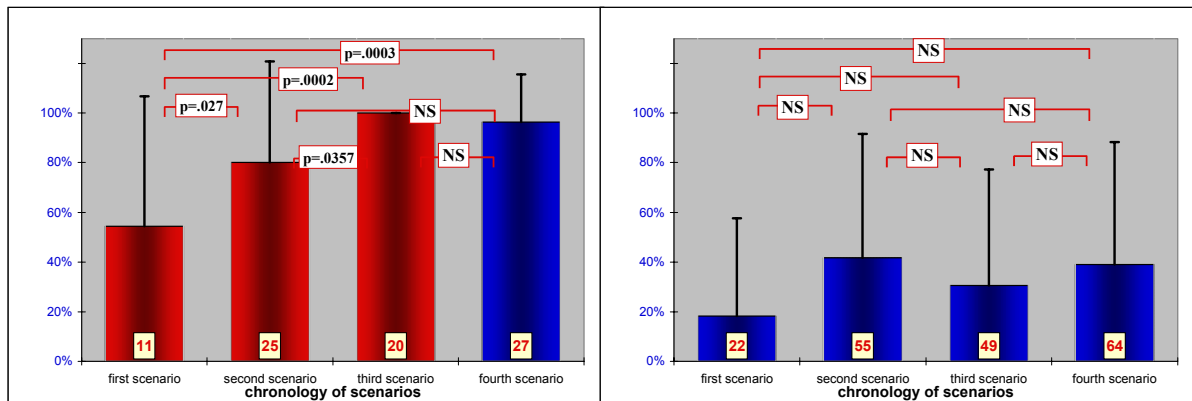


Figure 3: effect of training through successive scenarios on performance of DVI (left) and 3DS (right). The red columns identify the one statistically different than the others (Fisher test shown in red brackets).

4.4 Performance with DVI and 3DS with regards to workload

It appeared that there is no significant effect of workload on event performance (when they are all aggregated) ($F^4_{266}=1.27$; $p=0.28$; NS).

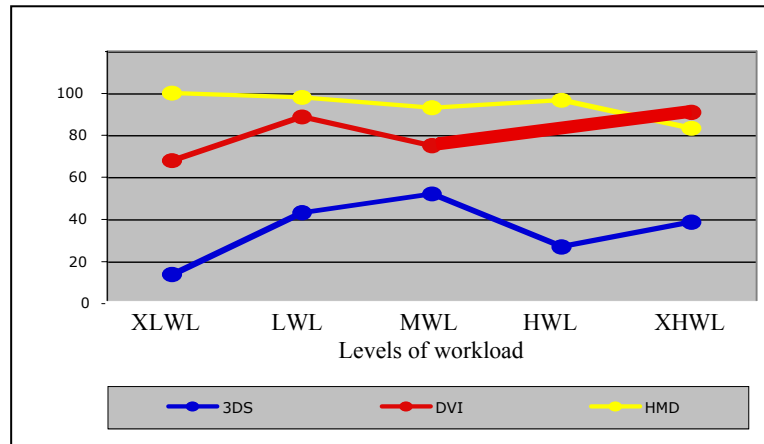


Figure 4: evolution of HMI suite activity performance with regards to workload levels.

When focusing on activity performances, though, there is a decrease in performance with the increase of workload ($F^1_{657}=5.04$; $p=0.025$), but only when Xtra Low, & Low levels are grouped as well as High & Xtra High levels. When analyzing HMI suite related activities with regards to workload (figure 4), it appears that 3DS and DVI performance vary differently. 3DS (in blue) rises from a very low performance to an approximate 50% at medium workload level and then recedes, while DVI (in red) seems to have an improvement of its performance with workload. Note the absence of DVI commands at HW Level. Because performance aspects do not seem all that descriptive, pilots use over workload levels needs to be looked into. To examine this, activity counts were made on DVI use by pilots with regards to workload.

4.5 DVI and 3DS use with regards to workload

It seems that pilot use of DVI varies according to increased workload (figure 5). As a surprise, at higher levels, all pilots but one, use DVI. The reconnaissance level was good but not 100% and usually pilots, at the highest workload levels, are not inclined to use non-fully reliable systems.

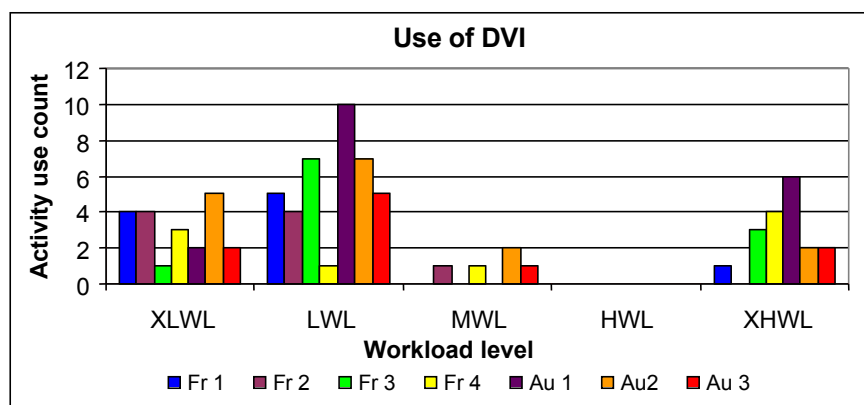


Figure 5: use of DVI with regards to pilots and workload level (F = French pilots; A = Australian pilots).

The surprising lack of use of DVI at the high workload (HWL) level can be explained three ways:

- First, pilots cannot, for some reason, delegate enough resources at that specific level for DVI use. It is a possibility, but then why would they be able to do it at the highest level?
- Second, a collection bias for events including DVI. The evaluator missing events with DVI at the high workload level. But then, this would not have happened for *all* of the pilots and over *all* runs.

- Third, and it is the explanation that is favored, the so called high level work load phase is always likely to be very short because workload increase is very abrupt and can jump directly from Xtra Low, Low & Medium to Xtra High. For instance, pilots consider enemy engagement as Xtra High workload, the build up phase before that is very likely to be short and not suitable for DVI commands.

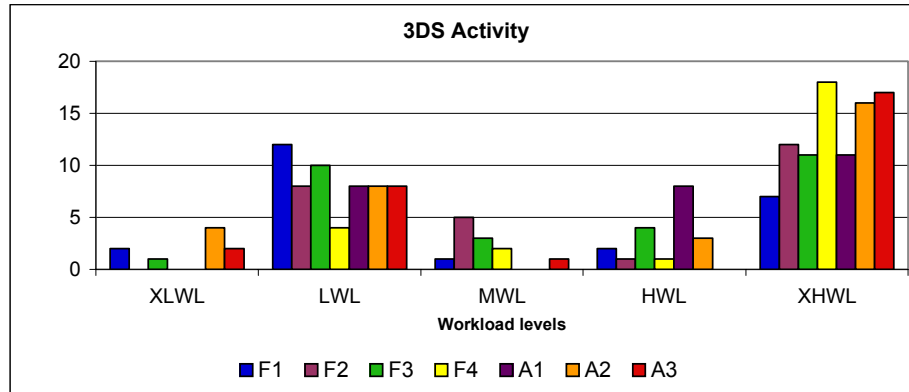


Figure 6: 3DS activity count with regards to pilots and workload level.

3DS use needs to be associated to DVI use, because, though it is imposed on pilots most of the time, they could also use it as a request. For instance, the "wingman" 3DS was on demand via DVI.

Figure 6 shows the total of 3DS events per pilots with regards to workload. It reveals two peaks of exposure, first at low WL figuring the 3DS from friendly parties and second at the exposure to enemy 3DS for the highest level of WL.

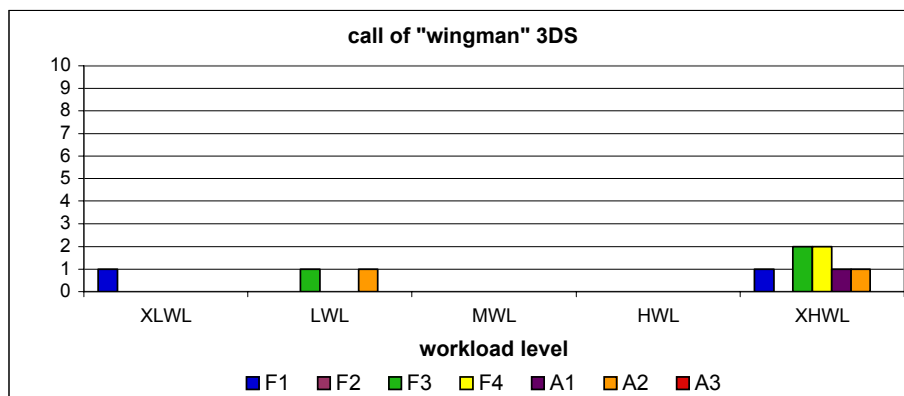


Figure 7: number of DVI "call wingman" the pilots used to localize wingman with 3DS.

3DS contribution to performance remains low whatever the workload, signifying low mastery (<45% success) besides low involvement due to poor expected gain. Such a low effectiveness needs to be commented; the only pilot, who localized well, used it with great advantage for his spatial awareness. However, the majority of the pilots were only able to localize moderately to poorly and 3DS performed for them accordingly. The reasons for this are unclear, as a number of factors appear to be involved. For example, some compromises needed to be made in the implementation of the 3DS in the simulator that impoverished the 3D stimulus. However this was the case for all pilots and hence cannot account for the variability witnessed. Possible factors are multiple but can be grouped with regards to technology, sound physics and human psychophysiology

On the other hand, DVI comes as a surprise as its performance (both activity and event) increases with workload; most pilots use it at the highest level. In fact, the increase of performance with workload happens for pilots that know they are efficient with DVI. Figure 7 shows the on demand ("wingman") 3DS and that they concentrate at the highest level.

Anecdotal reports suggested that it would be impossible to use it at higher workload levels. However pilots used it both for routine commands at low workload, and cost effective commands at higher workloads. Implications are that such strategies of use should be taken into account for training. At higher WL it gave a clear advantage to pilots in selecting functions immediately and directly with no regards to its hierarchical status.

5.0 CONCLUSION

In realistic simulated air combats, operational pilots had a relative easy appropriation of technology they had never used before. Their strategy of use was workload dependant, yet still efficient at the highest workload levels especially for DVI. Throughout missions, pilots displayed an appropriate cognitive resource management strategy mostly based on an effective "trade-off-guided" resource allocation of inputs.

Interestingly, new and resource hungry interface systems are not especially discarded at highest levels of workload when pilots think their use brings a fair trade-off. Such high level resource management strategy should be implemented in novel HMI design.

- [1] Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 63-102). New York, NY: Academic Press.
- [2] Flanagan, J. C. (1954). The critical incident technique. *Psychological bulletin.*, 51(4), 327-359.
- [3] Dix, A., Finlay, J., Abowd, G., & Beale, R. (1998). *Human-Computer Interaction, second edition*: Prentice Hall.

Design and Evaluation of Warning Sounds in Frigate Control Centres

Johan S. van Balken
Sander J. van Wijngaarden

TNO Defence and Security
PO Box 23, 3769 ZG Soesterberg
The Netherlands

{johan.vanbalken, sander.vanwijngaarden}@tno.nl

ABSTRACT

In the ship control centre of the Air Defence and Command Frigates of the Royal Netherlands Navy, auditory warnings are used to indicate critical system conditions and dangerous situations. By means of subjective experiments, the suitability of the applied auditory warnings was assessed, and possible improvements were evaluated. It appeared that for one of the most urgent warning conditions (platform alarm) a less suitable auditory signal was applied, the perceived urgency of which is relatively low, while it is also easily confused with other signals. A straightforward way to modify the perceived urgency (through manipulation of the signal inter-onset interval) was derived from the experimental results.

1.0 INTRODUCTION

Auditory warning signals are intended to draw the attention of somebody (user, operator, driver, pilot,...) to a certain event. A wide range of auditory warning signals have been designed for an equally wide variety of applications, some serving their purpose better than others.

Several guidelines for designing auditory warnings have been reported previously (e.g. [1]). Some researchers have focussed on auditory aspects of warning signal design ([2]), while others have focussed on ergonomic and cognitive aspects ([4]). However, no “fool proof” procedure to design optimal warnings has been suggested yet. This has to do with the complexity of the process (in which different design aspects sometimes impose conflicting requirements), but probably also with the fact that it is essentially a *design* process, featuring traits of art as well as science.

However, by adopting careful procedures for evaluating (candidate) audio signals for warning applications, auditory warning systems can sometimes be optimised considerably. This paper describes the signal evaluation approach adopted by TNO for a single case, namely the Ship’s Control Centre of the Royal Netherlands Navy’s Air Defence and Command Frigates. Although a description of the entire optimisation approach for a complex audio interface (including evaluation of interactions with communications links and “live” speech communication) would be beyond the scope of this paper, a description of the most relevant evaluation tests can be given. These tests are concerned with the perceived urgency, perceptual dissimilarity and audibility of the signals.

2.0 DESCRIPTION OF WARNING CONDITIONS AND SIGNALS

The ship control centre contains working stations for several operators. Each working station produces auditory warnings which can be heard by all operators. Most of these warning signals are generated by the IMCS, the Integrated Monitoring and Control System, which is the interface between the operators and the ship’s machinery. The warning signals evaluated in the current study had already been selected

independently by the designers of the IMCS. Our purpose was to evaluate how each of these signals fits its specific purpose, in the context of the total set of signals and other sounds in the control centre.

2.1. Warning conditions

The ship control centre features a large range of warning and alarm conditions, varying from notifications that a new user has logged onto the IMCS to alarms signalling failure in vital parts of the machinery. The warning signals are structured into three categories, derived from the relative importance of the associated system events. This structure is given in table 1.

Table 1. The events and conditions for which auditory warnings are generated.

		Alert group	Description of condition
Alarm	Alarm	NBCD Alarm	Alarm for Nuclear, Biological or Chemical danger or damage.
		Platform Alarm	Alarm for defects or problems with machinery on board.
	Warning	NBCD Warning	Warning for Nuclear, Biological or Chemical danger or damage.
		Platform Warning	Warning for possible defects or problems with machinery on board.
Error	Error	IMCS Error	An error in the IMCS system because of drop outs in sensors or network components.
Event	Event	Operation Event	Notification of a non-critical condition, marking the execution of an operation or the transition in system states.
		Attention Event	Notification of a change in control of the IMCS.
		Propulsion Warning I	Indication that the requested propulsion condition can not be met.
		Propulsion Warning II	General alarm of the Propulsion Control System.

The conditions listed in table 1 are subdivided into three main categories or *alert groups*. An *alarm* indicates an problem in the machinery or warns for attacks. An *error* is a problem in the IMCS itself, indicating that it is possible that an alarm condition can not be raised when one should be. A change in the operation of the IMCS is marked by an *event*. The propulsion warning conditions also result in auditory warnings, but these are not generated by the IMCS; they have not formally been assigned an alert group category. Informally, their urgency is ranked just above or below “IMCS error”.

The subdivision of alert groups into categories implies differences in *urgency* between the various warning conditions. These differences in urgency should be reflected by the choice of warning signals; each signal should suggest an appropriate amount of urgency, while urgent signals will be allowed to be more intrusive than those of lesser urgency. The appropriateness of the warning signals associated with table I was evaluated in a subjective comparison procedure. Also, it was evaluated whether the chosen signals were sufficiently dissimilar, and whether these signals were audible in the acoustic control centre environment.

2.2. Warning sounds

Different sounds were proposed to be used for each of the alert groups. These sounds are described in table 2. Most sounds are played in a loop and stop when the alert condition is acknowledged by the operator in a computer program. Events do not require such a reaction and are only repeated 2 or 5 times.

Table 2. Description of the sounds to each of the alert groups. A short verbal description is given, as well as the fundamental frequency and duration of the sound and the inter-onset interval (IOI) between two successive sounds and the number of repetitions.

Alert group	Description of associated sound	f_0 (Hz)	dur (s)	IOI (s)	n
NBCD Alarm	buzzer with linear frequency sweep up	400-670	0.5	0.5	∞
Platform Alarm	beep (dominating component: 5.5 kHz)	1000	0.1	0.2	∞
NBDC Warning	6-step-fading synthetic tone	370	0.75	0.75	∞
Platform Warning	sharp buzzer	300	0.25	0.55	∞
IMCS Error	four gradually fading beeps (1 beep:) (complete signal:)	1000	0.15	0.3	∞
			1.2	1.2	
Operation Event	chime	800	0.52	0.52	5
Attention Event	chime	800	0.52	0.52	2
Propulsion warning I	triangular wave beep	2000	0.5	1.0	∞
Propulsion warning II	triangular wave beep	2000	0.15	1.0	∞

3.0 METHODS

By means of three different subjective experiments, the suitability of the applied auditory warnings was assessed, and possible improvements were evaluated. These experiments were aimed at evaluating the perceived urgency of the signals, their perceptual dissimilarity and their audibility.

3.1. Participants

Sixteen university students, 18-25 years of age, participated in the experiments. All participants reported normal hearing. They were not familiar with operations on board of frigates, nor were they in any way introduced to the IMCS; prior knowledge of the sounds or systems would potentially have biased the outcome of the results, since participants would be inclined to adjust their responses towards the current situation. The participants were paid for their attendance.

3.2. Stimulus Material

Three sets of stimuli were synthesized, based on the original signals:

1. The signals of table 2, occurring only once without repetitions
2. These same signals, but with repetitions added (looped) according to the timings of the original set
3. A set of signals based on the IMCS error signal of table 2, but with different repetition rates and temporal structures

Signal sets 1 and 2 are smaller than the original set, since some were based on the same (chime) sound, but a different number of repetitions.

3.3. Pairwise comparison (urgency)

A pairwise comparison test was carried out to determine the relative urgency of the sounds. Participants were presented with a pair of sounds, which they could play back by means of two buttons. Only one sound at a time could be played, but they could switch back and forth between sounds at will. Once they felt that they had heard enough to form an opinion (“which alarm signal sounds more urgent?”), they indicated their choice by pressing one of two “decision” buttons.

All possible pairs of sounds were presented once, in random order. From the participants’ responses, a so-called preference matrix was constructed. This matrix indicates how often each of the signals was chosen over its competitor, the competitor being each other signal in the set.

Relative urgencies were calculated by taking z-scores for each of the cells, and summing row-wise ([5]). This yields a metric that indicates how often each of the signals was perceived to be “the more urgent one”, using the assumption that the underlying statistical distribution is normal. Variance was calculated by means of a jackknifing procedure: for each person the urgencies per sound were determined by subtracting the urgencies calculated from the matrix with the responses of all participants from those with one person left out. The significance of differences in urgency scores was determined by a binomial test. The pairwise comparison experiment was carried out with signal sets 1, 2 and 3.

3.4. Triadic comparison (dissimilarity)

A triadic comparison test was performed to determine the relative dissimilarity between the sounds. The setup was similar to the pairwise comparison, only groups of three sounds were presented. All sounds could be re-played by means of different buttons. For each triad participants had to successively point out the most similar pair and the most dissimilar pair by means of pressing buttons representing pairs of sounds.

All triads were presented once, in random order. From the participants’ responses a so-called dissimilarity matrix was formed. For every pair of sounds a dissimilarity score was calculated by adding 1 unit for each “most dissimilar” rating and subtracting 1 unit for every “most similar” rating. The most similar pairs were determined by finding the pairs with the lowest scores. For the triadic experiment signal set 1 was used.

Dissimilarity matrices can be subjected to multidimensional scaling techniques to explore the dimensions underlying the participants’ similarity judgments. For brevity the results of such an analysis (which is routinely performed as part of this type of experiments) is not given here. Instead, the analysis presented here (in section 4) is based on the raw dissimilarity scores, converted into equivalent z-scores.

3.5. Audibility

Audibility was measured by determining the threshold of audibility when presenting the warning sounds in background noise. Participants could control the overall sound level of the looped sounds with a slider and had to adjust the loudness to make the sounds just inaudible.

All sounds were presented twice. The background noise was recorded in the control room of the frigate. The audibility results are reported here as the average signal-to-noise ratio corresponding to the threshold of audibility, where signal and noise levels are expressed as A-weighted sound pressure levels. The audibility threshold was determined for signal set 2.

4.0 RESULTS

4.1. Perceived urgency

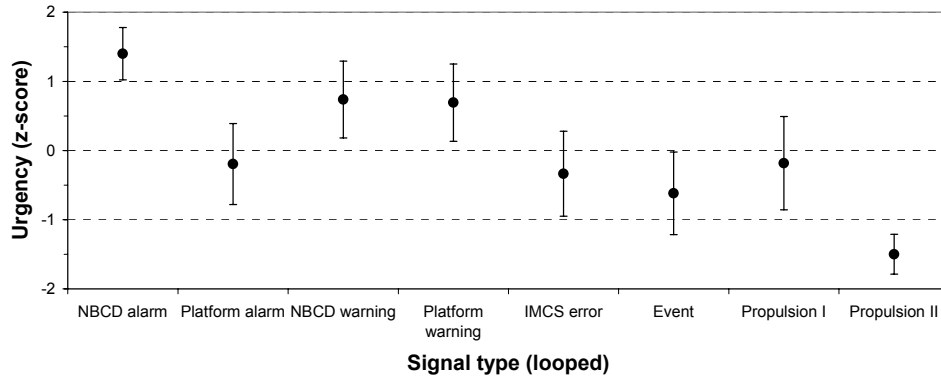


Figure 1. Perceived urgency of the looped warning sounds (signal set 2). The error bars represent one standard deviation.

Figure 1 shows that the perceived urgency rated by the participants matches the urgency of most of the alarm conditions, except for platform alarm. The platform alarm sound is judged to be less urgent than the NBCD warning ($p < 0.05$). The Propulsion II sound is least urgent of all the sounds. Propulsion II is less urgent than Event ($p < 0.05$).

4.2. Perceived urgency - tempo and structure

The urgency decreases with a longer inter-onset interval (Fig. 2). The three signals with an average inter-onset interval of 400 ms were judged to be more urgent than the signals with an average interval of 700 ms ($p < 0.01$).

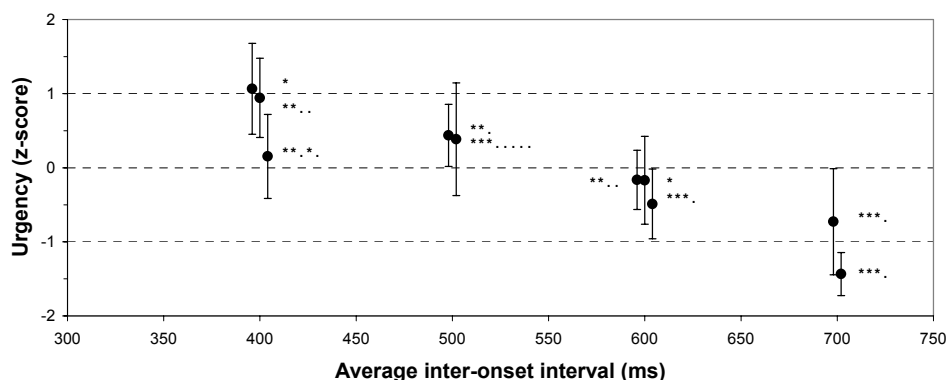


Figure 2. Urgency of IMCS beep-patterns (signal set 3) versus their average inter-onset intervals. The structure of the repetition patterns is indicated by the codes next to the points: each character denotes a period with the same duration, a '*' marks a 150 ms IMCS error beep, a '.' marks a silence. Error bars represent one standard deviation.

4.3. Perceptual dissimilarity

Dissimilarity scores were calculated for all pairs of signals out of signal set 1. Table 3 lists the five most similar pairs.

Table 3. The five most similar pairs. Similarity is expressed the z-score of the similarity scores, and the associated cumulative normal distribution for that value.

Pair		z-score
IMCS error	Platform Alarm	2.52
IMCS error	Platform Warning	1.42
Event	Propulsion I	1.34
Platform Alarm	Propulsion I	0.96
Platform Alarm	Platform Warning	0.92

The signals for “IMCS error” and “platform alarm” in set 1 are most similar. A remarkable result is that “platform alarm,” one of the more urgent alert groups, is listed three times in the table with the five most similar pairs.

4.4. Audibility

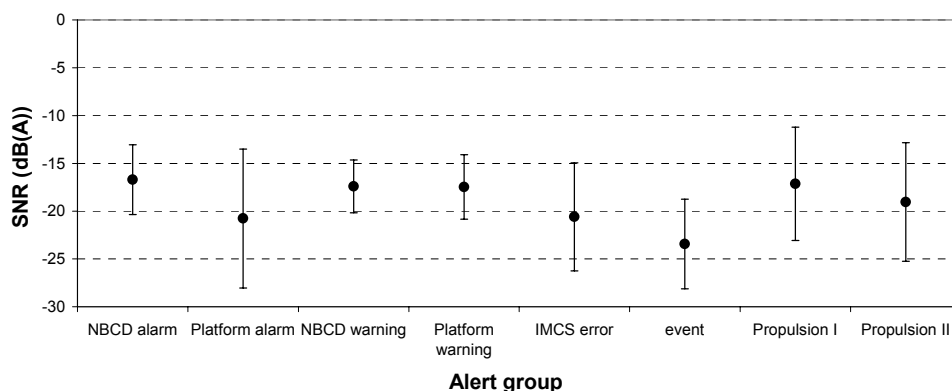


Figure 1: The audibility threshold for the warning sounds (signal set 1), expressed as the A-weighted signal-to-noise ratio (SNR). Error bars represent one standard deviation .

The average audibility threshold over all sounds lies at -19.0 dB(A) SNR. Per sound, every threshold lies within 5 dB from this average.

5. DISCUSSION AND CONCLUSIONS

Most of the signals used in the control centre were designed adequately. However, it appears that for one of the most urgent warning conditions (*platform alarm*) a less suitable auditory signal was chosen. The perceived urgency of the platform alarm sound is relatively low. Moreover, it is easily confused with a the IMCS error signal. Both sounds consist of 1000 Hz beeps, although the IMCS error signal consists of a series of 4 fading beeps, while the platform alarm uses the beeps continues. These two signals show up as the most similar pair in the triadic comparison experiment. We recommended to change either of the signals to avoid recognition problems.

Figure 2 shows that a mismatch in perceived urgency can be repaired relatively easily by adjusting the inter-onset interval. In case of the platform alarm, this means that we recommended increasing the tempo of the signal.

The audibility thresholds are satisfactory for all signals. It is commonly recommended that warning signal levels are adjusted to 15 – 25 dB above the detection threshold ([2],[3]). In this case, this is feasible with warning signal levels that do not exceed the level of the background noise.

REFERENCES

- [1] Edworthy, J. Loxley, S. & Dennis, I. (1991). "Improving auditory warning design: relationship between warning sound parameters and perceived urgency." *Human Factors* 33, 205-231
- [2] ISO/DIS 7731 (2002). *Ergonomics – Danger signals for public work areas – Auditory danger signals*. ISO, Geneva, Switzerland.
- [3] Patterson, R.D. (1982). *Guidelines for Auditory Warning Systems on Civil Aircraft*. (CAA Paper 82017). London, UK: Civil Aviation Authority
- [4] Stanton, N.E. & Edworthy, J. (1999). *Human Factors in Auditory Warnings*. (Ashgate, Aldershot, UK)
- [5] Torgerson, W.S. (1958) *Theory and methods of scaling*. (Wiley and Sons, New York, USA).



Spatial Audio Displays for Improving Safety and Enhancing Situation Awareness in General Aviation Environments

Brian D. Simpson, Douglas S. Brungart, Robert H. Gilkey, and Richard L. McKinley

Air Force Research Laboratory, AFRL/HECB, Bldg 441, 2610 Seventh Street,
Wright-Patterson AFB, OH 45433-7901 Tel: (937) 255-3660 ext 462 Fax: (937) 656-7680
Wright State University, Department of Psychology, 335 Fawcett Hall,
Dayton, OH 45435 Tel : (937) 775-3337 Fax : (937) 775-3347

brian.simpson@wpafb.af.mil

ABSTRACT

Most current displays in general aviation (GA) environments employ, at best, relatively simple audio displays that do not fully exploit the operator's sensory processing capabilities, and thus may not allow pilots to take full advantage of the available information. This paper describes a series of four experiments conducted in an actual flight environment at the USAF Test Pilot school that evaluate the utility of spatial auditory displays as a navigation aid and an attitude indicator. Performance was measured in tasks that required pilots to fly in the direction of a spatial audio "navigation beacon," and use an auditory artificial horizon display to detect changes in attitude and maintain straight and level flight when no visual cues were available. The results from these studies indicate that spatial audio displays can effectively be used by pilots for both navigation and attitude monitoring. In addition, because the stimulus employed in the audio horizon display was self-selected music, pilot acceptance was high and it was possible to provide ongoing information to the pilot without the risk of annoyance. Although performance with standard display configurations remains superior, spatial audio displays may be used to supplement these existing displays, and to provide cues when critical information is either obscured or missing, such as in low-visibility conditions or when the pilot's attention is focused on another task. Thus it appears that spatial audio may be used to support situation awareness and improve overall safety in the general aviation environment.

1.0 INTRODUCTION

Aviation professionals have been considering the use of spatialized auditory displays (3D audio) in military cockpits for target acquisition, threat avoidance, and the separation of multiple talkers [1,2,3,4] and in general aviation (GA) cockpits for taxiway navigation and collision avoidance [5,6,7,8] for many years. However, despite the long-term interest in this technology, nearly all studies have been conducted in simulator environments, and the few flight tests that have been published have yielded primarily qualitative data [3]. This paper reports results of a flight test of spatial auditory display concepts that have been developed to enhance pilot situation awareness (SA) and potentially improve safety in GA flight environments.

The accident rate in GA environments is extremely high relative to the rates in scheduled commercial air-carrier aircraft, and, in most cases, pilot error is cited as a primary or contributing factor. This suggests the need for new display technologies that increase the pilot's knowledge of the condition of the aircraft and the status of the surrounding airspace (situation awareness, SA). On the other hand, simply providing additional visual instrumentation to a pilot who is only trained for VFR operations may do little to increase SA. In this

Simpson, B.D.; Brungart, D.S.; Gilkey, R.H.; McKinley, R.L. (2005) Spatial Audio Displays for Improving Safety and Enhancing Situation Awareness in General Aviation Environments. In *New Directions for Improving Audio Effectiveness* (pp. 26-1 – 26-16). Meeting Proceedings RTO-MP-HFM-123, Paper 26. Neuilly-sur-Seine, France: RTO. Available from: <http://www.rto.nato.int/abstracts.aps>.

context, the use of spatialized auditory displays is particularly appealing. The allure of 3D audio is based on at least three facts: 1) The visual system is already overloaded. Pilots must monitor, in addition to the out-the-window view, a variety of visual flight instruments which report the condition and location of the plane and status of the airspace, and further must consult charts and other printed material in certain situations. Auditory displays have the potential to add new information without further straining visual resources or to provide redundant information to reinforce visual displays. 2) Hearing is a 4π steradian sense. That is, the auditory system can respond to events (e.g., traffic) above, behind, and below the pilot (i.e., in locations that are outside the visual field of view). 3) Hearing is a 24-hr sense. That is, the auditory system is always “on,” even in sleep, and thus has the potential to alert a pilot even when distracted by other tasks. Indeed, the auditory system is designed to respond to even subtle changes in the auditory environment and direct visual attention to the source of the new event. The task of the display designer is to exploit this natural potential of the auditory system in a manner that does not distract or annoy the pilot under normal flight conditions.

This paper reports the results of flight tests of two displays to help pilots control an aircraft with limited visual information: an auditory navigation aid and an auditory attitude indicator. Both displays provide usable information about the orientation of the aircraft in a way that allows them to be consciously used when desired and to attract attention when needed, without acting as an ongoing annoyance within the cockpit.

2.0 GENERAL METHODS

2.1 Subjects

Three members of Class 04A at the United States Air Force Test Pilot School served as subject pilots. The pilots had between 8 and 15 years of flight experience in various aircraft.

2.2 Apparatus

A twin-engine turboprop passenger and cargo aircraft (C-12C Huron test aircraft) was used as the flight environment for data collection. A pallet and rack containing the hardware necessary for stimulus generation, presentation, and data collection were securely mounted approximately in the middle of the aircraft, where the experimenter had sufficient access to monitor the status of the equipment and run the experimental software.

The primary component of the auditory display system was a PC-based laptop computer that functioned as the experimental control computer and the audio server. The server used the Sound Lab (SLAB) API developed at NASA-AMES [9]. SLAB enabled the system to generate spatialized auditory stimuli such that sounds presented over headphones appeared as if they came from actual locations in space external to the pilot’s head. This spatialization process is accomplished by modifying an arbitrary signal to contain the direction-dependent acoustic transformations (the head-related transfer functions, HRTFs) that a sound undergoes as it travels from a source to the eardrum. An audio mixer (4-channel Allen & Heath Phoenix) mixed the spatial audio cues with the aircraft intercom system, and the sounds were sent to a stereo headset (Bose AH-TS ANR headset modified to transduce a stereo signal) worn by the subject pilot.

The spatialization of auditory stimuli was updated in real time in one of two ways, depending on the experimental condition: 1) with respect to the orientation of the pilot’s head or 2) with respect to the position and attitude of the aircraft. Spatialization with respect to the pilot’s head was achieved using an inertial headtracking system consisting of a microelectromechanical system (MEMS) inertial measurement unit (IMU), a Global Positioning System (GPS) receiver (Garmin Industries), and a PC-based laptop computer. Spatialization with respect to the aircraft was achieved using a Time, Space, and Position Information (TSPI) GPS-Aided Inertial Navigation Reference (GAINR) system. This system served as the primary source of

truth data regarding the aircraft's position and attitude. The C-12C aircraft was also instrumented with a separate data acquisition system.

2.3 Procedure

On each day of data collection, a pre-test briefing was given in which weather conditions, traffic, and any issues regarding the experimental system were discussed. If a decision was made to proceed with data collection on that day, a full test of experimental instrumentation was conducted before the session commenced.

During data collection, the subject pilot was in the left-hand seat of the aircraft, the safety pilot was in the right-hand seat, and the experimenter was behind them, adjacent to the experimental control computer. In each session, four experiments were conducted, one baseline experiment on auditory localization and three experiments examining the utility of 3D audio as a navigation aid or as an attitude indicator. All four experiments were conducted in a single test flight, and two such flights were conducted for each pilot - one for the GAINR-coupled 3D audio condition and one for the headtracker-coupled 3D audio condition. In order to complete all conditions, there were a total of 8 flights lasting approximately 2 hours each. All tests were conducted at airspeeds of approximately 170 Knots Indicated Airspeed (KIAS) at an altitude of approximately 12,000 feet, though these values varied slightly depending on weather and traffic conditions. Only one flight took place on each day, and all flight tests were conducted under day visual meteorological conditions (VMC).

3.0 EXPERIMENT I. AUDITORY LOCALIZATION

In order to assess the relative benefit of spatial audio cues for various flight tasks, it was important to first obtain a baseline measurement of the system performance and the localization ability of the subject pilots. A simple auditory localization task was implemented to examine these issues.

3.1 Specific Methods

3.1.1 Stimuli

The stimulus utilized in the auditory localization task consisted of an ongoing train of broadband noise bursts. The azimuthal direction of the stimulus was generated in reference to the aircraft heading at the time the trial was initiated. For example, if the aircraft magnetic heading was 180° and the stimulus was generated at 3 o'clock, the stimulus had a magnetic azimuth of 270°. The stimulus could originate from any of 12 clock positions (30° spacing) and at any of 3 elevations (-30°, 0° and +30°, referred to as low, mid, and high, respectively). The stimuli were spatialized relative to the orientation of the pilot's head (headtracker-coupled 3D audio) or the position and attitude of the aircraft (GAINR-coupled 3D audio), depending on the condition.

3.1.2 Procedure

The auditory localization test consisted of two tasks. First, a baseline auditory localization test was conducted while the aircraft was on the ground and stationary, with the engines running. Next, the same localization task was conducted in-flight.

In both tasks, the experimenter initiated the start of a trial by entering an input on a graphical user interface (GUI) on the experimental control computer, at which point the auditory cue was presented to the subject pilot. The subject pilot's task was to determine the direction of the cue and respond by stating a clock position and elevation (e.g., "2 o'clock, high"). The experimenter entered this response on the computer, then initiated the start of the next trial. One auditory cue was presented from each of 12 clock positions at each of the 3 elevations, for a total of 36 trials, the order of which was randomized for each subject pilot. For the

in-flight portion of this experiment, the subject pilot first had to fly the aircraft to a desired altitude and achieve a desired airspeed at a safe heading. Once accomplished, the subject pilot was required to maintain straight and level unaccelerated flight while localizing the auditory stimuli. Due to hardware constraints, the localization tasks were completed for the GAINR-coupled 3D audio condition in the first session (i.e., the first flight) and for the headtracker-coupled 3D audio condition in the second session (i.e., the second flight) for all subject pilots.

