LiU-ITN-TEK-A-16/044--SE

A Laser Triangulation Approach for Optical Audio Reconstruction of Phonograph Records

Kristofer Janukiewicz

2016-09-22

Department of Science and Technology
Linköping University
SE-601 74 Norrköping, Sweden

Institutionen för teknik och naturvetenskap
Linköpings universitet
601 74 Norrköping

LiU-ITN-TEK-A-16/044--SE

A Laser Triangulation Approach for Optical Audio Reconstruction of Phonograph Records

Examensarbete utfört i Medieteknik
vid Tekniska högskolan vid
Linköpings universitet

Kristofer Janukiewicz

Handledare Joel Kronander
Examinator Jonas Unger

Norrköping 2016-09-22

Master of Science Thesis in Media Technology
Department of Science and Technology, Linköping University, 2016

A Laser Triangulation Approach for Optical Audio Reconstruction of Phonograph Records

Kristofer Janukiewicz

LINKÖPING UNIVERSITY

Master of Science Thesis in Media Technology

A Laser Triangulation Approach for Optical Audio Reconstruction of Phonograph Records

Kristofer Janukiewicz

LiTH-ITN-EX--YY/NNNN--SE

Supervisor: **Joel Kronander**
ITN, Linköpings Universitet
Mattias Johannesson
SICK IVP

Examiner: **Jonas Unger**
ITN, Linköpings universitet

Media and Information Technology
Department of Science and Technology
Linköping University
SE-601 74 Norrköping, Sweden

Abstract

This thesis introduces a method for contact-free optical audio reconstruction of phonograph records using laser triangulation. The reconstruction is done using a 3D-profile camera and by scanning the surface of record in a single circumference. The depth-map created by the camera is then used to decode the audio information stored in the record. To evaluate the quality of the decoded audio information, it is tested against digital copies of the same record in order to analyze the correlation between it and the extracted sound.

The result of this thesis presents a decoded and high correlated audio which is recognizable to an original digital copy. The method is well suited for fast real-time or faster than real-time decoding implementations.

Acknowledgments

My thanks goes to all the people involved in this project. Not the least to my supervisor at SICK IVP, Mattias Johanesson, who have been a great help to construct the setups, exchange ideas and introduce me to the company. But also forming this thesis idea and choosing me to do it.

A lot of thanks goes to my supervisor, Joel Kronander, at Linköping University who has guided and gave me a lot of introductory ideas on were to start and handle this thesis but also in finalizing it. A big thanks to my examiner, Jonas Unger, who approved of this thesis.

A loving thanks to Tiffani Höglind, who have put up with me during this thesis project.

A thanks to my thesis-colleagues at SICK IVP who have brought great joy during this thesis work: Jens Edhammer, Mikael Zackrisson and Richard Bondemark.

And of course, a huge thanks to everyone at SICK IVP for letting me use their equipment, experience and technology for this thesis.

Contents

1 Introduction

Libraries across the world have large collections of old music records that need to be digitized for preservation. For example the National Library of Sweden has, approximately, 146 000 such records [Vinnova, 2014]. Reading and digitizing these with conventional methods takes prohibitively long time and furthermore risks destroying the old artifacts, and therefore; more efficient methods are requested.

SICK[1] develops high-performance 3D profiling products, that uses a laser triangulation technique, which has the potential to read and decode audio information on records.

This thesis deals with contact-free reading of music on 78 rpm phonograph records using a 3D profiling camera provided by SICK IVP[2]; furthermore investigates the challenges laser triangulation will have on these kind of records and possible ways to overcome them.

1.1 Optical Audio Reconstruction of Phonograph Records

Contact-free reading of phonograph records is a technology that has existed in the past few decades. The topic is well researched and analyzed, however there is still room for new approaches and optic technologies to be improved/invented.

Achieving a contact-free reading is advantageous; as it does not degrade or ruin the information stored in a record and it also has the potential to do it more rapidly than a traditional stylus. Meaning; digitizing a single record quicker than the total audio time recorded into a disc. For a huge library, this is crucial as each

[1] `https://www.sick.com/us/en`
[2] `https://www.sick.com/se/sv/w/sick/`

78 rpm disc can store up to a few minutes of audio information on each side and $33\frac{1}{3}$ almost an half hour of music on each side. Some older records, in the libraries, are fragile or degraded to a point where the traditional stylus is not an option, as it could further damage the stored information.

In this thesis, the above topics are analyzed using a 3D-profiling approach, in contrast to the more common 2D approach. Creating a depth map of each scanned record directly with the camera opens up new possibilities of decoding the information stored in a phonograph record.

1.2 Related Work

Creating an optical audio reconstruction of records is a well researched topic where different methods and approaches have been studied and proposed for a few decades.

In a historical perspective, the first contact-free approach for reading records was presented by William K. Heine [Heine, 1976]. The proposed procedure is to use a laser beam to create an interference/diffraction pattern of the records groove walls and the information of magnitude and displacement to decode the track. His research was very beneficial to achieve an interest in optical decoding of phonograph records and William K. Heine also concluded that a higher fidelity was reached using a laser instead of a regular stylus on the record, as it was less affected by warps and physical irregularities. Although William K. Heines research was a breakthrough in the area of optical audio reconstruction, his prototype could only read and playback records in real-time as the laser beam, in his system, followed a groove as an ordinary stylus in a turntable would. The proposed method was intended for commercial usage and not for a decoding at an industrial case, where it is beneficial to scan a larger area of the record at the same time and decode a record faster than its playback time. His research also came at an unfavorable historical timing as the compact disc was just around the corner and made studies on laser turntables of gramophone discs less interesting for researchers.

Although the compact disc was introduced in 1982, there still existed some researchers and companies interested in a commercialized optical audio reconstruction system for records. The ELP Corporation [ELP, 1997] introduced a commercial laser turntable which can read black vinyl LP-records microgrooves with the help of 5 lasers analogically; keeping track of a groove like a stylus would, just as William K. Heine proposed. The technology was patented by Robert E. Stoddard [Stoddard, 1989] and later together with Robert N. Stark [Stoddard and Stark, 1990]. The laser turntable was a revolutionary invention as anyone could potentially read their vinyl records optically. Although it was a great invention; it was not flawless, the system is very expensive, it is sensitive to dust and abnormalities while playing a record and the system is only recommended to use on black vinyl records. Just like in the case of [Heine, 1976] it can only play records in real-time as it was only intended for commercial use. In the case of this thesis, we have large collections of records and some in bad condition that needs to be scanned and decoded in quick succession for digital preservation, which means that the system presented by [ELP, 1997] and [Heine, 1976] is not feasible for the case presented in this thesis.

A few years after the laser turntable was introduced, some researchers began experimenting on new approaches for optical audio reconstruction of records. One of them is Ofer Springer who believed that image processing on the surface of a record could be used for an optical audio reconstruction. In [Springer, 2002] the Digital Needle project was presented by Springer. He scanned a 2D-image of a record and implemented a decoder to track the groove spiral. Unfortunately, the results were barely recognizable due to the high amount of noise present in the decoded audio. With his proposed methods it was proven that it was possible to extract the audio information from a record using 2D image processing of the surface and by scanning an area of the surface and reconstructing the audio information from a scan, a faster than real-time decoding of a record can be achievable as the system is no longer using a laser stylus that follows the grooves to reconstruct the audio.

The first 3D reconstruction approach of a record surface was presented in [Fadeyev and Haber, 2002]. It was done with a 3D profile scanning through the use of a laser confocal microscope. In comparison to a 2D approach, presented in [Springer, 2002], a 3D reconstruction had the benefits of analyzing the entire surface structure of a record. The techniques used in [Fadeyev and Haber, 2002] were successful in decoding a record using a 3D profile scanning but was deemed to be slow by the author, as scanning and reconstructing a full side of a single record could take a whole day and thus still not suitable for decoding and digitally preserve a lot of records in quick succession.

As both 2D and 3D optical audio reconstruction from records were now possible to achieve; research is still done in both areas to find the best, fastest and suitable approach depending on the condition of the records and the goal of the reconstruction process. In [Stotzer, 2006] a new contact-free 2D-reconstruction was introduced under the name VisualAudio concept. A photo of the record was saved into a large film, the film was at least as big as the record. On the film, they used a line scan camera to detect groove wall edges. The VisualAudio concept introduced stereo readings from the walls of e.g. $33\frac{1}{3}$rpm records using a 2D-reconstruction technique but also took on the challenge of reconstructing records in really bad conditions. As the method presented in [Stotzer, 2006] proved that a 3D-reconstruction of the records surface is not needed to decode stereo audio (as stereo grooves modulate in both vertical and horizontal axis), noise was still present in the decoded audio.

In [Tian and Barron, 2006] and [Tian and Barron, 2011] another contact-less 3D-reconstruction of the audio information in records was introduced. By using Computer Vision techniques, they were able to reconstruct a depth map from 2D-microscopic photos of the grooves in $33\frac{1}{3}$rpm records. They used the depth maps to decode the sound from the walls of the grooves with surface orientation. This project is still under development and the authors are focusing on improving the algorithms to achieve a real-time implementation as it is still fairly slow due to the computer vision techniques used on the 2D photos. In comparison to

[Tian and Barron, 2006] and [Tian and Barron, 2011], this thesis is fairly similar but that the 3D-reconstruction is done by using a fast and industrially developed laser triangulation system. This means that it should be possible to achieve a faster and similar result, as the 3D-reconstruction is done by the laser triangulation system; but since a laser triangulation technique has not been studied and published before to achieve an optical audio reconstruction of records, limitations of such a system is not known. It can although be assumed that limitations of using a laser triangulation system is rooted to the achievable laser width, the scanning devices resolution & scanning frequency but also their sensitivity to irregularities found in a record and also the size of the information in the record.

1.3 Project's Contribution

As mentioned in section 1.1, the approach of this thesis is to reconstruct audio from a record using a 3D-profile camera. The camera uses a Sheet of light triangulation technique [IVP, 2015, p.7] (also known as: Laser Plane Triangulation or Laser Stripe Triangulation) to create a 3D-profile of the scanned record.

The advantage of using a 3D depth map to reconstruct a record, compared to 2D, is the complete analysis of the whole records surface. Knowing the height differences in the grooves makes it possible to detect the walls in the grooves and also their respective movement in the discs rotation for stereo sound; but also height irregularities in degraded discs and the chosen turntables vertical impact on the system.

The contribution of this thesis is to give an insight of an experimental approach using a 3D-profile camera to reconstruct audio from 78 rpm discs and also how this system can be further optimized and modified for $33\frac{1}{3}$ rpm and 45 rpm records.

With the proposed setup and methods we show that a fast decoding of a record can be achieved but also highlighting its limitations and strengths. We believe our approach could be used for an industrial scale decoding system in order to preserve and digitize many records that are in bad conditions in libraries and collections around the world.

1.4 Proposed Method Pipeline

A method pipeline is proposed in order to achieve an optical audio reconstruction using a laser triangulation approach on phonograph records. Each method of the pipeline is shortly summarized in this heading, to achieve an understanding of their individual part of the presented audio reconstruction system. More in depth explanation of each method can be found in section 4. An illustration of the method pipeline can be seen in figure 1.1.

1. **Pre-data process** - Before any data of the records surface is achieved, necessary measurements needs to be done on the system.

(a) **Record measurements** - Using a laser triangulation approach, it is necessary to know the length of an approximated sample in the record in order to achieve a high resolution scan. This can be calculated by knowing the frequency range of the record.

(b) **Laser & camera measurements** - Measuring the hardware limits of the laser triangulation system and comparing it to the record measurements is necessary to confirm a possible scan of the record.

(c) **System / Hardware setup** - To perform a scan, the hardware system needs to be constructed; this also opens up the possibility of testing different setups of the camera-laser system.

(d) **Record scanning** - A scan of the records surface is initiated. To know the different values of the system, a general scanning formula can be assumed.

2. **Post-data process** - After the scanning is completed we end up with an intensity- and range-data map of the scanned record.

   (a) **Pre-processing of scanned range data** - The data is assumed to contain noise of different kinds and needs to processed in order to achieve a robust decoding process. In this thesis we present a method of reducing noise using a 2D median filter supplemented with an 2D anisotropic diffusion.

   (b) **Groove segmentation** - Each data will contain a set of grooves, that is; a circumference of the groove in the record. Each groove is segmented using a fast path segmentation algorithm.

3. **Processes performed on each groove** - When each circumference of the groove is segmented, the decoding of the audio information is initiated. Each groove contains a set of samples, calculated in the pre-data process, and these samples contain the audio information of the record.

   (a) **Finding local extremas** - Each sample is assumed to contain two local maximums and a local minimum as each sample has the form of a valley. The local extrema positions are extracted in each sample of each groove. The extremas are then fitted using a polynomial fitting function to achieve finer extrema positions, this is performed due to low resolution of each sample.

   (b) **Simulating stylus wall-positions** - From the finer extrema positions, an intersection line is created to simulate a stylus touching the walls in each sample. The wall positions are considered to be the left and right channel of the signal.

   (c) **Decoding wall-positions to sound signal** - A derivation is then used on the separate left and right wall positions to center the signal around 0, this achieves a hearable sound of the scanned record.

4. **Signal processing** - The last step of the system is to achieve a hearable and continuous sound signal of the scanned record. The data in this step is 1D-signals in separated set of groove circumferences. Each 1D signal also contain a left and right audio channel due to the left and right wall positions.

   (a) **Sound stitching** - Each grooves audio signal is stitched together using cross-correlation to find the best lag position between the ending of the first groove into the beginning of the next groove. The signals are then stitched together in the found lag-position using a cross-fade. This is performed until all signals are stitched into one long single signal.

   (b) **Mono / Stereo** - In this step, we have two 1D signals containing the same information in two channels. The signal can either be playbacked and assumed done in this step if a stereo sound is the goal. To achieve mono, a simple mean of both the channels can be done.

   (c) **Post-processing of decoded sound** - This step is marked with a dotted line in the figure 1.1 as it considered a subjective quality enhancement step of the resulting audio signal. Different filters can be used to achieve a better sound quality. In this thesis we use a simple moving average filter to reduce high frequency noise and a maximum suppression filter to remove click-defects.

   (d) **Evaluation** - An evaluation of the decoded sound is created to analytically measure its quality. The evaluation is done by creating two ground truth cases and comparing them with the decoded sound using a shape envelopment and Pearson's product-moment correlation.

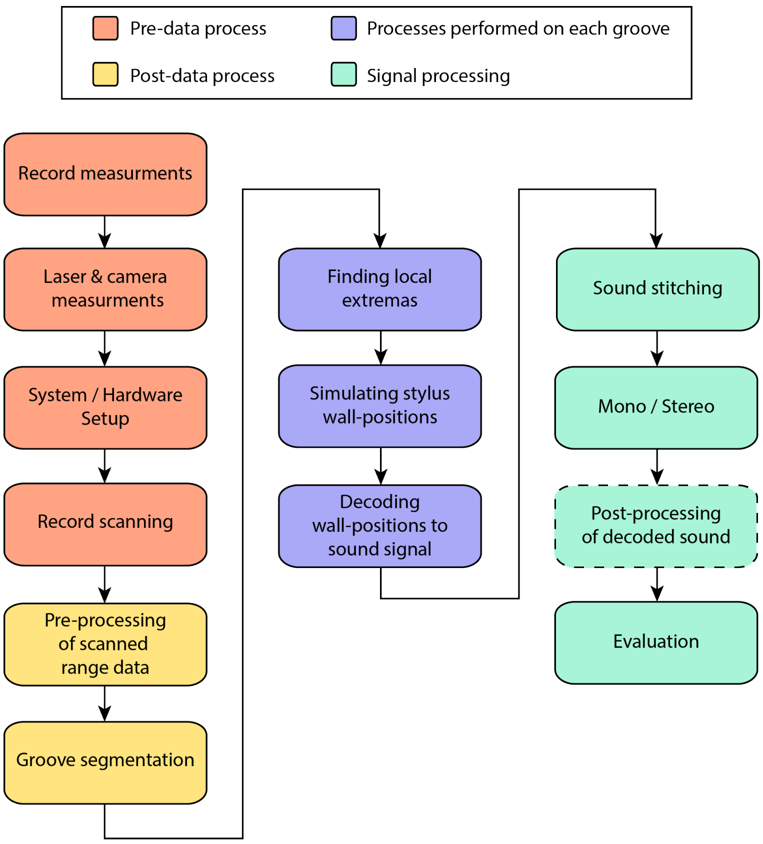


Figure 1.1: *The method pipeline used in this thesis for an optical audio reconstruction of phonograph records.*

2

Background

This chapter introduces and gives a brief background to the analogue audio recording formats. This chapter also presents a more in depth groove explanation.

2.1 A Brief History of Audio Recording

The history of audio recording began as early as 1857 by the inventor of the Phonoautograph, Léon Scott. The Phonoautograph was able to store the sound waves by making a small needle, stuck to a big horn, scratch the surface of a cylinder. Speaking or making sound through the horn, made the needle vibrate while the cylinder was rotated. The Phonoautograph could not reproduce the sound it recorded. It took two decades before it was possible to reproduce recorded sound as in 1877 Thomas Edison invented the Phonograph which could record and reproduce the sound from a tinfoil cylinder. Although a great technical innovation, cylinders were not a practical medium. Therefore, In 1887, Emile Berliner invented the Gramophone which uses discs instead of cylinders to record and reproduce sound. Although the Gramophone and the disc records had a though start due to the medium competing against the famous Edison's cylinders, it soon took over the market as the disc were cheaper to produce, tougher and were easier to store.

The discs were now, by the early 1900s, the standard audio storing medium; but with many new competitors appearing in the record market with their own standards and Gramophones, a standard recording format had to be achieved between the different companies and countries. In 1925 the standard rotations speed was settled around 78 rpm, also with the introduction of electrically powered turntables. Before 1925 discs could have varying speeds (from 60-130 rpm) and sizes. Why the 78 rpm was chosen as a standard is never explained [Read, 1952]. The 78 rpm disc has the potential to store at least 2 minutes of music into each side.

A few years later a new recording medium was invented; the magnetic tape, invented in 1928 by Fritz Pfleumer. Even if the magnetic tape was more practical than the ordinary disc, the discs stayed quite popular with the consumers and due to its popularity a new disc-format was introduced in 1931, the long play disc or its shorter name: LP-disc. The LP-disc rotates a slower pace, at a 33 1/3 rpm and has smaller grooves which in turn also needed a smaller stylus. The newly invented disc also has the potential to store at least 10 minutes of music into each side. They were also cheaper to produce as they were coated with a new plastic texture; vinyl. But because of the lack of affordable playback equipment and the Great Depression. LP did not make a breakthrough until 1948; when it was released by the Columbia Record Company as the microgroove.

Only a year after, in 1949, the Radio Corporation of America released a new format; the 45 rpm single or EP-discs as they were called. It used the same technique as the regular 33 1/3 rpm discs and could be played on the same machines, but of course with a faster rotation speed. The 45 rpm never became as popular as the 33 1/3 rpm discs as its intended use was for more public machines e.g. Jukeboxes, as the discs usually only stored singles (one track) on each side.

From the invention of the first Phonoautograph, it took almost a century before digital sound was commercially introduced. In 1982 the compact disc (CD) was invented. Storing and recording sound had never been easier, the discs were tougher and smaller than the LP-discs, they could store a whole album and sometimes even more content digitally and reproduced an almost perfect sound. CDs were also cheap and easy to manufacture. With this, the era of turntables and LP-discs came to an end as the analog medium quickly dropped in popularity. LP-discs are still produced and sold to this day, but in much smaller quantities as it is still believed by enthusiast and audiophiles that the analog medium is superior in reproducing recorded audio.

