Optical Audio Reproduction for Stereo Phonograph Records by Using White-light Interferometry and Image Processing

Beinan Li

Music Technology Area
Department of Music Research
Schulich School of Music
McGill University, Montreal

Submitted February 2011

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy

Abstract

This dissertation presents an optical approach for reproducing stereo audio from the stereo disc phonograph records (LPs). Since the late nineteenth century, as one of the most influential recording technologies, the phonograph recording has enjoyed its popularity and produced numerous cylinders and discs that carry speeches, music, and all kinds of audio cultural heritage. The preservation of phonograph sound recordings is thus of world-wide concern. This research provides an alternative approach to digitizing the stereo disc phonograph records, potentially for long-term preservation, by optically acquiring the 3D disc record surface profile and extracting the audio signals from the record surface profile images by using software algorithms.

The dissertation discusses the workflow of optically reproducing stereo audio from the stereo disc phonograph records by using the white-light interferometry technique. This workflow includes the acquisition of the 3D disc record surface profile by using a commercial white-light interferometry microscope, the extraction of the record groove undulations, which encodes the stereo audio information, by using our custom image processing algorithms, and finally the reproduction of the stereo audio signal from the groove undulations through signal processing. The workflow is evaluated with a test stereo record containing standard sinusoid signals and a musical record. The quality of the optically-reproduced audio is quantitatively evaluated and compared with that of the audio digitized by a turntable.

The dissertation contains three main parts. The first include an introduction to the general background of the optical audio reproduction for the stereo disc phonograph records and the review of the phonograph recording technology, the previous efforts in optically reproducing audio from the cylinder and disc phonograph records, and the relevant optical techniques including the white-light interferometry. The second part focuses on our complete workflow for optically reproducing the stereo audio from the stereo disc phonograph records. This is followed by the evaluation of our workflow and the output audio quality. The dissertation concludes by introducing the challenges and the possible directions in the future development of our optical audio reproduction workflow.

Résumé

Cette thèse présente une nouvelle approche de reproduction optique d'enregistrements phonographiques stéréo. L'enregistrement phonographique s'est imposé, vers la fin du XIX^{ème} siècle, comme la technologie d'enregistrement de référence partout dans le monde. Il existe donc une pléthore de cylindres et autres disques où ont été gravés discours, morceaux de musique, et autres artefacts culturel sonores. La préservation de ces enregistrements sonores phonographiques est donc une préoccupation mondiale. Le présent travail de recherche propose une approche alternative de numérisation des enregistrements phonographiques stéréo en vue de leur éventuelle préservation. En effet, à partir de l'acquisition optique du profil (en trois dimensions) de la surface d'enregistrement du disque, les signaux audio peuvent être reconstruits grâce à nos algorithmes d'analyse d'images.

Cette thèse examine les étapes de la reproduction optique audio stéréo à partir d'enregistrements phonographiques sur disques stéréo en utilisant l'interférométrie en lumière blanche. Ces étapes comportent: l'acquisition du profil de la surface d'enregistrement d'un disque 3D en utilisant un microscope commercial interférométrique en lumière blanche ; l'extraction des ondulations du sillon, qui encode l'information audio stéréo en utilisant nos algorithmes de traitement d'images ; et finalement, la reproduction du signal audio stéréo depuis les ondulations du sillon par des techniques de traitement du signal. Le processus complet est évalué sur un enregistrement stéréo test comprenant des signaux sinusoïdaux et un enregistrement musical. La qualité de l'audio reproduit par voie optique est évaluée de façon quantitative et comparée avec celle de l'audio numérisé de manière « traditionnelle », à l'aide d'une platine.

Cette thèse s'articule en trois parties. La première comporte une introduction des principes nécessaires à la reproduction d'enregistrements phonographiques stéréo par voie optique. Plus précisément, les principes de la technologie d'enregistrement phonographique sont passés en revue ; l'état de l'art des efforts de reproduction optique des enregistrements phonographiques sur disques et cylindres est présenté ; et enfin, les techniques optiques pertinentes incluant l'interférométrie en lumière blanche sont décrites. La deuxième partie livre une présentation détaillée du processus de reproduction optique que nous avons développé. Dans la troisième partie, l'évaluation quantitative de

la qualité de la restitution du signal audio obtenue par notre procédé est aussi décrite. La thèse se conclue sur un bilan des défis et des directions possibles dans le futur développement de notre approche de reproduction des signaux audio par voie optique.

Acknowledgements

I would like to thank the effort and support of many people who contributed to the completion of this dissertation.

First and foremost, I must extend my deepest gratitude to my thesis advisor, Professor Ichiro Fujinaga. As a resourceful and dedicated teacher, his valuable advice, attentiveness, and inspiring general research methodologies have made my doctoral study at McGill University a remarkable experience. I must also thank the former team members of this research: Simon de Leon, Yannick Fonjallaz, and Jacquat Claude, who contributed their effort into the initial background research of this dissertation. I am grateful to my colleagues, Dr. Mathieu Bergeron, John Ashley Bourgoyne, Johanna Devaney, Andrew Hankinson, Jason Hockman, Jessica Thompson, and former colleagues, Dr. Ian Knopke, Dr. Catherine Lai, Dr. Cory McKay, Dr. Laurent Pugin, Greg Eustace, Rebecca Fiebrink, Daniel McEnnis, and Jordan B. L. Smith, who provided me assistance in various aspects during my degree program. I would like to express my special thanks to David Curtis, Cynthia Leive, Cathy Martin, Brian McMillan, Melanie Preuss, Andrew Senior, Patrick Dupuis, and other librarians and clerks in the Marvin Duchow Music Library at McGill University. During the course of my research, they generously granted me the permission to conduct the research experiments for this dissertation in the special collection room of the Music Library and kindly offered me various resources and support for my work.

I also wish to acknowledge all those individuals who have contributed to the extraordinary journey of my doctoral study. I would like to show my sincere gratitude to my teachers: Professors Philippe Depalle, Stephen McAdams, Gary P. Scavone, and Marcelo M. Wanderley in the Music Technology Area, Professors Doina Precup and Clark Verbrugge from School of Computer Science at McGill University, and Professor Douglas Eck from Université de Montréal for their guidance in my coursework and research. I am grateful to Professor Martha De Francisco who offered me the opportunity to work in the webcast program and gain experience in sound recording technology. I am thankful to my fellow graduate students in the Music Technology Area, especially Stephen Sinclair and Bertrand Scherrer, and researchers in the Centre for Interdisciplinary Research in Music Media and Technology, who supplied me with inspirations in arts and

science. Furthermore, I am thankful to the care and support that I have received from my friends in Montreal: Jianming Dai, Matthew Hassler, Clarisse Manigand, Linxi Jiang, Rui Li, Shinichi Noguchi, Menglai Shang, Xinyi Xu, Mingfei Yuan, and Henry Zhang, who have made my life at Montreal memorable.

I would like to thank the Canada Foundation for Innovation and the Daniel Langlois Foundation for their financial support for conducting the research of this dissertation.

Finally, I would not have been able to finish this dissertation without my family. I am forever indebted to the unconditional love and constant encouragement from my parents Dehua Li and Shuqing Jiao, my stepmother Guifen Li, and my stepsister Liruxin Wan. My special thanks go to my beloved fiancée Yajing Zhu.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

1.1 About this Chapter

This dissertation presents an optical approach for reproducing stereo audio from stereo disc phonograph records. In this chapter, first the general background of the dissertation is introduced; the research objectives of the dissertation are then discussed and the research contributions explained. Finally the chapter organization of the dissertation is presented.

1.2 General Background

Sound propagates in air in the form of a longitudinal wave. To create a phonograph recording of sound, the energy of the sound wave is transmitted through atmospheric pressure variations, received by a pressure sensor, and converted to mechanical motions of a cutting stylus, which cuts a modulated groove into the surface of a record. To play the recorded signal from the record, this process is reversed: A playback stylus follows the groove and its mechanical motions are converted back to atmospheric pressure variations to be perceived by ear. This is the basic principle of phonograph recording and playback technologies.

Since the late nineteenth century, the phonograph recording technology has been one of the most influential recording technologies, and numerous cylinders and discs that carry speech, music, and all kinds of audio cultural heritage have been produced. As such, the preservation of phonograph sound recordings is of concern to institutions worldwide. Digitally preserved audio recordings have advantages such as better resistance to noise and distortion. This is especially noted in the *US National Recording Preservation Act of 2000* (2000), which makes transition from analog preservation to digital preservation is indicated as an official policy decision of the American government. Although it is standard practice to digitize phonograph records by using high-quality mechanical systems such as turntables, this practice, however, cannot digitize or restore severely damaged or broken shellac records (mainly used in 78 rpm recordings) and cylinders. The required physical contact by mechanical systems may also degrade the physical condition of phonograph records. The *US National Recording Preservation Act of 2000* indicates

that studies are needed to outline methodologies and standards needed in the analog-to-digital transition.

To bypass the disadvantages of turntable-based approaches, optical audio reproduction (OAR) represents a promising alternative solution for digitally archiving phonograph records. This approach extracts the audio information cut into the record surface through optical methods without physical contact with the record. This approach is recognized as a viable alternative to traditional mechanical reproduction methods, and has received official supports from the governments. For example, the US Library of Congress has recently sponsored OAR research (Fadeyev and Haber 2003).

The concept of the OAR was conceived of and patented back in the late 1920s (Brock-Nanestad 2001; Friebus 1929). A company called Finial Technology produced the first working laser turntable and presented it at the 1989 Winter Consumer Electronics Show (Hodges 1989). The laser turntable uses a reflected laser beam to replace the mechanical stylus to reproduce sound. Since then, a number of OAR research efforts have been made based on various optical technologies (Iwai et al. 1986; Poliak et al. 1992; Nakamura 1997; Petrov et al. 1997; Fadeyev and Haber 2003; Fadeyev et al. 2005; Stotzer et al. 2004; Tian 2006; Boltryk et al. 2007; Li et al. 2007). The common procedure of these existing OAR frameworks is as follows:

1) The physical characteristics of the record surface are captured by using optical instruments (microscopes in most cases or laser) without touching the record surface.
2) The groove undulations are isolated and extracted from the captured record surface characteristics.
3) Finally the groove undulations are converted into audio.

Of these efforts, Li et al. (2007), which this dissertation is based upon, is the first to successfully apply OAR to 33rpm stereo phonograph records using 3D surface information. Compared to digitization based on the mechanical reproduction, such as the turntables, a few advantages of the above procedure can be identified:

1) OAR techniques are contactless and thus introduce no damages to the records.
2) Without having to use a stylus or other physical instruments to trace the record grooves, OAR has the capability of capturing the images of the broken pieces of a

damaged record and recovering the audio information by merging the record grooves, which mechanical systems are unable to achieve.

3) If the specification of the output digital audio specification can be controlled by configuring the image resolution and quality parameters, audio recordings can then be alternatively archived in the form of images as an alternate medium for audio archiving, broadening the spectrum of archiving solutions for audio cultural heritage.

1.3 Research Objectives

This dissertation focuses on reproducing the stereo audio from the stereo LP (long-playing) records by using a specific high-precision 3D optical surface measurement (or optical surface metrology) technique, the white-light interferometry (WLI) (Wyant 1974), without contacting the record surface. This is outlined in the three primary research objectives of this dissertation.

First, we create a semi-automatic workflow that efficiently acquires the 3D surface profile of a stereo LP record with a generic WLI microscope (Veeco 2003) and stores the result as images. Measuring the surface information of a record with a generic microscope can be time-consuming. This is because the microscope, with a lens that is usually much smaller than the record surface area, needs to scan the entire target area, make multiple measurements to obtain an audio signal of a meaningful duration. Each measurement is then saved as a three-dimensional image. This process needs to be carefully designed to be highly automated, consume the shortest possible amount of time, and achieve the best possible measurement quality. The resulting images need to be aligned with each other so that they can be consolidated to provide the complete target record surface information.

Secondly, we devise image-processing algorithms to extract stereo audio information from the record surface images and reproduce sound, while minimizing the undesired artefacts in the resulting audio such as noise and distortion. The modulated groove cut into the record surface bears the audio information of a phonograph records (detailed in Section 2.2). Efficient image-processing algorithms need to be designed to extract groove information from the acquired record surface profile. Noise and distortions can be present in the resulting images, for example, dusts and scratches can cause signal

discontinuity. Although after the audio signals are extracted from the scanned images, their quality can be enhanced by applying digital audio signal processing techniques, such techniques are not the focus of this dissertation. Instead, it is our goal to remove the noise and distortions as much as possible before the audio is reproduced, for example, to remove the artefacts caused by dusts and scratches in the image domain.

Finally, we present the results of the above surface acquisition workflow and image-processing algorithms to obtain a segment of stereo audio from a standard test LP record and a music LP record, respectively, and evaluate the reproduced audio quality. Several standard quality evaluation parameters are to be measured such as signal-to-noise ratio and total harmonic distortion. The audio results from the OAR process are compared with the audio digitized from a turntable.

1.4 Research Contributions

This dissertation represents, to our knowledge, the first approach for reproducing stereo audio signals from the stereo LPs using the WLI technique, which has not been reported before. More specific contributions include:

1) A full OAR solution for stereo LP records is given, including the acquisition of the 3D surface profile of the record using WLI, and the audio reproduction from the acquisition result using image processing.
2) The record surface acquisition in this dissertation uses no custom mechanical hardware except a generic WLI microscope (Veeco 2003) with an optional motorized X-Y stage (to automatically move the phonograph record) for the acquisition of the record surface information. Most existing OAR approaches (e.g., Stotzer et al. 2004; Tian 2006; Boltryk et al. 2007) involve the use of custom hardware to rotate the target during the acquisition so that record grooves can be unwrapped into one-dimensional signal more conveniently. The groove in our approach is tracked by using purely software-based methods instead.
3) The dissertation presents algorithms to automatically detect and remove a few common types of noise and distortion that phonograph records display (see Section 2.2.3): Image-processing algorithms are provided to remove noises introduced by dusts and surface damages, utilizing the obtained 3D surface

information; an algorithm for automatically detecting and minimizing the pitch fluctuation artefact is also presented.

1.5 Summary

The dissertation is organized into seven chapters. Chapter 2 presents a literature review, including an overview of phonograph recording technology; the physical characteristics of the record grooves and their recording principles, existing OAR research works, and major optical surface metrology techniques adopted by these works. Chapter 3 focuses on the acquisition of the 3D surface profile of the disc phonograph records by using the adopted WLI microscope. Chapters 4 and 5 present the image processing chain for reconstructing stereo audio from scanned record surface profile images in detail. Chapter 4 focuses on the groove undulation extraction and Chapter 5 the audio reproduction. Chapter 6 is devoted to the evaluation of the reconstructed audio. Finally, the conclusions and future perspectives are discussed in Chapter 7.

Chapter 2 Literature Review

2.1 About this Chapter

Optical Audio Reproduction (OAR) strives to extract the recorded audio information from the surface topology of the phonograph records. In this chapter, we first introduce the phonograph records and the phonograph recording technology, which helps understand the physical characteristics of the records and how the audio information is encoded into surface topology of the records; then several important generic optical techniques for obtaining the record surface profiles (optical surface metrology) are presented; finally the previous OAR research efforts are reviewed.

2.2 Brief History of Phonograph Recording Technology

The earliest known sound recording device, the phonautograph (Figure 2.1), was invented by Édouard-Léon Scott de Martinville in 1857 (Hoffmann and Ferstler 2005, 296). This device can record a visual image of the sound by drawing lines that represent the sound waves onto a piece of paper rolled around a cylinder, but without the capability of reproducing the sound. In 1877, Thomas Edison invented the cylinder phonograph (Figure 2.2), which was capable of both recording and replaying monaural (monophonic, or mono) audio via a piece of thin tinfoil sheet attached to a rotating cylinder (Edison 1877). Charles Summer Tainter and Alexander Graham Bell experimented from the Volta laboratory experimented with vertically cut disc recording technology in 1880s (Newville 2009). The first disc phonograph system presented to the public, the Gramophone (Figure 2.3), was produced by Emile Berliner in 1887 (Berliner 1887). It was capable of recording and playing monophonic audio. In 1931, the stereophonic phonograph recording technology was invented by Alan Blumlein (1931). Later in 1948, Columbia Records introduced the first stereo LP (long-playing) phonograph disc system to the commercial market (Hoffmann and Ferstler 2005, 305). Quadraphonic phonograph disc recordings were introduced in 1970, which provides four-channel surround sound. The stereo LP became the most popular sound recording medium prior to the introduction of the compact disc (CD) in the 1980s.

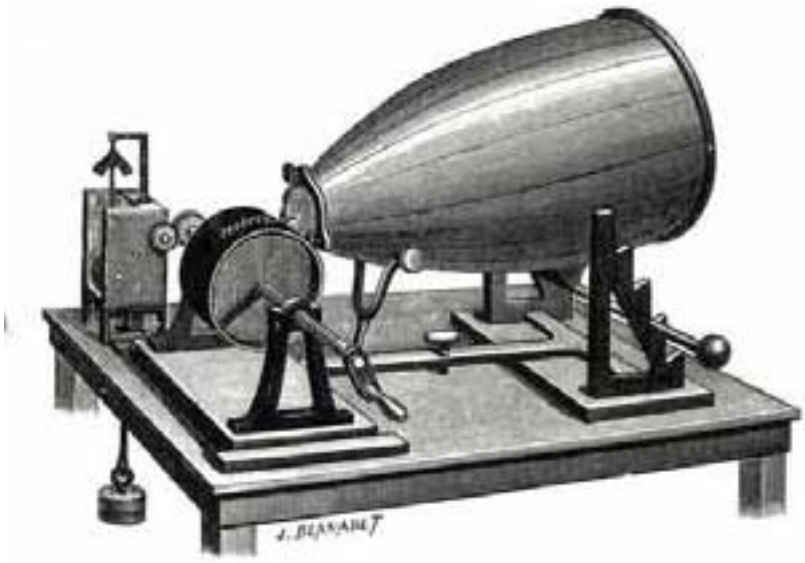


Figure 2.1: The phonautograph (image from Stotzer 2006, 14)

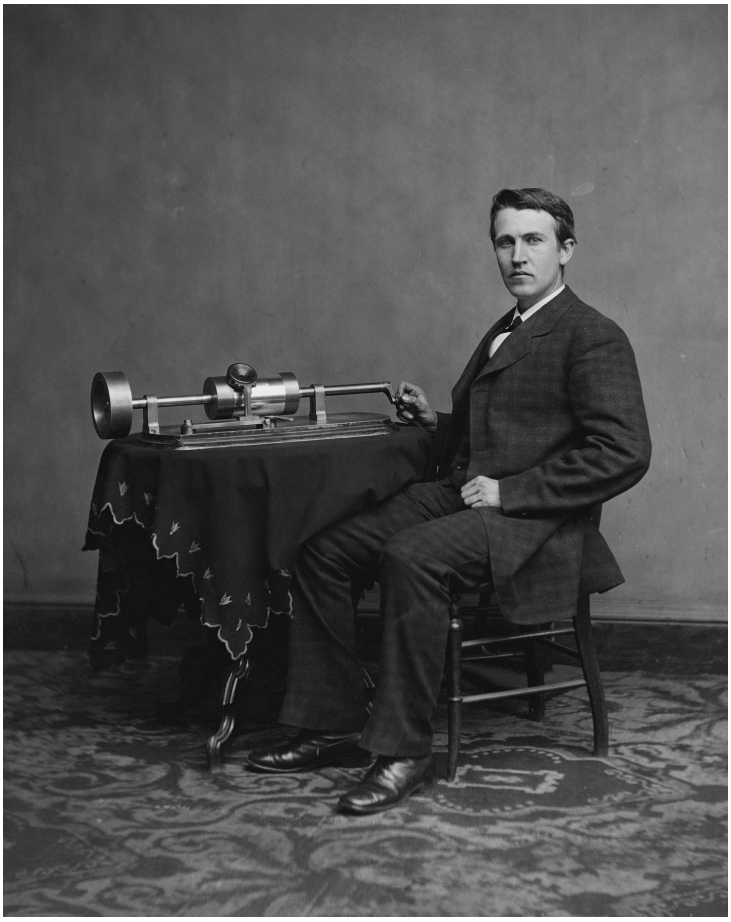


Figure 2.2: Edison and his early tinfoil cylinder phonograph (image from Wikipedia 2010)

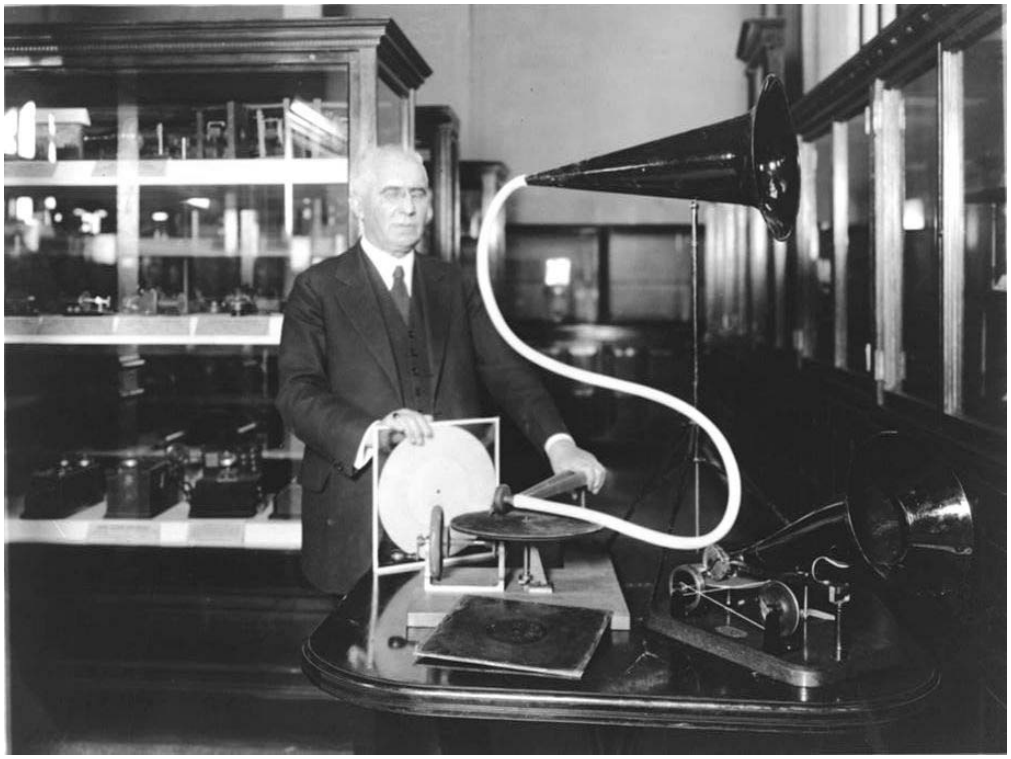


Figure 2.3: Berliner and his Gramophone (image from Library of Congress 1930)

2.3 Phonograph Recording Technology

As noted in 2.2, phonograph records include both cylinder and disc recordings. In this section the physical characteristics and the recording mechanisms of these two types of records are presented.

2.3.1 Cylinder Recordings

Although cylinder recordings are not the focus of this dissertation, we nevertheless present a brief overview. Interested readers are referred to Nascè (2009) for a detailed review of cylinder recordings.

2.3.1.1 Record Material

As shown in Figure 2.4, the surface of the earliest cylinder records was made of tinfoil; the audio information of a cylinder recording is cut into the record surface. Tinfoil is a vulnerable material that is easy to wear and tear. Audio recorded on a tinfoil cylinder became incomprehensible after a few playback cycles. In 1881, Tainter and Bell introduced the wax cylinder record, also known as Graphophone (Figure 2.5), which better resisted wear and tear and provided more playback cycles than the tinfoil cylinder.

Mass-produced cylinders made of celluloid (Figure 2.6) became available in the late 1890s. Compared to wax, celluloid has a few advantages including the better durability and better resistance to mould growth. However, it can still develop cracks and is vulnerable to shrinkage. Production of celluloid cylinder ended in late 1920s (Hoffmann and Ferstler 2005, 262).

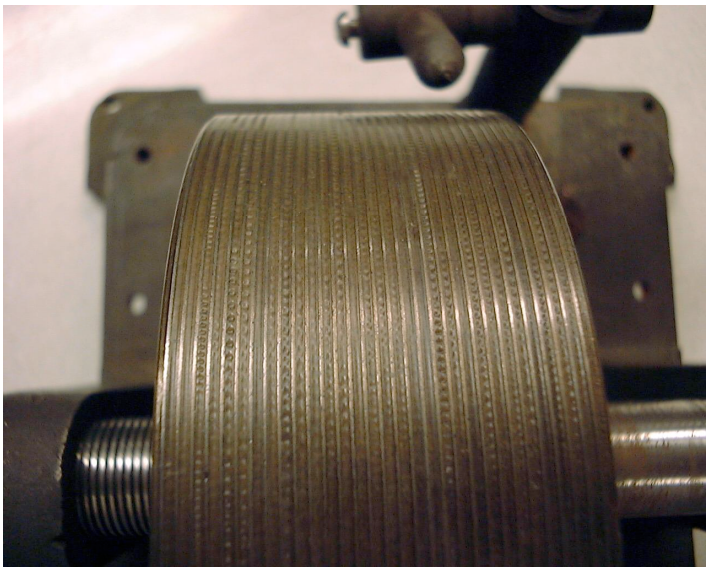


Figure 2.4: Tinfoil cylinder record mounted onto an Edison phonograph with grooves carved into its surface (image from Smithsonian National Museum of American History 2011)

Figure 2.5: Two-minute wax cylinder records, 2-inch in diameter and 4-inch long (image from Sage 2010)

Figure 2.6: An Edison Blue Amberol cylinder record made of celluloid (image from Bruderhofer 2010)

2.3.1.2 Record Manufacturing

The manufacturing method of the cylinder recordings was introduced around 1900 (Hoffmann and Ferstler 2005). The manufacture of a cylinder recording comprised these steps:

1) A soft brown wax cylinder was used to directly record the sound, which served as the master recording.
2) A gold mould was then made by using high-voltage gold electrodes to produce a gold layer over the master cylinder surface.
3) The mould was used to produce second-generation metal sub-moulds.
4) Finally the sub-moulds were used to mass-produce the derivative wax cylinders.

2.3.1.3 Recording Principle

Figure 2.7 shows an early wax cylinder recorder. The cylinder is horizontally mounted on a cartridge with its axis parallel to the ground. A horn with a diaphragm attached to its mouthpiece is used as a sound pressure sensor. The cutting stylus is coupled with the diaphragm at the end of the horn mouthpiece. The groove of a cylinder recording is cut with vertical modulation, also known as “hill-and-dale” modulation. The cutting stylus tip is initialized to penetrate into the cylinder surface to a fixed depth, which defines the unmodulated cut. During recording, the cylinder is rotated at a constant angular speed and

the stylus moves up and down, perpendicular to the record surface. The amount of the vertical stylus movement corresponds to the diaphragm vibration that represents the sensed sound pressure. As the recording continues, the cutting stylus advances horizontally along the cylinder axis while the cylinder rotates. In this manner, a mono analog signal is continuously cut into the circumvolution of the cylinder surface in the form of a spiralling groove.

The recording method is an acoustic one, in other words, the captured sound pressure is directly used for groove modulation during the recording and the stylus movement is directly used as the playback signal, without electrical transducer or amplification being involved.

Figure 2.7: Early wax cylinder recording apparatus (image from Nascè 2009, 11)

2.3.1.4 Groove Characteristics

The shape of the groove in wax cylinders is determined by the shape of the cutting stylus and how deep the stylus cuts into the surface. The cross-section of the cylinder groove is roughly “U”-shaped. The groove depth is determined by the modulation depth, which had to be large enough to overcome the surface noise (discussed in 2.2.4) because no amplification was used in the acoustic recording method. The depths of the cylinder grooves are proportional to the modulation depths. The maximum modulation depth ranges between 6.8 to 102µm (Nascè 2009, 17).

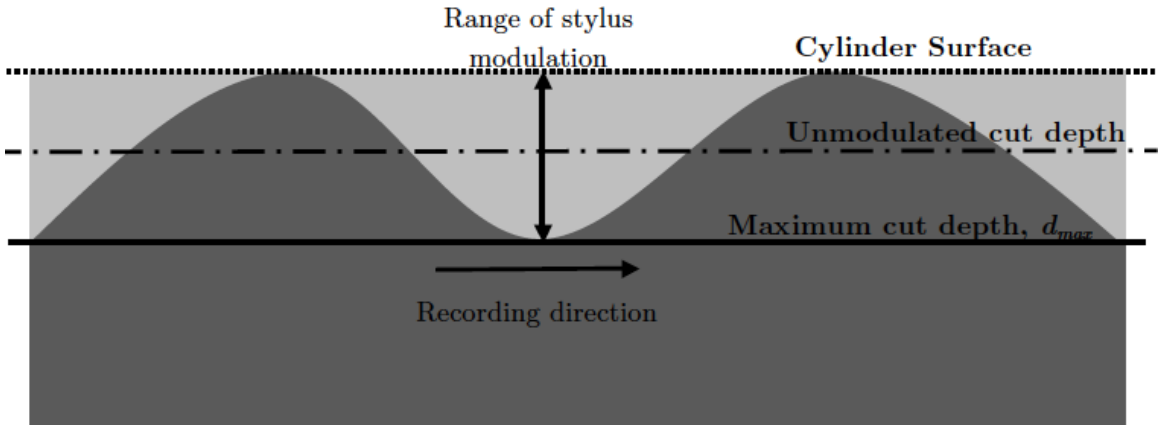


Figure 2.8: The longitudinal section of the cylinder groove with the maximum vertical modulation range (image from Nascè 2009, 17)

2.3.1.5 Audio Duration

The recording duration of a cylinder record is determined by the size, the rotation speed, and the groove density (or pitch) of the cylinder. For example, a celluloid cylinder of the length 102mm, with a rotation speed of 160rpm (revolutions-per-minute), and a pitch range of 39–79tpcm (tracks-per-cm), provides 2–4minutes of audio (Hoffmann, F., and H. Ferstler. 2005, 258–63). The audio duration of a record can be increased by using a longer cylinder, denser grooves, or a lower cylinder rotation speed. These three properties vary depending on the type of the cylinders. The detailed record geometry and duration statistics of the cylinder recordings are listed in 2.3.2.8.

2.3.1.6 Frequency Response

Cylinder recordings have characteristic frequency responses, which is largely determined by the horn-diaphragm component used in the acoustic recording. The small diaphragm is an inefficient transducer to receive the surrounding sound pressure and convert it to the stylus movement, and therefore the horn, with its smoothly decreasing cross-sectional area, is used to increase the sound transmission for the recording and playback. In technical terms, the horn acts as an impedance adapter to match the low impedance of the surrounding air to the high impedance of the diaphragm. The special impedance introduced by the horn-diaphragm component thus introduces a filter to the audio reproduction chain. As a consequence, the audio signals reproduced by cylinder recordings are typically band limited. The horn-diaphragm interaction determines the low frequency band limit of cylinder recordings while the mass of the diaphragm determines

the high frequency band limit (Fadeyev et al 2005). The pass band is usually between 150Hz and 5kHz (Nascè 2009, 18).

2.3.1.7 Record Geometry

The geometric characteristics of the tinfoil, wax, and celluloid cylinder are shown in Table 2.1.

Material type	Tin-foil cylinders	Wax cylinders	Celluloid cylinders
Production period	1877–1880	1880s–1911	1900–1929
Cylinder diameter (cm)	10–18	3–13	6
Cylinder length (cm)	5–10	10–22	10
Groove density (tpcm)	4–6	39–157	39–79
Rotation speed (rpm)	60	80–185	160
Audio duration (mins)	< 1	2–8	2–4
Total groove length (m)	11–13	45–570	72–144
Groove surface area (m^2)	~0.03	0.012–0.043	0.0183

Table 2.1: Geometric characteristics of cylinder recordings (From Hoffmann, F., and H. Ferstler. 2005, 258–63; Nascè 2009, 13)

2.3.2 The Disc Recordings

Disc phonograph recordings became the dominant recording type after the cylinder record production ceased in the late 1920s (Hoffmann and Ferstler 2005, 262). Figure 2.9 and Figure 2.10 show the typical disc phonograph record and the turntable, respectively. Unlike the cylinder recordings, which are exclusively monophonic, disc recordings may be mono, stereo, or quadraphonic. They share much of characteristics in the recording, playback, and manufacturing processes but differ primarily in the groove cutting methods.

Figure 2.9: A typical disc phonograph record (image from Wikipedia 2011)

Figure 2.10: A disc turntable atop a receiver (image from Wikipedia 2011)

2.3.2.1 Record Material

Three major materials have been used to produce most disc phonograph recordings, namely the shellac, the acetate, and the polyvinyl chloride (PVC or vinyl). Stotzer (2006) presents a good review of this topic, summarized below.

Shellac the earliest disc record material. It is a compound containing an organic material extracted from the secretions of the insect Coccus Lacca, fillers, binder, lubricant, colorants, etc. Although the combination of these ingredients were expected to provide optimal audio quality and facilitate manufacturing process, the organic material

in the compound was found to promote fungus growth, which can cause the audio quality of the records to degrade over time. Shellac was mainly used in pre-1950's record production.

Acetate records are made of glass or aluminum base, covered with a layer of the mixture of the cellulose nitrate lacquer and softening agents, making the records easy to cut. The colour of the lacquer records is usually black, although other colours may be found such as green, yellow, or even transparent. As mentioned in the introduction of the cylinder recordings, record surface roughness accounts for the surface noise in the audio. Compared to shellac records, acetate records provided better audio quality because of their finer graininess. However, the acetate records are the most fragile among the three major materials. Acetate records were mainly produced between 1930 and 1955, and also in the 1970s as instantaneous records (see next section) (Stotzer 2006, 17).

Since the 1940s, vinyl has gradually replaced acetate and shellac, and became the dominant material in disc record manufacturing by the mid-1950s. Vinyl records are made of a compound of PVC, stabilizers, fillers, colorant, and additives. The PVC also provides finer graininess than shellac and thus vinyl records produce less surface noise than the shellac ones. In addition, vinyl is the most stable material among the three major ones since it is resistant to crackling, fungus attack, and high humidity levels, all of which degrade audio quality. Vinyl records are also difficult to break.

It is worth mentioning that besides the above three major materials, wax, celluloid, and aluminum discs were all once used, although not as widely adopted as the above three major materials (Stauderman 2003)

2.3.2.2 Record Manufacturing

Two kinds of recording process are available: direct-cut and pressed records. Direct-cut discs are used for recording events such as speeches. Each direct cut record is available as a single copy of the recorded content and can be played back immediately after the recording is finished. Direct-cut records may also be referred to as the instantaneous records.

Pressed records are used for mass production. They are manufactured according to the following steps (Stotzer 2006):

1) A direct record with a lacquer coating is cut during the recording.

2) A metal negative master is made for the record by electroforming.
3) A positive master is made from the negative master to serve as the mold.
4) The pressing masters, or stampers, are made out of the mold.

The stampers are used to press the shellac or vinyl records.

2.3.2.3 Recording Principles

2.3.2.3.1 Recording Principles of Mono Discs

Berliner's Gramophone was the first disc phonograph recording and playback system for mono audio (Hoffmann and Ferstler 2005, 450). Berliner introduced the Gramophone partly for "overcoming" a pitfall of the cylinder recording technology. Vertical cutting could lead to signal loss when the recording stylus rises above the record surface. "… when the diaphragm moves outwardly, the stylus will leave the record surface entirely, so that part of each vibration will not be recorded at all. This is more particularly the case when loud sounds are recorded, and it manifests itself in the reproduction, which then yields quite unintelligible sounds. It is the object of my invention to overcome these difficulties …" (Berliner 1887, 2). To overcome this problem, Gramophone used a flat disc as the recording medium. A cutting stylus, with its tip remaining below the disc surface, then moved horizontally to cut grooves into the disc surface. This method is called the lateral modulation, or "side-to-side" modulation. During the recording, the disc is placed horizontally on a disc-cutting lathe and rotated at a constant angular speed. The cutting stylus, mounted on the cutting head, is gradually shifted from the outer rim of the disc to its center whenever it finishes cutting a revolution of groove. This produces a continuous spiralling groove in the disc surface surrounding the disc center. Both sides of the disc records can be used for recording. Figure 2.11 shows a disc-cutting lathe for stereo disc records, which shares many characteristics with the mono disc-cutting lathe. During the playback, the disc is rotated at the same speed as it was cut on a turntable, the playback stylus, mounted on a pivoted pickup arm (or tone-arm), slowly shifts to the disc center with its tip staying inside the groove. The movements of the stylus inside the grooves are then decoded into the audio signals.