3.2 Results

A 3-pole coordinate system [10] is employed here to examine the data from the auditory localization task. The coordinates of this system, Left/Right (L/R), Front/Back (F/B), and Up/Down (U/D), provide a particularly useful framework for examining auditory localization data because localization in each of these spatial dimensions appears to be mediated by different physical cues. In this system, the azimuthal position of a sound source is described in terms of a L/R coordinate, providing a measure of laterality, a F/B coordinate, providing a measure of how far a sound is in front of or behind a listener. Dividing azimuth in this way ensures that accurate performance in the L/R dimension will not be obscured by front-to-back or back-to-front reversals, which are not uncommon in spatial audio displays. For instance, a sound at 0° azimuth, 80° elevation that was judged to be at 180° azimuth, 80° elevation would suggest an error of 180° in azimuth using a traditional 2-pole coordinate system. However, in the 3-pole system, this error shows up as a 20° error in the F/B dimension, and no error in the L/R dimension. Finally, in this system the U/D coordinate is simply the elevation angle of the sound source.

Results from the localization tasks are shown in Figure 1, where mean angular errors in localization are plotted for each individual pilot, as well as the overall mean, in the L/R dimension (leftmost column), F/B dimension (middle column) and U/D dimension (rightmost column). In each panel, the black bars represent the data from the GAINR-coupled condition and the white bars represent the data from the headtracker-coupled condition. The top row depicts the results from the ground localization task and the bottom row depicts the results from the in-flight localization task. In both the ground and in-flight tasks, localization errors in the L/R dimension were generally the lowest, consistent with previous results in the literature [11]. Moreover, localization performance was best when the audio cues were coupled to the headtracker (i.e., when the cues were spatialized with respect to the orientation of the listener's head), with performance improvements of approximately 3° to 7° over the GAINR-coupled condition. This head-referenced display allowed the pilot to use head motion to turn and face the source directly, thus bringing the source into the “auditory fovea,” where spatial acuity is greatest.

The difference across conditions was even more pronounced in the F/B dimension, where mean localization errors were reduced from approximately 37° in the GAINR-coupled display to 13° in the headtracker coupled display. The large F/B errors in the GAINR-coupled condition likely represent front-back confusions, which were greatly reduced when the pilot utilized the dynamic localization cues associated with head motion. Although similar cues could be available to the pilot in the GAINR-coupled condition by changing the heading of the aircraft, the pilots were required to maintain a straight and level course and in general were not exposed to those dynamic cues.

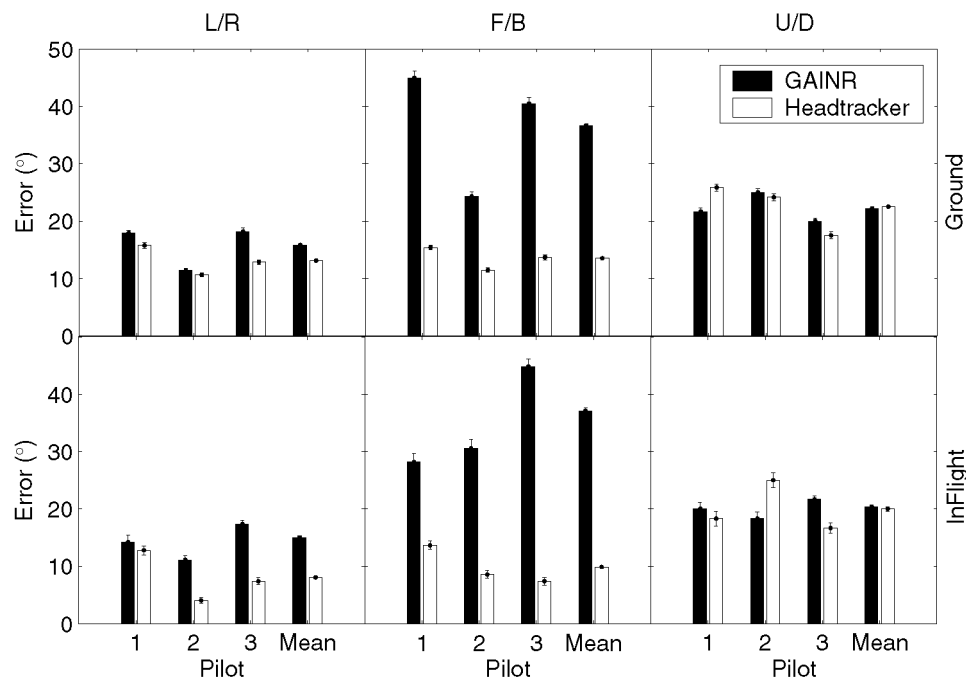


Figure 1: Localization errors for the individual pilots and the mean in the Left/Right, Front/Back, and Up/Down dimensions for the ground localization task (upper panels) and the in-flight localization task (lower panels). Both the GAINR-coupled (black bars) and headtracker-coupled (white bars) 3D audio conditions are shown. Error bars indicate ± 1 standard error.

Localization in the U/D dimension was only slightly better than chance and was independent of the 3D audio condition. These results are not altogether surprising because the cues believed to mediate spatial hearing in the U/D dimension – fine-structure spectral cues in the HRTF – are highly individualized. Previous literature on spatial audio suggests that U/D localization improves when the display employs the listener's own HRTFs [12]. However, because the current display employed a generic set of HRTFs, those pilot-specific cues were not employed.

Performance for the in-flight localization task (bottom row) was not substantially different from the baseline ground localization task, although errors were slightly lower overall. This may be attributable to a learning effect, as the in-flight task always followed the ground task.

4.0 EXPERIMENT II. SPATIAL AUDIO AS A NAVIGATION AID

Once it was established that the 3D audio display could adequately generate the cues required to support accurate sound localization, the utility of this display as a flight navigation aid was evaluated. In Experiment II, a task was implemented in which spatialized audio cues were provided as a navigation aid to direct the flight path of pilots along a given heading.

4.1 Specific Methods

4.1.1 Stimuli

The stimulus employed in the navigation task was either a nonspatialized (diotic) verbal cue providing a command heading to the pilot (e.g., "Set Course 3-1-5"), which served as a baseline measure of performance, or a spatialized audio cue that appeared to originate from a particular direction. The spatialized

cue consisted of the same noise stimulus used in Experiment I, interrupted at periodic intervals by the spatialized phrase “Set Course.” Note that no numerical heading cue was presented verbally in the 3D audio conditions. As in the localization task, the spatialized stimuli were positioned relative to the orientation of the pilot’s head (headtracker-coupled) or relative to the position and attitude of the aircraft (GAINR-coupled), and only one condition was tested in each block.

4.1.2 Procedure

Before the start of each data collection session, the subject pilot was required to maintain straight and level unaccelerated flight, during which the heading depicted on the Horizontal Situation Indicator (HSI) and the current time were determined to see if any discrepancy between the GAINR system and HSI existed. Once everything was checked, the experimenter initiated the start of a trial on the control computer, at which point the command heading was presented. The task of the subject pilot was to fly the aircraft from the initial heading to the command heading. When the pilot believed this had been accomplished, he indicated this to the experimenter, at which point the trial ended. In each experimental session, 20 navigation trials were collected, 10 in which the nonspatialized verbal cue was employed and 10 in which the spatialized audio cue was employed. Each cue was tested on alternate trials, and the command heading was generated randomly on each trial. In the event that this command heading was determined to be unacceptable due to traffic, planning, or other reasons, a new command heading was generated. As before, spatial audio cues were GAINR-coupled in the first experimental session for each subject and headtracker-coupled in the second session.

4.2 Results

Results from the navigation task are shown in Figure 2. For each pilot, the mean difference between the command heading and the final heading (angular error) is plotted for each condition. The overall mean is also shown. As can be seen, performance in the baseline verbal cue condition was always the best. Pilots were, on average, able to guide the aircraft to within 1°-2° of the command heading. This level of performance is not surprising given the fact that pilots regularly use this method for navigation, and the meaning of such a display is unambiguous. On the other hand, angular errors in the GAINR-coupled and headtracker-coupled conditions were substantially higher (8°-9°). One might expect performance in the headtracker-coupled condition to be slightly better than performance in the GAINR-coupled condition – a head-referenced display would enable the pilot to use exploratory head movements to determine the direction of the command heading with a high degree of accuracy. However, in order to truly minimize errors, the pilot had to be certain that the orientation of his head was perfectly aligned with that of the plane, and that is a nontrivial task. Moreover, the accuracy of the headtracker progressively degraded throughout the trials. Both of these issues likely contributed to these larger than expected errors in this condition.

Despite the larger errors found in the headtracker-coupled condition, the pilots were able to get to their perceived command heading in a fast and efficient manner. Taking into account the differences in initial heading across trials, the results indicate that the pilots approached the command heading at similar rates in the verbal cue and headtracker-coupled cue conditions (1.25 deg/sec and 1.28 deg/sec, respectively), but much more slowly in the GAINR-coupled display (0.79 deg/sec). It is likely that the pilots had to initiate small aircraft maneuvers in the GAINR-coupled condition in order to resolve front/back confusions, which would result in comparatively longer trial durations. Nevertheless, as shown in Figure 3, the pilots were able to fly a relatively direct course to the command heading in all conditions.

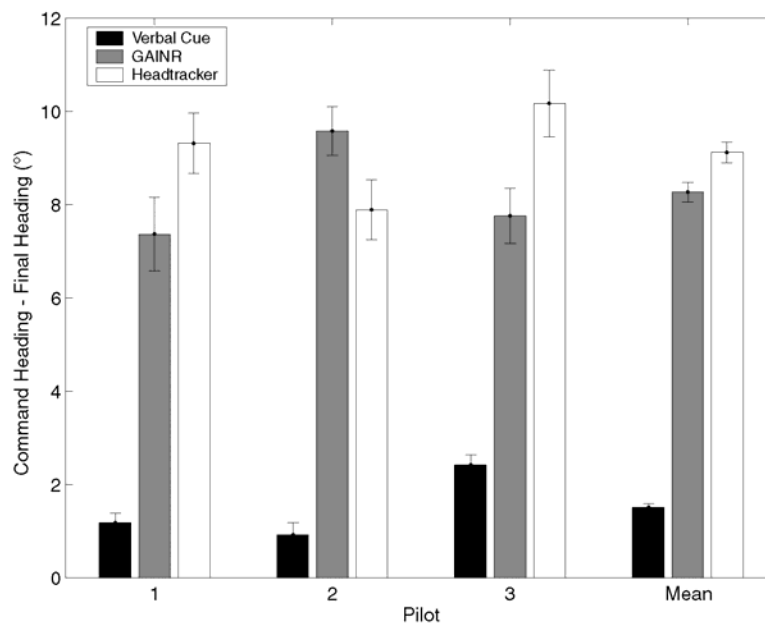


Figure 2: The difference between the initial heading and the command heading (in degrees) in each of the display conditions. Mean data are shown for each individual subject pilot as well as the mean of all three subjects. Error bars indicate ± 1 standard error.

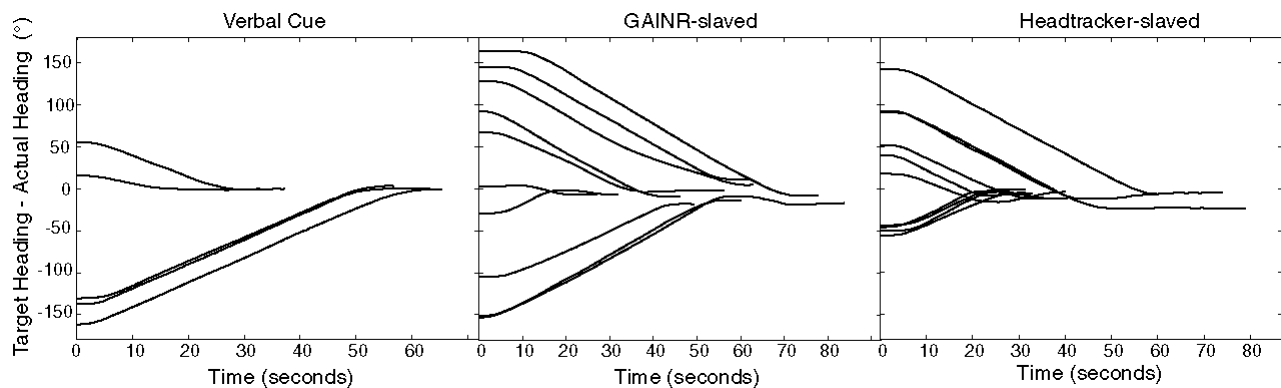


Figure 3: Flight paths as a function of time in each display condition for a single representative pilot.

5.0 EXPERIMENT III. SPATIAL AUDIO AS AN ATTITUDE INDICATOR: DETECTING CHANGES IN ATTITUDE

For Experiment III, an audio artificial horizon was developed. This display was designed to provide continuous attitude information to the pilot in order to enhance situation awareness.

5.1 Specific Methods

5.1.1 Stimuli

For this task, an audio artificial horizon cue was employed that provided the subject pilot with a continuous stream of information about the roll and pitch of the aircraft. Although the concept of an audio artificial horizon is not new, previous attempts have employed only relatively simple auditory stimuli such as tones in noise that vary in frequency, modulation rates, etc. [13]. These displays, while informative, can require a great deal of training to be effective. More importantly, they are not well tolerated by pilots because the ongoing nature of the sounds proves to be an annoyance and potentially a distraction. Because the goal is to provide *continuous* information about the attitude of the aircraft, these displays may be unsatisfactory.

In order to address these issues, an alternative audio display was developed that employs music as a continuous auditory attitude indicator. It was believed that such a display would be widely accepted by the pilot population, as it is not uncommon for pilots to listen to music over long periods on cross-country flights. For this task, the audio signal was obtained by allowing each pilot to select an audio CD with music they would enjoy over the course of the flight. The music on the CD was converted into a monaural '.wav' file by summing together the left and right audio channels, and then this '.wav' file was processed to add pitch and roll information and played back to the pilot over stereo headphones during in-flight testing. Roll was indicated by panning the music from the left ear to the right ear in response aircraft attitude. This was accomplished by decreasing the level of the signal in the ear associated with the higher of the two wings, with a maximum ILD of 48 dB when the magnitude of the roll was 30 degrees or greater. Pitch information was indicated in two ways. First, a spectral cue was provided in which the signal had a low-frequency emphasis when the aircraft was pitched down and a high-frequency emphasis when the aircraft was pitched up. This was implemented by passing the signal through a harmonic filter with peaks spaced every 350 Hz (low-frequency emphasis) or every 2500 Hz (high-frequency emphasis). A second cue for pitch related to the apparent image width of the audio signal. This manipulation was accomplished by varying the interaural cross-correlation of the signal such that level flight resulted in a perfectly correlated signal that was heard as a relatively compact auditory image; any deviation in aircraft pitch from level resulted in a lower correlation value and thus a wider auditory image. This decorrelation was achieved by convolving the left ear signal with a flat-spectrum 128-point random noise and the right ear signal with a time-reversed version of exactly the same 128-point random noise. These two cues (the frequency shaping and the interaural decorrelation) were scaled to produce a gradual transition from an unfiltered, perfectly correlated signal in a straight and level orientation to a maximally shaped, uncorrelated signal for a plane pitched up or down by 10 degrees. Because this display was intended to provide information about the attitude of the aircraft, it was believed that the most appropriate display would be one that was plane-referenced (i.e., coupled to the GAINR).

5.1.2 Procedure

In this experiment, the task of the subject pilot was to identify the direction of a change in the aircraft's attitude from level flight. After controlling the aircraft for a short time in order to become familiar with the dynamics of the audio horizon display, the subject pilot donned a blindfold and the safety pilot took over the controls of the aircraft. Data collection commenced when the subject pilot indicated that he was ready, at which point the experimenter started the trial and the safety pilot began to adjust the attitude of the aircraft (pitch up, pitch down, roll left, or roll right). The dimension and direction of change were chosen randomly from trial to trial. The attitude changes were always made at a rate of 1 degree per second, thus providing a repeatable and reliable stimulus. This slow rate also minimized any vestibular or "seat-of-the-pants" cues indicating motion. The trial ended when the subject pilot identified the direction of change with a verbal response (e.g., "Roll Left") and the experimenter entered the response on the control computer. If the maneuver brought the aircraft to the boundaries set as the limits for safe attitude ($\pm 30^\circ$ of roll or $\pm 10^\circ$ of

pitch), the trial was terminated. The audio condition was alternated from trial to trial, with the artificial horizon presented on the first trial.

5.2 Results

The results from the Change in Attitude task are shown in Figure 4, where the percent correct identification of the direction of attitude change is plotted for each actual attitude change condition, averaged across dimension (i.e., roll [left and right] and pitch [up and down]) and across subjects. The black bars represent performance in the task when no audio cue is given and the white bars indicate performance when the audio horizon cue is provided. As can be seen, the percentage of correct identifications improved substantially in both pitch and roll when the audio horizon was provided, with mean values of nearly 90% with the horizon cue as compared to approximately 43% with no audio cue. Moreover, pilots correctly identified the change in attitude when the aircraft pitched up 100% of the time, suggesting that this was a very robust cue. Overall, it appears that the cues for pitch may have been more salient and provided a slightly larger benefit relative to the baseline than the cues for roll.

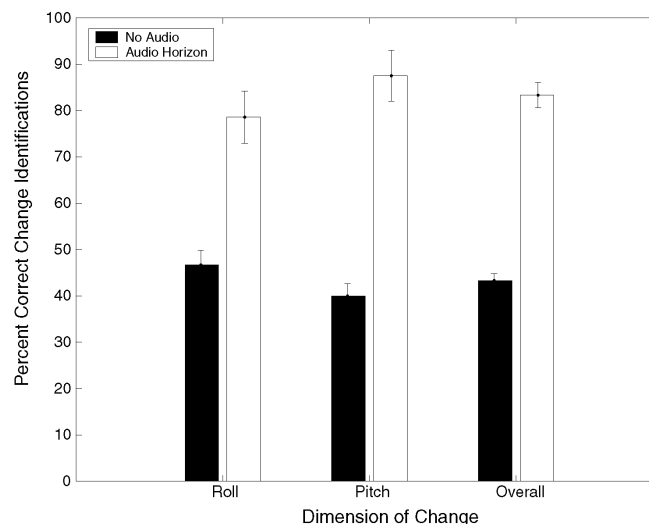


Figure 4: The percentage of correct identifications of the dimension and direction of change in the aircraft attitude. The black bars represent the baseline no-audio condition and the white bars represent performance with the artificial auditory horizon. Mean correct detections, averaged across dimensions, are also shown. Error bars indicate ± 1 standard error.

Thresholds for the correct change in detection in each direction are shown in Figure 5. Threshold values for roll were reduced by nearly 7° with the audio horizon cue and threshold in pitch were reduced more than 3° . Recall that the response method in this task was verbal report, which is not ideal for time critical measures. Because of the inevitable delay between the subject pilot's response and the experimenter's input, there is reason to believe that the threshold values reported here underestimate the true performance of the pilots with this display. While these improvements are substantial, it is difficult to assess whether they are

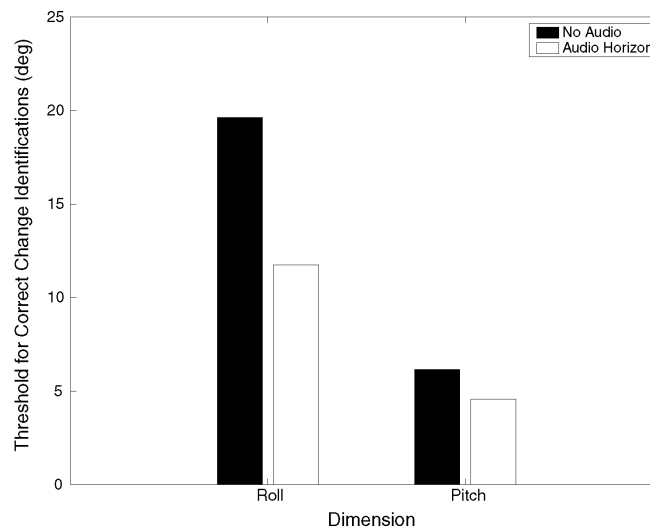


Figure 5: Mean threshold values for correct identification of the direction of change in roll and pitch, averaged across each of these dimensions. The black bars represent the baseline no-audio condition and the white bars represent performance with the artificial auditory horizon.

sufficient in all situations. Nevertheless, in the absence of visual information, the audio horizon provides a much more reliable cue for aircraft attitude than simple “seat-of-the-pants” cues.

6.0 EXPERIMENT IV. SPATIAL AUDIO AS AN ATTITUDE INDICATOR: MAINTAINING STRAIGHT AND LEVEL UNACCELERATED FLIGHT

In Experiment IV, the audio horizon display, in conjunction with a spatialized navigation aid, was evaluated as a means of supporting the pilot’s ability to maintain straight and level flight in the absence of any out-the-window views or any visual instrument reference.

6.1 Specific Methods

6.1.1 Stimuli

The audio horizon cue utilized in Experiment III was also utilized in Experiment IV as an aid to the pilot for maintaining level flight. In addition, the audio cue used to indicate a command heading in Experiment II was used to provide a navigation aid to the pilots for maintaining straight flight. The navigation aid was always presented from a direction consistent with the aircraft’s heading at the start of the trial.

6.1.2 Procedure

In this task, the subject pilot was required to fly the aircraft while blindfolded and maintain straight and level flight. Before the start of each trial the subject pilot entered a trim input on an acceptable heading. He then donned the blindfold and data collection was initiated. Each trial was 5 minutes in duration, during which the safety pilot monitored the status of the aircraft. If the aircraft flew beyond the attitude limits of this ($\pm 30^\circ$ of roll and $\pm 4^\circ$ of pitch), the safety pilot took control of the aircraft and the trial was paused until straight and level flight was regained, at which point the trial resumed. Two such trials were conducted on each flight, one in which no audio cue was provided, and one in which the audio horizon display and the navigation aid were provided. The order of conditions varied across subjects.

6.2 Results

The ability of the pilots to maintain straight and level flight is shown in Figure 6. Here, RMS errors in roll (left panel) and pitch (right panel) are plotted for each individual subject and the overall mean in each display condition. Note that the ordinates on the two panels differ.

As can be seen, the audio horizon display (white bars) led to relatively stable performance in both roll and pitch. RMS errors for roll were extremely low for Pilots 2 and 3 (approximately 2° - 3°) but higher for Pilot 1 ($\approx 11^{\circ}$). Moreover, RMS errors in pitch were much lower overall than those for roll ($\approx 1^{\circ}$). Note that the data from the baseline no-audio condition (black bars) are merely shown for comparison purposes. That is, the subject pilots found that they were generally unable to identify the attitude of the aircraft without the audio horizon and thus made few control inputs during this test condition. Therefore, the results shown reflect only the effectiveness of the initial trim input and the stability of the aircraft, and provide little information about pilot performance. The impact of this can be seen more clearly in Figure 7, which depicts the flight paths in the baseline no-audio condition (left panels) and 3D audio condition (right panels). In each panel, the deviation from straight and level in roll (top row) and pitch (middle row) are plotted as a function of flight time, as is the deviation from initial heading (bottom row). The dashed lines in each panel indicate the boundaries of safe travel in that particular dimension (± 30 in roll and ± 4 in pitch). The discontinuities in the data represent instances where the boundaries of safe travel were breached, the trial, and then restarted after straight and level flight was regained.

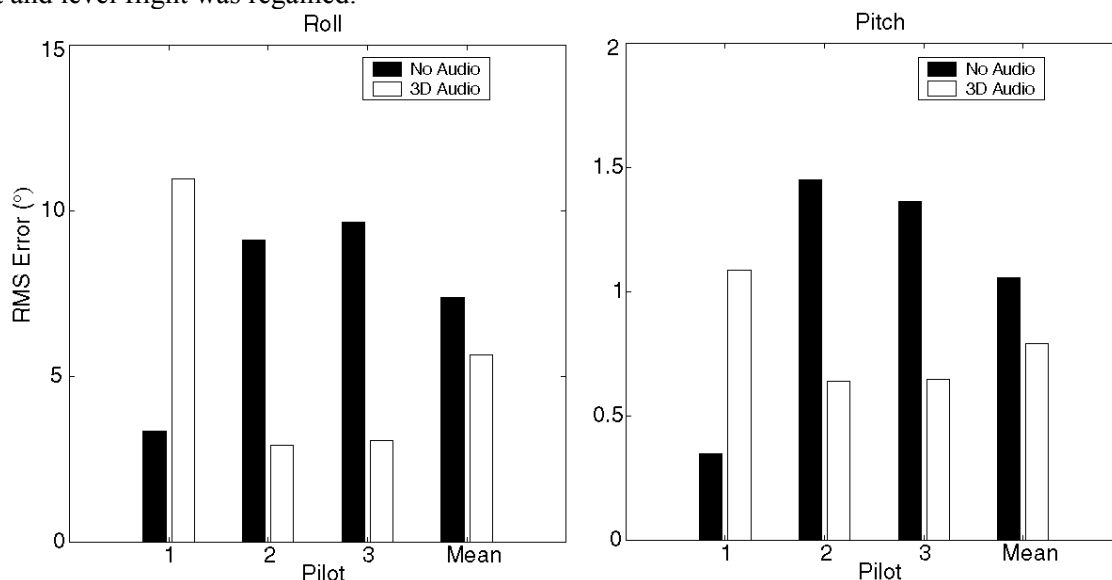


Figure 6: RMS errors between the ideal path and the actual path in the task requiring the subjects to maintain straight and level unaccelerated flight. The black bars represent the data in the no-audio condition and the white bars represent the data when the auditory artificial horizon display was provided.

As can be seen in Figure 7, the flight paths in the no-audio condition appear to be smooth relative to the flight paths in the 3D audio condition in roll, pitch, and heading – evidence that the pilots initiated very few control inputs in the no-audio condition. The surprisingly good performance for Pilot 1 in this no-audio condition, achieved with few inputs, further supports the spurious nature of the data from this condition. It is important to note that such results artificially reduce the difference between performance in the two display conditions and serves to underestimate the true benefit of the audio horizon display.

The right-hand panels show the deviation from straight and level flight when the artificial auditory horizon cue was provided. In both pitch and roll, Pilots 2 and 3 deviated little from the ideal path and never reached the boundary of safe travel. Pilot 1 reached this boundary only once in each dimension. Performance in the heading was also relatively good for Pilots 2 and 3. These results suggest that this display could be used effectively by pilots as an attitude indicator.

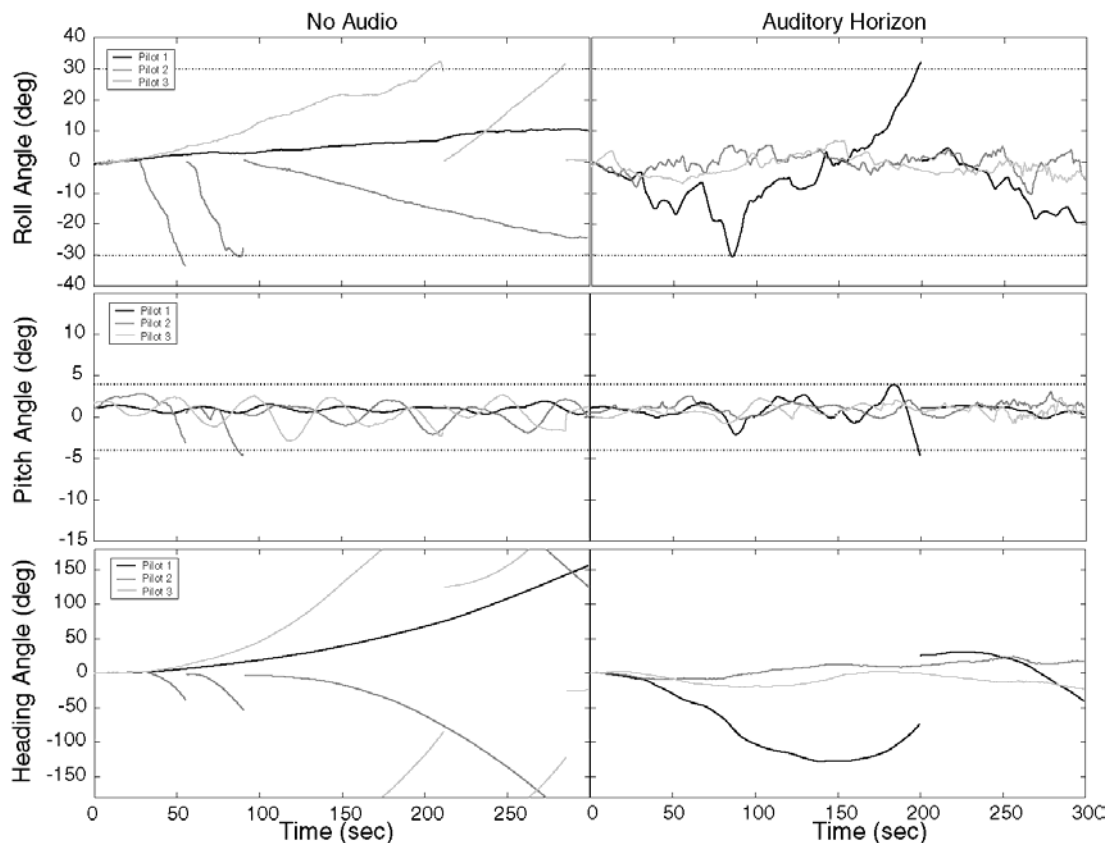


Figure 7: Flight paths travelled during the straight and level task. The left-hand panels depict the data from the trials when no audio cue was provided, and the right-hand panels depict the data when the artificial auditory horizon display was provided. The flight data are broken down into roll (top row), pitch (middle row) and heading (bottom row).

7.0 SUMMARY AND CONCLUSIONS

Overall, the results from these four experiments indicate that spatial audio displays may be useful as a means of providing information to a pilot about events in the relevant airspace and the status of the aircraft. This appears to be particularly true when displays and cues accessing other modalities are missing, obscured, or ambiguous.

The localization task served as a baseline measure of both system performance and the pilots' spatial hearing ability. However, spatial audio cues of this type may be used to provide information to the pilot regarding nearby air traffic and other obstacles. Indeed, in both basic laboratory experiments and immersive synthetic environment experiments, 3D audio has been demonstrated as an effective means of reducing visual

target acquisition times by providing an audio cue that is co-located with the visual target [14,15]. However, it appears that the spatialization in such a display must be referenced to the pilot's head in order to be effective. That is, the current experiment demonstrated that a plane-referenced display (the GAINR-coupled display) resulted in larger localization errors overall and front/back ambiguities that made it more difficult for the pilot to determine the true location of the stimulus. Nevertheless, the headtracker-coupled 3D audio display still must be improved if it is to become a system upon which the pilots may truly rely. The headtracking system exhibited degraded performance over time, and thus influenced localization performance negatively. In addition, localization in the U/D dimension with 3D audio was always near chance, independent of the spatialization reference. Although it may be sufficient to cue the pilot to a general location, the ambiguity of target elevation may prove to be unacceptable to pilots. Improving the fidelity and reliability of the spectral cues (which are believed to mediate localization in this dimension) by employing better head-related transfer functions, or indeed the pilot's own head-related transfer functions, would likely lead to much better performance overall, and greater pilot acceptance.

The potential to use 3D audio cues as navigation aids has also been demonstrated. Although performance with 3D audio never reached that achieved with a verbal cue and HSI (a display system in use throughout aviation today), it is nevertheless the case that 3D audio may be used in addition to such a display, or in place of this display in the event that the pilot does not have access to the HSI to get to an approximate heading, after which other cues (e.g., out-the-window visual cues) may be employed. As in the earlier task, a 3D audio navigation aid is most useful when the sounds are displayed with respect to the orientation of the pilot's head rather than with respect to the aircraft attitude and position. Additional cues for distance, such as a speech distance cue based on vocal effort [16], would be a useful addition if the task is to navigate to specific waypoints.

The experiments have also shown that the pilots are able to use the auditory artificial horizon display for detecting attitude changes in the aircraft. It is clear that pilots were, in general, able to identify the change of attitude that took place on a given trial, and could use this information to maintain straight and level flight when no visual cues were available. This display may prove to be particularly useful for mitigating the effects of spatial disorientation by notifying the pilot that an unusual or undesirable attitude has been assumed by the aircraft. However, in both cases, the parameters of the display may need to be adjusted, or supplemented in some way in order to reduce the size of attitude change required for detection and identification.

One issue that is of great concern with displays in general, and auditory displays in particular, is pilot acceptance. The audio horizon display allows pilots to effectively choose their own stimulus – that is, select their own music. Thus, acceptance of this display was high. Because it was not perceived as a disturbance or annoyance, it could remain on continuously and effectively provide ongoing information about the attitude of the aircraft. It is important to note, however, that because the spectral content of the music varies across individual pieces of music, the display could be rendered more or less useful, or perhaps even misleading, for pitch discrimination. Additional work must be conducted to overcome this potential shortcoming. Redundant displays, such as verbal cues to indicate the crossing of a safe-flight boundary, may prove to be useful.

Although the current displays examined in these four experiments may not lead to performance in all tasks and in all situations at the same level as can be achieved with standard visual instruments and out-the-window views, the potential for 3D audio to be a useful addition to cockpit display systems is clear. 3D audio may be used in conjunction with visual displays to redundantly display critical information to the pilot, or it may be used when the standard visual displays are unavailable or unreliable. The utility of a display that can capture attention at all times and from all locations without putting a strain on the already overworked visual system cannot be overestimated. Although there are many aspects of the current display system that must be further refined, it is nevertheless the case that significant advancements have been made in this study, and it is clear that 3D audio can be a part of an overall display system to support pilot situation awareness in flight environments.

8.0 REFERENCES

- [1] Veltman, J. A., Oving, A. B., & Bronkhorst, A. W. (2004). 3-D Audio in the fighter cockpit improves task performance. *International Journal of Aviation Psychology*, 14, 239-256.
- [2] Bronkhorst, A. W., Veltman, J. A., & Van Breda, L. (1996). Application of a three-dimensional auditory display in a flight task. *Human Factors*, 38, 23-33.
- [3] McKinley R. L., Ericson M. A., & D'Angelo W. R. (1994). 3-dimensional auditory displays: development, applications, and performance. *Aviation, Space, and Environmental Medicine*, 65, A31-38.
- [4] McKinley, R. L., & Ericson, M. A. (1997). Flight Demonstration of a 3D Auditory Display. In R. H. Gilkey and T. R. Anderson (Eds.) *Binaural Hearing in Real and Virtual Environments*. (pp. 683-699). New Jersey: Lawrence Erlbaum Associates.
- [5] Begault, D. R. (1993). Head-up auditory displays for traffic collision avoidance systems advisories: A preliminary investigation. *Human Factors*, 35, 707-717.
- [6] Foyle, D. C., Andre, A. D., McCann, R. S., Wenzel, E. M., Begault, D. R., & Battiste, V. (1996). Taxiway navigation and situation awareness (T-NASA) system: Problem, design, philosophy, and description of an integrated display suite for low-visibility airport surface operations. *SAE Transactions: Journal of Aerospace*, 105, 1411-1418.
- [7] Begault, D.R., & Pittman, M. T. (1996). Three-dimensional audio versus head-down traffic alert and collision avoidance system displays. *International Journal of Aviation Psychology*, 6, 79-93.
- [8] Simpson, B. D., Brungart, D. S., Gilkey, R. H., Cowgill, J. L., Dallman, R. C., Green, R. F., Youngblood, K. Y., & Moore, T. J. (2004). 3D audio cueing for target identification in a simulated flight task. *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting*, New Orleans, pp.1836-1840.
- [9] Wenzel, E. M., Miller, J. D., & Abel, J. S. (2000). Sound Lab: A real-time, software-based system for the study of spatial hearing. *Proceedings of the 108th Convention of the Audio Engineering Society. Paris* New York: Audio Engineering Society.
- [10] Kistler, D. J., & Wightman, F. L. (1992). A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. *Journal of the Acoustical Society of America*, 91, 1637-1647.
- [11] Wightman, F. L., and Kistler, D. J. (1989b). Headphone simulation of free-field listening. II: psychophysical validation. *Journal of the Acoustical Society of America*, 85, 868-878.
- [12] Wenzel, E. M., Arruda, M., Kistler, D. J., & Wightman, F. L. (1993). Localization using nonindividualized head-related transfer functions. *Journal of the Acoustical Society of America*, 94, 111-123.

- [13] Lyons TJ, Gillingham KK, Teas DC, Ercoline WR, Oakley C. (1990). The effects of acoustic orientation cues on instrument flight performance in a flight simulator. *Aviation, Space, and Environmental Medicine*, 61, 699-706.
- [14] Perrott, D. R., Saberi, K., Brown, K., & Strybel, T. Z. (1990). Auditory psychomotor coordination and visual search behaviour. *Perception and Psychophysics*, 48, 214-226.
- [15] Bolia, R. S., D'Angelo, W. R., & McKinley, R. L. (1999). Aurally aided visual search in three-dimensional space. *Human Factors*, 41, 664-669.
- [16] Brungart, D. S. (2000). Speech-Based Distance Cueing in Virtual Audio Displays. *Proceedings of the Human Factors and Ergonomics Society 44th Annual Meeting*, San Diego, pp.3714-3717.