2.2 Disc Formats

In 1963, the Record Industry Association of America introduced record standard formats for all produced records [RIA, 1963]. The records were, from this date forward printed in the following diameter sizes:

- 7" or more exactly (6 7/8" + 1/32")
- 10" or more exactly (9 7/8" + 1/32")
- 12" or more exactly (11 7/8" + 1/32")

And recorded in the following rotation speeds:

- 78 rpm
- $33\frac{1}{3}$ rpm
- 45 rpm

2.3 Groove

The audio information in a record is stored in a spiral, valley-formed groove; going from the edge, to the center of the circular disc, figure 2.1. A small stylus traverses in this groove (figure 2.2) at high speed creating a low but audible sound due to the small modulations present in the groove. The sound is, by the construction of a phonograph player, amplified enough to be heard loud and clear as the stored audio information. To get a better understanding of the groove structure; A 3D-illustration of a typical groove segment can be seen in figure 2.3.

78 rpm discs were made with only horizontal modulation. This single dimension modulation limits the audio storage to mono; single channel audio (see figure 2.4) and the information stored in both groove walls are identical. N.B. The y-axis presented in the figures represent the θ rotation position of the spiral groove shape.

Introducing one more dimension for the stylus to move increases the number of channels in a record to two; stereo. In stereo records, both sides of a groove can consist of different audio information. The modulation of the groove is in both horizontal and vertical and does not need to be identical on both sides, compared to mono, see figure 2.5. A 3D-reconstruction of stereo discs is therefore favorable as the modulation in depth (Z-axis) can be analyzed, in contrast to 2D-reconstruction processes.

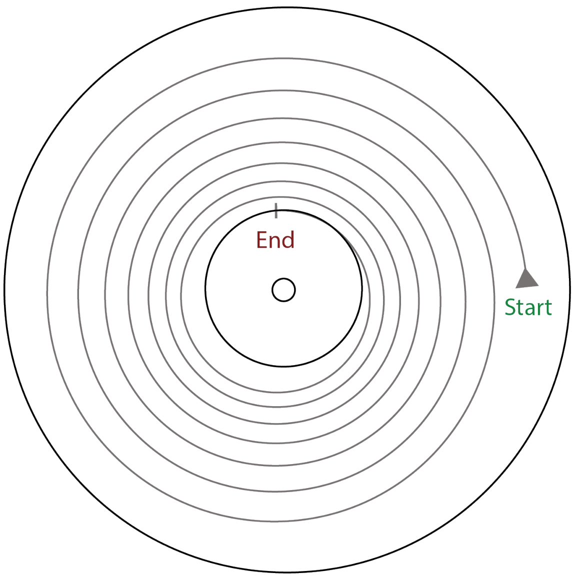


Figure 2.1: *A 2D-illustration in the top view of a gramophone disc. N.B. The groove in this illustration are greatly exaggerated for a better visibility.*

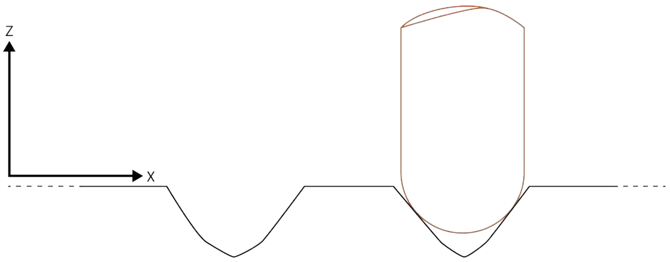


Figure 2.2: *A 2D-slice illustration of the surface in a record and a needle traversing in the rightmost groove valley.*

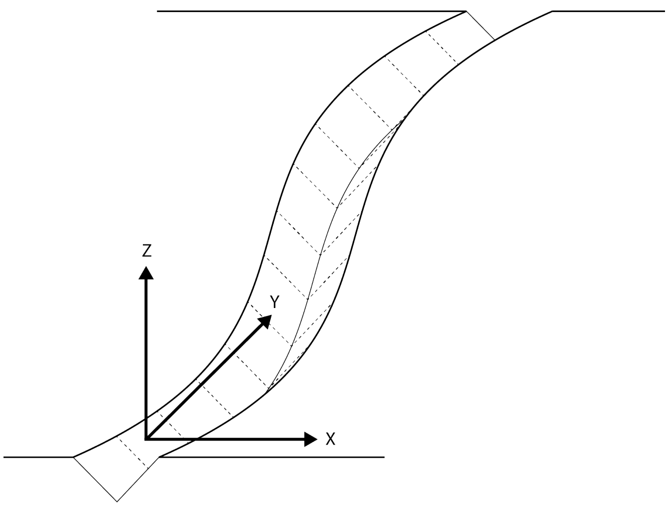
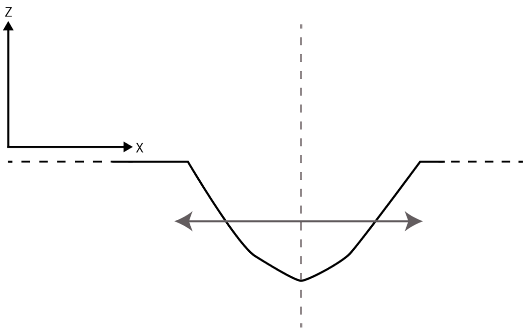
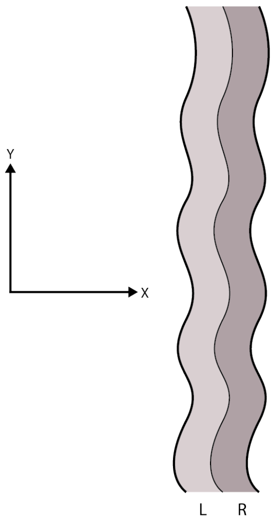


Figure 2.3: *A 3D-segment illustration of a groove in the surface of a record.*

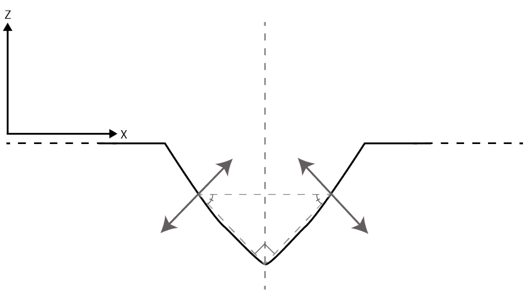


(a) *The horizontal modulation in XZ.*

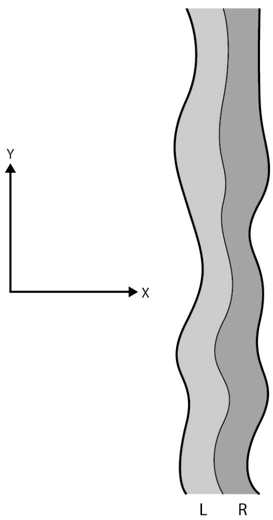


(b) *The horizontal modulation in XY.*

Figure 2.4: *A mono records horizontal modulation in the groove illustrated in XZ- and XY-domain.*



(a) *The horizontal and vertical modulation in XZ.*



(b) *The horizontal and vertical modulation in XY.*

Figure 2.5: *A stereo records horizontal and vertical modulation in the groove illustrated in XZ- and XY-domain.*

2.4 Error in analog decoding

The audio information stored in the grooves of a disc is purely analogue and does not reproduce the stored audio flawlessly. Some general analog decoding errors that can affect a turntable will also affect a scan with a camera; the observed ones are as following:

- **Off-axis horizontal modulation** created due to that the hole in the record is not correctly centered, see figure 2.7 for an illustration. The reason that it can be assumed that the hole is not correctly centered is because when the records were printed, the hole was made manually by a factory worker. The off-axis horizontal modulation for the record presented in table 3.3 in millimeters can be seen in figure 2.6. This is described as the wow defect in [Fadeyev and Haber, 2002, p.6] and [Stotzer, 2006, p.74]. A solution to this is presented in [Fadeyev and Haber, 2002, p.12] and [Stotzer, 2006, p.110], correcting the position of grooves by knowing the radial displacement from the center of the record. In this thesis; this horizontal modulation also creates an unwanted vertical modulation explained below.

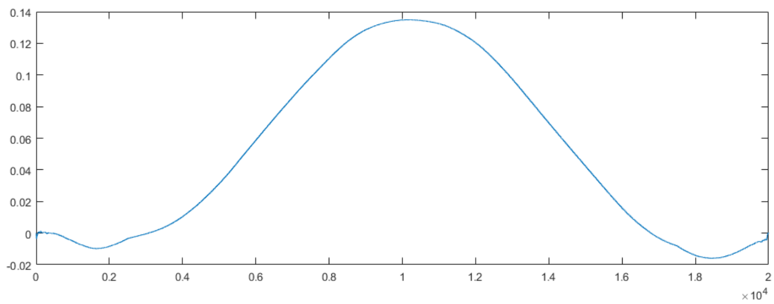
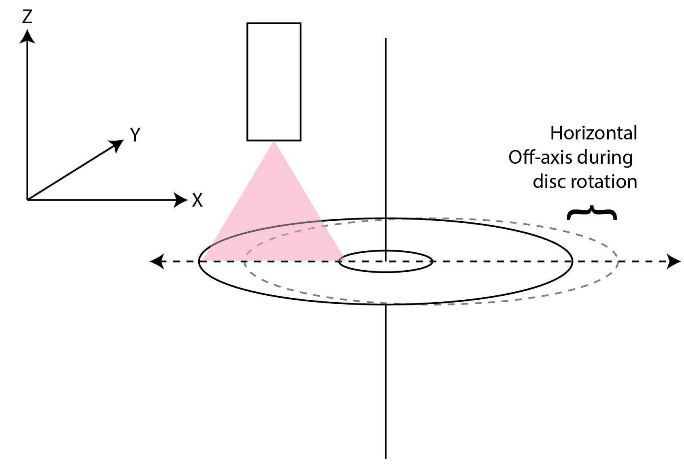
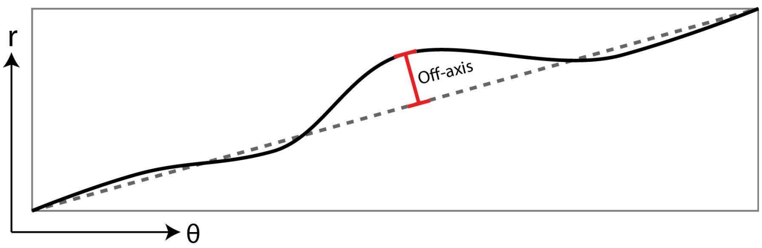


Figure 2.6: *By subtracting the horizontal positions of the grooves minimum position with the linear circumference position (the first and last horizontal position of a circumference) it is possible to measure the horizontal offset in millimeters produced on a scanned dataset. We can see that during every circumference; the horizontal offset maximizes at around 0.13 millimeters. This modulation defect is also hearable on each circumference using a regular turntable and stylus as a repeating small hissing sound.*

- **Off-axis vertical and horizontal modulation** created due to that the record is uneven, that the turntable is uneven or that the camera is not positioned exactly straight to the surface of the record, see figure 2.8 for an illustration. This can create sampling resolution irregularities as the groove is going back and forth from the camera perspective when its rotating around its own axis due to the **Off-axis horizontal modulation** explained above. A general approach to fix this modulation defect is to do as [Stotzer, 2006] and create a very stable turntable which effectively removes vertical modulation. A regular turntable and stylus is not affected by this defect as the



(a) When the disc rotates in the turntable it creates a horizontal off-axis modulation due to that the hole in the disc is not correctly centered.



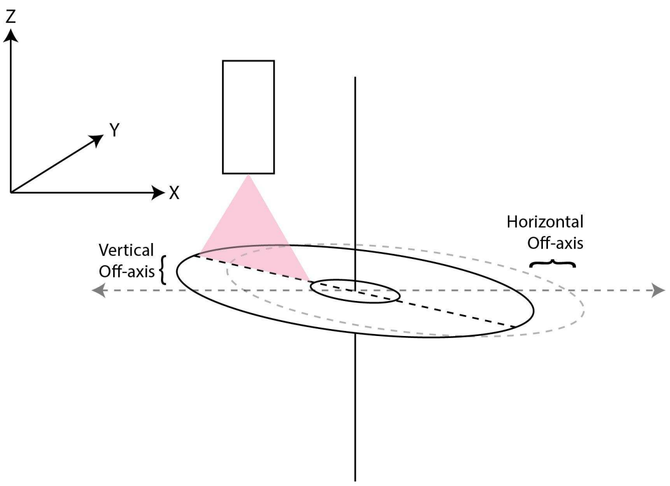
(b) The off-axis modulation creates off-axis defected grooves in the dataset. The dotted line represents a grooves position in the dataset if there were no horizontal modulation during rotation.

Figure 2.7: *An off-axis horizontal modulation created due to that the hole in the record is not correctly centered. For this thesis, a more correct off-axis modulation is illustrated in figure 2.8.*

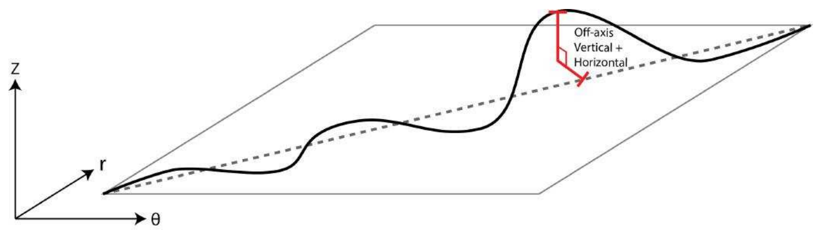
stylus is traveling through the grooves with the help of gravity, pushing the stylus down against the record even if it is vertically irregular.

- **Scratches and dirt** can affect the laser triangulation system proposed in this thesis. Physical irregularities can create unwanted light reflections and speckles in a scanned data. It is therefore important to clean the record as much as possible before a scan to remove as much dust and dirt as possible. Scratches, which should not appear as frequent as dust, has to be processed through a post-processing step. A regular stylus would simply push away the dust and dirt from the grooves but a scratch can make the stylus jump out of the groove.

- **Velocity / Sample rate artifact** is a regular analog error. As all records are



(a) In addition to the off-axis horizontal modulation due to the hole in the disc is not correctly centered. An off-axis vertical modulation is observed when the disc rotates in the turntable.



(b) The vertical off-axis modulation creates another dimension in the grooves position in the dataset. A perfect scan would result in the dotted line where there is no vertical or horizontal modulation in the scanned dataset.

Figure 2.8: *An off-axis horizontal and vertical modulation created due to that the hole in the record is not correctly centered followed by that the record, turntable or camera is uneven; making it look like the disc 'wobbles' in the z-dimension.*

recorded through the use of analog techniques there is a risk that the record was recorded slightly faster or slight slower than the standard format. This is not a hearable defect but can potentially lead to some sample rate artifact through a digital decoding; that is, an interval of a calculated sample rate needs to be evaluated to find the best sample rate for the record.

3

Hardware Setup

During this thesis three different lab setups were tested in order to scan a disc with a 3D profiling camera. In the following headings a more in depth information is provided of the tested setups and the hardware used.

3.1 Camera

The camera used in this thesis is the SICK Ranger E50[1], see fig.3.1, which uses laser triangulation (described in section 3.1.1) to create a 3D profile of scanned objects.

Figure 3.1: *A picture of SICK's Ranger camera.*

The Ranger camera has the potential to scan with a frequency of 30 KHz with binary algorithms and up to 2 KHz with Hi3D (3D-profile scans) which uses grayscale inputs [IVP, 2015]. Modifying the scan frequency of the camera is

[1] `https://www.sick.com/de/en/product-portfolio/vision/3d-vision/ranger/c/g138563`

equivalent to changing the rotational speed of the turntable; as in adjusting the desired frame rate of the record which is further discussed in section 4.6.2.

The cameras parameters are further modified using the software Ranger Studio (described in section 3.1.3).

3.1.1 Sheet of Light Triangulation

The laser triangulation technique used in the Ranger E camera is the Sheet of Light Triangulation. Which means that the scanned object is illuminated with a laser line in a horizontal direction to the perspective of the camera. The Ranger E camera then determines the height in each point of the object, illuminated by the laser line, by locating the vertical location of its cross section [IVP, 2015, p.20]. By then moving the object (a rotation of the record, in this thesis) a sequence of groove depth profiles are generated over a wide area on the surface of the record. The depth profile together with the XY-axis creates a 3D-reconstruction of the records surface. The Sheet of Light is illustrated in figure 3.2.

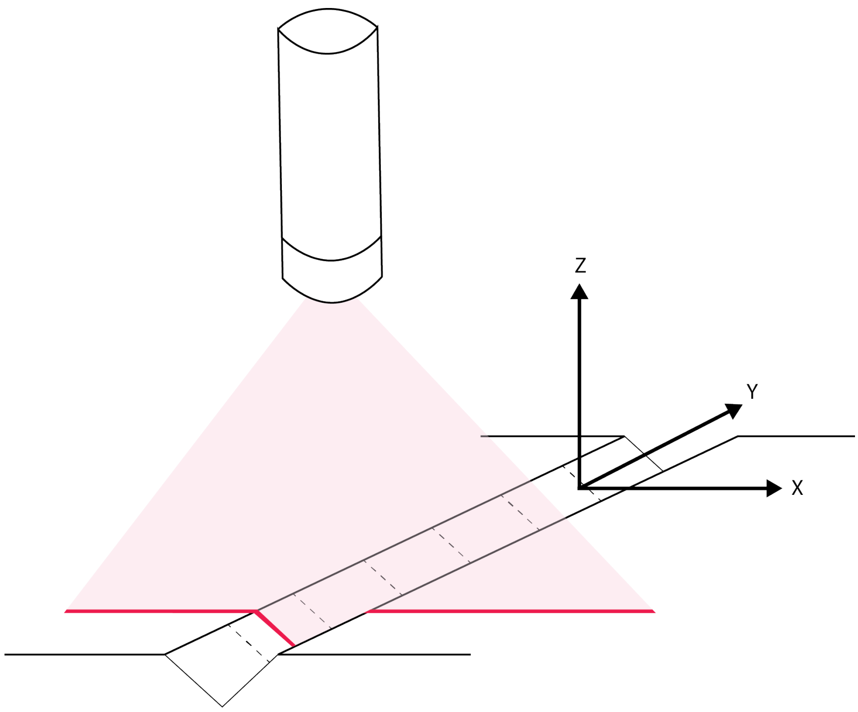


Figure 3.2: A Sheet of Light illustration performed on the surface of a record.

3.1.2 Camera Lens

The camera lens used in the setup for a 78 rpm disc, can be seen in the illustration 3.3 and photo 3.4. The lens, mounted onto the Ranger cameras sensor, is a modifiable kit (parts) manufactured by Schneider Optics[2] and provided by SICK IVP. The lens was not suitable for $33\frac{1}{3}$ rpm LP-discs as the grooves were deemed too small. Further testing on LP-discs requires a higher grade of magnification.