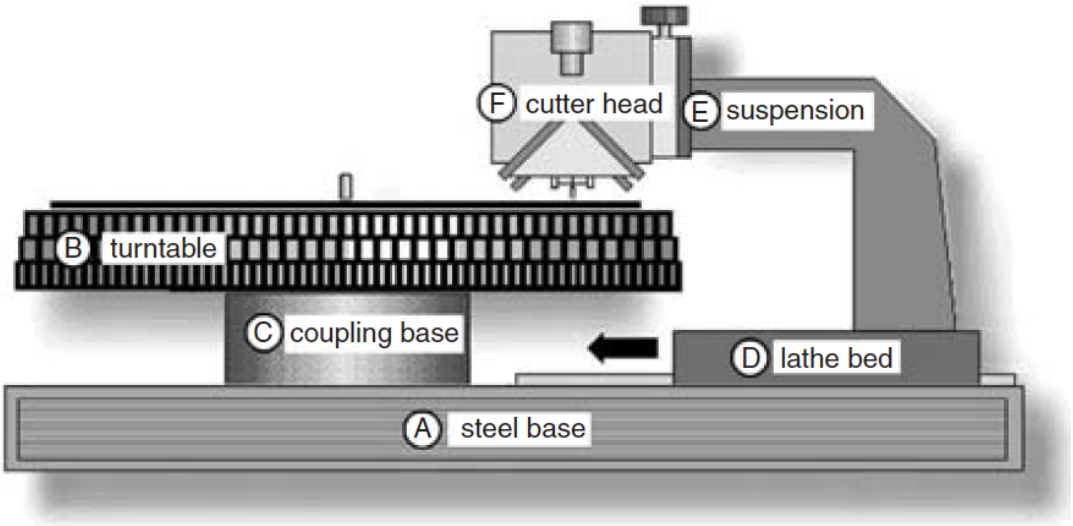


Figure 2.11: A stereo disc-cutting lathe (image from Huber and Runstein 2005, 573)

Early Gramophone recording devices used the horn-diaphragm-based acoustic recording method similar to that of the cylinder recordings. Following the invention of the microphone and the amplifier in 1925, the electrical recording method replaced the acoustic method (Hoffmann and Ferstler 2005, 298). The sound pressure sensed by the microphone is first converted to an electric signal. The electric signal is then used to control the stylus movement. The electric amplification is used to control the groove modulation amplitude during the recording, and the volume of the reproduced sound during the playback.

2.3.2.3.2 Recording Principles of Stereo Discs

To record for a stereo disc record, a spiralling groove is cut into a flat disc with a stereo cutting head (Figure 2.12). However, the stylus can move both horizontally and vertically so that the groove is the result of both the vertical and the lateral modulations. During the recording, the sum of the left and right channel signals control the lateral modulation, while their difference signal controls the vertical modulation. Therefore the left and right channel stereo audio information is encoded into a function of the groove depth information and the distance between the groove trajectory and the disc center. Stereo discs are recorded by the electric method.

If the left channel and the right channel signals are in-phase, only the lateral modulation is used. Therefore, the stereo disc system can record and play mono disc recordings but the mono disc system cannot record or play stereo disc recordings.

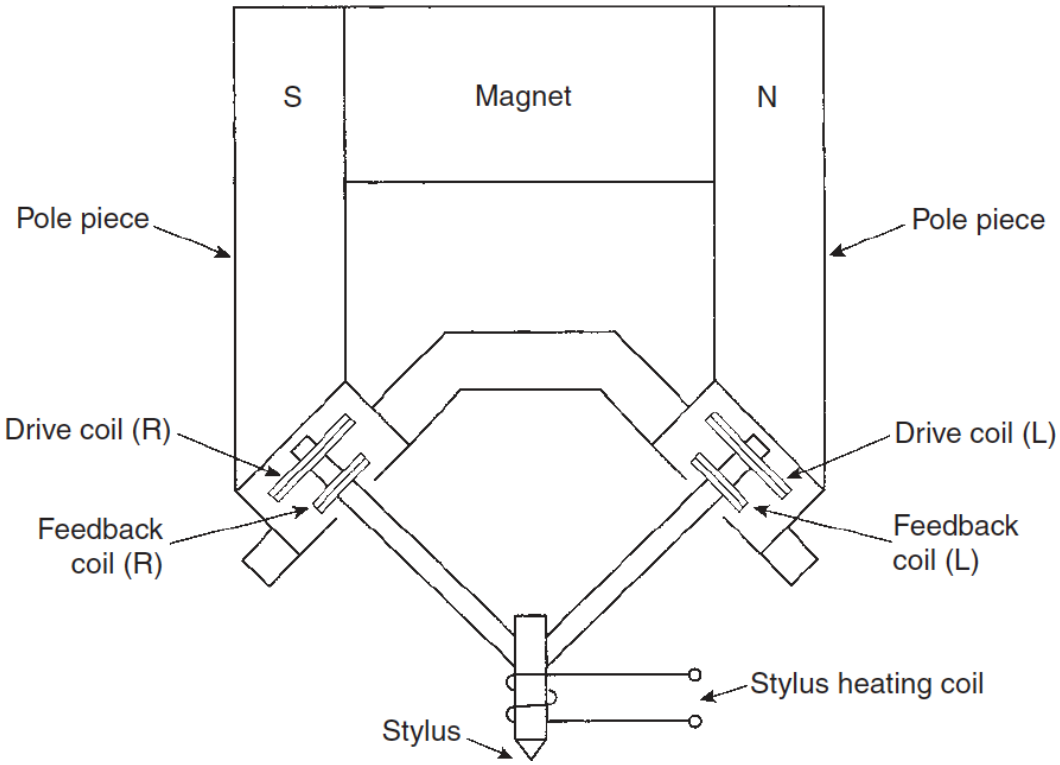


Figure 2.12: A stereo cutting head (image from Huber and Runstein 2005, 573)

2.3.2.3.3 Recording Principles of Quadraphonic Discs

Inoue et al. (1970) invented the CD-4 quadraphonic phonograph disc system that is capable of reproducing 4 discrete channels of audio (front-left, front-right, rear-left, and rear-right). The system uses the same mechanical systems for cutting and replaying the stereo recordings, but encodes the stylus modulation control signals differently. The system diagram and the audio encoding scheme are shown in Figure 2.13. The system first uses transducers to obtain four electric signals corresponding to the four channels; then divides them into two groups: the left group that contains the front-left and the rear-left channels, and the right group the front-right and the rear-right; the sum and the difference of each group are then obtained; frequency and phase modulations are then used to combine the sum and difference into a single signal. In this new compound signal, the sum signal stays in the low frequency band and the difference signal is modulated to

the non-audible high frequency band (45kHz) that does not overlap with the frequency band of the sum signal (Figure 2.14). Finally, the two compound signals are encoded into the stereo groove as if they were true stereo channel signals. During the playback, with the help of filtering and frequency demodulation, the difference signals are separated from the sum signals and the original four channel signals are decoded from the compound left and right signals. The quadraphonic disc system is able to record and play the stereo disc recordings. While the stereo disc system can play quadrophonic discs, it cannot record quadraphonic discs.

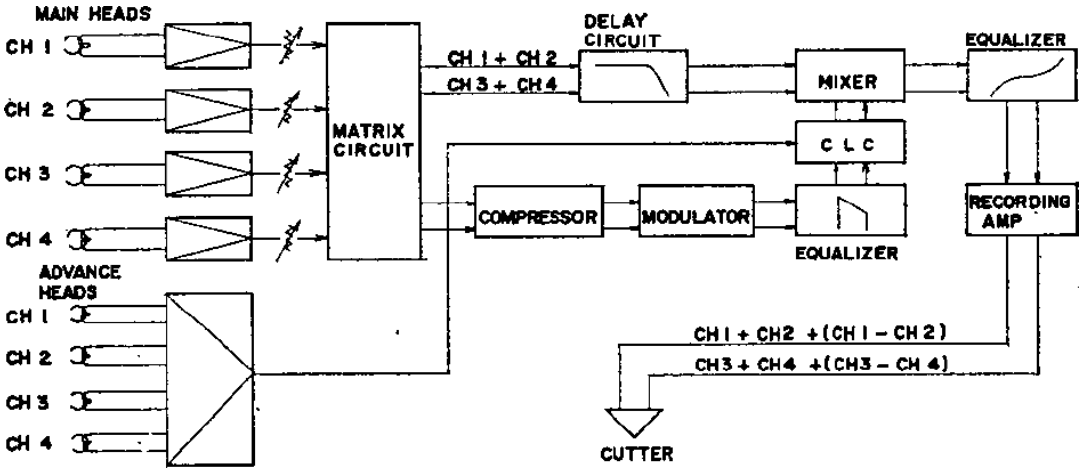


Figure 2.13: The block diagram of the CD-4 disc cutting system (image from Inoue et al. 1970, 577)

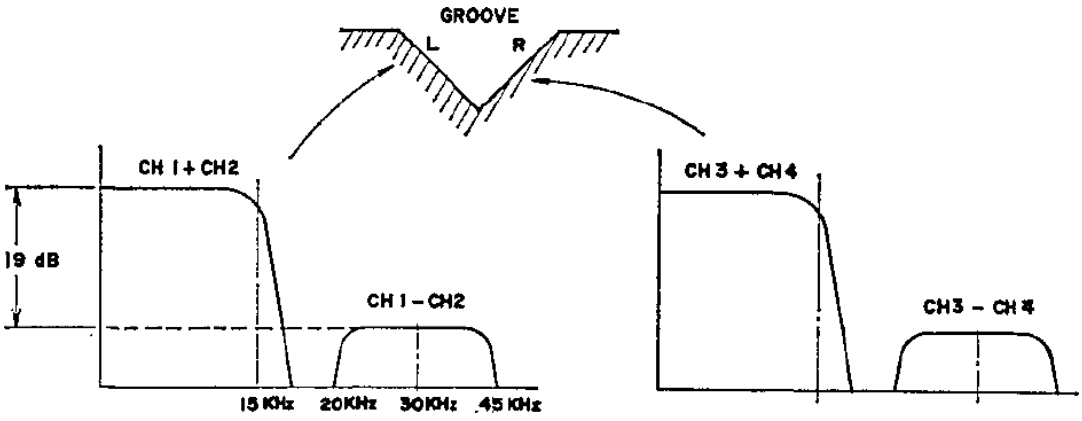


Figure 2.14: The frequency spectrum of the CD-4 system (image from Inoue et al. 1970, 577)

2.3.2.4 Groove Characteristics

Figure 2.15 illustrates the cross-section of a typical disc groove. They are different from the cylinder recordings in that the cross-sections of the disc grooves, including the mono,

stereo, and quadraphonic discs, are roughly "V"-shaped. The spacing between adjacent groove revolutions is referred to as the groove ridge. The bottom of groove represents the moving trajectory of the cutting stylus tip. On both sides of the groove bottom, there are two slope surfaces between the bottom and the two groove ridges. These two surfaces are referred to as groove sidewalls in this dissertation. Viewing from the groove cross-section, the angle from each sidewall to the vertical direction is 45 degrees for both sidewalls, often called 45/45 system. Mono, stereo, and quadraphonic disc grooves all use the 45/45 system.

Mono grooves (i.e., laterally-cut grooves) have constant width and depth because the cutting stylus stays at the same vertical location in the record surface. When no sound pressure is detected by the pressure sensor (a horn or a microphone), the trajectory of the groove bottom represents the unmodulated groove position. When sound pressure is detected, lateral deviations of the groove bottom from its unmodulated trajectory is related to the amplitude of the sound signal. This relationship is defined through the constant amplitude and constant velocity schemes discussed in 2.3.2.6. In comparison, the width and depth of the stereo disc groove are variable throughout its duration because the lateral opening of the groove becomes wider as the cutting stylus moves deeper, and vice versa. Since the quadraphonic disc cutting system reuses stereo hardware, the grooves from both systems are very similar to each other (Inoue et al. 1970).

The detailed statistics of the disc groove characteristics are shown in 2.2.2.7.

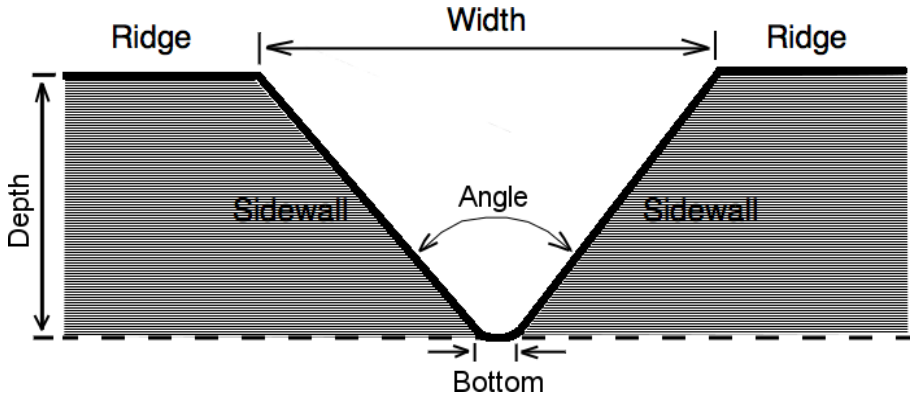


Figure 2.15: Cross-section of a typical disc grooves

2.3.2.5 Audio Duration

The recording duration of a disc record is determined by the size, the rotation speed, and the groove density of the disc.

In terms of the groove density, the disc phonograph recordings can be categorized into coarse-groove records (78rpm) and microgroove records (33rpm). The latter records have denser groove revolutions and lower rotation speeds than the former and therefore offer a longer audio duration.

The rotation speed of the earliest turntables varied greatly. By 1908, 78rpm was selected to be the standard rotation speed of the mono disc record system (Hoffmann and Ferstler 2005, 300). Mono disc phonograph records are thus often called 78rpm records. The 78rpm records are cut with coarse grooves and are usually either 10inch (25cm) or 12inch (30cm) in diameter. A 12-inch 78rpm record provides up to five minutes of audio per side (Stotzer 2006, 18).

The 33 1/3rpm (often called 33rpm) microgroove records were produced in 1925 to meet the needs of the film industry: Soundtracks for motion pictures were played with a turntable. Since a reel of film provides about 11minutes of content, it was desired that one side of the disc record provide the same duration, and thus 33 1/3rpm was determined as the standard rotation speed for the microgroove recordings. Vinyl, as a finer-grained material than the shellac and the acetate, is used for producing 33rpm records to allow denser grooves to be cut. The 33rpm records can provide up to 25 minutes of audio content per side and are often called the LP (Long-Playing) records. Initially, LPs were used for mono audio. In 1948, Columbia Records introduced the stereo LPs.

Other rotation speeds are available. The 16 2/3rpm (often called 16rpm) records were introduced during the 1950s to attempt to provide longer audio durations than LPs. Although 40cm-diameter 16rpm records can provide up to 40minutes of audio per side with coarse grooves and 60minutes with microgrooves, the resulting audio quality is limited, for example, bandwidth is limited to up to 3kHz due to the surface noise caused by the low rotation speed. Therefore the 16rpm records were usually used for speech audio only. The 45rpm microgroove records were introduced in 1949 according to a fixed commercial specification: to provide 5.5minutes of microgroove audio (Stotzer 2006, 19).

2.3.2.6 Equalization

The grooves of phonograph recordings are usually not cut with flat frequency response. Equalization is used in the recording and the playback, and is used to achieve a compromise between the following two contradictory requirements for the groove modulation amplitude:

1) The grooves have to be modulated at sufficiently high amplitudes for all frequency components to avoid being masked by the surface noise, which is mainly introduced by the rough groove surface (see 2.3.3).
2) The groove undulation has to be as small as possible to maximize the audio duration by ensuring a high groove density.

Cylinder recordings and early 78rpm disc recordings encoded the sound pressure directly into the positional information of the groove undulation without filtering. The modulation amplitudes are therefore the same for all frequency components of the sound. This is called the constant amplitude scheme. This scheme has the advantage that it can easily control the overall groove modulation amplitudes. This was in favour of Requirement 1. However, the introduction of the electric recording technology broadened the frequency responses of disc recordings and the constant amplitude scheme could not maintain the high groove density for high-frequency components of the audio signals because they require larger excursions of the cutting stylus. A new cutting scheme, called constant velocity scheme, was then invented to solve this dilemma, and ensure that the higher-frequency components produce smaller groove undulations so that they are comparable with those by the lower-frequency components. This is accomplished by maintaining a constant stylus movement speed for all frequency components using a velocity-sensitive device, the magnetic transducer. Despite its advantages, the constant velocity scheme exhibits an inherent flaw; the resulting groove undulation for the higher-frequency components may be too small to meet Requirement 1. A pre-emphasis equalization was thus introduced to amplify the higher-frequency components while attenuating the lower ones before the groove cutting process. This meant that inverted equalization (de-emphasis) had to be performed during the playback to restore the modified frequency content to its original.

The constant amplitude scheme and the constant velocity scheme both have their advantages and disadvantages. At one point, they were proposed as components of a hybrid solution for different frequency bands (Maxfield and Harrison 1925). In this solution, a bass turnover frequency is specified; during the recording, the constant amplitude scheme is applied to frequency band below the turnover frequency with an attenuation equalization applied; the constant velocity scheme is applied to frequency band above the frequency with an emphasis equalization imposed. Later, it was discovered that towards the higher end of the frequency range, the surface noise level increased and masked the high-frequency signals (Galo 1996). An additional treble turnover frequency was then defined, above which the third cutting scheme is applied with a greater emphasis equalization, the constant acceleration scheme. This scheme provided increasingly large groove amplitude to meet Requirement 1 and 3 without violating Requirement 2 and 4. During the playback, these equalizations were inverted.

The Recording Industry Association of America (RIAA) standardized the pre-emphasis and de-emphasis equalizations in 1954. The standard specifies a pair of equalization curves for the recording and playback, respectively (Figure 2.16). Three frequency turnovers are used: A low bass turnover at 50.05Hz, a bass turnover at 500.5Hz, and a treble turnover at 2122.06Hz. The only difference between the RIAA equalization and the equalization proposed by Maxfield and Harrison (1926) is the additional low bass turnover, below which the constant velocity scheme is used to avoid low frequency rumble noise (discussed later in this section). Figure 2.17 provides a simplified view of this technique.

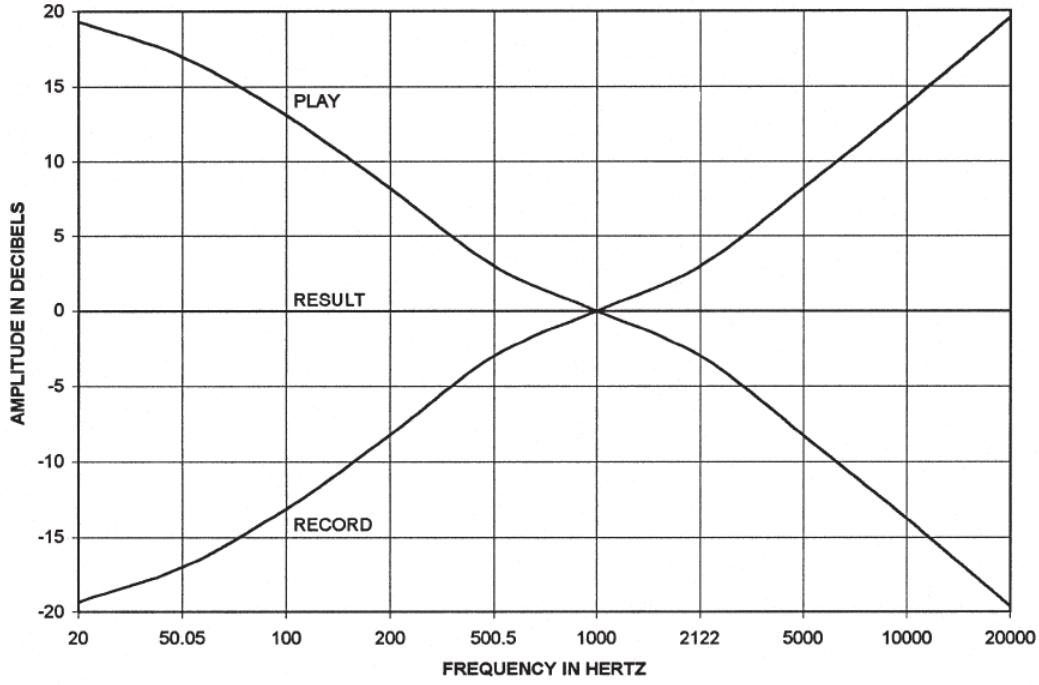


Figure 2.16: The RIAA equalization curve (image from Galo 1996, 192)

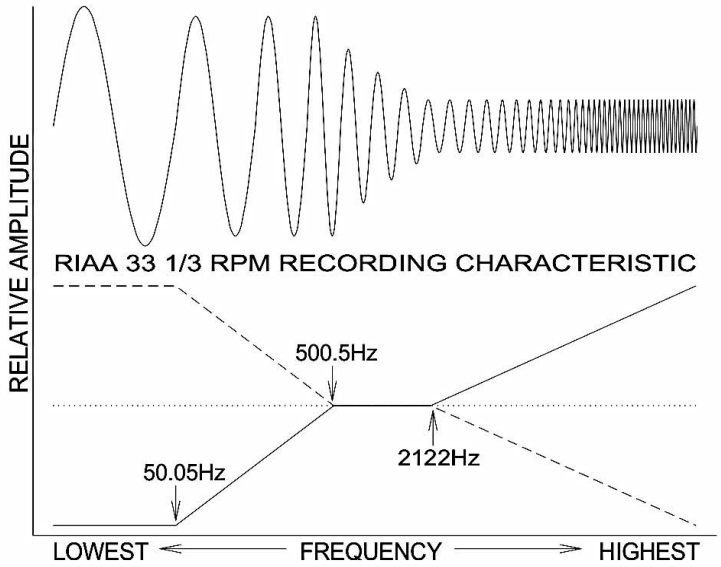


Figure 2.17: The simplified illustration of RIAA equalization curve, showing the frequency ranges for constant amplitude, constant velocity, and constant acceleration cutting schemes (image from Galo 1996, 196)

2.3.2.7 Frequency Response

Early disc recordings used the same acoustic recording method as the cylinder recordings, but with the adoption of electric recording, the subsequent disc recordings do not share the characteristic frequency response described in 2.3.1.6. Various reports regarding the bandwidth that the disc recordings can provide exist. For mono records, Vinyl Records Collector (2010) reports that the acetate records provides only around 2kHz bandwidth. For stereo LP records, a few examples include 7Hz–25kHz (Gray 1997); Glackin and Ricker (2002) reports that the high-frequency response of the high-end stereo LP disc recordings can reach 122kHz. The frequency response of the CD4 quadraphonic disc recordings can reach 45kHz (Inoue et al. 1970).

2.3.2.8 Record Geometry

Fadeyev and Haber (2003, 7) and Stotzer (2006, 26–7) give their measurements of groove geometrical characteristics of the 78rpm and the 33 1/3rpm stereo LP records. They are presented in Table 2.2 and Table 2.3.

Parameter	78 rpm (10" diameter)	33 1/3 rpm (12" diameter)
Groove width (µm)	152–203	25–76
Groove depth (µm)	~74	~15
Bottom width (µm)	38–58	3.8
Groove angle (degree)	82–98	87–92
Grooves per mm	3.78–5.35	7.87–11.81
Groove ridge width (µm)	178–254	83–127
Maximum groove amplitude (µm)	102–127	38–51
Total groove length (m)	152	437
Grooved surface area (m^2)	0.0386	0.055

Table 2.2: Geometry statistics of the disc grooves (converted to metric units from Fadeyev and Haber 2003, 7)

Parameters	78rpm	33 1/3rpm
Groove width (μm)	150–200	25–100
Groove depth (μm)	40–80	25–100
Bottom width (μm)	20.32–63.5	6.35
Groove angle (degree)	82–98	85–95
Grooves per mm	2.8–4.72	6.85–11.81
Groove ridge width (μm)	211–357	84–140
Maximum groove amplitude (μm)	75	25

Table 2.3: The geometry statistics of disc grooves (Stotzer 2006, 26–7)

2.3.3 Noise and Distortions

Phonograph recordings exhibit noise and distortions introduced during the recording, manufacturing, playback, and the storage. Cylinder recordings and disc recordings share some common artefacts, which include:

1) Surface noise: Often referred to as “hiss,” is caused by the rough inner surface of a groove, produced from the friction between the cutting stylus and the rough surface of the record.
2) Pitch fluctuation: Often referred to as “wow,” is caused by the fluctuation in the rotation speed of the phonograph records during either recording or playback. For cylinder recordings, pitch fluctuation can arise from several factors: A cylinder record is not necessarily a perfect cylinder shape and has a variable radius along its longitude; the record surface could become warped with repeated playback or long-term storage; the rotation driver device may not guarantee a constant rotation speed—for example, the earliest cylinder was rotated by hand, which easily introduced speed fluctuation. Due to these causes, there is no guarantee that a constant rotation angular speed can be maintained at all times throughout the entire cylinder surface. For disc recordings, wow can be caused by a few factors. The disc rotates around a spindle passing through its center hole, but the spindle center does not necessarily match the disc center. A warped record surface due to repeated playback or long-term storage can also lead to rotation speed fluctuation.

Pitch fluctuations could be slow or fast. The fast ones are sometimes referred to as "flutter" (Stotzer 2006, 24).

3) Signal discontinuity: Often referred to as "clicks" or "pops," signal discontinuity can be the result of a few factors. Tiny segments of the groove can be damaged by wear and tear over repeated playback or scratches. During long-term storage, the grooves can become overlaid by dust and debris. Mould and fungus can grow on the record surface and erode grooves. More serious damages can result in long missing segments of the groove where the audio signals are completely lost. For cylinder and stereo disc recordings, signal loss can be introduced when the vertical movement of the cutting stylus rises above the record surface (discussed in 2.2.2.3). Although laterally-modulated mono disc recordings do not suffer from this recording pitfall, the playback stylus may still accidentally leave the groove and result in signal interruptions.
4) Tracing distortion: When the cutting stylus cannot move fast enough due to inertia to cut the high-frequency groove undulations or the undulation curve is smaller than the cutting stylus tip, the frequency content will be distorted (as mentioned in 2.3.2.6,). Since the shape of the recording stylus and that of the playback stylus do not necessarily match each other, the playback stylus may not follow the groove undulation produced by the cutting stylus in the exact same manner. The reproduced audio may thus exhibit complex distortions.
5) Mechanical resonance: Disc and cylinder recording and playback hardware may introduce low-frequency noise due to mechanical vibrations in the motors. This is often referred to as "rumble".

Also, disc recordings have their own characteristic noise and distortions:

- Tracking error: The playback stylus of a turntable is mounted on the tonearm. Tonearms include the linear tracking tonearm (Figure 2.18) and the pivoted ones (Figure 2.19). Pivoted tonearms are more commonly used. As illustrated in Figure 2.20, with a pivoted tonearm there is an angle between the tonearm and the radius that the stylus is supposed to move along. The angle changes as the tonearm shifts from the outer to the inner groove revolutions. An angular error is introduced

during the playback so that the playback stylus often does not move linearly along the disc radius, which introduces special harmonics into the reproduced audio.

Figure 2.18: A turntable with its linear tracking tonearm (Eugenio, 2010)

Figure 2.19: A turntable with its pivoted tonearm (Wikipedia, 2011)

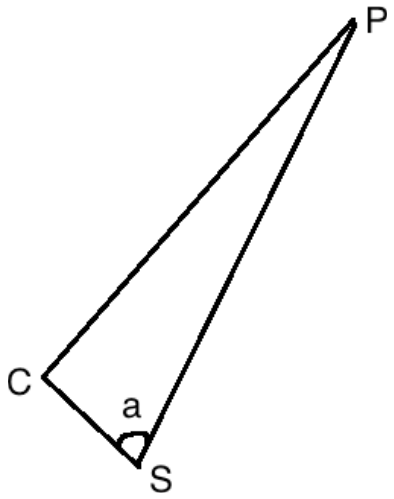


Figure 2.20: Tracking error illustration: Point *C* represents the disc center, Point *P* the tone-arm pivot, and Point *S* the stylus; Angle *a* is the angle between the tone-arm and the Radius *SC*

The OAR approaches have the potentials to reduce the above noise and distortions of the traditional phonograph recording and playback systems. Before discussing the OAR research, it is necessary to provide a summary of the optical surface metrology techniques adopted by previous OAR research and the one used in the research of this dissertation.

2.4 Optical Surface Metrology

Optical surface metrology measures small-scale surface texture, including surface roughness and waviness, using optical methods without contacting the surface being measured. Its common applications include the analysis of smooth engineering surfaces, biological cell analysis, and medicine inspection. Although all optical surface metrology techniques involve projecting a light beam onto the target surface and interpreting the images formed by the reflected light, they differ greatly in exact technique. In this section we introduce several important techniques that have been used by OAR for retrieving 3D groove information, namely laser triangulation, confocal microscopy, and white-light interferometry (WLI). The important technical parameters that affect the performances of the techniques in OAR are noted.

2.4.1 Laser Triangulation (Ray Tracing)

Shown in Figure 2.21, a laser beam is projected onto the target surface. The diffusely reflected beam passes through a focusing lens system and is received by a light intensity

sensor. The sensor detects the position of maximum light intensity of the reflected beam, from which the surface height can be derived by triangulation.

Laser triangulation can obtain 3D surface information with a vertical resolution of less than 1μm. This is relatively low resolution compared to confocal microscopy and WLI.

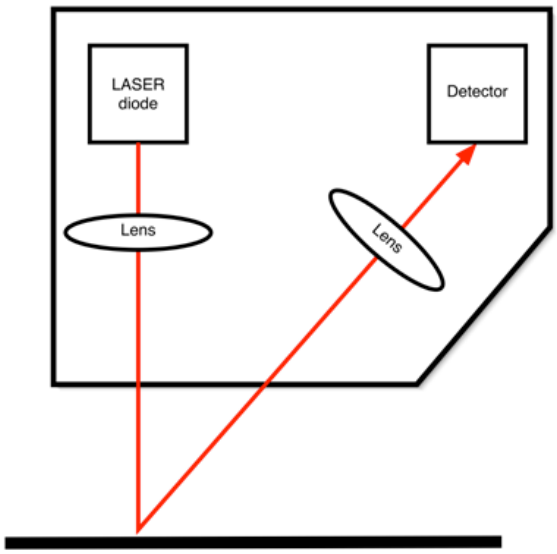


Figure 2.21: Laser triangulation (image from Nascè 2009, 64)

2.4.2 Confocal Microscopy

As seen in Figure 2.22, a basic confocal microscope comprises a light source, an objective lens, a beam splitter (usually a di-chromatic mirror that reflects lights of certain wavelengths but allows the other lights to pass it through), and a photosensor with an optical aperture (a pinhole) mounted in front of it. During the measurement, a light beam (e.g., a red-light laser for a confocal laser microscope) is reflected by the beam splitter, focused through an objective lens, and forms a small spot onto the target surface. The reflected light is propagated back, focused through the same objective lens, and is split by the beam splitter from the incidental light to reach the pinhole. The pinhole and the photosensor form a point-light detector and reject all but the reflected light in focus. This configuration prevents the out-of-focus reflected light from interfering with the in-focus light, resulting in a sharp image. The surface vertical position is then derived from the lens position corresponding to the maximum light intensity received by the photosensor.

The confocal laser microscope can obtain 3D surface information with a lateral resolution of 1μm and vertical resolution of 10nm. The disadvantage of confocal microscopy had been that due to the use of a point-light detector, the time required for scanning a large surface is usually long: for example, 160 hours for 10mm length of surface (Nascè 2009, 68). However, latest development in this technology has introduced multi-point line sensors that can perform faster scans. For instance, the MPLS-180 produced by STIL can measure 180 points (with an inter-point spacing of 10μm) simultaneously and a scan speed of 324k points/second (STIL 2011). Note that due to the inter-point spacing, the line sensor needs at least two scans to acquire a complete linear surface without missing areas.

The white-light confocal microscope uses white light as the light source; the white light is then split into several confocal systems corresponding to different wavelengths. The highest intensity among the results of these sub-systems is then used as for calculating the surface height. In 2.4.4, the comparison between the white-light confocal microscope and the laser confocal microscope is presented.

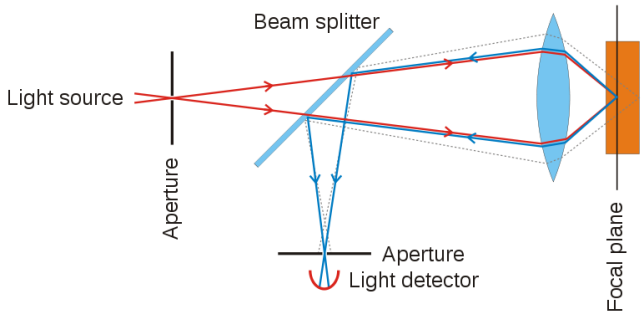


Figure 2.22: Principle of confocal microscope (image from Wikipedia 2011)

2.4.3 White-light Interferometry

Interferometry is an optical surface metrology technique that exploits optical interference for accurately measuring distance, displacements, and vibrations. It has various applications such as testing of optical systems, measurements of gas flows, plasmas, temperature, pressure, electrical and magnetic fields, rotation sensing, microscopy, and high-resolution spectroscopy (Hariharan 1992). White-light interferometry (WLI) is a special interferometry using white light as the light source. To understand WLI, we first introduce interferometry.

2.4.3.1 Principles of Interferometry

Light is a type of transverse electromagnetic wave. Light beams, when superposed onto each other, can create interference: Two light beams combine in such a way as to produce a resultant intensity that is either greater or less than the sum of the individual intensities, known as the constructive and destructive interference, respectively (Hariharan 1992; Morgan 1953). The resultant intensity can be derived as follows:

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2}\cos(\Delta\varphi) \qquad \text{Equation 2.1}$$

Where I_1 and I_2 are the separate intensities of the two light beams, and $\Delta\varphi$ the phase difference between them. If the two beams are derived from a common source, we then have:

$$\Delta\varphi = \frac{2\pi}{\lambda}\Delta p \qquad \text{Equation 2.2}$$

That is, the phase difference between two lights from the same source is linearly proportional to their optical path difference (OPD). Thus the resultant sinusoidally varying intensity forms alternating bright and dark fringes (intensity patterns, or interferogram) when projected onto a photosensor. The bright and dark fringes are the results of the constructive and the destructive interferences (Figure 2.23). Based on Equation 2.2, if the wavelength of the light source is known, the intensity patterns can be measured to derive the OPD (the distance information) between the two beams. Interferometry exploits this property of interference to accurately measure surface textures. To obtain sustained and distinguishable interference patterns, light beams used by an interferometer should satisfy the following conditions (Morgan 1953):

1) The phase difference from the two light beams must remain constant within the propagation time and spatial range, which correspond to the measurement period and gauge range. The propagation time and spatial range are called the coherence time and the coherence length, respectively. They are proportional to the wavelength of the light source if the light beams propagate in the same medium.
2) The light waves must have the same frequency, same wavelength, and nearly equal amplitudes.
3) The two interfering waves must possess the same polarization state.

4) The wavefronts of the interfering waves must be traveling in the same direction or span only a small angle between each other.

Figure 2.24 shows a basic interferometer (the Michelson interferometer). A light beam from a coherent light source with known wavelength is propagated onto a beam-splitter and is divided into two beams, which satisfy the aforementioned Requirements 1–3; the two beams are directed onto a reference surface and the target surface and then reflected back; and finally the two beams are directed into the same direction towards an intensity-sensitive photosensor (Requirement 4), and superposed, forming fringes measured by the sensor. In interferometry-based surface metrology, the interferometer adjusts the relative positions between the target surface and the reference surface and obtains a series of fringe patterns at each reference position. The position of the target surface relative to the reference surface determines an OPD between the two beams, which can be subsequently measured through various methods. The fringe pattern that presents the best intensity contrast, or, zero OPD between the reference beam and the target beam, is found. Finally the surface height can then be derived from the position of the reference surface that corresponds to the selected fringe pattern, and the optical path length of the reference beam.

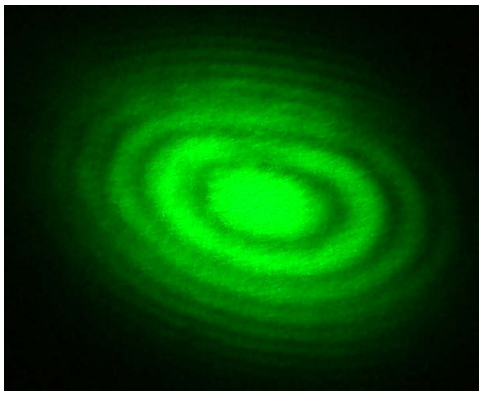


Figure 2.23: Interference pattern observed through a Michelson interferometer (image from Wikipedia 2010)

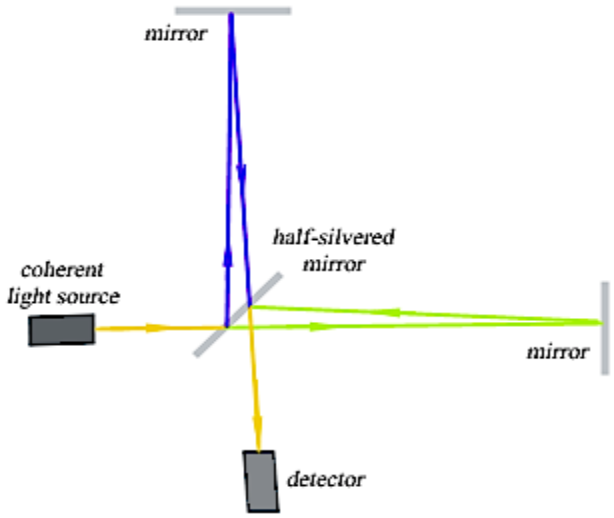


Figure 2.24: The principle of a basic interferometer (image from Wikipedia 2010)

2.4.3.2 White-Light Interferometry vs. Phase-Shift Interferometry

To acquire the surface profile by finding the zero-OPD between the reference and the target surface, there are two approaches in measuring the OPD between the split beams: the phase-shift interferometry (PSI) and white-light interferometry (WLI), sometimes also called "vertical-scan interferometry" (VSI).