An Algorithm for the Accurate Localization of Sounds

Justin A. MacDonald

Army Research Laboratory
Human Research and Engineering Directorate
Attn: AMSRD-ARL-HR-SD
Aberdeen Proving Ground, MD 21005
USA

jmacdonald@arl.army.mil

ABSTRACT

A computer-based algorithm that localizes sounds in near-real time has been developed. The algorithm takes input from two microphones and estimates the position of the sound source relative to the microphone array. The algorithm requires no a priori knowledge of the stimuli to be localized. The accuracy of the algorithm was tested using binaural recordings from a pair of microphones mounted in the ear canals of an acoustic mannequin. Sounds were played at 5 degree steps around the mannequin and the outputs were recorded at the entrance to each ear canal. These recordings were fed into the algorithm that estimated the location of the incoming sound on the horizontal plane. The algorithm utilizes a Head-Related Transfer Function (HRTF) to estimate the location of incoming sound stimuli. Although the HRTF of the acoustic mannequin was used, any HRTF can be inserted into the algorithm, allowing for predictions about individual performance differences. The results of this effort have been highly encouraging: the algorithm was able to identify accurately the location of a variety of sounds, committing an average of 2.9 degrees of unsigned localization error. Better than chance performance was found in noisy conditions of up to a -10 dB signal-to-noise ratio. The initial purpose of this algorithm is to predict the localization performance afforded by different types of combat helmets. Current and future encapsulating helmet designs are likely to impede localization performance, and an accurate localization model would be an invaluable tool in the helmet selection process. Adapting the model for use as a highly accurate machine-based localizer is an additional goal of this line of research. Applications for this technology include target tracking on unmanned vehicles, sniper detection, and machine-assisted sound localization.

1.0 INTRODUCTION

Many protective military ensembles include partially- or fully-encapsulating headgear which has a negative effect on sound detection, localization, and auditory distance estimation. This problem has prompted the Army to search for a model of human sound localization that can predict the effects of different headgear ensembles on the auditory localization ability of the soldier.

Constructing an accurate model of human sound localization is a complex task. When the listener estimates the location of an incoming sound many factors other than the stimulus itself can affect the judgment. Listeners are likely to incorporate visual information into their interpretation of the sound, which can alter the listener's perception of the auditory spatial information. Moreover, familiarity with the environment and prior knowledge of the type of sound and its amplitude and frequency characteristics can help to improve the

MacDonald, J.A. (2005) An Algorithm for the Accurate Localization of Sounds. In *New Directions for Improving Audio Effectiveness* (pp. 28-1 – 28-10). Meeting Proceedings RTO-MP-HFM-123, Paper 28. Neuilly-sur-Seine, France: RTO. Available from: <http://www.rto.nato.int/abstracts.aps>.

localization performance of the listener. In addition to a relatively accurate method to estimate the location of sound sources, any predictive model of human performance should incorporate both a learning mechanism as well as a component to model the interaction between visual and auditory information about the possible location of a sound source.

The initial phase of model construction has focused upon the development of an algorithm to produce location estimates from inputs to each of the ears. Development of this algorithm serves two purposes. First, the algorithm is intended to function as the basis for a more complete model of human sound localization that mirrors processes used by human listeners to produce location judgments. Second, the algorithm can be used as the basis of a computer-based “sound localizer” aiding the user in the accurate determination of sound source position. Such a system would have a number of commercial and military applications, such as inclusion into the navigation system for autonomous robotic vehicles or as an aid to the soldier to improve sound localization on the battlefield. Such a system would benefit from a fast and highly accurate localizing algorithm, ideally considerably more accurate than humans. Therefore, the initial goal of this project was to develop a localization algorithm that maximizes localization accuracy while minimizing computational requirements. The algorithm should maintain reasonable functionality in noisy environments, and provide accurate location estimates without utilizing any prior knowledge of the sound stimulus including its typical level and frequency characteristics.

1.1 Development of the Localization Algorithm

Several previous attempts have been made to develop sound localization algorithms. Berdugo, Doron, Rosenhouse, and Azhari [1] utilized an array of seven microphones to estimate the azimuth and elevation of a 20-second speech stimulus played from a loudspeaker in a quiet environment. Estimates were computed using the differences in time-of-arrival observed at each of the seven sensors. The mean unsigned localization error was found to be approximately 5° . While this system was found to be relatively accurate, it is less than ideal for the current application. First, testing of the algorithm was conducted using stimuli that were 20 seconds in duration, which should provide the algorithm with considerable data with which to make a localization estimate. Considering that humans are quite able to localize sounds that are less than 500 ms, any algorithm used as a basis for a model of human sound localization must generate accurate estimates using considerably less data. Secondly, the use of seven sensors in the localization system is clearly inappropriate for modelling a two-sensor system (the human). Viera and Almeida [2] constructed a localization system with two sensors that estimated the location of the sound based upon the time delay between the sensors. Testing required the system to localize sounds located at 0° elevation and between $+60$ and -60 degrees azimuth. The system averaged approximately 9° of unsigned localization error. While this algorithm utilizes only two sensors and is therefore appropriate as a model of human localization, it is less accurate than a human observer. Moreover, any two-sensor system that localizes sounds based solely upon the time delay will be susceptible to frequent front/back confusions. Any given time delay between the ears will restrict the subset of possible locations to the set of points that lie upon the surface of a cone extending outward from one of the sensors [3]. Even if source locations are restricted to those on the horizontal plane and the source distance is held constant, time delay information will not uniquely determine the location of the stimulus. Additional information about the frequency characteristics of the incoming sound must be utilized to resolve this ambiguity. Fortunately, the head and torso of the listener provide these cues: the effect of the head and torso upon an incoming sound wave depends upon the direction of the waveform. The catalogue of effects of the head and torso upon incoming sound sources makes up the Head-Related Transfer Function (HRTF; see [3], [4]) of the listener. These frequency differences can be used to eliminate the front/back ambiguity.

For this reason the localization algorithm developed by the author was based upon the HRTF, which incorporates both time delay and frequency cues to the location of a sound source. The algorithm was implemented using the HRTF of the Knowles Electronics Mannequin for Acoustics Research (KEMAR), although the HRTF of any listener could have been used. Two possible methodologies were considered during the development of the localization algorithm: the Inverse Localizer and the Cross-Channel Localizer. Both utilize the HRTF of the KEMAR and require no previous knowledge of the sound stimulus to be localized.

1.2 The Inverse Localizer

Consider a sound that originates directly in front of the right ear (at $+90^\circ$) on the horizontal plane. The waveform that reaches each of the ears will be subject to the head and torso effects summarized as the HRTF when the sound originates at $+90^\circ$ azimuth and 0° elevation. Assume that the HRTF at $+90^\circ$ is represented in the time domain as a pair of Finite Impulse Response (FIR) digital filters, denoted by $H_{Left}^{(+90,0)}$ and $H_{Right}^{(+90,0)}$. Suppose that microphones were placed in each of the ears to record the incoming waveform. Let R_{Left} and R_{Right} be digital recordings of the sound obtained at the left and right ears, respectively. R_{Right} could be filtered by the inverse of $H_{Right}^{(+90,0)}$ to obtain the original sound before its alteration by the head and torso. Filtering R_{Left} by the inverse of $H_{Left}^{(+90,0)}$ would result in a copy of the original unaltered stimulus identical to that obtained from the right ear. However, if R_{Left} and R_{Right} were not filtered by the above functions but instead by the inverse of $H_{Left}^{(-90,0)}$ and $H_{Right}^{(-90,0)}$, respectively, the original stimulus would not result in either case. In fact, this operation would lead to considerably different waveforms.

This simple idea suggests a method by which the inverse of the HRTF could be used to estimate the location of sound sources. Consider a sound that originates from direction θ and elevation ϕ in relation to a point P , where $-180^\circ < \theta, \phi \leq +180^\circ$. Imagine that the centre of the head of the listener is at P . The task of the inverse localizer is to provide the estimates $\hat{\theta}$ and $\hat{\phi}$ based upon the recordings R_{Left} and R_{Right} . In this case, the inverse localizer estimates the location of the sound source as follows:

$$\min_{\hat{\theta}, \hat{\phi}} \sum \left(R_{Left} * \left[H_{Left}^{(\hat{\theta}, \hat{\phi})} \right]^{-1} - R_{Right} * \left[H_{Right}^{(\hat{\theta}, \hat{\phi})} \right]^{-1} \right)^2, \quad (1)$$

where $*$ is the convolution operator and $\left[H_{Left}^{(\hat{\theta}, \hat{\phi})} \right]^{-1}$ and $\left[H_{Right}^{(\hat{\theta}, \hat{\phi})} \right]^{-1}$ are the inverses of $H_{Left}^{(\hat{\theta}, \hat{\phi})}$ and $H_{Right}^{(\hat{\theta}, \hat{\phi})}$, respectively.

This procedure is not without its difficulties, however, most of which arise when computing the inverses of $H_{Left}^{(\hat{\theta}, \hat{\phi})}$ and $H_{Right}^{(\hat{\theta}, \hat{\phi})}$. These filters include both the time delay and frequency effects caused by the head. Time delays are implemented as differences in the onsets of the filters for the left and right ears, and inverting the filters requires compensating for this delay. More importantly, the inversion of an FIR filter can lead to unpredictable results. Constructing the minimum-phase inverse results in a filter whose magnitude is only an approximate inverse of the original [5]. In addition, the phase response must also be inverted; this can be accomplished using the procedure outlined in Greenfield & Hawksford [6]. This procedure yields an inverse

filter that is roughly three times larger than the original. For these reasons, the development of the inverse localizer was postponed in favour of another methodology that does not require the calculation of inverse filters.

1.3 The Cross-Channel Localizer

As before, the task of the localization algorithm is to provide the estimates $\hat{\theta}$ and $\hat{\phi}$ based upon the recordings R_{Left} and R_{Right} . The cross-channel localizer chooses $(\hat{\theta}, \hat{\phi})$ as follows:

$$\min_{\hat{\theta}, \hat{\phi}} \sum \left(R_{Left} * H_{Right}^{(\hat{\theta}, \hat{\phi})} - R_{Right} * H_{Left}^{(\hat{\theta}, \hat{\phi})} \right)^2. \quad (2)$$

In words, the cross-channel localizer filters each of the recordings by the HRTF of the opposite ear and compares the resulting waveforms. To understand the reasoning behind this method, let $O^{(\theta, \phi)}$ be a digital recording of the sound originating from (θ, ϕ) recorded at P with the listener absent. Then $R_{Left} \approx O^{(\theta, \phi)} * H_{Left}^{(\theta, \phi)}$, and $R_{Right} \approx O^{(\theta, \phi)} * H_{Right}^{(\theta, \phi)}$. If the recording from the left ear is convolved with the HRTF of the right ear, then

$$R_{Left} * H_{Right}^{(\theta, \phi)} \approx \left(O^{(\theta, \phi)} * H_{Left}^{(\theta, \phi)} \right) * H_{Right}^{(\theta, \phi)} = \left(O^{(\theta, \phi)} * H_{Right}^{(\theta, \phi)} \right) * H_{Left}^{(\theta, \phi)} \approx R_{Right} * H_{Left}^{(\theta, \phi)}. \quad (3)$$

This follows from the commutability and transitivity of the convolution operator. The cross-channel localizer takes advantage of this relation by choosing the values of $(\hat{\theta}, \hat{\phi})$ so that the squared differences between the leftmost and rightmost terms in Equation 3 are minimized.

This method is preferable to the inverse localizer for several reasons: first, the construction of inverse filters is unnecessary; only the original HRTF is required to localize sounds. Second, the FIR filter that represents the HRTF is approximately one-third the size of an accurate inverse filter. Minimizing the size of the filters used in convolution will have a tremendous impact on the computational resources required. Accordingly, an initial implementation of the cross-channel localizer was developed and subsequently tested in signal-to-noise (S/N) ratios between 40 and -40 dB to assess its performance in noisy environments. As this is only a preliminary test of the model, the positions of all sound sources were limited to the horizontal plane.

2.0 EXPERIMENT METHOD

2.1 Stimuli. Eight naturally-occurring sounds were chosen as the test signals. They included the sounds of breaking glass, a barking dog, a closing door, the insertion of an M-16 magazine, a cough, dripping water, a camera in operation, and machine gun fire. Sounds ranged from 400 to 600 ms in duration and were stored in a 16-bit Microsoft WAV format with a sampling rate of 44.1 kHz.

2.2 Stimulus presentation apparatus. Stimuli were presented using the Army Research Laboratory Human Research and Engineering Directorate's RoboArm 360 system (see Figure 1 below). This system consists of a speaker attached to a computer-controlled robotic arm. The stimuli were output through a Tucker-Davis Technologies (TDT) System II DD1 D/A converter, amplified using a TDT System 3 SA1

amplifier, and presented from a GF0876 loudspeaker (CUI, Inc.) at the end of the robotic arm. Stimuli were presented at approximately 75 dB (A) measured one meter from the loudspeaker. The arm positioned the loudspeaker at 5° intervals around the KEMAR (a total of 72 positions). The loudspeaker was located 1 meter from the centre of the head of the KEMAR and at 0° elevation for all stimulus presentations.

2.3 Recording apparatus. EM-125 miniature electret microphones (Primo Microphones, Inc.) were mounted in foam inserts in the artificial pinnae of the KEMAR. Inputs to the microphones were amplified by a TDT System 3 MA3 microphone amplifier before being sent to a TDT System II DD1 A/D converter. The digital output of the DD1 was sent to a computer for storage in a 44.1 kHz, 16-bit Microsoft WAV format. A total of 576 recordings were made, one for each position/sound combination. These WAV files served as the sounds to be localized by the cross-channel localizer. Each recording was filtered to remove the frequency effects introduced by the recording system.



Figure 1 – The RoboArm 360 system. The computer-controlled robotic arm can place the loudspeaker to 1° precision.

2.4 HRTF measurement. The HRTF of the KEMAR was measured using the same presentation and recording apparatus detailed above. Maximum-Length Sequence (MLS; see [8]) stimuli were presented at 5° intervals around the head of the KEMAR and the resulting waveforms were recorded at each of the ears.

An Algorithm for the Accurate Localization of Sounds

These recordings were digitized and processed to determine the HRTF of the KEMAR at each location. Each HRTF was stored as a 256-tap FIR filter and the frequency effects introduced by the recording system were removed.

2.5 Localization task. The cross-channel localizer applied the measured HRTF of the KEMAR to the binaural recordings to estimate the location of each sound to 5° precision. The algorithm estimated the location of the sound source using a version of Equation 2 that was implemented in MATLAB. A random sample of Gaussian noise was added to each recording to obtain signal-to-noise ratios ranging from 40 to -40 dB in 10 dB increments. The localizer was required to localize each recording 10 times; a different sample of Gaussian noise was added on each attempt. This resulted in a total of 51,840 localization attempts (9 S/N ratios X 576 recordings X 10 trials each).

3.0 RESULTS

The mean unsigned error for each S/N ratio is shown in Figure 2. The means were obtained by averaging over all stimuli and locations within a given S/N ratio. Not surprisingly, mean error increased as the S/N ratio decreased. The cross-channel localizer maintained an error rate of less than 5° for S/N ratios of 40, 30, and 20 dB, and even performed well above chance performance at -10 dB.

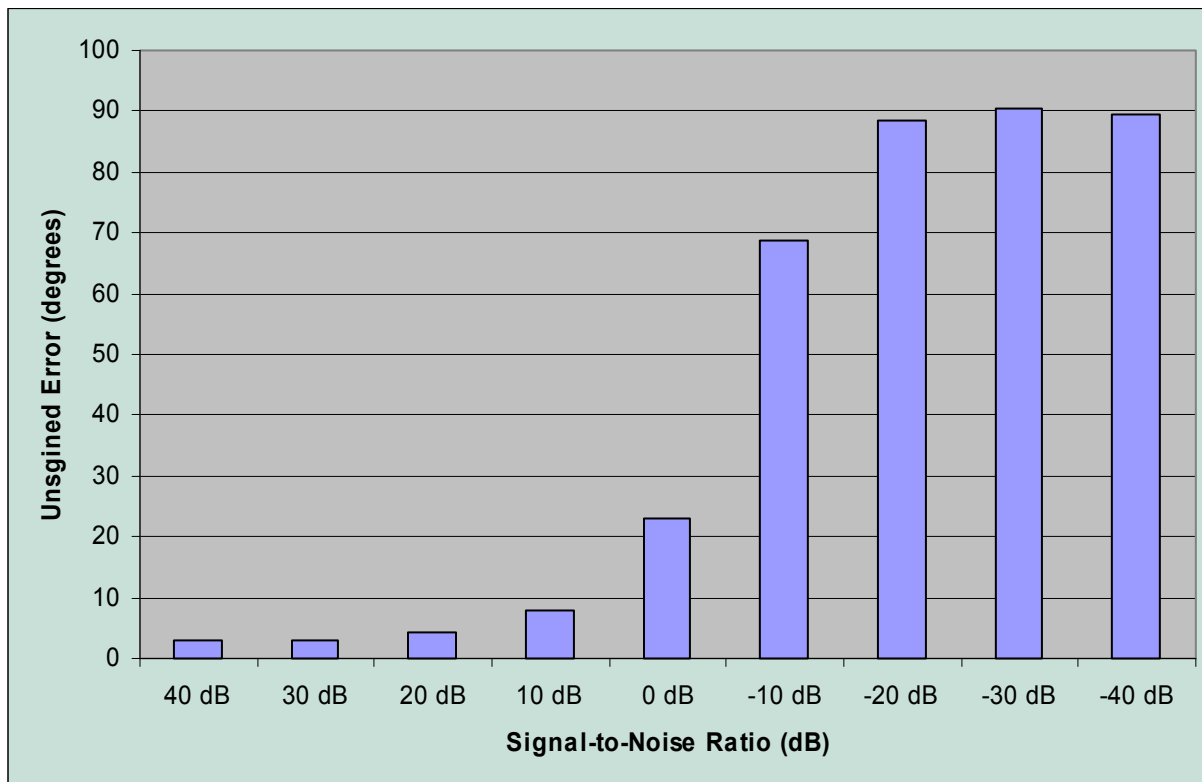


Figure 2 – Mean Unsigned Error by S/N Ratio

The proportions of front/back reversals are shown in Figure 3. For the purposes of this experiment, a reversal occurred when the estimated and actual locations of a sound were on opposite sides of the interaural axis. Sounds located at $+90^\circ$ and -90° were omitted from this analysis for obvious reasons. The localizer correctly resolved the front/back ambiguity for more than 95% of trials at 40 and 30 dB. Performance did not deteriorate to chance level until the S/N ratio reached -20 dB.

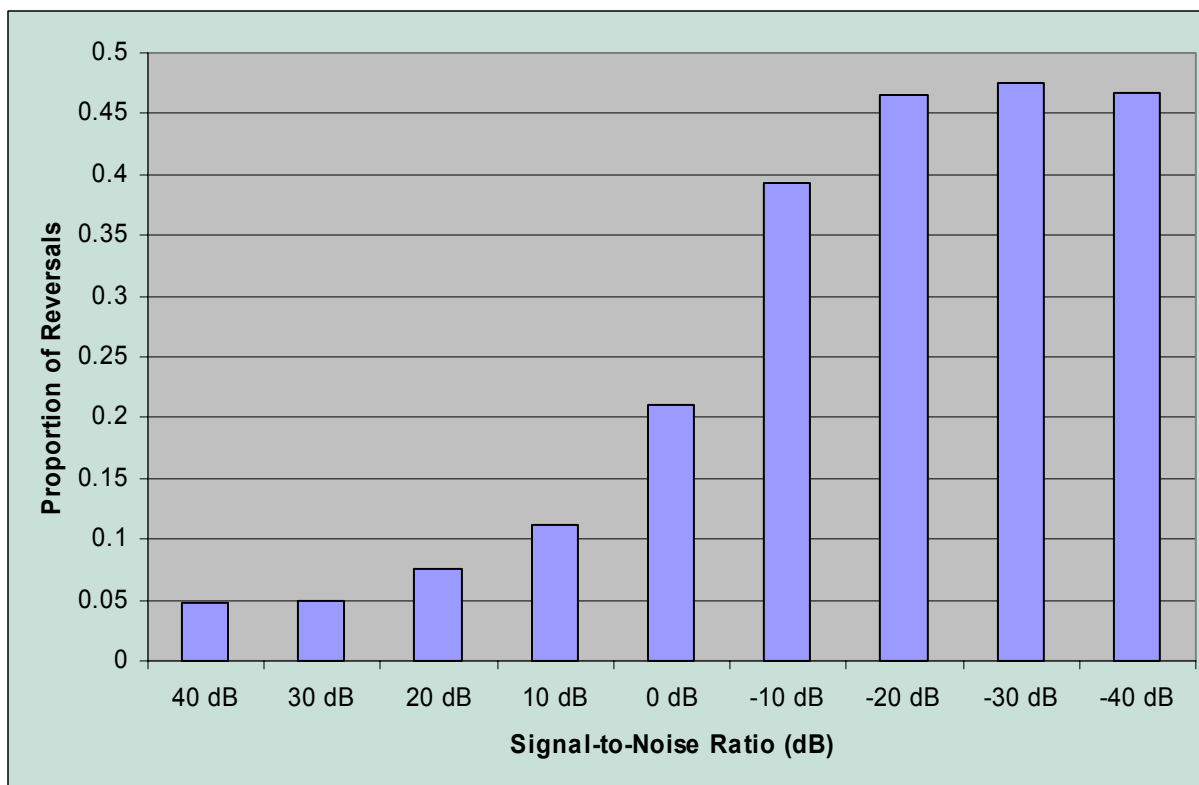


Figure 3 – Proportion of Reversals by S/N Ratio

Finally, the mean error corrected for front/back reversals is illustrated in Figure 4. If the localizer committed a reversal on a trial, the estimated location was flipped across the interaural axis and the unsigned error was computed using this corrected estimate. This is the performance that could be expected if the domain of possible sound source locations was restricted to those locations in the front hemisphere. The mean corrected error was below one degree for S/N ratios of 10 dB and higher, and fair performance was also obtained at -10 dB.

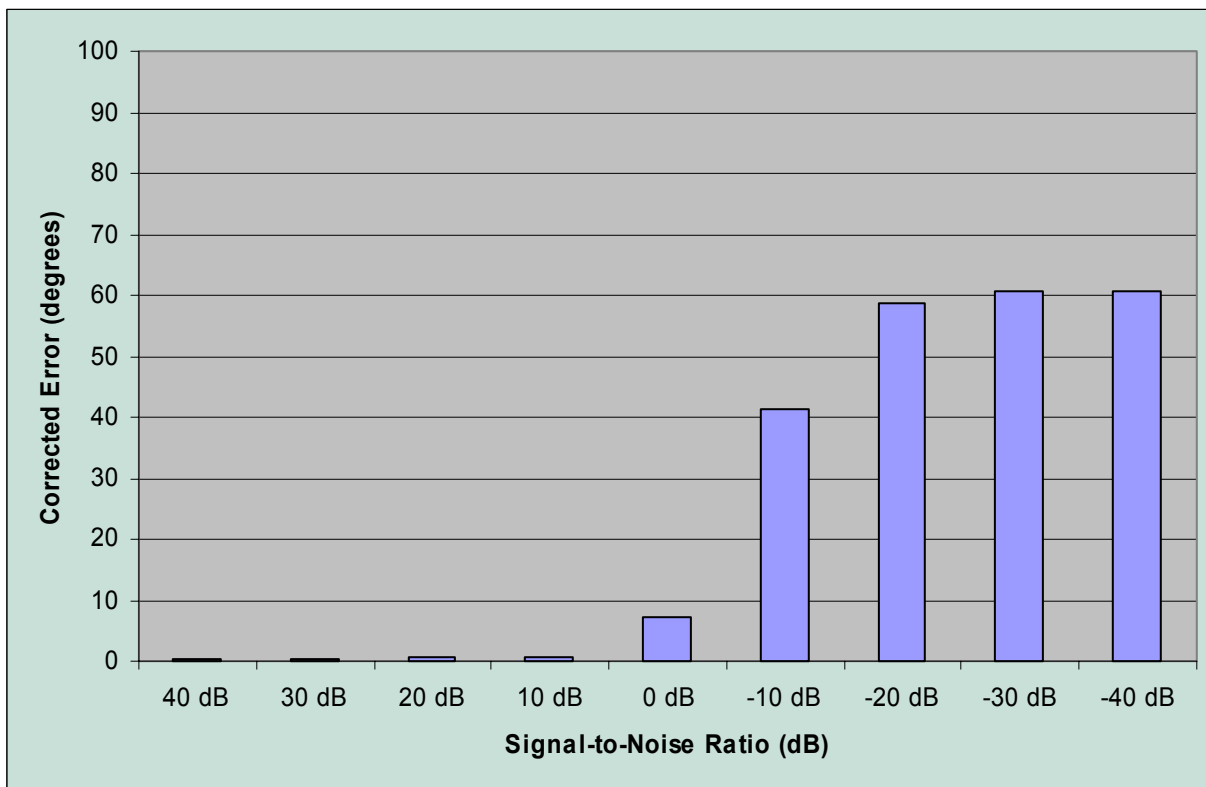


Figure 4 – Mean Corrected Error by S/N Ratio

4.0 DISCUSSION

The cross-channel localization algorithm clearly provides an effective means to estimate accurately the direction of a sound source. The localizer utilizes input from only two sensors and requires only moderate computational resources. The algorithm maintains accuracy under noisy conditions, and indeed performs at a better-than-chance level at S/N ratios of -10 dB and higher. This high level of performance is obtained without previous knowledge of the sound stimuli, instead utilizing relying solely upon HRTFs. The inclusion of both time delay information as well as frequency effects related to the head and torso serves to minimize the proportion of reversals. In fact, a localizer relying upon the time delay information alone would likely commit reversals on 50% of trials in all conditions. In contrast, the cross-channel localizer maintained reversal rates of less than 10% for S/N ratios 20 dB and higher.

Considered solely as part of a computer-based localization system, the cross-channel localizer shows considerable promise for military and commercial applications. The algorithm would be ideal for inclusion into a system to aid human sound localization. The algorithm could also be used for navigation or target tracking purposes, or possibly sniper detection. Future work will continue the testing of the algorithm with sounds at positive and negative elevations. In addition, the effect of adding input from a third microphone to the algorithm will be investigated. Three sensors may help to minimize front-back reversals as well as improve elevation discrimination.

The cross-channel localizer described in this paper should perform well as the basis of a model of human sound localization. Preliminary testing suggests that it performs similarly to humans when the accuracy of the localizer is reduced by adding noise to the stimulus. The algorithm is also well-suited to predict individual differences: performance can be compared when using the HRTFs of several different listeners. Future modelling work will focus upon predicting the localization performance afforded by different Army headgear designs, as well as the inclusion of a learning mechanism in the model. The eventual goal is a model that can predict performance in both sound localization and auditory distance estimation tasks in various environments and for variety of headgear combinations.

5.0 REFERENCES

- [1] Berdugo, B., Doron, M. A., Rosenhouse, J., & Azhari, H. (1999). On direction finding of an emitting source from time delays. *Journal of the Acoustical Society of America*, 105, 3355-3363.
- [2] Viera, J. & Almeida, L. (2003). A sound localizer robust to reverberation. *Proceedings of the 115th Convention of the Audio Engineers Society*, Paper 5973.
- [3] Blauert, J. (1989). Spatial hearing. MIT Press, Cambridge, Massachusetts.
- [4] Wightman, F. L. & Kistler, D. J. (1989). Headphone simulation of free-field listening. I: Stimulus synthesis. *Journal of the Acoustical Society of America*, 85, 858-867.
- [5] Oppenheim, A. V. Schaefer, R. W., & Buck, J. R. (1999). Discrete-time Signal Processing. Prentice Hall Publishers, Upper Saddle River, New Jersey.
- [6] Greenfield, R. & Hawksford, M. O. (1991). Efficient filter design for loudspeaker equalization. *Journal of the Audio Engineering Society*, 39, 739 – 751.
- [7] Rife, D. D. and Vanderkooy, J. (1989). Transfer-Function Measurement with Maximum-Length Sequences. *Journal of the Audio Engineering Society*, 37, 419-444.

Localization Performance of Real and Virtual Sound Sources

M.Sc.E.E. Jan Abildgaard Pedersen

AM3D A/S
Riihimäkivej 6
DK-9200 Aalborg
Denmark

M.Sc.E.E. Torben Jørgensen

Terma A/S
Hovmarken 4
DK-8520 Lystrup
Denmark

E-mail: jap@am3d.com / toj@terma.dk

ABSTRACT

This paper describes how a 3D-Audio system for use in fighter aircrafts was evaluated in an experiment, by comparing localization performance between real and virtual sound sources. Virtual sound sources from 58 selected directions were evaluated, while 16 of these directions were also evaluated using real sound sources, i.e. loudspeakers. 13 pilots from the Royal Danish Air Force and 13 civil persons were used in the test. The localization performance was split into a constant and a stochastic difference between the perceived direction and the desired direction (stimulus). The constant difference is a localization offset and the stochastic difference is a measure for the localization uncertainty. Stimuli length of both 250 ms and 2 s enabled investigation of the importance of head movements, i.e. using head tracking. Real and virtual sound sources could be located with an uncertainty of 10° and 14° degrees for azimuth while the uncertainty for elevation was 12° and 24° (real and virtual sound sources). No significant localization offset was found for azimuth, while an average offset for elevation of 3° – 6° degrees was found using long stimuli. A significant difference between the localization offset obtained in different directions was found – especially for elevation, where the offset was found to have a strong correlation to the stimuli elevation.

1 INTRODUCTION

3D-Audio is used in fighter aircrafts to enhance the situational awareness. This includes 3D-Audio indication of the directions of detected approaching missiles as well as channel separation of several simultaneous sound signals, e.g. speech from ground control, wingman and alarm signals. The 3D-Audio system included headphone playback system, head tracker and a digital signal processor (DSP). This paper describes the design and the results of a psychoacoustic test, which was used to evaluate such a 3D-Audio system. In particular the localization performance was evaluated, i.e. the capability of positioning sound sources in 3D space at predetermined positions, which enables listeners to localize sound sources at desired positions with some uncertainty. The experiment focused only on the localization performance in relation to direction - distance was not considered. Separate work focusing on distance perception has been presented earlier by other authors [1], [2], [3].

Localization performance has been a subject for research for many years and a long list of authors has presented work including localization performance of real sound sources, phantom sound sources and

virtual sound sources [4], [5], [6], [7], [8], [9], [10], [11]. Many of these papers focus on the localization performance of virtual sound sources, i.e. binaural sound reproduction systems, where test subjects wear headphones and the signals fed to each ear has been recorded using a dummy head or has been processed by head-related transfer functions (HRTF) [5]. Either the recording or the processing will enable listeners to localize sound sources in full 3D as if the listener was present at the recording position or at the synthesized listening position. Head tracking enables the signals fed to the two ears to change accordingly to head movements, so that the perception of the virtual sound source remains while moving the head. The perception of the virtual sound source not only remains because of the head tracking system, it is improved dramatically [5], [10], [11]. Especially front-back ambiguity is almost solved in 3D-Audio systems, which includes a head tracker [11].

When evaluating the localization performance of 3D-Audio systems different strategies has been proposed. One way is to determine the difference between the direction of stimuli (desired direction) and the perceived direction. This will show how large the errors can be when trying to position a virtual sound source in 3D space, i.e. an absolute measure of the angle between the desired and observed position. In a given application this could be compared to the actual needs or requirements for localization performance. However an obvious reference would be real life localization performance, i.e. localization performance of real sound sources [9]. Most 3D-audio systems are implementing an approximation to a real sound field, which means that if the localization performance of the 3D-Audio system is approximately the same as for real sound sources, then the localization performance of the 3D-Audio system is optimal.

The experiment described in this paper included a listening test, where the same test subjects evaluated the localization performance of both a 3D-audio system and a setup of real sound sources. Care was taken to ensure that all conditions were equal for the two evaluations. The only difference was the fact that test subjects did wear headphones when they evaluated the 3D-Audio system, which was not the case when they evaluated the real sound sources.

2 STRATEGY FOR THE EXPERIMENT

2.1 Purpose of the experiment

The purpose of this experiment was to provide information about the localization performance of a 3D-Audio system in relation to usage in “Mission Critical” systems. Specifically the performance measures for systems, where alarm signals are given a direction corresponding to objects that need attention, were of interest. It was a desire to determine the localization performance in situations that resemble the real life use of an auditory display for “Mission Critical” systems. The localization performance for virtual sources was compared to the localization performance for real sound sources. In this experiment headphones were used for virtual sound sources and loudspeakers were used for real sound sources.

2.2 Psychoacoustic method

First the subjects heard one sound signal from one direction either via headphones (virtual sources) or via one loudspeaker (real sources) and were then asked to point to the direction, where they perceived the sound source, i.e. the single stimuli method. Simply pointing to the direction where they perceived the sound source to come from was an intuitive method, which should decrease the need for very long and intensive training sessions.

A laser pointer was mounted on the pointing device in such a way, that a light dot gave visual feedback to the subject to ensure, that the answered direction was indeed the direction, where they perceived the sound

to originate from. Different and independent random sequences of directions and sound source type (real or virtual) were used for each subject.

The main experiment was preceded by a structured training session, which introduced the test subjects to the concept of direction of sound sources as well as running test sessions using both virtual and real sound sources. The aim of the training sessions were to minimize the increased variance introduced by using subjects, who were naive in the task, methods and procedures used in the experiment.

Visibility of the real sound sources, i.e. loudspeakers, would have introduced a bias into the experiment, because subjects would only point to directions, where they could see a loudspeaker – not in directions in between. This bias was avoided by installation of an acoustically transparent curtain (side, top and bottom) surrounding the subjects in such a way that no loudspeakers were visible.

2.3 Localization Performance and definition of direction

The localization performance was split into a constant and a stochastic difference between the perceived direction and the desired direction, i.e. stimulus. The constant difference is a localization offset and the stochastic difference is a measure for the localization uncertainty. A given direction relative to the head of the listener was characterised by the two Euler angles: azimuth and elevation.

Azimuth defines a rotation about the z-axis in a coordinate system, where origo is the centre of the listeners head, z-axis pointing upward and y-axis pointing straight forward, i.e. x-axis pointing to the right hand side. Azimuth is 0 for a direction straight ahead, i.e. in the direction of the y-axis. Azimuth is positive when turning in the direction to the left (counter clockwise), e.g. azimuth is $+90^\circ$ for a direction directly to the left of the listener and 270° for a direction directly to the right of the listener.

Elevation is a rotation around the x-axis, where a positive angle means upward and a negative angle means downward relative to the horizontal plane for the listeners head, e.g. $+90^\circ$ is right above the listeners head and -90° is directly below the listener. Elevation is 0° in the horizontal plane.

2.4 Virtual and real sound sources

The perceived direction of a sound will be different from the desired/real direction of the presented sound. This is true both when using virtual and real sound sources. This means, that when judging the localization performance of a 3D-Audio system (virtual sources), one could compare the results to results obtained using real sound sources placed in the desired direction relative to the subject.

A large number of directions were used for virtual sound sources because the 3D-Audio system was the main focus of the experiment. Azimuth and elevation were considered to be two independent variables in the experiment, which meant that separate selection of values for azimuth and elevation were performed. At first 107 directions were considered for use in the experiment for virtual sources, however due to left/right symmetry of the basic set of directions this number was lowered to 58 asymmetric directions, which basically covered the same directions as the basic set. This decision was made to lower the time consumption of the experiment and because verification of left/right symmetry was not part of the purpose of this experiment.

For practical reasons, a significant lower number of real sound sources were used: 16 loudspeakers. These 16 directions for the real sources were a subset of the 58 directions used for virtual sources. A test session was either a loudspeaker session or a virtual source session, so that test subjects did not have to put on headphones or remove them during a test session.

2.5 Controlled parameters in the experiment

The controlled parameters of the experiment were:

- Azimuth angle
- Elevation angle
- Test subject
- Stimuli
- Sound source type: virtual or real

3 DESIGN OF THE EXPERIMENT

3.1 Stimuli

3.1.1 Spectral and temporal characteristics

White noise was selected as the basic stimuli. Both short and long bursts were used: 250 ms and 2 s. Short bursts of white noise were used to prevent subjects from being able to turn their head as part of localizing the sound sources. This stimulus was used to evaluate the static localization performance. Earlier work has proved that the head movement has a dramatic impact on the localization performance, i.e. the difference between static and dynamic sound localization performance [5], [10], [11], [13]. Dynamic localization performance, where then evaluated using the long noise bursts, where subjects had time to use head movements as part of the localization task. During the training sessions only long bursts were used. If a test session lasted longer than 15 min. a break was automatically inserted by the program, which controlled the test. After a break of approximately 5 min. the test session was continued.

Training sessions:

- Long bursts of white noise: 2.00 s.

Experiment sessions:

- Short bursts of white noise: 0.25 s.
- Long bursts of white noise: 2.00 s.
- Maximum 15 min. without breaks

The level of the stimuli was adjusted to $84 \text{ dB SPL(LIN)} = 75 \text{ dB SPL(A)}$ at the listening position while playing the noise signal through a loudspeaker. A dummy head, “Valdemar” [5], [6], [7], were placed on the listening position and used to ensure the same reproduction level using loudspeakers and headphones.

