

Reproduction of sinusoidal signals from cylinder recordings measured via non-contact methods.

Antony Nasce, Peter Boltryk, Martyn Hill, John W McBride

Electromechanical Research Group, University of Southampton, UK

A non-contact method of sound reproduction has great potential for sound archives, aiming to transfer the audio content from early sound recordings such as wax cylinders, which may otherwise be ‘unplayable’ via conventional stylus methods. If non-contact, optical methods are to be considered a viable solution for sound archives, a method for quantifying the quality of the reproduced audio signal needs to be developed. In this paper, methods for recovering the audio signal from a discrete surface map of a cylinder recording measured via non-contact surface metrology are described. A test cylinder recording, encoded with sinusoids provides the basis for a signal quality analysis. Non-contact and conventional stylus methods of sound reproduction are then compared using the same test cylinder. It is shown that non-contact methods appear to have distinct advantages over stylus reproduction, in terms of reduced harmonic distortion and lower frequency modulation.

PACS numbers: 43.38.Md, 43.38.Ne, 43.58.Ry

Nomenclature

$\Delta\theta$ – circumferential grid spacing.

Δx - linescan grid spacing.

$z(x_i, \theta_j)$. – the discrete cylinder surface, composed of linescan measurement.

$z_j(x)$. – linear profile measurement (linescan), measured along the x -axis at rotational index j .

$\tilde{Z}_j(k)$ - discrete Fourier transform of the j -th linescan sequence. (k represents spatial frequency).

λ_x – groove pitch (inter-groove spacing)

g_n – groove seed (the n -th groove encountered along the linescan)

$\tau_{x,\theta}$ – average groove phase shift estimation vector.

$A(x, \theta)$ – stylus trajectory vector.

$z_d(t)$ – the radial displacement track formed by the cutting tool.

1. INTRODUCTION

Sound reproduction from early mechanical recordings via non-contact, optical methods has been the topic of increasing interest in recent years [1-9]. A non-contact method ensures that no further damage or pressure is applied to the often fragile grooves, wherein the sound was originally encoded. In general, two non-contact reproduction strategies have been proposed in the literature:

- I. Real-time optical playback [1-4].
- II. Full surface mapping via optical sensors, followed by post-processing [5-9].

In (I), accurate sound reproduction requires robust tracking of the grooves with a laser, in situ. This strategy can be problematic in cases where the recording is damaged or deformed.

An alternative transfer strategy (II), uses optical sensing to measure the full surface topology of cylinder recordings. The philosophy behind this technique is aimed at preservation of the full recorded surface, not real-time playback.

The advantage of mapping the full surface topology is that sound can be extracted from anywhere within the groove, any number of times, at any point in the future. Tracking of the grooves in post-processing can therefore be more accurate than in real-time, especially in cases where the surface is highly deformed or damaged. Furthermore, this technique has potential for recovering sound from broken artefacts, by scanning fragments of the recording, and then stitching data sets together.

In this paper, we focus on sound recovery from cylinder recordings, although the methods described can also be applied to other vertically cut media such as phonograph tinfoil recordings and early Berliner discs.

1.1 CYLINDER RECORDINGS

Cylinder recordings, represent the earliest mechanical means for archiving and playing back sound. Recording systems and cylinder technology developed over many years, but the basic recording principles remained the same (see fig. 1). In recording mode, a cylinder ‘blank’ is set to rotate with constant angular velocity, $\dot{\theta}$. A cutting tool (stylus), which is coupled to a diaphragm-horn arrangement advances along the cylinder axis (x -axis), via a lead screw with constant pitch, λ_x . Prior to recording, the cutting stylus is forced into the outer cylinder surface, producing an unmodulated cut of fixed radial depth. During recording, incident acoustic pressure variations are channelled via a horn from the surrounding air, towards a thin diaphragm, forcing the stylus to cut undulations which are normal to the surface, (in the radial plane), about its mean cut depth. This radial modulation, $z_d(t)$ is often referred to as “hill-and-dale” or “vertical”, as opposed to the majority of monophonic flat disc recordings, which are encoded with a “side-to-side” or “lateral” modulation. The encoded groove follows a helix of pitch λ_x , with cross-section dependent on the stylus tip geometry. Cylinder recordings were typically recorded with a pitch of 100-200 turns per inch (t.p.i) at a rotational speed of 100-160rpm.

Modern electrical reproduction of cylinder recordings is achieved by tracing over the surface modulations, with a contacting stylus. The stylus transducer converts the physical “hill-and-dale” modulations into an electrical signal. Sound reproduction can also be achieved by mapping the full cylinder surface via non-contact, optical sensor, and then estimating $z_d(t)$, in post-processing.

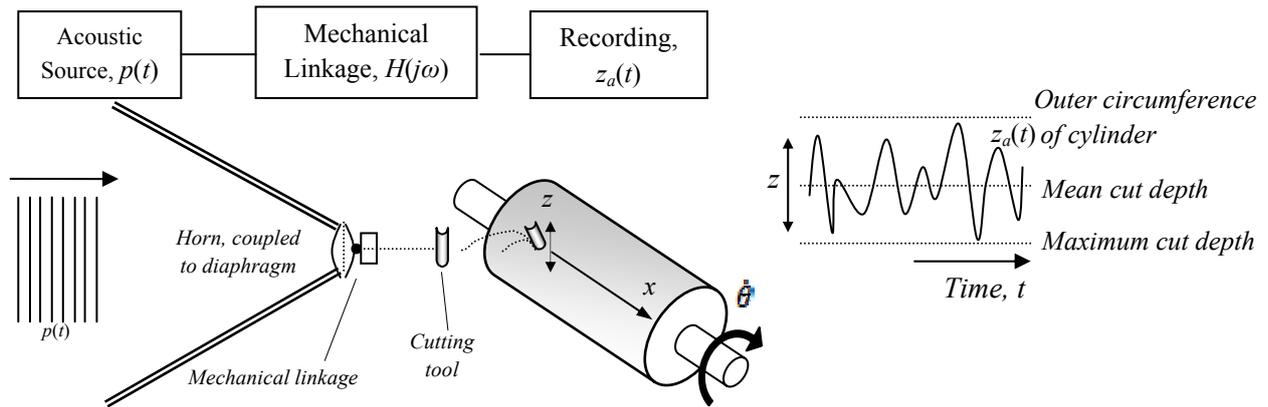


FIG. 1. Overview of the recording process for an acoustically recorded cylinder.

1.2 CYLINDER SURFACE MEASUREMENT AND DEFINITIONS

Figure 2 shows a cylinder recording mounted onto a non-contact measurement system. The methodology, which is typical of the approaches described in the literature [9,18] uses a single-point optical sensor to measure the full surface topology of the cylinder recording. A motion system is used to traverse the sensor linearly along the cylinder's axis, forming a 'linescan'. The cylinder is mounted on a rotatable mandrel; allowing the sensor to map the continuous cylinder surface $z(x, \theta)$, via a succession of linescan measurements. An additional linear stage provides a third axis of motion orthogonal to the linescan axis, which controls the sensor's stand-off distance from the artefact to keep the surface within the sensor's gauge range.

The discrete grid over which the cylinder is mapped is defined by the two spatial increments, Δx (in microns) and $\Delta \theta$ (in degrees), which are the linescan and rotational increments of the motion system respectively. We denote the discrete surface height matrix by $z(x_i, \theta_j)$. The j -th linescan, denoted by $z_j(x)$, is the linear scan measured along the x -axis at rotational index j . A typical linescan and surface map from a Blue Amberol cylinder is shown in figures 3 and 4.

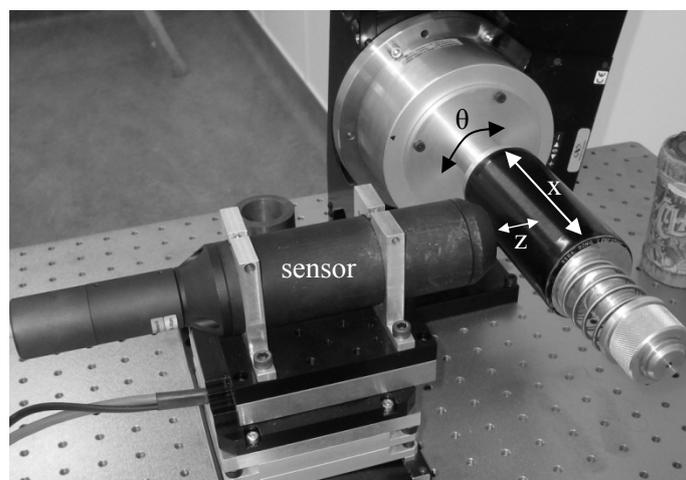


FIG. 2. Cylinder recording mounted on measurement system. The three measurement axes are shown.