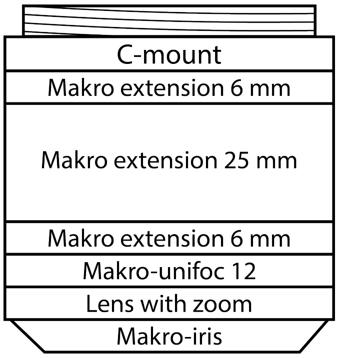


Figure 3.3: *An illustration of the camera lens used for the 78 rpm discs. All parts are constructed by Schneider Optics.*

Figure 3.4: *A photo of the Schneider Optics camera lens kit used for the 78 rpm discs.*

3.1.3 Ranger Studio

Ranger Studio is a configuration tool, provided by SICK IVP, where it is possible to adjust camera parameters, record and visualize camera data. Recorded data

[2] `https://schneideroptics.com/`

can be saved in a proprietary binary format (.dat) with formatting information in XML (.xml). For application development, in Ranger Studio, there is also an API, iCon, available in C++ and .NET. The version of the Ranger Studio used in this project is 5.1.

A screenshot of the software can be seen in figure 3.5. The buffered data (scans) is saved into .dat-files which can in turn be used in any preferred programming language; e.g: Matlab, C++, Python.

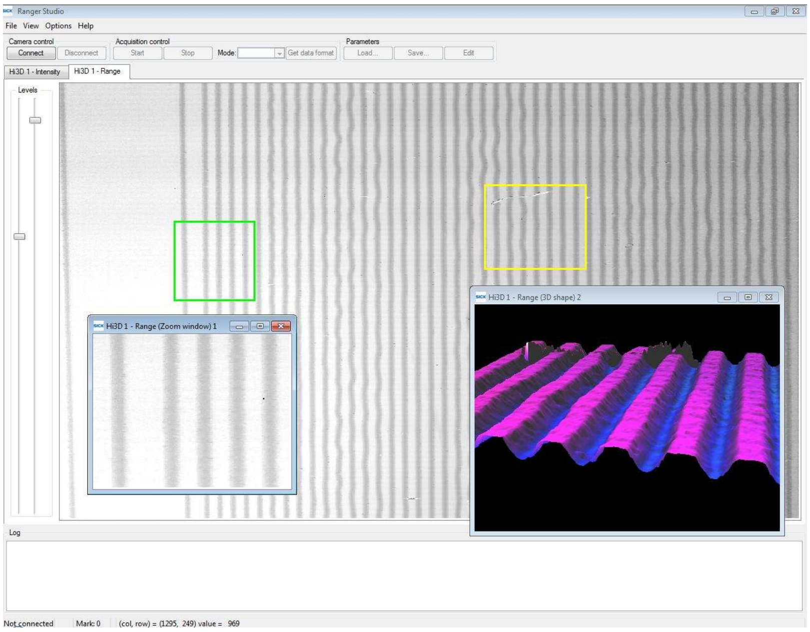


Figure 3.5: *An example screenshot of the software Ranger Studio. The scanned data is stored into slices in the y-axis as buffer. Two different magnification tools are shown, the 8-bit zoom (left) and the 3D zoom (right).*

3.1.4 Laser

In this setup; a fiber coupled laser system ZFSM/ZFMM, manufactured by Z-Laser[3] was used, see figure 3.6. This laser can achieve a thin line width of less than 10µm. A small line width, of the laser, is important due to the small modulations of the grooves which needs to be registered in each scan.

To achieve a thin line width, a short wave length in the laser is used. The wavelength in this setup is 450 nm (blue). A short wave length also reduce the speckle dimensions in the sensor. The speckles should be smaller than a pixel, registered by the sensor in the camera.

[3] `http://www.z-laser.com/en/products/product/machine-vision-lasers/zfsm-zfmm/fiber-coupled-lasers/`

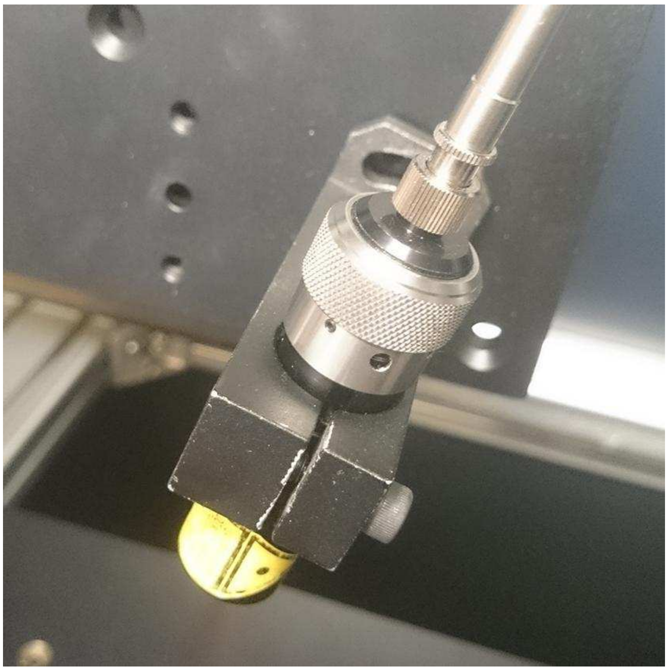


Figure 3.6: *The Laser used for this setup is manufactured by Z-Laser and provided by SICK IVP.*

3.2 Arduino Controlled Turntable

To scan the disc, a rotation mechanism is required. In this thesis; a rebuilt turntable is used. The turntable is connected to an Arduino Teensy; which is programmed to control the rotation of the turntable engine from a computer through a USB-connection. By controlling the rotation speed of the discs, different sample rates can be tested for varying setups and camera parameters. A photo of the turntable can be seen in figure 3.7.

Figure 3.7: *A photo of the Arduino Teensy connected turntable provided by SICK IVP.*
N.B. Electricity and the USB is not connected in this photo.

3.3 Photo of the Setup

The setup, as a whole, can be seen in the following photo, figure 3.8. It should be noted that this is one of many possible setups with the laser and camera. See section 3.4 for all three standard setups.

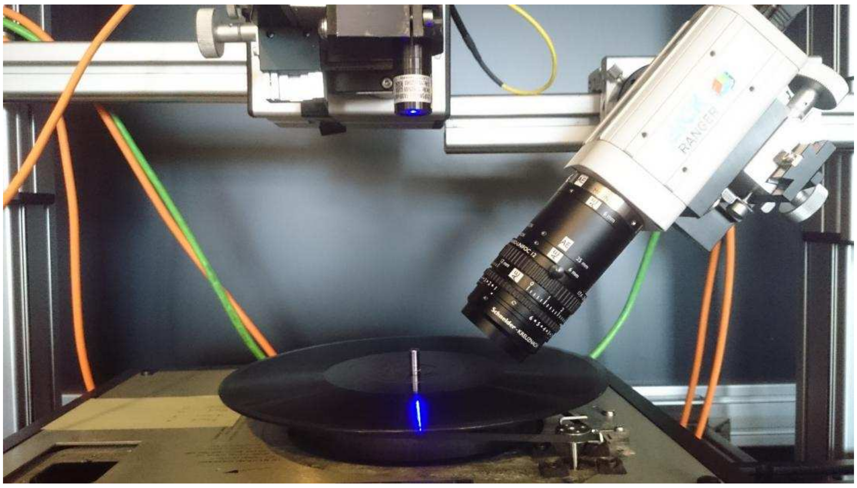


Figure 3.8: *A photo of the whole setup when a disc is scanned. This is one of standard setups (reversed ordinary, see section 3.4) where the laser is located straight above the object/disc and the camera is angled (around 30°).*

3.4 Illustrations of the setup

This section shows the three tested standard setups used in this thesis as illustrations based on the manual [IVP, 2015].

- **Ordinary** - The camera is located vertically above the scanned object. This setup achieves the highest resolution when measuring range but can result in miss-registers. A miss-register due to the measurements are made in the laser plane and without a proper calibration to align the measured data to a vertical measurement plane orthogonal to the motion artifacts/noise which is introduced with a vertical movement in the record. An illustration of this setup can be seen in figure 3.9.
- **Reversed Ordinary** - The laser is located vertically above the scanned object. This setup does not result in any miss-register but has a slightly lower depth resolution. An illustration of this setup can be seen in figure 3.10.
- **Specular** - Neither the camera nor the laser is located vertically above the scanned object. This setup is useful if the scanned object usually does return a small amount of light e.g. glossy, matte or dark objects. The miss-register, in this setup, are also made in the laser plane as explained in the ordinary setup above. An illustration of this setup can be seen in figure 3.11.

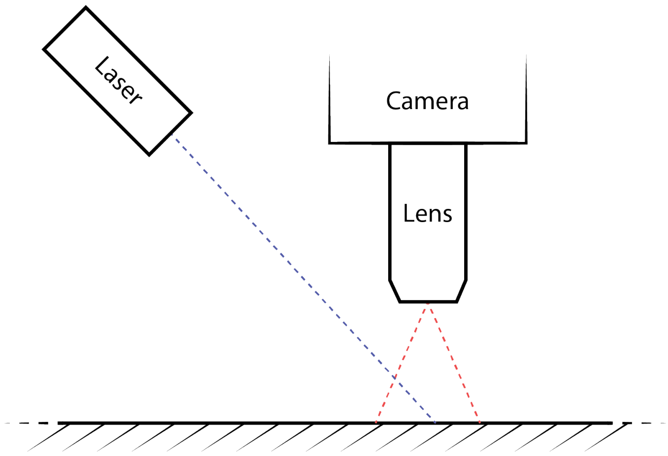


Figure 3.9: *The ordinary setup. The camera is located vertically above the scanned object as explained in [IVP, 2015, p.31].*

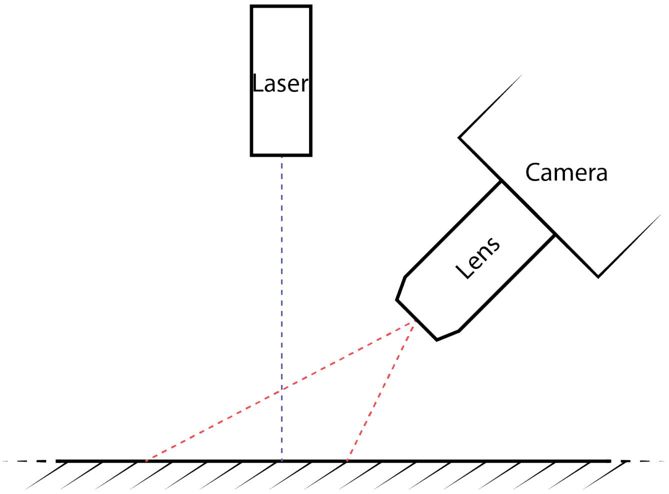


Figure 3.10: *The reversed ordinary setup. The laser is located vertically above the scanned object as explained in [IVP, 2015, p.31].*

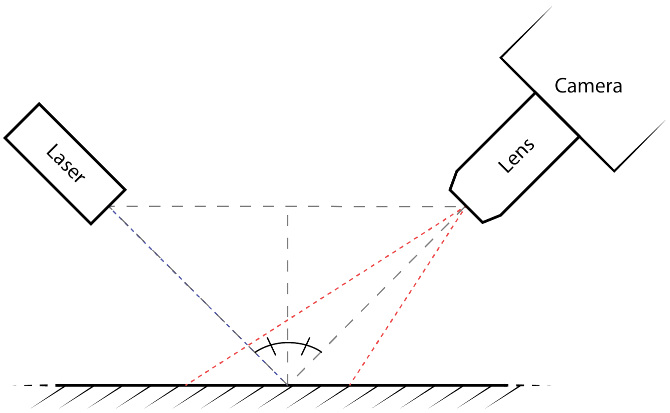


Figure 3.11: *The specular setup. The camera and the laser is not located vertically above the scanned object as explained in [IVP, 2015, p.31].*

3.5 Simulating a Phonograph Record with Archimedes' Spiral

A simple simulation of records were created in addition to this thesis. The role of the simulation is to measure the record and test measurements done in Ranger Studio. The simulation was created with an Archimedes' Spiral where the spirals radius represented groove bottoms,

$$r = a + b\theta \tag{3.1}$$

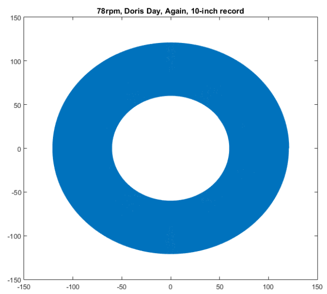
The parameter a decides the number of circumferences of the spiral and b the spacing between each turn, which is determined from a millimeter paper (see figure 3.14) and a typical scan.

The amount of turns, a, is calculated as,

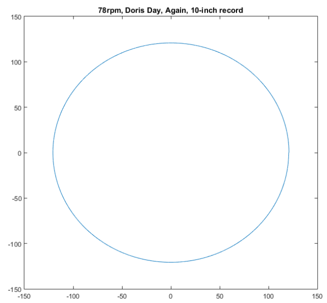
$$a = t_{audio} \cdot RPM_{record}/60 \tag{3.2}$$

The spiral starts at outermost radius and continues into the center depending on the length in time of the stored audio information, t_{audio}. While RPM_{record} is the rotation speed format of the disc.

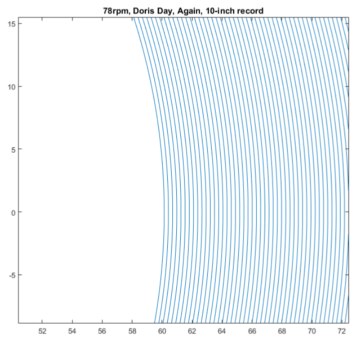
If measured correctly, the spiral stops at the radius of the innermost groove which can be examined with a simple ruler on the groove surface. Plots from the simulation can be seen in figure 3.12. The simulation also works on LP-discs if every songs outermost groove radius is determined. A plot of a simulated LP-disc can be seen in 3.13.



(a) *The whole record from start to end (167s).*



(b) *The first circumference of the simulated groove.*



(c) *A zoomed in image of the groove turns.*

Figure 3.12: *The simulated record generated with the use of Archimedes' spiral and only the measured outermost groove radius.*

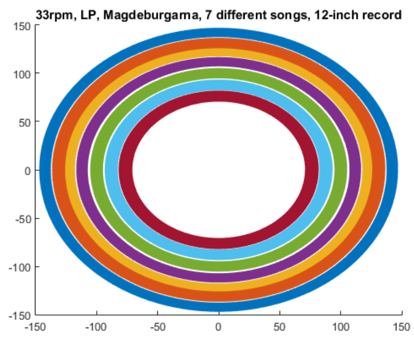


Figure 3.13: *Simulating a LP-disc with Archimedes' spiral.*

3.6 Measurements

The following tables present the initial measurements done in this thesis: table 3.1, **??**, 3.2 and 3.3. The measurements were calculated with the use of the simulation described in section 3.5 and a millimeter paper photographed by the camera (see section 3.1) and the lens (see section 3.1.4), see figure 3.14.

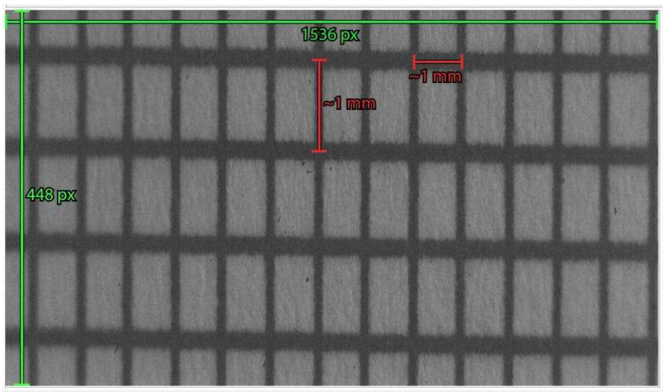


Figure 3.14: *A photo of 1 millimeter sized squares through the lens with a Ranger E camera. The photo is taken in Ranger Studio and measurements illustrated afterwards.*

Table 3.1: *Camera setup measurements (Ordinary/Reversed)*

Camera angle	0° / 30°
Width in pixels	1536 px
Width in mm	13.5 mm
Width in μm per pixels	8.8 μm/px
Height in pixels	448.0 px
Height in mm	4.0 mm
Height in μm per pixels	8.8 μm/px

Table 3.2: *Laser measurements*

Wave length	450.0 nm
Laser width	19.5 μm

[4] http://www.45worlds.com/78rpm/record/db2561http://www.45worlds.com/78rpm/record/db2561

Table 3.3: *Disc measurements done on the record 'Again' by Doris Day*[4].

Disc format	78 rpm
Printed year	1949
Disc format size	10"
Song time	167 s
Outer disc diameter	251.6 mm
Outermost groove diameter	241.8 mm
Innermost groove diameter	120.2 mm
Groove bottoms per mm	3.6848
Groove spacing, bottom to bottom in pixels	30.8776 px
Groove spacing, bottom to bottom in μm	271.383 μm
Width of groove at top in pixels	28.569 px
Width of groove at top in μm	251.10 μm
Groove length from start to end in m	123.4 m
Frequency Range [Browne and Browne, 2000, p.391]	14 KHz

4

Method

This chapter explains the proposed approach for decoding records using a laser triangulation technique. The following methods are meant to provide a general understanding of the decoding process which can be modified for stereo records by adding the Z-axis dimension to the modulation decoding.

The limitations of the provided lens, mentioned above, is the magnification. The grooves in $33\frac{1}{3}$ rpm discs are a lot smaller (by approximately a factor of 2-3 if compared with the measurements in [Stotzer, 2006, p.27]) and a higher grade of magnification is needed.

4.1 Scanning

The setup for the scanning process are described in chapter 3. Here it is also noted that the specular arrangement is not suitable for discs with the chosen camera, as the Ranger E camera has a rolling shutter and the algorithms for 3D extraction do not allow for a short enough exposure time. In the case of the specular setup; the results were poor due to saturation and were not further tested.

The scanning procedure begins by adjusting the camera and turntable parameters, adjusting the laser and camera position for best resolution (see figure 4.1), in every new attempted scan.

When one scan is completed (a single circumference of the record); approximately 13.5 mm, x_{fow}, (see table 3.1) in the records radius width can be decoded. With this width, approximately 49-50 grooves (table 3.3) should be present in the scanned buffer. If assumed that each groove represents one radial groove-spiral circumference in the record; the scanned time of the records audio information can be approximated.

Knowing the speed format of the disc RPS_{record} and the amount of groove-spiral circumferences c (as in; separate grooves in the buffer) the approximate

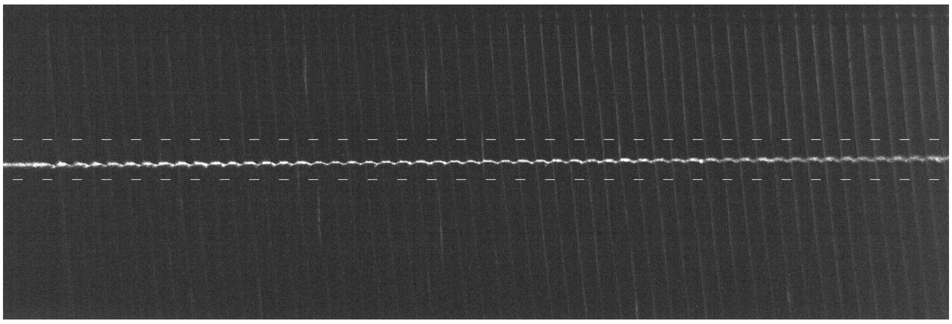


Figure 4.1: A sensor image in Ranger Studio. Adjusting the laser line and camera position for best resolution. In this image it is clearly visible that the resolution is best in the middle of the sensor and poorer in the left, right edge due to limitations of the camera lens.

audio time $\bar{t}_{audio}$ can be calculated as,

$$\bar{t}_{audio} = \frac{c}{RPS_{record}} \tag{4.1}$$

It can then be assumed that each scan has the potential to contain at least 37.6923 seconds of audio information and each separate groove circumference containing approximately 0.7692 seconds of audio information. If scanned in the start or end edge of the records radius, the time is assumed to be lower as lesser grooves will fit in a single scan. This can be seen in figure 4.2, where only 44-45 grooves are present which is approximately 33.8462 seconds of audio information.