3.1.2 Directions of virtual and real sound sources

Azimuth and elevation had to be independent since they were regarded as two controlled parameters of the experiment. Basically it was decided to use 15 values for azimuth and 9 values for elevation. This corresponds to the 107 directions, because azimuth are undefined for the two elevation values $+90^\circ$ and -90° . The values for azimuth angle were $0^\circ, 24^\circ, 48^\circ, 72^\circ, 96^\circ, 120^\circ, 144^\circ, 168^\circ, 192^\circ, 216^\circ, 240^\circ, 264^\circ, 288^\circ, 312^\circ$ and 336° . The values for elevation angle were $-90^\circ, -66^\circ, -44^\circ, -22^\circ, 0^\circ, 22^\circ, 44^\circ, 66^\circ$ and 90° . These 107 directions were then reduced to 58 simply by removing either the left or the right direction in a couple, which were left/right symmetric. The used 58 direction for virtual sound sources is shown in figure 3.1.

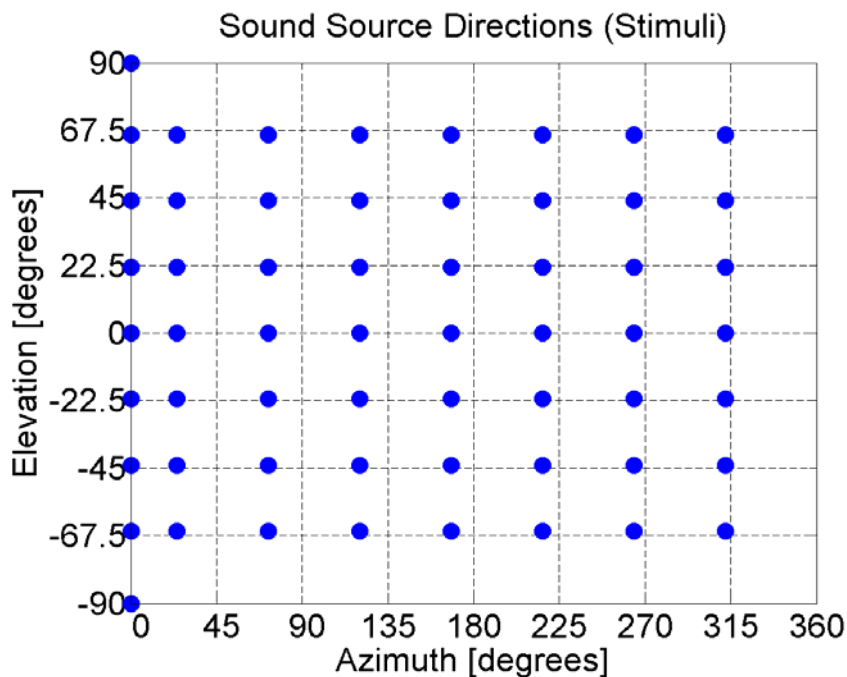


Figure 3.1: Directions for the 58 virtual sound sources without left/right symmetry.

This means that the number of values used for azimuth reduced to 8:

- Azimuth: $0^\circ, 24^\circ, 72^\circ, 120^\circ, 168^\circ, 216^\circ, 264^\circ, 312^\circ$
- Elevation: $-90^\circ, -66^\circ, -44^\circ, -22^\circ, 0^\circ, 22^\circ, 44^\circ, 66^\circ, 90^\circ$

16 of the 58 directions were then selected for real sound sources, i.e. loudspeakers were mounted in these 16 directions. Figure 3.2 shows these 16 directions in a similar way as in figure 3.1. Finally table 3.1 gives the complete list of the 58 directions including “Direction ID”, elevation angles, azimuth angles and an indication of “Real Source”, which means that such a direction were used both as a virtual and a real sound source.

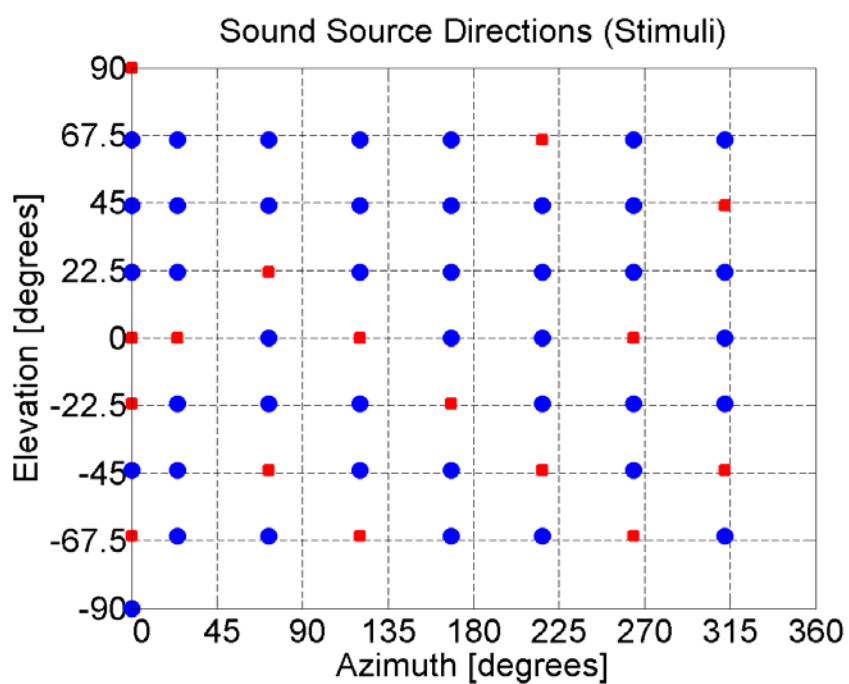


Figure 3.2: Directions for the 16 real sound sources (loudspeakers) shown as red squares, which are a subset of the 58 directions used for virtual sources (blue disks).

Table 3.1: Direction of all 58 virtual and 16 real sound sources.

Direction ID	Elevation	Azimuth	Real Source
1	-90°	0°	-
2	-66°	0°	X
3	-66°	24°	-
4	-66°	72°	-
5	-66°	120°	X
6	-66°	168°	-
7	-66°	216°	-
8	-66°	264°	X
9	-66°	312°	-
10	-44°	0°	-
11	-44°	24°	-
12	-44°	72°	X
13	-44°	120°	-
14	-44°	168°	-
15	-44°	216°	X
16	-44°	264°	-
17	-44°	312°	X
18	-22°	0°	X
19	-22°	24°	-
20	-22°	72°	-
21	-22°	120°	-
22	-22°	168°	X
23	-22°	216°	-
24	-22°	264°	-
25	-22°	312°	-
26	0°	0°	X
27	0°	24°	X
28	0°	72°	-
29	0°	120°	X
30	0°	168°	-
31	0°	216°	-
32	0°	264°	X
33	0°	312°	-
34	22°	0°	-
35	22°	24°	-
36	22°	72°	X
37	22°	120°	-
38	22°	168°	-
39	22°	216°	-
40	22°	264°	-
41	22°	312°	-
42	44°	0°	-
43	44°	24°	-
44	44°	72°	-
45	44°	120°	-
46	44°	168°	-
47	44°	216°	-
48	44°	264°	-
49	44°	312°	X
50	66°	0°	-
51	66°	24°	-
52	66°	72°	-
53	66°	120°	-
54	66°	168°	-
55	66°	216°	X
56	66°	264°	-
57	66°	312°	-
58	90°	0°	X

3.1.3 Processing signals to create virtual sound sources

Signals for right ear and for left ear were convolved in real time with the Head-Related Transfer Functions (HRTF) which corresponded to the direction of the desired sound source position relative to the orientation and position of the listeners head. This was enabled by fitting the headphone with a Head Tracker, which continuously did sent the positions and orientations to the program, which was performing processing of the signals. The used HRTF database did hold measurements from 11950 different directions, i.e. 2° angular resolution [5]. The complete real time 3D-Audio system including HRTF database, head tracker, convolution, binaural reverberation etc. was a product developed by AM3D A/S [12]. The processing also included equalization for the transfer function of the used headphone [5]. A calibration of the Head Tracker was performed before the start of each test session.

3.2 Physical setup

3.2.1 Reproduction system for real sound sources

The physical setup consisted of 16 real sound sources, i.e. loudspeakers, positioned in a distance of 2.13 m (7 feet) from a listening position, see figure 3.3. A curtain cylinder surrounded the listening position to avoid visibility of the loudspeakers. The light in the room was switched off during all tests and light was shined on the curtain from inside the cylinder, which ensured no visibility of objects outside the cylinder. The curtain cylinder was fitted both with a circular top and a circular bottom. The height of the curtain cylinder was 3.70 m and the diameter was 2.90 m.

The test subjects were standing on an adjustable platform, which elevated the ears of all subjects to a height of 2.15 m above floor level. Figure 3.3 shows the used elevation angles, while figure 3.4 shows a top view, where the used azimuth angles are shown. The physical setup was established in a large Television Studio at the facilities of AM3D A/S. The adjustable platform measured 30 cm by 30 cm, which constrained the possibility for test subjects to move away from the reference listening position. However the subjects were allowed to turn their whole body around in order to localize the sound and for precise pointing toward the position, where they perceived the sound to originate from.

The sound pressure responses of all 16 loudspeakers were measured at the listening position, and individual equalization were applied to each loudspeaker to ensure very similar response from all 16 loudspeakers. The resulting amplitude responses of all the loudspeakers were flat in the range from 140 Hz to 15 kHz.

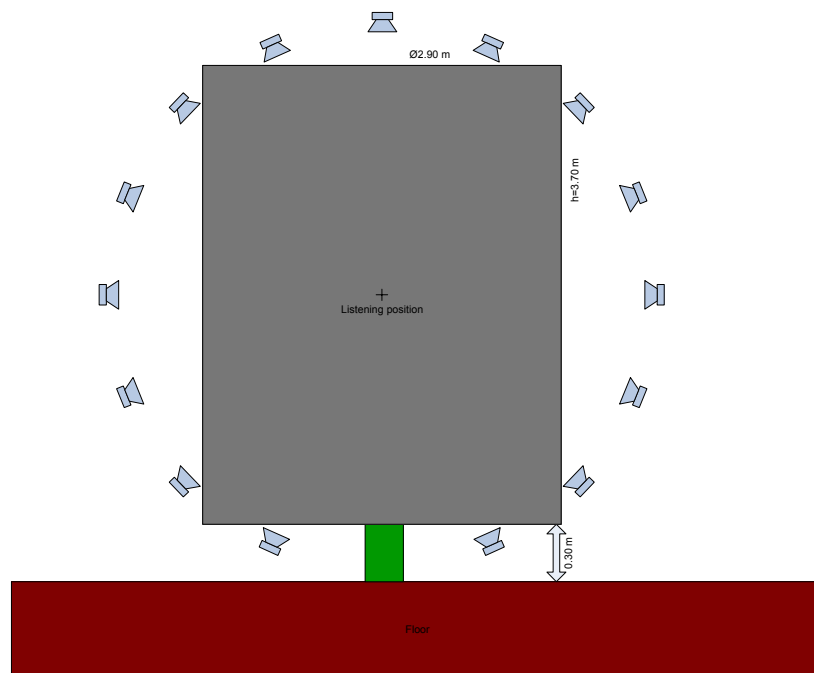


Figure 3.3: Side view of the physical setup, where 16 loudspeakers were positioned around a listening position in a distance of 2.13 m. A curtain cylinder (grey) surrounded the listening position. Diameter of the cylinder was 2.90 m and the height was 3.70 m. Both the top and bottom of the cylinder was fitted with curtain disks to close up the cylinder. Test subjects were standing on a platform (green) to ensure an ear height of 2.15 m.

Figure 3.5 shows the complete setup including the curtain cylinder and the loudspeakers mounted around it. This picture was taken from the left hand side, i.e. from the direction, where azimuth was +90 degrees. It follows from this, that the loudspeakers seen in the left side of the picture, were the loudspeakers directly in front of the listener, i.e. in the direction, where azimuth was 0 degrees.

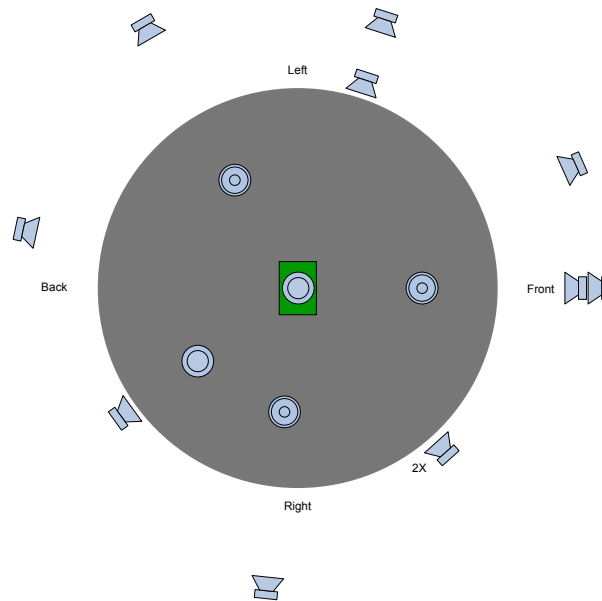


Figure 3.4: Top view of the physical setup, where 16 loudspeakers were positioned around a listening position in a distance of 2.13 m. A curtain disk (grey) forms the two ends of the closed cylinder. The diameter of the disks was 2.90 m. Three loudspeakers were pointing almost upwards (elevation = -66°), which is indicated by the upward facing loudspeakers. Two Loudspeakers were pointing downwards or almost downwards, which is indicated by the downward facing loudspeakers. All loudspeakers were positioned in a distance of 2.13 m from the listening position, but show up at different projected distances in this 2D top view diagram due to different values of elevation angle for the 16 different loudspeakers.



Figure 3.5: Picture of the curtain cylinder with loudspeakers mounted around it.

3.2.2 Reproduction system for virtual sound sources

The reproduction setup for virtual sources consists of signal processing inside the 3D-Audio system and the test subject wearing headphones, which were equipped with a head tracker. Based on the position and orientation of the test subjects head combined with the desired direction of the virtual sound source digital filters were applied to the short or long white noise burst as described in section 3.1.3. The selected headphone was a “Beyerdynamic – DT 990 PRO”, which has a specified frequency range of 5 Hz – 35 kHz. The head tracker updated the position and orientation of the headphones with a rate of 60 Hz. The maximum latency was 35 ms. Each headphone session started with a calibration of the head tracker by instructing the subject to look straight forward towards a fixed point on the curtain, which was in the same height as the eyes of the subject and in the direction defined as zero azimuth.

3.2.3 Pointing device

The test subjects had to show their perceived direction by pointing. For this a toy gun equipped with a tracking device was used. The test subjects were instructed to point by aiming at a target using both hands and straight arms. This ensured that the toy gun was held in a position directly in front of the test subjects, i.e. not of the either right or left hand side. The test subject were also instructed to turn their whole body as part of the localization task and when pointing they should be facing the position, where they perceived the sound to originate from. This should prevent subject from pointing and answering in an imprecise way, e.g. by shooting over their shoulder to point to a position behind them. Subjects were instructed to notice where they were pointing by observing where the laser dot shined on the curtain. The update rate of the tracker mounted on the toy gun was also 60 Hz, and this tracker was calibrated at the start of each session by instructing the test subject to point and aim at the same fix point on the curtain as described in section 3.2.2. This was performed at the same time as the calibration of the head tracker.

3.2.4 Automated test system

Presentation of stimuli in the predetermined directions, adjusting the signal processing according to the position and orientation of the head tracker, performing all the necessary signal processing and recording the answered directions by reading the data from the gun tracker was all performed by an integrated system, which was build around a PC including test system software and signal processing software, head tracker control and power amplifiers. Figure 3.6 shows the structure of this system. The system included a user interface which was used by the test operator to control the test.

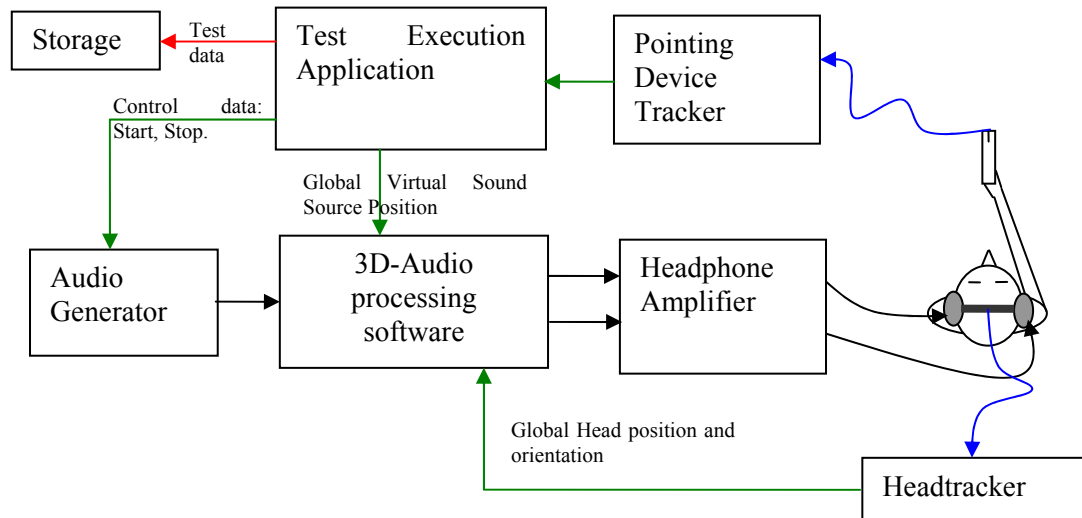


Figure 3.6: Structure of the automated test system

3.3 Procedure for the experiment

3.3.1 Selection of test subjects

The 3D-Audio system under test is targeted towards applications in fighter aircrafts. This led to the desire to test if fighter pilots have special requirements for a 3D-Audio system. In practise this was ensured by recruiting half of test subjects from the Royal Danish Air Force: 13 pilots from the Air Force Bases in Aalborg and Karup, Denmark. The other half of test subjects was civil persons, primarily students from Aalborg University, Denmark. This led to a total number of test subjects of 26 persons.

3.3.2 Screening and instruction of test subjects

All test subjects had a test of hearing threshold, i.e. audiogram. All test subjects were ensured to fall within a range of hearing loss between +20 dB to -10 dB, based on the standard hearing threshold (ISO 389). Then the test subjects had their picture taken, determination of leading eye, collection of personal facts (name, age, sex, experience in listening tests, ear height above the floor and total height) and they had to read a 2 page instruction, which described their task in the experiment. A short discussion between the test operator and the test subject should minimize the risk of misunderstandings. The height of the platform in the bottom of the curtain cylinder was then adjusted according to the measured height of the ears of the test subjects to ensure that the height of the ears when standing on the platform was 2.15 m in all cases. The subjects were not permitted to view the loudspeakers in the setup prior or during the experiment.

3.3.3 Test sessions

The whole experiment was split into 6 test session for each test subject:

- Headphone familiarization (training: 16 directions, 3 repetitions, long stimuli)
- Virtual sound sources, 2 s stimuli
- Virtual sound sources, 250 ms stimuli
- Loudspeaker familiarization (training: 16 directions, 3 repetitions, long stimuli)
- Real sound sources, 2 s stimuli
- Real sound sources, 250 ms stimuli

The order of these 6 sessions was made random for each person by random selection of 1 out of 4 possible sequences:

- Sequence A: Virtual sound sources first 2 s stimuli first
- Sequence B: Virtual sound sources first 250 ms stimuli first
- Sequence C: Real sound sources first 2 s stimuli first
- Sequence D: Real sound sources first 250 ms stimuli first

Following this terminology then sequence A is similar to the sequence given in the start of section 3.3.3.

Each sound source direction was used 3 times for each subject using both headphones and loudspeakers.

The total number of stimuli and answers were:

Headphone familiarization: 26 persons x 16 directions x 3 repetitions = 1248 stimuli

Virtual sources: 26 persons x 58 directions x 3 repetitions x 2 stimuli length = 9048 stimuli

Loudspeaker familiarization: 26 persons x 16 directions x 3 repetitions = 1248 stimuli

Real sources: 26 persons x 16 directions x 3 repetitions x 2 stimuli length = 2496 stimuli

Total: 14040 stimuli, which was equal to 39 hours of effective test (approx. 10 sec/stimuli)

This was equal to 1.5 hour of effective testing time for 1 test subject, but screening, instruction and pauses led to a total testing time for 1 test subject of approx. 2.5 hours. The automatic control software inserted a pause in any test session after each 15 min. of testing time. The 2.5 hour for 1 test subject corresponded to a total testing time of 65 hours.

3.3.4 One test session

The procedure for any one session was:

- The test subject was positioned in the reference position: standing on the platform in the centre of the curtain cylinder facing the fix point on the curtain, which indicated zero azimuth and zero elevation.
- The subjects calibrated both the head tracker and the gun tracker by aiming with two-handed grip and straight arms towards the fix point on the curtain and by looking straight ahead towards the same fix point. The calibration was effectuated by pressing the trigger on the toy gun. In the case of real sound sources then only the gun tracker was calibrated.

- Before each stimulus, the test subject reported “Ready” to the system by pressing the trigger after returning to the reference position facing the fix point.
- Each stimulus signal was presented
- Answered direction was given by the subject through aiming with two-handed grip and straight arms toward the perceived direction and pressing the trigger.

4 RESULTS

4.1 Basic data analysis

The localization error, calculated as the difference between the answered direction and the desired direction, was regarded as a random variable. The distribution was assumed to be a Normal distribution. Unbiased estimators for mean (average), μ , and variance, σ^2 , were calculated from the raw directional data obtained, and standard deviation, σ , of the error was simply calculated as the square root of the estimated variance [14]. A 95% confidence interval centred at the estimated mean (average) having a width equal to 3.92 times the standard deviation was then determined as a model describing the basic data [14].

The basic data analysis was performed using data from all 26 test subjects but limited to 1 specific direction, 1 length of stimuli and 1 specific reproduction method. Data from virtual sound sources, 2 s stimuli in the direction ID=28 are plotted in figure 4.1. Direction ID=28 is located in the horizontal plane, i.e. elevation=0°. The azimuth for the direction is 72° (to the left relative to front direction). For such a specific subset of data 78 answers were obtained (26 persons x 3 repetitions).

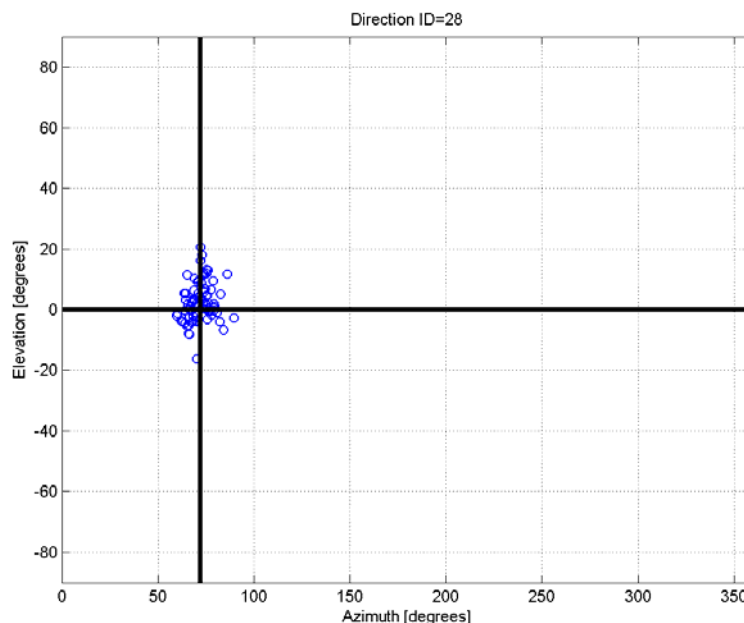


Figure 4.1: 78 answers for the direction indicated by the black horizontal and vertical lines. Each answer is indicated by a blue circle. Data were obtained using virtual sources and 2 s stimuli.

The azimuth and elevation values for each answer were then made relative to the stimuli direction by subtracting the stimuli azimuth and elevation. Figure 4.2 shows these relative answers, i.e. the localization errors. The red rectangle indicates 95% confidence intervals for azimuth and elevation, while the average azimuth error, μ_a , and average elevation error, μ_e , are indicated by black horizontal and vertical lines. The statistical parameters for these data are given in table 4.1.

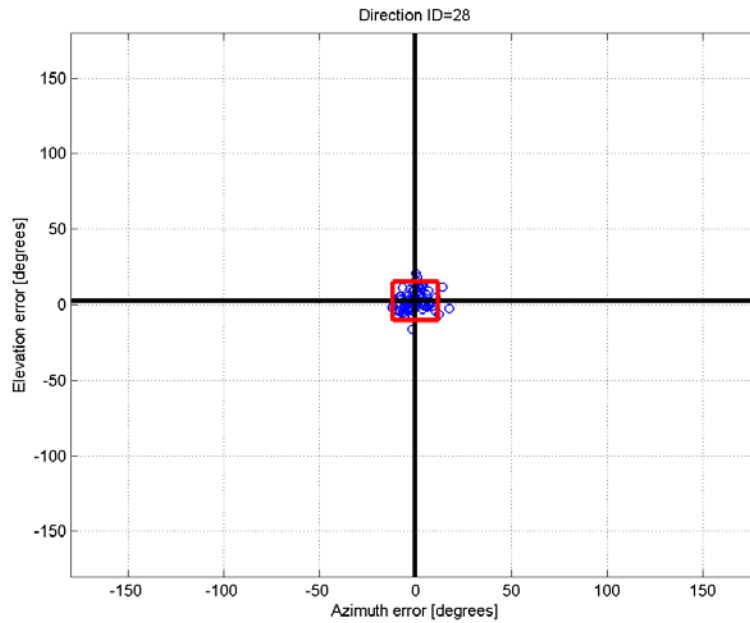


Figure 4.2: Localization error calculated from the 78 answers obtained using virtual sources and 2 s stimuli in direction 28 (azimuth=72°,elevation=0°). The red rectangle indicates 95% confidence interval for azimuth and elevation errors. Each localization error is indicated by a blue circle. Black horizontal and vertical lines indicate the global average azimuth and elevation errors.

Table 4.1: Statistical characteristics for the localization error obtained in the direction ID=28 using virtual sources and 2 s stimuli.

	Azimuth	Elevation
Average error (μ)	-0.1°	2.6°
Standard deviation (σ)	6.0°	6.5°
95% confidence interval	[-11.9° ; 11.6°]	[-10.1° ; 15.4°]

Equations 4.1 and 4.2 give the stochastic model for the localization error both for azimuth and elevation.

$$error_{azimuth} \in N(\mu_a, \sigma_a^2) \quad (4.1)$$

$$error_{elevation} \in N(\mu_e, \sigma_e^2) \quad (4.2)$$

4.2 Direction offset and uncertainty across all directions

The analysis process described in section 4.1 was repeated for all 58 directions used for virtual sources. Figure 4.3 shows the average azimuth error, μ_a , for all 58 directions, which shows that the average error, μ_a , must be modelled as a stochastic variable itself. This stochastic variable is also assumed to be a Normal distribution. The black line indicates the over all average for azimuth error, i.e. a global average azimuth error, $\mu_{a,global}$, and the 95% confidence interval for the average azimuth error is indicated by red horizontal lines.

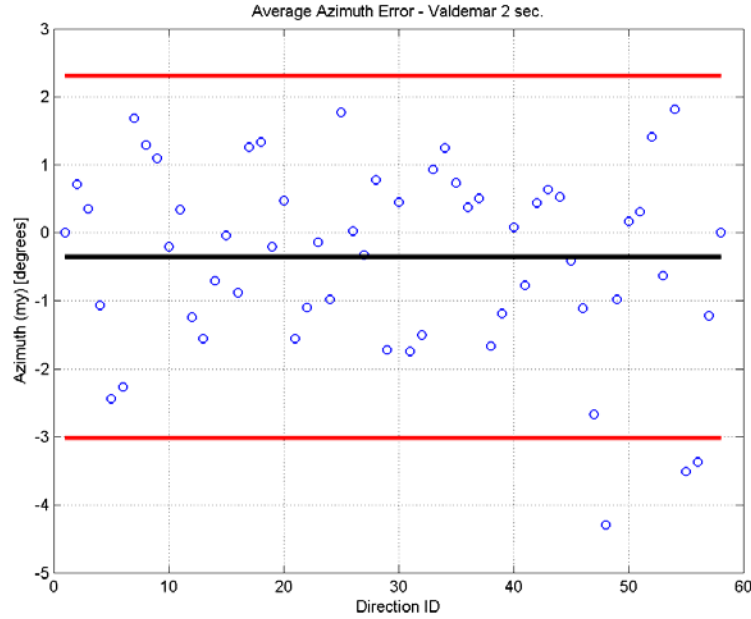


Figure 4.3: Average azimuth error calculated for the 58 directions used for virtual sound sources indicated by blue circles. The basic data were obtained using virtual sources and 2 s stimuli. The black horizontal line indicates the global average azimuth error, and the red horizontal lines indicate the 95% confidence interval for the average azimuth error.

Equation 4.3 gives the stochastic model for the distribution of average azimuth error, μ_a , at different directions, as shown in figure 4.3. The standard deviation of the average error is termed: $\sigma_{average,a}$. The global average azimuth error, $\mu_{a,global}$, was calculated as given in equation 4.4 based on the average azimuth error, $\mu_a(i)$, in all 58 directions.

$$\mu_a \in N(\mu_{a,global}, \sigma_{average,a}^2) \quad (4.3)$$

$$\mu_{a,global} = \frac{1}{58} \cdot \sum_{i=1}^{58} \mu_a(i) \quad (4.4)$$

The localization error can now be modelled as the sum of two stochastic variables, both Normal distributed, which in turn yields one Normal distributed stochastic variable with an increased variance. Equation 4.5 then gives the total stochastic model for the localization error, where $\sigma_{a,total}$ is the total

standard deviation of the azimuth error, which can be calculated using equation 4.6 [14]. This was done by assuming that the stochastic element within isolated directions was independent from the stochastic element between directions [14].

$$error_{azimuth} = N(\mu_{a,global}, \sigma_{a,total}^2) \quad (4.5)$$

$$\sigma_{a,total}^2 = \sigma_{average,a}^2 + \sigma_a^2 \quad (4.6)$$

Figure 4.4 shows the average elevation error for the 58 directions used for virtual sources, 2 s stimuli. Here it should be noted, that the average error seems to correlate to the direction ID. In fact the average elevation error was found to be a function of the stimuli elevation and the global average elevation error, $\mu_{e,global}$. Figure 4.5 shows the relationship between the average elevation error, μ_e , the stimuli elevation, S_e , and the global average elevation error, $\mu_{e,global}$, which is also described in Equation 4.7.

$$\mu_e = \alpha \cdot S_e + \mu_{e,global} = -0.25 \cdot S_e + 5.7^\circ \quad (4.7)$$

α is an “Elevation error factor”, which describes how large a fraction of an increase in stimuli elevation is reflected in an increased elevation error, i.e. larger stimuli elevation yields larger elevation error. Figures 4.6 and 4.7 show the standard deviations for azimuth errors and elevation errors as a function of stimuli direction ID. Figure 4.8 shows how the standard deviations for elevation errors increased when stimuli moved away from the horizontal plane (elevation=0°).

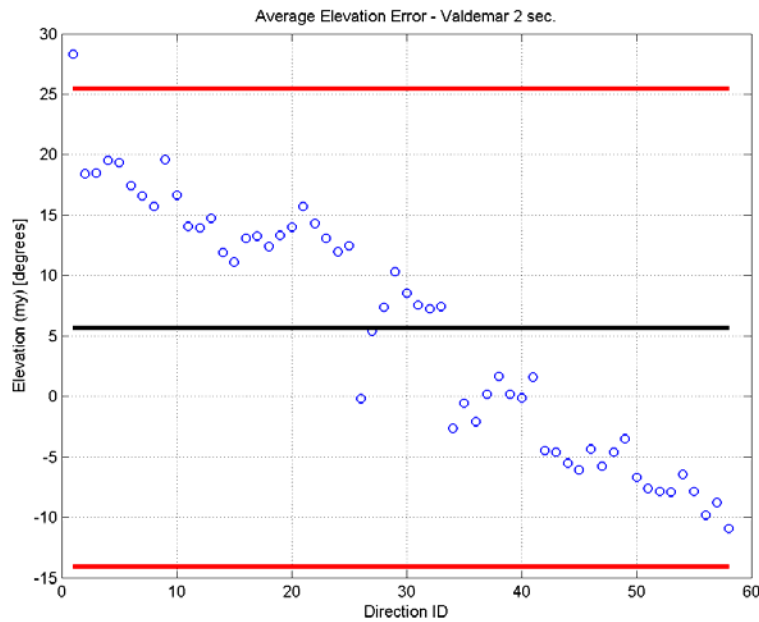


Figure 4.4: Average elevation error calculated for the 58 directions used for virtual sound sources indicated by blue circles. The basic data were obtained using virtual sources and 2 s stimuli. The black horizontal line indicates the global average elevation error, and the red horizontal lines indicate the 95% confidence interval for the average elevation error.

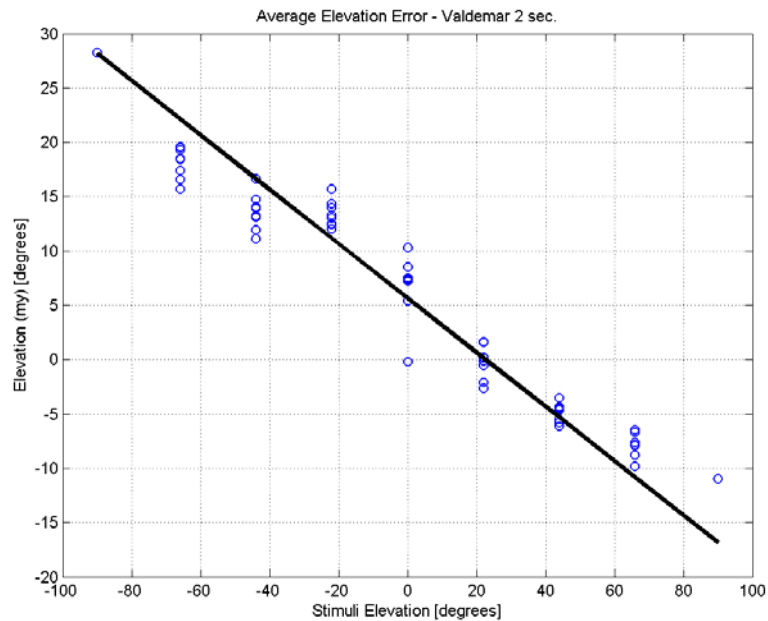


Figure 4.5: Average elevation error calculated for 58 directions plotted as a function of stimuli elevation. The basic data were obtained using virtual sources and 2 s stimuli. The black straight line indicates a simple relationship between average elevation error and stimuli elevation.

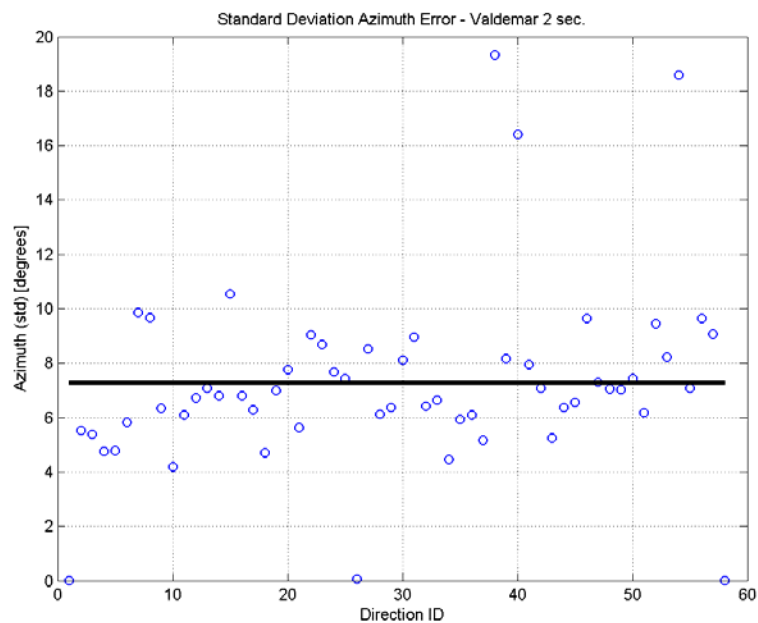


Figure 4.6: Standard deviations for azimuth error calculated for the 58 directions used for virtual sound sources indicated by blue circles. The basic data were obtained using virtual sources and 2 s stimuli. The black horizontal line indicates the average standard deviation across directions.