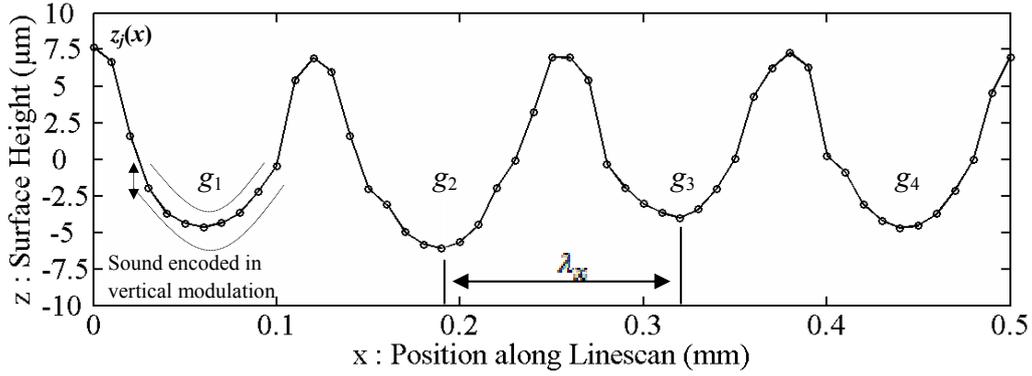


FIG. 3. A typical linescan of length $L = 0.5$ mm, measured from a Blue Amberol cylinder, ca. 1912 ($\Delta x = 10\mu\text{m}$). This cylinder was recorded with 200 t.p.i and has a groove spacing λ_x , of $127\mu\text{m}$. The *groove seed*, denoted by g_n is the n -th groove cross-section along the linescan.

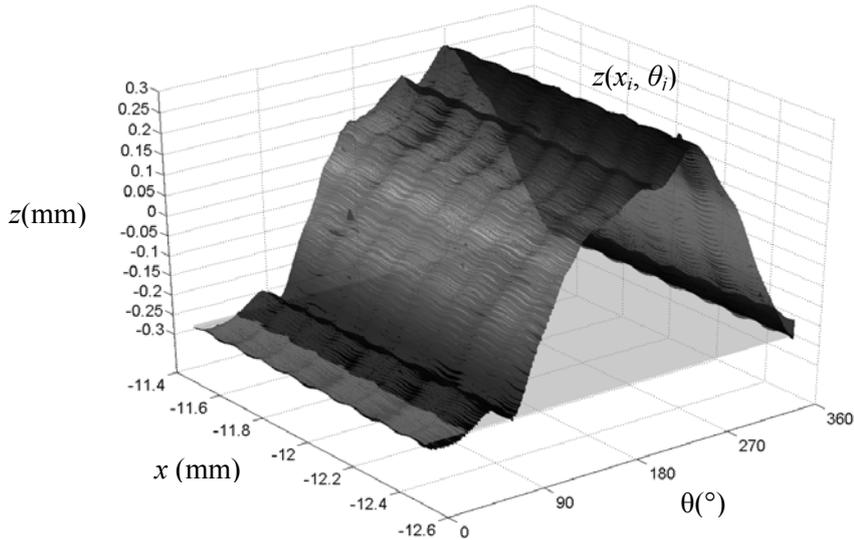


FIG. 4. Surface measurement of a Blue Amberol cylinder surface. The surface is built up through a succession of linescans. The large deviation in z is due to the lack of cylinder concentricity.

1.3 RESOLUTION REQUIREMENTS FOR SOUND REPRODUCTION

To ensure that sound can be accurately recovered from the cylinder, sampling resolution of the measurement system must be considered, to avoid aliasing and to optimise scanning time.

The choice of Δx determines the number of samples over which the groove cross-section is resolved along the linescan. This is dependent upon the cylinder pitch and complexity of the groove cross-section. For a cylinder recorded with 200 t.p.i, a linescan grid spacing of $\Delta x = 10\mu\text{m}$, gives approximately 13 samples per groove cross-section (See FIG. 3 for example). This sampling has been previously judged to provide sufficient resolution for sound reproduction [9].

The choice of $\Delta\theta$ (in degrees) determines the playback sample rate f_s , of the recovered audio signal. Assuming that the cylinder was recorded with a constant rotational speed of ω (in revolutions per minute), the estimated playback sample rate is given by:

$$f_s = \frac{6\omega}{\Delta\theta} \quad (1)$$

To avoid aliasing of the signal by the Nyquist criterion, we have:

$$f_s > \frac{f_{\max}}{2} \quad (2)$$

where f_{\max} is the maximum frequency limit of the cylinder record. For cylinder recordings, this upper limit is around 5 kHz [9,10]. For 160rpm cylinders, an angular sampling increment of $\Delta\theta = 0.1^\circ$ gives a playback sample rate of 9.6 kHz, which should be considered as an absolute minimum for accurate sound reproduction. Current archival recommendations [11] recommend a playback sample rate of 96 kHz, which is achievable with angular sampling of $\Delta\theta = 0.01^\circ$.

The third sampling dimension is the axial resolution of the sensor, which is the ability of the sensor to resolve displacements in the z -plane. It was shown [18] that in order to resolve the smallest of displacement amplitudes (high frequency, low amplitude) on cylinder recordings, an axial resolution in the order of 10nm (10×10^{-9} m) is required. The current measurement system uses a sensor with an axial resolution of 10nm.

2. REVIEW OF RELATED TECHNIQUES

2.1 SOUND RECOVERY FROM SURFACE MAPS OF CYLINDER RECORDINGS

Sound reproduction from discrete surface maps of cylinder recordings presents a new area of research, and there are currently no standardised methods for audio signal estimation in the literature.

In a previous collaborative study [9], between Lawrence Berkeley National Laboratory and the University of Southampton, a method was described for extracting sound from a discrete surface map of a Blue Amberol cylinder (c. 1912), measured via single-point optical displacement sensor, as described in Section 1.2.

In [9], the radial displacement track encoded by the cutting tool, $z_a(t)$ is found by initially estimating the local minimum height (minima) of each groove cross-section, for all linescans in the discrete surface, $z(x_i, \theta_j)$. Groove minima data is then re-organized into a time series, which follows a helical trajectory. A list of groove valleys (minima) is initially located via a neighbourhood search along each linescan, and outlying candidates are removed. A set of points in the k -th groove valley are then fitted to a quadratic function, $H_k(x) = A_k x^2 + B_k x + C_k$. The quadratic term (A_k) remains fixed, and is chosen such that the curvature of the parabola is an approximation to that of the original cutting tool, which produced the recording. The depth (minimum ordinate) of the fitted quadratic curve is then used to form a radial displacement track estimate, $z_a(t)$, corresponding to the encoded sound. The concept of three stand alone data streams were considered:

- The groove bottom (valley)
- The groove top (ridge)
- The groove bottom position with respect to the groove top position (top-bottom).

It was noted that the use of the ridge data stream, provided audible, albeit noisy sound content, but with interference from different times in the recording. This kind of temporal distortion was also noted for real-time playback systems in [2], when the width of the laser beam spot was large compared with the groove width. The use of ‘top-bottom’ subtraction in [9] was shown to remove low frequency structures (below 150Hz), due to surface form.

The audio signal reproduced in [9] was compared with a stylus transfer of the same recording. Visual comparisons of the time histories and spectra showed that the optical method provided an ‘accurate audio transcription’, compared with the stylus playback, but no quantitative analysis of the audio signal quality was presented.

2.2 SIGNAL QUALITY TESTING FOR NON-CONTACT SOUND REPRODUCTION

In order to obtain a quantitative assessment of sound quality, a signal quality analysis is presented in [12], to determine the feasibility of their optical methods of sound reproduction for flat disk recordings, and to make comparisons with stylus playback. The signal-to-noise ratio (SNR) and total harmonic distortion (THD) are determined through the use of test records, encoded with single frequency tones. In this way, the recovered signal is of known type, thus the SNR and THD be calculated by estimating the ratio of powers in the signal and noise bands.

A summary of discrete-time methods for estimating the signal-to-noise power ratio estimates is given in [13]. Two algorithms are described for estimating SNR of a noisy sinusoid from discrete-time data. The first is based upon estimating directly, the four parameters of a sine wave (amplitude, frequency, phase and dc-offset), the second is based on a spectrum averaging method, via the DFT.