Based on Equation 2.2, the OPD can be derived via the known wavelength of the light source and the phase difference between the reference and the reflected beam. Based on this principle, the PSI approach uses a monochromatic coherent light source of known wavelength and obtains the zero OPD by shifting the reference surface by a small known amount, causing a phase shift between the reference and the reflected beam. By measuring the phase difference between the resulting interferograms at each shift step, the zero phase difference in the fringe pattern is found by the photosensor. The zero OPD can be obtained and the surface height can thus be derived. However, the PSI is only suitable for measuring smooth surfaces with the maximum step size of $\lambda/4$, in which λ is the wavelength of the light source. This is because phase ambiguities exist when the target step size of the surface is of multiples of half-wavelength, resulting in several possible heights at which the zero-OPD can be found (Blunt 2006). In practice, red light is often used in PSI. As an example, the maximum measurable step size by the PSI mode of our WLI microscope is about 150nm when using a 633nm-wavelength red-light source

(Veeco Metrology Group 2003, 3). PSI is therefore not suitable for measuring rough surfaces with deep slopes such as the grooved surface of disc phonograph records (see Table 2.3).

To measure rough surfaces with a step size greater than the PSI can handle, the WLI approach provides a solution. By using a white light source—that is, a broadband light source—the interferometer objective vertically scans the target surface at various heights, capturing a series of interference intensity signal frames (see Figure 2.25) as each vertical location passes the focal plane of the objective, called correlogram (Harasaki et al. 2000). The correlogram is defined in the equation below (Kino and Chim 1990):

$$I(x,y,z) = a(x,y) + b(x,y)c\left[z - 2h(x,y)\right] \cdot \cos\left[2\pi\lambda_0 z - \alpha(x,y)\right] \qquad \text{Equation 2.3}$$

where $I(x,y,z)$ represents the interference intensity at the surface planar location (x,y) and height z, $a(x,y)$ the average intensity, $b(x,y)$ the modulation factor, $c(z)$ the envelope function defined by the spectral properties of the light source, $h(x,y)$ the height of the surface planar location (x,y), and $\alpha(x,y)$ the initial phase difference between the reflected beam and the reference beam. The zero OPD location is derived by using a coherence-peak-sensing approach (Balsubramanian 1980), which finds the vertical-scan frame that corresponds to the interference peak of the correlogram. Since a white-light source is a polychromatic and has short time coherence, the interference occurs only in a limited range around the coherence plane, which is defined by the zero OPD. This results in a compact correlogram, from which an accurate surface height corresponding to zero OPD can be obtained. The white light source also ensures no phase ambiguity, as that exhibited by PSI (Blunt 2006). The surface step size that can be resolved by WLI with vertical scanning is therefore not limited by the λ/4 threshold.

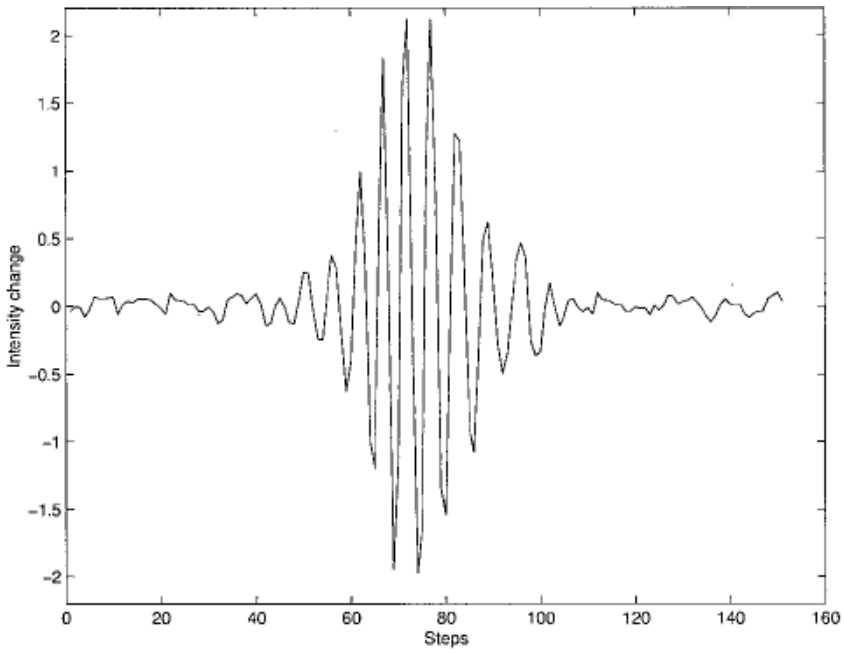


Figure 2.25: A typical interference correlogram obtained by the WLI at a single planar location across the height range of the coherent length of the white-light source (image from Olszak 2000)

WLI is the adopted optical surface metrology technique in this dissertation. The actual optical surface metrology hardware is introduced in Chapter 3.

2.4.4 Comparison of the Optical Surface Metrology Techniques

A comparison of several measurement parameters in the aforementioned optical surface metrology techniques is given in Table 2.4. Nascè (2009, 67–9) provides benchmarking of the laser triangulation (LT), the confocal laser microscope (CLM), the white-light confocal microscope with a point sensor (WLCM), and the MPLS-180 white-light confocal line sensor (WLCLS). We use most of the same parameters in our comparison. These parameters are as follows:

1) Axial resolution: The minimum resolvable surface displacement by the optical sensor of interest. In the case of WLI, the lateral and vertical resolutions are given separately.
2) Gauge range: The vertical range in which the sensor can reliably measure the height of the surface.

3) Angular tolerance: A common limitation of optical surface metrology techniques is that they cannot measure a surface with an overly steep slope. When the target surface has a slope steeper than the angular tolerance of the sensor, the reflected light beams can be obstructed and fail to reach the sensor. The measurement would thus fail to acquire the slope surface height information.
4) Spot size: For single-point sensors such as the confocal microscope, obtaining the measurement result of the minimum resolvable lateral surface area and averaging over its neighbouring area is taken to yield the measurement result as the sensor moves across the spot. This parameter defines the lateral resolution of such surface metrology techniques based on single-point sensors, and is therefore not applicable to WLI because it is not a point-sensor.
5) Sampling frequency: The number of measurements the optical instrument can make per second.
6) Scan time: The time taken to measure and store surface heights. Note that no evaluation setup such as the mean depth of the target surface is given for evaluating scan time in Nascè (2009). The scan time of WLI for a 10mm deep surface in Table 2.4 is derived from a vertical scan speed of 5μm/sec.

Sensor	Axial resolution (nm)	Gauge range (mm)	Angular tolerance (degrees)	Spot size (μm)	Sampling frequency (Hz)	Scan time for 10mm-long linear surface (hours)
WLI	Lateral: 100 Vertical: 3	8	25	N/A	60	0.56
WLCM	10	0.35	27	7	1000	22
WLCLS	Lateral: $1x10^{4}$ Vertical: 125	0.5	N/A	4	1800	$8.56x10^{-7}$
CLM	10	0.6	17	2	1400	160
LT	1000	10	17	30	2000	19.5

Table 2.4: Comparison of several important optical surface metrology techniques (the data other than WLI are from Nascè 2009, 68 and STIL 2011; the WLI data are from Veeco Metrology Group 2003). Note that the angular tolerance is not part of the specification of WLCLS

With WLI, the Charge-Coupled Detector (CCD) camera of our microscope provides a 60 frame-per-second (fps) scan speed.

Most optical sensors have limited fields of view (FOV), so that to measure a large target surface an optical sensor has to move over the surface. At each scan location, a single vertical measurement is made. Note that CCD camera is used in WLI, which can simultaneously measure an array of points across the target surface.

Nascè (2009) reports that due to the low vertical and lateral resolution, the LT usually yields noisy measurement results for record grooves; CLM needs to trade high resolution with low sampling frequency; WLCM is superior to LT and CLM, which offers the best combination of axial resolution, angular tolerance, and sampling frequency for measuring the grooves of cylinder recordings. In comparison, WLI has the advantage of higher vertical resolution and faster scanning speed.

In the context of scanning grooved phonograph record surface, one shared disadvantage of all the optical surface metrology techniques is that the reflected light beam can be trapped inside the groove sidewall and cause the measurement of the portion of the sidewall to fail. It is therefore difficult for an optical surface metrology technique to acquire the entire disc groove sidewalls of the 45/45 system. For example, the groove sidewalls are often missing in the scanned images by WLI (see Chapter 3).

2.5 Previous OAR Approaches

In this section, the previous OAR approaches are reviewed. As introduced in 1.2, most OAR frameworks follow the same high-level three-step procedure to reproduce the phonograph audio, even though they operate on different types of recordings such as cylinders or mono discs. Approaches for each of these steps vary significantly in terms of the adopted optical techniques. In this section, we discuss the previous OAR approaches in terms of their targeted types of recordings, adopted optical techniques, general audio reproduction workflows, and achievements. They are divided into four categories: optical playback systems, approaches based on 2D imaging, approaches based on confocal microscopy, and approaches based on standard optical microscopy.

2.5.1 Early Optical Playback Systems

The first category of optical playback systems attempted to recognize and optically measure the grooved portion of the record surface, decode the audio information in the measurement result, and play it back in real-time. The measurement requires optically tracking the groove in the manner resembling the stylus of a phonograph playback device. The ray tracing method described in 2.4.1 is mainly used to measure the groove surface profile. This category of systems has been previously referred to collectively as the optical stylus (Stotzer 2006), the optomechanical method (Petrov 1992), or the optical turntable (Stoddard and Stark 1988). In this dissertation, these systems are referred to as the early optical playback systems. Below we briefly introduce a few important of the early systems in terms of:

1) target media types, e.g., the cylinder or the disc recordings
2) optical representation of the groove undulation
3) methods of locating and tracking the groove from its beginning to its end without leaving it
4) properties of the light source

According to Brock-Nanestad (2001), the concept of extracting audio from phonograph records by using a light beam dates back to the late 1920s. Friebus (1929) conceived of an OAR concept system intended to be used for both cylinder and disc phonograph records: A steady light beam from a light source is cast onto the groove sidewalls and the reflected light is received via a photosensor mounted to the eye-piece of a microscope. The curvature along the groove sidewall causes the intensity of the reflected light to fluctuate. Such intensity fluctuations thus represent the groove undulation and are subsequently converted into intermittent electric current by the sensor. This is then passed on to an amplifying system and finally generates analog audio. This concept does not specify the portion of the groove cross-section that needs to be illuminated and measured, for example, at which vertical location of the groove sidewall the curvature is measured, the groove tracking method, and the properties of the light source.

Williams (1930) proposed a system for both cylinder and disc recordings. The proposed system illuminates the entire groove cross-section with a focusing lens system, and uses a diaphragm aperture to control the amount of reflected light that reaches the

photosensor, so that as the sensor traverses along the record grooves, the succession of light flux passing through the aperture represents the amplitude of the groove undulations. It is not clear how deep the device focuses itself into the groove and how it tracks the groove.

Gradually, more systems started to address the groove-tracking problem. In order to maintain the measurements inside the groove, servomechanisms were adopted which exploited the optical information received by the photosensor as the feedback to detect the boundaries between the groove and the other portions of the record surface. Alexanderson (1934) proposed a system specifically for cylinder recordings. The system uses a pair of photosensors to measure the undulations of both groove sidewalls at a specific depth and the sum of the received intensities of the reflected lights from both sidewalls are interpreted as the amplitude of the mono audio at the groove position. The difference of the intensities is used as the groove tracking control signal. The difference should remain zero for the mono recordings if the pair of sensors remains focused onto each groove sidewall. The positions of the sensor pair are constantly adjusted to maintain the zero difference so that the groove is properly tracked. The tracking is driven by a cylinder-like custom rotary device.

Rabinow and Morse (1961) targeted both mono and stereo disc recordings. In contrast to the aforementioned approaches, which measured the groove undulation at a specific depth, their proposed system scans back and forth laterally across the entire groove cross-section and tried to cover its entire depth range. To ensure that the breadth of the scan remain inside the groove, a photosensor, placed with a special angle towards the record, is used to detect the reflected lights whenever the scan reaches the top ridge of either side of the groove cross-section. When reaching the ridge on one side of the groove cross-section, the scan is then reversed towards the other side. Such a scan method therefore also acts as the groove tracking mechanism. The scan is performed while the record is rotated by a turntable.

As can be seen from the above example, researchers also came to the idea of reusing part of the regular mechanical phonograph playback systems, for example, keeping the turntables as part of the groove tracking mechanism while replacing the stylus with the optical measurement device. Dahlen (1965) proposed a system for both

mono and stereo disc recordings. It replaces the stylus at the tip of the tonearm of a turntable with a non-contact optical device. As the record rotates on the turntable, the optical device illuminates both the cross-section and the top ridges of the groove, and inserts a polarized filter to eliminate spurious light reflections, thus creating high-contrast light-and-dark intensity images from the reflected light. The bright regions correspond to the groove top ridges while the dark ones the sidewalls. The boundary of the light and dark regions represent the groove undulation. These images are then transmitted to a pair of photoelectric sensors via optical fibres to create electric currents. The sensor pair also serves as the groove tracking mechanism to keep the device aimed at the center of the groove. This is achieved by constantly checking whether both sensors receive light intensity images that are symmetrical relative to the center of the groove and maintaining such symmetry by shifting the tone-arm according to minimize the detected intensity difference from both sensors.

Early OAR research documents give little or no description of the light source properties. Later, monochromatic light sources (mainly the laser) are described. Heine (1974) described the use of a laser interferometer for mono, stereo, and quadraphonic phonograph recordings, in which the groove sidewalls are illuminated with a laser beam, consisting of several light waves. As monochromatic source, the laser beam wavefronts are initially in-phase; the groove curvature alters the spatial relationship of the wavefronts, resulting in interference and diffraction patterns that encode the groove undulation. This information is then converted to electrical signals for reproducing sound. The system also adopts a servo system to lock the incident laser beam within the groove region while the turntable rotates the record.

The combination of the laser light source and a groove-tracking servo system was adopted as the basic framework of many subsequent OAR systems such as Tanaka et al. (1980), Iwai et al. (1985), Stoddard (1986), and Stoddard and Stark (1988). The latter two patents eventually turned into the first working laser turntable, the Finial LT-1, produced by Finial Technology and presented at the 1989 Winter Consumer Electronics Show (Hodges 1989). This turntable handles mono and stereo disc phonograph recordings. It uses a dual-beam optical technique (Stoddard and Stark 1988). When playing an LP, the optical unit is placed at a fixed height from the record; the laser beams are divided into

two separate groups: data and tracking beams. As the record rotates, the data beams are used to determine the lateral and vertical undulations of the grooves, while the tracking beams track the lateral and vertical positions of the optical unit relative to the record surface as a servomechanism. Commercial laser turntables were made available (Finial Technology 1986; ELP Corporation 1997) but were never mass-produced for the consumer market (Hodges 1989). The framework inspired by the laser turntables received a few additional efforts in the 1990s (Poliak 1992; Petrov 1997).

In addition to audio reproduction applications, it is worth mentioning that the optical servo system was also used for improving the stylus playback precision in the mechanical phonograph systems by optically monitoring the stylus movements (Rabinow and Fischer 1963). Early optical playback systems differ from the later OAR systems in that:

1) They aim at playing the phonograph audio back in real-time instead of preserving the records; the extracted groove topology is thus not preserved.
2) When acquiring the record surface information, they measure only the groove portion and need to track the groove in a manner similar to the mechanical playback systems use a stylus.

These approaches thus cannot handle broken or damaged records where tracking the grooves is impossible.

2.5.2 OAR with 2D Imaging

2.5.2.1 The VisualAudio Project

Stotzer et al. (2004) created a system (VisualAudio) that performs a multistep OAR to provide an archiving solution for 78rpm disc phonograph records. Figure 2.26 shows the workflow of the system. First, the entire side of a 78rpm disc record is photographed into a film with a custom-built large 2D analog camera. The film that contains the 2D surface image of the record is preserved as an analog copy of the record. When the audio is needed to be reproduced, the film is then placed onto a custom-built rotary stage; as the stage rotates, an overhead CCD camera mounted on a microscope continuously captures the film portion passing beneath its field-of-view (FOV) into high-resolution digital image; the camera periodically shifts perpendicularly to the rotation whenever a revolution of the film is captured, until the full groove-bearing area of the film is

captured. Next, the resulting 2D digital images are analyzed by software to extract the groove undulation and then decoded into audio. The advantage of the photography step is reported to be able to keep an analog copy of the record for fast archiving purposes before making decisions on the eventual format specification for digitally archiving the audio. Furthermore, extracting the groove undulation by software bypasses the need to track the groove directly during scanning. This system is thus not a real-time playback device but stresses its archiving usage. In Chapter 3 of Sotzer (2006) detailed description of their OAR system and performance analysis about the resolution, noise, and distortions of the system is given.

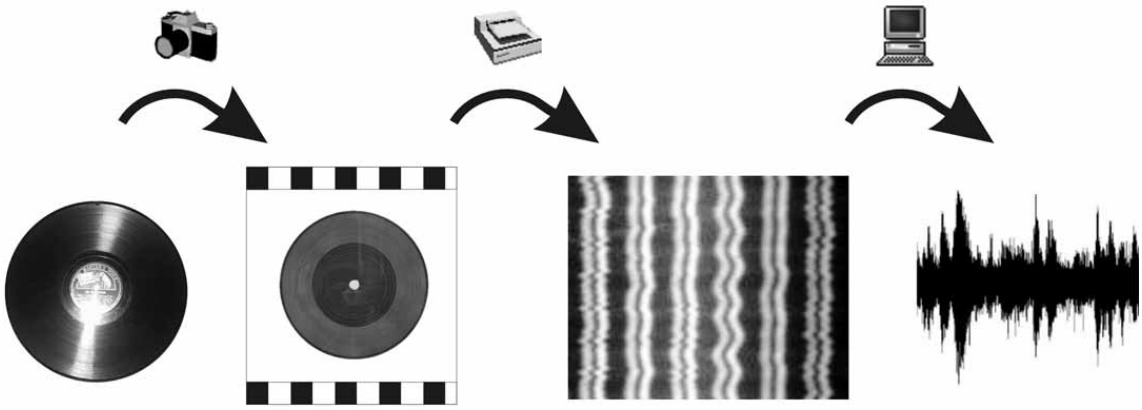


Figure 2.26: The workflow of the VisualAudio system for archiving and reproducing audio from 78rpm disc records (image from Stotzer 2006, 3)

The camera apparatus used in the analog photography step is a custom-built optical instrument of three parallel layers (see Figure 2.27): The film is placed on the top layer, the middle layer consists of a camera with 1:1 magnification, and the bottom layer contains a light source to illuminate the record. The goal of this step is to capture the image of the entire area of the record surface, in which the groove ridges and the sidewalls can be clearly distinguished. To photograph the entire record surface with the homogeneous lighting condition across the record, the size of the illumination system is chosen to have a diameter of 83cm (Stotzer 2006, 37), twice the size of the largest possible record disc in their research, 40cm (Stotzer 2006, 34).

Stotzer (2006) took a few measures to ensure the sharpness of the resulting image on the film: He used the monochromatic source to avoid chromatic aberration of the camera that may lower the sharpness of the image; he chose the blue light because its

short wavelength results in sharper images; the adopted film is orthochromatic and sensitive to only the blue light; the film in use is black-and-white, which is said to provide better sharpness than the colour films and lower graininess in the resulting image.

Stotzer (2006) made the boundaries (edges) between the groove ridges and the sidewalls as the groove undulation. To obtain high-contrast edges between the groove ridges and sidewalls in the film, he used a spiral-shaped fluorescent tube as the light source, of which the incident light is specially angled to only illuminate the groove ridges, leaving the sidewalls to be dark regions. Then, to avoid the diffusion halo due to long exposure time can blur the edge between the groove ridges and sidewalls in the resulting film; to avoid this, the shortest possible exposure time is selected (15 seconds).

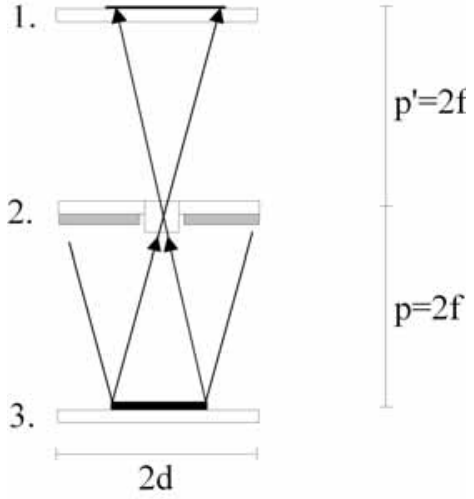


Figure 2.27: Camera apparatus used in the analog photography step of the VisualAudio system (image from Stotzer 2006, 35)

In the film scanning step, the overhead camera samples lines of pixels at a fixed sampling rate (lines-per-rotation) while the film is rotated on the rotary tray. The sampling line is along the horizontal radius of the tray. The camera and the rotary tray together effectively unwrap the spiralling groove in the film image so that it becomes straight in the digitally captured images. The goal was to obtain a high-resolution digital image of the entire groove in the film. The camera sampling rate and the tray rotation speed were chosen to maximize the coverage without missing any portion as the disc rotates. The consecutively acquired images in one revolution and those of the adjacent revolutions overlap so that the redundancy guarantees that there should be no discontinuity in the obtained groove.

The scanning device maintains a constant image sampling rate by using a phase-locked loop to remove fluctuations in the camera shutter speed, and a constant tray rotation speed by adopting a high-inertia tray that resists the vibration of the driving motor. The result of this step is a series of digital grayscale images, each containing the unwrapped groove shown as light-and-dark regions on the negative film. The dark regions correspond to the groove ridges and the light ones the sidewalls, respectively (Figure 2.28).

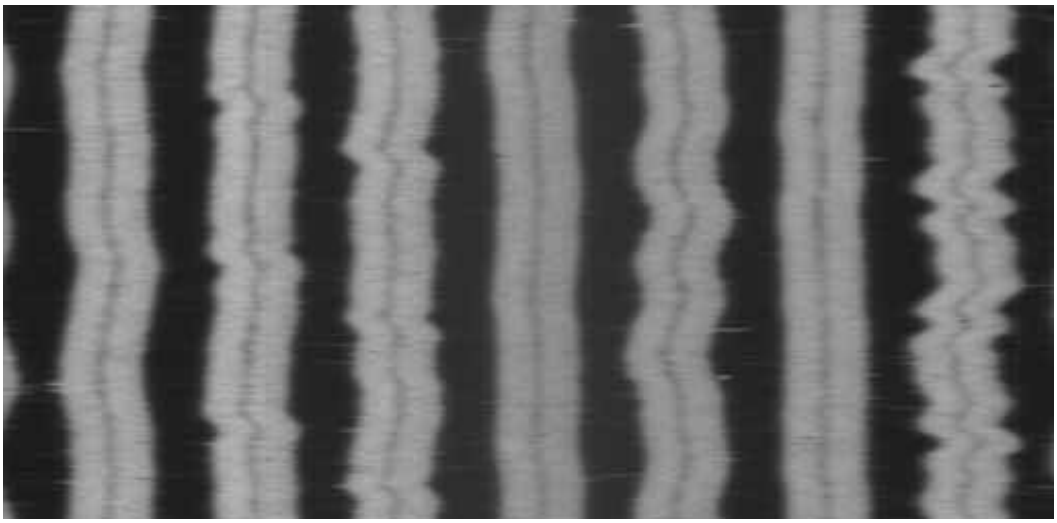


Figure 2.28: A portion of the grooves of a 78rpm disc record in the scanned digital images (image from Stotzer 2006, 109)

In the software analysis step, the groove undulations are extracted from the digital images and the mono audio is obtained from the groove undulation. In the images, the undulation is represented by the boundary between the light-and-dark regions and located by using edge detection algorithms for 2D grayscale images. The edge between the ridge and sidewall is then extracted, resulting in a 1D time-series. Since the grooves of the disc recordings are mostly cut with the constant velocity scheme the audio signal is then derived from the first-order time derivative of the 1D undulation series. Finally, the de-emphasis filtering, implemented with the digital filters, is applied to obtain the output audio.

Stotzer evaluated the image resolutions provided in the analog photographing step and the overall image resolution of the acquisition system. He also derived the theoretical audio sampling rate based on the image sampling rate in the film scanning step.

The image resolution in the analog photographing step is determined by both the film resolution and the camera resolution. The adopted film has a resolution of 600 lp/mm (line pairs per mm), equivalent to a digital resolution of 0.83μm/pixel (Stotzer 2006, 30). According to Stotzer, a high camera resolution requires a large depth of field, which in turn needs a large lens size. However, a large lens introduces diffraction and the camera geometric aberration (blurring) that may reduce the resolution of the outer rim of the resulting record image. To avoid this, he uses a large diffraction-limited aperture lens for the camera (Stotzer 2006, 37), which provides a depth of field of 70 μm—long enough to cover the mean depth of the 78rpm records. The resulting camera pixel resolution is 58.8 lp/mm or about 8.5μm (Stotzer 2006, 46).

In practice, Stotzer claims that the system image resolution is derived by considering both the resolution of the film, the resolution of the cameras, and different types of blurs introduced in the multi-step solution. These blurs include camera aberration, shading blur caused by illumination, motion blur caused by tray rotation in the film scanning step, and image sampling blur when using rectangular pixels to represent groove arcs. The worst-case system image resolution is estimated to be 24.6μm along the direction of rotation (Stotzer 2006, 50).

In the film scanning step, according to Sotzer, the image sampling rate while the film is being rotated is determined by the fixed camera line rate LPS_{camera} (1.31×10^4 lines-per-rotation), the rotation speed of the tray RPM_{tray} (12rpm or 6rpm), and the RPM of the record. The equivalent audio sampling rate for 78rpm records is thus (Stotzer 2006, 140):

$$S_{audio} = \frac{LPS_{camera} \cdot RPM_{record}}{RPM_{tray}} = \begin{cases} 85.2kHz, RPM_{tray} = 12 \\ 170.4kHz, RPM_{tray} = 6 \end{cases} \qquad \text{Equation 2.4}$$

To evaluate the audio quality obtained by the OAR system, Stotzer analyzed the noises and distortions that affect the extracted audio, including the signal discontinuity caused by record surface damages, the noisy edges caused by illumination variation and saturation, the noise from the film artefacts, the CCD camera noise and saturation, and the pitch fluctuation caused by off-axis rotation and the non-horizontal rotation of the rotary tray similar to the wow artefact described in 2.3.3. The detailed analysis of their impact on the audio quality can be found in Chapter 4 of Stotzer (2006). He also tried to restore damaged groove portions in the scanned digital images. The damaged portions were

detected by assuming that the pixels of the corrupted groove are outliers that are not correlated with the surrounding well-conditioned groove pixels. These can be located by using gradient masks called the corrupted pixel map. Interpolations were then used to fix these outlier pixels. After converting to audio, various filtering is performed in the audio domain to remove low and high frequency noises in the extracted 1D undulation signals before the audio is finally generated.

To evaluate the audio quality of his OAR system, the full solution is applied to a test 78rpm record containing a sine signal of known frequencies and signal level (300Hz with +15dB gain). The reproduced audio is evaluated by examining the signal-to-noise ratio (SNR). These parameters are measured for 1) one-second audio signal and 2) the entire extracted audio signal. The best SNR obtained is 19dB for 78rpm records, lower than the SNR that typical turntable systems provide (40dB). As mentioned previously, the equivalent audio sampling rate from the film scanning step is up to 170kHz for the 78rpm records. Different image acquisition system parameter configurations and the groove undulation extraction algorithms are evaluated in terms of the SNR, the total harmonic distortion (THD), and the standard deviation (STD) of the reproduced audio. It is desired that the audio display high SNR, low THD, and low STD. These results can be found in the Chapter 8 of Stotzer (2006).

The time taken for Stotzer's system to scan the film for a 78rpm record is 1.5 times of the audio duration of the record and the image processing time 2.5 times. The total time cost for reproducing the audio is thus 4 times the audio duration. The required data storage space for the extracted images is not clear in the literatures.

2.5.2.2 Efforts from Lawrence Berkeley National Laboratory

Lawrence Berkeley National Laboratory (LBNL) applies a two-step 2D imaging approach to reproduce mono audio from 78rpm disc recordings (Fadeyev and Haber 2003). First, the disc is fixed onto an X-Y object stage and laterally scanned by a video zoom microscope with CCD camera (Figure 2.29). As a result, the record surface of interest is captured by a series of overlapped 2D gray-scale images (Figure 2.30). In the second step, the images are merged and analyzed by software to extract the complete groove undulations and finally the audio is derived from the undulations.

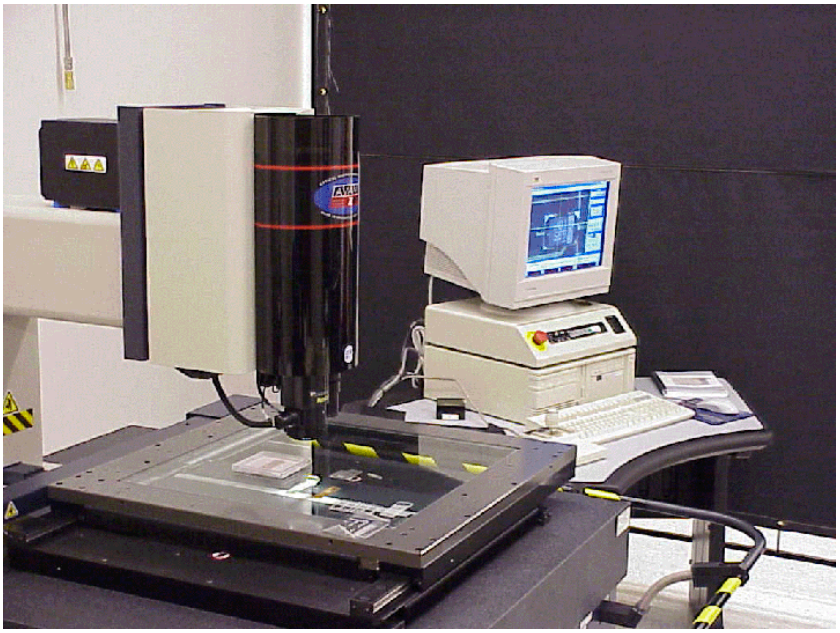


Figure 2.29: The video zoom microscope used by Fadeyev and Haber (2003) (image from Fadeyev and Haber 2003)

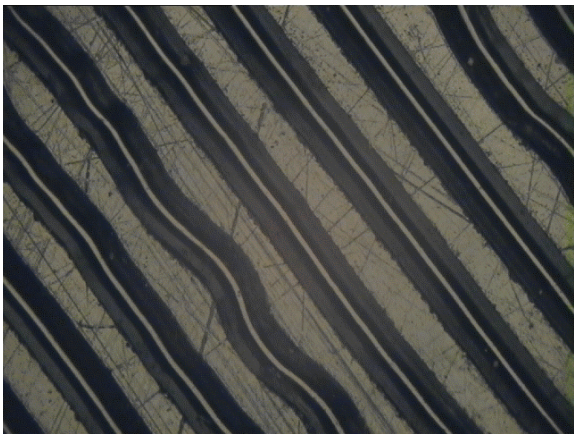


Figure 2.30: The obtained 2D groove image of a 78rpm disc record using the microscope in Figure 2.29 (image from Fadeyev and Haber 2003)

In the record surface acquisition step, a light illuminates the record surface area of interest to create light-and-dark intensity patterns in the captured images. The groove ridges and bottom portions appear bright while the sidewalls, scratches, and debris, dark. The microscope used to acquire the groove surface images provides a pixel resolution of $0.26 \times 0.29\ \mu m^2$ with an FOV of $700 \times 540\ \mu m^2$. After the surface area of interested is measured, the stage moves to the next adjacent location for the microscope to measure it. A typical image containing segments of adjacent groove revolutions is shown in Figure

2.30. Unlike early OAR systems and Stotzer's system, Fadeyev and Haber use no custom systems to rotate the records during the scanning—the groove segments in the resulting images are thus not as straight in all the obtained images as in those from Stotzer's system (Figure 2.28). Also, different from Stotzer's approach, Fadeyev and Haber illuminates the groove ridges and bottom so that such areas are the bright portions in the resulting images, an approach that differs from Stotzer.

In the image analysis step, Fadeyev and Haber derive the continuous groove bottom from the acquired images as the cutting stylus trajectory. They then treat the edge between the bottom and the sidewall as the groove undulation of a mono disc record. The acquired images are first converted into a polar coordinate system (R, φ) centered at the disc center, where R is the distance from the disc center to the pixel of interest, and φ the azimuth. The groove undulation is then defined as the radial deviation of the groove bottom from the unmodulated trajectory about the origin (the disc center). The boundaries between the illuminated groove bottom and the sidewalls are then extracted through 2D-image edge detection. The distances from these boundary points to the disc center result in a 1D edge radius series, which are then treated as the groove undulations. Fadeyev and Haber tried to reduce the noise in the undulation by assuming that the groove sidewalls on both sides of the groove bottom are equally wide. The inner sidewalls are bounded by the inner edges of the groove ridge and bottom, while the outer sidewalls are bounded by the outer edges of the groove ridge and bottom. The sidewall widths determined by these two pairs of edges on both sides of the groove bottom are then compared to filter out noises. The groove undulation segments in obtained images are then merged into a continuous groove. Small mechanical shifts in the stage movements can lead to slightly mis-aligned images, which can cause discontinuities in the extracted groove undulation and eventually results in audible clicks in the audio output. This is corrected by correlating the overlaps between the adjacent images in the measurement sequence and eliminating the shifts. Similar to Stotzer's system, the audio information is encoded in the groove undulation according to the constant velocity scheme. Numerical differentiation is then performed by first fitting a piece-wise polynomial on the 1D undulation signal and then taking its first-order time derivative. Different from Stotzer's approach, the RIAA de-emphasis is not routinely performed on

the final audio output, but is used in the evaluation of the audio quality in Fadeyev and Haber's approach.

Since the estimation of the disc center position is often inaccurate, the directly extracted groove undulation radii may introduce pitch fluctuation similar to the wow noise discussed in Section 2.2.3. Fadeyev and Haber solved this problem by fitting the obtained raw undulation signal to a spiral curve to minimize the center estimation error.

The reproduced audio quality is qualitatively compared to the turntable-digitized audio and the remastered CD audio of the same content. The quality of the reconstructed sample was judged by Fadeyev and Haber to be better than the turntable version, but poorer than the CD version. The audio reproduced from the OAR system was found to have fewer clicks and pops and lower continuous noise level than those produced by the turntable. However, it was found to contain background hiss and unwanted low frequency modulation not found in the CD version (Fadeyev and Haber 2003, 16).

The file size of an image produced by this OAR system is about 1-megabyte data (Fedeyev and Haber 2003, 8). With the reserved overlap for correcting image mis-alignment, this requires 10^5 images to capture the complete grooves of a 10-inch 78rpm record, corresponding to 100 gigabytes of data (Fadeyev and Haber 2003, 8). It takes their system 50 minutes to scan about 1 second of recorded audio (Fadeyev and Haber 2003, 12).

In 2006, the system was improved by adopting a line scan CCD to accelerate the scanning process, a laser servo system to keep the microscope in focus on warped disc surfaces, and other software improvements (Cornell et al. 2007).

2.5.3 OAR with Confocal Microscopy

2.5.3.1 Efforts from Lawrence Berkeley National Laboratory

Using a low-mass confocal scanning probe, Fadeyev et al. (2005) from LBNL built a contactless surface metrology system for wax cylinders (Figure 2.31). Similar to 2D-imaging-based approaches, the audio reproduction workflow also comprises two steps: 1) record surface profile acquisition and 2) groove undulation extraction and audio reproduction.

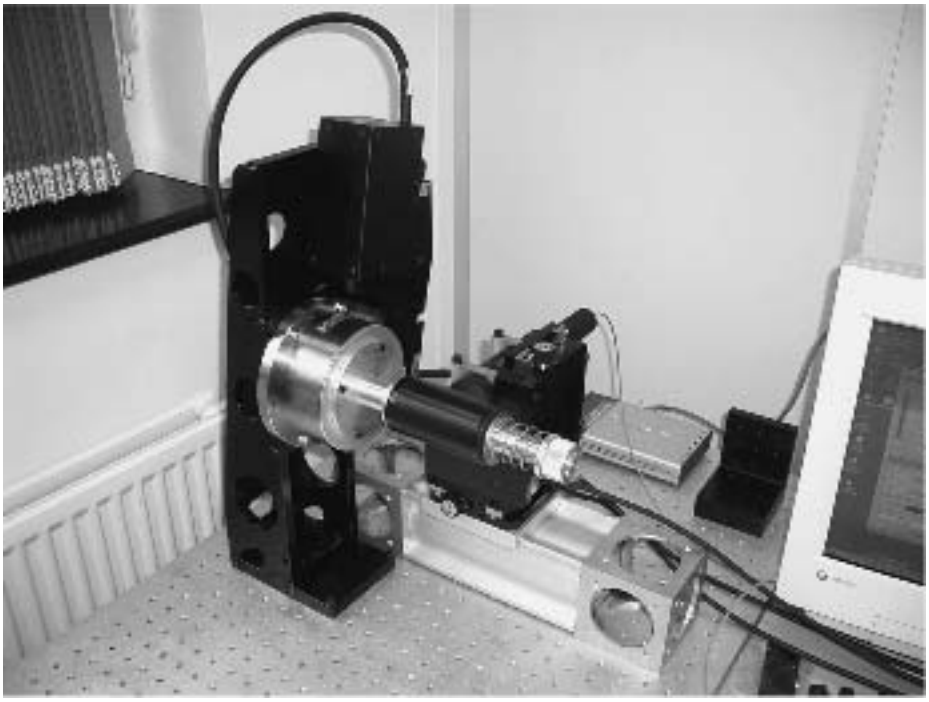


Figure 2.31: The cylinder scanning apparatus used in Fadeyev et al. (2005)

In the surface acquisition step, their goal was to obtain the 3D record surface profile containing both the ridge and the bottom of the cylinder groove by using a confocal microscope. The undulation containing the vertically modulated audio signals (see Section 2.2.1.2) can then be derived from the height information. This requires that the probe scan angularly along the circumference of the cylinder and linearly along the cylinder axis. They use a custom-built scanning apparatus to control the relative movement of the optical probe and the cylinder, and report that there are two optional scanning modes:

1) Angular-first mode: Similar to a cylinder player, as the cylinder rotates, the probe first scans a groove track along cylinder circumference and upon the completion of a full revolution, shifts linearly along the cylinder axis to scan the next groove track.
2) Linear-first mode: The cylinder first remains stationary; the probe first scans linearly along the cylinder axis, obtaining the groove portion of the same angular position across all the groove tracks and returns to the first track upon completion; the cylinder then rotates to the next angular position for the probe to start the next linear scan.