A scanned buffer contains two datasets: range, see figure 4.2, and the intensity data, see figure 4.3. To scan the whole disc of 167 seconds t_{total} of audio information, (table 3.3) the minimum amount of scans N_{scans} needed can be calculated as a function in time,

$$\lceil N_{scans} \rceil = \frac{t_{total}}{\bar{t}_{audio}} \tag{4.2}$$

$$\frac{167}{37.6923} = \lceil 4.4306 \rceil = 5 \text{ number of scans.} \tag{4.3}$$

The number of minimum scans N_{scans} can also be calculated knowing the radius of the outermost r_{outer} and innermost r_{inner} groove, see table 3.3,

$$\lceil N_{scans} \rceil = \frac{r_{outer} - r_{inner}}{2 \cdot x_{fow}} \tag{4.4}$$

$$\frac{241.8 - 120.2}{2 \cdot 13.5} = \lceil 4.5037 \rceil = 5 \text{ number of scans.} \tag{4.5}$$

The time to make each scan is dependent on the rotation speed of the turntable and can be calculated as,

$$t_{scan} \approx \frac{N_{scans}}{RPS_{turntable}} \tag{4.6}$$

The $RPS_{turntable}$ tested in this thesis ranged from $0.025 - 0.2$.

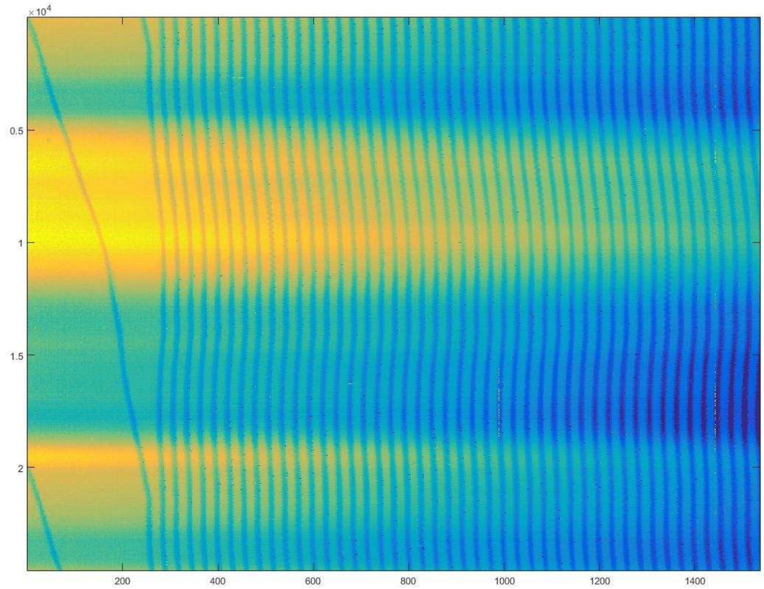


Figure 4.2: *A typical buffer containing the range data from one scan. In this scan the beginning (edge) of the disc is visible. This scan contains approximately 36 seconds of music from the song Again by Doris Day. The large dark and light areas shows that the disc is modulating vertically. An horizontal modulation is also visible as the grooves are not directly linear. The vertical and horizontal modulation is visible in the grooves and is explained in section 2.4. Observe that this figure is compressed in the tangential direction.*

To scan the whole disc; the camera needs to be moved closer to the center of the disc in every scan until the whole record has been scanned. With a perfect lens and system, the movement to the center should be a camera width size (from pixels to millimeters, see section 3.1 inwards which should result in approximately 5 scans, equation 4.3.

The lens used in this thesis acquired good resolution in $\frac{2}{5} - \frac{4}{5}$ of the screen size in each scan and rapidly lost resolution in the edges, see figure 4.1. Therefore more scans are suitable for best decoding resolution, depending on the lens. In

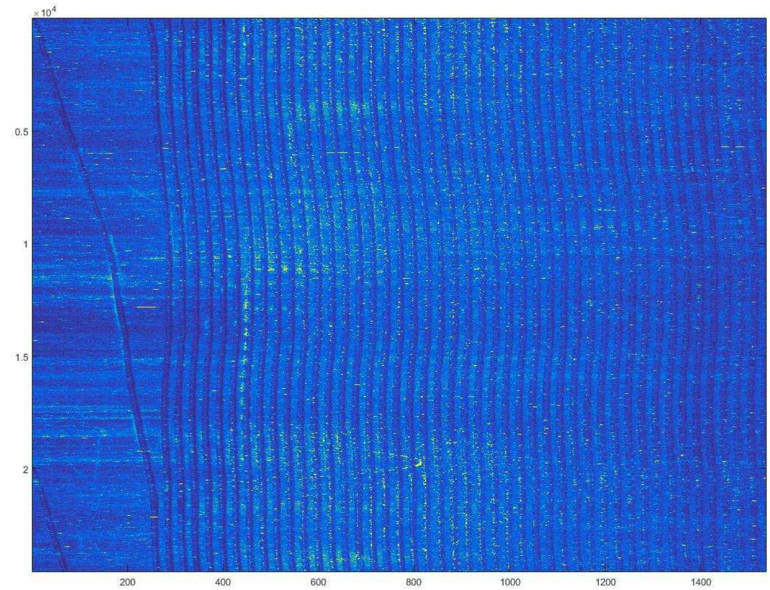


Figure 4.3: *The same buffer as in figure 4.2 but containing the intensity data of the scan. Specular reflections (bright dots) are more visible in this data. Observe that this figure is compressed in the tangential direction.*

this thesis; the movement inwards is approximately half a camera width size, which means that at least 10 scans are needed to scan the whole record in full resolution.

Analyzing the disc and its audio information is also crucial to achieve a good scanning, and later, decoding result. In table 3.3 it is noted that the scanned disc has the potential to have audio information in the frequency range of 14 KHz. The Nyquist sampling theorem states that the desired sampling rate *fs* should be at least 28 KHz for this disc. Knowing the desired sampling rate and the amount of audio time $\bar{t}_{audio}$ the grooves contain, see equation 4.1, the number of samples $n_{samples}$ in a single circumference can be calculated,

$$n_{samples} = fs \cdot \bar{t}_{audio} \tag{4.7}$$

$$28000 \cdot 0.77 \approx 21538, \text{ samples in a single groove circumference.} \tag{4.8}$$

Since the grooves have a spiral shape, it can also be noted that each groove closer to the center have a shorter circumference distance. This results in that the tangential speed of the stylus decreases linearly as it gets closer to the center and also that the samples are sampled closer in tangential direction to retain the constant sampling rate *fs* of the record.

To know the distance between each sample in tangential direction, the length of each groove circumference needs to be known. To calculate the length of a spiral based on the outer radius, l, the following equation can be solved,

$$l(\alpha, \beta) = \int_{\alpha}^{\beta} \sqrt{(r_{outer} - g_{spacing} \cdot \theta)^2 + (g_{spacing})^2} d\theta \tag{4.9}$$

Where r_{outer} is the outermost groove radius, which can be found in table 3.3, and $g_{spacing}$ is the groove spacing, which is also found in table 3.3. The α and β represents the desired length interval of the spiral from outer to inner circumference. E.g. calculating the length of the first groove of the spiral; $\alpha = 0$ and $\beta = 2\pi$.

Knowing how to measure the length of the spirals (the groove) in the record, it is now possible to calculate the amount of samples per desired length format (e.g: millimeters) in the desired radius of the disc,

$$f(\alpha, \beta) = \frac{1}{l(\alpha, \beta)} \cdot n_{samples} \cdot \frac{(\beta - \alpha)}{2\pi} \tag{4.10}$$

Where $f(\alpha, \beta)$ returns the samples per desired length format in the radius interval α to β. The $l(\alpha, \beta)$ can be solved using the equation 4.9 and $n_{samples}$ can be found using equation 4.8. The term $\frac{(\beta - \alpha)}{2\pi}$ is a multiplicative factor (amount of circumference) to the amount of samples in the radius interval.

Having a formula to measure the samples per desired length format and knowing that the needle should be moving linearly slower as it gets closer to the center of the disc; a general solution can be assumed to specify e.g. samples per millimeter in the entire disc, from the outermost to the innermost radius. But to measure such a general solution, the innermost radius in radiance needs to be known; the last groove circumference of the disc. It can be approximated with equation 4.1 but instead calculating the variable c, multiplying it with 2π, and also knowing the total time of the audio information on the disc,

$$c_{last} = RPS_{record} \cdot t_{total} \cdot 2\pi \tag{4.11}$$

Where c_{last} is the last circumference of the disc in radiance.

It is now possible to calculate the samples per desired length format on the first and last circumference: $f(0, 2\pi)$ and $f((c_{last} - 2\pi), c_{last})$ and assume the linear increase of the samples per desired length format on the entire disc; in this thesis we use samples per millimeters and the measurements for the disc, presented in table 3.3, are plotted in figure 4.4.

Knowing the samples per millimeter we can simply calculate the millimeter per sample, which is crucial for the laser setup, by taking the inverse of the values plotted in figure 4.4. The result is plotted in figure 4.5.

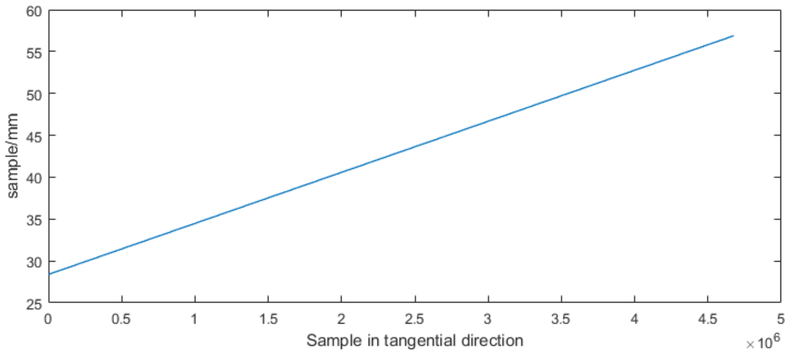


Figure 4.4: *A plot of the samples per millimeter, y-axis, in the disc presented in table 3.3 where the x-axis represents every sample in the disc from the outermost to the innermost sample. It can be seen that in the outermost spiral the sample per millimeter is around 30 while it linearly increases until it reaches the innermost radius where it is approximately 55 samples per millimeter.*

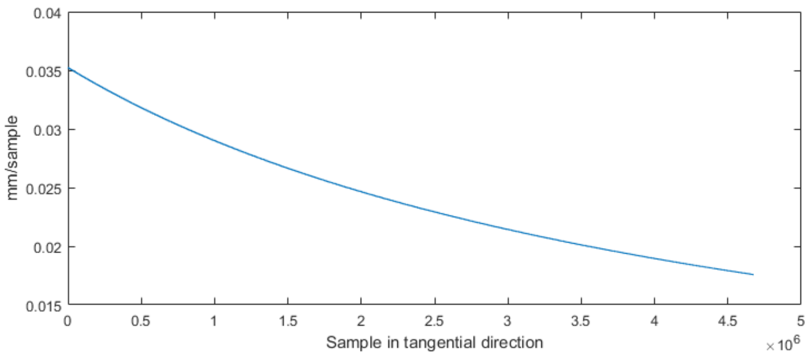


Figure 4.5: *A plot of the millimeters per sample, y-axis, in the disc presented in table 3.3 where the x-axis represent every sample in the disc from the outermost to the innermost sample. From this measurement it is possible to approximate that the outermost sample have a length of 35* µm *in tangential direction of the groove and the innermost have a length of 17* µm.

4.1.1 Laser width and scanning frequency

To achieve a high resolution scan with the camera, the laser width should be smaller or as close as possible to the length of each sample in the scan sequence to minimize smearing defects of the samples. The length of each sample is measured and calculated in section 4.1 and plotted in figure 4.5 for the disc presented in table 3.3. Knowing that the laser width is 19.5 µm, from table 3.2, it is safe to assume that the laser width is thin enough for the sample lengths in this disc. As mentioned, the laser width is smaller than the sample size, a smearing of the samples are avoided, see an illustration in figure 4.6. But having a thin laser increases the missed patches size between each scan, which is seen as a down-sampling of the original audio information, but avoids smearing. A note here is that this assumption is done by assuming that the integration period of the camera to laser is zero; an unrealistic simplification as some smearing will always occur (even if the integration period is extremely close to zero). Therefore, a safer approach to scanning the data is to assume that smearing occurs on every scanned sample and that is why a thinner laser width is used than the assumed sample length.

A reduction in the audio quality of the scanned disc cannot be avoided fully, but a down sampling of the audio information is favorable before smearing.

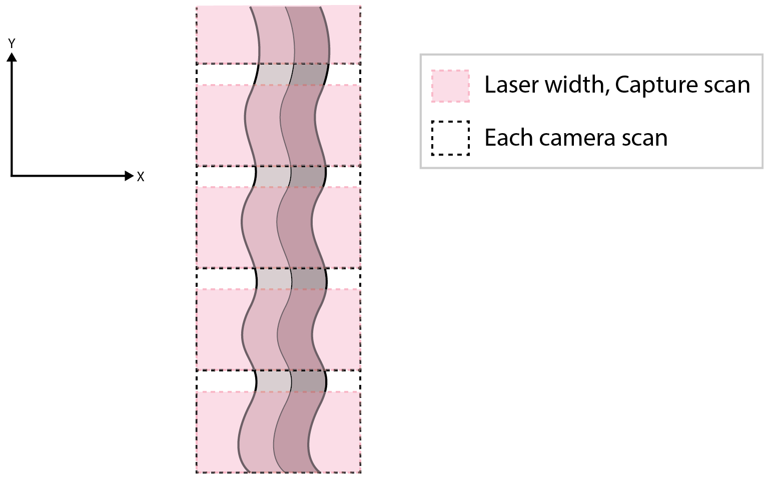


Figure 4.6: *An illustration of the laser width in each scan. N.B. The groove modulation is greatly exaggerated.*

As it is assumed that the width of the laser is thin enough to scan the audio information in the disc. But to achieve a good scan, the camera frequency needs to be adjusted for the desired speed of the scanning turntable. The correct scanning frequency, C_f of the camera can be calculated as,

$$C_f = RPS_{turntable} \cdot n_{samples} \tag{4.12}$$

The equation can be adjusted depending on if we have a fixed camera frequency or rotation speed of the turntable and $n_{samples}$ can approximated from equation 4.8,

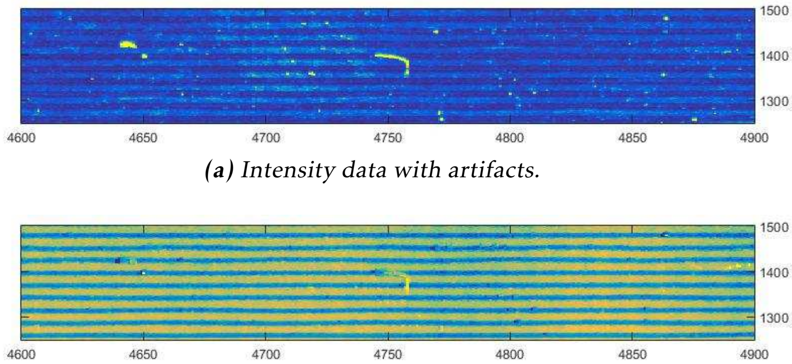
$$RPS_{turntable} = \frac{C_f}{n_{samples}} \tag{4.13}$$

4.2 Artifacts in Scans

Artifacts, or 'noise', of varying types can appear in the scans. An artifact in this thesis are areas or pixels in the scanned dataset that reflects light in irregular and unwanted directions, creating very dark or light areas. This section explains which types of artifacts that appeared during the scans and the next section, 4.3 explains how they are removed.

The varying types of artifacts that appears in the scans of this thesis are the following:

- **Physical Uncleanness** as dust, small hair, dirt etc. The most of these can be removed by cleaning the record before scanning. These small particles are seen as artifacts that potentially creates depth differences or shadows the grooves in a 3D-reconstruction laser triangulation approach. If not removed, the dirt will create unwanted high frequency 'tics'-defects in the decoded information. A dust particle is visible in figure 4.7.



***(a)** Intensity data with artifacts.*

***(b)** Range data with artifacts.*

***Figure 4.7:** Samples from the dataset showing typical artifacts found after the scans. In the middle of the images: a typical dust-particle is visible accompanied by specular irregularities and speckles.*

- **Physical Irregularities** as scratches or broken grooves/samples. These irregularities are also visible as artifacts since they reflect the light in irregular directions. Physical irregularities are hard to find in the records used in this thesis as they were probably tiny and were deemed as physical uncleanness, due to their nature of reflecting light. This system is not robust enough to fix greater physical irregularities as shown in [Stotzer, 2006, p.26, fig.2.6].

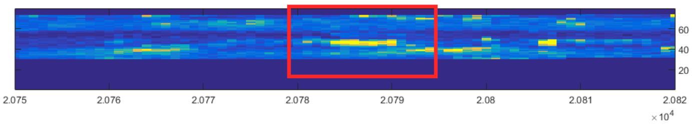
- **Specular irregularities and speckles** which were the majority of the artifacts in the scans. These can be seen as small white dots in different shapes (usually a single pixel) over the whole scanned dataset. Specular irregularities happens when close too 100% of the light in surface is reflected into the camera. With a vertically modulating system, explained in section 2.4, and an observation in the scanned dataset; this appears to happen more frequently in areas closer to the laser where the vertical modulation of the disc is closer to its maximum. The speckles appear in the scanned dataset where the constructive interference from the laser creates an intensity maximum. These are formed on the surface of the sensor and are unfortunately inevitable in a laser system. Speckles gives the scanned intensity dataset a very "grainy" look, and in some cases generates a very high intensity.

The specular irregularities and speckles can be seen in figure 4.7. If the specular irregularities and speckles are not handled, they will create an additive and irregular noise on the decoded audio information of the disc.

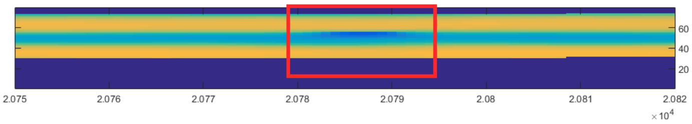
- **Shadowing** occurs when irregular light occurs on one of the two walls of a groove. That is; one of the walls, which reflects a lot of light, the reflected light hits the second wall and is registered in the camera sensor as a mix of the laser reflection from the second wall and the bright reflected light from the first wall to the second wall. The sensor then calculates the mean of these bright reflections, which creates a so called 'double reflection', creating a shadow defect in the range data; which are observed as deep valleys in the range dataset.

  Shadowing is hard to detect, as the decoding still sees the groove as a valley, and does not seem to occur as frequently as the other artifacts. A shadowing defect can be seen in figure 4.8. The shadowing defects creates quick and irregular modulation which can be heard as 'tics' or 'clicks' in the decoded information.