Estimates for SNR using spectrum averaging techniques are given in [12], for various test discs, recorded with a 300 Hz tone. The SNR is defined as the ratio between the signal power P_{sig} and noise power P_n , for tracks containing a single frequency, f_0 . The SNR is expressed in decibels and is given by:

$$SNR = 10 \log_{10} \left(\frac{P_{sig}}{P_n} \right) \quad (3)$$

The signal and noise powers are derived from discrete power spectral estimates of the recovered tonal waveform, where P_{sig} is calculated over a frequency range of, $\pm f_{margin}$ around the centre frequency, f_0 . (In [12], $f_{margin} = 5$ Hz). The noise power P_n is the power in the remaining frequency bins, outside of the range $f_0 \pm f_{margin}$, from 100 Hz up to 10 kHz.

Another method of measuring the quality of sinusoidal signals is to examine THD. The spectrum of an ideally recovered sinusoid should contain a single peak at the fundamental frequency. In practice, geometric distortion of the sine wave can be introduced at time of recording or at time of measurement. Harmonic distortion can also be introduced by errors in the signal recovery method itself and this effect must be minimised. In [12] the following definition for THD is given:

$$THD = 10 \log_{10} \left(\frac{P_{harm}}{P_{sig}} \right) \quad (4)$$

Where P_{sig} is defined as in Eq. (3) and P_{harm} is the power of the harmonics ($2f_0, 3f_0, \dots$) upto 10 kHz. Harmonic peaks are also considered on a fixed frequency width of $\pm f_{margin}$ around the peak frequency, with $f_{margin} = 5$ Hz.

3. METHODOLOGY

3.1 PROPOSED METHODS OF AUDIO SIGNAL RECOVERY

In previous work [9], the estimation of stylus trajectory and groove depth were not considered independently, but as a single process. For damaged cylinder surfaces, it is useful to visually assess the estimated trajectory before extracting the sound, to ensure that the trajectory is coherent (no skipping of grooves occurs) and complete (no regions containing sound are missed). For this reason, we consider sound recovery in two independent stages – trajectory estimation and audio signal estimation. A new approach for estimating the stylus trajectory and two groove depth estimation methods are now described.

3.1.1 Trajectory Estimation

In [14], a method of rapidly approximating the trajectory produced by the cutting tool was introduced, by modelling the linescan as a periodic signal, and measuring the relative shift between linescans around the circumference. Reasons for adopting this strategy include the reduction in computation time and the ability to observe and account for any potential phase offsets between adjacent linescans, caused by measurement error.

Through the recording process, the linescan has a periodic structure, with adjacent grooves separated by a constant distance λ_x , corresponding to the groove pitch. In most cases, the groove spacing, λ_x can be derived from historical records or by observation, but for completeness and records of unknown type, it can be found numerically by spectral analysis of the linescan.

The discrete Fourier transform (DFT) of the j -th linescan, is given by the complex sequence $\tilde{Z}_j(k)$, where k denotes the k -th spatial frequency bin. The fundamental spatial frequency of the linescan, k_0 is related to λ_x via the reciprocal relationship of period and frequency and is determined by calculating the expectation for all linescans:

$$k_0 = E [\max | \tilde{Z}_j(k_p) |], \quad j \in \theta : [0^\circ, 360^\circ] \quad (5)$$

where k_p is the range of possible frequency bins in which k_0 is valid, which for cylinder recordings of the era is between 3.9 and 7.9 mm^{-1} .

Fig. 5 shows that if we observe a single groove valley at linescan at $\theta = 0^\circ$ and track its shift along the x -axis around the cylinder circumference until linescan at $\theta = 360^\circ$, the x -position of the valley propagates through a distance of λ_x . For a cylinder recorded with constant pitch, this shift is approximately equal for all grooves. An estimate of this shift vector, $\tau_{x,\theta}$ can therefore be used to form an approximate stylus trajectory.

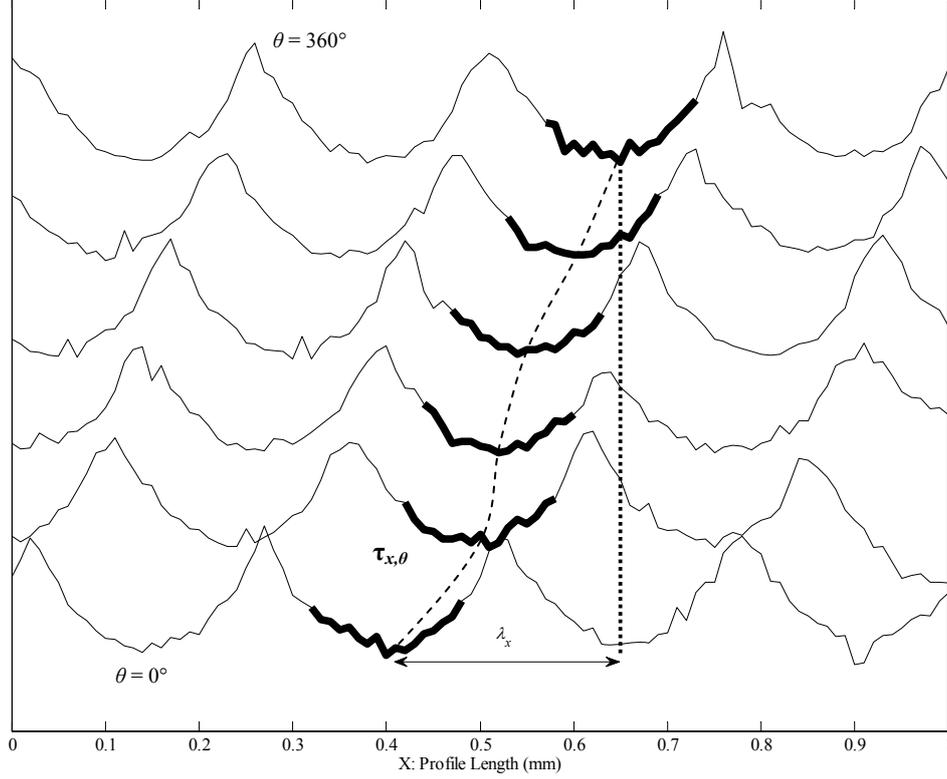


FIG. 5. Linescan measurements taken at six angular intervals from 0 to 360° around the cylinder circumference. The path traced by $\tau_{x,\theta}$, shows the shift along the x -axis of the highlighted groove valley. This shift vector remains approximately equal for all groove valleys and can be estimated by calculating the relative phase shift between adjacent linescans.

The shift vector $\tau_{x,\theta}$, can be found by first estimating the relative phase (shift) between linescans from 0 to 360°. With k_0 determined from (5), the phase estimate of the j -th linescan is given by:

$$\varphi_j = \arctan \left[\frac{\text{Im}\{\tilde{Z}_j(k_0)\}}{\text{Re}\{\tilde{Z}_j(k_0)\}} \right], j \in \theta : [0^\circ, 360^\circ] \quad (6)$$

The result from Eq. (6) is given in radians, thus in order to obtain $\tau_{x,\theta}$, the phase angle vector φ_j must be scaled such that a phase shift of 2π radians is equal to the distance λ_x . It should be noted that an equivalent trajectory estimate may also be obtained using cross-correlation between linescans, but in this example, a Fourier-based method is described.

The resultant shift vector $\tau_{x,\theta}$ is then rounded to the nearest grid position along the x -axis, so that the discrete surface data can be referenced. Fig. 6a shows an example of $\tau_{x,\theta}$, overlaid onto the surface of a cylinder surface map.