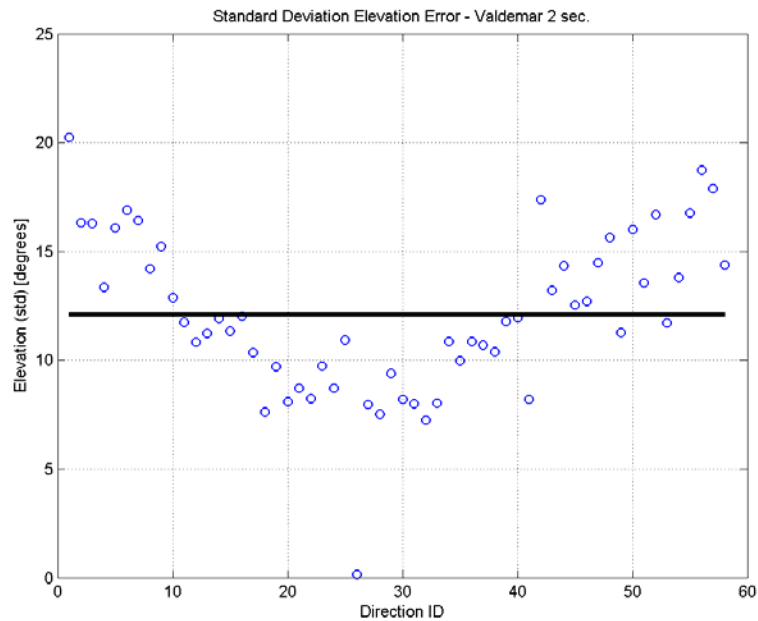


Figure 4.7: Standard deviations for elevation error calculated for the 58 directions used for virtual sound sources. The basic data were obtained using virtual sources and 2 s stimuli. The black horizontal line indicates the average standard deviation across directions.

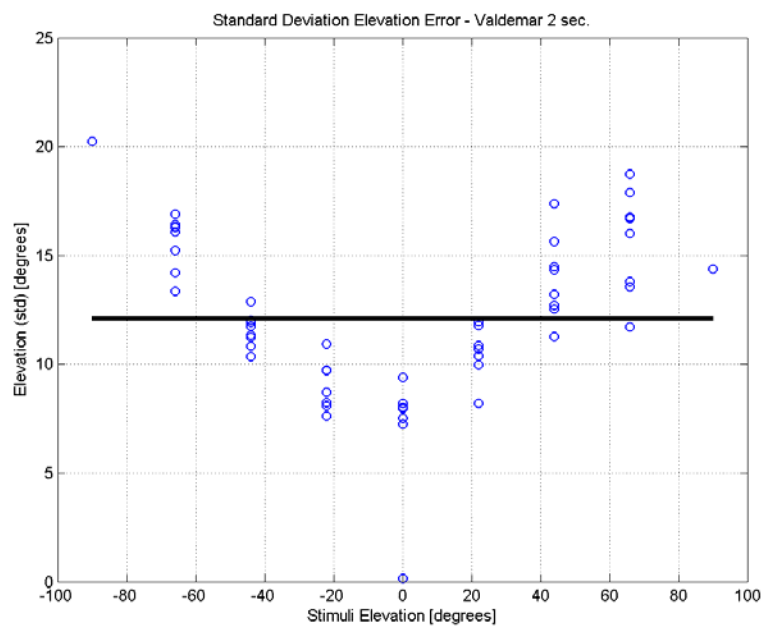


Figure 4.8: Standard deviations for elevation error calculated for 58 directions plotted as a function of stimuli elevation. The basic data were obtained using virtual sources and 2 s stimuli. The black horizontal line indicates the average standard deviation across directions.

Localization Performance of Real and Virtual Sound Sources

Figure 4.9 shows all 4524 answers obtained using virtual sound sources and 2 s stimuli. The black horizontal and vertical lines show the global average for azimuth error, $\mu_{a,global}$, and elevation error, $\mu_{e,global}$. The red rectangle indicates 95% confidence intervals for azimuth and elevation, which are based on the statistical parameters given in table 4.2.

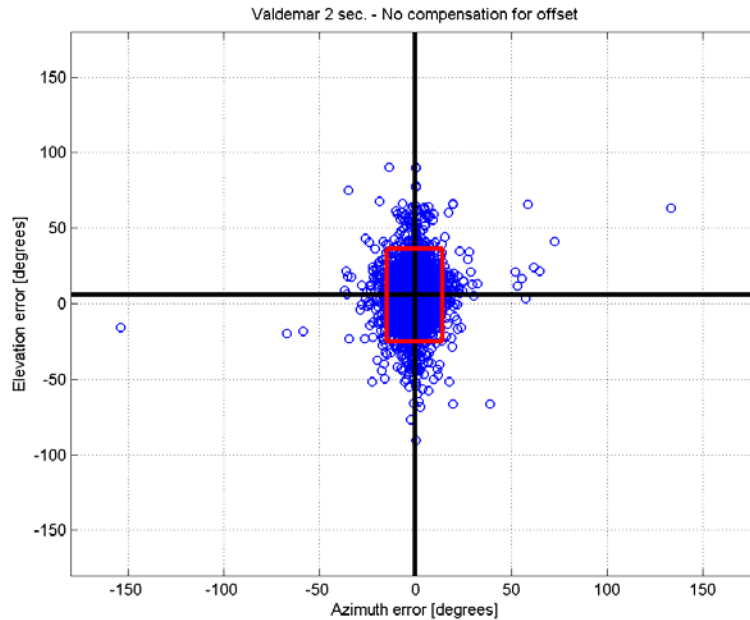


Figure 4.9: Localization error calculated from 4524 answers obtained using virtual sources and 2 s stimuli in all 58 directions. The red rectangle indicates 95% confidence interval for azimuth and elevation errors. Each localization error is indicated by a blue circle. Black horizontal and vertical lines indicate the global average azimuth and elevation errors.

Table 4.2: Statistical characteristics for the localization error obtained using virtual sources and 2 s stimuli.

	Azimuth	Elevation
Average error (μ_{global})	-0.4°	5.7°
Standard deviation (σ_{total})	7.4°	15.7°
95% confidence interval	$[-14.8^{\circ}; 14.1^{\circ}]$	$[-25.2^{\circ}; 36.5^{\circ}]$

The localization offset was then removed by calculating the localization error as the difference between the answered directions and the average of the answered directions for a given stimuli direction. This is equivalent to compensating for the localization offset in the different directions, see figures 4.3, 4.4 and 4.5. It follows from this that the average localization error becomes zero in all stimuli directions, and the global average localization error vanishes to zero. The standard deviation of the average localization error also becomes zero because the localization error was constant (zero) in all stimuli directions. Though the

total standard deviation for the localization error, σ_{total} , now becomes equal to the standard deviation for the localization error in the individual directions, σ , because no increase in variance is found due to variation of the average localization error across stimuli directions. Figure 4.10 shows all 4524 answers obtained using virtual sound sources and 2 s stimuli after compensation for the localization offset found in the 58 individual stimuli directions. The black horizontal and vertical lines show the global average for azimuth error, $\mu_{a,global}$, and elevation error, $\mu_{e,global}$, which are both trivially equal to 0° after compensation. The red rectangle indicates 95% confidence intervals for azimuth and elevation, which are based on the compensated statistical parameters given in table 4.3.

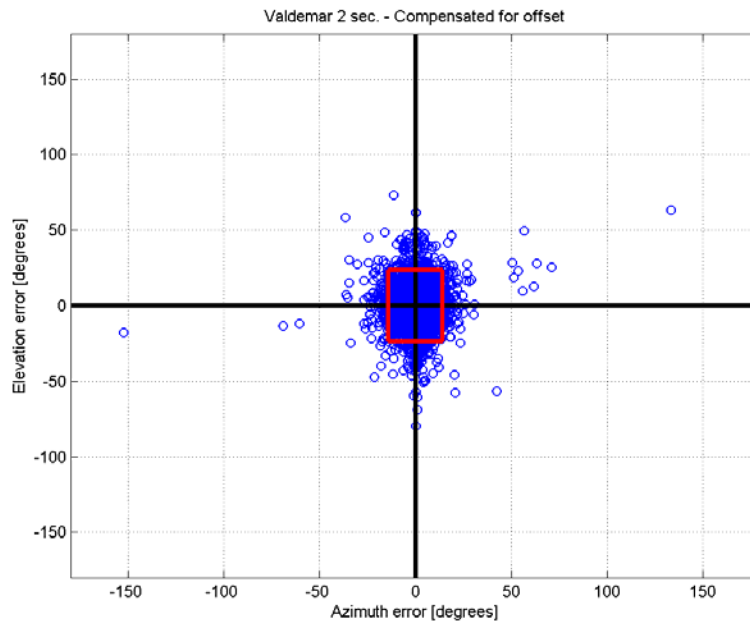


Figure 4.10: Compensated localization error calculated from 4524 answers obtained using virtual sources and 2 s stimuli in all 58 directions. The red rectangle indicates 95% confidence interval for azimuth and elevation errors. Each localization error is indicated by a blue circle. Black horizontal and vertical lines indicate the global average azimuth and elevation errors

Table 4.3: Statistical characteristics for the localization error obtained using virtual sources and 2 s stimuli compensated for the localization offset found in each of the 58 stimuli directions.

	Azimuth	Elevation
Average error (μ_{global})	0.0°	0.0°
Standard deviation (σ)	7.3°	12.1°
95% confidence interval	$[-14.2^\circ ; 14.2^\circ]$	$[-23.7^\circ ; 23.7^\circ]$

4.3 Accuracy of physical setup and directional data recording

A special session was run in order to determine the accuracy of the physical setup and the system for recording the directional data, i.e. the positions of the real sound sources (loudspeakers) and the tracker system, which determines the position and orientation of the toy gun. A test subject was placed on the platform inside the curtain cylinder and was instructed to point to each of the 16 loudspeakers one at a time. The test subject was assisted in this task by a person outside the curtain, who ensured that the laser dot shined in the centre of the loudspeaker (+/- 0.5 cm). The light in the room was turned on during this session to enable visibility of the loudspeaker from inside the cylinder. The obtained directional data was then analyzed in a similar way as described in section 4.2. Table 4.4 gives the results of this analysis.

$\sigma_{average}$ is a measure for the accuracy of the positions of the loudspeakers relative the theoretical correct positions as defined in table 3.1. σ on the other hand is a measure for the accuracy of the system for recording the directional data, i.e. gun tracker. The uncompensated confidence interval gives the total localization uncertainty including both the uncertainty of loudspeaker positions and the gun tracker. The compensated confidence interval is the uncertainty only for the gun tracker. This compensation for localization offset was described in section 4.2.

Two methods were tried out for determining the answered direction:

- Direction given by the orientation of the toy gun
- Direction given by a straight line passing through the centre of the head of the test subject and the calculated position of the laser dot on the curtain.

Similar results were obtained by the two methods, which led to usage of the orientation of the toy gun for all directions in this paper.

Table 4.4: Statistical parameters for accuracy of physical setup and directional data recording.

	Azimuth	Elevation
Average localization error (μ_{global})	-0.1°	-0.9°
Standard deviation for the average ($\sigma_{average}$)	1.6°	2.2°
Standard deviation in each stimuli direction (σ)	0.9°	0.6°
Total standard deviation (σ_{total})	1.8°	2.3°
95% confidence interval	[-3.7° ; 3.5°]	[-5.4° ; 3.6°]
95% confidence interval (compensated)	[-1.8° ; 1.8°]	[-1.2° ; 1.2°]

4.4 Summary of results for sound source type and stimuli length

Table 4.5, 4.6, 4.7 and 4.8 gives a summary of the results obtained using the 4 combinations of long/short stimuli and virtual/real sound sources.

Table 4.5: Summary of analysis results obtained using virtual sources and 2 s stimuli.

	Azimuth	Elevation
Average localization error (μ_{global})	-0.4°	5.7°
Elevation error factor (α)	-	-0.25
Standard deviation for the average ($\sigma_{average}$)	1.4°	10.1°
Standard deviation in each stimuli direction (σ)	7.3°	12.1°
Total standard deviation (σ_{total})	7.4°	15.7°
95% confidence interval	[-14.8° ; 14.1°]	[-25.2° ; 36.5°]
95% confidence interval (compensated)	[-14.2° ; 14.2°]	[-23.7° ; 23.7°]

Table 4.6: Summary of analysis results obtained using real sources and 2 s stimuli.

	Azimuth	Elevation
Average localization error (μ_{global})	0.1°	2.9°
Elevation error factor (α)	-	-0.125
Standard deviation for the average ($\sigma_{average}$)	1.3°	4.8°
Standard deviation in each stimuli direction (σ)	4.9°	6.2°
Total standard deviation (σ_{total})	5.1°	7.8°
95% confidence interval	[-9.9° ; 10.0°]	[-12.4° ; 18.2°]
95% confidence interval (compensated)	[-9.6° ; 9.6°]	[-12.1° ; 12.1°]

Table 4.7: Summary of analysis results obtained using virtual sources and 250 ms stimuli.

	Azimuth	Elevation
Average localization error (μ_{global})	-2.8°	0.9°
Elevation error factor (α)	-	-0.50
Standard deviation for the average ($\sigma_{average}$)	13.6°	20.5°
Standard deviation in each stimuli direction (σ)	21.4°	19.2°
Total standard deviation (σ_{total})	25.4°	28.1°
95% confidence interval	[-52.5° ; 46.9°]	[-54.2° ; 55.9°]
95% confidence interval (compensated)	[-42.0° ; 42.0°]	[-37.7° ; 37.7°]

Table 4.8: Summary of analysis results obtained using real sources and 250 ms stimuli.

	Azimuth	Elevation
Average localization error (μ_{global})	0.7°	4.4°
Elevation error factor (α)	-	-0.125
Standard deviation for the average ($\sigma_{average}$)	6.0°	8.7°
Standard deviation in each stimuli direction (σ)	8.8°	11.2°
Total standard deviation (σ_{total})	10.7°	14.2°
95% confidence interval	[-20.2° ; 21.6°]	[-23.5° ; 32.2°]
95% confidence interval (compensated)	[-17.3° ; 17.3°]	[-22.0° ; 22.0°]

4.5 Pilots vs. civil persons

Separate analysis has been performed on data from the 13 pilots and from the 13 civil persons to investigate if differences could be found. Table 4.9 gives the uncompensated confidence intervals and table 4.10 gives the confidence intervals, which are compensated for localization offset.

Table 4.9: 95% confidence intervals – no compensation for localization offset.

	Azimuth	Elevation
virtual sources – 2 s – pilots	$[-13.1^{\circ}; 9.5^{\circ}]$	$[-26.1^{\circ}; 35.2^{\circ}]$
virtual sources – 2 s – civil persons	$[-15.2^{\circ}; 17.5^{\circ}]$	$[-24.2^{\circ}; 37.7^{\circ}]$
real sources – 2 s – pilots	$[-8.0^{\circ}; 6.7^{\circ}]$	$[-11.6^{\circ}; 17.9^{\circ}]$
real sources – 2 s – civil persons	$[-10.3^{\circ}; 11.9^{\circ}]$	$[-12.8^{\circ}; 18.3^{\circ}]$
virtual sources – 250 ms – pilots	$[-57.2^{\circ}; 48.2^{\circ}]$	$[-52.4^{\circ}; 57.5^{\circ}]$
virtual sources – 250 ms – civil persons	$[-47.6^{\circ}; 45.3^{\circ}]$	$[-55.6^{\circ}; 54.1^{\circ}]$
real sources – 250 ms – pilots	$[-21.6^{\circ}; 21.6^{\circ}]$	$[-25.3^{\circ}; 36.0^{\circ}]$
real sources – 250 ms – civil persons	$[-18.4^{\circ}; 21.1^{\circ}]$	$[-21.2^{\circ}; 28.0^{\circ}]$

Table 4.10: Compensated 95% confidence intervals.

	Azimuth	Elevation
virtual sources – 2 s – pilots	$[-11.0^{\circ}; 11.0^{\circ}]$	$[-22.7^{\circ}; 22.7^{\circ}]$
virtual sources – 2 s – civil persons	$[-15.9^{\circ}; 15.9^{\circ}]$	$[-24.2^{\circ}; 24.2^{\circ}]$
real sources – 2 s – pilots	$[-6.9^{\circ}; 6.9^{\circ}]$	$[-11.4^{\circ}; 11.4^{\circ}]$
real sources – 2 s – civil persons	$[-10.6^{\circ}; 10.6^{\circ}]$	$[-12.0^{\circ}; 12.0^{\circ}]$
virtual sources – 250 ms – pilots	$[-41.1^{\circ}; 41.1^{\circ}]$	$[-38.3^{\circ}; 38.3^{\circ}]$
virtual sources – 250 ms – civil persons	$[-40.5^{\circ}; 40.5^{\circ}]$	$[-36.0^{\circ}; 36.0^{\circ}]$
real sources – 250 ms – pilots	$[-16.4^{\circ}; 16.4^{\circ}]$	$[-24.0^{\circ}; 24.0^{\circ}]$
real sources – 250 ms – civil persons	$[-16.9^{\circ}; 16.9^{\circ}]$	$[-18.8^{\circ}; 18.8^{\circ}]$

4.6 Front/back reversals

Front-back ambiguity has been found to be a significant problem for 3D-Audio systems, however systems including a head tracker have been shown to almost solve this problem [11]. The front-back ambiguity or front/back reversal occurs when a sound event is perceived in a position in front of the person when the

sound source in fact is positioned behind the person. The characteristics of the misperceived position relative to the true position is a direct front-to-back or back-to-front mirroring in a vertical plane that passes through the two ears of the person (front and back hemisphere). A special analysis to uncover this problem has been performed, where the localization error was calculated to be the smaller of two possibilities:

- The difference between the perceived direction and the stimulus direction
- The difference between the mirror image of the perceived direction and the stimulus direction

Table 4.11 gives the results for this analysis, where front/back% is a measure of how many times the mirrored position was used compared to the total number of positions, i.e. how often front/back reversals were identified.

Table 4.11: Localization offset compensated 95% confidence intervals for azimuth, where front/back reversals mirrored back to positions closer to the stimuli directions.

	Azimuth - normal	Azimuth - mirrored	Front/back%
virtual sources – 2 s	[-14.2° ; 14.2°]	[-12.7° ; 12.7°]	5.1%
real sources – 2 s	[-9.6° ; 9.6°]	[-8.2° ; 8.2°]	4.2%
virtual sources – 250 ms	[-42.0° ; 42.0°]	[-27.2° ; 27.2°]	21.3%
real sources – 250 ms	[-17.3° ; 17.3°]	[-14.3° ; 14.3°]	9.1%

5 DISCUSSION

One of the interesting findings in the results was the strong correlation between average elevation error and stimuli elevation, which was shown in figure 4.5. Equation 4.7 formulated a simple relationship between stimuli elevation, S_e , and the average elevation error, μ_e . The global average elevation error, $\mu_{e,global}$, formed an offset for the average elevation error as seen in equation 4.7. One way to exploit this knowledge is to compensate the stimulus direction in such a way, that taking the average elevation error into account will yield a perceived direction, which closely approximates a desired elevation, D_e . Equation 5.1 states the relationship between the elevation error, $error_e$, stimulus elevation, S_e , and the answered elevation, A_e . The average elevation error, μ_e , can then be calculated from equation 5.2, where $E[x]$ is the expected value operator [14].

$$error_e = A_e - S_e \quad (5.1)$$

$$\begin{aligned} \mu_e &= E[A_e - S_e] \\ \Downarrow \\ \mu_e &= E[A_e] - S_e \end{aligned} \quad (5.2)$$

The expected value of the answered elevation, i.e. the average answered elevation, $E[A_e]$, should be equal to the desired elevation, D_e . Combining this with equation 4.7 leads to equation 5.3, which gives the

relationship between the desired elevation, D_e , and the stimuli elevation, S_e . Isolating the stimuli direction, S_e , leads to equation 5.4, which specifies the needed stimuli direction for a given desired elevation in order to compensate for the found relationship between average elevation error and stimuli direction.

$$D_e = \mu_e + S_e = \alpha \bullet S_e + \mu_{e,global} + S_e = (1 + \alpha) \bullet S_e + \mu_{e,global} \quad (5.3)$$

$$S_e = \frac{D_e - \mu_{e,global}}{1 + \alpha} \quad (5.4)$$

Both the global average elevation error, $\mu_{e,global}$, and the elevation error factor, α , can be found in tables 4.5, 4.6, 4.7 and 4.8. Using equation 5.4 for calculating the needed stimuli leads generally to stimuli elevation angles larger than the desired elevation angles, e.g. for a desired elevation angle of 60° require a stimuli elevation angle of 72.4° using virtual sources and 2 s stimuli (table 4.5).

However stimuli elevation angles can't go above $+90^\circ$ or below -90° . This fact imposes a limit onto the range of possible desired elevation angles. Setting the stimuli elevation, S_e , to $+90^\circ$ and -90° in equation 5.3 leads to a lower and upper limit for desired elevation angles to -61.8° and $+73.2^\circ$. The impact of this statement is that positions near the “poles of the sphere”, i.e. -90° elevation and $+90^\circ$, represents a special challenge, which require further research.

Why should stimuli elevation angle be larger than the desired elevation angle? One possible answer could be that humans are more specialized in localizing in the horizontal plane due to everyday life, where the majority of sound events take place in the same level above the ground as our ears. This combined with the fact that the human ears are located in a way, which favours the discrimination from left to right rather than down to up, may explain why humans has a default localization near the horizontal plane?

The results given in table 4.5, 4.6, 4.7 and 4.8 supports a statement saying that: “The more a human is in doubt about where the sound originates from – the more restricted range exists for the answered elevation angles”. E.g. using virtual sources and short stimuli (250 ms) yields a possible range for the desired elevation angles to $[-44.1^\circ ; 45.9^\circ]$. The experience is that sound localization becomes more difficult for shorter stimuli and for virtual sources compared to real sources. Table 5.1 gives the range of possible desired elevation angles in the 4 cases originating from tables 4.5, 4.6, 4.7 and 4.8.

Table 5.1: Ranges for possible desired elevation angles.

	Possible desired elevation angles
virtual sources – 2 s	$[-61.8^\circ ; 73.2^\circ]$
real sources – 2 s	$[-75.9^\circ ; 81.7^\circ]$
virtual sources – 250 ms	$[-44.1^\circ ; 45.9^\circ]$
real sources – 250 ms	$[-74.4^\circ ; 83.2^\circ]$

Using azimuth and elevation to specify a direction in 3D may represent a problem: when calculating the average elevation error near the “poles of the sphere”, then the answered directions are limited/clipped to

the range $\pm 90^\circ$. E.g. for stimuli elevation angle = $+90^\circ$, where test subjects will point to directions distributed in an area around the “pole of the sphere”. Any position away from the “pole of the sphere” will yield a negative elevation error – the elevation error can’t be positive in this situation by the definition of elevation angle. It follows from this, that even if the answered directions are evenly distributed around the “pole of the sphere”, then the average elevation error will not be zero, i.e. it will be negative. This led to a special analysis, where the average localization error wasn’t split into azimuth and elevation error – it was performed as an average of complex numbers, which represented the localization error in polar form. However this didn’t change the result dramatically, but the standard deviation for elevation error did decrease.

Figure 4.9 and 4.10 confirms in a graphical way the fit between the obtained localization errors and the stochastic model including the calculated parameters. Figure 4.10 shows how compensation for the localization offset do decrease the overall localization uncertainty in a significant way. It follows from this, that it is very important to separate the variance due to different localization offsets in different direction from the variance due to localization uncertainty in each direction.

Table 4.4 shows that both the accuracy of the physical setup and the accuracy of the directional data recording system were sufficient compared to the results obtained in the experiment.

Standard deviation for azimuth using real sources and 2 s stimuli was found to be 4.9° (table 4.6), which is comparable to results obtained by other authors: 1.5° (in the stimuli range $\pm 30^\circ$) [4], 8° (average absolute localization error) [9] and 1° to 3° [8]. The standard deviation for elevation errors using real sources and 2 s stimuli was found to be 6.2° , which again could be compared to the average absolute localization error found by [9]: 8° . Differences in choice of stimuli, stimuli length, physical setup, range and distribution of sound source positions may be possible explanations for these differences. It should be noted, that this experiment covered the full sphere of directions, i.e. combinations of azimuth and elevation covering the full sphere.

Virtual sources and stimuli length of 2 s yielded a standard deviation of 7.3° for azimuth and 12.1° for elevation. The average absolute localization error for virtual sources was found by [9] to be 11° and 10° by [13]. Using short stimuli (250 ms) this was found to increase to 21.4° for azimuth and 19.2° for elevation. This compares to the data presented by [13]: 18° , which was a test restricted to the horizontal plane.

The standard deviations obtained for pilots generally were found to be smaller compared to civil persons. A significantly lower standard deviation was found for azimuth errors for pilots compared to civil persons: 5.0° vs. 8.0° . The ability to focus on a task at hand and to use the input from the auditory system may be an explanation for this result?

The front/back reversal rate was found to be 4.2% (2 s stimuli) and 9.1% (250 ms stimuli) for real sources, which is lower than the results obtained by [9]: 12%. However this shows that even when using real sources the auditory system has a significant rate of front/back reversals. The rate for virtual sources was found to be 5.1% (2 s stimuli) and 21.3% (250 ms). In [9] the front/back reversal rate was found to be 20% for virtual sources and in [13] the rate was found to be 4% (2 s stimuli) and 10% (250 ms). It is clear, that short stimuli, i.e. no head movement, greatly increases the problem with front/back reversals both for real and virtual sources, however the increase was substantially larger for virtual sources.

The main results of the paper show that the uncertainty for azimuth for virtual sources does compare to the real sources. However the uncertainty for elevation is significantly higher for the virtual sources. Actually the uncertainty for elevation for virtual sources (2 s stimuli) compares well with real sources for short stimuli (250 ms). The uncertainty both for azimuth and elevation increased for shorter stimuli, where head movements can’t be used in the localization task. This was particularly clear for virtual sources.

6 CONCLUSION

The localization uncertainty was much higher for short stimuli (0.25 sec.) compared to long stimuli (2 sec.). The long stimuli enabled head movements to be used in the localization task.

Pilots had a lower localization uncertainty compared to civil persons. Head movements, i.e. long stimuli, greatly reduced the problem of front/back reversals, especially for virtual sound sources.

No significant localization offset was found for azimuth, while an offset for elevation of 3 – 6 degrees was found using long stimuli. A significant difference between the localization offsets (average error) obtained in different directions was found – especially for elevation, where the offset was found to have a strong correlation to the stimuli elevation. It follows from this, that a significant part of the uncertainty for elevation can be removed by compensating for these individual offsets in the different directions (Compensated for localization offset).

A simple formula was devised for calculating the needed stimuli elevation from a desired elevation angle, which generally introduces larger stimuli elevation angles compared to the desired elevation angle.

Due to the needed compensation of elevation angle a restricted range of desired elevation angles was determined, which showed a problem of how to position sound sources near the “poles of the sphere”, i.e. near elevation = +90° and -90°.

The results of the analysis of the obtained directional data compared well to results presented by other authors under the observation that significant differences existed between the different experiments: stimuli, stimuli length, reproduction setup, range and distribution for azimuth and elevation angles and sound source type. This experiment covered the full sphere of directions: both azimuth and elevation.

The localization uncertainty for virtual sound sources was found to be larger than for real sound sources, especially for elevation. However the localization performance of virtual sound sources was comparable to the performance of real sound sources:

95% confidence interval (compensated for offset):

Real Sound Sources:	Azimuth=[-9.6 ; 9.6]	Elevation=[-12.1 ; 12.1]
Virtual Sources:	Azimuth=[-14.2 ; 14.2]	Elevation=[-23.7 ; 23.7]

7 ACKNOWLEDGMENT

The authors would like to thank Douglas S. Brungart, Wright-Patterson Air Force Base, for taking part in the early stage discussions about the physical setup for the experiment. Aalborg University is thanked for contributing to the experiment by providing loudspeakers and power amplifiers. The authors would also like to thank their colleagues from AM3D A/S and Terma A/S for contributing to the work, setting up and carrying out the experiment, for inspiration, for their experience and for decisive discussions: Michael Holmberg, Steffen Vang Madsen, Peter Theilgaard Hounum, Preben Schmidt Nielsen, Kim Eriksen, Simon Attermann Brinch, Robert Stepien and Per Schmock. Finally the authors would like to thank the 26 test subject for taking part in the experiment, i.e. 13 pilots from the Royal Danish Air Force, from the Air Force Bases in Aalborg and Karup, Denmark, and 13 civil persons, primary students in Acoustics at Aalborg University, Denmark.

8 REFERENCES

- [1] Douglas S. Brungart: “A Speech-Based Auditory Distance Display”, presented at the AES 109th convention, Los Angeles, 2000, September 22–25. Preprint number 4816.
- [2] S. H. Nielsen: “Auditory Distance Perception in Different Rooms”, *Journal of the Audio Engineering Society*, Volume 41, 1993, pp. 755–770.
- [3] A. W. Bronkhorst and T. Houtgast: “Auditory distance perception in rooms”, *Nature*, Volume 397, 1999, pp. 517–520.
- [4] Sylvain Choisel and Karin Zimmer: “A Pointing Technique with Visual Feedback for Sound Source Localization Experiments”, presented at the AES 115th convention, New York, 2003, October 10–13. Preprint number 5904.
- [5] Pauli Minnaar: “Simulating an Acoustical Environment with Binaural Technology – investigations of binaural recording and synthesis”, Ph.D. thesis obtained at Aalborg University, Denmark, September 2001.
- [6] Pauli Minnaar, Søren Krarup Olesen, Flemming Christensen and Henrik Møller: “Localization with Binaural Recordings from Artificial and Human Heads”, *Journal of the Audio Engineering Society*, Volume 49, No. 5, 2001, pp. 323–336.
- [7] H. Møller, D. Hammershøi, C. B. Jensen and M. F. Sørensen: “Evaluation of artificial heads in listening tests”, *Journal of the Audio Engineering Society*, Volume 47, No. 3, 1999, pp. 83–100.
- [8] J. Blauert: “Spatial hearing: The psychophysics of human sound localization”, MIT Press., Cambridge, USA, Revised edition, 1997.
- [9] A. W. Bronkhorst: “Localization of Real and Virtual Sound Sources”, *Journal of the Acoustical Society of America*, Volume 98, No. 5, 1995, pp. 2542–2553.
- [10] D. R. Begault, E. M. Wenzel, A. S. Lee and M. R. Anderson: “Direct comparison of the impact of head tracking, reverberation and individualized head-related transfer functions on the spatial perception of a virtual speech source”, presented at the AES 108th Convention of the Audio Engineering Society, Paris, France, 2000, February 19–22. Preprint number 5134.
- [11] F. L. Wightman and D. J. Kistler: “Resolution of front-back ambiguity in spatial hearing by listener and source movement”, *Journal of the Acoustical Society of America*, Volume 105, No. 5, 1999, pp. 2841–2853.
- [12] Company web site for AM3D A/S, including explanation of technology: www.am3d.com
- [13] Douglas S. Brungart *et al.*: “The interaction between head-tracker latency, source duration, and response time in the localization of virtual sound sources”, Proceedings of ICAD 04- Tenth Meeting of the International Conference on Auditory Display, Sydney, Australia, July 6–9, 2004.
- [14] Sheldon M. Ross: “Introduction to Probability and Statistics for Engineers and Scientists”. ISBN 0-471-60815-7.

Improving Multitalker Speech Communication with Advanced Audio Displays

Douglas S. Brungart and Brian D. Simpson

Air Force Research Laboratory, AFRL/HECB, Bldg 441, 2610 Seventh Street,
Wright-Patterson AFB, OH 45433-7901 Tel: (937) 255-3660 ext 422 Fax: (937) 656-7680

douglas.brungart@wpafb.af.mil

ABSTRACT

Historically, most of the metrics that have been used to evaluate the effectiveness of military communications systems have focused on measuring the intelligibility of a single talker in the presence of a continuous noise masker. However, many critical military operations involve complex communications tasks that require listeners to monitor, process, and respond to two or more simultaneous speech signals. Many factors can influence performance in such tasks, including the relative levels of the competing talkers, the similarities between the voices of the competing talkers, and, in audio displays that allow the different channels of speech to be spatially separated, the apparent locations of the target and masking voices. In this paper, we present an overview of the factors that can influence speech intelligibility in multitalker listening environments, and compare and contrast them to the factors that influence intelligibility in the speech-in-noise situations that are usually used to evaluate military communications systems. We also discuss the intelligibility benefits that can be achieved with advanced audio displays that use either dichotic or binaural processing to spatially separate the apparent locations of multiple simultaneous channels of speech. Finally, we describe a spatial configuration that has been shown to maximize the benefit of spatial separation in a listening configuration with up to seven simultaneous speech signals.

1.0 INTRODUCTION

Many critically important military tasks require listeners to monitor and respond to speech messages originating from two or more competing talkers. A classic example of this kind of task occurs in military air traffic control centers, where controllers are required to communicate critical information to and from multiple simultaneous aircraft while maintaining an acute awareness of the relative positions of all the aircraft in their assigned area. In command and control centers, commanders are also faced with the task of assimilating information obtained from multiple sources and using this information to coordinate multi-unit actions against enemy assets. Military pilots face an even more difficult situation where they may need to communicate with other personnel in their own aircraft, other aircraft in their own formation, command and control personnel in AWACS aircraft and at ground-based command centers, and ground-based target spotting personnel near the site of an airstrike. In all of these situations, a well-designed multitalker speech display could improve the overall performance of the operator, not only because it may reduce the chances of a potentially deadly miscommunication, but also because it may reduce the overall workload associated with multitalker listening and allow the operator to attend to other critical tasks.

Many researchers have commented on the substantial benefits that audio display technology can provide in a multitalker communication environment. Some of the earliest efforts in this area used spectral manipulations to enhance the segregation of multiple talkers in a monaural audio channel. For example, in a

Brungart, D.S.; Simpson, B.D. (2005) Improving Multitalker Speech Communication with Advanced Audio Displays. In *New Directions for Improving Audio Effectiveness* (pp. 30-1 – 30-18). Meeting Proceedings RTO-MP-HFM-123, Paper 30. Neuilly-sur-Seine, France: RTO.
Available from: <http://www.rto.nato.int/abstracts.aps>.

three-talker system, speech segregation may be enhanced by high-pass filtering one talker, low-pass filtering a second talker, and all-pass filtering the third talker (Spieth, Curtis, & Webster, 1954; MIL-STD-1472E, 1997). More recent efforts have used virtual audio displays to spatially separate the competing speech channels (Crispien & Ehrenberg, 1995; Ericson & McKinley, 1997; Begault, 1999). To this point, however, no consensus has been reached on the design parameters that are most important in determining the effectiveness of multitalker speech displays. In part, at least, this lack of consensus is a result of the extreme complexity of the multitalker listening problem. Performance in such tasks depends on a wide variety of factors, including: 1) the signal-to-noise ratio in the communications system; 2) the number of competing talkers; 3) the voice characteristics of the talkers; 4) the relative levels of the talkers; 5) the apparent spatial locations of the talkers; and 6) the listener's *a priori* knowledge about the listening environment. A further complicating issue is the variety of methodologies that has been used to examine these factors; procedural variations often make it difficult to compare the results of different multitalker listening experiments. In this paper, we present the results of a number of experiments that have used the Coordinate Response Measure (CRM) to examine the impact that different audio display design parameters has on performance in a multitalker communications task. This allows a comparison of the relative importance of each of these parameters that can be used as a guide in the design of multitalker speech displays.

2.0 EXPERIMENTAL METHODOLOGY: THE COORDINATE RESPONSE MEASURE

All of the experiments described in this paper were conducted using the Coordinate Response Measure (CRM). This speech intelligibility test was originally developed to provide greater operational validity for military communications tasks than standard speech intelligibility tests based on phonetically balanced words (Moore, 1981). In the CRM task, a listener hears one or more simultaneous phrases of the form "Ready, (call sign), go to (color) (number) now" with one of eight call signs ("Baron," "Charlie," "Ringo," "Eagle," "Arrow," "Hopper," "Tiger," and "Laker"), one of four colors (red, blue, green, white), and one of eight numbers (1-8). The listener's task is to listen for the target sentence containing the pre-assigned call sign (usually "Baron") and respond by identifying the color and number coordinates contained in that target phrase. Although the CRM was originally intended to measure speech intelligibility with a noise masker, its call-sign-based structure makes it ideal for use in multitalker listening tasks. The embedded call sign is the only feature that distinguishes the target phrase from the masking phrases, so the listener is forced to attend to the embedded call signs in all of the simultaneous phrases in order to successfully extract the information contained in the target phrase (Spieth et al., 1954; Abouchacra, Tran, Besing, & Koehnke, 1997). In this regard, it is similar to many command and control tasks where operators are required to monitor multiple simultaneous channels for important information that may originate from any channel in the system. However, because the simple sentence structure and test words provide no syntactic information to the listener, the CRM may not be representative of performance in all communications tasks.

The experiments were conducted using the corpus of CRM speech materials that has been made publicly available in CD-ROM format by researchers at the Air Force Research Laboratory (Bolia, Nelson, Ericson, & Simpson, 2000). This CRM corpus contains all 256 possible CRM phrases (8 call signs X 4 colors X 8 numbers) spoken by eight different talkers (four male, four female). The experiments described in the following sections were conducted using this corpus. In all cases, the stimulus consisted of a combination of a target phrase, which was randomly selected from all of the phrases in the corpus with the call sign "Baron," and one or more masking phrases, which were randomly selected from the phrases in the corpus with different call signs, colors, and numbers than the target phrase. These stimuli were presented over headphones at a comfortable listening level (approximately 70 dB SPL), and the listener's responses were collected either by using the computer mouse to select the appropriately colored number from a matrix of colored numbers on the

CRT or by pressing an appropriately marked key on a standard computer keyboard. Each of the following sections discusses a different factor that influences speech intelligibility in a multitalker listening environment.