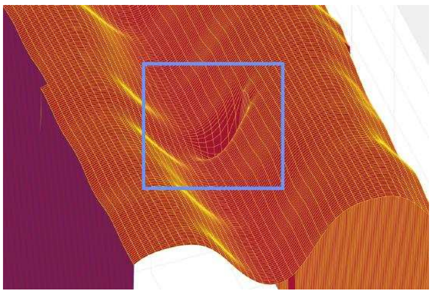
- **Hot pixel effect** during the scanning process the system was operating at its limits, there were little light, and for the chosen sensor, a long exposure time. In a CMOS sensor (the sensor used in this thesis) some pixels have a larger dark current, giving a higher signal even without illumination. In a column with such a "hot" pixel the result is that; if there is little light in the scene, the extracted position will always be in this position. In the scanned range data this can be seen as streaks of 3D-data at a fixed height appearing in a column where the light is very low. The hot pixel have 1 pixel width and are represented as dotted vertical line in the range dataset, which can be seen in figure 4.9.



(a) In the middle of this groove, bottom (left) wall is reflecting a lot of light.

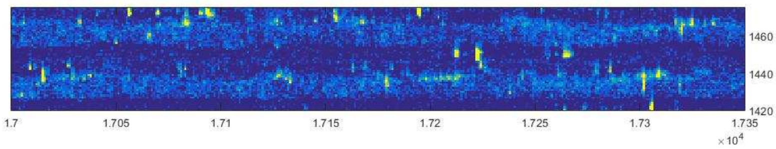


(b) The strong light reflection created in the bottom (left) wall is shadowing the groove next to the top (right) wall.

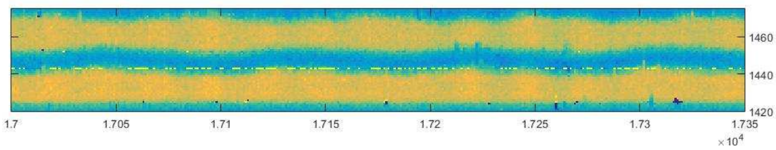


(c) In a 3D-mesh created from the range data, a deep valley can be seen in the top right wall due to the shadowing defect.

Figure 4.8: *A shadowing defect as seen in the intensity data, range data and a 3D mesh of the range data.*



(a) Intensity data with a hot pixel line.



(b) Range data with a hot pixel line.

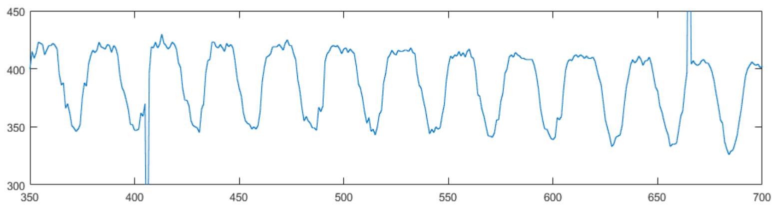
Figure 4.9: *The hot pixel line is slightly visible in the middle of the scanned range data.*

4.3 Artifact Removal

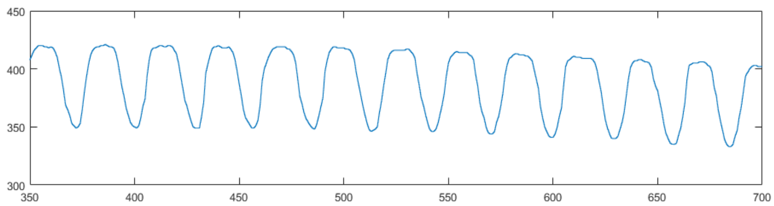
Before decoding the scanned information from the camera; the artifacts mentioned in section 4.2 need to be handled as far as possible to increase the quality of the sound and the robustness of the system. To achieve a good initial data for the decoding, two preprocessing techniques are used on the range dataset to remove as much artifacts as possible. The two preprocessing methods used are the following:

1. **Median filtering** - A median filter is performed in large patches (9-13 pixels depending on the dataset) in the x-axis and in a small patches (3 pixels) or none in y-axis of the range data buffer. Due to the large width of the grooves, a large patch in x-axis is a beneficial method to remove scan irregularities in each scan. The smaller patch in y-axis (or the θ rotation position of the grooves in the dataset) is to preserve high modulation frequencies but in the same operation remove very high modulations in time.

   The median filter removes artifacts created from: Physical Uncleanness, Physical Irregularities, Specular Irregularities and hot pixels.



(a) A 2D-profile plot of the range data before using the median filter. Note that there is large 'drops' and 'jumps' in the plot.



(b) The same 2D-profile plot of the range data after using the median filter. Not only is the large 'drops' and 'jumps' smoothed out but the overall structure of the grooves are smoother but still keeping their overall structure.

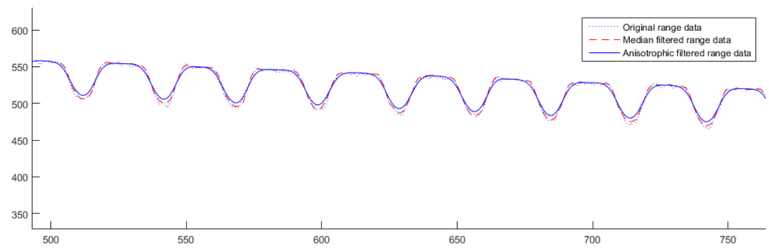
Figure 4.10: *Two 2D-profile plots of the range data before and after using the median filter.*

2. **Anisotropic Diffusion** - Smooths out high frequency range data with an iterating 2D anisotropic diffusion. The anisotropic diffusion preserves the edges (groove information) and smooths irregularities in X- and Y-axis.

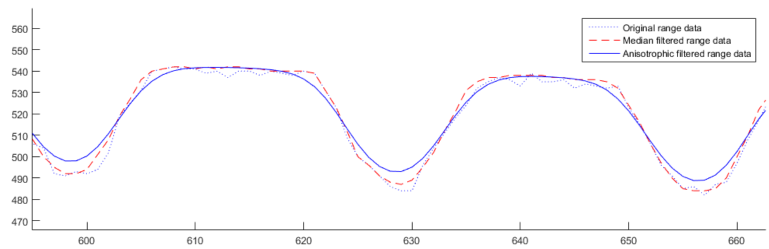
Depending on the scanned data, the amount of iterations might differ from 6-16 iterations or fewer depending on the data set and how well the median filter performed. Its preferred to use as few iterations as possible, since the anisotropic diffusion smooths out the data in the y-axis. Too many iterations will have the same effect on the decoded audio information as a low pass filtering.

The anisotropic diffusion used in this thesis is a Matlab implementation based on [Kovesi, 2002] which in turn is based on the theories of [Perona and Malik, 1990].

As seen in figure 4.11 and 4.13 the median filter alone do a great job in removing artifacts in the range data and creating well fitted grooves; but not in all cases as sometimes the median filter creates irregularities on its own due to the large median filter patch done in the x-axis. This is where the anisotropic diffusion can assist in forming the data into smooth curves. This can be seen in figure 4.12, as the median filter created a 'spike' in the right groove wall. The anisotropic filter smoothed it out.



(a) Zoomed out plot of the range data before and after filtration.



(b) Zoomed in plot of the range data before and after filtration.

Figure 4.11: *A plot of a typical scan in the range data, XZ-dimension.*

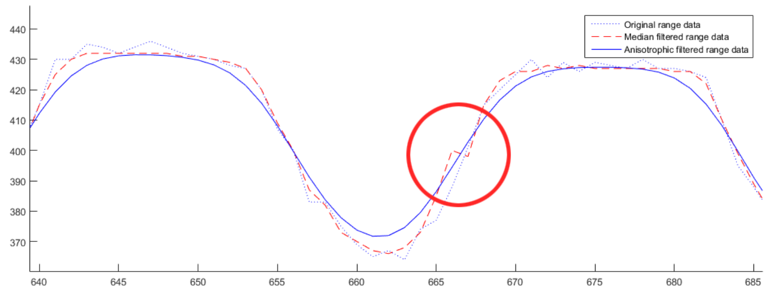
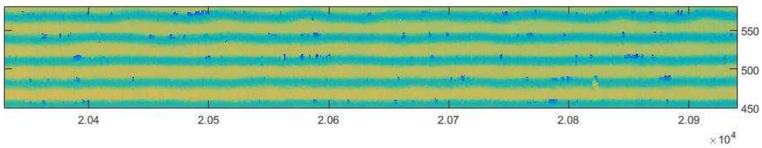
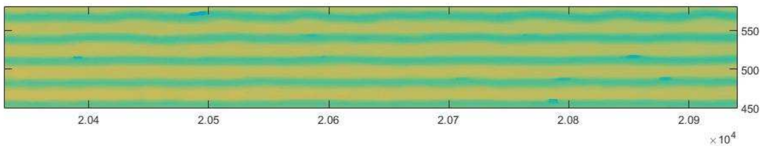


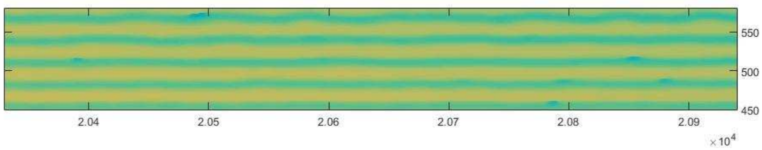
Figure 4.12: *A single irregularity was formed on the right wall in this sample after the median filter. These kind of artifacts happens frequently and if not handled, creates high horizontal modulation which potentially results in a 'click' or a 'tic' sound defect.*



(a) *A segment from the range data buffer were a lot of noise is present.*



(b) *The same segment as in figure 4.13a after using the median filter.*



(c) *The sample from 4.13b after 8 iterations of anisotrophic diffusion.*

Figure 4.13: *The preprocessing steps performed on the buffer to remove artifacts and smooth out irregularities with the anistrophic diffusion. The preprocessing is performed in the order from (a) to (c).*

4.4 Segmentation

After the artifact removal steps; segmenting the grooves from each other is done in order to decode the information of each groove separately. The grooves, as seen in figure 4.2, are formed as nearly vertical valleys in each buffer. Each groove starts in the first y-position (top) and ends in the last y-position (bottom). This statement is not true in the real case; as there is in fact only one groove which continues as a spiral from the edge of the disc to the center. But in this thesis; we state that each vertical valley is a different groove. Doing this lessens the amount of necessary scans and opens up the possibility of parallelising the decoding process of the entire record; groove per groove.

In this thesis, segmentation of the grooves in the range data was done with a path segmentation algorithm explained in section 4.4.1.

4.4.1 Path Segmentation

The idea of Path Segmentation is that a groove has a start position in the first values of the range buffer. The starting position is each grooves horizontal position of the found local minimums. Each horizontal starting position is then tracked, vertically, through each groove; finding the shortest route to the end of the buffer. A complete groove is considered a groove which begins in the first row of the buffer and ends in the last row. An incomplete groove is a groove which does not start or end in the first respective last row in the buffer. The grooves which are deemed incomplete are removed and never segmented or decoded. As these incomplete grooves only appears in the leftmost and rightmost edges of the buffer; the removed incomplete grooves are deemed redundant, as new scans (with a movement of the camera) will include the removed grooves again as a whole.

The first step in the path segmentation algorithm, is to down sample the range data in y-axis. This is an appropriate method in order to speed up the segmentation process and is possible as the grooves horizontal modulation is small (±1 pixel for every 10-40 rows, may differ for different records). The down sampling is done with a fixed step-size, in this thesis its chosen as 20 pixels. For every step, the row is extracted from the range data to create a temporary down sampled data.

The next step is to find the peak-positions in every row of the range buffer; using a peak detection algorithm. The peak-positions are the local minimums; the lowest range values in each groove-valley, see figure 4.14.

With the minimums found in every row of the down sampled buffer. The next step is to track each of the start positions through all found minimum-positions, creating a path; to find the shortest route to the end of the buffer. To handle incomplete grooves; grooves which start at the top of the buffer but never finishes to the bottom. A small threshold is set, the threshold checks if the grooves path has been paired with a minimum to far away from its path. As the path should be nearly vertical in a discs groove. If the path has been paired with a minimum position to far away from its former positions; the path search stops and a groove is considered incomplete. The minimum search step is illustrated in figure 4.15.

The minimum search has now generated a path for every complete groove found. The paths are now up sampled by repeating each position with the down sampled step size in y-axis. Around every step of the paths; a region in a mask, created for each groove, is set to ones. The width of each region should be, in pixels, larger than the width of a groove (see section 3.6) while the height is the

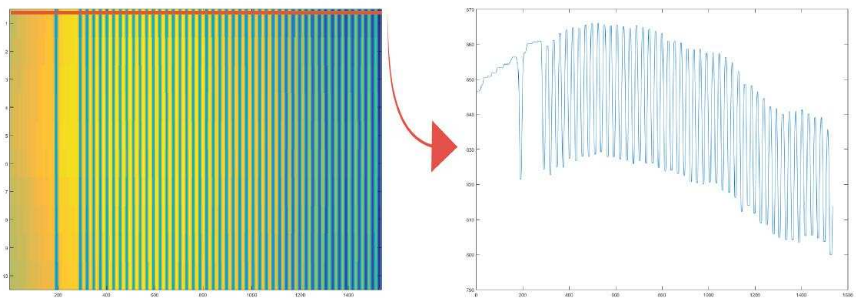


Figure 4.14: *Plotting the first row in the range-buffer displays the dark areas of each groove as a local minimum.*

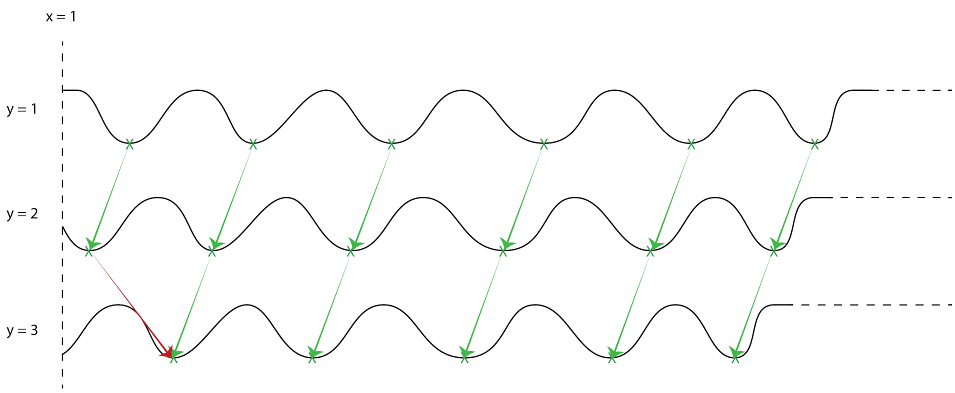


Figure 4.15: *The green arrows shows the searched path from found minimums in every row of the down sampled and processed range buffer. The red arrow shows what happens when a groove is incomplete; it will find the wrong closest minimum. N.B. The distance is poorly illustrated in this figure; a wrong minimum search has a much larger distance than a correct guess.*

down sampled step size. The region width is exaggerated to include the local maximums of each groove found at the top of the left and right wall in each groove (except the first two grooves from the outer edge, as they have a wider spacing than the inner grooves).

A path segmented groove can be seen in figure 4.16. After segmenting each groove; a decoding of the range buffer can be initiated.

Having the grooves segmented, enables the possibility to calculate the amount of audio information in time captured by the scans. To determine the time, in the record of each groove, the following equation can be solved,

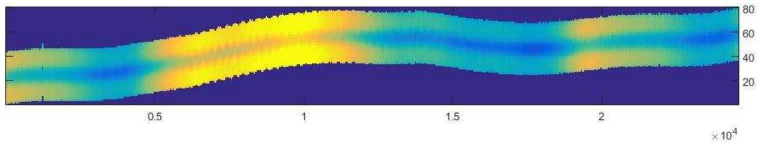


Figure 4.16: A segmented groove with the use of path segmentation algorithm. The step size is visible in the edges of the segmented information.

$$t_{scan} = \sum_{j=1}^{k} \frac{y_k}{fs} \tag{4.14}$$

Where y_k is the amount of row samples in each groove k, fs is the sample rate (is explained in section 4.6.2) and t_{scan} is the total number of seconds of audio information in the scan.

4.5 Decoding

In this section, and the following subsections, the decoding procedure of the segmented range data grooves is described. In this section, referring to a groove means a segmented groove after path segmentation and a sample is a single row in a groove. That is; a groove consists of samples in each row.

4.5.1 Groove Extremas

In this thesis, the groove walls are chosen for tracking and decoding the audio information in each groove. This is to simulate a real needles movement in the grooves as general as possible and if further development on these methods would be done on stereo records.

To counter the macroscopic vertical modulation, finding the extremas in order to track the vertical shape of each groove is introduced. The idea is to create vertical scale invariant solution to the macroscopic vertical modulation defect.

After using an anisotropic diffusion, section 4.3, to smooth out all the samples in the range dataset, it is assumed that every sample have an 'U'-shape, see figure 4.17.

In this shape; each sample should have at least two local maximums and one local minimum. The two local maximums should be located at the top of both left and right wall and the local minimum somewhere in between these maximums. As in the path segmentation step, section 4.4.1, a peak detection is used to find the local minimum in each sample. When the minimum is found, the sample is split in a left and right side to find local maximums, with the peak detection, in each part. This is done in every sample and groove. A groove with found extremas can be seen in figure 4.18.

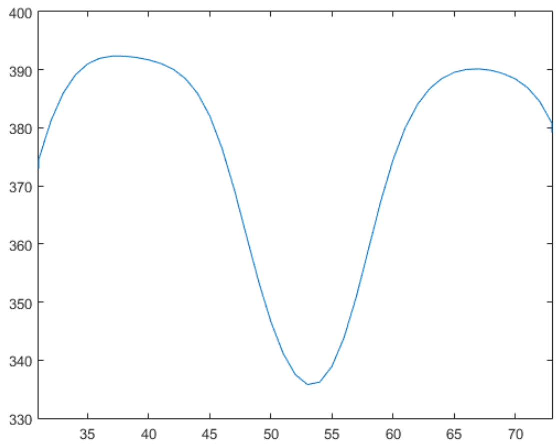


Figure 4.17: *A typical sample. A sample have an 'U'-shape and formed like a valley.*

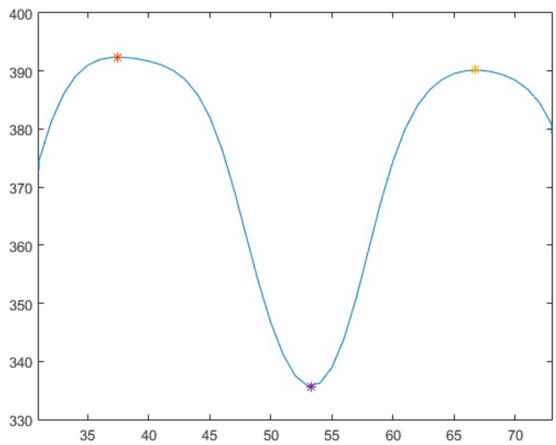


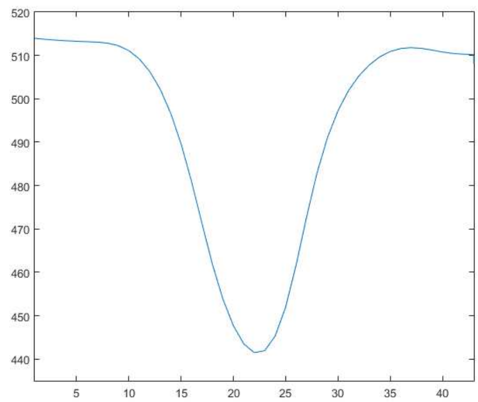
Figure 4.18: *The same sample as in figure 4.17 but with found local maximums and minimum marked as stars.*

4.5.1.1 Special Shaped Samples

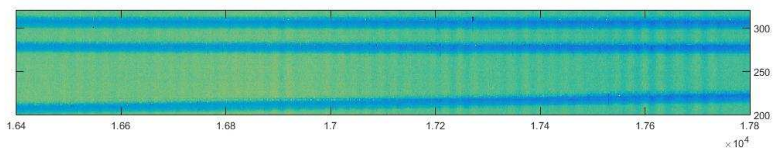
All samples have a single local minimum, but not every sample have two or even a single local maximum. This tend to happen in the samples found in the grooves next to the edge of the record, which is logical since the grooves there are further away from neighboring grooves which in turn does not create the typical valley walls. The edge grooves are handled as special cases. As can be seen in figure 4.19, the edge sample walls have a linear shape and a very slight vertical modulation.