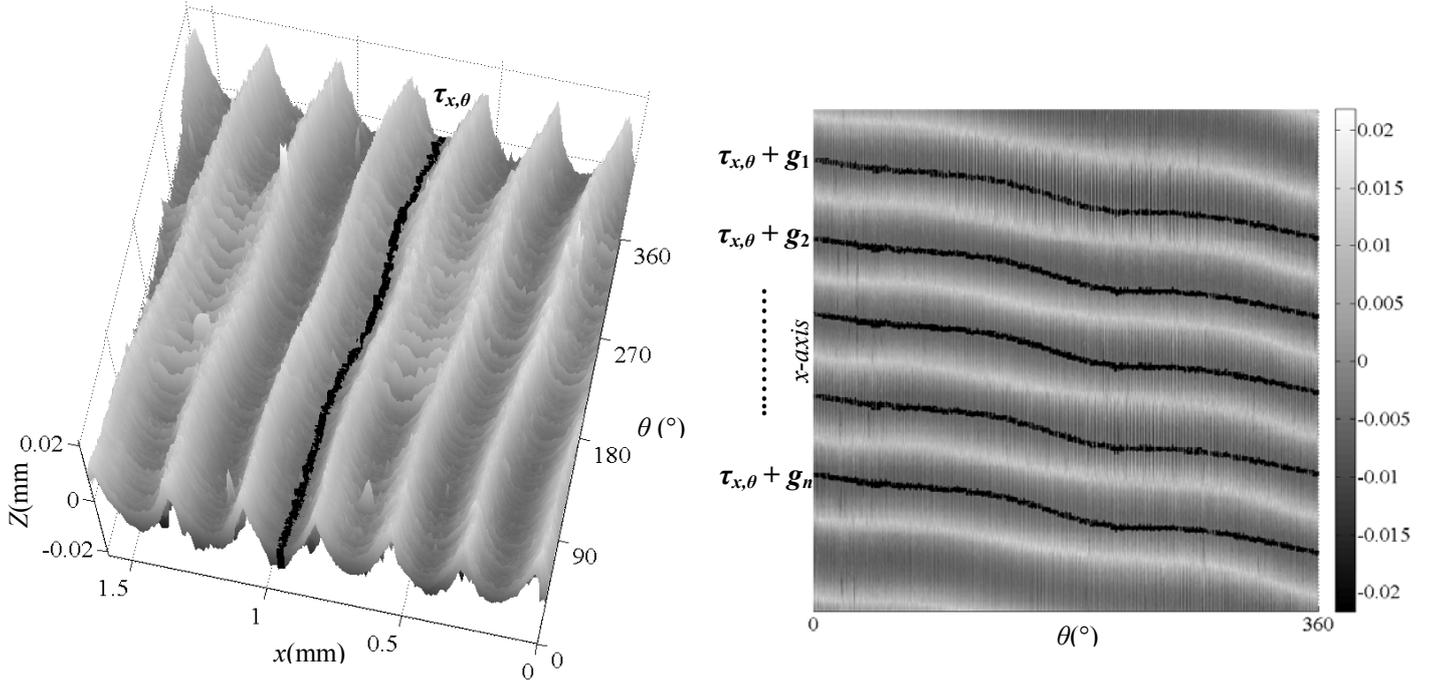


FIG. 6 : (a) Scaled, discretised shift estimation vector $\tau_{x,\theta}$, overlaid onto a 3-D surface plot of the test cylinder data. (b) Formation of stylus trajectory vector $A(x,\theta)$, using the phase shift estimate $\tau_{x,\theta}$ and groove seed positions, $g_1, g_2 \dots g_n$.

This shift vector $\tau_{x,\theta}$ can then be used as a ‘template’ for segmenting the surface into groove regions (groove seed). The n -th groove seed, g_n (see Fig. 3) is found by locating the medial axis of each groove valley at 0° . For smooth, symmetric groove cross-sections, the medial axis, is indexed by the local minimum point of the groove and can be found via a simple neighbourhood search.

In cases where the groove is rough and/or asymmetric (see Fig 7. for example), the medial axis does not always lie at the local minimum point. To ensure that each groove is indexed by its medial axis, an appropriately designed Savitzky-Golay (SG) polynomial smoothing filter is applied to the linescan $z_j(x)$ to form the filtered linescan $z_j'(x)$. The local minima of this SG filtered linescan are then used to locate the groove seed positions.

SG filters were initially developed [17] to identify the relative widths and heights of spectral peaks noisy spectrometric data. They work by a local polynomial regression (of degree P) on a distribution (of at least $P+1$ equidistant samples) to determine the smoothed value for each point. A carefully designed SG filter can be used as a means of locating the medial axis of the groove valley.

Fig. 7 shows that with a cubic SG filter ($P = 3$), and window length $L = 25$ (approximately equal to the number of samples across each groove cross-section), the position of minima in the filtered linescan, $z_j'(x)$ are located closer to the middle of the groove valley (medial axis) than minima found using the raw linescan $z_j(x)$.

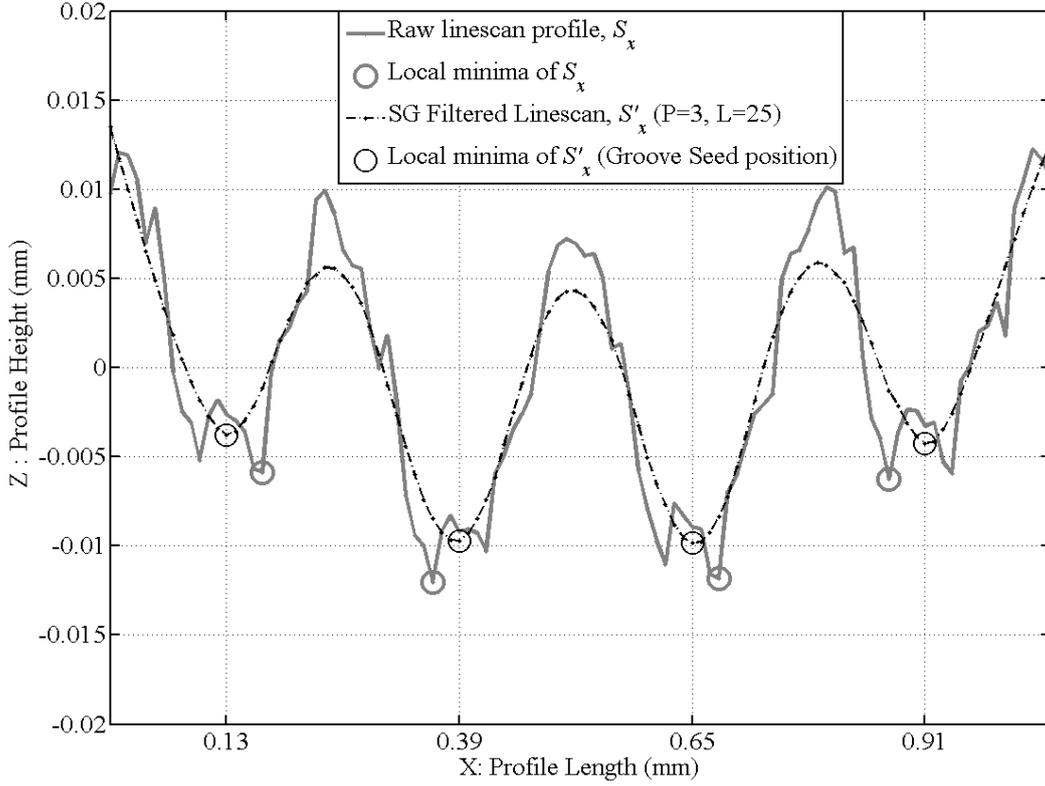


FIG. 7. The abscissa of local groove minima from the raw linescan profile, $z_j(x)$ varies from one groove to the next due to the roughness of the cutting tool. The abscissa of minima from the filtered linescan $z'_j(x)$ can be seen to be spaced equidistantly, by a distance of $\lambda_x = 0.254$ mm, which is the expected spacing for a cylinder recorded at 100tpi. The local minima of $z'_j(x)$ form the groove seed positions, $g_1, g_2 \dots g_n$.

A stylus trajectory vector, $\mathbf{A} = A(x, \theta)$, is then constructed (see Fig. 6b) by positioning $\tau_{x, \theta}$ at the groove seed positions, $g_1, g_2 \dots g_n$, by:

$$A(x, \theta) = \begin{bmatrix} [\tau_{x, \theta} + g_1]^T & [\theta]^T \\ \vdots & \vdots \\ [\tau_{x, \theta} + g_n]^T & [\theta]^T \end{bmatrix} \quad (7)$$

3.1.2 Audio Signal Estimation

Trajectory estimation provides an approximation to the helical path traced by a reproduction stylus. The radial displacement signal, $z_a(t)$ which corresponds to the encoded sound is recovered by a discrete estimate of the groove depth along the trajectory vector \mathbf{A} , defined in Eq. (7).

Proposed schemes for deriving a discrete estimate for $z_a(t)$, include:

- i. Single point data streams (using raw Z -data, no averaging)
- ii. Averaging samples across the groove cross-section [14].
- iii. Stylus model, fitting based methods [9].
- iv. Rule-based methods.

In [14], the notion of a ‘groove matrix’ was introduced. The groove matrix, denoted by \mathbf{G}_i is a matrix of size n by m , where n is the number of groove-cross sections (corresponding to the number of

samples in the time domain) and m is the number of samples across the groove-cross section, found along the trajectory vector \mathcal{A} .

Representing the groove depth by a single discrete data point (i) is not ideal for rough groove cross-sections, as the selection of the sample is very sensitive to noise. Scheme (ii) was investigated in [14] by averaging data streams at the groove bottom and side-wall. The cylinder in question was a moulded Blue Amberol cylinder. In this case, the groove shape was symmetric, and the abscissa of groove minima remained relatively unchanged from one groove to the next. A comparative examination of three different signal estimates in the time and frequency domain showed that the SNR increased as the estimate approached the bottom of the groove cross-section.