3.0 FACTORS INFLUENCING PERFORMANCE IN MULTITALKER DISPLAYS

3.1 Signal-to-Noise Ratio:

One factor that influences the performance of any audio display is the overall noise level in the output signal. In the case of a speech display based on radio communications, three different kinds of noise contribute to this overall noise level: 1) ambient noise in the environment of the talker that is picked up by the microphone that records the talker's voice; 2) electronic noise or distortion in the transmission channel (wireless or wired); and 3) ambient noise in the environment of the listener. Intelligibility is determined by the ratio of the target speech signal to this overall noise level.

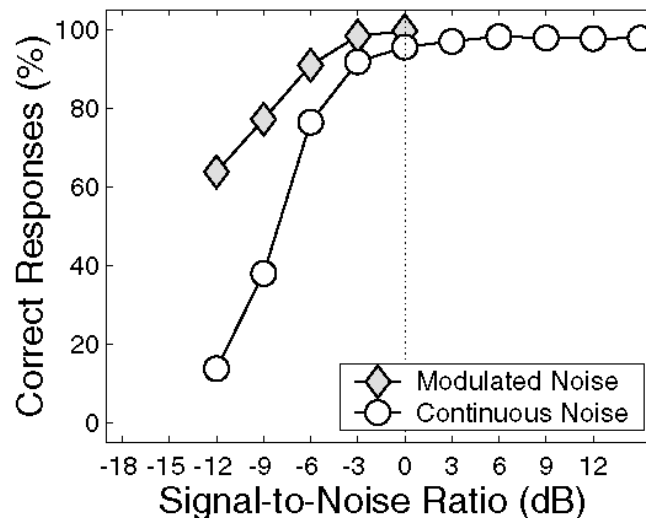


Figure 1: Percentage of correct color and number identifications for a CRM target phrase masked by continuous or modulated speech-shaped noise signal. Adapted from Brungart (2001).

The effects of signal-to-noise ratio (SNR) on speech perception are well documented, and, in many cases, it is possible to use the Articulation Index (AI) or the Speech Transmission Index (STI) to make a quantitative prediction of speech intelligibility directly from the acoustic properties of the noise and speech signals (Kryter, 1962; Steeneken & Houtgast, 1980). In general, the sensitivity of speech intelligibility to the SNR depends on the phonetic structure, vocabulary size, and context of the speech signal. Although the CRM phrases provide no contextual information (it is impossible to predict the color or number in a CRM phrase from any of the other words in the phrase), they are limited to a small vocabulary of colors and numbers. This allows listeners to perform well in the CRM task even at very low SNRs. Figure 1 (adapted from Brungart, 2001a) shows performance in the CRM as a function of SNR (calculated for each stimulus as the ratio of the RMS level measured across the entire individual speech utterance in the stimulus to the long-term RMS level

of the individual noise sample in the stimulus) for a continuous speech-shaped noise (circles) and for a speech-shaped noise that has been modulated to match the envelope of a speech signal from the CRM corpus (diamonds). In each case, both the target speech and the noise were presented diotically, i.e., with the same audio signal presented simultaneously to both ears. The results show that performance in the CRM task is nearly perfect in continuous noise when the SNR is 0 dB or higher, and that performance with a noise masker that is modulated to match the amplitude variations that occur in speech is reasonably good ($> 80\%$) even at an SNR of -6 dB. It should be noted, however, that these surprisingly good results are a direct result of the small vocabulary size in the CRM corpus. The most demanding speech materials (nonsense syllables) require an SNR of approximately +20 dB in the speech band (200 Hz - 6100 Hz) to achieve near-perfect identification performance (Kryter, 1962). Thus an ideal multitalker speech display should be able to achieve an SNR of +20 dB in the frequency range from 200 Hz to 6100 Hz (measured from the overall RMS levels of the speech and noise signals). It should be noted that the relative importance of each frequency range to speech intelligibility has been thoroughly documented in the literature on Articulation Theory. This information is invaluable when tradeoffs between bandwidth and SNR become necessary in the design of communications systems.

3.2 Number of Competing Talkers:

One obvious factor that can affect the performance of a multitalker speech display is the number of competing talkers. As a general rule, performance in a multitalker listening task decreases when the number of talkers increases. Figure 2 (adapted from Brungart, Simpson, Ericson, & Scott, 2001) shows how performance in the CRM task changes as the number of interfering talkers increases from 0 to 3. The data are shown for different same-sex talkers presented at the same level diotically over headphones (Brungart et al., 2001). When no competing talkers were present in the stimulus, performance was near 100%. The first competing talker reduced performance by a factor of approximately 0.4, to 62% correct responses. The second competing talker reduced performance by another factor of 0.4, to 38% correct responses. And the third competing talker reduced performance by another factor of 0.4, to 24% correct responses. Thus we see that CRM performance in a diotic multitalker speech display decreases by approximately 40% for each additional same-sex talker added to the stimulus.

These results clearly show that it is advantageous to reduce the number of simultaneous talkers in a multitalker speech display whenever it is practical to do so. Possible ways to achieve this reduction range from simple protocols that reduce the chances of overlapping speech signals on a radio channel (such as marking the end of each transmission with a terminator like “over”) to systems that allow only one talker to speak on a radio channel at any given time to sophisticated systems that queue incoming messages that overlap in time and play them back to the listener sequentially. However, none of these solutions is appropriate for complex listening situations where a single communication channel is in nearly constant use by two or more simultaneous talkers or where a listener has to monitor two or more communications channels for time-critical information that might occur on any channel. For these situations, the designers of speech displays must rely on other cues to help users segregate the competing speech messages.

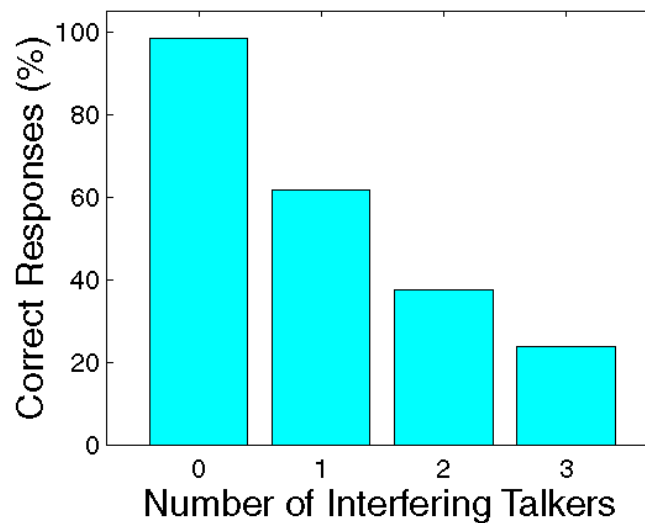


Figure 2: Percentage of correct color and number identifications for a CRM target phrase masked by 0, 1, 2, or 3 simultaneous same-sex masking phrases. All of the competing talkers were presented diotically at the same level. Adapted from Brungart et al. (2001).

3.3 Voice Characteristics:

Differences in voice characteristics provide one audio cue that can be used to segregate competing speech signals. The voices of different talkers can vary in a wide variety of ways, including differences in fundamental frequency (F0), formant frequencies, speaking rate, accent, and intonation. Talkers who are different in sex are particularly easy to distinguish, because on average female talkers have F0 frequencies about two times higher and substantially shorter vocal tracts than male talkers. The shorter vocal tracts of female talkers cause their format center frequencies to be approximately 1.3 times higher than those of male talkers.

Figure 3 (adapted from Brungart et al., 2001) illustrates the effect that differences in voice characteristics can have on a listener's ability to segregate a target speech signal from one, two, or three interfering talkers. The target and masker talkers were randomly selected from the corpus within each block of trials. Thereby, no information about the target or masker talkers' voice characteristics was provided to the listeners. The white bars show performance when the interfering talkers were different in sex than the target talker. The gray bars show performance when the masking phrases were spoken by different talkers who were the same sex as the target talker. The black bars show performance when the target and masking phrases were all spoken by the same talker. In all cases, performance was best when the interfering talkers were different in sex than the target talker and worst when all the phrases were spoken by the same talker. In situations where it is possible to control the voice characteristics of the competing talkers in a multitalker speech display, the characteristics of the competing voices should be made as different as possible. One example of a situation where this should be relatively easy to accomplish is in the use of computer-generated voice icons in an audio display. Consider, for example, a cockpit display where one voice icon might be used to indicate an engine fire and another might be used to indicate a terrain warning. Because the relative priority of these two

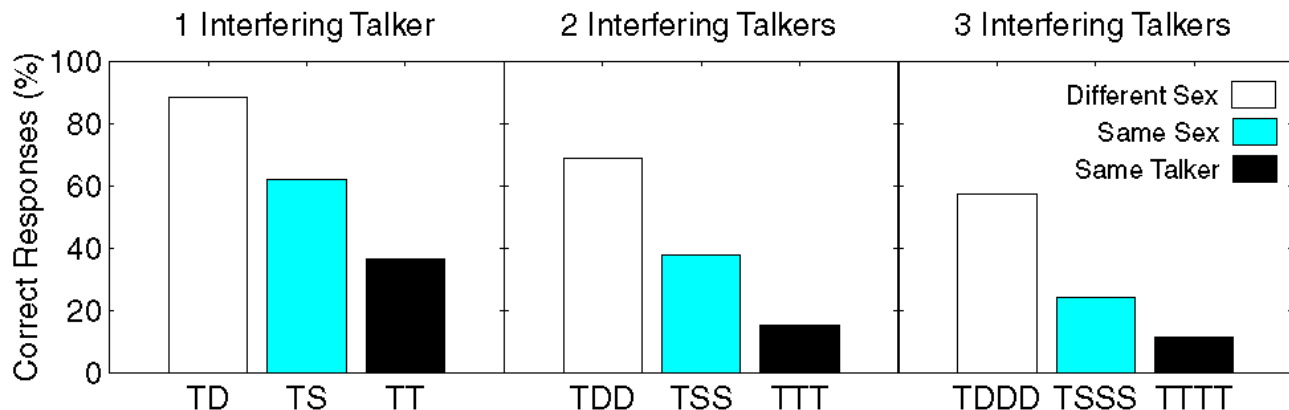


Figure 3: Percentage of correct color and number identifications for a CRM target phrase masked by 1, 2, or 3 simultaneous masking phrases. The white bars show performance with masking talkers who were different in sex than the target talker (the TD condition). The gray bars show performance with different masking talkers who were the same sex as the target talker (the TS condition). The black bars show performance when the target and masking phrases were all spoken by the same talker (the TT condition). All of the competing talkers were presented diotically at the same level. Adapted from Brungart et al. (2001).

warnings can vary with the situation, both of these warnings must be presented to the pilot as soon as they occur. If the two warnings are pre-recorded in both male and female voices, the display system can act to ensure that the two warnings are spoken by different-sex talkers. This would make it easier for the pilot to attend to the warning with greater immediate relevance.

In audio displays that are designed to present externally-generated voice communications rather than internally-generated audio icons, it is much more difficult to control the vocal characteristics of the competing talkers. One possible option is to perform some kind of real-time or near-real-time audio processing on the different competing voice signals to make them more distinct. It may be possible to achieve this result by manipulating the parameters used to reconstruct the voice in communication systems that use low-bandwidth parametric vocoders. For example, the fundamental frequencies (F0s) of the two talkers could be manipulated to introduce a difference between the two competing talkers in real time. Assman and Summerfield (1990) have shown that a difference in F0 of 1/6th of one octave is sufficient to produce a significant improvement in intelligibility. However, this approach also has a major drawback: it may make it substantially more difficult (or impossible) for the listener to use voice characteristics to determine the identity of the talker. Thus the segregation efficiency that is gained by introducing differences in voice characteristics may be more than offset by the reduction in a listener's ability to correctly identify the target talker. A good rule of thumb might be to restrict the use of voice modification to situations in which speaker identification is not important and avoid the use of voice modification when accurate speaker identification is critical. Note also that care must be taken to ensure that voice characteristics such as formant frequencies are not changed enough to degrade the intelligibility of the speech.

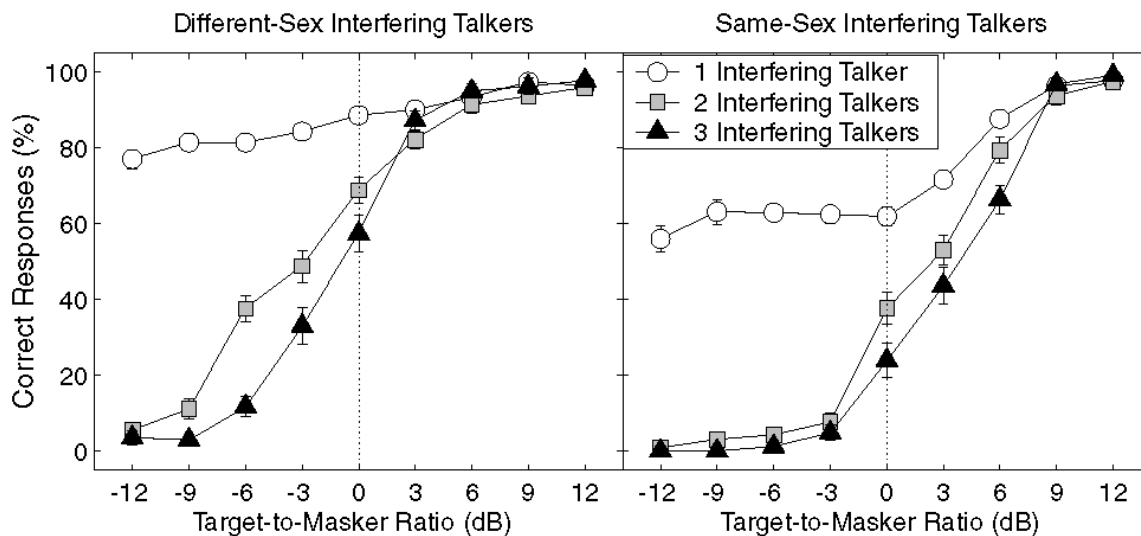


Figure 4: Percentage of correct color and number identifications for a CRM target phrase masked by 1, 2, or 3 interfering talkers. The results are shown as a function of the target-to-masker ratio (TMR), which is the ratio of the level of the target talker to the level of each of the other interfering talkers in the stimulus (note that all the interfering talkers were presented at the same level). The left panel shows performance with different-sex interfering talkers; the right panel shows performance with same-sex interfering talkers. The error bars show the 95% confidence intervals of each data point. Adapted from Brungart et al. (2001).

3.4 Target-to-Masker Ratio:

Another factor that has a strong influence on a listener's ability to segregate competing speech signals is the level of the target talker relative to the competing talkers. In general, it is much easier to attend to the louder talker in a multitalker stimulus than to the quieter talker in a multitalker stimulus. This is illustrated in Figure 4, which shows performance as a function of the target-to-masker ratio (TMR) for 1, 2, or 3 interfering talkers. In this context, TMR is the ratio of the overall RMS level of the target talker to the overall RMS level of each of the interfering talkers in the stimulus. Thus, when the TMR is 0 dB, all of the talkers in the stimulus are presented at the same level. The results in Figure 4 show that performance is substantially improved when the target talker is the most intense talker in the stimulus (TMR > 0 dB).

Clearly a substantial improvement in speech intelligibility can be achieved by increasing the level of the target talker relative to the levels of the other talkers in the stimulus. Unfortunately, this also degrades the intelligibility of the other talkers in the stimulus. Because it is usually difficult or impossible for the audio display designer to identify the target talker in the stimulus, there is no way to automatically determine which talker should be amplified relative the others. One alternative approach is to allow the listener to adjust the relative levels of the talkers and thus increase the level of the talker who is believed to be the most important in the current listening situation (Spieth et al., 1954). This ability is provided by current multichannel radio systems, which typically have adjustable volume-level knobs for each radio channel. It should be noted, however, that a potential drawback of this approach is that the listener will miss crucial information that is spoken by one of the low-level talkers in the stimulus: the data in Figure 4 show that performance decreases rapidly with TMR when there are two or more interfering talkers and that listeners essentially receive no semantic information from the low-level talkers when the TMR falls below -6 dB, or below 0 dB for same-sex talkers.

Improving Multitalker Speech Communication with Advanced Audio Displays

The data for the situation with one same-sex interfering talker (open circles in the right panel of Figure 4) have some interesting implications for the design of two-channel communications systems. In this condition, listeners were apparently able to selectively attend to the quieter talker in the stimulus. Consequently, performance in this condition did not decline when the TMR was reduced below 0 dB. Performance did, however, improve rapidly when the TMR was increased above 0 dB. Although one might intuitively expect that two equally important communications channels should be presented at the same level, the data in Figure 4 (adapted from Brungart et al., 2001) suggest that this is a poor strategy. When a level difference is introduced between the two channels, performance improves substantially when the target talker occurs on the louder channel, but is unaffected when the target talker occurs on the quieter channel. Thus, overall performance in the CRM task improves substantially when the speech stimuli are presented at levels that differ by 3-9 dB. These data are also consistent with results of a previous experiment that examined performance as a function of TMR with a different call-sign-based task (Egan, Carterette, & Thwing, 1954). Note that this strategy appears to improve performance only with same-sex or same-talker interfering speech signals, and that it provides much less benefit with different-sex interfering talkers where differences in voice characteristics seem to dominate any segregation cues provided by differences in the levels of the two talkers. The introduction of level differences may also fail to improve intelligibility in noisy environments, where the less-intense talker may be masked by ambient noise. Level differences should also be avoided in cases where there is more than one interfering talker and intelligibility falls off rapidly with decreasing TMR (Figure 4). Further investigation is needed to explore these level-difference segregation cues in more detail.

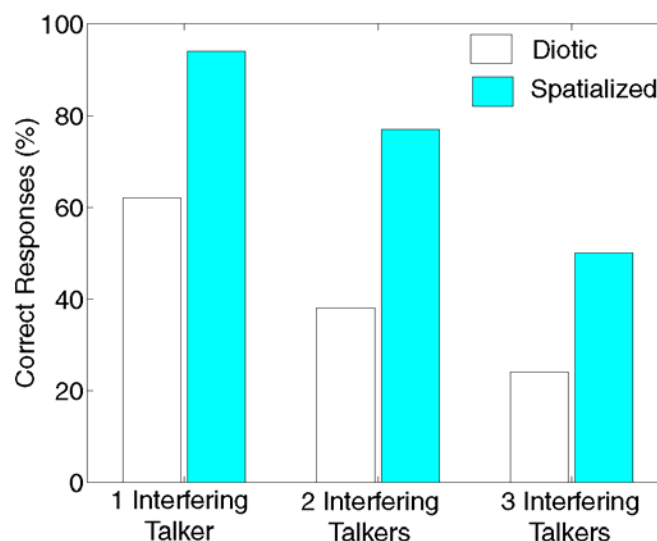


Figure 5: Percentage of correct color and number identifications for a CRM target phrase masked by 1, 2, or 3 same-sex interfering talkers. The white bars show results for a diotic condition where the competing talkers were not spatially separated (adapted from Brungart et al., 2001). The gray bars show performance where the competing talkers were spatially separated by 45° (talkers at 0° and 45° with one interfering talker; -45°, 0° and 45° with two interfering talkers; and -45°, 0°, 45° and 90° with three interfering talkers). The spatialized results have been averaged across all the different possible target talker locations in each configuration.

3.5 Spatial Separation:

In intercom systems where it is possible to use stereo headphones to present a binaural audio signal to the listener, substantial performance benefits can be achieved by using a virtual audio display to spatially separate the apparent locations of the competing sounds (Crispien & Ehrenberg, 1995; Abouchacra et al., 1997; Ericson & McKinley, 1997; Nelson, Bolia, Ericson, & McKinley, 1999). Figure 5 shows the effect of spatial separation on overall performance in the CRM task with one, two, or three same-sex interfering talkers. These data have been taken from an experiment where the target and masking talkers were always separated by 45 degrees (Ericson et al., 2004). In the case with one interfering talker, spatial separation increased performance by approximately 25 percentage points. In the cases with two or three interfering talkers, spatial separation nearly doubled the percentage of correct responses. These results clearly illustrate the substantial performance advantages that spatial separation in azimuth can produce in multitalker audio displays.

3.5.1 Target Location

In listening situations where the target and masking talkers are spatially separated, the actual level of intelligibility achieved at any given time will vary according to where the target talker is located and how large the angular separation is between the target talker and the other simultaneous talkers on the intercom. The left panel of Figure 6 shows two possible spatial configurations for a three-channel spatial intercom: a 3-Talker Close configuration, where the competing talkers are separated by 15 degrees, and a 3-Talker Far configuration, where the talkers are separated by 90 degrees. The middle panel of the figure shows mean performance in the CRM task as a function of the location of the target talker (Brungart et al., 2005). Note the following three main points from these results:

- Overall performance is generally better for widely-separated talker locations than it is for closely-spaced talker locations.
- Performance is generally better when the target talker is located at the leftmost or rightmost talker location than when it is located at a central location. This occurs because the target talker is the most intense talker in either the left ear or the right ear when it is located at the extreme leftmost or rightmost position, but is less intense than at least one talker in both ears when it is located at a central location (e.g., see Zurek, 1993).
- Performance is generally better when the target talker is located in the right hemisphere than when it is located in the left hemisphere. This is believed to be related to the specialization of speech processing centers in the left hemisphere of the human brain (Bolia et al., 2001).

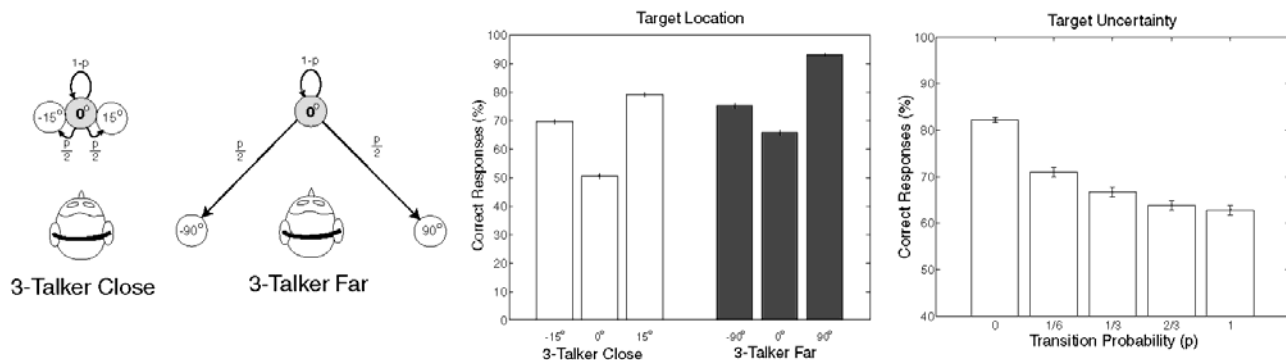


Figure 6: Effects of Spatial Location and Target Uncertainty on Multitalker Listening Performance. The left panel of the figure illustrates two 3-talker spatial listening environments, one with the competing talkers separated by 15 degrees and one with the competing talkers separated by 90 degrees. The middle panel shows the percentage of correct responses in the CRM task as a function of the target talker location in each of these configurations. The right panel shows performance as a function of the level of uncertainty the listener has about the location of the target talker on any given trial, which is expressed by the probability p that the target talker will change from its current location to a different location at the end of any given trial in a block of trials in the experiment (as illustrated in the form of a Markov diagram by the arrows in the left panel of the figure). The data have been adapted from Brungart and Simpson, 2005.

3.5.2 Target Uncertainty

Another factor that influences performance in multitalker communications tasks that involve a spatialized intercom system is the amount of *a priori* information the listener has about where the most important information will originate at any given time. In general, listeners benefit more from the spatial separation of competing channels of speech when they know which talker to listen to than when they have to scan their attention across multiple talkers to determine where the most pertinent information is located. One way to measure the effect this *a priori* information has on performance is to perform a multitalker listening experiment where the likelihood that the target talker will move from its current location to a new location at the end of any given trial is systematically varied across blocks. The arrows in the left panel of Figure 6 illustrate this kind of dynamic listening environment in the form of a Markov diagram where the number associated with each arrow represents the probability that the system state will change in a particular way at the end of a single trial of the experiment. The bars in the right panel of Figure 6 show how performance varies in the 3-talker CRM task as a function of the total transition probability p of the target talker moving from its current location to a new location at the end of each trial. These results show that listeners perform best when the target talker is located in a fixed position ($p=0$), that their overall performance degrades substantially (by about 12 percentage points in this case) when a small amount of uncertainty is added to their system, and that their performance degrades more gradually as additional uncertainty is added up to the point where the target talker location changes in every trial of the experiment ($p=1$) (Brungart and Simpson, 2005).

3.5.4 Headtracking

A final factor that has the potential to influence performance in spatialized multitalker speech displays is the integration of a real-time headtracking device that updates the apparent locations of the competing talkers in the system to compensate for the movements of the listener's head. This kind of head-coupling is known to improve the fidelity, realism, and localizability of virtual sound sources (e.g., see Wallach 1940), so it might be reasonable to expect it also to improve performance in multitalker listening tasks. However, recent experiments in our laboratory (Brungart and Simpson, 2004) have shown that real-time headtracking provides, at most, a modest performance benefit in multitalker listening tasks, and that these benefits are limited to those cases where the target talker location changes infrequently (i.e., the transition probability is low) and the competing talkers are widely separated in azimuth. Because headtracking is quite expensive to implement, these results suggest that it is not likely to be cost effective to implement in military communications applications that do not already have access to real-time head position information for some non-communication related purpose (such as a helmet-mounted display in an aircraft cockpit).

4.0 AN OPTIMIZED SPATIAL CONFIGURATION FOR SEVEN TALKERS

4.1 Choosing an Optimal Spatial Configuration for a Multitalker Display:

Table 1: Summary of locations used in previous multitalker speech displays

Study	# of Talkers	Talker Locations
1) Cherry (1953)	2	Non-spatial (left ear only, right ear only)
2) Triesman (1964)	3	Non-spatial (left ear only, right ear only, both ears)
3) Moray et al. (1964)	4	Non-spatial (L only, 2/3 L+1/3 R; 1/3 L + 2/3 R; R only)
4) Khoenke et al. (1998)	4	-90, -45, +45, +90 azimuth
5) Spieth et al. (1954)	3	-20, 0, 20 azimuth or -90, 0, 90 azimuth
6) Drullman & Bronkhorst (2000)	4	-90, -45, 0, +45, +90
7) Yost (1996)	7(3)	-90, -60, -30, 0, +30, +60, +90 azimuth
8) Hawley et al. (1999)	7(2-4)	-90, -60, -30, 0, +30, +60, +90 azimuth
9) Crispian & Ehrenberg (1995)	4	-90 az, +60 el; -30 az, +20 el; -30 az, -20 el; -90 az, -60 el
10) Nelson et al. (1999)	8(2-8)	6: -90, -70, -31, +31, +70, +90 7: -90, -69, -45, 0, +45, +69, +90 8: -90, -69, -45, -11, +11, +45, +69, +90 azimuth
11) Simpson et al. (1998)	8(2-8)	7: -90, -69, -135, 0, +135, +69, +90 8: -90, -69, -135, -11, +11, +135, +69, +90 azimuth
12) Ericson & McKinley (1997)	4	-135, -45, +45, +135 azimuth (w/headtracking)
13) Brungart & Simpson (2002)	2	90 degrees azimuth, 1 m; 90 degrees azimuth, 12 cm

Although a number of researchers have demonstrated the advantages of spatial filtering for multitalker speech perception, very little effort has been made to systematically develop an optimal set of HRTF filters capable of maximizing the number of talkers a listener can simultaneously monitor while minimizing the amount of interference between the different competing talkers in the system. Most systems that have used HRTF filters to spatially separate speech channels have placed the competing channels at roughly equally-spaced intervals in azimuth in the listener's frontal plane. Table 1 provides examples of the spatial separations used in previous multitalker speech displays. The first three entries in the table represent early systems that used stereo panning over headphones rather than head-related transfer functions to spatially separate the signals. This method has been shown to be very effective for the segregation of two talkers (where the talkers are presented to the left and right earphone), somewhat effective for the segregation of three

Improving Multitalker Speech Communication with Advanced Audio Displays

talkers (where one talker is presented to the left ear, one talker is presented to the right ear, and one talker is presented to both ears), and only moderately effective in the segregation of four talkers (where two talkers are presented to the left and right ears, one talker is presented more loudly in the left ear than in the right ear, and one talker is presented more loudly in the right ear than the left ear). However, these panning methods have not been shown to be effective in multitalker listening configurations with more than four talkers.

The other entries in the table represent more recent implementations that either used loudspeakers to spatially separate the competing speech signals or used HRTFs that accurately reproduced the interaural time and intensity difference cues that occur when real sound sources are spatially separated around the listener's head. The majority of these implementations (entries 4-8 in Table 1) have used talker locations that were equally spaced in the azimuth across the listener's frontal plane. One implementation (entry 9 in Table 1) has spatially separated the speech signals in elevation as well as azimuth, varying from +60 degrees elevation to -60 degrees elevation as the source location moves from left to right. And two implementations (entries 10 and 11 in Table 1) have used a location selection mechanism proposed by Nelson et al. (1999) that selects talker locations in a procedure designed to maximize the difference in source midline distance (SML) between the different talkers in the stimulus.

Recently, a new talker configuration has been proposed in which the target and masking talkers are located at different distances (12 cm and 1 m) at the same angle in azimuth (90 degrees) (entry 13 in Table 1). This spatial configuration has been shown to work well in situations with only two interfering talkers, but it has not been tested with more than two competing talkers.

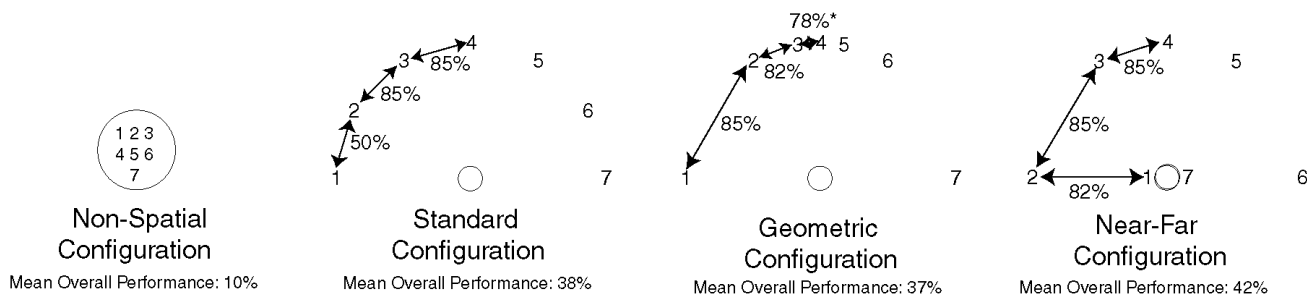


Figure 6: Comparison of performance for different spatial configurations in a multitalker speech display system with seven simultaneous talkers. The leftmost panel shows a standard monaural configuration where all of the talkers are presented diotically and appear to originate from the center of the listener's head. The next panel shows a standard spatial configuration where the talkers are equally spaced 30 degrees apart in azimuth. The third panel shows a geometric configuration where the talkers are spaced closely together in front of the listener (where spatial resolution is greatest) and further apart to the sides (where spatial resolution is worst). The right panel shows a hybrid near-far configuration with five geometrically-spaced far-field talkers plus one nearby talker in each of the listener's ears. The arrows show mean performance in a two-talker task CRM with competing talkers at each end of the arrows, and the overall performance numbers at the bottom show performance in a CRM task with seven simultaneous same-sex talkers (adapted from Brungart and Simpson, 2003).

While it is possible to make theoretical arguments in favor of each of these possible talker configurations, we know of no previous studies that have objectively measured speech intelligibility as a function of the placement of the competing talkers. However, recent results in our laboratory have shown that equal spacing in azimuth cannot produce optimal performance in systems with more than five possible talker locations. The reason for this stems from the fact that listeners are generally more sensitive to changes in the spatial locations of objects in front of them than they are to changes in the spatial locations of objects located off to the sides (Mills, 1950). Thus it turns out that, in a seven-talker configuration with the talkers spread out equally every 30 degrees in the listener's front hemifield (the "standard" configuration in Figure 6), listeners are generally good at segregating speech signals that happen to occur from adjacent talkers near 0 degrees azimuth [as indicated by the 85% correct performance level achieved in a two-talker listening test with talkers at locations 3 and 4 (indicated by the arrow in the figure)], but they are very poor at segregating adjacent competing talkers at lateral locations (as indicated by the 50% performance level for the 2-talker CRM task with talkers at locations 1 and 2). This inadequate spatial resolution at lateral source positions suggests that better overall performance might be achieved with a "geometrically" spaced source configuration with closely-spaced talker locations near the midline and larger angular separations between the talkers at more lateral locations (third panel of Figure 6). Such a configuration does indeed result in more uniform (and better) segregation performance for pairs of adjacent talker locations, but unfortunately it does not result in improved performance when all seven competing talkers are active at the same time (indicated by the overall performance scores underneath each panel in Figure 6).

Significantly better overall performance *can*, however, be achieved with the seven-talker hybrid near-far configuration shown in the rightmost panel of Figure 6. This configuration takes advantage of the fact that listeners can use interaural level differences to distinguish between nearby and distant talkers located along the interaural axis (Brungart and Simpson, 2002), and provides a configuration that produces both a high level of performance for pairs of adjacent talkers in a two-talker CRM task and a roughly 10% improvement in overall performance (relative to the standard configuration) in a seven-talker CRM task. Our research indicates that this configuration represents a near-optimal configuration for a multitalker display with more than five simultaneous talkers.

4.2 The Benefits of Spatial Audio in a Realistic Military Communications Environment:

While there is general consensus that spatialized audio can improve multitalker listening performance in laboratory settings, it is not necessarily always clear that these benefits will transfer in a useful way to operational military environments, where the communications situation is dynamically changing and the number of simultaneous talkers varies across time. Also, while it is indisputably true that spatialized audio can improve multitalker intelligibility relative to a monaural speech display that provides the same signal to both of the listener's ears, it is not at all obvious that spatialized displays will provide a compelling benefit over *dichotic* speech displays that provide the listener with the option of hearing competing radio channels in the left ear only, the right ear only, or in both ears simultaneously (which effect causes them to be perceived in the "center of the head"). In order to test these situations, we conducted an experiment in our laboratory that compared the performance of monaural, dichotic, and spatialized speech displays in a seven-talker CRM listening task where each talker was 50% likely to be active on any given trial and the target phrase had a 25% probability of changing to a different talker at a different location at the end of each

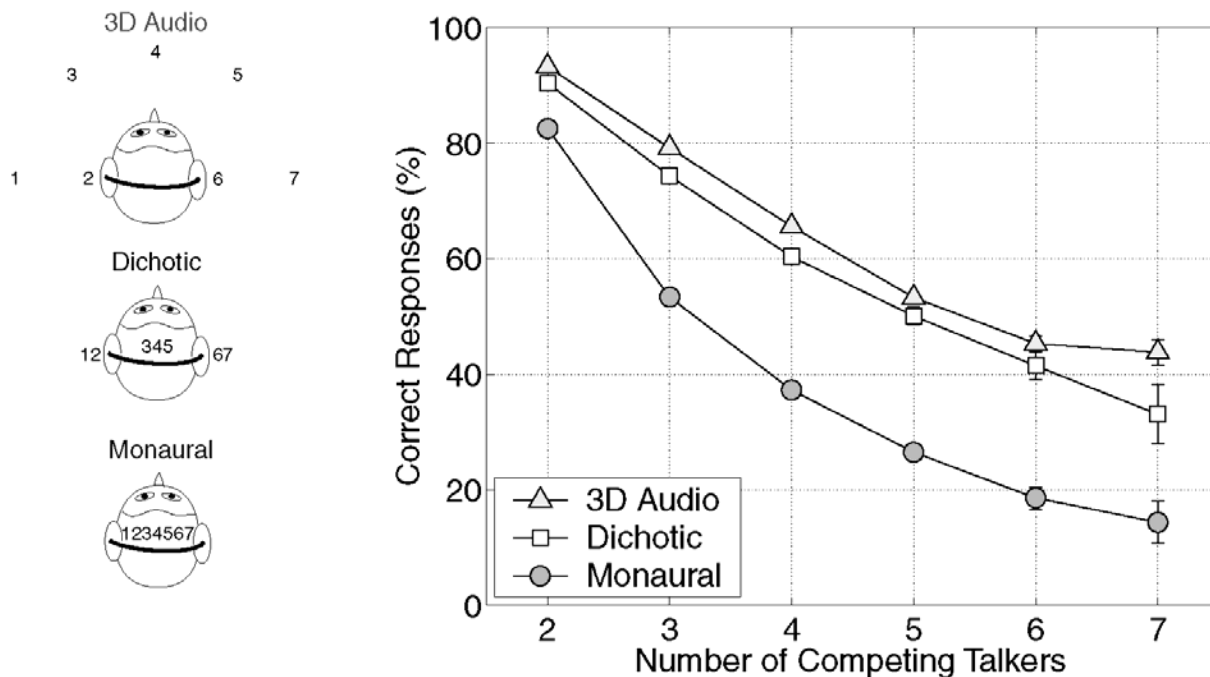


Figure 7: Comparison of overall performance with spatialized, dichotic, and monaural audio displays in a seven-talker CRM listening task with all male talkers. Each talker was 50% likely to be active in each trial, and there was a 25% chance that the target phrase would move to a different talker location at the end of each stimulus presentation. The results have been plotted as a function of the number of active talkers in the trial. The error bars represent the 95% confidence intervals for each data point.

stimulus presentation. The results demonstrate that the spatialized audio display improved performance by about 30 percentage points for all the listening situations with more than two competing talkers. They also show that the 3D Audio display consistently improved performance by about 5 percentage points relative to the dichotic speech display, even in situations where only a small number of the speech channels were active. This seems to confirm that the additional expense of upgrading a dichotic speech display into a spatialized speech display is justified in situations where accurate communications are necessary to ensure successful accomplishment of the mission.