They compared the two acquisition modes to select the more efficient one. Since linear-first mode scans across the cross-sections of the groove tracks at each cylinder angular increment, the groove ridge and bottom profile at each angular increment is captured more easily than with the angular-first mode. Linear-first mode is therefore adopted in the system. However, there is an extra cost of reorganizing the scanned data into a time series, and a reset time cost for the probe to return to the starting end of the cylinder at each angular increment. After the scanning is complete, the groove tracks are unwrapped into flattened series of images that contain the surface height profile along the cylinder axis. The obtained 3D surface profile of a portion of a cylinder record is shown in Figure 2.32.

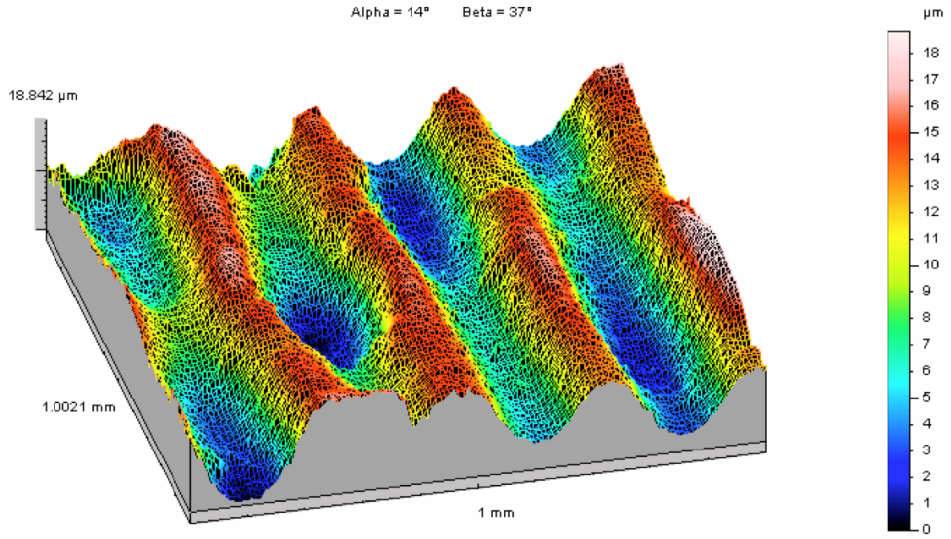


Figure 2.32: 3D surface profile of a portion of a wax cylinder record scanned by the apparatus in Figure 2.31 (image from Fadeyev et al. 2005)

In the image analysis step, the software iterates across the resulting data from the linear scanning across cylinder axis and finds the local minima (groove bottom) and maxima (groove ridges) of the measured surface heights at each angular increment; these local height extrema from different groove tracks at individual cylinder angular increments are then reorganized and extracted as time-series. According to Fadeyev et al., the cylinder grooves are cut with parabolic-shaped cutting stylus, and the height extrema is then fitted to a parabolic function by a piece-wise polynomial-fitting to extract the groove undulations. The undulation can be derived from the time-series of the distances

between the groove ridge and the bottom (top-bottom-combined scheme), or the bottom heights only (bottom-only scheme). These two options are discussed later.

The audio is reproduced after various filtering is performed on the undulation time-series. Missing points from the corrupted grooves are linearly interpolated from the neighbouring points. A high-pass filter is applied to remove frequency components below 20Hz. A low-pass filter with a cut-off frequency around 5kHz is also applied to remove the artefacts caused by the groove surface discontinuity introduced from the cutting process of the adjacent grooves. They assume that no dust or debris is present in the grooves and do not try to clean the records or correct serious errors such as large scale discontinuities created by dust or debris.

To evaluate the reproduced audio, they used an Edison Blue Amberol cylinder (Fadeyev et al. 2005, 491). Qualitative comparisons were made between the audio derived from the groove bottom-only, the top-bottom-combined schemes, and mechanically-digitized (via stylus) audio signals. It was discovered that the audio results from the three methods are qualitatively similar to each other. The bottom-only version contained more clicking noise the other two, while the mechanically-digitized version contained more unwanted low-frequency content.

The adopted confocal microscope from 2003–2005 has a lateral resolution of 100nm, and a vertical resolution of 10nm. The circumferential angular step size is 0.01º (5µm for the test cylinder), corresponding to an equivalent digital audio sampling rate of 96kHz for 160rpm cylinder recordings (Fadeyev et al. 2005, 492). The linear scan speed is 3mm/s, resulting in a scan time of 1.13 hours to capture 1 second of recorded audio and 185.2 hours for the entire cylinder record. In the reported experiment, audio from 10mm along the cylinder axis with full rotation is reproduced, which corresponds to about 29.25 seconds of recorded audio).

In 2007, a new white-light confocal line sensor (see Section 2.4.2) was introduced into their system (Haber 2008). Since the sensor can measure 180 points simultaneously, the total time for scanning an entire Edison cylinder can be reduced to of order one hour or less.

2.5.3.2 The Sound Archive Project

Boltryk et al. (2007) and Nascè (2009) from the Sound Archive Project at the University of Southampton used a confocal-microscopy based OAR system similar to that of Fadeyev et al. (2005) for cylinder recordings. The microscope used was a white-light confocal microscope. The scanning device and the scanning procedure were largely the same as the aforementioned approach of Fedyev et al. (2005). Significant differences between the two approaches include: The optical sensor in use was a white-light confocal sensor; the distance between the sensor and the cylinder surface was adjustable to allow for vertical scan; the system aimed at preserving the entire groove contour including the ridges, sidewalls, and bottom of the grooves.

Like the LBNL system, this system provides 0.01º angular scanning resolution, equivalent to 96kHz audio sampling rate for the 160rpm cylinders (Nascè 2009, 80). The vertical resolution of the system is 10nm (Nascè 2009, 81). Nascè estimated that the vertical dynamic range of the test cylinder groove (containing 1kHz sine wave) is 10µm and that the equivalent audio bit-depth that the system can provide is under 10-bit (Nascè 2009, 81). Experiments were made for wax cylinders in both good and poor physical conditions, and tinfoil cylinder recordings. The OAR results for the recordings are qualitatively compared with versions digitized from mechanical playback systems. The OAR system was found to have better signal fidelity and less frequency distortion than mechanical systems. With the recordings of poor conditions, the optically reproduced audio version presented better signal continuity than that from stylus-based reproduction. Apparently, manual adjustments to help the microscope follow the correct groove trajectories were needed (Nascè 2009, 117).

Although mainly focused on cylinder recordings, Nascè (2009) also introduces a system with the same white-light confocal microscopy for 78rpm recordings and conducts preliminary experiments. A custom-built rotary scanner was used, which resembles a turntable. As the disc rotates, the microscope performs a linear scan starting from the disc center towards its outer edge. According to Nascè, the system has a theoretical vertical resolution of 2nm, capable of achieving 16-bit dynamic range. Since this is limited by the size of the white-light confocal sensor spot (7µm), the best achievable dynamic range is apparently less than 15bit (Nascè 2009, 90). The system is reported to a 96kHz equivalent

audio sampling rate (Nascè 2009, 91). The goal of Nascè (2009) was to preserve the full surface profile of the phonograph records. However, since the white-light sensor cannot extract the full groove sidewall (see detail in Section 2.3.4), instead of acquiring the full groove cross-section data, only groove bottom and top ridges were measured. The boundaries between the measured top ridges and the non-measurable sidewalls were then treated as the groove undulation and extracted through edge detection. Although the theoretical approach is introduced and a portion of extracted groove image is presented, the full workflow and audio rendering process are not discussed in their published papers.

The audio signals reproduced from sine waves were evaluated in terms of the SNR and THD. For 1kHz test sine signal, the best SNR achieved was around 31dB and the best THD was below 2.5% (Nascè 2009, 168). The result is compared with a stylus-digitized version. The stylus-digitized version had a better SNR but showed a higher THD than the OAR version.

2.5.4 OAR with Standard Optical Microscopy

Tian (2006) used an approach based on the standard optical microscopy for reproducing audio from 78rpm records. His goal was to "reproduce the audio signal from gramophone records using 3D scene reconstruction of the micro-grooves." He modeled the disc record groove undulation (both mono and stereo) as the groove sidewall orientation fluctuation along the groove spiral. He first obtained 2D intensity images at different heights of the record surface by using a standard wide-field microscope (as opposed to the confocal microscopes, which uses narrow-field sensors) and focusing it onto different vertical locations of the groove sidewalls. Then he derived the 3D information of the grooves from the intensity images by using software image processing algorithms. The capability of handling LPs is claimed, but its implementation details have not been published.

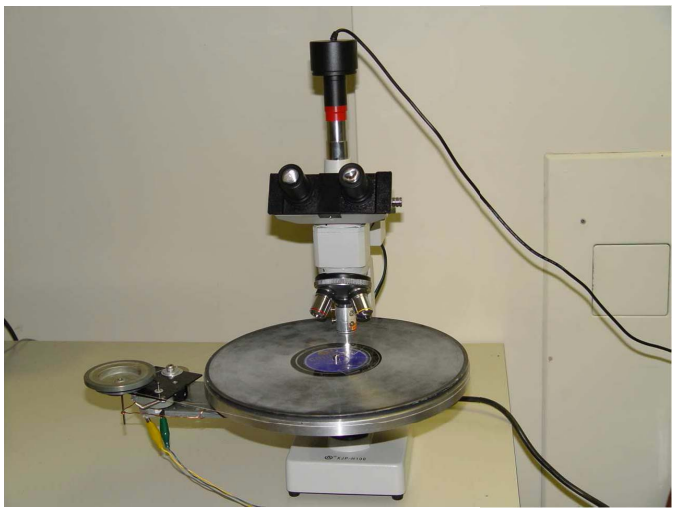


Figure 2.33: The disc scanning apparatus used in Tian (2006, 71)

In the record surface measurement step, Tian's goal was to obtain 2D intensity images of the record surface at different vertical positions of the same groove angular position along the groove orientation, which is required for the subsequent image processing step to derive the groove sidewall orientation. The metrology apparatus used consists of a wide-field optical microscope with a custom-enhanced dual-focal objective, digital camera mounted on the microscope, and a custom-built rotary stage (Figure 2.33). During the scanning, the record was placed on the slowly-rotating stage, while the record surface profile was captured and saved as gray-scale intensity images when the target surface passed beneath the camera FOV. The imaging principle of the standard optical microscope lens system is shown in Figure 2.34, which conforms to the relationship below:

$$\frac{1}{u}+\frac{1}{v}=\frac{1}{f} \qquad \text{Equation 2.5}$$

in which f, v, and u denotes the focal length of the microscope, the object distance (the distance between the microscope lens and the record surface to measure), and the image distance (the distance between camera and the microscope lens), respectively. When u is fixed, multiple surface images at different vertical positions (variable v) can be obtained by adjusting the distance between the objective lens and the record to bring the target surface into focus.

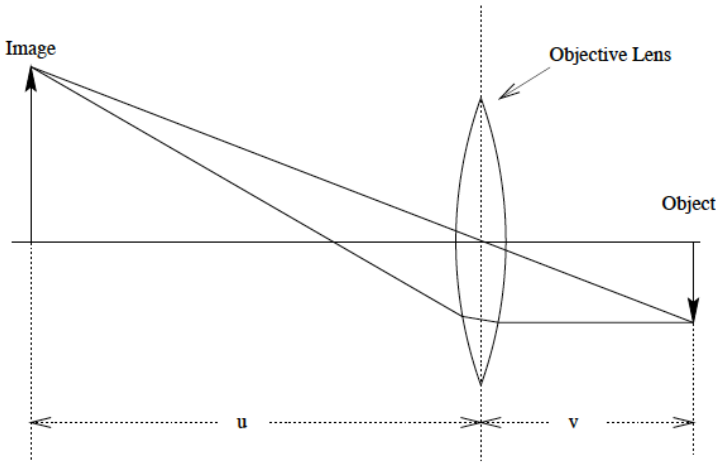


Figure 2.34: The illustration of the standard optical microscope light path (image from Tian 2006, 72)

Tian explained a dilemma that prevented his regular microscope from capturing the groove sidewall orientation. To derive the sidewall orientation, the microscope lens was expected to provide a large depth-of-field in order to measure a sufficiently large depth range. However, to accurately measure the groove vertical information, a higher magnification was desired, which in turn demanded that the record be placed at a close distance to the microscope (a shorter focal length and depth-of-field). For example, in order to have the acceptable depth measurement error (7-8%) when measuring a 78rpm record of 120μm groove depth range, the maximum permitted focal length was 1.7mm (Tian 2006, 72). In Tian's experiments, the 4X low magnification objective corresponds to a 32mm focal length with a depth-of-field of 45μm, while the 40X high-magnification one a 3.9mm focal length with a depth-of-field of 1μm (Tian 2006, 73). Neither lens can satisfy the aforementioned contradictory requirements. To solve this dilemma, Tian then inserted a thin cover glass between the microscope objective lens and the record surface to create an additional focus, resulting in a dual-focal optical imaging system (Figure 2.35). This customized microscope was then capable of bringing two portions of the groove sidewall surface at different vertical positions into two foci within the same FOV. According to Tian (2006), although the system was incapable of capturing the full depth range of the grooves, it was able to acquire sufficient surface vertical information to derive the groove sidewall orientation. An acquired image is thus a 2D intensity image with multiple in-focus sub-areas that can provide 3D groove surface information (Figure

2.36). The in-focus intensity patterns are at different vertical position and nearly parallel to each other along the groove track. However, it is unclear in the literature how the groove was tracked either through hardware or software.

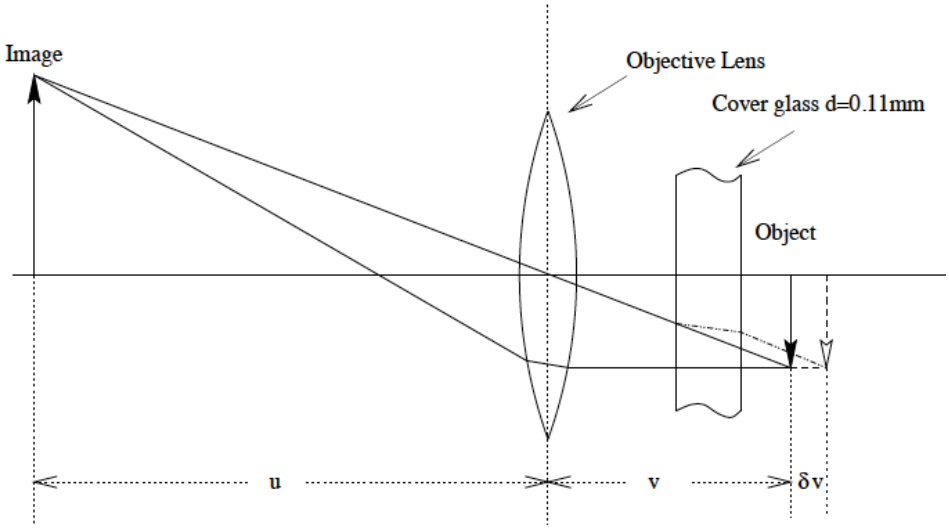


Figure 2.35: The illustration of the dual-focal lens system used in the disc scanning apparatus in Tian (2006, 75)

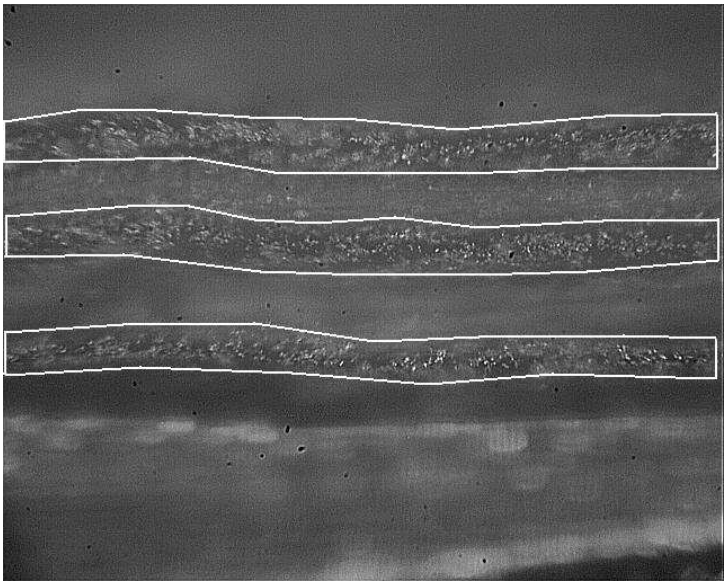


Figure 2.36: The groove image captured by the dual-focal disc scanning apparatus in Figure 2.35. The top and the bottom in-focus areas are at the same vertical location of the groove sidewalls while the one in the middle is at a lower vertical position (image from Tian 2006, 76)

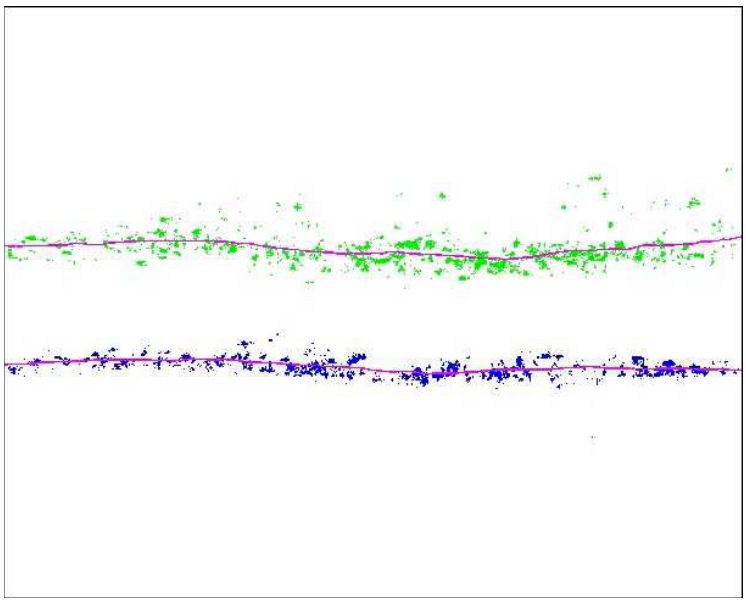


Figure 2.37: The intensity patterns and the thinned parallel centroid lines derived from the groove image shown in Figure 2.36 (image from Tian 2006, 82)

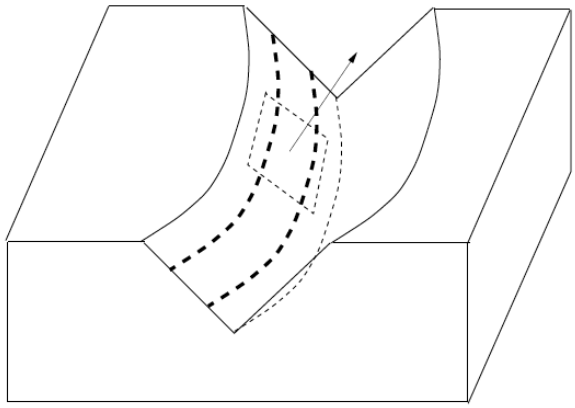


Figure 2.38: The illustration of the groove sidewall orientation derived from the sidewall 3D information in Figure 2.37 (image from Tian 2006, 83)

In the image-processing step, the groove undulation was extracted and then the audio was obtained from the undulation. The groove undulation is defined as the sequence of the groove sidewall orientation (normal of the sidewall plane) at each angular increment of the groove relative to the disc center, instead of the edges of either the groove ridge or bottom. To derive the groove sidewall orientations, first the in-focus intensity patterns in the intensity images were identified by finding the optical flow through Hough transform. The resulting stripe-like patterns were then thinned, and finally the local plane spanned by the two thin intensity line patterns was derived and its normal

was determined as the sidewall orientation at the local angular position of the groove (Figure 2.37 and Figure 2.38).

Tian claims to achieve the equivalent audio sampling rate of 2404.71kHz (Tian 2006, 81). This requires about 5×10^4 images of an FOV of 640×480 pixels to represent a two-second audio signal (Tian 2006, 54). The image acquisition time is unclear in the literatures. It takes two days for four computer workstations to generate the image representation of the two-second audio. No automations were used in their system and most experiments were conducted by hand.

2.6 Summary

This chapter reviewed the literature on the phonograph recording technology, the optical surface metrology, and the previous OAR approaches from other researchers were reviewed. The means by which audio is encoded into grooves discussed in Section 2.2 provided clues for devising optical methods to extract the groove surface profile accurately and efficiently. The optical surface metrology techniques and the previous OAR approaches demonstrate how audio can be extracted by optically measuring the record surface profile. These OAR approaches exploit various optical surface metrology techniques and aim at different recordings. Early OAR systems for disc recordings aimed at real-time playback and require tracking the grooves while playing the record. They thus cannot handle broken or seriously damaged records. Later systems exploited generic surface metrology hardware with custom additions to acquire record surface profile images, which are then analyzed to extract and track the groove by using software methods. To the best our knowledge, no efforts have been made to reproduce stereo audio from 33rpm microgroove LP records by using the WLI techniques. In the next chapter, we present our record surface profile acquisition system using WLI.

Chapter 3 The Record Image Acquisition

3.1 About the Chapter

We adopt a generic WLI (white-light interferometer) microscope to acquire the 3D surface profile of a stereo LP (long-playing) record as a set of images. This chapter first briefly introduces the image acquisition hardware and software. We then describe the technical details of measuring a single FOV (field of view) of the record surface area and a custom workflow for acquiring images of larger areas of the record surface by using the acquisition hardware and software. Finally, the procedure for stitching these images together in order to correctly track the record grooves is discussed.

3.2 The Image Acquisition System

The Wyko NT8000 series WLI microscope (Veeco Metrology Group 2003) is selected as our image acquisition system. It consists of a microscope, an X-Y sample stage, and a computer. Wyko Vision, the microscope measurement and analysis software package, is pre-installed on the computer.

3.2.1 The Wyko NT8000 Series Microscope

Figure 3.1 shows the Wyko NT8000 series microscope used to perform the research described in this dissertation. The most important components are the Modular Optics Assembly (MOA), the motorized sample stage, and the vibration-isolation table. Figure 3.2 shows the microscope with a disc record loaded on the stage.

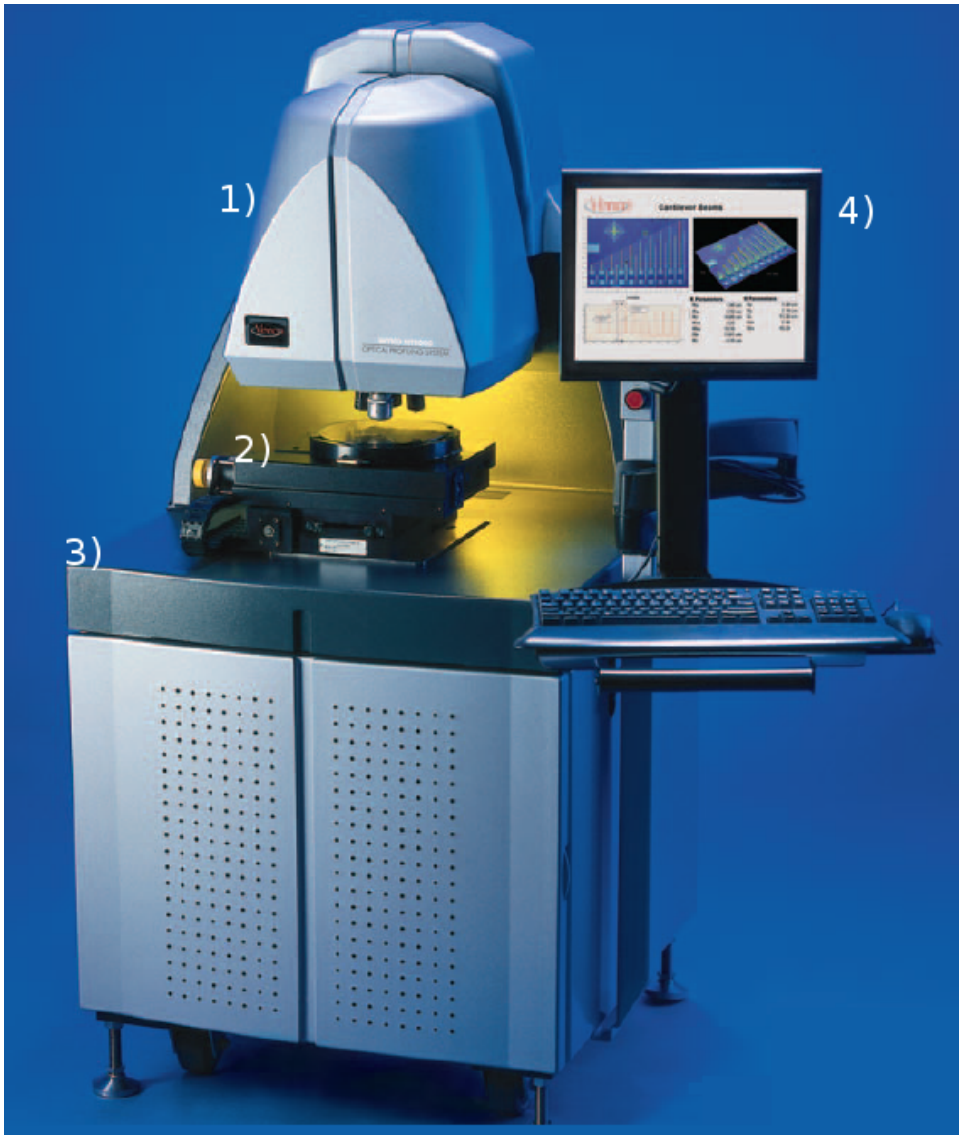


Figure 3.1: The Wyko NT8000 series microscope, including 1) the MOA, 2) the sample stage, 3) the vibration-isolation table, and 4) the control computer with the pre-installed Vision software (image from Veeco Instruments Inc. 2006)

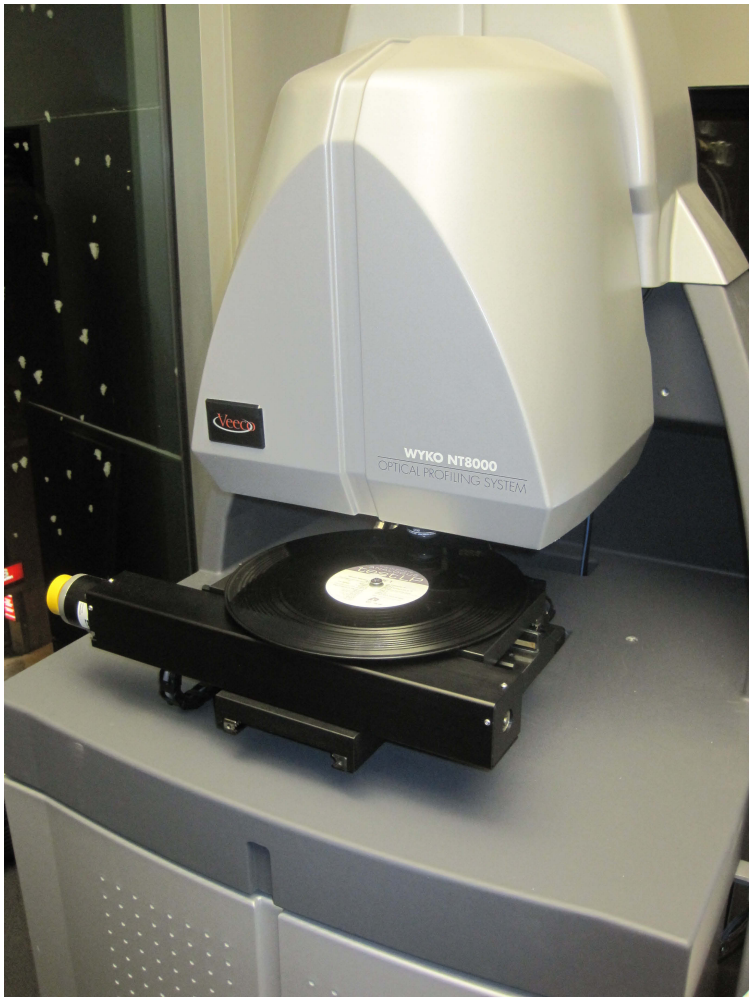


Figure 3.2: Our NT8000 microscope with a disc record loaded on the stage

3.2.1.1 The Modular Optics Assembly (MOA)

The MOA is the core measurement component of the microscope. It contains a CCD camera (640x480 pixels) and a Multiple Magnification Detector (MMD), which holds the white-light interferometer objectives with FOV lenses. The objectives are attached to the bottom of the MOA on a turret (Figure 3.1).

Our microscope has two optional objectives: a Michelson interferometer objective of 5X magnification, and a Mirau interferometer objective of 50X magnification. They support both PSI (phase-shifting interferometry) and VSI (vertical-scanning interferometry) modes. In addition, our microscope is equipped with three FOV lenses, which can be used in combination with either objective to increase or decrease the overall magnification of the microscope. The captured surface profile image within an FOV is a 640x480 pixel matrix. The real-world size (or the FOV size) of the image is determined

by the overall magnification of the microscope. See Table 3.1 for the available magnification options and the corresponding real-world FOV sizes. Using the VSI, the vertical resolution for both interferometers is 5nm, and with the PSI is 0.1nm (Veeco Instruments Inc. 2006).

Objectives		**5X**	**50X**
FOV	**0.55X**	2.53 x 1.90	0.25 x 0.19
	1X	1.27 x 0.95	0.13 x 0.10
	2X	0.63 x0.48	0.06 x 0.05

Table 3.1: Field of view size (mm x mm) with available objective-FOV combinations of our Wyko NT8000 series microscope (Veeco Metrology Group 2003)

The Michelson and Mirau interferometers share the basic operation principles described in 2.4.3. According to Kino and Chim (1990), the Michelson interferometer (Figure 3.3) has the advantage of being insensitive to measurement distortion caused by spherical and chromatic aberrations of the objectives in the vertical direction because the lenses used in the reference and test beam paths are identical, and the aberration effects are cancelled out during the differentiation of the test beam and the reference beam. They do, however, suffer from mechanical vibrations that may be introduced into both the reference and the test beams along their long beam paths. On the other hand, as shown in Figure 3.4, the Mirau interferometer overcomes the mechanical vibration problem by using a special arrangement of the reference mirror and the microscope objective so that the reference and test beams share a common path over most of their lengths. In this manner, any disturbance introduced to both beams is largely cancelled out during differentiation. The Mirau interferometer thus provides higher resolution than the Michelson interferometer. However, as its size increases, the Mirau interferometer suffers more from objective spherical aberration than the Michelson interferometer. The FOV size of the Mirau interferometer is therefore usually smaller than that of the Michelson interferometer.

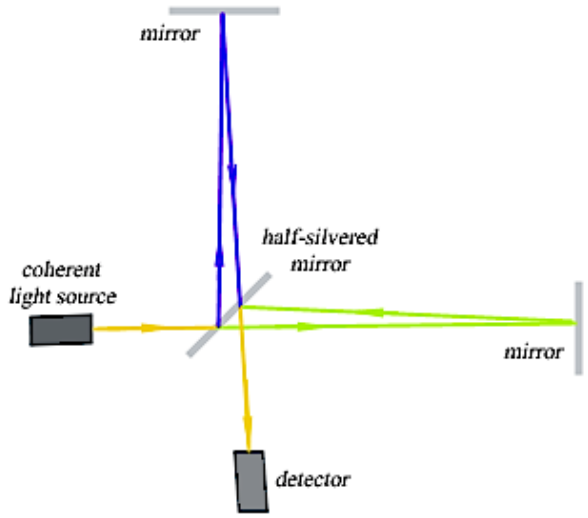


Figure 3.3: The Michelson interferometer (image from Wikipedia 2011)

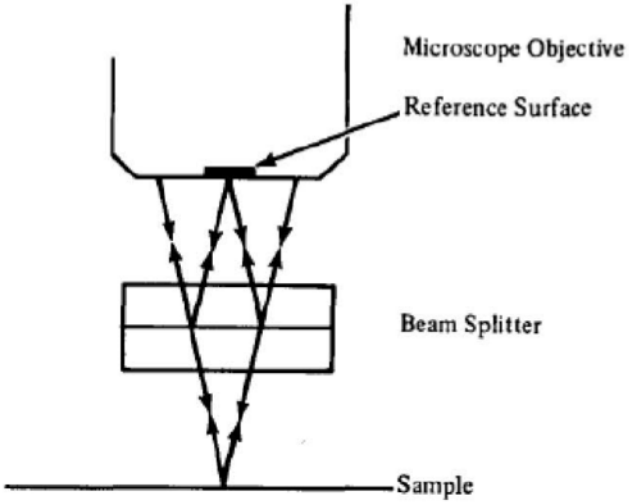


Figure 3.4: The Mirau interferometer and its disturbance cancellation mechanism (image from Bhushan et al. 2002)

3.2.1.2 The Motorized Sample Stage

The motorized sample stage (Figure 3.1) is the component that hosts the target sample and allows the microscope to measure the different portions of a sample surface larger than its FOV sizes by transporting the target portion into the FOV of the microscope. The sample stage used in our microscope system is an X-Y stage that moves front-and-back or left-and-right. This is different from many previous OAR approaches (Stotzer 2006; Tian 2006; Nascè 2009), which adopt rotary stages to unwrap the record grooves while

acquiring the record surface profile. Our type of stage was also adopted by Fadeyev and Haber (2003). At the time when we purchased the microscope, no stitching software (see Section 3.5) was available with the rotary stage from the microscope manufacturer. We thus decided to adopt the X-Y sample stage to acquire the record surface profile via a 2D linear scanning trajectory and then unwrap the grooves in the scanned images by using software algorithms. The technical details of the stage in use are shown in Table 3.2.

Bi-directional repeatability	**Accuracy**	**Flatness of travel**	**Straightness of travel**
±5μm	±5μm / 25mm	2μm / 25mm	2μm / 25mm

Table 3.2: The technical specification of the motorized sample stage of our microscope (Veeco Instrument Inc. 2006)

3.2.1.3 The Vibration Isolation Table

There is a vibration-isolation table (Figure 3.1) underneath the stage with a nitrogen-gas-filled cushion to ensure that the stage remains level even if mechanical vibrations occur during the measurements. The presence of the cushion, however, introduces another issue: when the stage moves from one location to another, the inertia could cause the stage above the cushion to become slightly imbalanced. As soon as this occurs, the vibration-isolation table is designed to automatically rebalance the stage, which can take a short time to complete. Any measurement made during the rebalancing period may therefore yield inaccurate height results. To avoid this problem, a small time interval needs to be reserved between consecutive measurements over different target record locations. According to our experiments, the minimum-required waiting interval is approximately 1.5 seconds.

3.2.2 The Custom Fixture for the Disc Phonograph Records

It is desirable that the disc record remain still relative to the stage at all times during the record surface profile acquisition in order to avoid measurement errors. We therefore secured the record by using a screw to tightly attach it to the stage through the center spin-hole (see Figures 3.5–3.6). We also placed two washers on both sides of the record to further prevent the disc from drifting (Figure 3.7). The washer underneath the record also prevents direct contact with the stage on the other side. After fixing the disc record

onto the stage, we can start making measurements to acquire the surface profile image of the record.

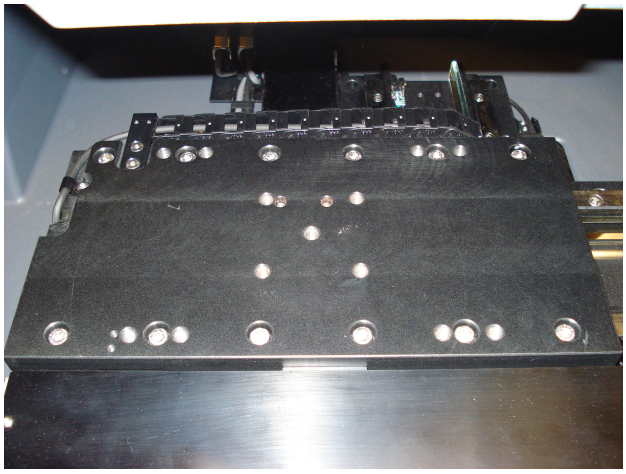


Figure 3.5: The screw holes on the sample stage of our microscope

Figure 3.6: The disc record fixed onto the stage with the screw and washers

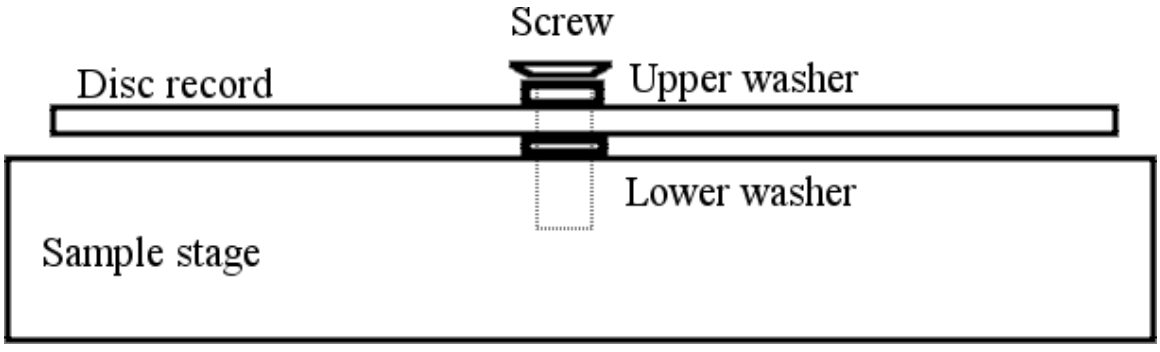


Figure 3.7: The disc hosting apparatus used in our acquisition system

3.2.3 The Wyko Vision Software Package

All the measurement operations of the microscope are performed by the bundled Wyko Vision software package (Version 3.6). The important functionalities used in the research described in this dissertation can be summarized into the following categories: manual measurement control; VSI parameter configuration; data storage; data visualization and basic analysis; advanced sample navigation; grid-based automatic measurement; and image stitching. These functionalities are introduced in the remaining sections of this chapter in the context of the tasks associated with acquiring record surface profile images. We discuss the use of these functionalities and our own strategies and algorithms; how to manually obtain a record surface profile image (Section 3.3); how to semi-automatically acquire a large area surface profile as a composite of FOV-sized images that covers the desired groove range (Section 3.4); and how to stitch these images together (Section 3.5).