One approach would be to measure a fixed distance from the minimum position to the left and right side of each groove sample to create maximum wall positions. This solution can be troublesome to use as there is not always the same amount of samples in both the left and right wall surrounding the minimum position.

The approach used in this thesis is to pick maximum positions in the left and right edges of each groove sample where local maximum positions could not be found. This is due to that the horizontal position is not important for these cases, as only the maximums vertical position is used to track the walls in the groove. The solution to these cases are more understandable after reading section 4.5.2 where the wall position is defined from the vertical maximum positions.



(a) *A typical edge sample with hard to define local maximums.*



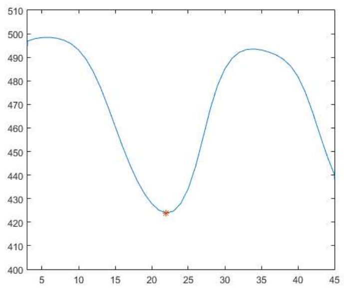
(b) *Edge grooves with visible long distances to neighboring grooves.*

Figure 4.19: *Grooves in the edge of the record are handled as special cases due to their lack of definable local maximums.*

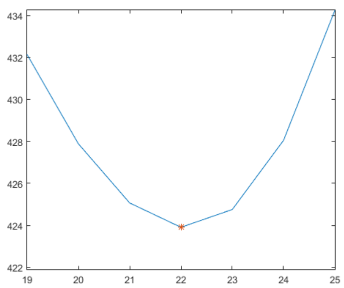
4.5.1.2 Finer Groove Extremas

As can be seen in figure 4.17, the samples have a rather low width resolution. The low resolution results in that the dynamic amplitude of the audio information is suppressed, due to the low resolution modulation directly from the scans.

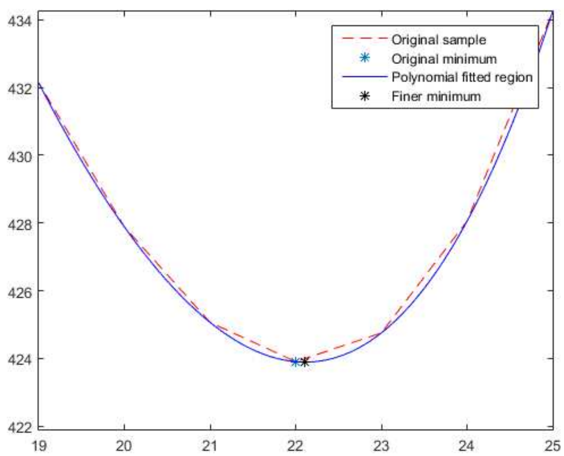
The idea is to counter the scanned low resolution by up sampling the found extremas and redefine new up sampled extrema positions. The up sampling is done by creating a small region (in this thesis the region is 7 pixels in width) around each extrema in every sample and create polynomial curve fitting of at least 3 degrees. The finer extrema position exists in the roots of the polynomial fitted function. The method of finding the finer extrema positions is visualized in figure 4.20.



(a) *A typical profile plot of a groove with its found local minimum position marked as a small star.*



(b) *A region of 7 pixels width are cut out around the minimum, the minimum position is marked as a small star in the lower middle section of the plot.*



(c) *A polynomial curve fitted function is created in this region and plotted on top of the low resolution plot (dotted line). The roots of the polynomial function has a new finer minimum position. As can be seen in the image, the minimum position have moved slightly to the right of its original position.*

Figure 4.20: *The step by step method of finding finer extrema positions in the samples.*

4.5.2 Intersection Line

Having extracted two local maximums and one local minimum in each sample, an intersection line D is created. The line D is created to simulate where a real needle would hit the left and right wall of each sample.

In order to create the line D, the following equations were used to determine the line vertical start and end position,

$$D_{L,i} = (y_{max,L,i} - y_{min,i}) \cdot d + y_{min,i} \tag{4.15}$$
$$D_{R,i} = (y_{max,R,i} - y_{min,i}) \cdot d + y_{min,i} \tag{4.16}$$

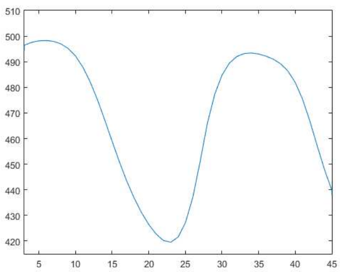
where y is the vertical position in sample i and d is a scale constant to determine at which height the line should be created. Where d is,

$$0 < d < 1 \tag{4.17}$$

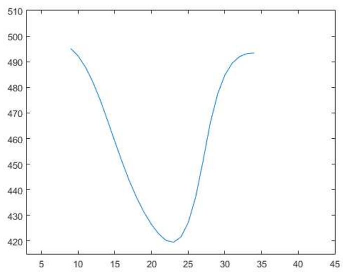
The line D vertical starting position is $D_{L,i}$ and end position $D_{R,i}$ in each sample. The horizontal positions for the line start and end is the corresponding left- and right maximum horizontal position.

A line, D, should now cross every sample depending on the local maximums and minimum at a scale height d. To extract the wall-position with the line, another region is created around each sample. The region size is cut out from the left to the right maximum horizontal position. This region is created to avoid any unnecessary intersection of the line and the sample; as only two intersections should happen: one in the left wall and one in the right wall. The position extraction is visualized in figure 4.21.

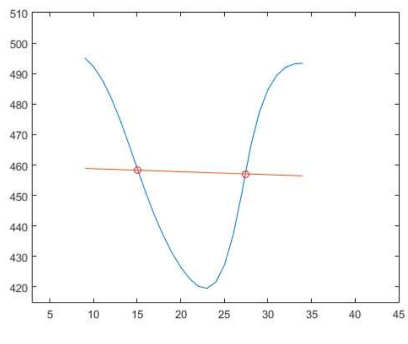
After extracting all the positions; they can, as a test, be plotted in the grooves e.g. 3D-mesh, 2D-XY-plot and 2D-XY-plot on extracted dataset. See figure 4.22, 4.23 and 4.24.



(a) *A typical sample.*



(b) *The region from the samples local maximum in left and right is cut out.*



(c) *The line D is inserted. The sample intersections of the line are the tracked wall positions. The scale constant d in this figure is* 0.5.

Figure 4.21: *Extracting the wall positions using an intersecting line D.*

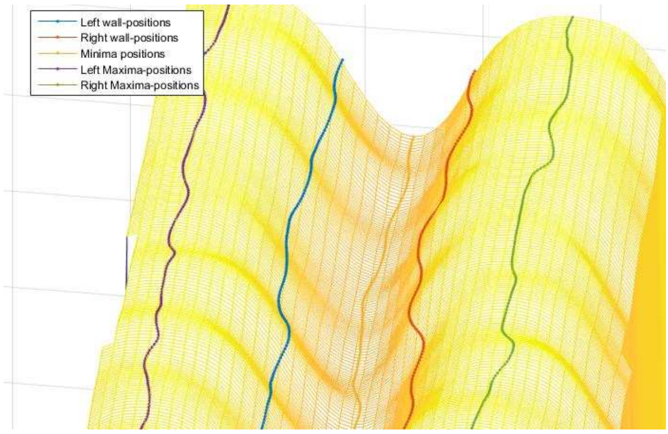


Figure 4.22: *A 3D-mesh plot using the range dataset with plotted sample extremas and the wall-positions found using the intersection line.*

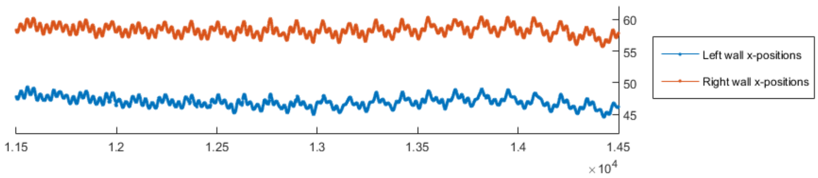


Figure 4.23: *A 2D-plot of the samples left and right horizontal wall position in XY-domain.*

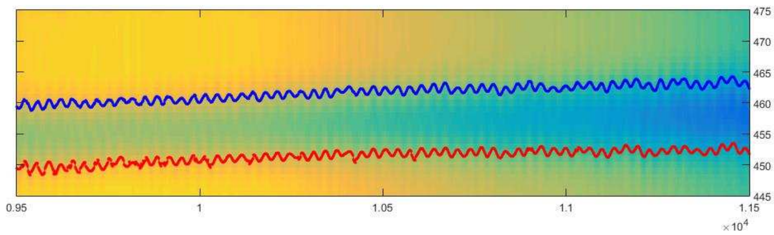


Figure 4.24: *A 2D-plot of samples left and right horizontal wall position plotted onto the dataset they were extracted from.*

4.5.3 Stylus Position in Sample

Using an intersecting line, as explained in section 4.5.2, is not the only approach for tracking wall positions of the grooves samples. An alternative approach is to simulate an actually sized stylus in the grooves. Tracking the positions it would intersect in the groove walls. To create a simulated stylus; this thesis measurements were used (section 3.6). But since this thesis lack the size of an actual stylus, the radius is borrowed from [Stotzer, 2006, p.27]. We assume that the stylus radius is 80µm. Knowing its radius, it can be transformed into pixel values for the samples by comparing to section 3.6,

$$\frac{80}{8.8} \approx 9.1 \text{ pixels in radius} \tag{4.18}$$

Also knowing that the scanned disc is a mono disc and assuming that the samples should have constant size a constant m can be created, which represents the intersection width in the walls. This is illustrated in figure 4.25.

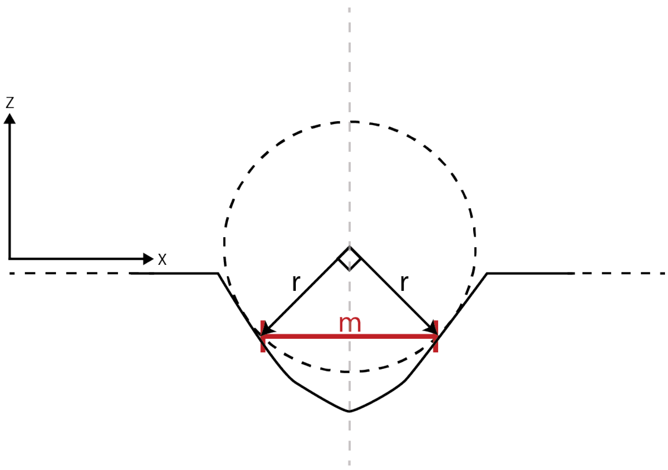


Figure 4.25: *Assuming that size of the grooves are constant and intersection width m could be calculated.*

The intersection width m is calculated as,

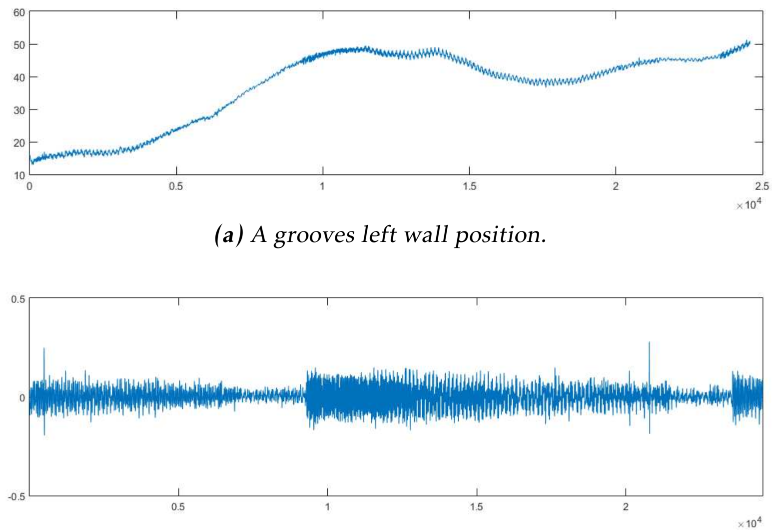
$$m = \sqrt{2r^2} \tag{4.19}$$

An intersection line, as explained in section 4.5.2, is then moved up- and downwards during a minimization process in every sample. That is; until the width of the intersection line is minimized to the constant m. The wall positions are then extracted at the intersection position.

The extraction line was found to be more robust, as such, fitting a constant width stylus in the grooves was not further tested in this thesis.

4.5.4 Modulation to Sound

The audio information stored in a record is in XYZ-domain for stereo and XY-domain for mono. In this thesis, the decoding is done by a derivative of the magnitude position. The magnitude position for stereo is in vertical, horizontal (XZ) and for mono only horizontal (X). The derivative (modulation) along each sample row (Y) will then be the raw audio signal from the record for each groove. The raw audio signal is referred to the pre-filtered audio signal in the record and is split up into as many pieces as there were grooves and each groove consists of a left and right wall audio signal (the horizontal modulation in the left, respectively, right wall position). The signal, before and after derivation can be seen in figure 4.26.



*(**a**) A grooves left wall position.*

*(**b**) A grooves left wall position after derivation.*

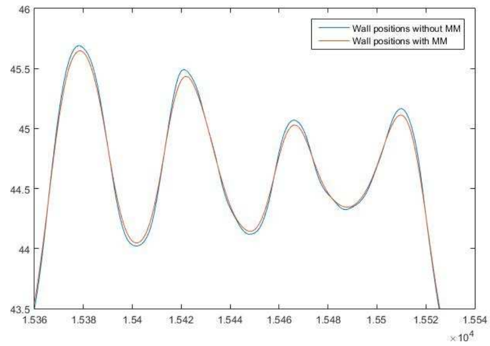
Figure 4.26: *A grooves wall position before and after derivation into an audible audio signal.*

4.5.5 Filtering

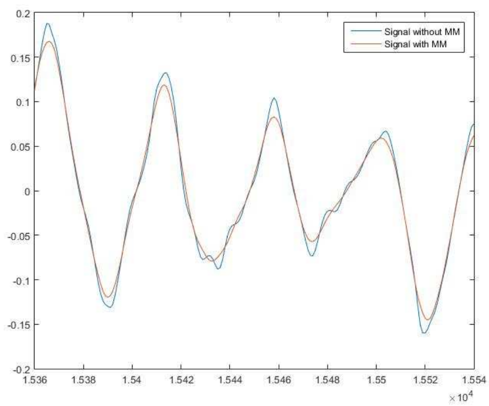
Before and after the raw audio signal has been extracted, as explained in section 4.5.4, two filtering methods are used to increase the quality of the signal from every groove. A **Moving Average filter** on the wall positions before derivation and a **Maximum Suppression filter** for click removals on the post derivative signal.

4.5.5.1 Moving Average - Smoothing

When listening to the raw audio signal, a lot of high frequency noise is hearable. The idea to use a moving average filter on the wall positions is to decrease high frequency modulation in the Y-axis before derivation occur. Doing a subjective listening to the audio signal, before and after, a quality increase of the audio can be heard. The result of the filter is visualized in figure 4.27.



(a) *A segment of left horizontal wall positions with and without moving average filter.*

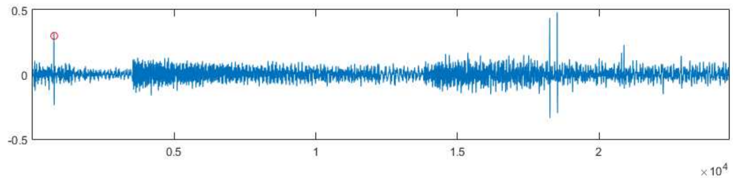


(b) *A segment of derivated left horizontal wall positions with and without using a moving average filter on the extracted wall positions. As can be seen in the middle of the figure, some high frequency modulations has been smoothed out.*

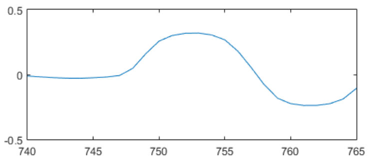
Figure 4.27: *Plots of the signal, with and without the use of a moving average filter.*

4.5.5.2 Maximum Suppression filter - Click Removal

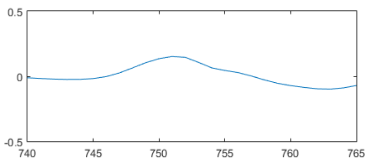
After the horizontal modulation has been derivated into an audio signal in each groove, a basic click removal is performed using maximum suppression. The maximum suppression is defined by creating a threshold in each grooves audio signal. The threshold is the median of the sorted and absolute valued audio signal. If a sample in the absolute valued audio signal is above this threshold, its set to zero. The zeros in the audio signal is then handled by using a normalized convolution [Fisher, 2003], to try repairing them; effectively removing high 'clicks'-defects in the audio. The before and maximum suppression is used can be seen in figure 4.28.



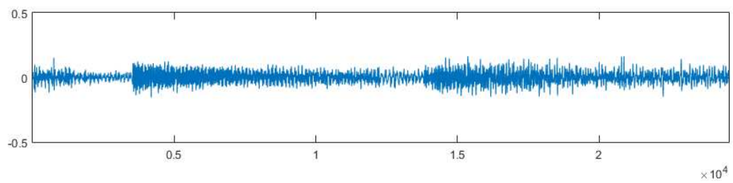
*(**a**) A grooves audio signal before click removal. The red circle in the left side is the zoomed in 'click' in figure 4.28b.*



*(**b**) The zoomed in 'click' from figure 4.28a*



*(**c**) The zoomed in 'click' after maximum suppression and normalized convolution.*



*(**d**) The groove audio signal after click removal.*

Figure 4.28: *Plots showing the click removal in the audio signal.*

4.6 Sound Stitching

The last step in the decoding of the audio signal from the grooves is to stitch the separated groove information together into two long tracks, a left and right channel; as every groove consists of both left and right horizontal wall modulations.

In this thesis, it is assumed that the scans are not perfect and a short padding is left at the end of each groove; more than a full circumference has been scanned. The padding at the bottom of a groove, G_j is the beginning of the next groove, G_{j+1}, to the right and so on (see figure 4.29), the padding is unknown.

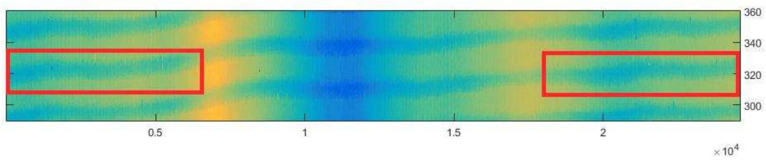


Figure 4.29: *Somewhere in the red area to the right and in the red area to the left begins the next groove.*

To approximate where to stitch the decoded samples together, the following three constants needs to be known:

- $RPS_{turntable}$ - The rotation speed of the turntable.
- c_f - The camera scan frequency.
- s - and the number of scanned rows in the original dataset.