4.0 CONCLUSIONS

In the design of military intercom systems intended to accommodate more than one simultaneous channel of speech communications, virtual synthesis techniques that spatially separate the apparent locations of the competing talkers are an effective and affordable way to improve communications efficiency and enhance warfighter performance. The data from Figure 5 show that spatially separating same-sex competing talkers by 45° produced a 25-35 percentage point increase in overall performance in the CRM task. In terms of the other

factors examined in this paper, this is roughly equivalent to: 1) reducing the number of competing talkers in the stimulus by 1 to 1.5 talkers (Figure 2); 2) replacing the same-sex interfering talkers with different-sex interfering talkers (Figure 3); or 3) increasing the target-to-masker ratio by 3-9 dB (Figure 4). However, spatial separation has substantial advantages over these other techniques. The biggest advantage is that spatial separation improves the intelligibility of all the talkers in the stimulus, while the other techniques tend to increase the intelligibility of only one of a few selected talkers. Reducing the number of talkers in the stimulus increases the intelligibility of the remaining talkers at the expense of losing all the information from the eliminated talker. Replacing the same-sex interfering talkers with different-sex talkers provides a benefit only for the talker who is different in sex from the other talkers in the stimulus. Increasing the target-to-masker ratio increases the intelligibility of one talker but generally reduces the intelligibility of the other talkers in the stimulus when there are more than two talkers. Only spatial separation is able to improve overall performance across all the talkers in a three- to four-talker stimulus. Spatial separation is also relatively inexpensive to implement in multitalker speech displays. Many of the benefits of spatially separating speech signals can be obtained with relatively simple digital signal processing techniques that do little more than introduce interaural time differences (Carhart, Tillman, & Johnson, 1967) and interaural level differences (Bronkhorst & Plomp, 1988) into the different communications channels of the system. The listener-specific pinna-related spectral details that are required to produce realistic, localizable, externalized virtual sounds in non-speech virtual displays (Wenzel, Arruda, Kistler, & Wightman, 1993) simply do not provide any additional benefit to speech intelligibility in multitalker listening tasks for presentation in azimuth (Nelson et al., 1999; Drullman & Bronkhorst, 2000). Similarly, real-time headtracking devices are not required to achieve good intelligibility in multitalker speech displays (the data shown in Figure 5 were collected without any head tracking). If a communications system or intercom is capable of processing audio signals in the digital domain, it may be possible to implement an effective speech segregation algorithm in software for little or no additional cost. The only restriction is that the system must be capable of producing a stereo output signal: no reliable spatialization cues are possible in a system with only one analog output channel, and unfortunately this can be a severe impediment in efforts to attempt to upgrade legacy intercom systems to accommodate spatial audio. However, on the basis of the results presented here, it is clear that spatial audio should be given serious consideration both in the design of new military intercom systems and in the upgrading of existing systems to accommodate evolving C4ISR requirements.

5.0 REFERENCES

- Abouchacra, K., Tran, T., Besing, J., & Koehnke, J. (1997). Performance on a selective attention task as a function of stimulus presentation mode. Proceedings of the Midwinter Meeting of the Association for Research in Otolaryngology, St. Petersburg Beach, Florida.
- Assman, P. F. & Summerfield, Q. (1990). Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies. *Journal of the Acoustical Society of America*, 88, 680--697.
- Begault, D. R. (1999). Virtual Acoustic Displays for Teleconferencing: Intelligibility Advantage for 'Telephone-Grade' Audio. *Journal of the Audio Engineering Society*, 47, 824--828.
- Bolia, R., Nelson, W., Ericson, M., & Simpson, B. (2000). A speech corpus for multitalker communications research. *Journal of the Acoustical Society of America*, 107, 1065--1066.

- Bolia, R. S., Nelson, W. T., & Morley, R. M. (2001). Asymmetric performance in the cocktail party effect: Implications for the design of spatial audio displays. *Human Factors*, 43, 208–216.
- Bronkhorst, A. & Plomp, R. (1988). The effect of head-induced interaural time and level difference on speech intelligibility in noise. *Journal of the Acoustical Society of America*, 83, 1508–1516.
- Brungart, D. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *Journal of the Acoustical Society of America*, 109, 1101–1109.
- Brungart, D. & Simpson, B. (2001). Optimizing multitalker speech displays with near-field HRTFs. In *Proceedings of the 2001 International Conference on Auditory Display*, Espoo, Finland, July 29-August 1, 2001, pp. 169–174.
- Brungart, D. & Simpson, B. (2002). The effects of spatial separation in distance on the informational and energetic masking of a nearby speech signal. *Journal of the Acoustical Society of America*, 112, 664–676.
- Brungart, D.S., Simpson, B.D., Kordik, A.J. and McKinley, R.L. (2003). The impact of headtracking on intelligibility in a multitalker display. *Proceedings of HFES 2003*, Denver, Colorado, October, 2003.
- Brungart, D.S., and Simpson, B.D. (2003). Optimizing the spatial configuration of a seven-talker speech display. *Proceedings of the 2003 International Conference on Auditory Display*, Boston, Massachusetts, July 6-9, 2003, pp. 188-191.
- Brungart, D., Simpson, B., Ericson, M., & Scott, K. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *Journal of the Acoustical Society of America*, 110, 2527–2538.
- Carhart, R., Tillman, T., & Johnson, K. (1967). Release of masking for speech through interaural time delay. *Journal of the Acoustical Society of America*, 42, 124–138.
- Cherry, E. (1953). Some experiments on the recognition of speech, with one and two ears. *Journal of the Acoustical Society of America*, 25, 975–979.
- Crispien, K. & Ehrenberg, T. (1995). Evaluation of the 'Cocktail Party Effect' for Multiple Speech Stimuli within a Spatial Audio Display. *Journal of the Audio Engineering Society*, 43, 932–940.
- Drullman, R. & Bronkhorst, A. (2000). Multichannel speech intelligibility and talker recognition using monaural, binaural, and three-dimensional auditory presentation. *Journal of the Acoustical Society of America*, 107, 2224–2235.
- Egan, J., Carterette, E., & Thwing, E. (1954). Factors affecting multi-channel listening. *Journal of the Acoustical Society of America*, 26, 774–782.
- Ericson, M.A., Brungart, D.S. and Simpson, B.D. (2004). Factors that influence intelligibility in multitalker speech displays. *International Journal of Aviation Psychology*. 14(3), 313-333.

- Ericson, M. & McKinley, R. (1997). The intelligibility of multiple talkers spatially separated in noise. In *Binaural and Spatial Hearing in Real and Virtual Environments*. Edited by R.H. Gilkey and T.R. Anderson (Erlbaum, Hillsdale N.J.), pp. 701-724.
- Hawley, M., Litovsky, R., and Colburn, H. (1999). "Speech intelligibility and localization in a multi-source environment," *J. Acoust. Soc. Am.* 105, 3436-3448.
- Kryter, K. (1962). Methods for calculation and use of the articulation index. *Journal of the Acoustical Society of America*, 34, 1689-1697.
- MIL-STD-1472E (1997). Department of Defense Design Criteria Standard. Department of Defense.
- Mills, A. (1958). On the minimum audible angle. *Journal of the Acoustical Society of America*, 30, 237-246.
- Moore, T. (1981). Voice communication jamming research. In AGARD Conference Proceedings 331: Aural Communication in Aviation, pp. 2:1--2:6. Neuilly-Sur-Seine, France.
- Moray, N., Bates, A. and Barnett, T. (1964). Experiments on the four-eared man. *Journal of the Acoustical Society of America*, 38, 196-201.
- Nelson, W. T., Bolia, R. S., Ericson, M. A., & McKinley, R. L. (1999). Spatial audio displays for speech communication. A comparison of free-field and virtual sources. *Proceedings of the 43rd Meeting of the Human Factors and Ergonomics Society*, 1202--1205.
- Spieth, W., Curtis, J., & Webster, J. (1954). Responding to one of two simultaneous messages. *Journal of the Acoustical Society of America*, 26, 391--396.
- Steeneken, H. J. M. & Houtgast, T. (1980). A physical method for measuring speech-transmission quality. *Journal of the Acoustical Society of America*, 67, 318--326.
- Tun, P. & Wingfield, A. (1994). Speech recall under heavy load conditions: Age, predictability and limits on dual-task interference. *Aging, Neuroscience, and Cognition*, 1, 29--44.
- Triesman, A. (1964). The effect of irrelevant material on the efficiency of dichotic listening. *American Journal of Psychology*, 77, 533-546.
- Wenzel, E., Arruda, M., Kistler, D., & Wightman, F. (1993). Localization using non-individualized head-related transfer functions. *Journal of the Acoustical Society of America*, 94, 111--123.
- Zurek, P. M. (1993). Binaural advantages and directional effects in speech intelligibility. In G. Studebaker & I. Hochberg (Eds.), *Acoustical factors affecting hearing aid performance, Second Edition*. Portland: Allyn and Bacon.
- Yost, W. (1997). "The cocktail party problem: Forty years later," in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson. Erlbaum, Hillsdale, NJ, pp. 329-348.



Detection and Identification of Simultaneous Communications in a Simulated Flying Task

Ken McAnally, Russell Martin, Jodie Doman, Geoff Eberle & Simon Parker

Air Operations Division, Defence Science and Technology Organisation,
PO Box 4331, Melbourne 3001
Australia

ken.mcanally@dsto.defence.gov.au

SUMMARY

Operators of military flight vehicles are often required to attend to more than one source of communications signal. Previous research has shown that the intelligibility of a speech message in a background of speech distractors is improved if the signals are presented using a 3-dimensional (3-d) audio display rather than the traditional diotic configuration. However, whether infrequent target messages (e.g., callsigns) are more reliably detected in a continuous monitoring task with high temporal uncertainty when using a 3-d audio display has not been examined. This study examined participants' ability to detect a target callsign and identify a colour/number combination associated with it while engaged in a 20-minute, simulated formation-flying task. Participants were required to monitor 5 communications channels in each of which messages were presented at random intervals. (On average, 2.4 channels were simultaneously active.) Thirty targets were presented over the 20-minute period. There were three audio display conditions: diotic, all channels in front, and channels separated in azimuth (3-d). Detection of target callsigns was significantly higher in the 3-d condition compared to the other conditions. Detections and false alarms were combined to calculate sensitivity and criterion measures using signal detection theory. Sensitivity was significantly higher in the 3-d condition compared with the other conditions, but there were no differences in criterion. Also, consistent with previous results, correct identification of the target number/colour combination was significantly higher in the 3-d condition compared with the other conditions.

1.0 INTRODUCTION

Previous research has shown that spatial separation of competing speech signals increases intelligibility compared with that associated with co-located talkers (see [1] for a review). A popular paradigm for conducting research on intelligibility under conditions of simultaneous communication is the coordinate response measure (CRM)[2]. A speech corpus comprising spoken messages of the form "Ready (callsign) go to (colour) (number) now" has been released to facilitate research using this paradigm [3]. In the most commonly adopted form of the paradigm, listeners are instructed to attend to the message addressed to a given callsign (e.g., Baron) and respond to the colour/number combination contained in that message, while ignoring distracting messages. This task involves a high degree of temporal certainty because the target and distractor messages are begun simultaneously in order to maximise informational masking. However, in most operational scenarios the timing of target messages is highly uncertain and operators have to monitor communications channels over long periods.

This study adopted a paradigm in which the timing of messages was random and the frequency of target messages was low, such that at any time there was high uncertainty about the presence of a target message. Participants were engaged in a simulated formation-flying task to ensure that their background workload was similar to that potentially present in operational environments.

McAnally, K.; Martin, R.; Doman, J.; Eberle, G.; Parker, S. (2005) Detection and Identification of Simultaneous Communications in a Simulated Flying Task. In *New Directions for Improving Audio Effectiveness* (pp. 31-1 – 31-6). Meeting Proceedings RTO-MP-HFM-123, Paper 31. Neuilly-sur-Seine, France: RTO. Available from: <http://www.rto.nato.int/abstracts.aps>.

2.0 METHODS

2.1 Participants

Six male volunteers aged from 26 to 46 years (mean age 34.5 years) recruited from the Defence Science and Technology Organisation participated in the study. Four had some previous flight experience, and all had experience with PC-based flight simulators. All had normal hearing and normal or corrected-to-normal vision.

2.2 Experimental Design

Each of three conditions of audio display (diotic, in-front, and 3-d) was evaluated during two 20-minute flight simulations for each participant. In the diotic condition, all stimuli were presented without any spatial processing. In the in-front condition, all stimuli were filtered by the individual's head-related transfer functions (HRTFs) for the location directly in front (i.e., 0 degrees of azimuth and elevation). In the 3-d condition, each of five channels of audio was spatialised to a different location (-90,0 -40,0 0,0 40,0 or 90,0 degrees of azimuth, elevation) using individualised HRTFs.

The order of conditions was balanced across participants using all possible orders. Conditions were also balanced within participants by presenting the second block of conditions in reversed order. Differences in performance across conditions were assessed using repeated-measures analyses of variance (ANOVAs) employing Greenhouse-Geisser corrections (where $df > 1$) and a criterion of significance of .05.

2.3 Flight task

Each participant completed six simulation runs, each around 20 minutes in duration. Participants were required to follow a lead aircraft (an F-111) that performed gentle (30-degree bank) turns and gentle climbs and dives. The measure of performance for the flight task was the proportion of time the participant was able to keep the lead aircraft within a 7.5-degree circular reticule centred on the ownship's x-axis. Workload was adjusted for each participant by changing the following distance (between 200 and 400 m) so that the lead aircraft was maintained in the reticule for up to about 95 % of the time of the simulation. One of six scripts specifying the track of the lead aircraft was randomly assigned to each simulation run.

The out-of-the-window display was generated and rendered at 60 Hz (Silicon Graphics, Onyx), and projected onto a 4-channel (i.e., left side, front, right side, above) cube display. Participants sat on a seat within the cube and flew using both the throttle and the stick. Head-down displays of flight instruments and a situation awareness display were presented.

2.4 Stimuli

Speech stimuli were sentences taken from a speech corpus [3]. Sentences were of the form "Ready Baron go to red three now", where the callsign (in this case Baron), the colour and the number varied across items.

Participants were required to monitor five channels of communication. The delay between messages in each channel was randomly selected from the range encompassing 1.3 to 4 seconds. This resulted in a talker density that varied from 0 to 5 simultaneous talkers with an average of 2.4. The target callsign to be monitored was "Baron". There were 30 target messages in each 20-minute flight simulation, with a minimum interval between target messages of 20 seconds. Thus, the frequency of target calls was low (only 2.5 % of all messages were targets) and temporal uncertainty was high. The talker, colour and number for the target calls were chosen randomly from those available in the corpus. All distractor messages were addressed to callsigns other than "Baron". The talker, callsign (other than "Baron"), colour and number for distractor messages were chosen at random.

To simulate the effects of transmission through a radio channel, stimuli were low-pass filtered at 4 kHz and clipped following amplification by 20 dB.

2.5 3-Dimensional Audio

Individualised HRTFs were recorded following the procedure previously described [4]. Briefly, miniature microphones were placed in the participant's ear canals and test signals were played from a loudspeaker that could be positioned in space around the participant. The recorded signals were then processed to derive the filtering effect of the head, torso and external ears (i.e., the HRTFs). The HRTFs were corrected for the transfer functions of the loudspeaker, microphones and headphones that were used later to present the audio stimuli. Each participant was able to accurately localise "virtual" stimuli in azimuth. The average azimuth error (adjusted for convergence at poles) was 8.5 degrees in the free field and 8.8 degrees using the 3-d audio display. This difference was not significant ($t[5] = 0.90$, $p = .41$).

2.6 Data Collection

A matrix of coloured buttons was placed close to the throttle for making responses to target communications signals. Response time was measured in ms from the initiation of a target call. The number of detected target calls, the number of false alarms, the proportion of correct colour/number response combinations and the reaction time to correct responses were calculated. The time window following a target message during which a response was accepted was 5 seconds. The rate of correct detections and false alarms were combined to calculate sensitivity (d') and response criterion using signal detection theory [5].

3.0 RESULTS

The proportion of detected target calls is shown for each display condition in figure 1a. The proportion of detected target calls differed significantly across display conditions ($F[1.8,9.0] = 6.9$, $p = .017$). Planned comparisons revealed that the proportion of detected calls was significantly higher in the 3-d condition compared with either the diotic or the in-front condition (diotic: $t[5] = 3.55$, $p = .016$; in-front: $t[5] = 3.50$, $p = .017$) and that the proportion of detected calls for the diotic and in-front conditions did not differ significantly ($t[5] = 0.12$, $p = .91$).

The average number of false alarms made in each 20-minute simulation is shown for each display condition in figure 1b. The number of false alarms was very low and did not differ significantly across display conditions ($F[1.4,6.8] = 0.49$, $p = .56$).

Sensitivity (d') and response criterion were calculated from the hit and false alarm rates using signal detection theory. Sensitivity is presented for each display condition in figure 1c. Statistical analysis revealed a significant effect of display condition ($F[1.1,5.7] = 9.23$, $p = .023$). Planned comparisons revealed that sensitivities in the diotic and in-front conditions did not differ significantly ($t[5] = 0.82$, $p = .45$) and that sensitivity in the 3-d condition was significantly higher than that in the diotic ($t[5] = 2.77$, $p = .039$) or in-front condition ($t[5] = 3.54$, $p = .017$).

The response criterion is shown for each display condition in figure 1d. Criterion did not differ significantly across display conditions ($F[1.3,6.6] = 0.30$, $p = .66$).

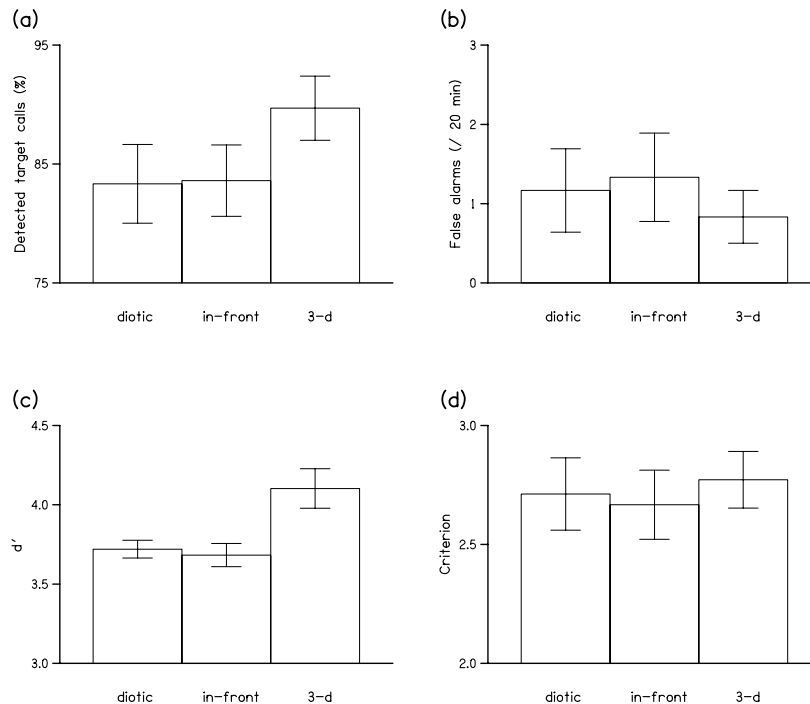


Figure 1. (a) Proportion of detected target messages, (b) number of false alarms, (c) sensitivity, and (d) response criterion for each audio display condition.

From the responses made to target calls, the proportion that correctly identified the colour/number combination was calculated. Participants made a similar proportion of correct responses in the diotic and in-front conditions, but more in the 3-d condition (figure 2a). Statistical analysis revealed a significant effect of display condition ($F[1.7,8.4] = 16.66$, $p = .001$). Planned comparisons revealed that the proportion of correct responses did not differ significantly between the diotic and in-front conditions ($t[5] = 0.91$, $p = .41$), but was significantly higher in the 3-d condition compared with either the diotic ($t[5] = 5.04$, $p = 0.004$) or the in-front condition ($t[5] = 5.82$, $p = .002$).

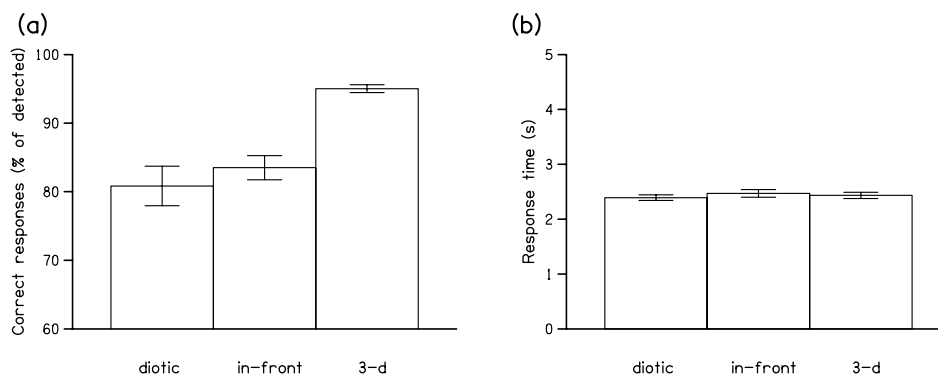


Figure 2. (a) Proportion of detected messages for which a correct response was made, and (b) reaction time for correct responses in each audio display condition.

Median response times were calculated for correct colour/number responses for each participant. Average median response times were close to 2.4 seconds (figure 2b) and did not differ significantly across display conditions ($F[1.7,8.8] = 3.78$, $p = .07$).

4.0 DISCUSSION

This study used a paradigm of high temporal uncertainty in the presence of a sustained background perceptual-motor task and found that the ability of listeners to detect a target callsign is improved when speech messages are distributed in azimuth. The study also found that the accuracy of responses was improved for spatially distributed messages, compared to messages presented via a traditional diotic display.

The detection of target messages has been studied previously using the CRM [6] and has been found to be significantly higher in spatialised audio conditions than in diotic conditions. However, in that study there was little temporal uncertainty because each trial was independent and contained messages that began simultaneously. In contrast, the present study was conducted under conditions of high temporal uncertainty.

An advantage for spatial presentation of simultaneous speech messages with respect to response accuracy has also been demonstrated in other previous studies employing the CRM under conditions of high temporal certainty. For an accurate response to be made in the CRM task, the callsign has to be identified, and features of the target voice (e.g., its location, timbre, or prosody) have to be used to link the target callsign and coordinates (i.e., colour and number). Previous work has shown that the predominant error involves responding to coordinates associated with a distracting talker [7], so the main difficulty encountered when listening to co-located talkers is linking callsigns and coordinates correctly rather than understanding the messages.

One previous study [8] has investigated the detection and identification of multiple communications in a simulated flying task. In that study, pilot's verbal responses to communications messages were more correct when communications were presented via a 3-d audio rather than a diotic display. However, performance in callsign detection and message identification were combined into a single response score so it is not clear how each of these tasks contributed to the result.

The task in the present study approximates those existing in operational environments more closely than do the tasks in many previous studies. Our results, therefore, provide a further demonstration of the potential for 3-d audio communications displays in operational environments.

- [1] Bronkhorst, A.W. (2000) The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. *Acustica-Acta Acustica* 86, 117-128.
- [2] Moore, T. (1981) Voice communication jamming research. In AGARD Conference Proceedings 331: *Aural Communication in Aviation* (pp. 2:1-2:6). AGARD: Neuilly-sur-Seine, France.
- [3] Bolia, R.S., Nelson, W.T, Ericson, M.A. & Simpson, B.D. (2000) A speech corpus for multitalker communications research. *Journal of the Acoustical Society of America* 107, 1065-1066.
- [4] Martin, R.L., McAnally, K.I. & Senova, M.A. (2001) Free-field equivalent localization of virtual audio. *Journal of the Audio Engineering Society* 49, 14-22.
- [5] Green, D.M. & Swets, J.A. (1966) *Signal detection theory and psychophysics*. Wiley, New York.

- [6] Nelson, W.T., Bolia, R.S., Ericson, M.A and McKinley, R.L. (1998) Monitoring the simultaneous presentation of spatialized speech signals in a virtual acoustic environment. Proceedings of the 1998 IMAGE conference.
- [7] Brungart, D.S. (2001) Informational and energetic masking effects in the perception of two simultaneous talkers. Journal of the Acoustical Society of America 109, 1101-1109.
- [8] Haas, E.C, Gainer, C., Wightman, D., Couch, M. & Shilling, R. (1997) Enhancing system safety with 3-d audio displays. Proceedings of the Human Factors and Ergonomics Society 41st annual meeting.

Comment améliorer la sélection et le traitement des messages verbaux ?

Marie Rivenez

Département Sciences Cognitives
IMASSA
BP 73
91233 Bretigny sur Orge, France

m.rivenez@imassa.fr

Chris Darwin

Department of Psychology
University of Sussex
Falmer, Brighton
BN1 9QG, UK

cjd@sussex.ac.uk

Anne Guillaume

Département Sciences Cognitives
IMASSA
BP 73
91233 Bretigny sur Orge, France

aguillaume@imassa.fr

RESUMÉ

L'objectif de cette recherche est d'améliorer la sélection des messages verbaux. Nous cherchons à déterminer les facteurs influençant le traitement d'un message verbal lorsque l'attention est portée sur un autre message. Nous mesurons le traitement non attentionnel des messages en utilisant un paradigme d'amorçage en écoute dichotique : les participants doivent détecter un mot cible appartenant à une catégorie spécifique, parmi une liste de mots présentés rapidement dans l'oreille focalisée (oreille gauche). Le mot amorce (identique d'un point de vue phonologique au mot cible) est présenté dans une phrase juste avant le mot cible dans l'oreille non focalisée (oreille droite). L'effet d'amorçage (TRs plus rapides lorsque l'amorce est le même mot que la cible que lorsque c'est un mot sans rapport) est un indicateur du traitement du message de l'oreille non focalisée. Nous tentons de déterminer le rôle de la différence de hauteur (Expérience 2 et 4) et de timbre (Expérience 3 et 4) entre les messages co-occurents sur le traitement du message non focalisé.

1. INTRODUCTION

L'objectif de cette série d'étude est de mettre en évidence les facteurs permettant de traiter un messages sans attention afin d'améliorer la sélection des messages verbaux.

Le rôle de l'attention dans le traitement de messages verbaux est très controversé. Des études antérieures montrent que l'attention n'est pas nécessaire pour traiter un message verbal ([1], [2], [3], [4], [5]). Ces auteurs soutiennent que les propriétés lexicales et sémantiques des mots peuvent être traitée sans attention.

Rivenez, M.; Darwin, C.; Guillaume, A. (2005) Comment améliorer la sélection et le traitement des messages verbaux ? Dans *Nouvelles orientations pour l'amélioration des techniques audio* (p. 32-1 – 32-12). Compte rendu de réunion RTO-MP-HFM-123, Communication 32. Neuilly-sur-Seine, France : RTO. Disponible sur le site : <http://www.rto.nato.int/abstracts.aps>.

Cependant, dans toutes ces études, l'attention n'était jamais parfaitement contrôlée. Holender (1986) [6] porte sa critique sur le fait que toutes ces études ont utilisé des mots isolés pour constituer le message non focalisé. Hors les mots isolés peuvent, plus facilement que de la parole continue, attirer l'attention des auditeurs. Plusieurs auteurs ont tenté de contrôler l'attention des auditeurs en utilisant de la parole continue pour le message non focalisé ([7], [8], [9], [10], [11]). Aucune de ces études ne montre de traitement des messages non focalisés. Toutefois, cette absence de traitement peut être engendrée, d'une part, par l'utilisation de mesures explicites (les auditeurs doivent explicitement dire ce qu'ils ont entendu du message non focalisé) et indirectes (mesures off-line) du traitement du message non focalisé et, d'autre part, l'utilisation de voix similaires pour les deux messages co-occurents.

En effet, la capacité à suivre un message présenté simultanément à un autre message est sérieusement dégradée lorsque les messages co-occurents sont prononcés avec la même voix ([12], [13]). Brungart (2001a) [12] et Brungart, Simpson, Ericson, & Scott (2001) [13] montrent que, lorsque deux messages co-occurents sont prononcés par la même personne, les auditeurs identifient correctement les phrases dans 50 % des cas, lorsque ils sont prononcés par deux personnes différentes mais de même sexe, les performances passent à 75 % et elles passent à 85 % lorsque les messages sont prononcés par des locuteurs de sexe différent.

Les paramètres déterminant dans la différenciation du sexe des voix sont la hauteur tonale, déterminée par la fréquence fondamentale et le timbre, déterminé par la longueur de tractus vocal qui modifie la fréquence des formants.

La hauteur tonale est un indice particulièrement utile pour ségréger des messages simultanés. Une différence de fréquence fondamentale permet d'aider la ségrégation de deux voyelles simultanées [14]. La fréquence fondamentale joue un rôle important dans le groupement des formants [15] [16] [17]. La localisation de sons de parole simultanés peut être améliorée par les relations de fréquence fondamentale : le groupement entre fréquence sur la base du délais interaural est rendu possible si les relations de fréquence fondamentale sont concordantes avec celles de localisation [18]. La capacité à suivre un message dans le temps, présenté simultanément à d'autres messages, est aussi largement améliorée lorsque les messages ont une fréquence fondamentale ou une longueur de tractus vocal différents ([19], [20]).

Dans cette recherche, nous tentons de mesurer l'efficacité de ces indices de groupement de hauteur tonale et de timbre sur le traitement de messages verbaux co-occurents. Nous faisons l'hypothèse que le traitement de messages non focalisés par l'attention dépend directement de la disponibilité de ces indices de groupement.

2. METHODE GÉNÉRALE

Afin d'évaluer le traitement d'un message non focalisé, nous utilisons un paradigme d'amorçage en écoute dichotique. Les participants doivent détecter un mot cible appartenant à une catégorie spécifique, parmi une liste de mots présentés rapidement dans l'oreille focalisée (oreille gauche). Le mot amorce (identique d'un point de vue phonologique au mot cible) est présenté soit isolément, soit dans une phrase inepte juste avant le mot cible dans l'oreille non focalisée (oreille droite). L'effet d'amorçage (TRs plus rapides lorsque l'amorce est le même mot que la cible que lorsque c'est un mot sans rapport) est un indicateur du traitement du message de l'oreille non focalisée. Il est calculé en soustrayant les essais pour lesquels la cible est précédée d'une amorce de répétition (condition amorce reliée) aux essais pour lesquels la cible est précédée d'un mot sans rapport (condition amorce non reliée).

L'amorce est toujours non focalisée par l'attention car elle est insérée dans une phrase. Le même locuteur est utilisé pour enregistrer tous les stimuli. Nous aplatissons la fréquence fondamentale du message contenant la cible tandis que les variations prosodiques du message contenant l'amorce sont conservées.

Afin de faciliter la sélection du message focalisé, celui-ci est présenté à un niveau sonore de 72 dB tandis que le message non focalisé est présenté à 60 dB.

Dans l'Expérience 1, nous testons l'effet des relations de fréquence fondamentale entre les messages sur l'effet d'amorçage par une amorce non focalisée. L'Expérience 2 évalue le rôle du timbre sur cet effet d'amorçage. L'Expérience 3 vise à mesurer l'effet conjugué des relations de hauteur tonale et de timbre entre les messages sur l'effet d'amorçage. Enfin, dans l'Expérience 4, nous cherchons à mesurer l'effet d'amorçage par une amorce focalisée, lorsque les deux messages ont la même voix.

3. EXPERIMENT 1

L'objectif de l'Expérience 1 est d'étudier le rôle des relations de hauteur tonale entre des messages co-occurents sur le traitement d'un message non focalisé en mesurant l'effet d'amorçage provoqué par une amorce non focalisée sur une cible focalisée.

3.1. Méthode

Afin de tester l'effet de la hauteur tonale sur le traitement d'un message non focalisé, nous manipulons la fréquence fondamentale (F0) du message focalisé et celle du message non focalisé. Les messages focalisé et non focalisé sont resynthétisés en utilisant l'algorithme PSOLA, sous le logiciel Praat, afin de maintenir le contour de hauteur original.

La fréquence fondamentale moyenne des deux messages est de 140 Hz ou de 180 Hz. Ils peuvent être dans une gamme différente de F0 (différence moyenne de 5 dT) ou dans la même gamme. La Table 1 résume les quatre conditions expérimentales utilisées.

F0	Différente	Similaire	F0	Différente	Similaire
Message focalisé	180Hz	140Hz	Message focalisé	180Hz	140Hz
Message non focalisé	140Hz	140Hz	Message non focalisé	140Hz	140Hz

Table 1: Conditions de relation de F0 entre les messages focalise et non focalise, utilisées dans l'Expérience 1.

Chaque participant est assigné à l'un des 4 groupes de relation de hauteur. Les amorces reliées et non reliées sont contrebalancées dans 2 groupes. Le facteur d'amorçage (condition amorce reliée vs condition amorce non reliée) est un facteur intra sujets tandis que le facteur de relation de hauteur est un facteur inter sujets.

3.2. Participants

80 étudiants de l'Université de Sussex (UK), rémunérés participent à cette expérience. Ils sont tous de langue maternelle anglaise et n'ont pas de trouble du langage, ni de l'attention, ni de l'audition.

3.3. Résultats

Une ANOVA à quatre facteurs est conduite pour l'analyse par participants déclarant les facteur d'amorçage (facteur croisé), de relation de hauteur (contrastant les conditions hauteur différente et hauteur similaire : facteur emboîté), de hauteur tonale du message non focalisé (contrastant les conditions de fréquence fondamentale 140 Hz et 180 Hz : facteur emboîté), et de contrebalancement (facteur emboîté). Une ANOVA à trois facteurs est calculée pour l'analyse par items, déclarant les facteur d'amorçage (facteur croisé), de relation de hauteur (facteur croisé) et de hauteur du message non focalisé (facteur croisé).

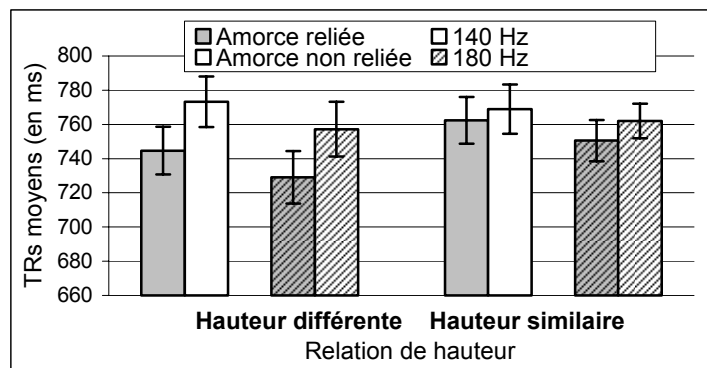


Figure 1: TRs (en ms), obtenus pour l'Expérience 1, moyennés sur les 80 sujets de l'expérience, pour les conditions amorce reliée (colonne grise) et amorce non reliée (colonne blanche) lorsque les messages ont une gamme de hauteur différente (pattern de gauche) et lorsque les messages ont une gamme de hauteur similaire (pattern de droite), et lorsque la fréquence fondamentale moyenne des messages non focalisés est de 140 Hz (colonnes vides) et lorsqu'elle est de 180 Hz (colonnes hachurées). Les barres d'erreur indiquent les erreurs types entre les participants.

Les résultats sont représentés dans la figure 1. Nous trouvons un effet principal du facteur d'amorçage : les TRs moyen sont plus rapides dans la condition amorce reliée (747 ms) que dans la condition amorce non reliée (765 ms). Cet effet d'amorçage de 19 ms est significatif ($F_1(1,72) = 12,35$; $p < 0,0001$; $F_2(1,56) = 22,89$; $p < 0,00001$).

L'effet d'amorçage varie entre les conditions de relation de hauteur : lorsque les deux messages sont dans des gammes de hauteur différentes, l'effet d'amorçage est de 28 ms ($F_1(1,72) = 14,23$; $p < 0,001$; $F_2(1,56) = 25,10$; $p < 0,00001$), tandis que lorsque les deux messages sont dans la même gamme de hauteur, l'effet d'amorçage n'est que de 9 ms et il n'est pas significatif pour l'analyse par participants ($F_1(1,72) = 1,43$; $p = 0,23$), et marginalement significatif pour l'analyse par items ($F_2(1,56) = 3,61$; $p = 0,06$). L'interaction entre les facteurs d'amorçage et de relation de hauteur est marginalement significative pour les participants ($F_1(2,72) = 3,32$; $p = 0,07$), et significative pour les items ($F_2(1,56) = 7,49$; $p < 0,01$).