3.3 Manually Obtaining a Record Surface Profile Image

In this section we briefly present the typical workflow for manually acquiring a surface profile image of a disc record. This dissertation focuses primarily on a stereo audio system test LP record (Analog Productions n.d.) that contains standard stereo test signals. The specific signal that we extract is two revolutions of groove (about 3.4seconds), of which the left channel contains silence and the right channel contains a 1kHz sine wave.

To obtain a record surface profile image, we first define our goal for this task. As discussed in Section 2.4.4, the groove sidewall angle of a stereo LP record usually exceeds the angular tolerance of the WLI. Consequently, the groove sidewalls are largely missing in the image acquired by our acquisition system. We nevertheless decided to explore the possibilities of reproducing the stereo audio signal by just acquiring the groove ridge and bottom. The goal of obtaining an FOV of the disc record surface image is therefore to capture just the groove ridges and the bottom within the FOV at a sufficient resolution for reproducing the audio within the shortest-possible amount of time.

To manually obtain a record surface image, the typical workflow is as follows:

1) Fix the record onto the host apparatus selected in Section 3.2.2, which protects the grooved disc surface from contacting the sample stage and prevents the disc from drifting during the measurements.

2) Select the combination of the objective and the FOV lens.
3) Navigate to the record surface location of interest.
4) Roughly focus the microscope onto the record groove.
5) Configure the VSI measurement parameters.
6) Launch the measurement.
7) Observe and save the resulting image.

Steps 2–7 are all performed with the Vision software. We next show a typical record surface profile image and introduce its basic properties, and then discuss the workflow for obtaining the image.

3.3.1 The Typical View of the Acquired Record Surface Profile Image

Figure 3.10 shows the 2D contour view of an FOV-sized record surface profile image under 10X magnification. The pseudo-colours correspond to the depth values of the surface. The red portion represents the groove ridges, the blue the groove bottom, and the black the groove sidewalls. As discussed in Section 2.4.4, the groove sidewall angle of a stereo LP record usually exceeds the angular tolerance of the WLI. Consequently, the groove sidewalls are largely missing in the image acquired by our acquisition system. The missing data is defined as invalid data, or not-a-number (NaN). According to our experiments, under 10X magnification there are usually 30% invalid data in a record surface profile image. However, it will be shown in Chapter 4 that the invalid data regions can still provide useful information for extracting the grooves from the images.

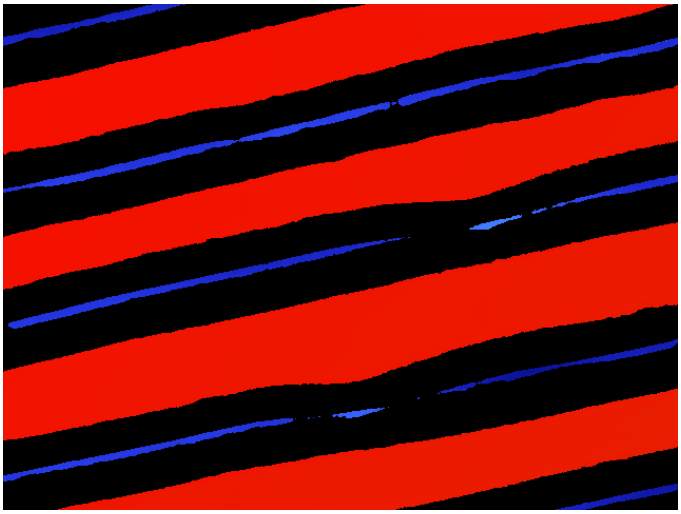


Figure 3.8: The typical 2D contour view of an FOV of record surface profile under 10X magnification

3.3.2 Selecting the Combination of the Objective and the FOV Lens

As examples, Figures 3.9–3.13 show the typical measurement result of one FOV record surface image under 5X, 10X, 20X, 50X, and 100X overall magnification.

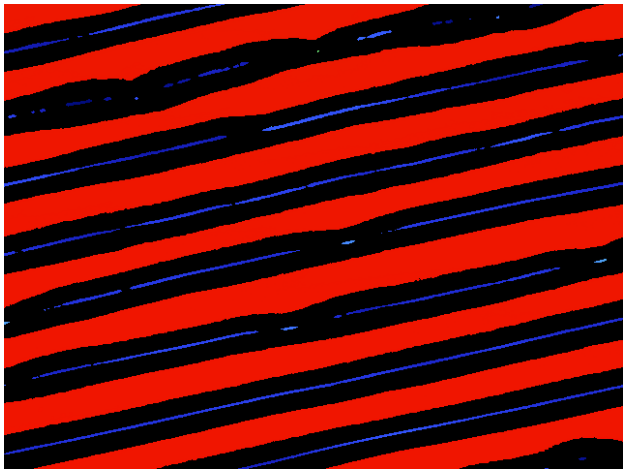


Figure 3.9: The acquired surface profile with 5X magnification

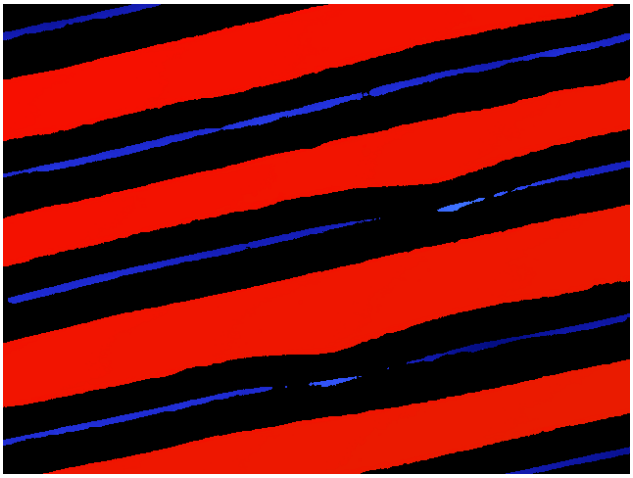


Figure 3.10: The acquired surface profile at the same stage position as in Figure 3.9 with 10X magnification

Figure 3.11: The acquired surface profile at the same stage position as in Figure 3.9 with 20X magnification

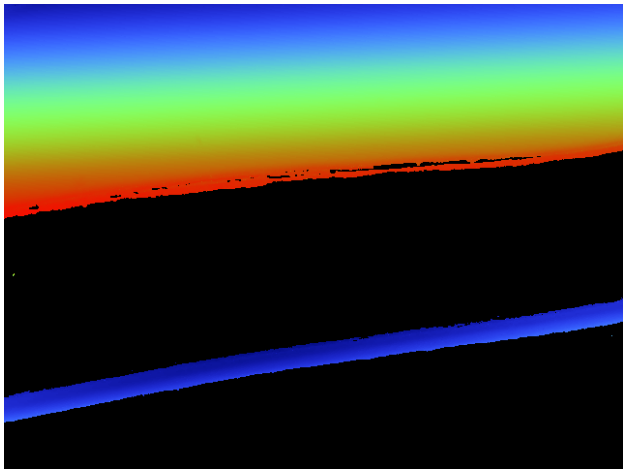


Figure 3.12: The acquired surface profile at the same stage position as in Figure 3.9 with 50X magnification

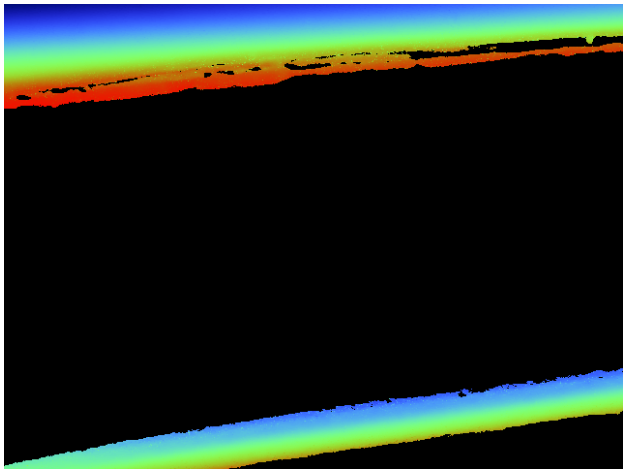


Figure 3.13: The acquired surface profile at the same stage position as in Figure 3.9 with 100X magnification

3.3.3 Navigating the Record Surface

The Vision software provides a graphical user interface (GUI) that includes the Intensity and Focus Window (Figure 3.14), which allows the user to navigate across the sample surface and focus the microscope on the surface of interest through mouse or keyboard commands. The Intensity and Focus Window also serves as the microscope FOV monitor, through which the magnified sample surface under the microscope objective can be observed.

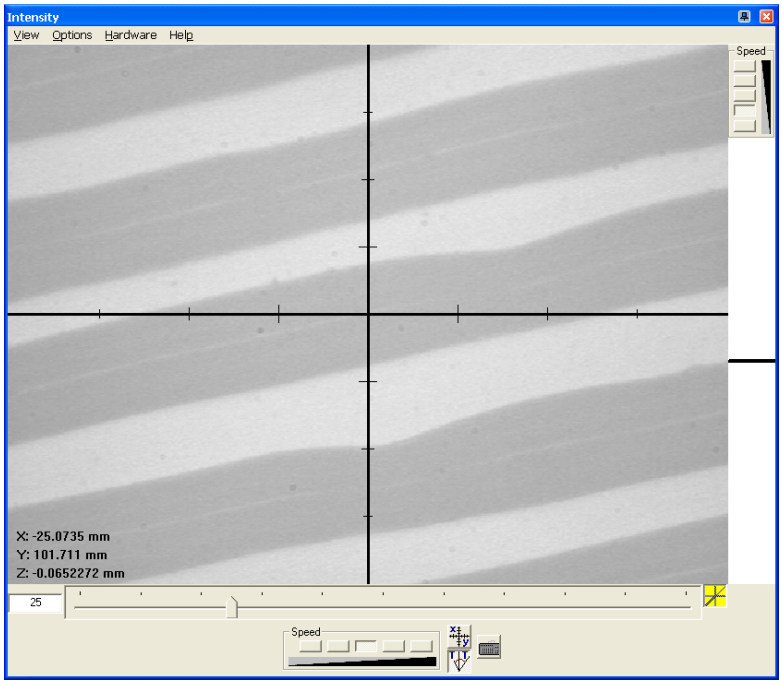


Figure 3.14: The Intensity and Focus Window of the Vision software

While monitoring the sample through the Intensity and Focus Window, the user can move the X-Y sample stage to bring the sample portion of interest into the microscope FOV. The Vision software offers two interaction modes within the Intensity and Focus Window: the sample navigation mode and the turret tip-tilt mode. The former allows the user to control the stage to navigate to any portion of the sample via either mouse gestures or keyboard shortcuts.

Manual navigation is useful when the sample area is small enough to explore and the location of interest is unknown. However, when the sample areas of interest are scattered across a surface significantly larger than the FOV, or repeated measurements are required for the same location, it is desirable that the user be able to specify the coordinates of the destination sample area and trigger the sample stage to automatically bring the location of interest into the FOV. The Vision software allows the user to perform this navigation through the Stage File feature.

The user can specify the X-Y coordinates of the location of interest in an ASCII file called the Stage File. This can be performed either by editing the text file with a standard text editor, or through the Stage File graphical interface (Figure 3.15). After this, it is then possible to move the stage to the destination by using the Stage File graphical

interface. Since the Stage File can be saved to the disk, sample locations of interest can be stored and retrieved.

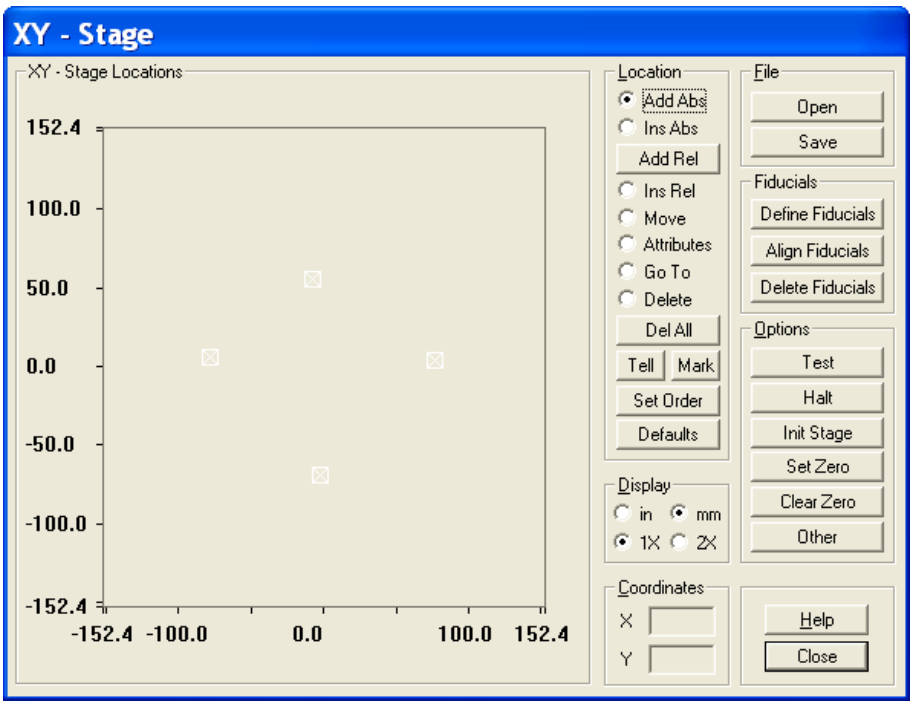


Figure 3.15: The Stage File GUI, with white dots representing the stage destinations

3.3.4 Focusing on the Groove

When a sample portion of interest is found, the user needs to start focusing the microscope by using mouse gestures within the focusing control panel of the Intensity and Focus Window. While focusing, the user needs to monitor the Intensity and Focus Window until high-contrast interference intensity patterns (fringes) become visible in the FOV (Figure 3.16). When the turret is at certain angles to the sample surface, the fringes may appear too narrow to observe (Figure 3.17). Under the other interactive mode of the Intensity and Focus Window, the turret tip-tilt mode, the user can adjust the relative angle between the objective in use and the sample surface until the well-contrast fringes become easy to see, as in Figure 3.16.

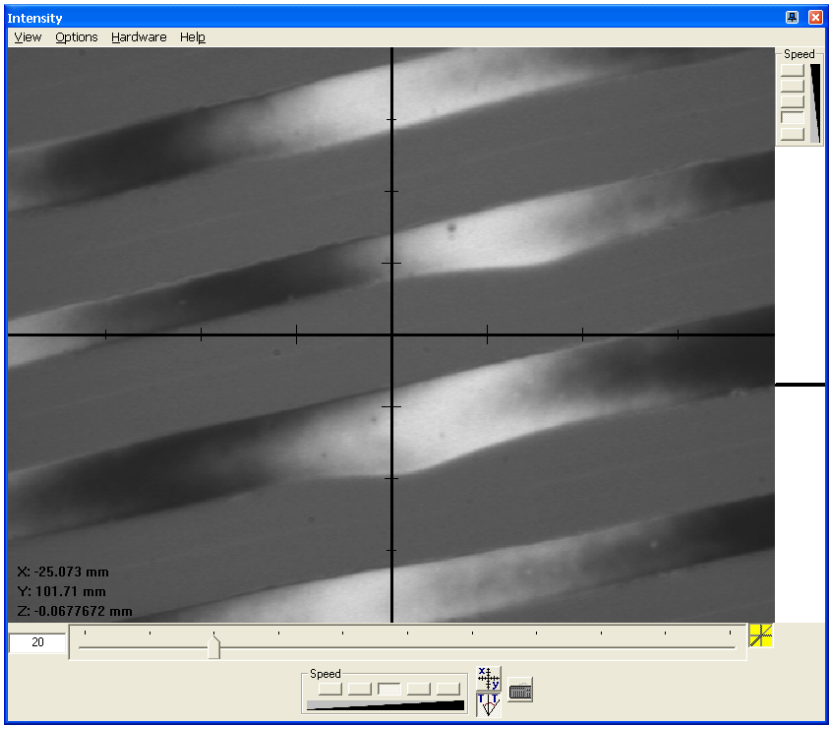


Figure 3.16: The fringe pattern that indicates that the microscope is in focus

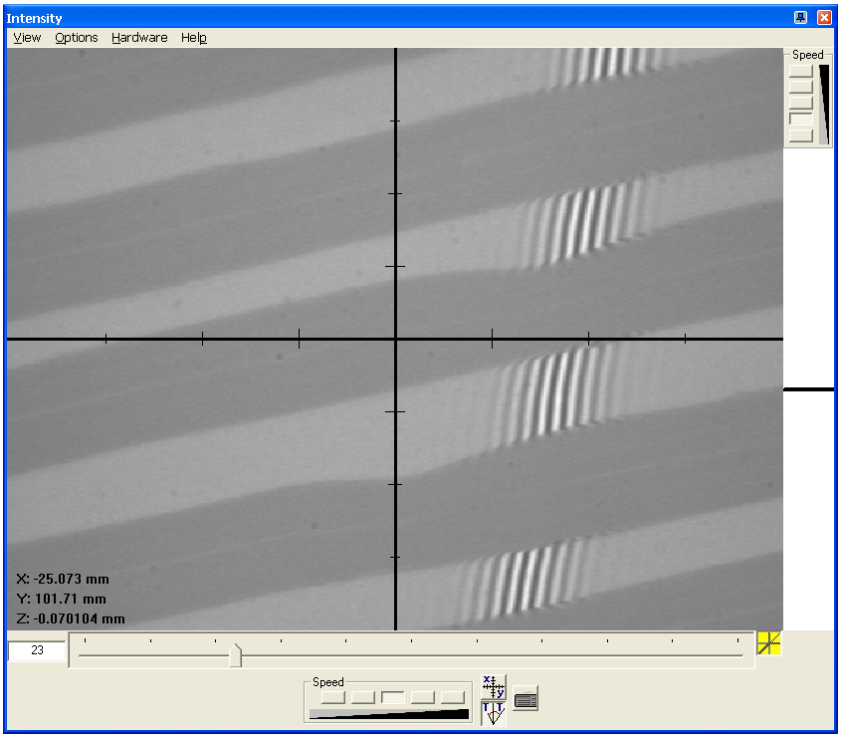


Figure 3.17: The fringe patterns of poor contrast due to inappropriate tilt of the turret

3.3.5 Configuring the VSI Parameters

Once the microscope is in focus, the user needs to manually configure the PSI or VSI measurement parameters before triggering to start a measurement to acquire the surface profile in the FOV of interest.

As described in Section 2.4.3.2, PSI is mainly used for measuring smooth surfaces while VSI is more suitable for measuring surfaces of large step sizes, such as the grooved record surface. VSI is thus used for the research described in dissertation. The important VSI parameters include: the light intensity, the scan range, the modulation threshold, and the scan speed.

The light intensity is used to set the intensity, ranging from 0 to 12VDC, of the light source, which is a tungsten halogen lamp (Veeco Instruments Inc. 2006). In order to obtain the fringes with a high contrast the ideal light intensity is expected to be as high as possible without saturating the interferometer, which can be monitored through the Intensity and Focus Window. Figures 3.18–3.20 demonstrate examples of the observed FOV and the measurement results under insufficient, sufficient, and saturated light intensities, respectively. Insufficient and saturated light intensities lead to data loss. It is usually suitable to use a fixed intensity to measure a flat surface. For a narrow and concave surface such as a record groove, however, a fixed light intensity may not ensure the best measurement results for both the top and bottom portions of the sample. As descried in Section 2.4, it is difficult for the light beams reflected from the groove bottom to reach the optical sensor; those beams that are able to reach the sensor often have low intensities and thus yield unsatisfactory results. Using a higher intensity helps better illuminate the bottom portion but tends to saturate the top portion. This indicates that, for narrow concave surfaces, a fixed intensity may not ensure the best measurement results for both the top and bottom portions of the sample. To solve this dilemma, the Vision software is capable of gradually increasing the light intensity from a lower value to a higher value, both specified by the user, as the microscope scans from the higher portion towards the lower portion of the sample.

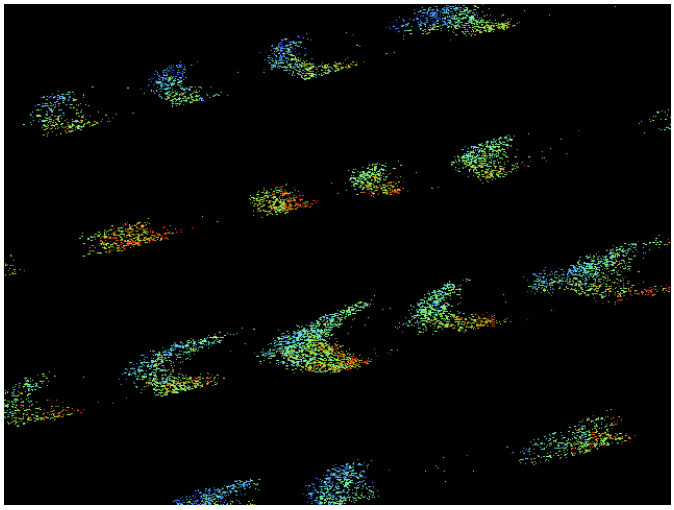


Figure 3.18: The surface acquisition result under an insufficient light intensity, leading to significant data loss

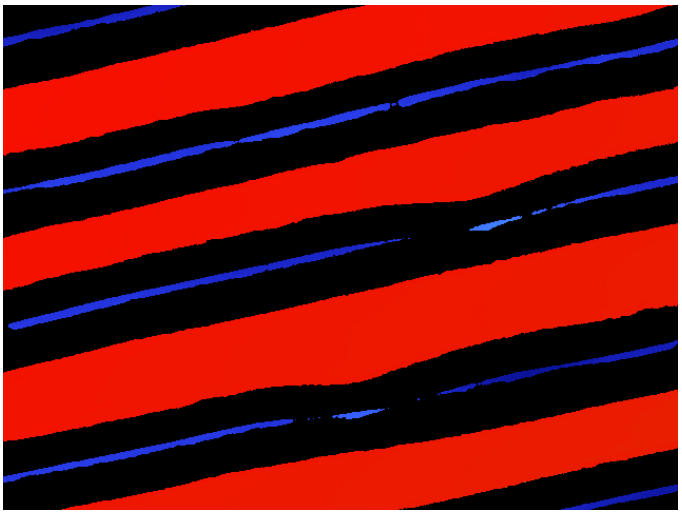


Figure 3.19: The same surface profile as Figure 3.18 acquired by using an appropriate light intensity

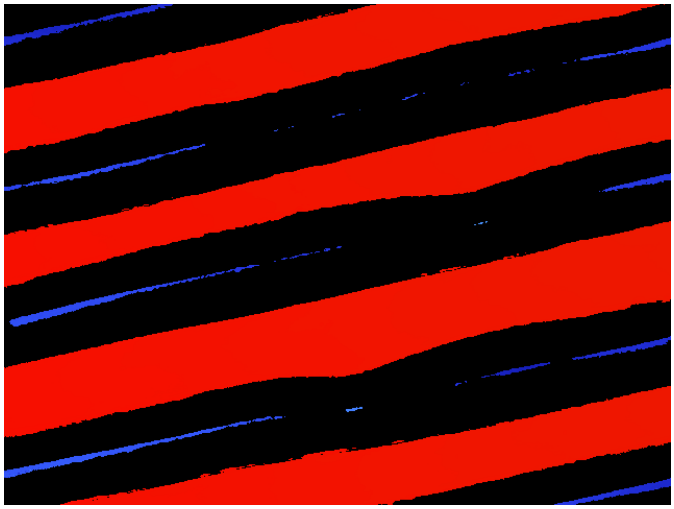


Figure 3.20: The same surface profile as Figure 3.18 acquired by using a saturated light intensity, leading to a partial loss of the groove bottom

The VSI scan range is the sum of two parameters: the back-scan length and the scan length. The focus position found by manually observing the fringes is not accurate and may be below the refined focus so, to avoid missing the refined focus during the downward scanning, the Vision software allows the turret to rise by a distance and start scanning from above the manually-found rough focus position. This distance is called the "back-scan length." The scan length is the distance from the rough focus position to the lowest scan position. The effective VSI scan range is thus the sum of the back-scan length and the scan length. When focusing on the groove bottom with insufficient back-scan length (Figure 3.21), the groove ridges will be missing from the acquired image (Figure 3.22). When focusing on the groove ridge with insufficient scan length (Figure 3.23), the groove bottom will be missing from the acquired image (Figure 3.24).

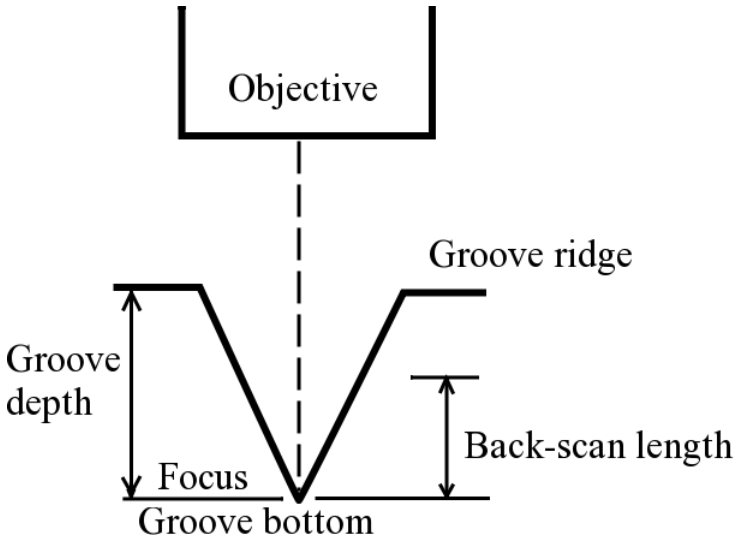


Figure 3.21: An insufficient back-scan length fails to cover the groove height range when focusing on the groove bottom

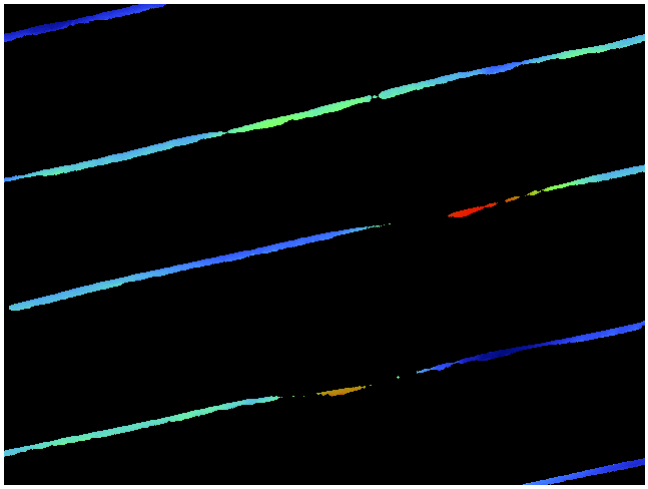


Figure 3.22: The insufficient back-scan length in Figure 3.19 leads to the loss of the groove ridges

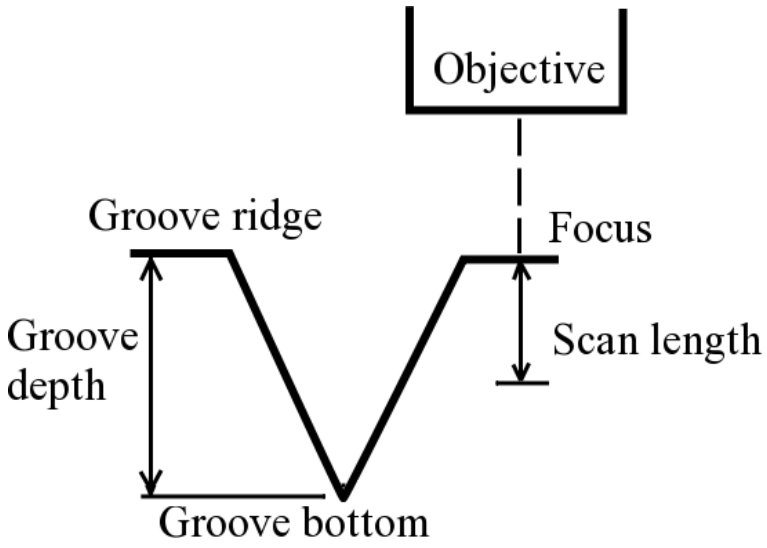


Figure 3.23: An insufficient scan length fails to cover the groove height range when focusing on the groove ridge

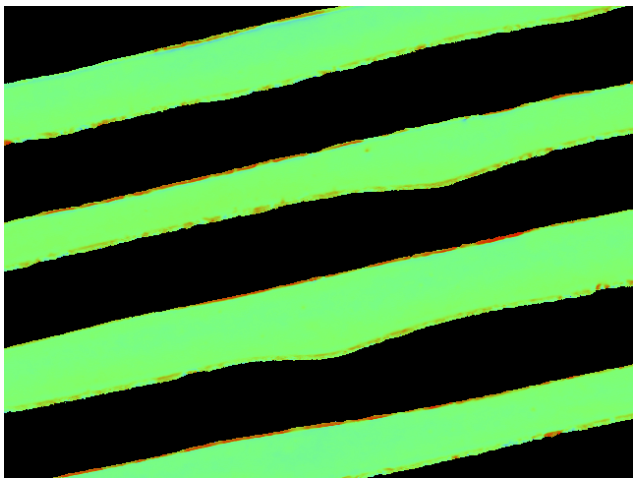


Figure 3.24: An insufficient scan length leads to the loss of the groove bottom, when focusing on the groove ridges

The modulation threshold controls the signal-to-noise ratio of the acquired surface profile image. High thresholds will therefore allow more data to be included in the result, but the result will be noisier than if lower thresholds had been used. Low thresholds reject more data and provide results that are cleaner than higher thresholds.

The vertical scan speed specifies how fast the VSI scan is performed. Fewer scan frames are used with higher scan speeds, resulting in faster measurement at the cost of measurement quality. Table 3.1 presents the speeds available in our microscope. See Figures 3.25–3.28 for the result of scanning one FOV sample under various scan speeds.

The adopted VSI parameters are listed in Chapter 6 when the experimental results are evaluated.

Scan speed options	**1X**	**3X**	**11X**	**23X**
Scan speed (µm/second)	5	15	55	115

Table 3.3: The available VSI scan speeds of our microscope

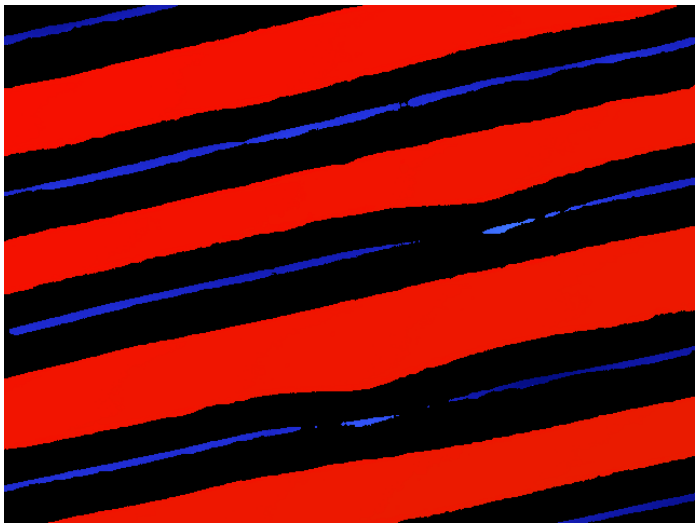


Figure 3.25: The surface profile acquired with 1X scan speed

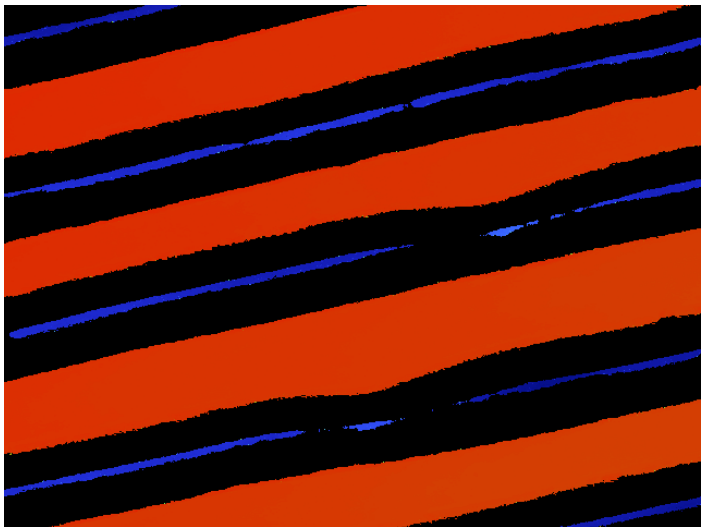


Figure 3.26: The same surface profile as in Figure 3.25 acquired with 3X scan speed

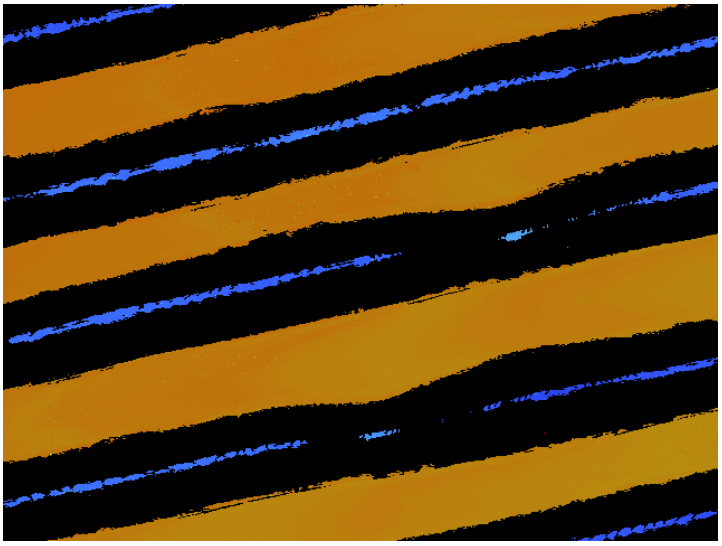


Figure 3.27: The same surface profile as in Figure 3.26 acquired with 11X scan speed

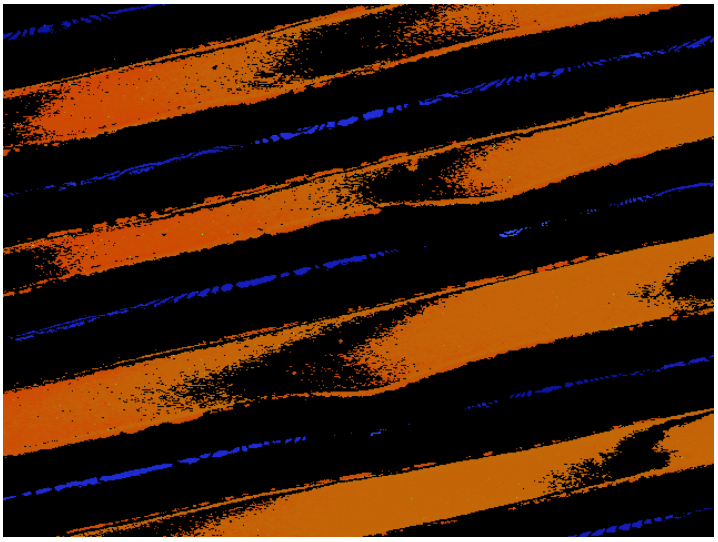


Figure 3.28: The same surface profile as in acquired with 23X scan speed

3.3.6 Observing Results with the Basic Data Visualization and Analysis Views

After configuring the VSI parameters, the user can have the Vision software perform a VSI measurement to acquire the record surface profile image of the FOV of interest. The immediate result is shown as a 2D contour plot on the computer screen with preliminary analytical data (Figure 3.29).