The parabola fitting method described in Section 2.1 and [9], based upon scheme (iii) seems well suited for cylinder recordings with a smooth, symmetric groove cross-section, as shown in Fig. 8b, where the quadratic fit agrees well with the groove shape. In some cases however (see Fig. 8a and 8c for example), the cylinder groove cross-section is irregular, asymmetric and does not resemble the quadratic fit.

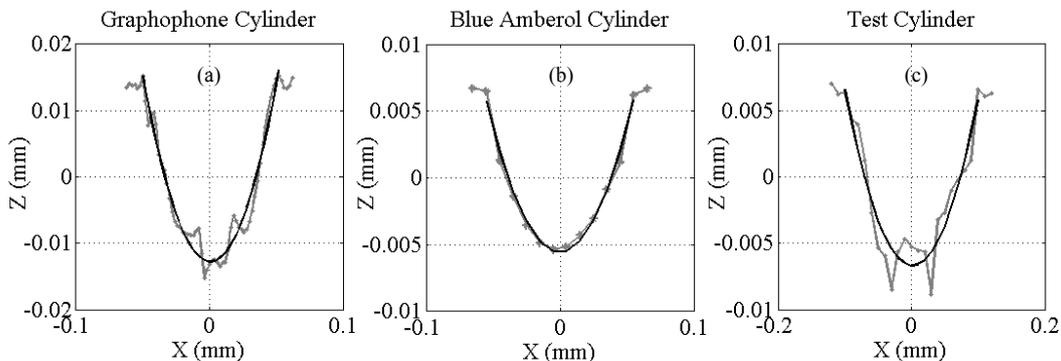


FIG. 8. This figure shows three groove valley regions. The smooth dark line in each plot shows a polynomial fit of order 2, which best fits the groove data, in a least squares sense. The Blue Amberol Cylinder (b) has strong agreement with the quadratic curve, whereas the Graphophone (a) and Test Cylinder (c) example shows a more complex groove shape, due to the cutting tool which produced the recording, and increased surface roughness/damage.

In this paper we compare two methods of estimating the radial displacement track, $z_a(t)$, based on schemes (ii) and (iii). Both methods use the trajectory $\mathcal{A} = \mathcal{A}(x, \theta)$, as defined by Eq. (7) as a guide for the approximate path followed by the cutting tool.

The cylinder used for test purposes has an irregular groove cross-section (as shown in Fig. 8c.), which is in contrast to the smooth groove cross-section described in [9] and [14]. The test cylinder, which contains signals of known type will allow a quantitative assessment of signal quality, and a comparison of the two different estimation methods for $z_a(t)$.

Method A provides the most simplistic estimate for the groove depth, based on averaging of samples across the groove cross-section. By investigating different length sample averages, it is possible to observe the effects of mean averaging for the recovered signal, in terms of SNR and THD.

In Method B, a Savitzky-Golay filter is used to firstly smooth the groove cross-section, to reduce the effects of noise. The local minima of the smoothed cross-section is then used as an estimate of the groove depth. Through the use of different combinations of filter polynomial order (P) and filter length (L), an optimal filter can be found by evaluating the SNR and THD of the recovered signal.

The best signal estimates derived by Method A and B can then be compared in terms of signal-to-noise ratio (SNR) and total harmonic distortion (THD). These methods are now described:

METHOD A: Un-weighted mean Sample averaging across the groove matrix

- A groove matrix \mathbf{G}_t of size n by m is formed from the trajectory vector $\mathbf{A} = A(x, \theta)$, where n is the number of samples in the time domain and m is the number of samples across the groove cross-section.
- Radial displacement estimates $z_a(t)$ are then derived from \mathbf{G}_t by an un-weighted mean averaging process across the n -th row of \mathbf{G}_t .
- The mean average is taken from the central column of the \mathbf{G}_t , which corresponds to the medial axis of the groove cross-section. The number of points included in the average varies from $N=1$ (central column of the \mathbf{G}_t only) to $N=25$ points (± 12 samples either side of the central column of \mathbf{G}_t).

METHOD B : Use of Savitzky-Golay Filters of to locate unique groove minima.

- Raw surface data is filtered along each linescan via SG filter of polynomial order P , and frame length L .
- Along the trajectory vector $\mathbf{A} = A(x, \theta)$, the radial displacement estimate $z_a(t)$ is found by locating the local minimum point of the filtered surface.
- Different SG filters of orders $P=\{1,3...9\}$ and frame length $L=\{11,13,...25\}$ are used to obtain unique groove minima at different positions in the groove cross-section.

3.2 METHOD FOR COMPARING AUDIO SIGNAL QUALITY

3.2.1 Test Cylinder

A wax cylinder recording was produced for purposes of assessing the quality of the recovered signals. The cylinder was recorded at Poppy Records [15]. The recording stylus was not driven by acoustic pressure, but was directly cut by an electrical transducer. Pure tones (sinusoids) in the range of 200Hz – 5 kHz were recorded onto a cylinder at 160rpm with a pitch of 100 t.p.i.

Following the recording process, the groove structure remained untouched by stylus or cleaning. The surface map produced by the optical scanning system used a measurement grid of: $[\Delta x = 10\mu\text{m}, \Delta\theta = 0.01^\circ]$. This sampling scheme gave approximately 25 data points per groove cross-section, and a playback sample rate $f_s = 96$ kHz. The recovered signals exhibited some frequency modulation and harmonic distortion, which is in part due to the nature of the recording process. Recovered signals from the 1 kHz tone region provide the basis for comparison of audio signal estimates.

3.2.2 SNR Analysis

The signal to noise ratio (SNR) is computed by locating the peak frequency and then calculating the total noise power in the remaining frequency bins, (considered to be noise). Definitions of the signal and noise powers, P_{sig} and P_n , as described in Eq. 3 are not immediately obvious for pseudo-sinusoidal signals which are frequency modulated and contain harmonic distortion. Care must therefore be taken when deciding what constitutes the signal and noise. For example, it is questionable whether energy from harmonics in the spectrum should be considered as part of the signal or noise band. The choice of frequency band in which to search for P_{sig} is complicated by the fact that the recovered signals were frequency modulated. Other factors affecting the calculation include window length and frequency resolution of the DFT.

The short-time Fourier transform was used to carry out spectral analysis on windowed segments of the 1 kHz tone. Each frequency window was of duration 0.125 seconds, with a frequency bin resolution of 2 Hz. The signal power P_{sig} is found by locating the peak in the power spectrum (fundamental frequency) around 1 kHz.

Frequencies which contribute to the noise power P_{noise} band were considered in two ranges:

- Below 1kHz : 200-950 Hz
- Above 1kHz : 1050 Hz- 10 kHz.

Harmonic frequencies above the fundamental were considered as undesirable and were also included in the noise band calculation. Frequencies below 200 Hz were not included in the noise band calculation, as these were attributed to the surface form and not the audio signal.

3.2.3 Total Harmonic Distortion Analysis

The THD is the ratio of the sum of the powers of all harmonics components above the fundamental frequency, to the power of the fundamental. Expressed as a percentage, the THD is given by:

$$THD(\%) = 100 \times \frac{\sqrt{F_2^2 + F_3^2 + F_4^2 + F_n^2}}{F_1} \quad (8)$$

F_n is the Fourier component of the n -th harmonic, where $n = 1$ denotes the fundamental frequency. In the case of the 1 kHz tone, the first four harmonics (2 kHz, 3 kHz, 4 kHz and 5 kHz) are included in this calculation, as harmonics above 5 kHz become lost in background noise. The lower the harmonic distortion, the more the recovered signal resembles a true sinusoid.

4. RESULTS

4.1 SIGNAL QUALITY COMPARISONS FOR METHODS A AND B

SNR and THD results for Method A are shown in Figure 9.

Figure 9 shows that the SNR increases as the number of points in the average window, N increases, but care must be taken when interpreting this trend. The best SNR for Method A comes from a 21pt mean average ($N = 21$), however this signal did not have the lowest THD, and upon listening to the signal, a small echo effect was noticeable. This kind of temporal distortion is not accounted for in the

spectrum averaging technique and is discussed in Section 4.2. The signal which had lowest THD distortion (and hence the ‘purest’ sinusoid) was from a mean average of $N = 15$.

Mean averaging is a low pass filtering operation, since its effect is to allow lower spatial frequencies to be retained, whilst suppressing higher frequency components. A larger mean average will remove more noise (high frequency), but can reduce the level of detail in the groove cross-section. The result of this for the recovered sinusoid signals is a reduction in peak signal power and increased harmonic distortion when N is large (greater than 19 samples). In addition, by listening to signals where N is large, temporal distortion is also present in the form of an echo. A mean average of $N = 15$ produced a signal with lowest THD, reasonably high SNR score, no noticeable temporal distortion, and was therefore considered to be the best signal produced by Method A.