3.4. Discussion

L'Expérience 1 montre un effet d'amorçage de 28 ms, lorsque la hauteur timbre des messages focalisé et non focalisé est différente. Cette expérience montre qu'il est possible d'obtenir un effet d'amorçage non attentionnel, en insérant l'amorce dans de la parole continue afin d'éviter les basculement attentionnel. Par ailleurs, il apparaît que les relations de hauteur entre les messages co-occurents sont déterminantes dans le traitement d'un message non focalisé. Lorsque l'attention n'est pas portée sur un message présenté simultanément à un autre message, il est nécessaire que les messages co-occurents aient une hauteur tonale différente pour que le message non focalisé soit traité. Nos données contredisent les conclusions de Holender (1986) [6] et Dupoux et al. (2003) [11] concernant l'incapacité des auditeurs à traiter un messages non focalisé par l'attention. Cette contradiction s'explique par le fait que les études passées, qui n'ont pas montré de traitement des messages non focalisé, ont toutes utilisé la même voix pour les messages co-occurents ([7], [8], [9], [10], [11]).

Les résultats de l'Expérience 1 suggèrent que la présence d'indices de groupement, permettant de ségréger correctement et de façon stable au cours du temps des messages simultanés, est nécessaire pour le traitement lexical des messages non focalisés. Si cette hypothèse est vraie, alors le fait de procurer n'importe quel autre indice de groupement (comme par ex. le timbre) aux auditeurs devrait leur permettre de traiter lexicalement un message non focalisé. Nous testons cette hypothèse en introduisant une différence de timbre entre les messages, sans qu'il n'y ait de différence de hauteur tonale.

4. EXPERIMENT 2

L'objectif de l'Expérience 2 est de tester le rôle d'un autre indice de groupement sur le traitement d'un message non focalisé à savoir le timbre.

4.1. Méthode

Nous utilisons la même méthode générale que celle décrite plus haut. Les messages focalisés sont prononcés par le même locuteur (F_0 moyenne = 140 Hz). Ils sont donc dans la même gamme de hauteur. Nous manipulons les relations de timbre entre les voix en changeant la longueur de tractus vocal (TV) des messages. La différence de longueur de tractus vocal peut être de 15 ou 30 % entre les messages. Pour chaque condition de différence de longueur de TV, nous allongeons de 15 % le message focalisé ou le message non focalisé. Afin d'obtenir une différence de longueur de 30 % entre les messages, cet allongement est associé à un raccourcissement de 15 % de l'autre message. La longueur de TV est changée à l'aide de l'algorithme PSOLA (voir [19] pour le détail des manipulations). Nous construisons ainsi 4 conditions de différence de TV. Étant donné que nous contrebalançons les amorces reliées et les amorces non reliées, nous obtenons 8 conditions expérimentales, chaque participant étant assigné à l'une de ces 8 conditions.

4.2. Participants

120 étudiants de l'Université de Sussex (UK), rémunérés participent à cette expérience. Ils sont tous de langue maternelle anglaise et n'ont pas de trouble du langage, ni de l'attention, ni de l'audition.

4.3. Résultats

Une ANOVA à trois facteurs est calculée pour l'analyse par items, déclarant les facteurs d'amorçage (facteur croisé), de différence de tractus vocal (facteur croisé) et de message auquel est appliqué l'allongement du tractus vocal (facteur croisé).

Nous trouvons un effet principal du facteur d'amorçage : les TRs moyens sont plus rapides dans la condition amorce reliée (783 ms) que dans la condition amorce non reliée (796 ms). Cet effet d'amorçage de 13 ms est significatif ($F_1(1,112) = 9,93$; $p < 0,005$; $F_2(1,54) = 16,16$; $p < 0,005$). Aucune interaction entre l'effet d'amorçage et les autres facteurs de l'expérience n'est trouvée. Les TRs sont représentés dans la figure 17, dans laquelle nous avons moyenné les conditions de message pour lequel est appliqué l'allongement du tractus vocal. L'effet d'amorçage ainsi que les barres d'erreur, pour les participants et les items, trouvés pour les deux conditions de longueur de tractus vocal, est représenté dans la figure 18. Cette figure indique que l'effet d'amorçage semble plus important pour une différence de longueur de tractus vocal de 30 % (16 ms) que de 15 % (10 ms). Cependant, l'absence d'interaction significative entre le facteur d'amorçage et celui de différence de longueur de tractus vocal montre que cette différence de taille d'effet d'amorçage n'est pas systématique.

Un effet principal du facteur de différence de timbre est trouvée pour l'analyse par items indiquant uniquement ($F_2(1,54) = 12,92$; $p < 0,001$) : les TRs sont significativement plus rapides lorsque la différence de longueur de tractus vocal entre les deux messages est de 15 % que lorsqu'elle est de 30 %.

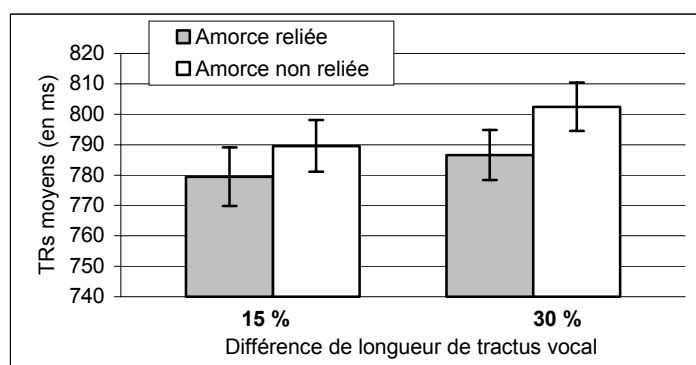


Figure 2: TRs (en ms), obtenus pour l'Expérience 1, moyennés sur les 120 sujets de l'expérience, pour les conditions amorce reliée (colonne grise) et amorce non reliée (colonne blanche) lorsque les messages ont une différence de longueur de tractus vocal de 15 % (pattern de gauche) et lorsque cette différence est de 30 % (pattern de droite), moyennés pour les deux conditions d'application de l'allongement du tractus vocal. Les barres d'erreur indiquent les erreurs types entre les participants.

4.4. Discussion

L'Expérience 2 montre un effet d'amorçage de 13 ms, lorsque les messages co-occurents diffèrent sur la dimension du timbre. Ces données sont cohérentes avec celles des études montrant que le timbre est un indice de groupement très robuste pour suivre un message dans le temps ([19], [20], [21]). Nous montrons qu'en l'absence d'une différence de hauteur tonale, une différence de timbre entre des messages co-occurents permet de traiter un message sur lequel l'attention n'est pas portée. Il semble donc que le traitement d'un message non focalisé dépende des mêmes facteurs que l'organisation perceptive des signaux de parole. Cette observation suggère que le traitement d'un message non focalisé présuppose qu'il soit bien ségrégué des messages co-occurents et que cette ségrégation dépend de la présence d'indices de groupement.

Les études sur la capacité à suivre un message présenté simultanément à d'autres messages montrent que les performances sont toujours meilleures lorsque les voix des messages co-occurents sont produites par des locuteurs de sexe différent, que lorsque le seul indice disponible est soit la hauteur, soit le timbre. Les paramètres principaux qui permettent de discriminer des voix de sexe différent sont la fréquence fondamentale et la longueur de tractus vocal. Il est donc possible que le traitement d'un message non focalisé soit maximisé en mettant en conjonction les indices de hauteur tonale et de timbre. C'est ce que nous faisons dans l'Expérience 3, dans laquelle nous utilisons des messages ayant à la fois une hauteur et un timbre différents.

5. EXPERIMENT 3

Les études sur les indices permettant de suivre un message présenté simultanément à d'autres messages, indiquent que les auditeurs tirent plus de bénéfice d'une conjonction des indices de fréquence fondamentale et de tractus vocal que de ces mêmes indices séparément ([20]). Darwin, Brungart, & Simpson (2003) montrent que les indices de hauteur et de timbre sont super-additifs, c'est-à-dire que la conjonction de ces indices engendre des performances meilleures que la simple addition des performances obtenues avec un seul de ces indices. Dans cette expérience, nous cherchons à tester l'effet de la conjonction de ces indices de groupement sur l'effet d'amorçage.

5.1. Méthode

Le matériel et la procédure sont les mêmes que ceux des expériences précédentes, sauf que les messages focalisé et non focalisé ont une hauteur et un timbre différents. Le message focalisé a une F0 aplatie de 180 Hz et son timbre est inchangé. Le message non focalisé a F0 moyenne de 140 Hz et sa longueur de TV est augmentée de 15 % (avec l'algorithme PSOLA). Ainsi les messages focalisé et non focalisé diffèrent du point de vue de leur F0 (5 dT d'écart) et de leur timbre (15 % d'écart). Les amorces reliées et non reliée sont contrebalancée dans 2 groupes. Les participants sont assignés à l'un de ces 23 groupes.

5.2. Participants

26 étudiants de l'Université de Sussex (UK), rémunérés participent à cette expérience. Ils sont tous de langue maternelle anglaise et n'ont pas de trouble du langage, ni de l'attention, ni de l'audition.

5.3. Résultats

Une ANOVA à deux facteurs est conduite pour l'analyse par participants déclarant les facteur d'amorçage (facteur croisé), et de contrebalancement (facteur emboîté). Un test t est conduit pour l'analyse par items, déclarant le facteur d'amorçage (facteur croisé).

Les résultats sont représentés dans la figure 3. Nous trouvons un effet principal du facteur d'amorçage : les TRs moyen sont plus rapides dans la condition amorce reliée (761 ms) que dans la condition amorce non reliée (793 ms). Cet effet d'amorçage de 32 ms est significatif ($F_1(1,24) = 12,77$; $p < 0,005$; $t(57) = 5,11$; $p < 0,00001$). Aucune interaction entre l'effet d'amorçage et le facteur de contrebalancement n'est trouvée.

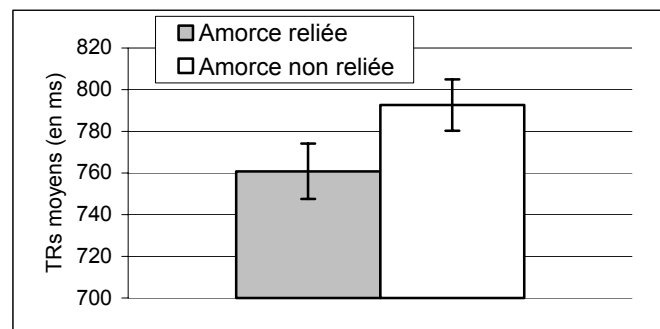


Figure 3: TRs (en ms), obtenus pour l'Expérience 2, moyennés sur les 26 sujets de l'expérience, pour les conditions amorce reliée (colonne grise) et amorce non reliée (colonne blanche). Les barres d'erreur indiquent les erreurs types entre les participants.

L'Expérience 3 a pour objectif de tester l'effet de la conjonction des indices de groupement de hauteur et de timbre sur l'effet d'amorçage. Aux vues des expériences passées, nous faisons l'hypothèse d'une augmentation de l'effet d'amorçage en mettant deux indices de groupement à disposition. Nous ajoutons cette section afin de comparer directement l'effet d'amorçage trouvé pour les conditions expérimentales dans lesquelles des indices de groupement sont disponibles. Nous comparons l'effet d'amorçage de la condition de différence de hauteur de l'expérience 1 (moyennée pour les deux conditions de différence de hauteur), la condition de différence de longueur de tractus vocal de l'Expérience 2 (moyennée pour les deux différences de longueur utilisées), et l'Expérience 3. Nous avons donc les conditions d'indice de groupement suivantes : (a) – Hauteur différente ; (b) – Timbre différent ; (c) – Conjonction d'indices de hauteur et de timbre.

Nous conduisons une ANOVA sur l'effet d'amorçage, pour l'analyse par participants déclarant le facteur d'indice de groupement (comparant les conditions a, b et c : facteur emboîté). Pour l'analyse par item, nous conduisons une ANOVA sur l'effet d'amorçage, déclarant le facteur d'indice de groupement (facteur croisé). Etant donné que certains items sont supprimés lors du nettoyage des données pour chaque expérience, il y a certaines valeurs manquantes dans cette dernière analyse.

Les résultats sont présentés dans la figure 20. Les analyses statistiques indiquent un effet principal du facteur d'indice de groupement significatif ($F_1(2,183) = 3,09$; $p < 0,05$; $F_2(2,108) = 3,14$; $p < 0,05$). Les comparaisons planifiées indiquent que l'effet d'amorçage est le même pour les conditions de conjonction d'indices (32 ms) et de différence de hauteur (28 ms) ($F_1 < 1$; $F_2 < 1$), mais il est significativement plus élevé que pour les condition timbre différent (13 ms) ($F_1(1,183) = 6,15$; $p = 0,05$; $F_2(1,54) = 9,88$; $p < 0,005$).

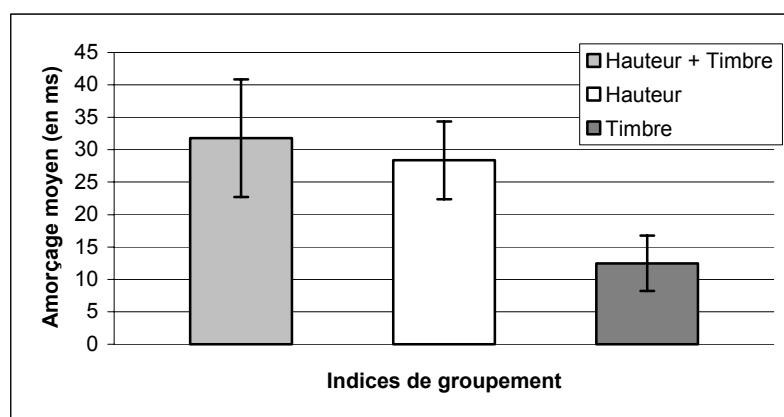


Figure 4: Effet d'amorçage (condition amorce non reliée – amorce reliée, mesuré en ms) obtenu pour les 3 conditions d'indice de groupement (différence de hauteur et de timbre – Expérience 3 –, colonne gris clair ; différence de hauteur – Expérience 1 –, colonne blanche ; différence de longueur de tractus vocal – Expérience 2 –, colonne gris foncé droite). Les barres d'erreur indiquent les erreurs types entre les participants.

5.4. Discussion

Les résultats de l'Expérience 3 montrent un effet d'amorçage de 32 ms, lorsque la hauteur et le timbre des messages focalisé et non focalisé sont différents. La comparaison de cette expérience avec celles contenant un seul indice de groupement (soit la hauteur, soit le timbre) indique que la conjonction des indices engendre un effet d'amorçage plus élevé (32 ms) que lorsque le seul indice est le timbre (13 ms). En revanche, l'effet d'amorçage n'est pas plus important que lorsque le seul indice présent est la hauteur (28 ms). Cette absence de différence suggère que seule la différence de hauteur tonale est à l'origine de l'effet d'amorçage trouvé dans l'Expérience 3.

L'Expérience 3 montre que le traitement d'un message non focalisé est principalement déterminé par la différence de hauteur tonale entre les messages co-occurents.

6. DISCUSSION GÉNÉRALE

L'attention est-elle nécessaire pour traiter des messages co-occurents ?

Nous montrons qu'il est possible d'obtenir un effet d'amorçage par une amorce non focalisée et insérée dans de la parole continue afin d'éviter les basculements attentionnels. Ce résultat s'oppose à Holender (1986) et Dupoux et al. (2003) qui suggèrent que le traitement de messages non focalisé n'est possible qu'avec des basculements attentionnels.

L'effet d'amorçage trouvé indique que lorsque des auditeurs doivent effectuer une tâche nécessitant une forte charge attentionnelle, ils sont capables de traiter un message sur lequel ils ne portent pas leur attention.

Quels facteurs rendent possible le traitement d'un message non focalisé ?

Nous montrons que le traitement de messages non focalisés dépend de la différence de hauteur (Expérience 1) ou de timbre (Expérience 2) entre les messages focalisé et non focalisé. Les indices de groupement de hauteur et de timbre permettent de ségréger les messages et de les maintenir séparés au cours du temps. Toutefois les auditeurs montrent une préférence pour l'indice de hauteur. En effet, lorsque la hauteur et le timbre sont associés, l'effet d'amorçage, mesurant le traitement du message non focalisé, n'est pas différent de celui trouvé lorsque seule la hauteur est disponible (Expérience 3).

La conjonction de ces indices ne permet pas d'augmenter l'effet d'amorçage, ce qui montre que ces indices n'ont pas un effet additif sur le traitement d'un message non focalisé. Ces données ne vont pas dans le sens de celles de Darwin et al. (2003) qui montrent que la hauteur et le timbre ont des effets super-additifs sur la capacité à suivre un message sélectionné dans le temps. Ces observations suggèrent que les indices de groupement ne sont pas utilisés de la même façon pour suivre et traiter un message focalisé et un message non focalisé. En effet, La différence entre notre expérience et celle des auteurs réside dans l'attention portée aux messages. Dans notre expérience, les indices de ségrégation permettent de suivre un message non focalisé, tandis qu'ils permettent de suivre un message focalisé dans les expériences de Darwin et al. (2003). Ces données suggèrent qu'en l'absence d'attention, la hauteur est préférentiellement utilisée pour suivre et traiter un message dans le temps mais que son efficacité ne peut être magnifiée en présence d'un autre indice de groupement tel que le timbre.

Il faut, néanmoins, noter que la différence de tractus vocal utilisée dans cette expérience est de 15 %, ce qui correspond à la condition engendrant un plus faible effet d'amorçage (10 ms), lorsque la longueur de tractus vocal est le seul indice disponible. Même si l'Expérience 7 n'indique pas d'interaction significative entre le facteur d'amorçage et le facteur de taille de la différence du tractus vocal, il est possible que lorsque la différence de hauteur tonale est de 5dT, l'ajout d'une différence de tractus vocal supérieure pourrait renforcer l'effet d'amorçage et ainsi maximiser le traitement du message non focalisé. Par ailleurs, les différences de hauteur (5 dT) et de longueur de tractus vocal (15 %) utilisées ne permettent pas un changement de sexe de voix. Pour opérer un changement de sexe de voix, il est judicieux d'associer la différence de longueur de tractus vocal de 15 % à une différence d'une octave. Il est donc possible que le traitement de messages non focalisé soit maximisé en changeant le sexe des voix.

7. CONCLUSION

Comment améliorer la sélection et le traitement de messages verbaux ?

Nous suggérons que lorsque le traitement d'un message non focalisé est possible, la sélection des informations pertinentes devrait être améliorée. En effet, il devrait être plus aisé pour l'opérateur de basculer son attention sur les messages importants s'il les a déjà préalablement traités. Les données exposées suggèrent que la présentation des messages verbaux doit tenir compte de la charge attentionnelle des auditeurs. Dans une situation à charge attentionnelle élevée (les auditeurs ne peuvent accorder beaucoup d'attention à certains messages), les indices de groupement, classiquement utilisés pour suivre des messages dans le temps (localisation ([22]), intensité ([18]), prosodie, hauteur, timbre), ne sont pas tous utilisés de la même façon que lorsque la charge attentionnelle n'est pas élevée (les auditeurs peuvent ont suffisamment de ressources attentionnelles pour traiter certaines messages). En effet, lorsque les messages sont prononcés par la même voix, les auditeurs ne peuvent tirer profit de la différence de localisation des messages ni de leur différence de niveau sonore pour traiter le message non focalisé. Il semble donc important de tenir compte des mécanismes perceptifs impliqués dans le traitement de la parole afin de maximiser l'efficacité de la sélection et du traitement des informations auditives.

Deux conclusions sont donc à retenir : (a) - Lorsque la charge attentionnelle des auditeurs est élevée, les messages non focalisés par les auditeurs peuvent être traités. Le niveau de traitement de ces informations reste toutefois à tester. (b) - Afin d'optimiser le traitement de messages non focalisés il est nécessaire et suffisant que les messages co-occurents soient dans une gamme de hauteur différente des messages focalisés. La présentation des informations verbales dans les environnements nécessitant une charge attentionnelles élevée devrait tenir compte de ces indications.

8. REFERENCES

- [1] Eich, E. (1984). Memory for unattended events: Remembering with and without awareness, *Memory and Cognition*, 12, 105-111.
- [2] Forster, P. M., & Govier, E. (1978). Discrimination without awareness? *Quarterly Journal of Experimental Psychology*, 30, 289-295..
- [3] Lewis, J. (1970). Semantic processing of unattended messages using dichotic listening. *Journal of Experimental Psychology*, 85, 225-228.
- [4] McKay, D. G. (1973). Aspects of a theory of comprehension, memory and attention. *Quarterly Journal of Experimental Psychology*, 25, 22-40.
- [5] Norman, D. A. (1969). Memory while shadowing. *Quarterly Journal of Experimental Psychology*, 21, 85-93.
- [6] Holender, D. (1986). Semantic activation without conscious identification in dichotic listening, parafoveal vision, and visual masking: A survey and appraisal. *The behavioral and Brain Sciences*, 9, 1-66.
- [7] Treisman (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology*, 12, 242-248.

- [8] Newstead, S. E., & Dennis, I. (1979). Lexical and grammatical processing of unshadowed messages: a re-examination of the McKay effect. *Quarterly Journal of Experimental Psychology*, 31, 477-488.
- [9] Treisman (1964c). Effect of irrelevant material on the efficiency of selective listening. *American Journal of Psychology*, 77, 533-546.
- [10] Wood, N., Stadler, M. A., & Cowan, N. (1997). Is there explicit memory without attention? A re-examination of task demands in Eich's (1984) procedure. *Memory and Cognition*, 25, 772-779.
- [11] Dupoux, Kouider, & Mehler, J. (2003). Lexical access without attention? Exploration using dichotic priming. *Journal of Experiment Psychology: Human Perception and Performance*, 29(1), 172-184.
- [12] Brungart, D. S. (2001a). Informational and energetic masking effects in the perception of two simultaneous talkers. *Journal of the Acoustical Society of America*, 109 (3), 1101-1109.
- [13] Brungart, D. S., Simpson, B. D., Ericson, M. A., & Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *Journal of the Acoustical Society of America*, 110 (5), 2527-2538.
- [14] Scheffers, M. T. (1983). Shifting vowels: Auditory pitch analysis and sound segregation. Ph.D. thesis. Groningen University, The Netherlands.
- [15] Darwin, C. J. (1981). Perceptual grouping of speech components differing in fundamental frequency and onset-time. *Quarterly Journal of Experimental Psychology*, 33A, 185-208.
- [16] Culling, J. F., & Darwin, C. J. (1993). Perceptual separation of simultaneous vowels: within and across-formant grouping by F0. *Journal of the Acoustical Society of America*, 93, 3454-3467.
- [17] Bird, J., & Darwin, C. J. (1998). Effects of a difference in fundamentele frequency in separating two sentences. In A. R. Palmer, A. Rees, A. Q. Summerfield, & R. Meddis (Eds), *Psychophysical and Psysiological Advances in Hearing*, Whurr: London, pp. 263-269.
- [18] Hill, N. I., & Darwin, C. J. (1993). Effect of onset asynchrony anf of mistuning on the lateralization of a pure tone embedded in a harmonic complex. *Journal of the Acoustical Society of America*, 93, 2307-2308.
- [19] Darwin, C. J., Brungart, D. S., & Simpson, B. D. (2003). Effect of fundamentele frequency and vocal tract length changes on attention to one of two simultaneous talkers. *Journal of the Acoustical Society of America*, 114 (5), 2913-2922.
- [20] Darwin, C. J., & Hukin, J. (2000a). Effectiveness of spatial cues, prosody and talker characteristics in selective attention. *Journal of the Acoustical Society of America*, 107, 970-977.
- [21] Darwin, C. J., & Hukin, J. (2000b). Effect of reverberation on spatial, prosodic and vocal tract size cues to selective attention. *Journal of the Acoustical Society of America*, 108, 335-342.
- [22] Culling, J. F., Hawley, M. L., & Litovsky, R. Y. (2004). The role of head-induced interaural time and level differences in the speech reception threshold for multiple interfering sound sources. *Journal of the Acoustical Society of America*, 116 (2), 1057-1065.



REPORT DOCUMENTATION PAGE																					
1. Recipient's Reference	2. Originator's References RTO-MP-HFM-123 AC/323(HFM-123)TP/60	3. Further Reference ISBN 92-837-1147-5	4. Security Classification of Document UNCLASSIFIED/ UNLIMITED																		
5. Originator Research and Technology Organisation North Atlantic Treaty Organisation BP 25, F-92201 Neuilly-sur-Seine Cedex, France																					
6. Title New Directions for Improving Audio Effectiveness																					
7. Presented at/Sponsored by The RTO Human Factors and Medicine Panel (HFM) Symposium held in Amersfoort, The Netherlands, 11-13 April 2005.																					
8. Author(s)/Editor(s) Multiple			9. Date April 2005																		
10. Author's/Editor's Address Multiple			11. Pages 442																		
12. Distribution Statement There are no restrictions on the distribution of this document. Information about the availability of this and other RTO unclassified publications is given on the back cover.																					
13. Keywords/Descriptors <table border="0"> <tr> <td>Active Noise Reduction (ANR)</td> <td>Intelligibility</td> </tr> <tr> <td>Active noise reduction system</td> <td>NIHL (Noise-Induced Hearing Loss)</td> </tr> <tr> <td>Auditory perception</td> <td>Noise reduction</td> </tr> <tr> <td>Bone conduction</td> <td>Signal processing</td> </tr> <tr> <td>Contact transducers</td> <td>Spatial auditory display</td> </tr> <tr> <td>Display devices</td> <td>Spatial hearing</td> </tr> <tr> <td>Hearing protection</td> <td>Speech intelligibility</td> </tr> <tr> <td>Impulse noise</td> <td>Speech recognition</td> </tr> <tr> <td>Individual differences</td> <td>Voice communication</td> </tr> </table>				Active Noise Reduction (ANR)	Intelligibility	Active noise reduction system	NIHL (Noise-Induced Hearing Loss)	Auditory perception	Noise reduction	Bone conduction	Signal processing	Contact transducers	Spatial auditory display	Display devices	Spatial hearing	Hearing protection	Speech intelligibility	Impulse noise	Speech recognition	Individual differences	Voice communication
Active Noise Reduction (ANR)	Intelligibility																				
Active noise reduction system	NIHL (Noise-Induced Hearing Loss)																				
Auditory perception	Noise reduction																				
Bone conduction	Signal processing																				
Contact transducers	Spatial auditory display																				
Display devices	Spatial hearing																				
Hearing protection	Speech intelligibility																				
Impulse noise	Speech recognition																				
Individual differences	Voice communication																				
14. Abstract Participants from eight nations reviewed (1) technical challenges for more effective protection of human hearing in adverse military environments, (2) efforts to optimize the portrayal of auditory information via helmets and headphones, and (3) research to improve speech intelligibility, especially via channels that encode and transmit multiple speech and data signals. This report summarizes progress and recommends seven areas for future research concentration.																					





BP 25

F-92201 NEUILLY-SUR-SEINE CEDEX • FRANCE
Télécopie 0(1)55.61.22.99 • E-mail mailbox@rta.nato.int



DIFFUSION DES PUBLICATIONS RTO NON CLASSIFIEES

Les publications de l'AGARD et de la RTO peuvent parfois être obtenues auprès des centres nationaux de distribution indiqués ci-dessous. Si vous souhaitez recevoir toutes les publications de la RTO, ou simplement celles qui concernent certains Panels, vous pouvez demander d'être inclus soit à titre personnel, soit au nom de votre organisation, sur la liste d'envoi.

Les publications de la RTO et de l'AGARD sont également en vente auprès des agences de vente indiquées ci-dessous.

Les demandes de documents RTO ou AGARD doivent comporter la dénomination « RTO » ou « AGARD » selon le cas, suivi du numéro de série. Des informations analogues, telles que le titre et la date de publication sont souhaitables.

Si vous souhaitez recevoir une notification électronique de la disponibilité des rapports de la RTO au fur et à mesure de leur publication, vous pouvez consulter notre site Web (www.rta.nato.int) et vous abonner à ce service.

CENTRES DE DIFFUSION NATIONAUX

ALLEMAGNE

Streitkräfteamt / Abteilung III
Fachinformationszentrum der
Bundeswehr (FIZBW)
Friedrich-Ebert-Allee 34, D-53113 Bonn

BELGIQUE

Etat-Major de la Défense
Département d'Etat-Major Stratégie
ACOS-STRAT – Coord. RTO
Quartier Reine Elisabeth
Rue d'Evère, B-1140 Bruxelles

CANADA

DSIGRD2
Bibliothécaire des ressources du savoir
R et D pour la défense Canada
Ministère de la Défense nationale
305, rue Rideau, 9^e étage
Ottawa, Ontario K1A 0K2

DANEMARK

Danish Defence Research Establishment
Ryvangs Allé 1, P.O. Box 2715
DK-2100 Copenhagen Ø

ESPAGNE

SDG TECEN / DGAM
C/ Arturo Soria 289
Madrid 28033

ETATS-UNIS

NASA Center for AeroSpace
Information (CASI)
Parkway Center, 7121 Standard Drive
Hanover, MD 21076-1320

FRANCE

O.N.E.R.A. (ISP)
29, Avenue de la Division Leclerc
BP 72, 92322 Châtillon Cedex

GRECE (Correspondant)

Defence Industry & Research
General Directorate, Research Directorate
Fakinos Base Camp, S.T.G. 1020
Holargos, Athens

HONGRIE

Department for Scientific Analysis
Institute of Military Technology
Ministry of Defence
H-1525 Budapest P O Box 26

ISLANDE

Director of Aviation
c/o Flugrad
Reykjavik

ITALIE

Centro di Documentazione
Tecnico-Scientifica della Difesa
Via XX Settembre 123
00187 Roma

LUXEMBOURG

Voir Belgique

NORVEGE

Norwegian Defence Research Establishment
Attn: Biblioteket
P.O. Box 25, NO-2007 Kjeller

PAYS-BAS

Royal Netherlands Military
Academy Library
P.O. Box 90.002
4800 PA Breda

POLOGNE

Armament Policy Department
218 Niepodleglosci Av.
00-911 Warsaw

PORTUGAL

Estado Maior da Força Aérea
SDFA – Centro de Documentação
Alfragide
P-2720 Amadora

REPUBLIQUE TCHEQUE

LOM PRAHA s. p.
o. z. VTÚLaPVO
Mladoboleslavská 944
PO Box 18
197 21 Praha 9

ROYAUME-UNI

Dstl Knowledge Services
Information Centre, Building 247
Dstl Porton Down
Salisbury
Wiltshire SP4 0JQ

TURQUIE

Milli Savunma Bakanlığı (MSB)
ARGE ve Teknoloji Dairesi Başkanlığı
06650 Bakanliklar – Ankara

AGENCES DE VENTE

NASA Center for AeroSpace Information (CASI)

Parkway Center, 7121 Standard Drive
Hanover, MD 21076-1320
ETATS-UNIS

The British Library Document Supply Centre

Boston Spa, Wetherby
West Yorkshire LS23 7BQ
ROYAUME-UNI

Canada Institute for Scientific and Technical Information (CISTI)

National Research Council
Acquisitions, Montreal Road, Building M-55
Ottawa K1A 0S2, CANADA

Les demandes de documents RTO ou AGARD doivent comporter la dénomination « RTO » ou « AGARD » selon le cas, suivie du numéro de série (par exemple AGARD-AG-315). Des informations analogues, telles que le titre et la date de publication sont souhaitables. Des références bibliographiques complètes ainsi que des résumés des publications RTO et AGARD figurent dans les journaux suivants :

Scientific and Technical Aerospace Reports (STAR)

STAR peut être consulté en ligne au localisateur de ressources uniformes (URL) suivant :

<http://www.sti.nasa.gov/Pubs/star/Star.html>

STAR est édité par CASI dans le cadre du programme NASA d'information scientifique et technique (STI)
STI Program Office, MS 157A
NASA Langley Research Center
Hampton, Virginia 23681-0001
ETATS-UNIS

Government Reports Announcements & Index (GRA&I)

publié par le National Technical Information Service
Springfield

Virginia 2216

ETATS-UNIS

(accessible également en mode interactif dans la base de données bibliographiques en ligne du NTIS, et sur CD-ROM)



BP 25
F-92201 NEUILLY-SUR-SEINE CEDEX • FRANCE
Télécopie 0(1)55.61.22.99 • E-mail mailbox@rta.nato.int



DISTRIBUTION OF UNCLASSIFIED RTO PUBLICATIONS

AGARD & RTO publications are sometimes available from the National Distribution Centres listed below. If you wish to receive all RTO reports, or just those relating to one or more specific RTO Panels, they may be willing to include you (or your Organisation) in their distribution.

RTO and AGARD reports may also be purchased from the Sales Agencies listed below.

Requests for RTO or AGARD documents should include the word 'RTO' or 'AGARD', as appropriate, followed by the serial number. Collateral information such as title and publication date is desirable.

If you wish to receive electronic notification of RTO reports as they are published, please visit our website (www.rta.nato.int) from where you can register for this service.

NATIONAL DISTRIBUTION CENTRES

BELGIUM

Etat-Major de la Défense
Département d'Etat-Major Stratégie
ACOS-STRAT – Coord. RTO
Quartier Reine Elisabeth
Rue d'Evère
B-1140 Bruxelles

CANADA

DRDKIM2
Knowledge Resources Librarian
Defence R&D Canada
Department of National Defence
305 Rideau Street
9th Floor
Ottawa, Ontario K1A 0K2

CZECH REPUBLIC

LOM PRAHA s. p.
o. z. VTÚLaPVO
Mladoboleslavská 944
PO Box 18
197 21 Praha 9

DENMARK

Danish Defence Research
Establishment
Ryvangs Allé 1
P.O. Box 2715
DK-2100 Copenhagen Ø

FRANCE

O.N.E.R.A. (ISP)
29, Avenue de la Division Leclerc
BP 72
92322 Châtillon Cedex

GERMANY

Streitkräfteamt / Abteilung III
Fachinformationszentrum der
Bundeswehr (FIZBW)
Friedrich-Ebert-Allee 34
D-53113 Bonn

GREECE (Point of Contact)

Defence Industry & Research
General Directorate, Research Directorate
Fakinos Base Camp, S.T.G. 1020
Holargos, Athens

HUNGARY

Department for Scientific Analysis
Institute of Military Technology
Ministry of Defence
H-1525 Budapest P O Box 26

ICELAND

Director of Aviation
c/o Flugrad, Reykjavik

ITALY

Centro di Documentazione
Tecnico-Scientifica della Difesa
Via XX Settembre 123
00187 Roma

LUXEMBOURG

See Belgium

NETHERLANDS

Royal Netherlands Military
Academy Library
P.O. Box 90.002
4800 PA Breda

NORWAY

Norwegian Defence Research
Establishment
Attn: Biblioteket
P.O. Box 25, NO-2007 Kjeller

POLAND

Armament Policy Department
218 Niepodleglosci Av.
00-911 Warsaw

PORTUGAL

Estado Maior da Força Aérea
SDFA – Centro de Documentação
Alfragide, P-2720 Amadora

SPAIN

SDG TECEN / DGAM
C/ Arturo Soria 289
Madrid 28033

TURKEY

Milli Savunma Bakanlığı (MSB)
ARGE ve Teknoloji Dairesi Başkanlığı
06650 Bakanlıklar – Ankara

UNITED KINGDOM

Dstl Knowledge Services
Information Centre, Building 247
Dstl Porton Down
Salisbury, Wiltshire SP4 0JQ

UNITED STATES

NASA Center for AeroSpace
Information (CASI)
Parkway Center, 7121 Standard Drive
Hanover, MD 21076-1320

SALES AGENCIES

NASA Center for AeroSpace Information (CASI)

Parkway Center
7121 Standard Drive
Hanover, MD 21076-1320
UNITED STATES

The British Library Document Supply Centre

Boston Spa, Wetherby
West Yorkshire LS23 7BQ
UNITED KINGDOM

Canada Institute for Scientific and Technical Information (CISTI)

National Research Council
Acquisitions
Montreal Road, Building M-55
Ottawa K1A 0S2, CANADA

Requests for RTO or AGARD documents should include the word 'RTO' or 'AGARD', as appropriate, followed by the serial number (for example AGARD-AG-315). Collateral information such as title and publication date is desirable. Full bibliographical references and abstracts of RTO and AGARD publications are given in the following journals:

Scientific and Technical Aerospace Reports (STAR)

STAR is available on-line at the following uniform resource locator:

<http://www.sti.nasa.gov/Pubs/star/Star.html>

STAR is published by CASI for the NASA Scientific and Technical Information (STI) Program
STI Program Office, MS 157A
NASA Langley Research Center
Hampton, Virginia 23681-0001
UNITED STATES

Government Reports Announcements & Index (GRA&I)

published by the National Technical Information Service
Springfield
Virginia 2216
UNITED STATES
(also available online in the NTIS Bibliographic Database or on CD-ROM)