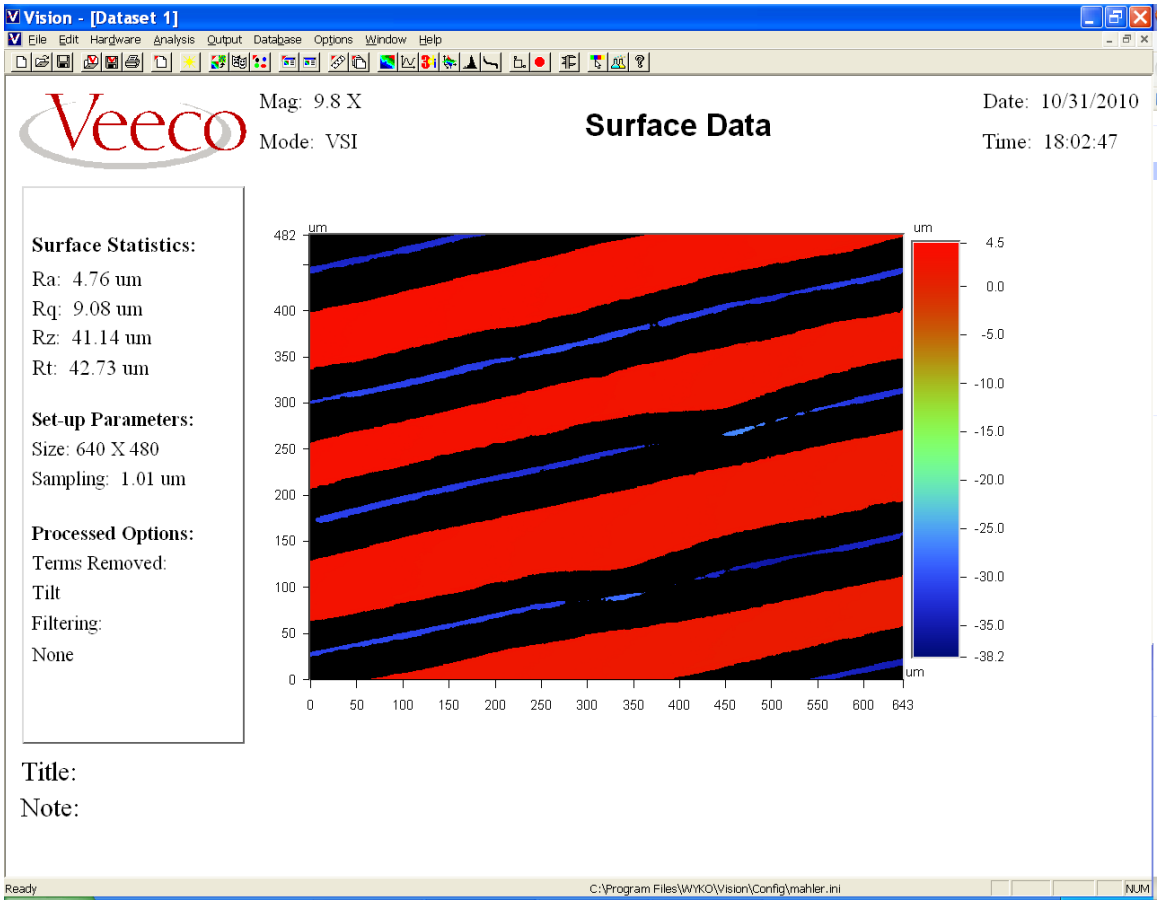


Figure 3.29: The 2D contour plot of the Vision software

The Vision software provides basic tools for visualizing the acquired surface depth information. Figures 3.29–3.31 show the 2D contour plot, the 2D cross-section plot, and the 3D view of the acquired sample surface. In the 2D contour plot and the 3D view, some generic statistical data are computed automatically, including the average roughness (R_a), the root-mean-squared roughness (R_q), the peak-to-valley distance (R_t), and the averaged peak-to-valley distance of the ten greatest such values (R_z). In the 2D cross-section plot, such values can be retrieved in terms of the cross-sections along the FOV X and Y dimensions through mouse interaction with the visualized FOV depth view.

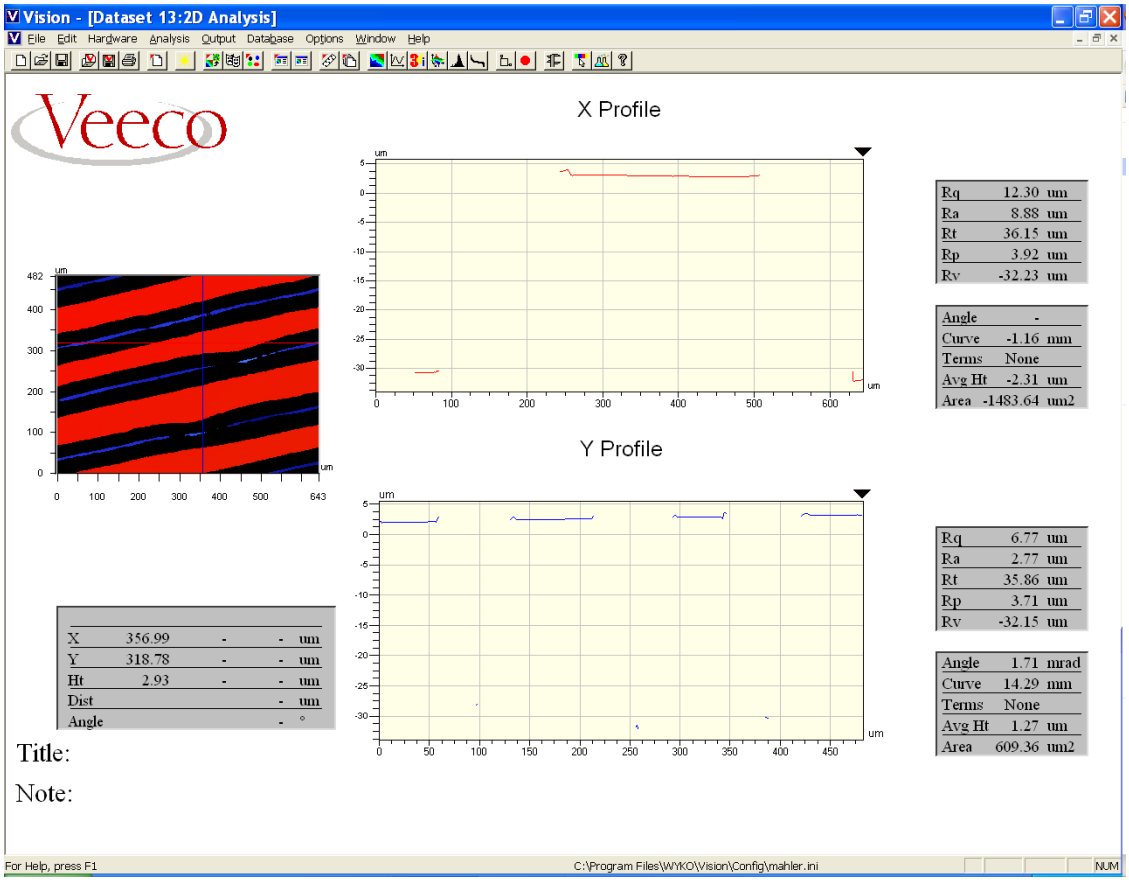


Figure 3.30: The cross-section view of the Figure 3.29 surface as seen using the Vision software

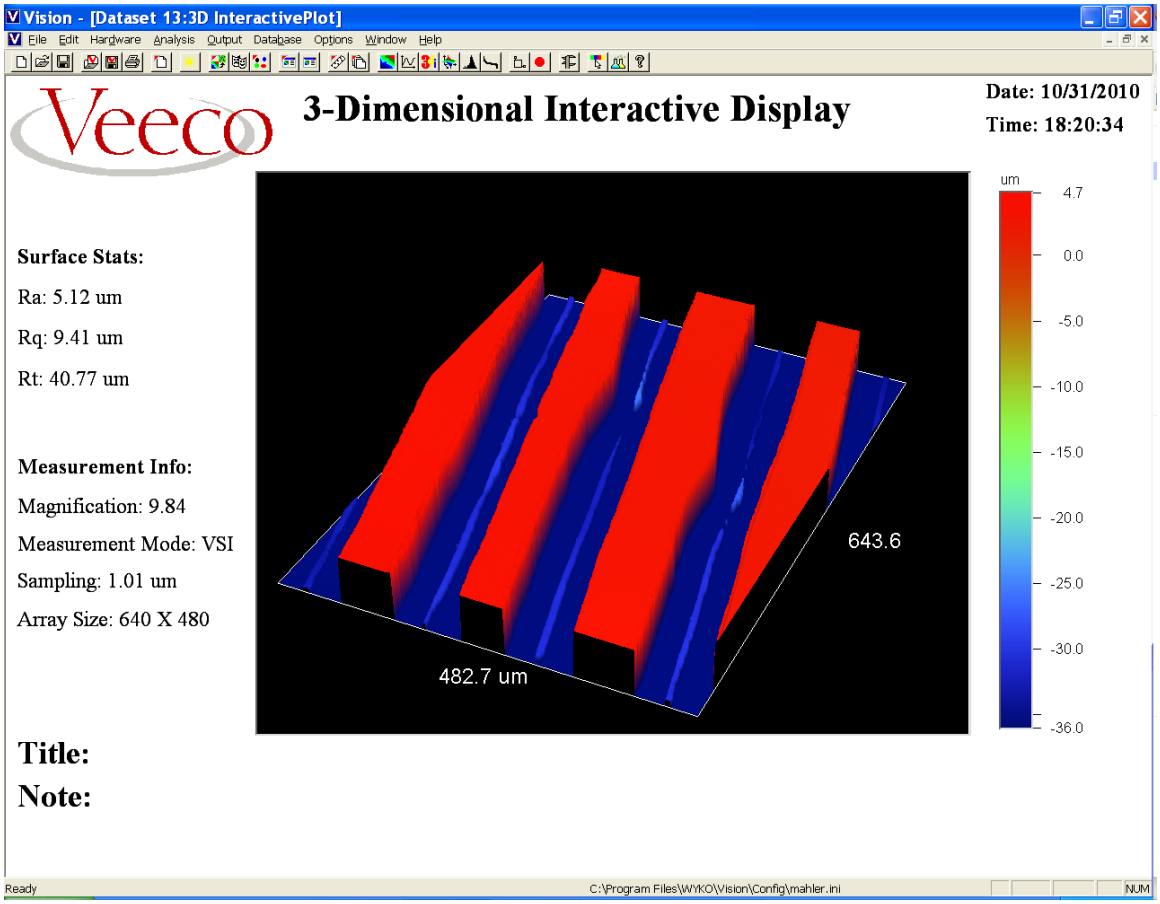


Figure 3.31: The 3D view of the same surface in Figure 3.29 of the Vision software

The data visualization and analysis tools provided by the Vision software are applicable to generic surface metrology tasks, and are limited to a single FOV. The reproduction of audio from the acquired record surface profile requires the use of our custom analytical software, which is described in Chapter 4.

3.3.7 Data Storage

After the measurement, the user can save the acquired image onto disk storage devices. The acquired surface profile of a single measurement can be saved as a data file onto the disk as either ASCII or binary. The file contains a header and a data array. The header contains metadata about the data array including the stage position corresponding to the scanned FOV, the size of the FOV, the wavelength of the light source, the lateral pixel resolution, and so on. The data array contains the heights of the pixels corresponding to the sample surface portion of the FOV size. Each height value is saved as a 32-bit floating-point number.

We used the binary format, called the OPD file format (Veeco Metrology Group 2003), as our data storage format because it is faster to save and load, and takes smaller storage space than the ASCII format. Table 3.4 compares the time and storage space taken for saving one FOV of data (under 10X magnification) in binary and ASCII formatted files.

	Storage space (megabytes)	**Saving time (seconds)**	**Loading time (seconds)**
Binary	1.5	< 1	<1
ASCII	10.5	20	5

Table 3.4: Time and space for saving and loading an FOV of surface measurement data in our acquisition system

3.4 The automatic Scanning of the Record Surface

Section 3.3 introduced the procedure for manually obtaining an FOV of disc record surface profile. Manual acquisition is only suitable for obtaining the record surface profile of small areas. In this section, we discuss the need for automatically acquiring record surface profile images, software support provided by the Vision software, and our custom workflow for automatically scanning large areas of the record surface.

3.4.1 The Need for Large-Scale Scanning

The record surface that contains the audio-bearing grooves is much larger than a typical FOV size that our microscope can provide. It is therefore necessary to perform a large number of measurements in order to acquire the record surface profile of interest. In order to understand this situation, the entire record geometry is reviewed in Table 3.5.

Note that these estimates, based on Table 2.2 and 2.3, which describe the record geometry, are conservative because they do not include factors such as the refocusing time and the overlap areas between consecutive surface areas of interest that need to be scanned (Section 3.4.2).

As can be seen from Table 3.5, even if we ignore the time that it takes to manually move the objectives to the measurement location of the record surface, the required time for scanning one side of a record is still very long. It is thus desirable to acquire the record surface profile automatically. Fortunately, the Vision software provides an automated measurement functionality, which is introduced next.

	2.75X	5X	10X	27.5X	50X	100X
FOV size (mm^2)	2.53 x 1.90	1.27 x 0.95	0.63 x 0.48	0.25 x 0.19	0.13 x 0.10	0.06 x 0.05
Number of Images	$1.14x10^4$	$4.56x10^4$	$1.82x10^5$	$1.16x10^6$	$4.23x10^6$	$1.83x10^7$
Storage space (gigabytes)	17.16	68.38	272.82	$1.74x10^3$	$6.35x10^3$	$2.75x10^4$
Time cost (days)	0.71	2.82	11.23	71.67	261.25	1133.33

Table 3.5: Theoretical time and storage costs for scanning one side of an LP record by using our acquisition system without image overlap

3.4.2 Grid-based Automatic Measurement

The Vision software allows the user to measure surfaces that are larger than one FOV size automatically. The user first defines the area of interest, which is then divided into small areas of the same size, called "cells." This process of automatically scanning a large area will be referred to here as "grid-based automatic measurement," or "GAM." The details of how to set up the GAM with Visions software will be explained next.

To measure a sample surface area larger than the FOV size, the Vision software allows the user to divide the sample surface area of interest into a grid of smaller areas (cells), each of which is of the same size as the FOV. Then, upon launching the automatic measurement, the sample stage automatically moves to the starting cell location and performs a single VSI measurement on each cell in sequence, and saves the results to disk. It is then up to the user to combine the results by stitching them together (discussed in Section 3.5).

To launch a grid-based automatic measurement (GAM), the user needs to define a grid that covers the target sample surface area, set up the global VSI measurement parameters that are applied to the measurement for each cell, and, finally, start the batch measurement.

The grid is created and configured through the use of an ASCII file called the "Stage-grid File," in which the user can specify the top-left corner location of the grid; the

number of rows and columns of the grid; the size of each cell of the grid; the spacing between the adjacent cells; and the cells that are disabled. The grid configuration in a Stage-grid File can be saved to disk and retrieved later. The default grid is a rectangular matrix of cells; by enabling or disabling certain cells in the grid, however, a scanning area of a specific geometrical layout can be defined. The grid can be created either by writing a Stage-grid File with a text editor or by using the Stage-grid File graphical interface in the Vision software (Figure 3.32). To be able to stitch the cell images resulting from the batch measurement together, the cell spacing should be specified as negative so that overlaps are reserved between the adjacent cells for Vision's built-in or custom stitching algorithms to utilize (see 3.2.3.7). Usually a 20% overlap is reserved (Figure 3.33). The required number of cells, N_{cell}, to cover the target surface area, S_{groove}, is thus:

$$N_{cell} \approx \frac{S_{groove}}{\left(1 - R_{overlap}\right)^2 S_{fov}} \qquad \text{Equation 3.1}$$

where $R_{overlap}$ is the inter-cell overlap ratio and S_{fov} the size of a single FOV, or, cell.

There are several built-in constraints associated with the GAM. Firstly, all of the individual cell measurements have to share the same VSI measurement parameter configuration. This means that the grid-based measurement cannot dynamically adapt the vertical scan range to the actual sample surface contour during the batch measurement. To fully capture the target surface 3D profile, the user must ensure that the chosen scan range spans the full depth range of the target surface area corresponding to the grid. This problem is further discussed in 3.4.3.

The second constraint is related to the vibration isolation table introduced in 3.2.1.3. A small time-interval (about 1.5 seconds) needs to be reserved between the measurements made to the consecutive cells in the GAM sequence, which permits the vibration isolation table to rebalance itself to ensure that the stage remains flat during the batch measurement. This is called the "stage rebalancing time" in this dissertation. The Vision software provides a metadata field in the Stage-grid File to specify a uniform inter-cell measurement delay that can be used as this interval. During a large-scale GAM, the accumulated waiting intervals can account for a significant portion of the overall processing time.

Finally, the X-Y sample stage iterates through the grid cells in a fixed sequence:

1) The stage starts from the top-left cell of the grid;
2) It then scans the cells in the same row in the column-by-column manner, starting from the left-most cell until the right-most cell.
3) After scanning an entire row of cells, the stage then returns to the left-most cell of the row below and repeats Step 2) for the row until it finishes scanning the bottom right cell of the grid.

Note that it takes extra time for the stage to move to successive cells in the scan sequence. When the consecutive cells are adjacent to each other, this excursion time is minimal; when they are not, the excursion time increases in a manner that is proportional to the inter-cell distances. In particular, it usually takes longer for the stage to proceed from the end of current row to the beginning of the next row than to move between intra-row columns. Such time costs are called the "stage excursion time" in this dissertation.

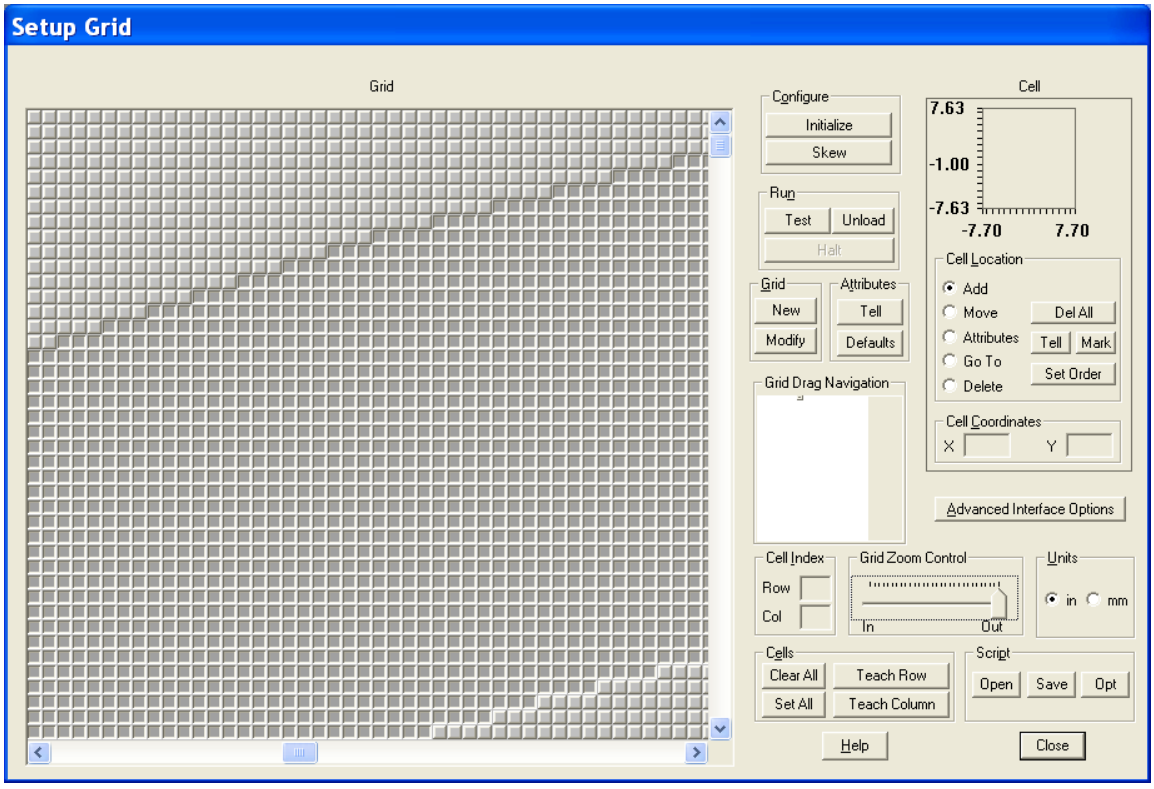


Figure 3.32: The Stage-grid File GUI: The pressed cells are the enabled ones

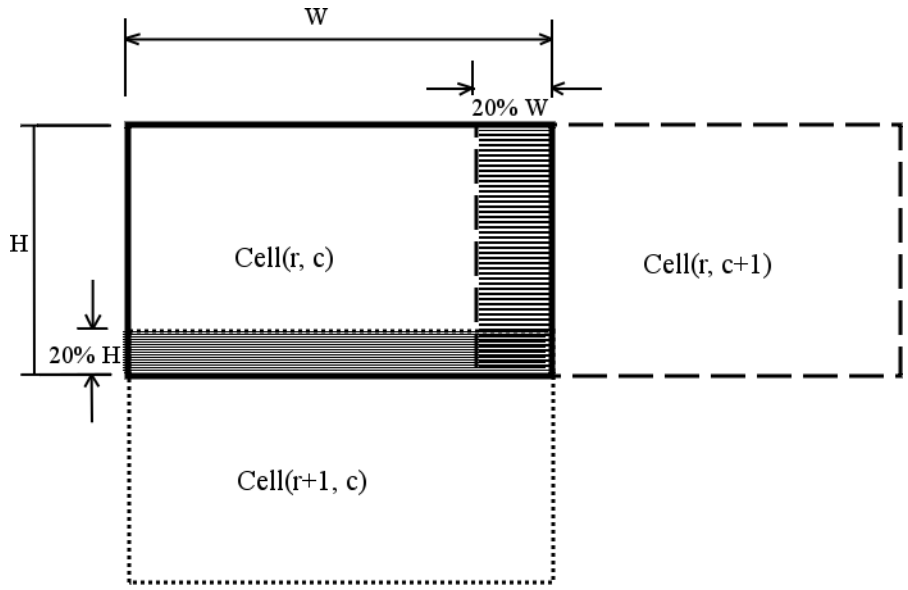


Figure 3.33: An illustration of the inter-cell overlap scheme: The shadowed areas are the overlaps of a cell at Row *r* and column *c* with two adjacent cells

3.4.3 Challenges in the Time Consumption of the Large-Scale Scanning

In theory, the large-scale scanning of the record surface can be fully automated through the GAM introduced in 3.4.2 by specifying a single grid that covers the entire target record surface area. In practice, however, this would lead to unnecessary time costs (discussed in 3.4.1). We therefore introduce a semi-automatic multi-GAM approach in 3.4.4 to reduce time cost without sacrificing scanning quality.

As indicated in 3.4.1, it would already be time-consuming to obtain the surface profile of the grooved record surface area by assuming that a common VSI parameter configuration is suitable for the entire record surface area during a large-scale scanning. However, other factors that contribute to the time consumption in the large-scale scanning of the record exist, namely height fluctuations in the record groove ridges, inter-cell waiting intervals, and the stage-returning time. They are related to the three systematic constraints of the GAM introduced in 3.4.2.

First of all, due to the warping of the disc surface, the record is not perfectly flat. That is, the groove ridge heights fluctuate across the disc surface area. Ideally, the VSI scan range for measuring an individual cell of the grid, L_g, should only span the actual groove depth in the cell. However, as mentioned in 3.4.2, the microscope cannot adapt its VSI scan range to the complex surface contour during the GAM and has to use a fixed

global scan range that covers the largest possible groove height range across the target surface. In general, such a scan range is greater than the local groove depths in most cells. According to our measurements, the overall ridge height range (the distance between the highest and the lowest groove ridges) is around 200μm for our test signal area, although the average groove depths are about 80μm, requiring about 300μm total scan range if taking into account the necessary back-scan. This would thus lead to an unnecessary time cost for the microscope to scan through the extra vertical range in addition to the actual groove depth range. We denote the time cost of scanning one FOV using the global VSI parameter configuration as T_{sg}, which has a value of 20seconds for the 300μm scan range under 3X scan speed. The total scanning time is thus:

$$t_{sg} = n \cdot T_{sg} \qquad \text{Equation 3.2}$$

where n represent the number of cells in the grid.

Secondly, the stage rebalancing time (see 3.4.2) between consecutive cell measurements is a necessary time cost to obtain correct surface profiles. We denote the stage rebalancing time between two consecutive cell measurements as T_{sr}, which is usually 1.5 seconds. The total stage rebalancing time is then:

$$t_{sr} = n \cdot T_{sr} \qquad \text{Equation 3.3}$$

where n represent the number of cells in the grid.

Lastly, the stage excursion time (see 3.4.2) determined by the fixed sample stage iteration trajectory becomes more significant as the row and column sizes of the grid grow. This suggests that the larger the grid is, the more significant the stage-returning time cost will be, which adds to the overall time cost of the GAM.

Let us assume that all columns in a row are continuous (each column overlaps with its neighbours). We denote the stage returning time needed for each row as T_{se}. According to our experiments, this is usually 3seconds for a typical row. The total stage returning time is thus:

$$t_{se} = n_r \cdot T_{se} \qquad \text{Equation 3.4}$$

where n_r represents the number of rows in the grid. Based on the above factors, the total time consumption for scanning a target area by using a GAM with a single grid, t_g, is represented in Equation 3.5:

$$t_g = t_{sg} + t_{sr} + t_{se} = n \cdot T_{sg} + (n-1) \cdot T_{sr} + n_r \cdot T_{se} \qquad \text{Equation 3.5}$$

From Equation 3.5, using a single-grid GAM, with 10X magnification, 3X scan speed, and 20% inter-cell overlap, the various time costs for acquiring the record surface profile for one side of a 30-cm (12-inch) LP record are listed in Table 3.6, respectively. The two situations (namely the theoretical and practical groove height ranges) are compared.

L_g	T_{sg}	t_{se}	t_{sr}	t_{sg}	t_g
80µm	5.33 seconds	43.20 minutes	4.95 days	17.60 days	22.58 days
300µm	20 seconds	43.20 minutes	4.95 days	65.98 days	70.96 days

Table 3.6: The time cost for acquiring disc record surface profile of one side of an LP record by using a single-grid GAM, corresponding to a grid of 2.85×10^5 cells and 789 rows

As can be seen in Table 3.6, it is time-consuming to use a single-grid GAM to acquire the disc surface profile. It is thus desirable to devise a custom scanning strategy to reduce the time cost without compromising the quality of record surface profile acquisition.

The stage rebalancing time, t_{sr}, is a fixed amount of time that is difficult to optimize further. Because the number or rows in a grid and the length of the cells in a row are largely determined by the size of the target area and the selected FOV, it is also difficult to optimize t_{se} (an ad-hoc optimization strategy in reducing the t_{se} for our test signal is discussed in 3.4.2.3). Therefore, the main goal of our custom scanning strategy is to reduce t_{sg}, the VSI scanning time.

3.4.4 A Time-Saving Workflow for Large-Scale Scanning

A time-saving workflow is devised for scanning large areas of disc phonograph records by using our acquisition system. The main idea of this workflow is to divide the target record surface into sub-areas, divide each sub-area into a grid, and perform the GAM for each sub-area by using the fixed vertical scan range that is adapted to the grid-wise groove ridge height range.

Because the groove ridge height fluctuations in the sub-areas are smaller than the entire target area (as discussed in 3.4.2.1), the VSI scanning time for each grid is reduced, leading to a reduction in the overall VSI scanning time. For the GAM for each sub-area grid, a coarse-to-fine workflow is applied to ensure that the chosen VSI scan range covers the height range fluctuations that are presented among the cell images of the grid.

3.4.4.1 The Grid Division for the Target Record Surface

There are many different solutions for dividing the target surface area into sub-areas. Note that the disc label area contains no grooves. This leads to unnecessary back-and-forth stage excursions across the label. The area is thus split into left and right halves (Figure 3.34) to restrict the stage excursions to within the left or right half of the annulus sub-areas without having to cross the label, thus reducing the stage excursion time. Then for simplicity, we divide both half areas into pies of an even angular range. Experimentation has indicated that an 8-grid division scheme (Figure 3.35) can ensure that a 100μm scan range covers the intra-grid groove ridge height range. Note that finer division schemes may further reduce the required intra-grid scan range; however, because manual work is needed to focus the microscope to the grid area, an overly fine grid division would require too much effort in the manual focusing. In practice, we thus usually adopt a moderate number of grids, such as the 8-grid division that we applied to the test signal.

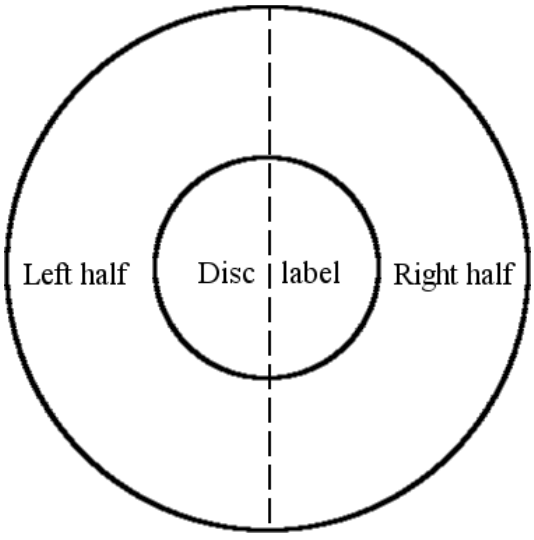


Figure 3.34: The splitting scheme (the dashed line) that helps to avoid unnecessary stage excursion across the disc label area

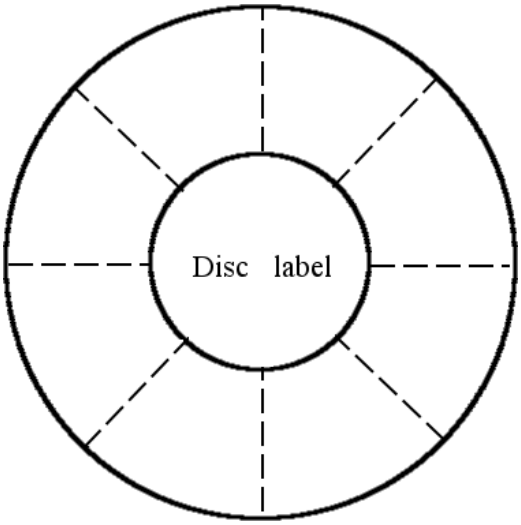


Figure 3.35: The grid division scheme used to acquire the test signal

3.4.4.2 The Coarse-to-Fine Scanning Workflow for an Individual Grid

After dividing the target record surface area into several grids, a coarse-to-fine scanning workflow is applied to each individual grid to obtain the surface profile of the grid. The workflow includes two steps: the coarse scan and the fine scan. In the coarse scan, we rapidly find the microscope focus position and the minimum scan range that covers the height range of most cells in the grid by using a fast VSI scan speed (at the cost of the acquisition quality). Then, in the fine step, by using the fixed focus position and the scan range found in the coarse scan, a GAM is performed to the entire grid with a slower VSI scan speed to obtain high-quality results.

A seed-cell is randomly selected from the grid as the first part of the coarse scan. We then manually focus the microscope onto the groove bottom, denoted as Z_s. Then, through trial and error, we find the scan range, L_g, that is long enough to span the groove height range. We denote the time taken for manually finding the scan range as T_m. Finally we randomly sample a portion of the cells in the grid (usually 10%), denoted as p, scan them using a fast VSI scan speed (usually 11X) with the scan range L_g. This takes time T_{sc} to scan each cell. The results can then be analyzed using our custom automatic algorithm to verify whether both the groove ridges and bottom for all sample-cells are fully captured. This takes time T_{ac} for each cell. See Figure 3.27 for an example of the record surface profile image obtained by a fast VSI scan speed. Although the acquisition

quality is low, it is sufficient for verifying whether both the ridge and the bottom of the groove are captured.

As mentioned in 3.3, the valid data represents the groove ridge and bottom while the invalid data the sidewalls; the invalid data is usually below a threshold portion of the total data in the FOV (30% under 10X magnification). The verification algorithm thus checks whether the portion of the invalid data in the obtained profile image is greater than the threshold. Because the manually-found focus is on the groove bottom, if L_s is shorter than the actual height range of the sample-cells the groove ridges tend to be largely missing in the coarse-scan results (Figure 3.22). This causes the invalid data percentage to well exceed the threshold portion. Note that under different overall magnifications the threshold may be different. In particular, when the overall magnification is much greater than 10X (so that each FOV contains exactly the groove sidewalls but no ridges) the invalid data will be the major portion of the data in the FOV. It would thus be impossible to verify the scan range by the aforementioned thresholding. In this case, a solution is to stitch several FOVs together and then apply the coarse step to the composite (the stitched image).

Ignoring the stage excursion time because the number of sample cells is usually small, the time spent in the coarse scan for a grid of sub-areas is denoted as:

$$T_c = T_m + p \cdot n_l \cdot \left(T_{sc} + T_{ac}\right) \qquad \text{Equation 3.6}$$

where n_l denotes the number of cells in the grid. Since, in practice, T_m is usually negligible, Equation 3.6 can then be simplified as:

$$T_c = p \cdot n_l \cdot \left(T_{sc} + T_{ac}\right) \qquad \text{Equation 3.7}$$

In the fine scan step, L_g is used as the scan range with the slower VSI scan speed, which takes time T_{sf} to finish one cell (usually 3X). The time taken for the fine scan is represented as:

$$T_f = n_l \cdot T_{sf} \qquad \text{Equation 3.8}$$

From the above analysis, the time taken to scan a grid of sub-areas of the target record surface area using our coarse-to-fine workflow can be represented as:

$$\hat{T}_{sg} = T_c + T_f = n_l \cdot \left[p \cdot \left(T_{sc} + T_{ac}\right) + T_{sf}\right] \qquad \text{Equation 3.9}$$

In practice, using 11X scan speed for the coarse scan and 3X speed for the fine scan, we have:

$$T_{sc} + T_{ac} \approx \frac{T_{sf}}{3} \qquad \text{Equation 3.10}$$

The fine scan time consumption can be represented as:

$$T_{sf} = \frac{L_s}{L_g} T_{sg} \qquad \text{Equation 3.11}$$

Considering Equation 3.10 and Equation 3.11, Equation 3.9 then becomes:

$$\hat{T}_{sg} = \left(1 + \frac{p}{3}\right) \frac{L_s}{L_g} \cdot n_l \cdot T_{sg} \qquad \text{Equation 3.12}$$

The total time consumption of the VSI scan time of our workflow is thus:

$$\hat{t}_S = \hat{T}_{sg} \sum_g n_g = \left(1 + \frac{p}{3}\right) \frac{L_s}{L_g} \cdot T_{sg} \sum_g n_g = \left(1 + \frac{p}{3}\right) \frac{L_s}{L_g} \cdot t_{sg} \qquad \text{Equation 3.13}$$

where t_{sg} is the time cost for scanning the entire target record surface area with a single GAM.

In practice, 10% random sampling is usually used in the coarse-scan verification (i.e., $p = 0.1$), L_s is 100μm under 8-grid division, and L_g is 300μm. The resulting time consumption of the coarse-to-fine VSI scan is about 34% of the time cost of the single-grid GAM for the test signal. Therefore, the devised coarse-to-fine workflow is very effective in reducing the overall time needed to acquire the record surface profile.

In our subsequent discussion and experiments, the VSI scan speed used in the estimation is 3X and the scan range is 80μm. As a trade-off between time consumption and acquisition quality, 10X overall magnification is used in our main experiment configuration.

3.5 Image Stitching

The result of large-scale acquisition of the disc record surface profile is an array of overlapping images, each corresponding to a sub-area of the target record surface area of interest. The composite of these images represents the entire target record surface profile. To extract the groove from these images, they first need to be seamlessly stitched together. Although the Vision software has built-in stitching functionality, its limitations

make it difficult to be applied to our needs. We therefore designed our custom image-stitching algorithm to perform this task. The need for our custom image-stitching algorithm is further explained in 3.5.1, and the stitching algorithm itself is discussed in 3.5.2.

3.5.1 The Need for Custom Image Stitching

After the record surface profile images are obtained by using the GAM, it is expected that each pair of horizontally adjacent images, $I_i(r,c)$, and $I_{i+1}(r,c+1)$, or vertically adjacent images, $I_i(r,c)$, and $I_j(r+1,c)$, must be seamlessly aligned, respectively, where r and c are the row and column cell indices in the grid. That is, it is expected that the image composite should not display any discontinuity in the groove patterns, based on their expected cell coordinates. However, slight mechanical drifts between the record and its hosting apparatus or between the hosting apparatus and the sample stage can occur during inter-cell stage excursions. As a consequence, images of adjacent cells may be captured at a location with a small offset from the expected cell location. After the direct composition, the two images will thus appear to be misaligned (Figure 3.36). Such misalignments would eventually result in the discontinuity in the extracted groove undulations and thus introduce noise into the reproduced audio signal. A stitching algorithm is therefore required to realign misaligned images and thereby correct the cell coordinates in the grid so that a seamless composite of the target area surface profile can be produced.

The Vision software has built-in stitching functionality that can stitch adjacent surface profile images into a single image and remove any possible misalignment. However, this functionality is only capable of stitching a small number of images. In other words, the size of the composite image that the Vision software can produce is limited. According to our tests, the composite image size supported by the Vision software is about 6.37×10^7 pixels at a time, which corresponds to 324 FOVs with 20% overlap, or an area of about 100 mm^2 under 10X magnification. This is usually not sufficient to cover the grooved record surface area for audio reproduction. The Vision software provides no application programming interface (API) for reusing the built-in

stitching functionality or for incorporating custom algorithms. We therefore developed our own stitching algorithm.

Next, we summarize the related research on image stitching.