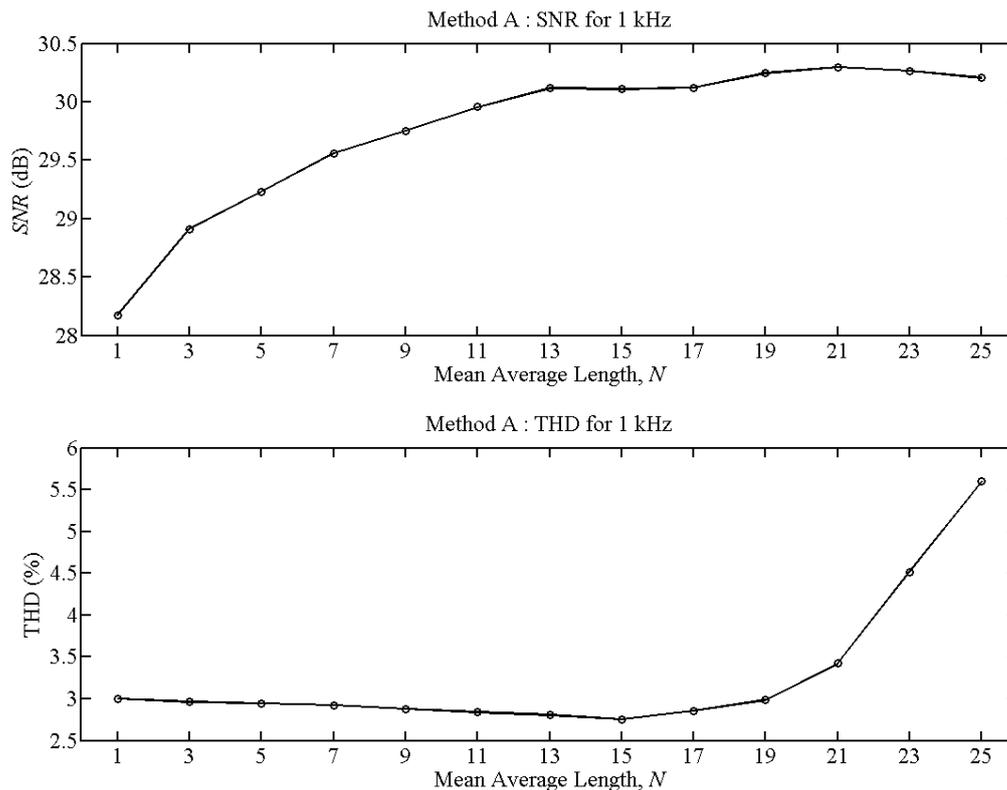


FIG. 9. SNR and THD results for Method A (mean averaging across the groove matrix).

SNR and THD results for Method B for the different SG filters are shown in Figure 11.

In general, the SNR decreased as the filter order P increased. This is due to the fact that higher order polynomial filters retain the higher spatial frequency content (roughness) of the groove cross-section (see Fig 10.), which results in increased high frequency noise for the audio signal. The highest SNR came from filter $[P=1, L=15]$. An SG filter of order $P = 1$ is essentially an un-weighted, moving average filter, similar to the mean averaging in Method A. This similarity can be observed by noticing that the THD curves for Method A and Method B, ($P = 1$) have similar trends. An echo effect was also noticeable for $[P = 1, L = 19-25]$, as was found with Method A for large N .

With the exception of the 1st Order filter ($P=1$), the THD scores remained relatively constant as L increased. The lowest THD scores came from the higher order filters, $[P=9, L=21]$ for example. This may be due to the fact that the unique groove minima produced by higher order SG filters tended

towards the positions of raw groove minima, as opposed to being positioned higher up in the groove cross-section, where raw displacement data does not exist, as can be seen in Fig. 11.

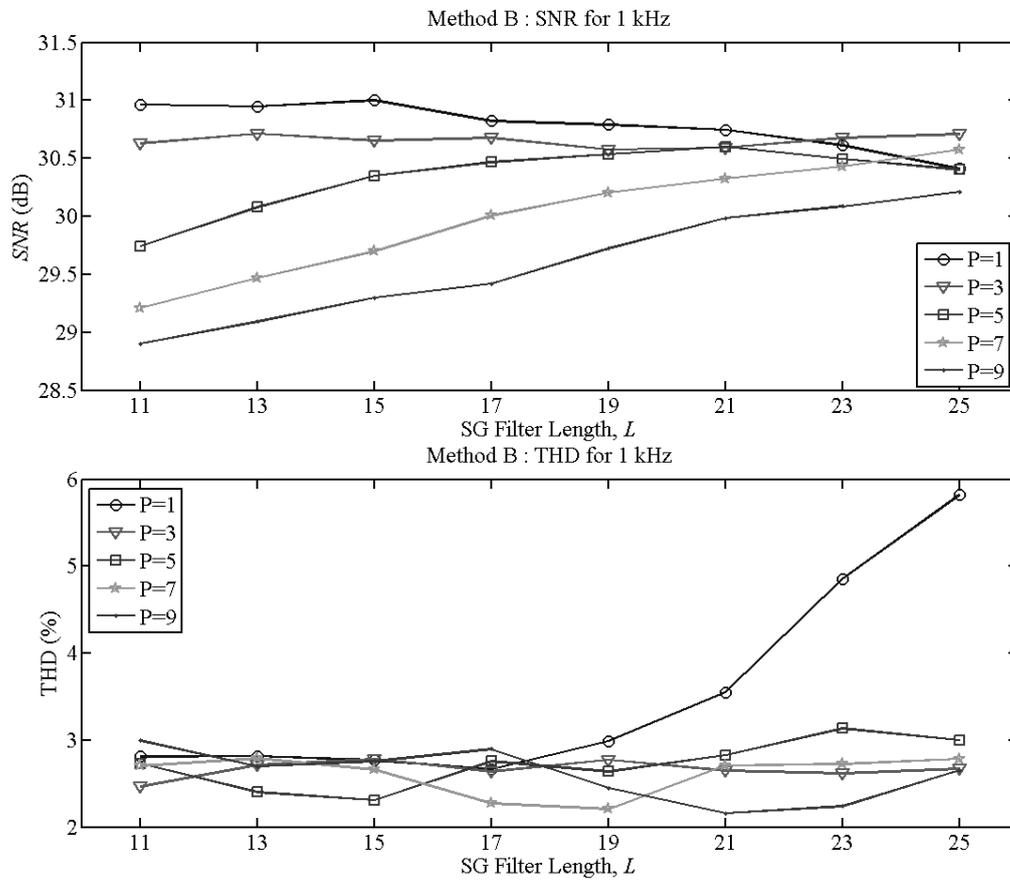


FIG. 10. SNR and THD Results for Method B (SG filtering).

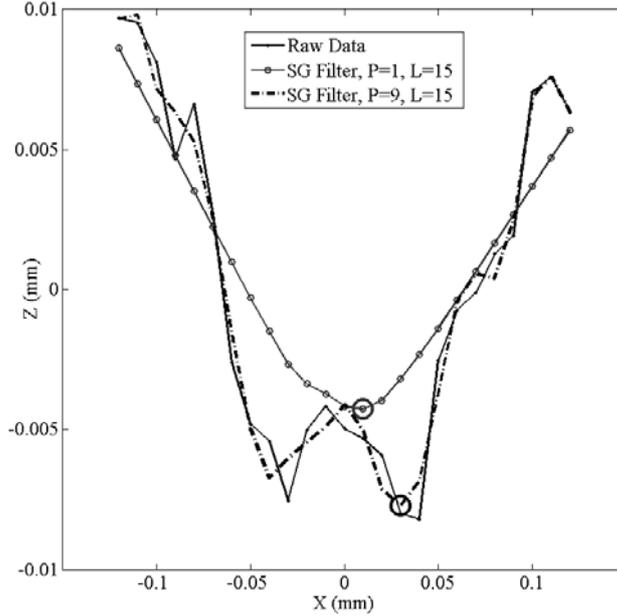


FIG. 11. 1st and 9th order SG filters ($L=15$), applied to the groove cross-section of test cylinder data. The 9th order filter retains more of the higher spatial frequencies, whereas the 1st order filter produces a smoother cross-section. Large circles indicate the position of unique groove minima from the filtered linescan, used to form the displacement estimate, $z_a(t)$.

A summary of the best SNR and THD results are given in Table 12. In both cases, an appropriately designed SG filter (Method B) gave better signal quality scores, compared with simple mean averaging (Method A). This suggests that mean averaging is not optimal for noisy, asymmetric groove cross-sections, as seen with the test cylinder data.