The number of samples for one circumference can then be calculated as,

$$\overline{s} = \frac{c_f}{RPS_{turntable}} \tag{4.20}$$

The number of padding samples, $\overline{p}$ at the start and end of each groove is,

$$\overline{p} = s - \overline{s} \tag{4.21}$$

The last $\overline{p}$ number of samples in G_j is assumed to be the perfect spot where the stitching should be performed. That is; if the scan frequency and the turntable are working in perfect conditions. It can be assumed that the system is not flawless and e.g. the arduino controlled turntable is not rotating at a constant speed during the whole scanning process. Instead we use the last $\overline{p}$ number of samples in G_j as an approximation on where the stitching can/should be performed with G_{j+1}.

If we assume that $(s - \overline{p})$ is the stitching position in G_j to the next groove G_{j+1}; a small interval (a one dimensional region of interest) is created around this position in G_j. The interval is put to 100 samples in both directions around the stitching position in G_j, in this thesis, creating an area of 201 samples in size H_j. An equally large interval is extracted from the beginning of G_{j+1}, creating

H_{j+1}. The idea is to find the best match between H_j and H_{j+1} in order to achieve a more accurate stitching position in G_j. Since there are two channels of audio information (left and right) represented in H_j and H_{j+1}, a simplification is done and the matching is only done on a single channel. In this thesis the left channel in H_j and H_{j+1} is used; which makes these two intervals into one dimensional discrete-time sequences.

To match H_j and H_{j+1} a cross-correlation is used. The cross-correlation has the potential to find if one of the intervals lags behind the other. If a lag is found between H_j and H_{j+1}, the stitching position is moved to the best correlation position found in H_j. The position is then transferred to G_j as position $\overline{p_c}$.

Just stitching G_j from position $\overline{p_c}$ to the start of G_{j+1} will result in a continuous and synced audio feedback; but a risk of such hard stitching can result in unwanted 'click'-defects between the stitched audio information. This is due to that the range-data in G_j is not always perfectly matched with G_{j+1} because of unwanted light differences during a circumference scan or modulation defects of the scanning system. A solution to remove the risk of 'click'-defects during stitching of the audio information is to create a cross-fading interval between the audio signals from the $\overline{p_c}$ position in G_j and the first position in G_{j+1}.

The cross-fading is done by first initiating a small cross-fading interval size. In this thesis we use 100 samples. Then the amount of samples in G_{j+1} are cut from the start to $(\overline{p_c} + 100)$ in both the left and right audio information. The last 100 samples, in both left and right audio information, of the now smaller G_j is then multiplied with a linear filter that goes from 1 to 0.5; this results in a fade out. The same procedure is done on G_{j+1} but instead on the first 100 samples and with a linear filter that goes from 0.5 to 1, fade in. The audio information in the cross-fade intervals of both G_j and G_{j+1} are calculated together additively and the rest of the sequence in G_{j+1} is added to the end of the G_j sequence; creating one continuous audio signal from two signals for each channel (left and right).

The stitching process continuous until it has been performed on every groove of the dataset; this results in that G_j constantly grows after each new groove while G_{j+1} will always be the size of the next groove.

After the stitching process, two long tracks are created from the whole scanned dataset, for the left A_L and right A_R channels from their corresponding wall positions.

4.6.1 Mono or Stereo

As mentioned in 4.6 there is now two long audio tracks of the scanned data; left A_L and right A_R channel. Two types of outputs can be done from here on; either mono or stereo. As the scanned record is a 78 rpm record, see section 3.6, the information in the disc is mono. If stereo is desired, this step is not necessary. To create a mono signal from the two channels A_L and A_R, the mean between each sample is computed,

$$A_{mono} = \frac{A_L + A_R}{2} \tag{4.22}$$

4.6.2 Calculating the Sample Rate

Before listening to the decoded track, the sampling rate needs to be calculated. To calculate the achieved sampling rate from the scans. Three variables needs to be known:

- $RPS_{turntable}$ - The rotation speed of the turntable.
- RPS_{record} - The coded rotation speed in the record.
- C_f - The cameras scan frequency.

The following equation can then be solved for the sampling rate *fs*,

$$fs = \frac{C_f \cdot RPS_{record}}{RPS_{turntable}} \tag{4.23}$$

Where *fs* is the sample rate of the scanned and decoded audio.

5
Result

The result of this thesis is to present a comparison of the decoded audio signal against ground truths. The ground truths used against the decoded 'Doris Day'-record (see section 3.6) is a digitally enhanced/mastered edition and a cell phone recording while playing the exactly decoded record on a non-electric gramophone player. All comparisons are done on a single scan; approximately the first 36 seconds of the scanned record.

The evaluation procedure used in this thesis was first proposed in [Tian and Barron, 2006]. It is done through the use of a technique called 'shape envelopment' on the ground truth and the decoded information. The idea of shape envelopment is to create a phase-mask of the structure in each of the compared signals to achieve the signals structure comparison instead of a frequency comparison, as we are more interested in achieving a decoded sound that has a fairly similar structure, or information, compared to the ground truth instead of a comparison done on each frequency compared to the ground truth frequency. To create a shape of each signal, we use Hilbert transform. The absolute value of the analytic signal, created by using Hilbert transform, is the enveloped shape of the original signal. These shapes are then compared with Pearson's product-moment correlation. The Pearson's correlation coefficient is then used as an indication of the correlation between the two signals. In order to have a comparable correlation between the two sounds, both tracks are synced with each other using cross-correlation. The ground truth is also resampled to the sample rate (as it had almost double the sampling rate achieved in this decoding) of the decoded audio.

Some subjective comparison of the decoded audio information is also presented in the results.

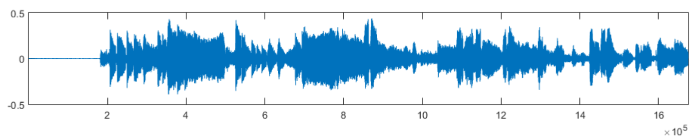
5.1 Ground truth comparison

As mentioned in the introduction of this chapter, two ground truth sets were used to evaluate the decoded audio signal. The first one is a digitally mastered version of the same song which is found on the scanned record. The second one is a cell phone recorded version of the scanned disc, using a non-electric gramophone turntable.

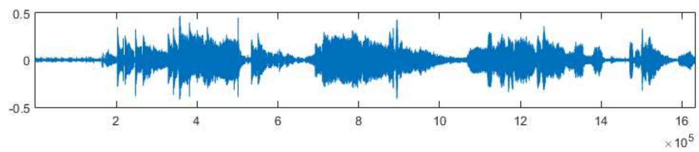
To better understand the evaluation process a comparison is done between the two ground truths, with the same technique used for comparison between a ground truth and the decoded audio. A side-by-side comparison of both ground truth signals can be seen in figure 5.1a and 5.1b.

- **Subjective comparison** - The digitally mastered version is flawless and reproduce the song without any noise present. The cell phone recorded version contains distinctive record background noise and clicks but reproduce the sound in the record clearly. Both the lyrics and instruments can be heard clearly in both sets, even if the signals look rather different from each other (figure 5.1a and 5.1b). The cell phone recorded version sounds a little slow compared to the mastered version, this can be due to the condition of the gramophone player used for playback.

- **Analytic comparison** - As mentioned, an analytic comparison was done between the two signals in the same way they are compared against the decoded audio. The maximum achieved correlation coefficient was 0.4257. The low correlation coefficient is probably due to that the cell phone recorded signal is constantly lagging behind due to that the record probably was played too slowly or that the rotation speed of the turntable was not constant in the gramophone turntable. A comparison done from the cell phone recorded audio and the decoded audio can therefore be biased due to a velocity defect.



(a) The signal of the mastered and enhanced signal of 'Doris Day - Again'.



(b) The cell phone recorded signal of 'Doris Day - Again'.

Figure 5.1: *A side-by-side comparison between the two ground truth signals.*

5.2 Decoded Sound from Record - Reversed Ordinary Setup

Depending on the chosen setup during the scans (see section 3.4), the decoded audio inherits different hearable characteristics.

Using the reversed ordinary setup (figure 3.10) results in a lower resolution of the records groove depth; which in turn increases the depth of each pixel in the scanned data. In theory, the reverse ordinary should return a better scanning result than the ordinary setup, as the scan is performed vertically over the groove; in comparison to the ordinary setup where the camera is angled and registers maximums and minimums in the grooves at an angled position.

Using this setup results in a periodic hissing (almost a whistle) sound during a whole decoded scan. The origin of this defect is not known, but as it is not found in a scan done with the ordinary setup; it is plausible that it has something to do with the vertical modulation and the depth resolution or it could be coincidence as the scans, with the different setups, where performed at different occasions.

The result of the scanned and decoded record using a reversed ordinary setup can be seen in figure 5.2 and 5.3. The following scan measurements were used for this result:

- The RPS of the turntable was 0.025.
- The camera scanned in 500 Hz.
- Which results in a sample rate of 26000 Hz.

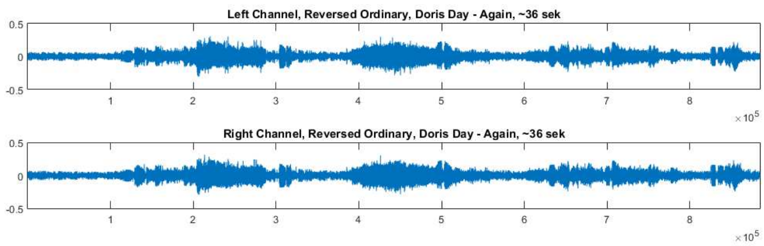


Figure 5.2: *The decoded audio from a single reversed ordinary scan presented in left and right channel.*

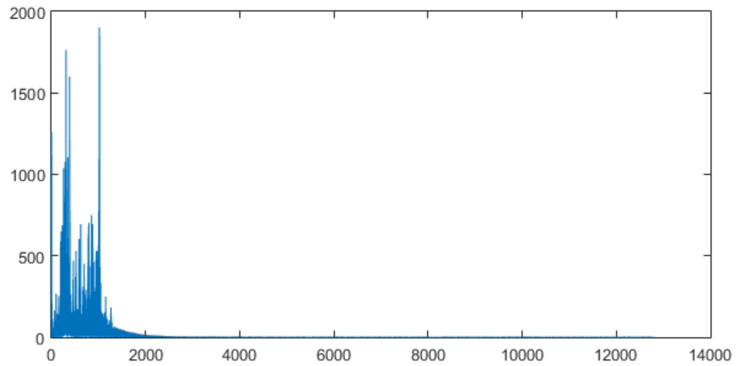


Figure 5.3: *The Fourier spectrum of the decoded audio in figure 5.2. The hissing sound appears between 500-1000 Hz in the spectra.*

5.3 Decoded Sound from Record - Ordinary Setup

Using the ordinary setup, figure 3.9, reduces the resolution in vertical modulation of the record and its effect on the decoded sound. This setup is preferable if the record is in mono. In this setup; the whistling/hissing heard in reversed ordinary, explained in section 5.2, is absent.

The result of the scanned and decoded record using an ordinary setup can be seen in figure 5.4 and 5.5. The following scan measurements were used for this result:

- The RPS of the turntable was 0.1.
- The camera scanned in 2000 Hz.
- Which results in a sample rate of 26000 Hz.

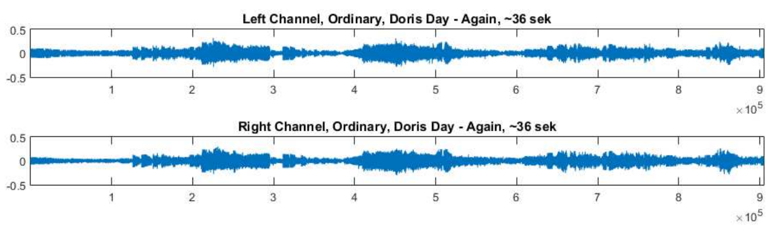


Figure 5.4: *The decoded audio from a single ordinary scan presented in left and right channel.*

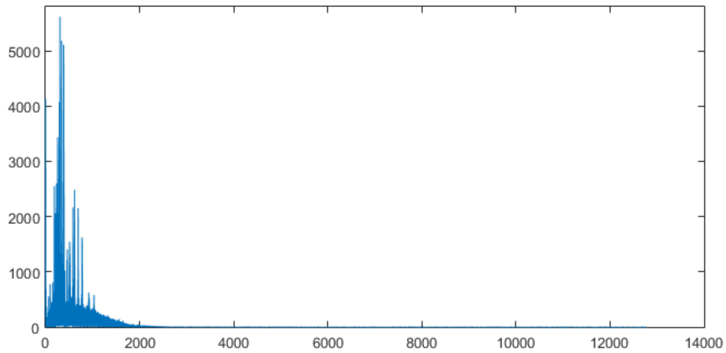


Figure 5.5: *The Fourier spectrum of the decoded audio in figure 5.4.*

5.4 Comparison to Digitally Mastered Audio

A comparison between the decoded audio is done against a digitally mastered/enhanced version of the same track. The mastered version is assumed to be a 'perfect' ground truth as it has a higher sound quality than the version recorded from a gramophone, see section 5.5. The mono-signal and Fourier spectrum of the mastered sound can be seen in figure 5.6.

A side-by-side comparison of the digitally mastered to above decoded tracks can be seen in figure 5.7.

In the the table 5.1, a comparison of using different setups and parameters in the scans against the ground truth are presented.

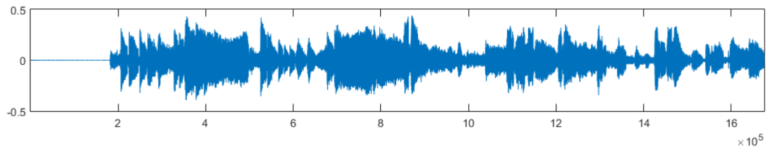
Table 5.1: *Comparison - Digitally mastered/enhanced edition to scans of 'Doris Day - Again'.*

Setup	Calc. fs	Acctual fs	Turntable RPS	Camera F.	Correlation Coeff.
R. Ordinary	3250	4168 Hz	0.200	500 Hz	0.2646
R. Ordinary	5200	5223 Hz	0.125	500 Hz	0.2490
R. Ordinary	6500	6542 Hz	0.100	500 Hz	0.3689
R. Ordinary	13000	13085 Hz	0.050	500 Hz	0.4267
R. Ordinary	26000	25617 Hz	0.025	500 Hz	0.7371
Ordinary	26000	25616 Hz	0.100	2000 Hz	0.7632

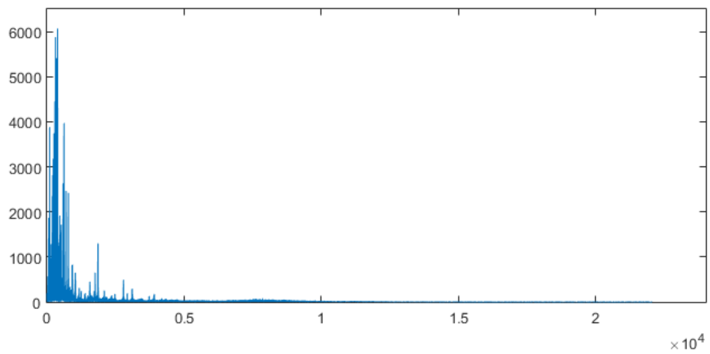
Where the calculated sampling rate fs can be approximated with the following formula,

$$fs = \frac{C_f \cdot \frac{RPM_{record}}{60}}{RPS_{turntable}} \tag{5.1}$$

Where C_f is the camera scan frequency, RPM_{record} is the scanned records RPM-format and $RPS_{turntable}$ is the RPS of the turntable during the scan.



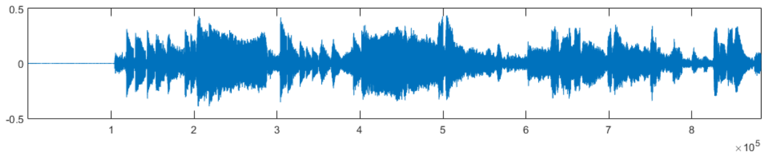
(a) *The audio signal from the mastered/enhanced version of Doris Day - Again (first ~36 sec.).*



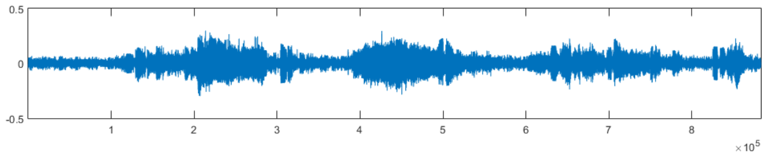
(b) *The Fourier spectra from the mastered/enhanced version of Doris Day - Again (first ~36 sec.).*

Figure 5.6: *Plots of the digitally mastered audio.*

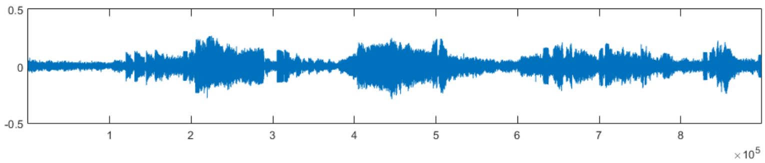
The actual sampling rate can be measured by doing tests in a large interval around the calculated sampling and finding the largest correlation coefficient and the sampling rate it was found on. This is done to counteract if the turntable is not rotating the disc at a constant speed during the whole scan and resulted in that the decoded sound was synchronized with the compared audio.



(a) The digitally mastered audio in mono.



(b) A decoded audio in mono using reversed ordinary setup.



(c) A decoded audio in mono using ordinary setup.

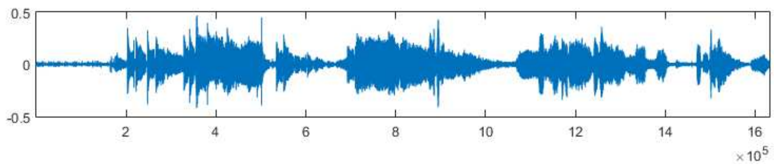
Figure 5.7: *A side-by-side comparison of decoded sound to the digitally mastered edition.*

5.5 Comparison to Cell Phone Recorded Audio from Gramophone Player

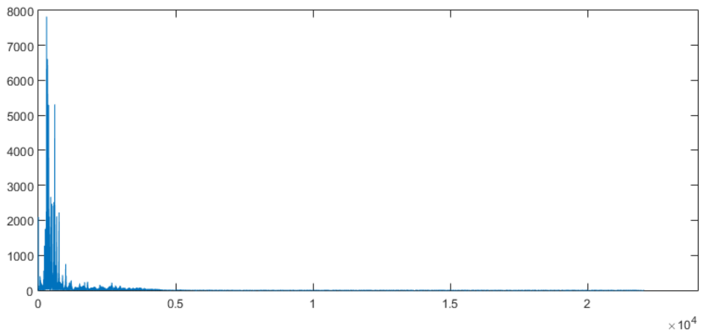
A comparison between the decoded audio is also done against a digitally cell phone recorded audio. The cell phone recording is done on the same record that is scanned through a non-electric gramophone player. In this comparison, the cell phone recording is assumed to be the ground truth. The mono-signal of the recording and its Fourier spectrum can be seen in figure 5.8.