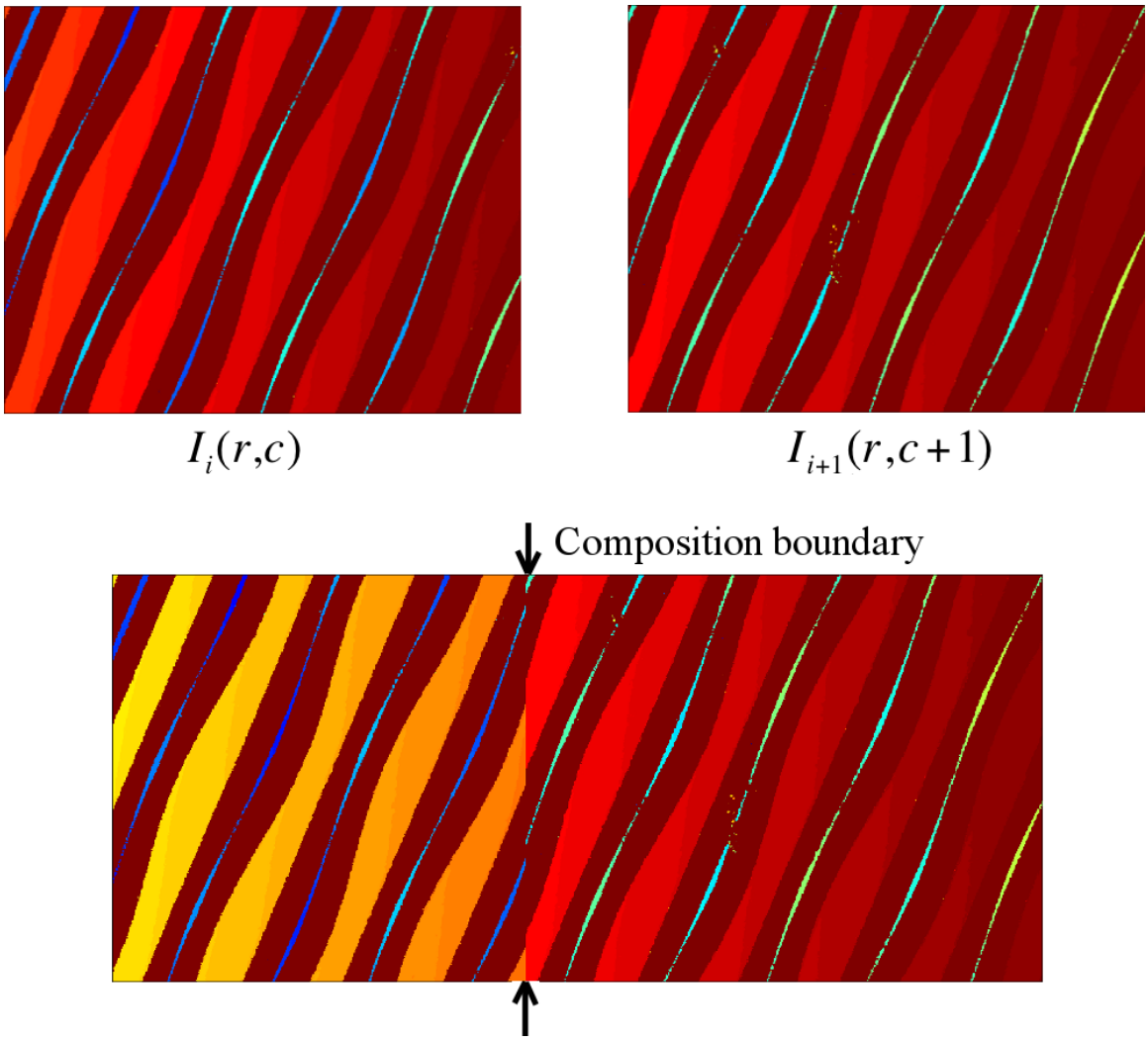


Figure 3.36: Misalignment between two adjacent cells during the GAM (contrasting colours are used in the composition image to clearly show the boundary)

3.5.2 Related Research on Image Stitching Workflows

Image stitching algorithms are widely used in the applications such as the creation of panoramic image mosaics and image stabilization in digital cameras and camcorders. Our stitching workflow is presented in the context of such related research. Szeliski (2006) gives a good review of existing image stitching algorithms. Szeliski defines the generic purpose of image stitching in terms of the need to “discover the correspondence

relationships among images with varying degrees of overlap and blend the images in a seamless manner."

An image stitching algorithm usually contains the following steps:

1) Determine the mathematical relationships between the pixel coordinates of the images that will be stitched together, which are usually called the motion models (Szeliski 2006, 4). See 3.5.2.1 for detail.
2) Find the locally-optimal pair-wise image alignments in target image collections by estimating the parameters of the motion models assumed in 1) by using either direct pixel-to-pixel comparison algorithms or feature-based algorithms (see 3.5.2.2).
3) Eliminate global misalignments among the locally-aligned image groups from 2) and determine in what order the images appear in the output composite image if it is unknown (see 3.5.2.3).
4) Perform necessary transformations to blend the aligned images into the final canvas image (see 3.5.2.4).

Our image stitching follows these basic steps, with the inclusion of adjustments made necessary by the problem domain of our image acquisition task. The images under consideration in image stitching tasks are usually 2D images. Also, the images involved in generic image stitching tasks are usually created by means other than the microscopes as in our OAR applications. This means that they can be taken under more complex conditions such that the resolutions, the exposures, and the camera lens distortions vary among the collection of images to be stitched. In our summary, we focus on the factors that relate directly to our image stitching task.

3.5.2.1 Motion Models

The misalignment between an image pair is usually modeled as the result of the pixel coordinates of one or both images being transformed by one or more types of geometrical motions. These motion models can be categorized into 2D motion models, 3D motion models, and lens distortions. The 2D motion models assume that image coordinate transformations are performed on the plane of the image of interest, including:

1) Translation: The image of interest is shifted along either or both the row and column dimensions.

2) Rotation: The image of interest is rotated around a specific point in the image plane.
3) Scaling: The image of interest is on a different scale from that of the other image.
4) Affine transform: The image of interest is transformed in a manner such that the original parallel lines remain parallel after the transform.
5) Perspective transform: The image of interest is projected onto a plane in such a manner that straight lines remain straight after the projection.

3D motion models assume that the scene that the image of interest are processed by one or more types of 3D coordinate transformations that share the same properties as the 2D motion models.

Lens distortions are associated with wide-angle lenses of cameras. The most common lens distortion is the radial distortion, where the coordinates in the observed images are displaced away or towards the camera center by an amount proportional to their distance to the center. This distortion is also referred to as the "geometrical aberration" (Stotzer 2006, 35).

We assume that the misalignment motion model for the acquired record surface profile images only involves 2D translations. The 2D translation is caused by the relative planar movement between the record, its hosting apparatus, and the sample stage. By compensating for the 2D translation alone, the misalignment can be fixed. The reasons why we ignore other types of geometrical distortions are:

1) The moving trajectory of the sample stage can be treated as a straight line. The straightness of the stage excursion is ±2μm error per 25mm of travel (see Table 3.2), which corresponds to ±0.05μm error during inter-cell excursions under 10X magnification (i.e., about 1 μm/pixel resolution). This represents sub-pixel precision. We have not discovered any rotational relative movements between the stage and the hosting apparatus either. Therefore, the record being scanned is not rotated during the large-scale acquisition process. In other words, the rotation transform does not apply to the acquired images.
2) The vibration isolation table ensures that the stage remains level during any GBAM processes. The microscope objective maintains the same posture (i.e., tilt)

during the entire scanning process. The vertical translation, affine transform, and more complex 3D motion models thus do not apply to the acquired images.

3) The surface profile images are obtained by using the same microscope with the same VSI configuration, which means the objective, the camera, and the light source used for each image are the same (see 3.2.1). There are therefore no scale, light intensity, or exposure differences amongst the images.
4) Due to the sub-millimeter FOV size, the geometrical aberration that is usually associated with the wide-angle camera objectives need not be considered.

3.5.2.2 Pair-Wise Image Alignment

Having determined the motion model to use for representing the alignment between a pair of images, we need to estimate the parameters of the model and correct possible errors. One approach is to perform a series of assumed geometrical transforms with different parameters (e.g., translation by various pixel distances) to one image relative to the other and, for each transform, to examine how much the pixels in their overlap regions match each other (i.e., present the smallest matching errors) until the best match is found. The best match is then regarded as the correct alignment of the image pair. The pixel-based alignment includes the following steps:

1) Select an error metric, against which the images are compared.
2) Apply a search technique to find the motion model parameters in the parameter space that generate the minimum error according to the selected error metric. In the case of the translation motion model, this involves using one image of the image pair as a template image $I_0(x)$ $\{x_i = (r_i, c_i)\}$, shifting it in the plane by candidate discrete pixel displacements $u = (u, v)$, the 2D translation motion model parameter, and finding the displacement where its overlap region is located in the other image $I_1(x)$.

Detailed discussion on various error metrics that handle various motion models can be found in Szeliski (2006, 16–8). Our custom stitching algorithm uses the normalized correlation error metric, which is often used in translation-based image registration. It corresponds to the cross-correlation of the normalized pixel intensities of the overlap regions of the image pair (Szeliski 2006, 18):

$$E_{corr}(u) = \frac{\sum_i [I_0(x_i) - \bar{I}_0][I_1(x_i + u) - \bar{I}_1]}{\sqrt{\sum_i [I_0(x_i) - \bar{I}_0]^2 [I_1(x_i + u) - \bar{I}_1]^2}} \quad \text{Equation 3.14}$$

where $\bar{I}_0$ and $\bar{I}_1$ are the mean intensity of the pair of image, respectively:

$$\bar{I}_0 = \frac{1}{N} \sum_i I_0(x_i) \quad \text{Equation 3.15}$$

$$\bar{I}_1 = \frac{1}{N} \sum_i I_1(x_i) \quad \text{Equation 3.16}$$

The normalization is used to eliminate the influence of the overall intensity difference or, in our case, the overall height difference, between two images.

In Step 2), the simplest search technique is an exhaustive search, which tests all possible alignments according to the possible motion model parameters. In practice, this approach can be time consuming. Various techniques are thus used to achieve accurate alignment more quickly. These techniques include: hierarchical motion estimation, Fourier-based alignment, incremental refinement, and parametric motion estimation. Readers can find details about these techniques in Szeliski (2006, 19–32).

Another important pair-wise image alignment approach is to first extract a small number of pixel groups (features) from the pair of images and then match such features from both images to coarsely align them. Finally, the parameters of the geometric motion models are obtained from refining the coarse alignment. Feature-based alignment approaches are often used for images susceptible to complex transforms because the extracted features can be insensitive to certain transforms that could compromise the estimation of motion-model parameters. This approach, however, is usually more difficult to implement than the pixel-based approach. We chose to use pixel-based alignment instead of the feature-based option for several reasons. First of all, all the images in our application contain only definitive patterns, namely the grooves. Also, as stated in 3.5.2.1, the camera and objective configurations and the VSI configuration for the images in our application are identical. Only 2D translation is involved in image misalignments, so pixel-based approaches provide a feasible solution. It is therefore unnecessary to use special features to overcome more complex transforms.

Secondly, although the search step in the pixel-based alignment can be time-consuming, the image overlap region for pixel-based alignment comparison is small, and thus the required time for the search step is short. In practice, the inter-image overlap reserved for a GBAM is 20%, corresponding to a 640x96 pixel overlap area with the neighbouring FOV of the adjacent rows, and a 128x480 pixel overlap area with the adjacent columns. According to our experiments, typical image misalignments by translation along either dimension of the image plane are shorter than 10μm (10 pixels under 10X magnification). On the other hand, FFT-based cross-correlation can further optimize the computation speed.

We chose to use the pixel-based alignment, the Fourier-based cross-correlation. Because the spatial cross-correlation between pixel matrices corresponds to the multiplication between their spectrums in the frequency domain (convolution), converting the intensities into frequency domain by using the Fast Fourier Transform (FFT), multiply, then converting it back to the spatial domain is faster than performing cross-correlation between the pixel matrices.

Figure 3.36 and Figure 3.37 show the compositing result of two neighbouring record surface profile images before and after applying our pair-wise alignment, respectively. The alignment yields an optimal displacement $\hat{u}$, by which the image I_0 is then translated to eliminate a possible seam between image pairs.

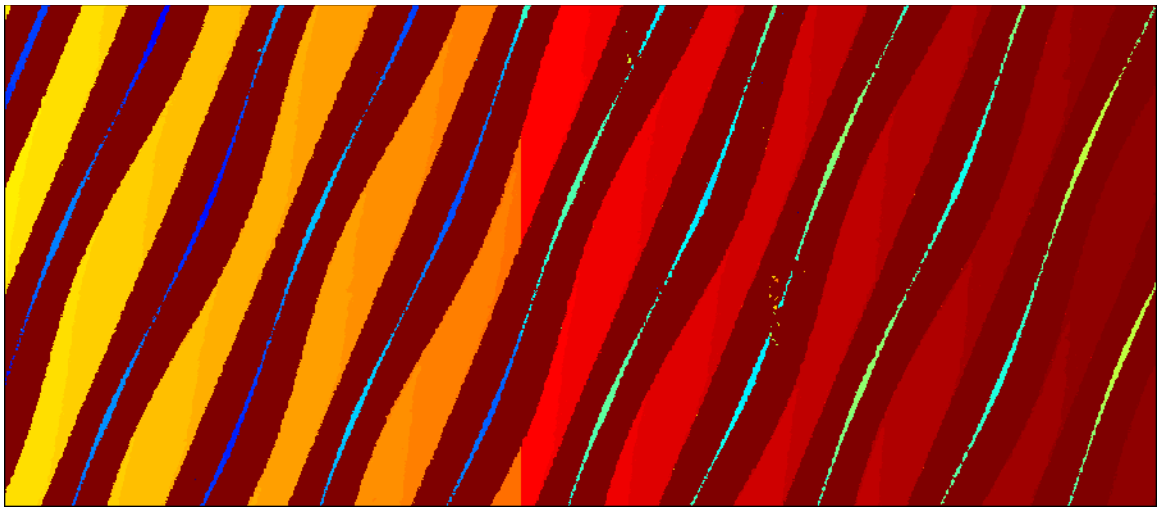


Figure 3.37: The image composition of the images in Figure 3.36 found by using only 2D translation

Using the above pair-wise alignment algorithm, large-scale pair-wise image alignments for the acquired record surface profile images are performed by using the following steps:

1) Intra-row alignments: First images in the same row of the grid are aligned pair by pair starting from the left-most pair of cells in the row using pixel-based alignment. The image on the right is aligned against the one on the left.
2) Inter-row alignments: Starting from the top row, randomly select a cell image $I_r(x)$ in the row that has a lower neighbour cell, $I_{r+1}(x)$. Then perform pair-wise alignment for the two images to align $I_{r+1}(x)$ against $I_r(x)$. The resulting displacement $\hat{u}_{r,r+1}$ is then applied to all the cells in the lower row. This process is repeated until the bottom-most row is aligned.

The inter-row alignments need to be refined by the global registration introduced in the next section.

3.5.2.3 Global Registration

The local pair-wise image alignments are performed in a locally-optimal manner. That is, each alignment strives to achieve the minimum error based on pixel or feature matching between the pair of images alone, without considering their positions relative to other images in the resulting composite image. Such a local alignment inevitably introduces some small error, and when there are a large number of images to stitch the small errors from the pair-wise image alignments can accumulate as the locally-aligned images are gradually integrated into the composite. The accumulated errors can become large enough to result in visible gaps in the output composite image, which is undesirable.

Global registration is thus performed to eliminate such gaps. Two general approaches are available:

1) The gap closing method, which stretches the existing alignments across all image pairs in small proportions based on the sizes of the observed global gaps
2) The bundle adjustment method, which simultaneously aligns all the images by distributing the alignment errors across the collection of images. See Szeliski (2006, 48–51) for details.

In our case, the displacements obtained from the inter-row image alignments, $\left\{\hat{u}_{r,r+1} \middle| r = 1,2,\cdots n_r\right\}$ (where r is the row index and n_r is the number of rows), once applied

to the corresponding rows of images, typically resulted in a gap between the last two rows to the point that they were not overlapped anymore.

The gap closing method was chosen to remove the gap and ensure seamless global alignment. We chose this approach over the bundle adjustments for the following reasons:

1) According to our experiments, the bottom-row gap described above is usually small, in the range of 0.15mm for a grid of 670 rows for the test signal. Therefore, if evenly distributed to each row-pair, the images in each row need to be shifted by 0.22μm (0.2 pixel under 10X magnification). This is small enough to fill the bottom-row gap without generating visible seams between other adjacent pairs of rows.
2) The bundle adjustment approach is usually more time-consuming than the gap closing method because the dynamic programming involved is computationally intensive for large numbers of images.

3.5.2.4 Blending of the Aligned Images

In a generic image stitching application such as camera panorama scene creation, the final step is to create the single image by blending the smaller images into a single canvas. In our case, such a step is not performed. After the global registration is completed, we do not blend the images into one single composite image because this is not necessary in order to be able to extract grooves from the images and use them to reproduce audio signals, as detailed in the coming chapters. Therefore only the images' 2D coordinates are corrected and stored for later usage in image processing (see Chapter 4).

3.6 Summary

This chapter introduced the hardware and software of our record surface profile image acquisition system. This system allows us to acquire a 3D record surface profile of any portion of the record surface area.

The built-in constraints of the microscope hardware and software create challenges in reducing the large-scale scanning time. A semi-automatic time-saving workflow was therefore devised in order to reduce the time needed for performing large-scale surface profile measurement.

The inevitable mechanical drifts that occur during the image acquisition process cause the acquired record surface profile images to become misaligned, and may cause degradation in the resulting audio. The built-in stitching functionality has difficulty handling a large number of images, and is not programmable. A custom stitching algorithm was therefore designed to correct possible alignment errors in the obtained images.

Chapter 4 The Groove Extraction

4.1 About the Chapter

Having acquired the surface profile images of the disc record by using our acquisition system and re-aligned the misaligned images, it is then necessary to recognize and extract the groove undulations from these images and stitch them together in the temporal order of the audio signal. In this chapter, we define a groove model corresponding to the pixel areas in the acquired record surface profile images and establish a groove-extraction workflow by using image-processing algorithms.

4.2 The Groove Model and the Groove Extraction Workflow

The acquired record surface profile images can be seen as a series of 2D rasterized images using the surface heights as their pixel values. It is necessary to define the groove model in order to divide the images into pixel regions that correspond to the audio-bearing grooves in the context of our OAR application. Also, in order to be able to stitch the segments of the groove undulations from the acquired large number of images together, it is necessary to define a groove model that applies to all the images so that the instances of the groove model from each image can be matched and connected with each other later. First, to be able to locate the pixels in the images and define their inter-relationships in terms of the spiralling groove structure, a 2D polar coordinate system is introduced; then we define the special pixel areas in the images as the *groove components*, which we strive to segment from the images.

4.2.1 The Polar Coordinate System

The Cartesian coordinate system determined by the microscope is the default coordinate system used for the pixels in the acquired images. The local origin of each image is at its lower-left corner while the positive directions of the row and the column coordinates are to the top and the right (Figure 4.1), respectively, and the coordinates of the origin in the microscope coordinate system is stored as metadata in the image file. Because the disc record groove is of an inward spiral shape revolving around the disc center, however, it is more convenient to model the pixel areas that represent the groove in a polar coordinate

system centered at the disc center. A polar coordinate system similar to the system described in Fadeyev and Haber (2003) is thus used (Figure 4.2, Equation 4.1–Equation 4.3), where r represents the distance between the pixel of interest and the disc center, θ the azimuth.

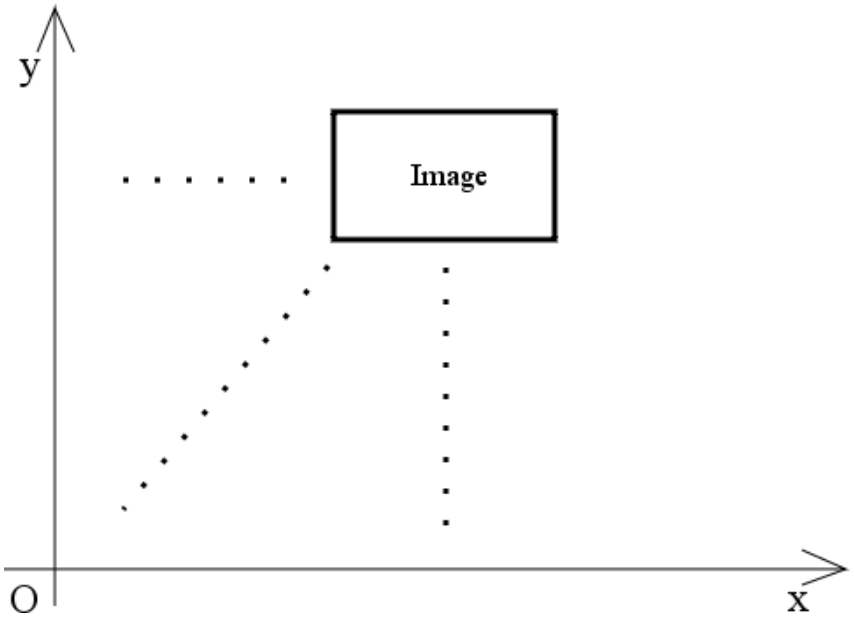


Figure 4.1: The Cartesian coordinate system used in the record surface profile images: The origin is a constant position determined by our microscope

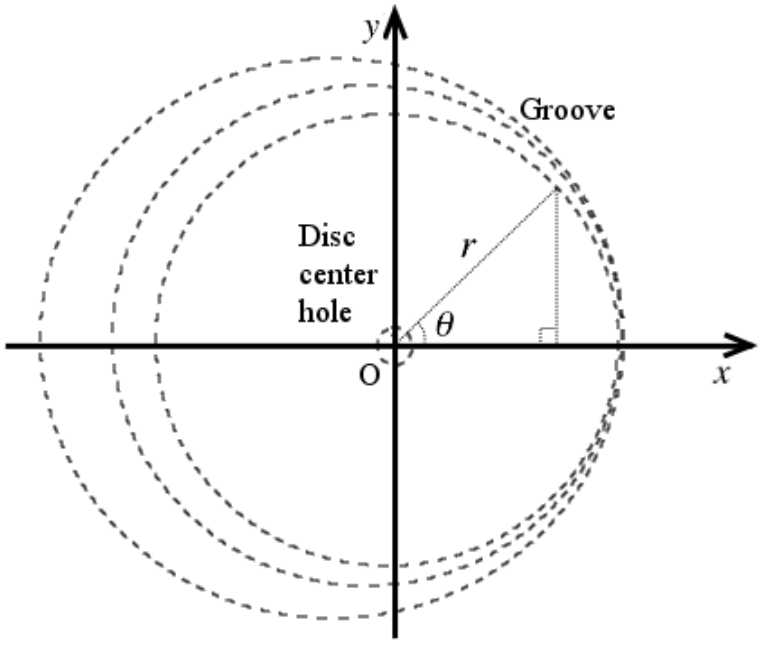


Figure 4.2: The polar coordinate system used in our groove model: The origin *O* is the center of the disc center hole

$$\begin{cases} x = r \cdot \cos\theta \\ y = r \cdot \sin\theta \end{cases}$$ Equation 4.1

$$r = \sqrt{x^2 + y^2}$$ Equation 4.2

$$\theta = \begin{cases} \arctan\left(\frac{y}{x}\right) & x > 0, y \geq 0 \\ \arctan\left(\frac{y}{x}\right) + 2\pi & x > 0, y < 0 \\ \arctan\left(\frac{y}{x}\right) + \pi & x < 0 \\ \frac{\pi}{2} & x = 0, y > 0 \\ \frac{3\pi}{2} & x = 0, y < 0 \\ 0 & x = 0, y = 0 \end{cases}$$ Equation 4.3

The coordinates of the center of the disc in the microscope Cartesian coordinate system is measured after the surface profile acquisition as a separate step. There were two optional methods:

1) Assuming that the disc is a perfect circle and the desired “disc center" is the center of it, estimate the center by sampling the coordinates of a few points on the disc brim and fit to the circle equation Equation 4.4.
2) Assuming the end of the lead-out groove, the innermost groove in which the stylus can remain circulating after the end of the playback content, is a perfect circle, estimate the disc center by sampling a few points on the lead-out groove bottom and fit their Cartesian coordinates to the circle equation (Equation 4.4).

$$\left(x - x_o\right)^2 + \left(y - y_o\right)^2 = r^2$$ Equation 4.4

From our experiments, both methods above tend to yield inaccurate disc center estimates. The possible reasons may include:

1) In reality, neither the disc brim nor the lead-out groove can be reduced to a perfect circle.

2) The disc brim and the lead-out groove can be seen as perfect circles centered at a position other than the true disc center.
3) The sampled points for estimating the disc brim or the lead-out groove contain errors.

The inaccurately estimated disc center would in turn yield inaccurate polar coordinates of the image pixels that are recognized as the groove undulations in the later image processing steps. These inaccuracies usually lead to the pitch fluctuation problem mentioned in Section 2.3.3. However, from our experiments, the roughly estimated disc center proved accurate enough for the purpose of groove segmentation and undulation extraction from the record surface profile images. We thus do not attempt to correct the disc center estimation errors until the groove undulation is extracted. The solution to the pitch fluctuation problem caused by the disc center estimation errors is presented in Chapter 5.

After the Cartesian coordinates of the estimated disc center, that is, the origin of the polar coordinate system, is acquired, the polar coordinates of all the acquired record surface profile images are obtained.

4.2.2 The Groove Components

In Figure 4.3, the groove ridges, sidewalls, and the bottom can be clearly seen; the groove sidewalls and the bottom together form the groove valley; we defined the groove ridge on the side of the valley that is closer to the disc center as the groove inner ridge, and the one closer to the outer rim of the disc the groove outer ridge (Figure 4.3). A groove revolution of interest has two undulations: the inner and outer undulations. The inner undulation is defined as the boundaries between the groove valley (i.e., the region that comprises both the sidewalls and the bottom) and the inner ridge; the outer undulation is defined as the boundaries between the groove valley and the outer ridge (Figure 4.3). Both groove undulations are required to eventually decode the stereo audio signal: The inner undulation is related to the left channel and the outer the right channel of the output stereo audio signal (see Chapter 5 for details). The groove bottom component is useful in removing noises, which is described in 4.5.

In a record surface profile image, the pixel sets that represent the groove ridges, sidewalls, valleys, and bottom are called the groove components in this dissertation, denoted as G_R, G_S, G_V, and G_B, respectively. The entire image pixel set is denoted as I. Note that there are usually multiple segments of any type of groove component in an image, which correspond to different groove revolutions (see Figure 4.3).

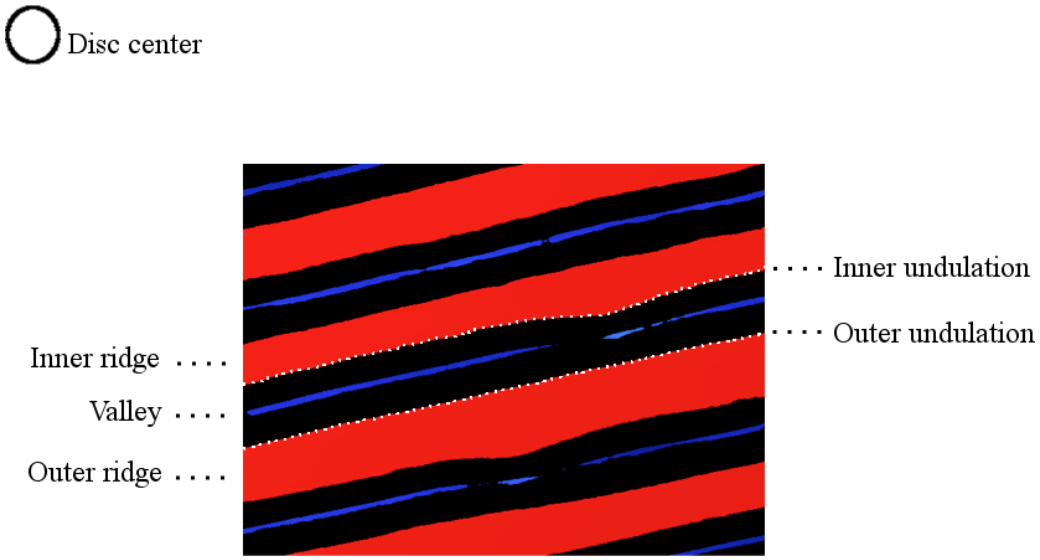


Figure 4.3: The valley, the inner and outer ridge, and the undulations of a record groove: The dash lines represent the undulations

The geometrical relationships between the groove components, represented by the pixel sets, include:

$$G_V = G_S \cup G_B \qquad \text{Equation 4.5}$$

$$I = G_R \cup G_V \qquad \text{Equation 4.6}$$

$$G_V = G_R{}' \qquad \text{Equation 4.7}$$

Equation 4.5 implies:

$$G_B \subset G_V \qquad \text{Equation 4.8}$$

$$G_S \subset G_V \qquad \text{Equation 4.9}$$

Equation 4.8 and Equation 4.9 indicate an important geometrical property of the groove components: The groove bottom and sidewalls are contained by the groove valley. Here, the containment geometrical relationship is defined as: The pixel set A is contained by the pixel set B if at any azimuth along the tangential direction of the pixel set A and B. the radius range of the radial cross-section of A is not beyond that of the radial cross-section of B (including the cases where the radial boundaries of both cross-sections overlap each other). Equation 4.6 and Equation 4.7 suggest that the valleys are separated by the ridges in the image.

Another important geometrical property of the groove components is their tangential continuity: For any type of groove component, only one instance of the type is supposed to be found on a single revolution of the groove. Note that the two sidewalls contained by the groove valley are considered to belong to a single component.

4.2.3 The Groove Undulation Extraction Workflow

Having defined the groove components and undulation in Section 4.2.2, we then extract the groove undulation as the boundary between the groove valley and its adjacent ridges. The workflow consists of the following steps:

1) The groove component segmentation: In each record surface profile images, segment the groove components: the ridges, the valleys, the bottom, and the sidewalls. In order to extract the boundary between groove ridges and valleys, such groove components have to be extracted first. See Section 4.3.
2) The coarse groove component indexing: Assign a number to all the pixels of each segmented component from Step 1) according to the groove revolutions to which they belong so that all the pixels of each component have the same index number. This is for continuously tracking the groove undulations along the groove revolutions across the images. See Section 4.4.
3) The noise removal and the indexing correction: The acquired images are often affected by various types of noise and damage so that the results of the coarse groove component indexing in Step 2) usually contain errors, which would affect the later undulation extraction and tracking. It is then necessary to correct the indexing errors by removing the noise, restoring the damaged groove components, and re-indexing the components. See Section 4.5.

4) The groove ridge-valley boundary extraction: Extract the edges between the groove ridge and valley components as the groove undulation for each indexed groove segment in each image. See Section 4.6.
5) The groove undulation tracking and stitching: Track the extracted edges of the indexed groove components across the images according to the groove continuity, and stitch them together to create the complete time series. See Section 4.7.

The steps of the workflow are discussed individually in the following sections.

4.3 The Groove Component Segmentation

In the groove component segmentation step, the four aforementioned types of groove components are recognized and separated from each other in the images. First, because the sidewalls usually cannot be captured during the image acquisition step, they are usually homogeneous invalid data (not-a-number, or NaN) as explained in Section 3.3. Thus, they are readily segmented from the rest of the image. The sidewalls are denoted as

$$G_s = \left\{ i(x,y) \middle| h(x,y) = NaN, i(x,y) \in I(x,y) \right\}$$

where $i(x,y)$ is the pixel subset of an image $I(x,y)$, and $h(x,y)$ the height value of a pixel in the image. Then a height thresholding process is performed to separate the groove ridges and the groove bottom, which is explained below.

As mentioned in 4.2, a record surface profile image can be seen as a 2D image with the surface heights as its pixel values. There are two conditions that enable the use of height thresholding to segment the ridges and the bottom of the groove:

1) The groove ridges are higher than the bottom in an image and the average groove depth is known (see Chapter 3).
2) No groove portions other than the ridges and the bottom are in consideration here because the groove sidewalls, which span the vertical ranges between the ridges and the bottom, are already segmented.

Then by finding a height threshold between the lowest groove ridge point and the highest bottom point in the image, the ridge and the bottom can be separated. The method that we use for finding the threshold is based on the finding that the ridges within an FOV is flat enough so that the height fluctuation range of the ridges ($\Delta\hat{H}_R$) is roughly known from

statistics, for example, 15μm on average for our tested LP records. We first find the peak height of the image, $\hat{h}_{\max}$, and then the threshold ($\hat{h}_{th}$) is thus:

$$\hat{h}_{th} = \hat{h}_{\max} - \Delta\hat{H}_R \qquad \text{Equation 4.10}$$

$$\Delta\hat{H}_R = k \cdot \Delta\hat{H}_g \qquad \text{Equation 4.11}$$

where $\Delta\hat{H}_g$ is the empirical groove depth (usually 80μm) and k the scaling factor (usually 1/6). The groove ridges belong to $G_R = \left\{ i(x,y) \middle| h(x,y) > \hat{h}_{th}, i(x,y) \in I(x,y) \right\}$ and the bottom $G_B = \left\{ i(x,y) \middle| h(x,y) < \hat{h}_{th}, i(x,y) \in I(x,y) \right\}$.

After separating the groove ridges and bottom, the groove valleys can then be separated by using the relationship from Equation 4.7. Then we have $G_V = \left\{ i(x,y) \middle| i(x,y) \notin G_R, i(x,y) \in I(x,y) \right\}$. According to our experiments with several records, this proves a robust approach to separate the groove ridges and bottom. After this step, the groove components G_R, G_V, G_B and G_S, are segmented. Figure 4.4–Figure 4.7 show a record surface profile image and its segmented groove components, respectively.

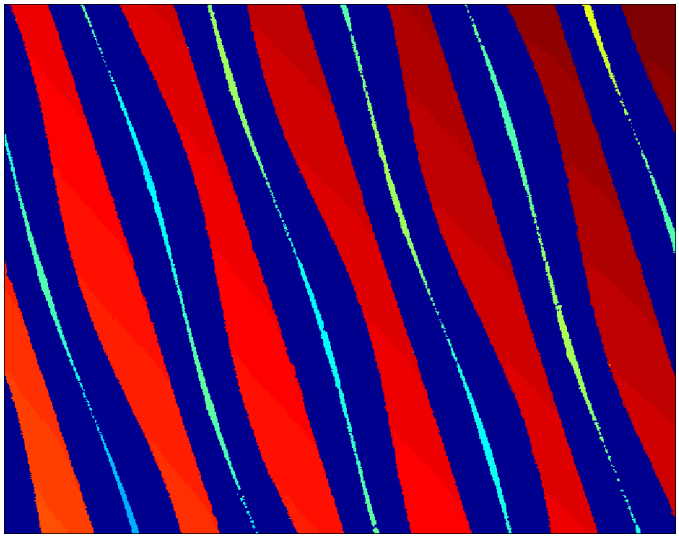


Figure 4.4: A record surface profile image

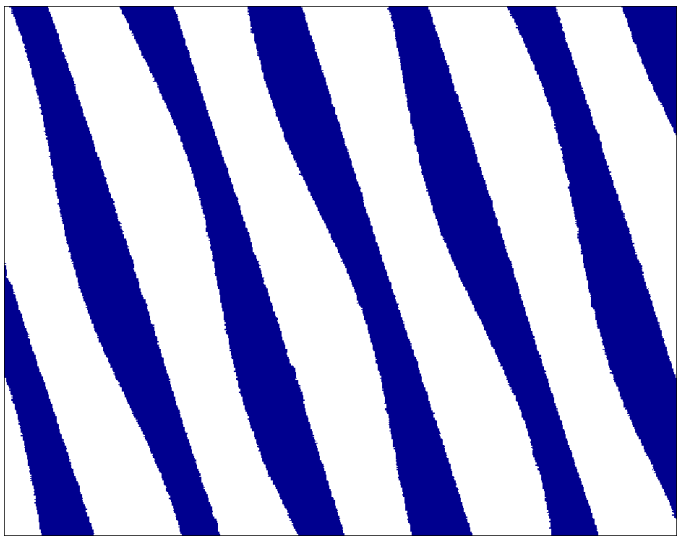


Figure 4.5: The ridge (the white portion) and the valley components (the blue portion) in the image in Figure 4.4

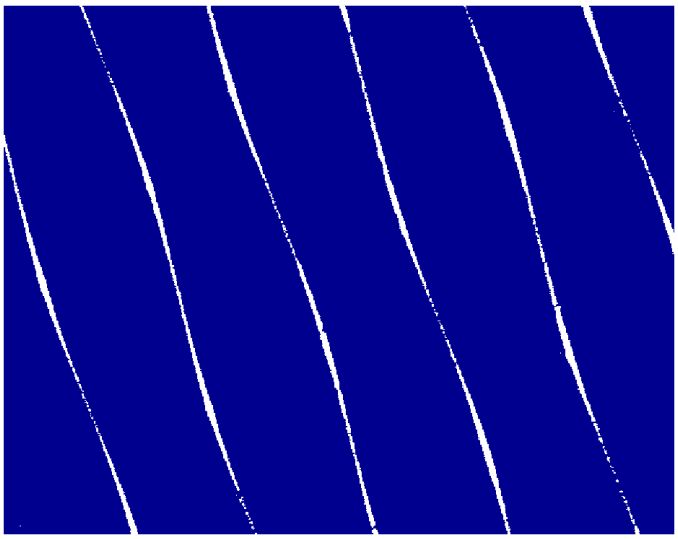


Figure 4.6: The groove bottom components (the white portion) in the image in Figure 4.4

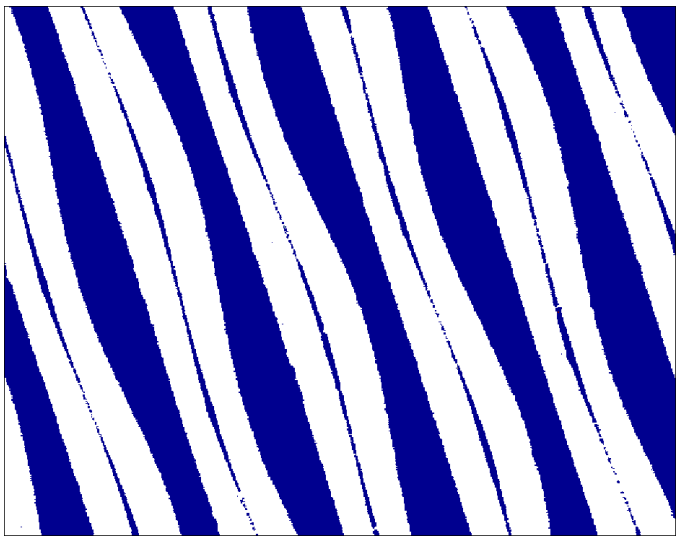


Figure 4.7: The groove sidewall components (the white portion) in the image in Figure 4.4

4.4 The Coarse Groove Component Indexing

Eventually the extracted groove undulations from each image need to be tracked and stitched together according to their temporal order in the groove spiral; during the tracking and stitching, given an undulation segment of interest, its successor segment needs to be selected from among the multiple segments of the adjacent images. The groove components therefore need to be indexed according to the groove revolutions that they belong to. The indexing is performed by using the connected-component labelling (CCL), which is an important image segmentation technique used to detect connected pixel areas in an image as objects and assign a unique label to each object (Ballard 1982, 369). A connected pixel area is defined as a pixel set in which a connected pixel path exists between any two pixels in the set and the path is composed of only the pixels within the set (Fujinaga 1997). The typical types of pixel connectivity used in CCL are shown in Figure 4.8. The pixels of an input binary image (Figure 4.9) can thus be labelled according to such connectivity and separated as connected components (CCs). The result is shown in Figure 4.10.