Frequency (Hz)	METHOD A (Mean Averaging)	METHOD B (SG Filtering)
Highest SNR (dB)	30.3	31.0
Lowest THD (%)	2.78	2.16

TABLE. 12: Summary of best SNR and THD results for Method A and Method B.

4.2 SOURCES OF NOISE AND DISTORTION

A time-frequency analysis shows that the recovered sinusoid signals exhibit the following types of noise and distortion:

Low Frequency Noise

The low frequency noise (below 200 Hz) is mainly due to the macroscopic surface form of the cylinder, which should be accounted for through an equalisation stage. In order to match the velocity response of a magnetic cartridge (as is common with stylus reproduction), an appropriate equalisation should be applied. In this paper we have only dealt with the raw displacement signals $z_a(t)$ which were not equalised or differentiated in any way.

Impulsive Noise

Impulsive noise can be caused by surface debris which was not cleaned from the cylinder prior to scanning. Any debris which lies above the imaging plane will therefore be measured by the sensor. This can result in transients for the recovered audio signal. In order to minimise this effect, any loose debris is removed from the surface prior to scanning wherever possible. Additionally, if the cylinder surface is measured in piece-wise segments (as is sometimes required for heavily warped surfaces [Boltryk]), impulsive noise can be introduced at segment joins if the segments are not appropriately aligned in software prior to sound extraction.

Temporal distortion

It is observed that displacement estimates which used groove cross-section data towards the tops of the groove exhibited an echo. This phenomenon was first noted in [1] for real-time optical playback of phonograph cylinders, when the width of their laser beam was wide compared with the groove width. Similarly in [9], the author notes that groove ridge data, provided audible, albeit noisy sound content, but with interference from different times in the recording. The test cylinder data allows us to quantify the interference observed from the ridge data stream. Figure 13 shows a time-frequency analysis (spectrogram) of data streams derived from the top and bottom of the groove cross-section. It appears that an echo is introduced due to the superposition of adjacent grooves at times t and $t+T$, where T is the time taken for one complete rotation (0.375 seconds in this case for 160rpm). To avoid introducing temporal distortion, it is therefore suggested that the use of displacement data from the groove ridge should be avoided.

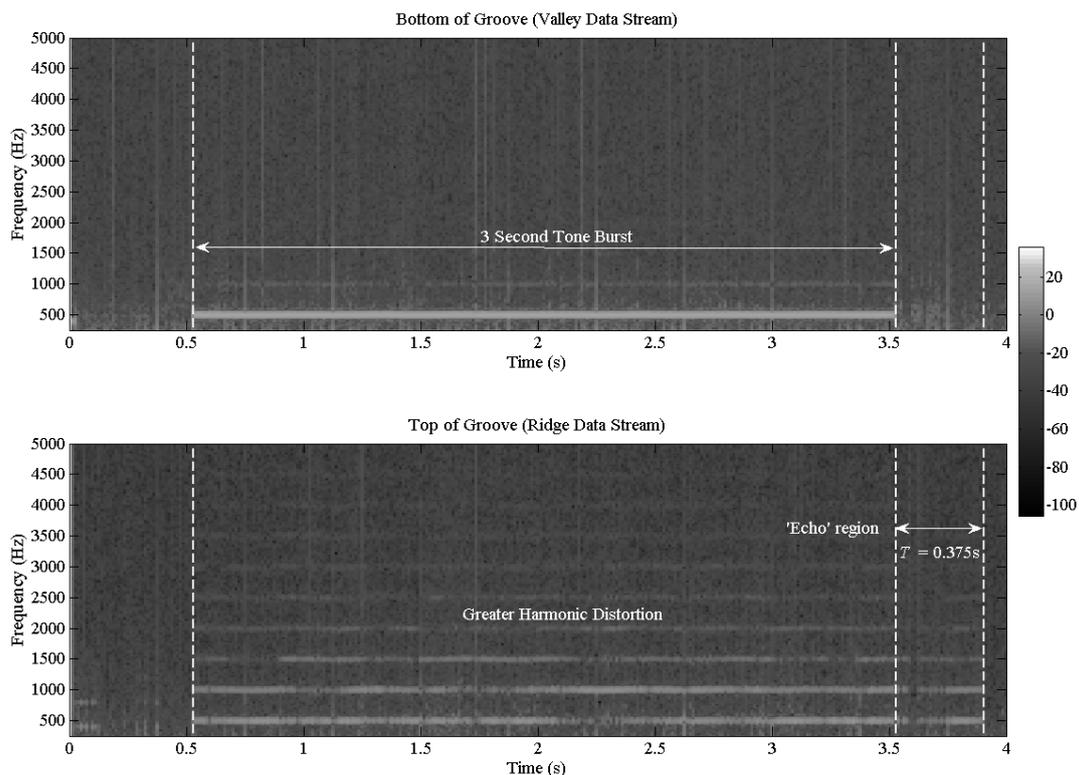


FIG. 13. Spectrogram analysis of signals recovered from the top and bottom of the groove cross section for 500 Hz tone. An echo is clearly visible in the ridge data stream due to the superposition of adjacent groove displacement data at times t and $t+T$. In addition, the ridge data stream has higher harmonic distortion.

4.3 COMPARISON OF STYLUS AND NON-CONTACT REPRODUCTION

In addition to optical reproduction, the test cylinder audio was also transferred using conventional stylus techniques at the British Library Sound Archive. Due to difficulties in matching the equalisation of the stylus and optical reproduction, direct signal quality comparisons are not presented in this paper. Subjective listening suggests that stylus reproduction has a higher SNR, in terms of a lower noise floor power and less impulsive noise. However, the accuracy of sine-wave reproduction in terms of harmonic distortion and frequency modulation appears to be greater with the non-contact method.

Harmonic Distortion

The spectrogram in Figure 14 compares the reproduction of a 2 kHz tone from both stylus and optical methods. Stylus reproduction appears to show increased harmonic distortion compared with the optical signal. This is likely due to the stylus-groove interaction, known as ‘tracing distortion’ [16], where the curvature of the reproduction stylus is large compared with the curve of the modulated groove. Non-contact reproduction is not affected by such dynamic effects.

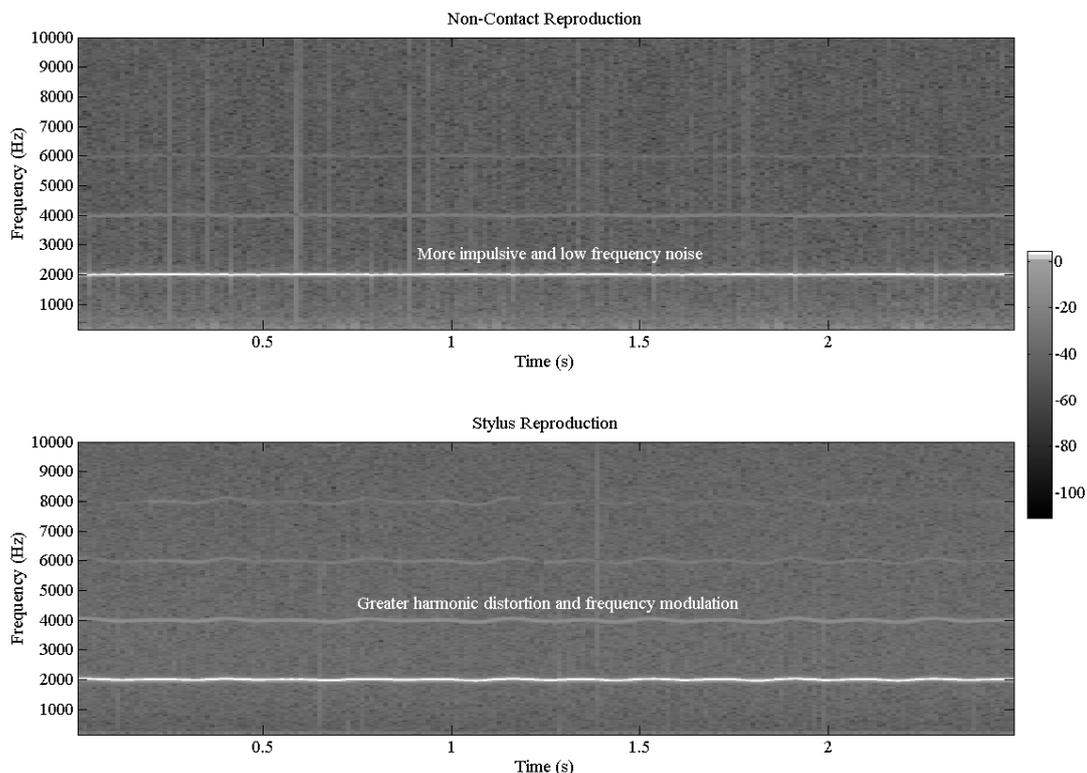


FIG. 14. : Spectrogram comparisons for non-contact and stylus reproduction of 2 kHz tone. Harmonic distortion is greater with stylus reproduction. The non-contact method has less harmonic distortion but more low frequency (below 500 Hz) and impulsive noise. In addition, frequency modulation is less apparent for the non-contact method.