A side-by-side comparison of the cell phone recording and the decoded signals, from the sections above, can be seen in figure 5.9.

In the the table 5.2, a comparison of using different setups and parameters in the scans against the ground truth are presented.



(a) *The audio signal from the cell phone recorded version of Doris Day - Again (first ~36 sec.).*

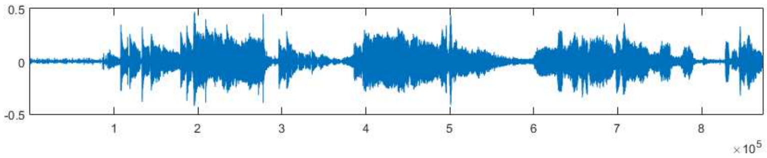


(b) *The Fourier spectra from the cell phone recorded version of Doris Day - Again (first ~36 sec.).*

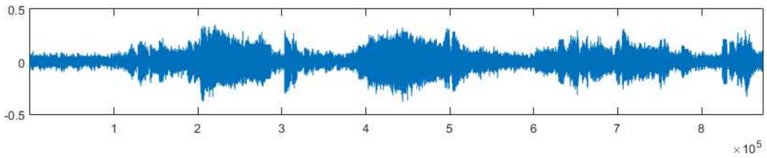
Figure 5.8: *Plots of the cell phone recorded audio.*

Table 5.2: *Comparison - Cell phone recorded version from gramophone player to scans of 'Doris Day - Again'.*

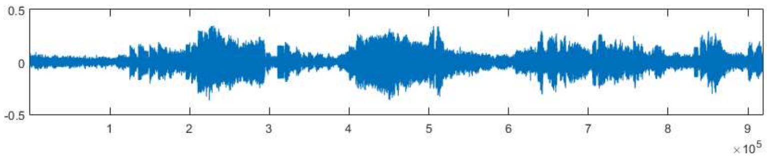
Setup	Calc. *fs*	Acctual *fs*	Turntable RPS	Camera F.	Correlation Coeff.
R. Ordinary	3250	2754 Hz	0.200	500 Hz	0.1404
R. Ordinary	5200	5529 Hz	0.125	500 Hz	0.2383
R. Ordinary	6500	6360 Hz	0.100	500 Hz	0.3366
R. Ordinary	13000	12790 Hz	0.050	500 Hz	0.4075
R. Ordinary	26000	25051 Hz	0.025	500 Hz	0.4416
Ordinary	26000	25013 Hz	0.100	2000 Hz	0.4604



(a) The cell phone recorded audio in mono.



(b) A decoded audio in mono using reversed ordinary setup.



(c) A decoded audio in mono using ordinary setup.

Figure 5.9: *A side-by-side comparison of decoded sound to the cell phone recorded version.*

6

Discussion

This chapter discusses benefits and limitations of the method mentioned in this thesis, a deeper noise analysis, a possible real-time approach and applicable future improvements. The discussion also mentions and analyses faults on the system and their influences on the result presented in this thesis.

6.1 Disc Modulation

A method to handle macroscopic vertical modulation in the scanned range data is presented in section 4.5.2. A similar method was never implemented for the global horizontal and vertical modulation due to time limitations of this thesis. But as described in both [Stotzer, 2006] and [Fadeyev and Haber, 2002], it is solvable. The defects goes under the name as 'wow-defect' or 'pitch-variation-defect'. The large horizontal and vertical modulations can be seen in figure 6.1 and smaller modulations in figure 6.2.

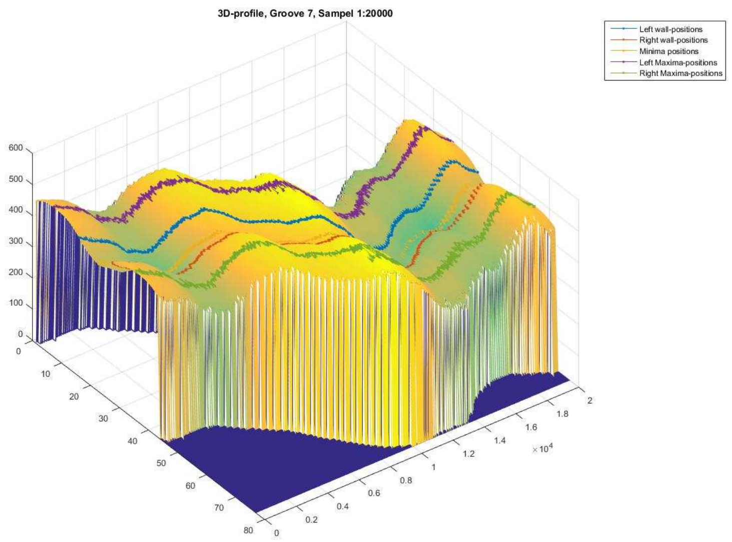


Figure 6.1: *A 3D-mesh of a segmented groove. Both the vertical and horizontal modulation is visible.*

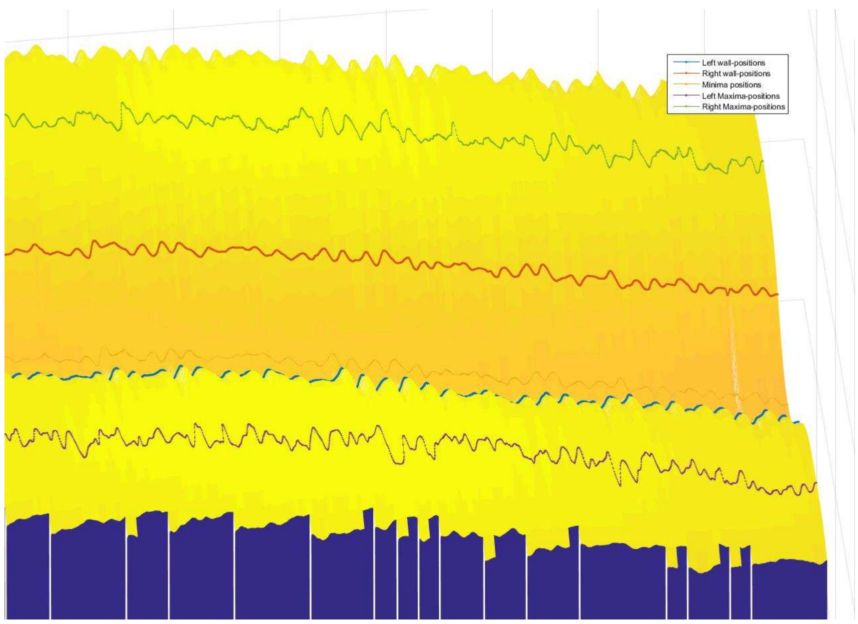


Figure 6.2: *A closer look at the 3D-mesh in figure 6.1 shows frequent vertical modulation. This is assumed to be the frequency of the turntable engine.*

6.2 Noise in the Beginning

As can be seen in figure 5.7 and 5.8, the beginning of each decoded scan have a lot of noise. This noise is also assumed to be continuing through the whole decoded signal.

Listening and comparing this noise in the decoded sound to the beginning of the cell phone recorded sound; a more faint but similar noise-periodicity is heard. The noise is assumed to be irregularities in the record or the horizontal modulation. It can also be assumed that the noise appears due to the vertical and horizontal modulation which is larger when closer to the edge of the record, illustrated in figure 2.8.

The noise found in the beginning potentially alters the comparison result negatively, presented in chapter 5.

6.3 Noise Analysis

Listening to the decoded sound, following noise/defects can be heard:

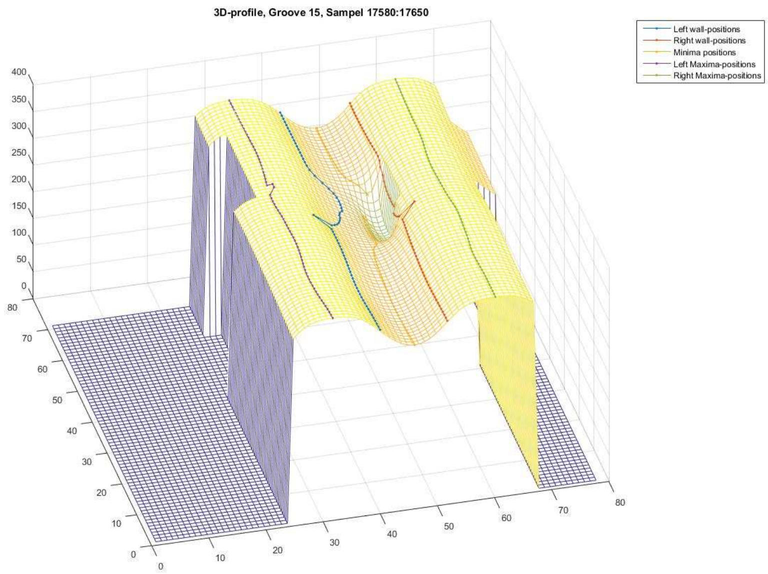
- **Clicks/Ticks** - Although these defects are handled using a median filter on the dataset and a maximum suppression they still seem to appear. This is probably due to the thresholding technique used in the maximum suppression; as it seems some of the 'ticks' gets through. Further development of this technique and using a histogram in segments to create thresholds can probably lower this defect more.

  It can also be assumed that groove shadowing is the cause for high frequency 'click'-defects as they can alter the grooves minimum a lot in few samples. The groove shadowing defect is hard to detect but can be reduced by using a median filter in the horizontal wall position before creating the extraction line. A shadowing defect can be seen in figure 6.3

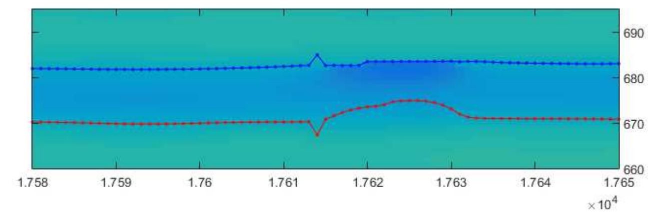
- **Periodic Whistling/Hissing** - Due to the periodicity, this noise is assumed to be created from the horizontal and vertical modulation of the spinning record during scans. The whistling/hissing appears in the scans done with the reversed ordinary setup and shows that the defect is potentially amplified with a lower vertical resolution.

- **Regular High Frequency Noise / Static Background Noise** - An additive noise is heard during the whole decoded record. The origin of this noise is not known but can be countered with a low-pass filter after analyzing the Fourier Spectrum. Tests done on a scanned record while its not spinning shows that there is a hearable noise present. It can be assumed that this noise is generally created by the hardware setup and its sensibility to vibrations and light.

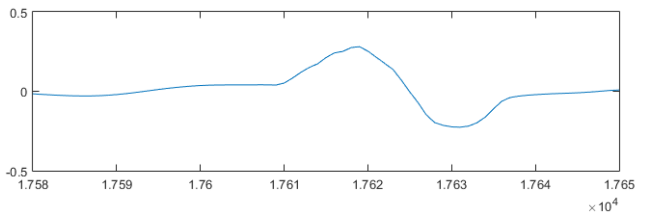
- **Irregular High Frequency Noise** - In an irregular pattern, an additive high frequency noise can heard along some samples in the decoded sound. This



(a) A 3D-mesh of a found click defect, groove shadowing.



(b) The same click as above, seen on the dataset and its corresponding horizontal wall-positions.



(c) The click after decoding to a signal.

Figure 6.3: *A click defect created from groove shadowing.*

noise tend to appear in areas of the audio where high frequency music appear. It is assumed that this noise appear withing high frequency areas during a steep horizontal modulation creating unusually high frequencies of the original signal. These irregular high frequency noise defects can be removed with a low-pass filter.

6.4 Sample Rate Artifacts

As can be seen in table 5.2, the correlation coefficient with the cell phone recordings are a lot lower compared to the mastered comparison, table 5.1. This is assumed to be directly linked to the condition of the non-electric gramophone. The gramophone player is old and the wind-up feather controlling the turntable is well-used. This in turn can create a non-constant rotation speed of the turntable making the comparison 'lacking'. If the turntable have a non-constant rotation speed, it creates sample rate artifacts/defects.

Sample rate artifacts also appear in the electric turntable (the one used for the scans), but were smaller. As can be seen in table 5.1 and 5.2.

To find the best, or most correct sample rate, the comparisons are looped in an interval around the calculated sample rate, section 4.6.2. The best sample rate, the one with the highest correlation coefficient is presented in the tables 5.2 and 5.1.

6.5 RIIA Playback Curves - Equalization

A post-processing method was introduced in the early stages of this thesis in order to increase the quality of the decoded sound, equalization using RIIA Playback Curves[1] in Audacity[2].

Using RIIA Playback Curves on the decoded sound increased the bass and treble frequencies. Although this was an interesting method, the output sounded more artificial than authentic and needs further testing for a future implementation.

6.6 Real-Time Approach

Reaching a faster than real-time decoding goal is possible for the methods presented in this thesis, if the physical scan time is not considered. That is; with scan time, it is currently impossible or hard to reach faster than real-time if the prototype scanning process is not automatized mechanically and improved.

To reach a fast than real-time goal, some of the methods ('bottlenecks') presented in this report needs to be modified. The finer extrema process is very time consuming, removing it and reducing the quality of the decoded sound, is an option for faster decoding.

Removing the anisotropic diffusion is also an option to reach a quicker decoding, but pushes down the robustness of the system with the intersection line, see

[1] `http://wiki.audacityteam.org/wiki/78rpm_playback_curves`

[2] `http://www.audacityteam.org/`

section 4.5.2. Optimizing the amount of needed anisotropic diffusion iterations can also speed up the system.

Removing the segmentation process and having a direct parallelized tracking and decoding from the first found minimum-positions in every groove is also a solution to increase the performance of the system. Another approach to increase performance could be to look for maximum positions in a constant groove-step and create an extraction line between many (or all) grooves at the same time.

The methods presented in this thesis are parallelizable; e.g. decoding per groove or tracking per sample. Further utilizing this on the CPU or GPU in a faster language e.g. C or C++, has the potential to greatly increase the systems performance. This thesis is done in Matlab using CPU parallelization (*parfor*), on each groove, as often as possible and reached some real-time results. The real-time optimized approach was, halfway through this thesis, abandoned as further testing was needed on the decoded and tracked information of the grooves; creating more slow/complex but detailed approaches. The quickest run, with an recognizable decoded audio, achieved a tracking and decoding speed of 1.2683 $s_{scanned}/s_{real}$ in Matlab. This fast tracking and decoding speed should be taken with a grain of salt, as it was obtained in the early stages of this thesis.

The time needed for each scan in the scanning process is limited to the amount of light in the scene. The Ranger-camera, which was used in this thesis, could reach a scanning speed of 4-5 KHz with favorable light conditions. The time to complete each scan is always affected by the speed of the camera as the rotation speed of the turntable will be adjusted to the scan speed and the desired sampling rate. As faster cameras are always in development, it is safe to assume that the time to scan each disc can be reduced by using faster cameras in the future.

6.7 Future Work

- More tests should be conducted to determine the best setup using a laser triangulation technique. As $33\frac{1}{3}$ rpm discs were never scanned or decoded, its too early to assume that a higher vertical resolution is necessary to decode stereo sound.

- Finding and obtaining a better lens for this type of task could result in better, less noisy, scans. Especially the lower scan quality found in the edges of the lens. A lens with a higher magnification would also make the $33\frac{1}{3}$ rpm disc scans possible.

- Methods to analyze/find and remove groove shadowing in the pre-processing stage is necessary. It is possible that these shadowing-defects appear more often than previously assumed and create many of the 'clicks'/'ticks'-defects found in the decoded sound.

- More tests needs to be done on the used turntable. Its possible that the rotation speed has a significant effect on the vertical modulation of the disc. That is; analyzing the modulations depending on the speed of the turntable.

- A comparison between the ground truth and the decoded audio using the Archimedes' spiral simulation can be implemented to further understand the modulation defect of the record.

- Handling the modulations effectively found in vertical and horizontal modulations of the disc. Creating counter horizontal modulations and resampling the samples which are affected by the vertical modulations. Implementing the methods mentioned in [Fadeyev and Haber, 2002] and [Stotzer, 2006].

- Building a more stable hardware setup for the scanning process. E.g. using a heavy rotation mechanism as presented in [Stotzer, 2006] to reduce the effect of trembling during the scanning process. It was observed during the scans that simply stomping in the same room or touching the hardware setup resulted in 'jumps' in the scan, reducing the overall quality of the scanned data by creating noise.

7

Conclussion

In this thesis a new method of doing a contact-less optical reconstruction of phonograph records has been developed and analyzed.

Using a laser triangulation technique and a 3D-profile camera proved to be an effective method of reconstructing audio from a phonograph records. The methods presented, in this thesis, leave a lot of room for improvements and has the potential to work in a faster real-time system. Meaning that a further research and optimization into these techniques could prove useful in scanning a lot of records in a quick succession.

More research and analysis must be put into the disc modulations and its effect on the decoded sound, using laser triangulation. Not only the pitch variation and 'wow'-defect but the resampling of vertically modulating samples.

Bibliography

Dimensional standards disc phonograph records for home use. 1963. URL `http://www.aardvarkmastering.com/riaa.htm`. Cited on page 10.

Ray B. Browne and Pat Browne. *The Guide to United States Popular Culture*. The University Wisconsin Press, 2000. Cited on page 28.

ELP. Elp corporation laser turntable. 1997. URL `http://www.elpj.com`. Cited on pages 2 and 3.

Vitaliy Fadeyev and Carl Haber. Reconstruction of mechanically recorded sound by image processing. 2002. Cited on pages 3, 14, 69, and 75.

Bob Fisher. Signal reconstruction in 1-d. 2003. URL `http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/PIRODDI1/NormConv/NormConv.html`. Cited on page 55.

William K. Heine. Disc phonograph record playback by laser generated diffraction pattern. 1976. Cited on pages 2 and 3.

SICK IVP. *Reference Manual, Ranger E/D*. 2015. Cited on pages 4, 17, 18, 23, and 24.

Peter Kovesi. Matlab and octave functions for computer vision and image processing. *School of Computer Science & Software Engineering*, 2002. URL `http://www.peterkovesi.com/matlabfns/`. Cited on page 40.

P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1990. Cited on page 40.

Oliver Read. The recording and reproduction of sound, revised and enlarged second edition. *Indianapolis: Howard W. Sams & Co., Inc., chapter 2, History of Acoustical Recording*, 1952. Cited on page 9.

Ofer Springer. Digital needle - a virtual gramophone. 2002. URL `http://www.cs.huji.ac.il/~springer/DigitalNeedle/index.html`. Cited on page 3.

R. E. Stoddard. Optical turntable system with reflected spot position detection. 1989. Cited on page 2.

R. E. Stoddard and R. N. Stark. Dual beam optical turntable. 1990. Cited on page 2.

Sylvain Stotzer. Phonographic record sound extraction by image processing. 2006. Cited on pages 3, 14, 29, 36, 52, 69, and 75.

Baozhong Tian and John L. Barron. Reproduction of sound signal from gramophone records using 3d scene reconstruction. 2006. Cited on pages 3, 4, and 59.

Baozhong Tian and John L. Barron. Using computer vision technology to play gramophone records. 2011. Cited on pages 3 and 4.

Vinnova. Utveckling av beröringsfri överföring av analoga skivinspelningar till digitala ljudfiler. 2014. URL `http://www.kb.se/om/projekt/Vinnova`. Cited on page 1.