Frequency Modulation

The degree of frequency modulation was also compared by locating the position of the peak frequency in the DFT for windowed segments of recovered tones. The standard deviation in measured frequency from peaks in the DFT was used as an indication of frequency modulation. Figure 15 shows

frequency modulation for all tones reproduced from the test cylinder by optical and stylus methods. At 5 kHz, the frequency modulation with stylus reproduction is approximately ± 40 Hz, and for optical reproduction is less than ± 10 Hz. The increased frequency modulation with stylus reproduction is largely due to the mounting of the cylinder on the mandrel. If mounted incorrectly, the surface speed of the cylinder will vary, meaning that for a constant playback sample rate a modulation in frequency occurs.

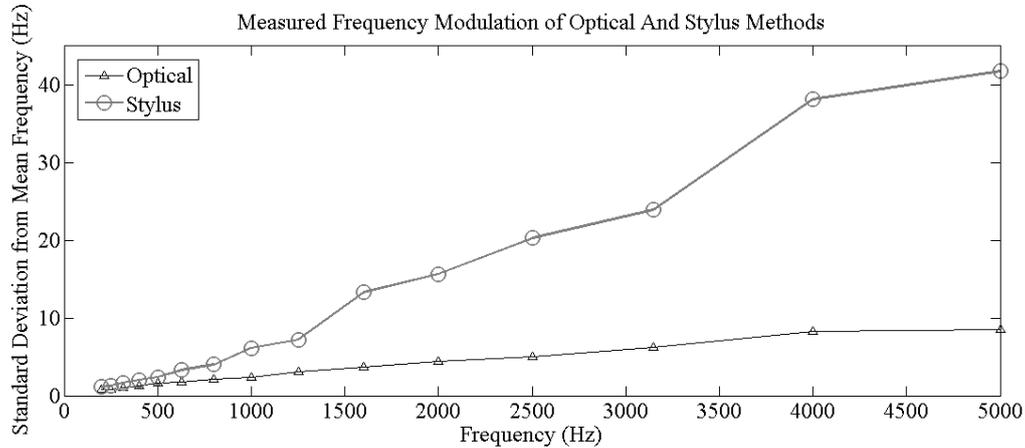


FIG. 15: This figure compares the frequency modulation of tones recovered from the test cylinder via stylus and optical methods.

5. CONCLUSIONS

The non-contact measurement system which has been developed allows for accurate sound reproduction from cylinder recordings. The advantage of mapping the full surface topology (as opposed to realtime optical playback) is that numerous estimates for the audio signal can be derived from a single data set in post-processing.

A stylus trajectory estimation method based on the shift in relative phase between consecutive linescans has been described. Such a ‘global’ method of trajectory estimation has advantages in terms of reduced processing time, when compared with ‘local’ tracking of individual grooves, and has been found to be robust for the test cylinder recording. In cases of cylinders which may be damaged or broken, local tracking of groove features may still be required.

A cylinder encoded with sinusoids provides a basis for assessing the quality of sound reproduction for signals of known type. Two different methods of estimating the radial displacement track from a discrete surface have been shown to produce variable audio signal quality, in terms of signal-to-noise ratio (SNR) and total harmonic distortion (THD). The use of appropriately designed SG filters (Method B) to produce an estimate for the groove depth gives a better audio signal estimate signal in terms of SNR, and THD, compared with un-weighted mean averaging across the groove matrix (Method A).

Care must be taken when obtaining an average for the groove depth. For example, the use of groove ridge data was shown to introduce temporal distortion, (echo). Averaging across the full groove cross-section to obtain an estimate for $z_a(t)$ is therefore not advisable as the inclusion of ridge data from adjacent grooves can result in echo and increased harmonic distortion.

Audio signal estimation is not restricted to the two methods described in this paper. To improve further the quality of the recovered signal, other rule- or model-based navigation methods may be used to avoid the introduction of outliers, (caused by debris etc.).

The audio signal recovered via conventional stylus reproduction exhibited more frequency modulation and harmonic distortion compared with the optical method. Although this is a feature of this particular transfer, (and not necessarily inherent for all stylus playback), it highlights some advantages of non-contact reproduction. For example, dynamic effects such as tracing distortion do not affect optical reproduction. Additionally, the non-contact method appears to be less affected by speed irregularities, caused with stylus reproduction if the cylinder is poorly mounted, and rotated at high speed.

6. REFERENCES

1. T. Iwai, T. Asakura, T. Ifukube, T. Kawashima. (1986). "Reproduction of sound from old wax phonograph cylinders using the laser-beam reflection method." *Appl. Opt* **25**, 597-604.
2. T. Ifukube, T. Kawashima, T. Asakura, et al. (1989). "New methods of sound reproduction from old wax phonograph cylinders." *JASA* **85**(4): 1759-1766.
3. T. Asakura, and J. Uozumi (1999). "Optical methods for Reproducing Sounds from Old Phonograph Records." *International Trends in Optics and Photonics, ICO IV*: 65-81.
4. S. M. Shanoylo, I. V. Kosyak, V. Petrov, A. A. Kryuchin. (2001). "Reading and processing of audio information reproduced from Edison phonograph cylinders by method of laser interferometry". *Proc. SPIE Vol. 4402*, p. 194-201, *Laser Techniques and Systems in Art Conservation*.
5. S. S. Cavaglieri, O. Johnsen, F. Bapst. (2001). "Optical Retrieval and Storage of Analog Sound Recordings." In *AES 20th International Conference, Archiving, Restoration, and New Methods of Recording*.
6. S. Stotzer, O. Johnsen, F. Bapst, C. Milan, R. Ingold. (2007). "Phonographic sound extraction using photography and signal processing." *Digital Signal Processing* **17**(2): 433-450.
7. S. Stotzer, O. Johnsen, F. Bapst, C. Sudan. (2004). "Phonographic Sound Extraction Using Image and Signal Processing." *Acoustics, Speech, and Signal Processing*, 2004. in *Proceedings ICASSP '04*, **4**(5): 289-292.
8. V. Fadeyev, and C. Haber. (2003). "Reconstruction of Mechanically Recorded Sound by Image Processing." *J. Audio Eng. Soc* **51**(12): 1172-1185.
9. V. Fadeyev, C. Haber, C. Maul, J. W. McBride, M. Golden (2005). "Reconstruction of Recorded Sound from an Edison Cylinder using Three-Dimensional Non-Contact Optical Surface Metrology." *J. Audio Eng. Soc.* **53**(6): 485-508.
10. H. Meulengracht-Madsen, (1976). "On the Transcription of Old Phonograph Wax Records" *J. Audio Eng. Soc* **24**(1): 27-32
11. IASA TC-04 (2004). "Guidelines on the production and preservation of digital audio objects".
12. S. Stotzer. (2006). "Phonographic Record Sound Extraction by Image Processing", PhD Thesis.
13. Y.-C Jenq., (1996). "Discrete-time method for signal-to-noise power ratio measurement." *Instrumentation and Measurement, IEEE Transactions on* **45**(2): 431-434.
14. A. J. Nascè, J.W. McBride, M. Hill, P. J. Boltryk. (2007). "Signal Processing Methods for the Recovery of Audio from Early Acoustic Cylinder Recordings, measured via non-contact optical sensor", *AES 31st International Conference : New Directions in High Resolution Audio*.
15. Poppy Records, <http://www.poppyrecords.co.uk/>
16. L. D. Willard (1941). "An Approximate Theory of Tracing Distortion in Sound Reproduction from Phonograph Records." *The Journal of the Acoustical Society of America* **12**(3): 462.
17. A. Savitzky, and M. J. E. Golay (1964). "Smoothing and Differentiation of Data by Simplified Least Squares Procedures." *Analytical Chemistry* **36**(8): 1629-1639.
18. P. J. Boltryk, M. Hill, J. W. McBride, A. J. Nascè (2008). "A comparison of precision optical displacement sensors for the 3D measurement of complex surface profiles." *Sensors and Actuators A: Physical* **142**(1): 2-11.
19. P. J. Boltryk, J. W. McBride, M. Hill, A.J. Nascè, Z. Zhao, and C. Maul (2008). "Non-contact surface metrology for preservation and sound recovery from mechanical sound recordings", *due for Publication in J. Audio Eng. Soc.*