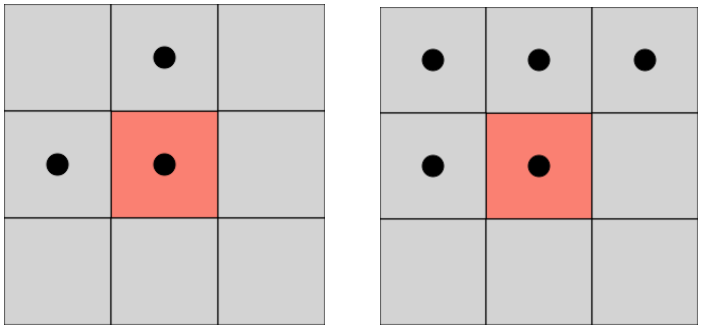


Figure 4.8: The 4-connectivity (left) and 8-connectivity (right) of the center pixel used by CCL (images from Wikipedia 2011)

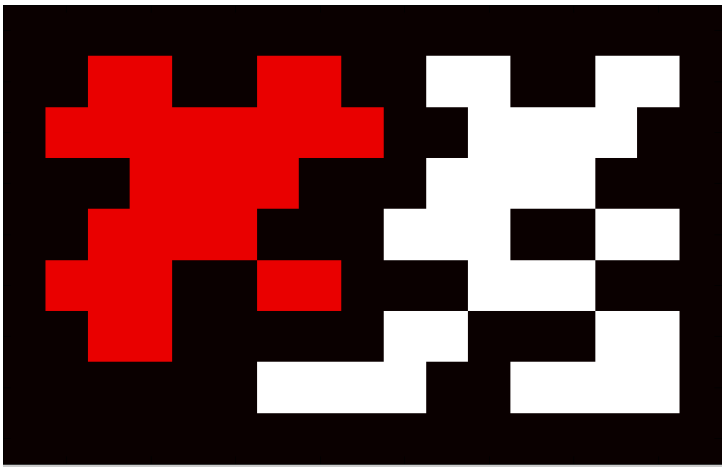


Figure 4.9: An input binary image for CCL: The pseudo-colours indicate the connected pixel areas (image from Wikipedia 2011).

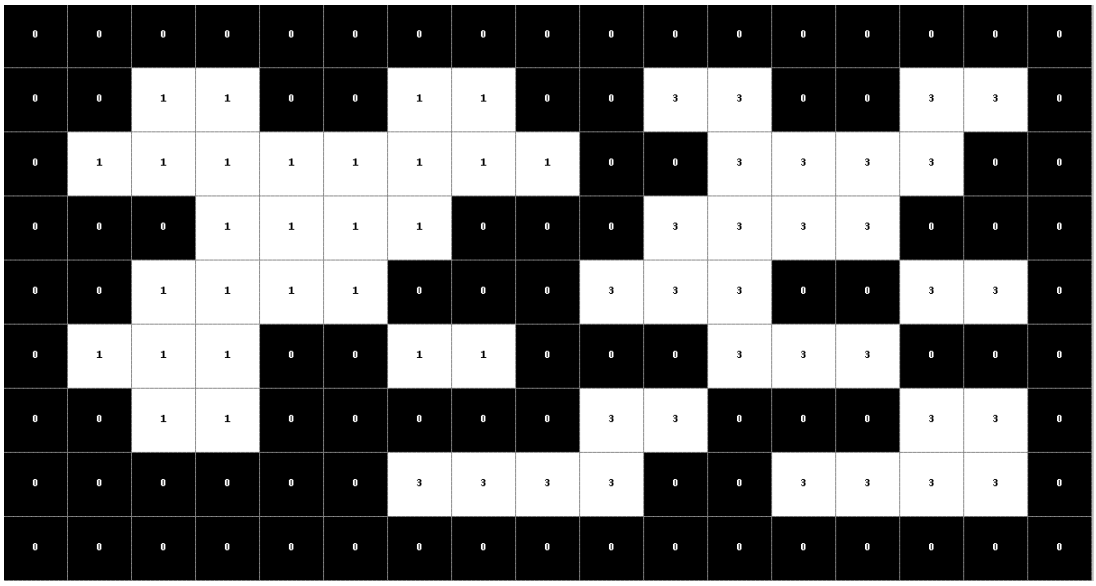


Figure 4.10: The connected pixel areas in Figure 4.9 that are labelled with indices (image from Wikipedia 2011)

As seen in Figure 4.5–Figure 4.7, the groove components all correspond to specific connected pixel areas; CCL is thus suitable to index the groove components in the record surface profile images. The actual CCL algorithm adopted here is the two-pass connected-component labelling (Haralick and Shapiro 1992, 28–48).

We perform CCL to index G_R, G_V, and G_B, respectively; note that the sidewall components are not required to be indexed. Before performing CCL, the height information of the image is first preserved. Then to index G_R, the images are binarized based on Equation 4.7 into G_R and G_R', and then processed with CCL, resulting in the connected components that represent the groove ridges, denoted as $C_R^{(i)}$, $i = 1,\cdots,N_R$, where N_R is the number of the groove ridge segments corresponding to different groove revolutions in the image. G_V can be indexed by inverting the binary image for G_R and perform CCL to the inverted binary image; the result is denoted as $C_V^{(i)}$, $i = 1,\cdots,N_V$, N_V is the number of the groove valley segments corresponding to the groove revolutions in the image.

From Equation 4.5–Equation 4.7, it can be derived that:

$$G_B = \left(G_R \cup G_S\right)' \qquad \text{Equation 4.12}$$

The image is then binarized into G_B and G_B'; CCL can then be performed on the binary image. The acquired groove bottom that is contained in the groove valley is often present as broken series of pixel areas. The immediate result CCs from CCL for G_B therefore usually do not share the same indices as their container valley (Figure 4.13). Later in Section 4.5, the groove bottom components need to be used in association with their container valley and are therefore better indexed with respect to their container valley. A special post-processing step is then performed on the raw groove bottom CCs to label the CCs that are contained by the same groove valley component as the valley CC's label. This is done by exploiting the containment relationship between the bottom and the valleys indicated in Equation 4.8. The result is denoted as $C_B^{(i)}$, $i = 1,\cdots,N_V$. See Figure 4.11–Figure 4.15 for the indexed groove components corresponding to those in Figure 4.5–Figure 4.7. The indices are colour-coded.

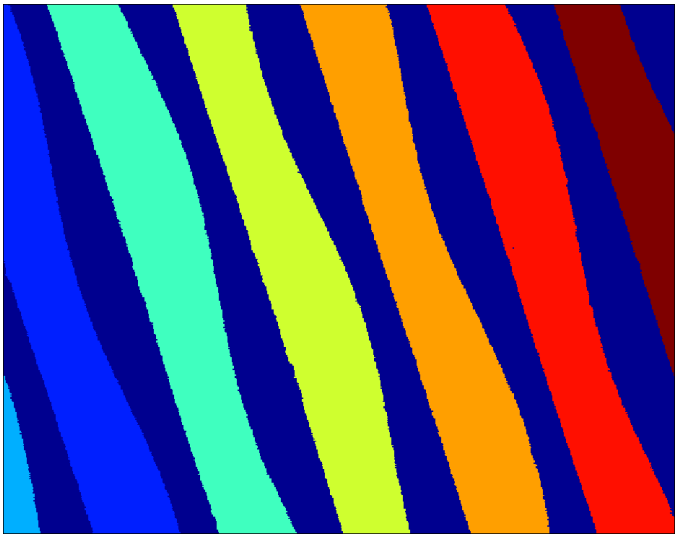


Figure 4.11: The indexed groove valley components in the image in the image in Figure 4.4

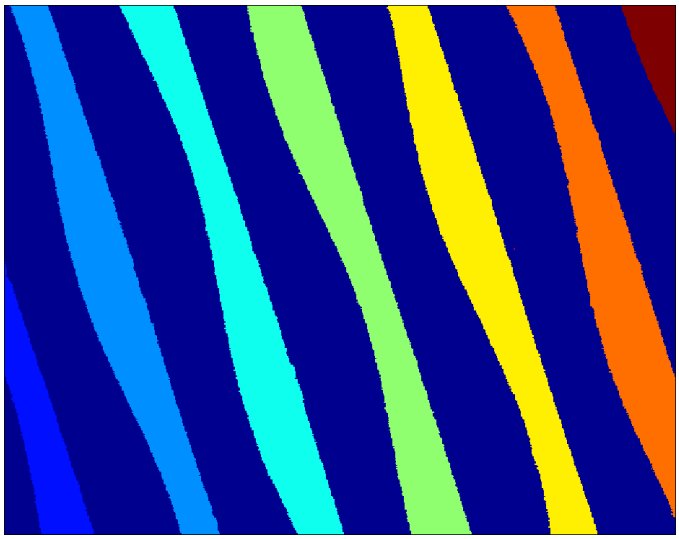


Figure 4.12: The indexed groove ridge components in the image in Figure 4.4

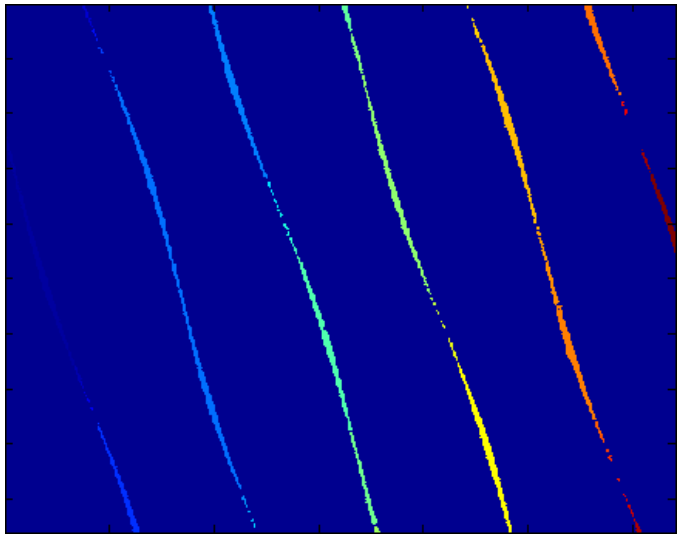


Figure 4.13: The immediate result of CCL on the groove bottom components for Figure 4.4

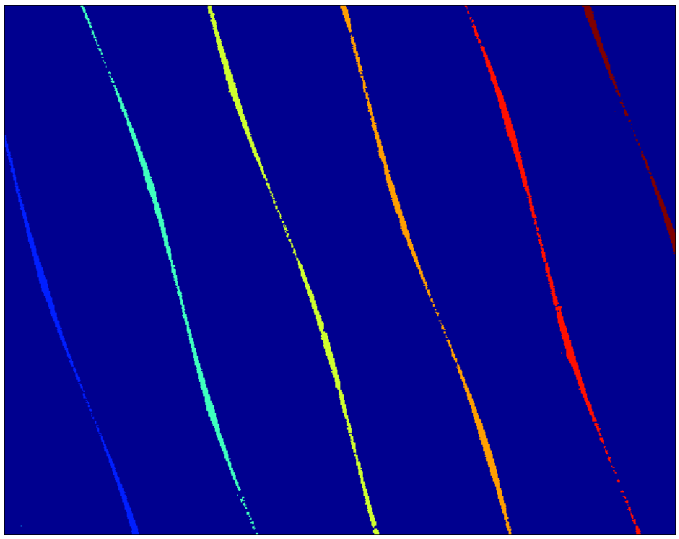


Figure 4.14: The groove bottom components in Figure 4.13 are further indexed using the labels of their container groove valley components

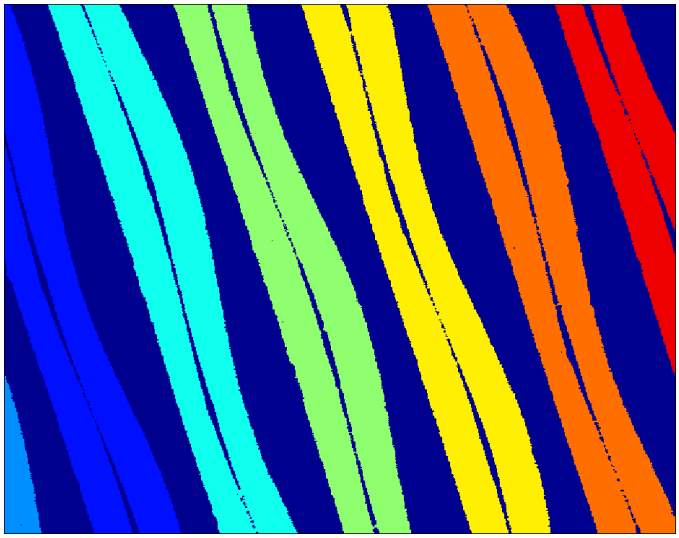


Figure 4.15: The indexed groove sidewall components in the image in Figure 4.4

4.5 The Noise Removal and Indexing Correction

The CCs obtained in Section 4.4 usually include not only the groove components but also the noise from dust and dirt on the record. In our experiments, although the record was thoroughly cleaned with a special vacuum cleaner for phonograph records in advance, the dust and dirt still accumulated on the record surface during the record surface profile acquisition process. In addition, the groove components are often affected by groove damages. For instance, some groove ridges can be broken due to the damage. As the consequences of these factors, there are usually two types of noise and defects among the coarsely indexed groove components: blob noise and the discontinuity in the groove components (see 4.5.2). Their negative influences on later processing and the solutions to remove them are discussed in the next sub-sections.

In the future, the record surface profile acquisition is recommended to be conducted in a dust-free environment in order to help reduce the effort for removing the noise such as the techniques presented in this section.

4.5.1 Removing Blob Noise

Blob noise refers to false groove components found among the indexed groove ridges and valleys. A typical example of blob noise in the groove valley components is shown in Figure 4.16. There are several types of such noise:

I. The groove ridges may have damaged or deformed portions that are well lower than the other portions of the ridges; they then tend to be wrongly segmented as part of the groove bottom or valley components (because the valleys are the complements of the ridges).
II. When the global VSI (vertical-scanning interferometry) modulation threshold is not high enough for certain cell-images during the large-scale image acquisition, the measurement errors may exist in the resulting images and present themselves as blob noise.
III. Dust and dirt narrower than the groove width may fall into the groove valley and the gaps of the damaged grooves. Since the dust and dirt are usually higher than the bottom, they tend to be wrongly segmented and indexed as part of the groove ridge components after height thresholding and CCL.
IV. The groove ridges may be damaged and the broken pieces may present themselves as blob noise.

According to our statistics, more than 80% of the acquired images of the test signal contain blob noise. If blob noise were recognized as the groove ridges or valleys, extra efforts would need to be spent on the later groove undulation extraction for these noise instances to obtain their boundaries and then filter them out, which adds to the processing time cost. Blob noise thus needs to be identified and removed. However, we do not attempt to remove all the types of blob noise listed above. This is due to two factors:

1) Some instances of the above Type IV blob noise are useful for reconnecting the broken groove components later (see Section 4.5.2).
2) The instances of the Type III blob noise are difficult to be differentiated from the Type IV ones because they share similar 2D shapes and heights.

These two types of blob noise thus need to be identified and preserved. In other words, only the first two types of blob noise need to be removed in this phase.

From Figure 4.16, one may assume that blob noise is usually small in size compared to true groove ridges or valleys and so CC-size thresholding could be used to identify and remove blob noises as small CCs; however, such heuristics proved to be not robust in practice due to the diversity of the noise sizes. We thus use heuristics based on the containment geometrical relationship between the groove components mentioned in

Section 4.2.2. In theory, the groove ridges and the groove valleys separate each other from different groove revolutions and none of them contain each other; therefore, any $C_V^{(i)}$ ($i = 1, \cdots, N_V$) contained by a $C_R^{(i)}$ ($i = 1, \cdots, N_R$) is an instance of blob noise and should be removed; although such a rule can be applied to the cases where ridge components are contained by a valley component, at this moment, we do not try to remove blob noise found in such cases (Type III and IV blob noise) because they can contain the true pieces of broken ridges as previously mentioned. After the discontinuous groove components are all reconnected later, any blob noise found contained by the groove valley components is removed.

This method proved to be a robust method for filtering out blob noise. When such a CC is detected, it is then re-labelled as part of its container ridge or valley component. An example of the result of the noise removal is shown in Figure 4.17.

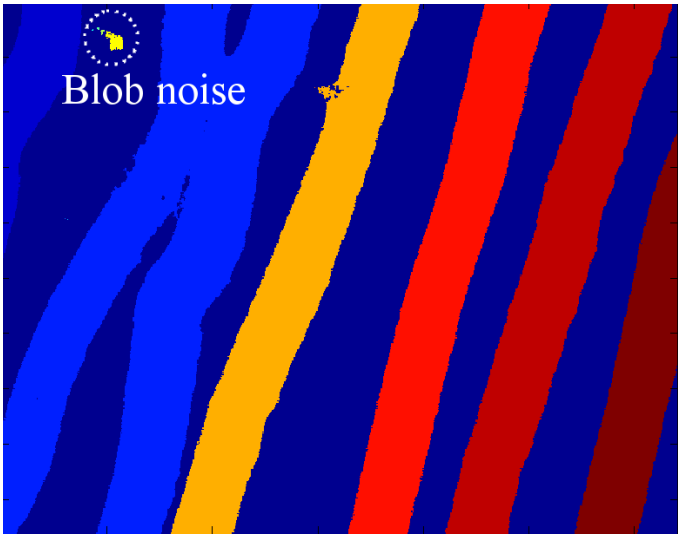


Figure 4.16: The indexed groove valley components that contain blob noise (Type I)

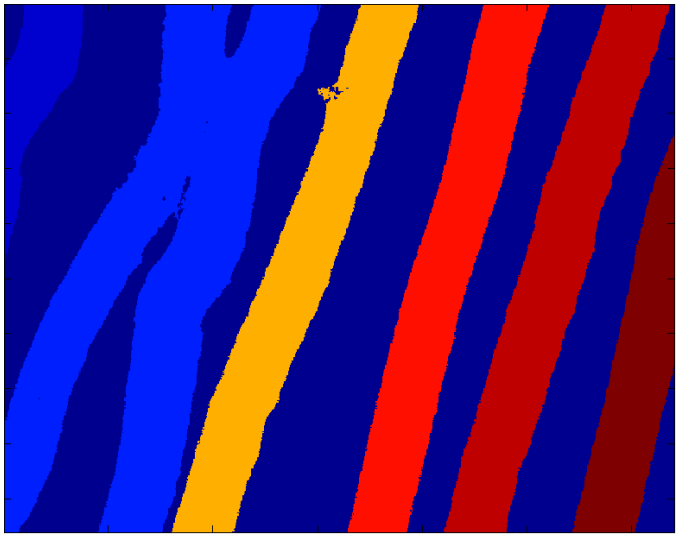


Figure 4.17: The re-indexed groove valley components in Figure 4.16 after removing blob noise

4.5.2 Reconnecting the Discontinuous Groove Components

Other than blob noise, the discontinuity in the groove components is another type of defect that creates difficulty in the later groove undulation extraction and tracking steps and thus need to be fixed. An example is shown in Figure 4.18–Figure 4.21, in which the groove ridge, valley, and bottom all contain discontinuous segments. The challenges that they introduce and the solutions to fix them are discussed in detail in the following sub-sections.

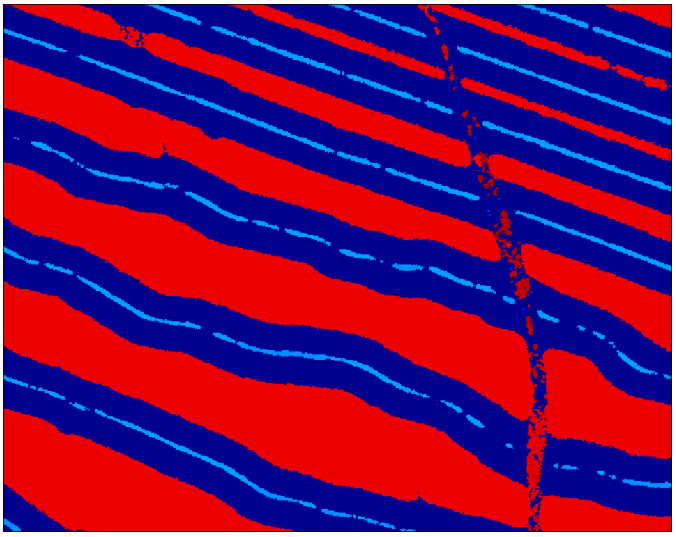


Figure 4.18: A record surface profile image that contains discontinuous groove components

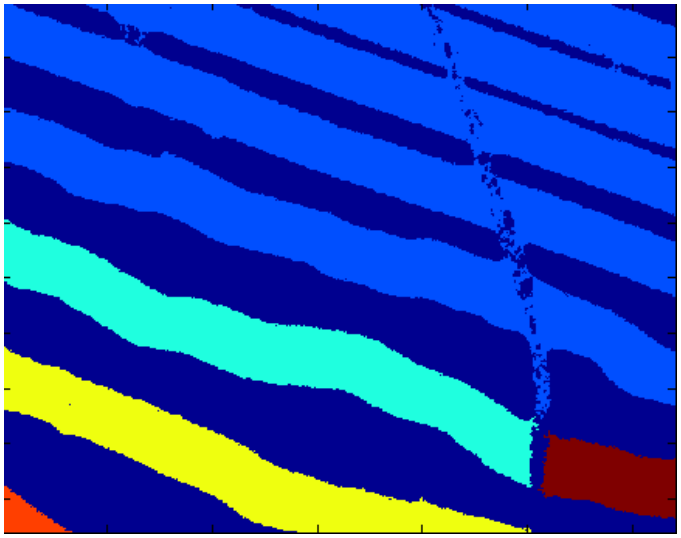


Figure 4.19: The indexed groove valley components affected by the discontinuity defects in the image of Figure 4.18

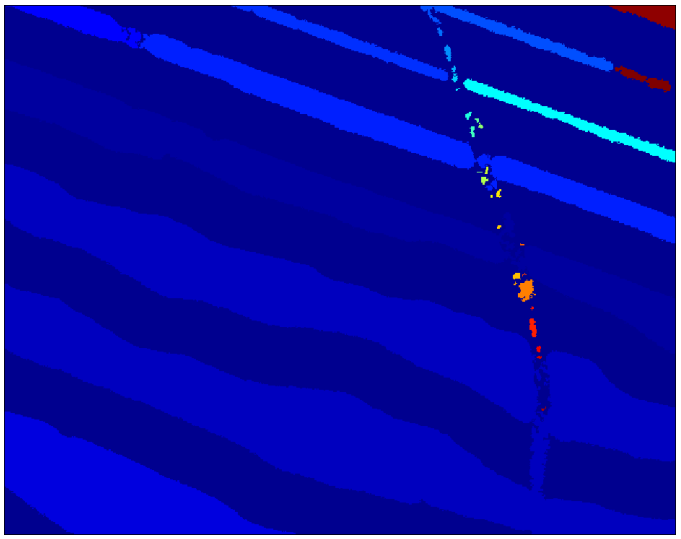


Figure 4.20: The indexed groove ridge components affected by the discontinuity defects in the image of Figure 4.18

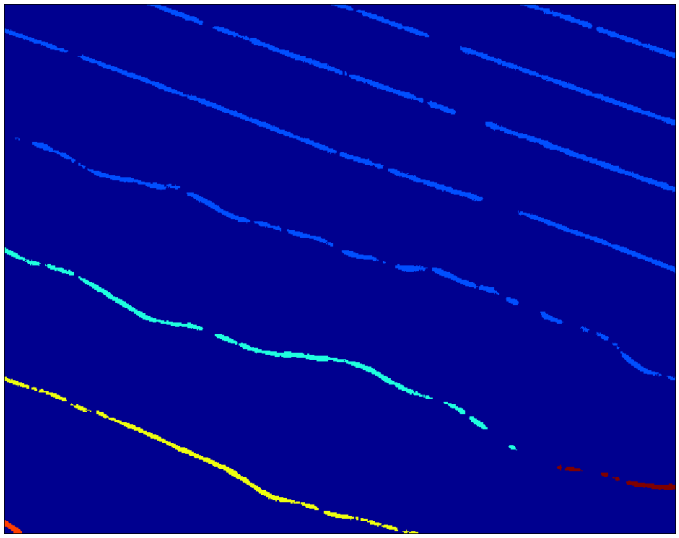


Figure 4.21: The indexed groove bottom components affected by the discontinuity defects in the image of Figure 4.18

4.5.2.1 The Challenges in Fixing the Discontinuous Groove Components

The discontinuities in the groove valleys and ridges are mainly due to the dust and groove damages.

The discontinuous groove valley components can be formed in the following manner. When a piece of dust or hair that is longer than the width of the groove happens to fall onto the record surface, because its height is comparable with the height of the surrounding groove ridge, it is recognized as part of the groove ridge during the height-thresholding (see Section 4.3); the valley is then recognized and indexed as two separate valley CCs after CCL (the lower part of Figure 4.19) and the ridges on its both sides are interconnected and thus indexed as a single ridge CC (the lower part of Figure 4.20).

The discontinuous groove ridge components can be formed via two different causes. First, the groove ridges may be damaged from external forces such as scratches. The gaps caused by the damage are captured as invalid data during the image acquisition process (see Section 3.3.1) and thus recognized as part of the groove valley (Figure 4.22); each damaged groove ridge becomes two separated CCs after CCL is performed (Figure 4.23) and the adjacent groove valley segments in turn become a single CC (Figure 4.24).

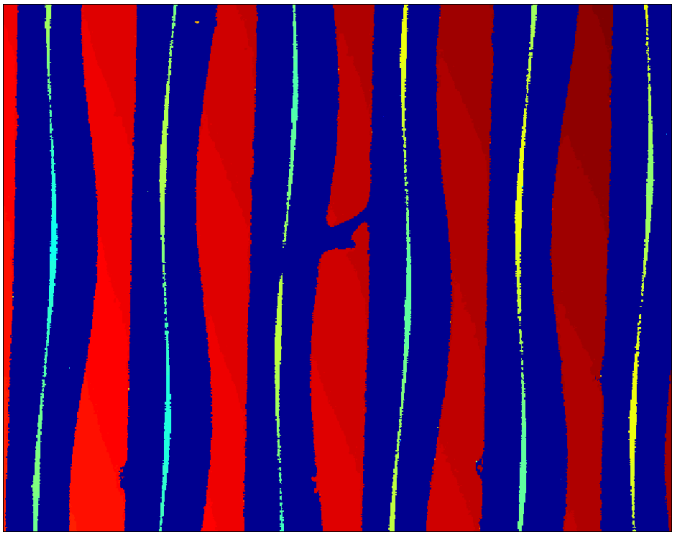


Figure 4.22: A record surface profile image that contains a broken ridge (the center one)

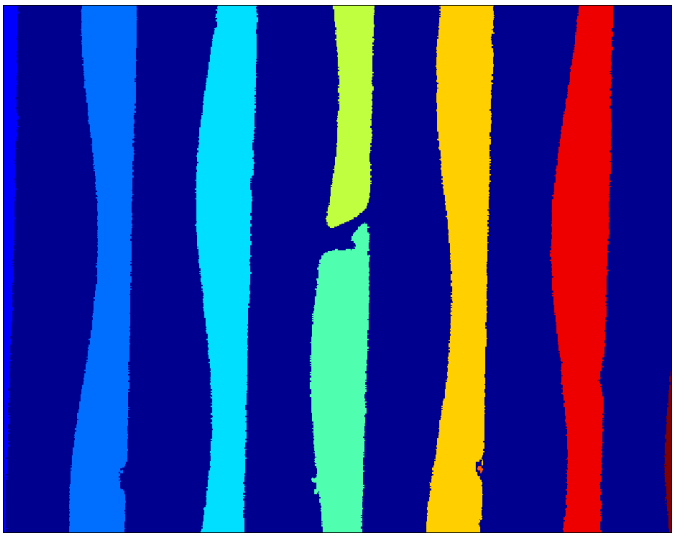


Figure 4.23: The broken ridge CCs (the center ones) corresponding to the ridges in Figure 4.21

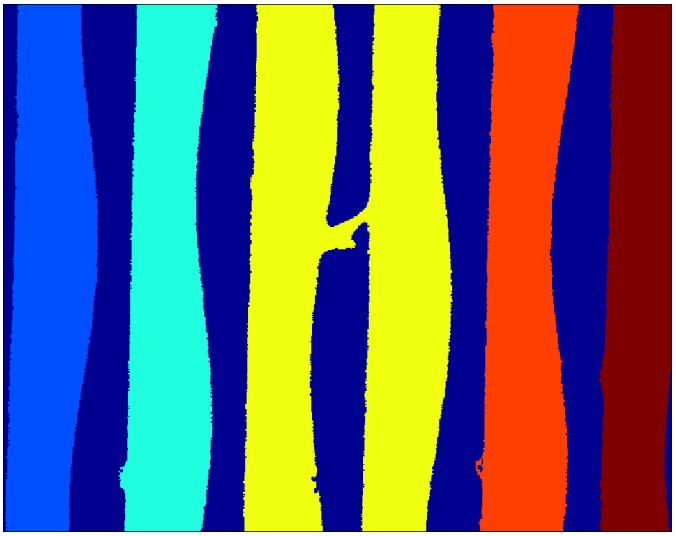


Figure 4.24: The interconnected groove valley CC (the yellow one) due to the broken ridge in Figure 4.22.

Secondly, a piece of dust or hair may fall onto the disc surface and happen to lie across a groove valley. Because the dust or hair is also a surface with slope, part of it results in invalid data after the image acquisition process and leads to the discontinuity in the groove ridges (the upper part of Figure 4.20), which results in the broken groove ridge components and a single groove valley CC similar to the first situation (the upper part of Figure 4.19).

The discontinuity in the groove bottom (Figure 4.21) is mainly caused by the acquisition failure in the record surface acquisition step due to the limited angular resolution provided by the white-light interferometry (see Section 2.4.3).

The discontinuity in the groove valleys and the groove ridges would cause errors in the groove undulation extraction step. A broken groove valley would simply interrupt the groove undulation tracking; as for the broken ridges, because they cause the adjacent groove valleys to be indexed as a single CC, during the groove undulation tracking and stitching step, the undulation segment of interest would be stitched to a successor CC that belong to a wrong groove revolution (Figure 4.25). These defects would cause discontinuity in the output audio and thus need to be detected and fixed. Note that the discontinuity in the groove bottom does not create problems in the groove undulation extraction or tracking.

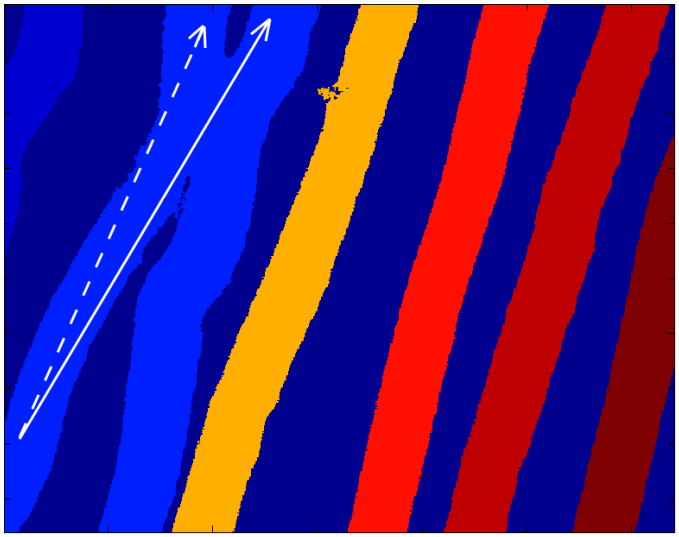


Figure 4.25: A possible groove tracking error shown in a groove valley components: The dash line represents the correct tracking direction, the solid line the wrong direction caused by the discontinuous ridge components

One may assume that the discontinuities in the three types of groove components could be fixed in arbitrary orders. However, it was found in practice that a specific order for fixing the three types of groove components is required for overcoming certain challenges in fixing the discontinuity in the groove ridges, which are explained below.

Let us focus on the problem of the discontinuity in the groove ridges. Several candidate heuristic solutions to detect the broken groove ridges were proposed and tested.

First, one may find that the interconnected groove valleys are usually wider in the radial direction (i.e., towards the disc center) than a regular groove valley; valley-width thresholding might be useful for distinguishing the interconnected groove valleys. That is, any groove valleys that are wider than a specified threshold can be taken as the interconnected ones. However, this method was not robust according to our experiments. For example, some dust-affected groove may have a groove valley that is wider than usual yet still not interconnected with the adjacent groove valleys (Figure 4.26); the width-based thresholding thus may generate false-alarms for these cases. As a detection algorithm, this method thus requires additional processes to rule out the false alarms.

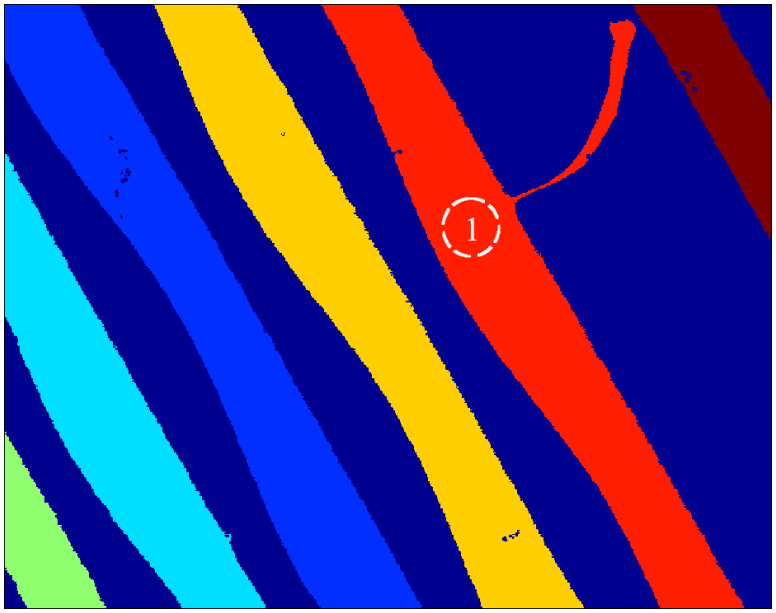


Figure 4.26: Groove valley 1 is affected by dust and wider than usual but is not interconnected with any other groove valleys

Second, the broken groove ridges become contained by the interconnected groove valley, while a regular groove valley does not contain any ridges; therefore, by detecting whether a valley CC contains one or more ridge CCs, one can detect the broken ridges or the interconnected valleys. To fix the detected defect, more heuristics are needed to identify which ridge CCs among all belong to the same revolution and should be reconnected when more than two valleys are interconnected. For example, in the upper part of Figure 4.19–Figure 4.20, there are several ridge segments contained in the single interconnected valley component (illustrated in Figure 4.27). This requires clustering the ridges contained by the valley according to the tangential continuity property of the groove components. Because the shapes of the dust and dirt vary, the topology of the broken ridges proved diverse (see Figure 4.20) and it proved difficult to rule out the CCs that are part of the noise when attempting to reconnect the ridge CCs. Because the false-ridge CCs in such cases are mainly caused by dust or hair on top of the record surface, one may assume that such areas are significantly higher than the regular groove ridges. In practice, however, it is not a robust heuristic. According to our experiments, the heights of the false-ridge CCs fluctuate along the length of the hair; these CCs thus cannot be removed with simple height thresholding without wrongly removing the true groove ridges. For example, in Figure 4.20, the false-ridge CCs that belong to the hair spreading

across the grooves may be lower than the true groove ridges if they are inside the valley area as the hair slightly warps downwards inside the valleys.

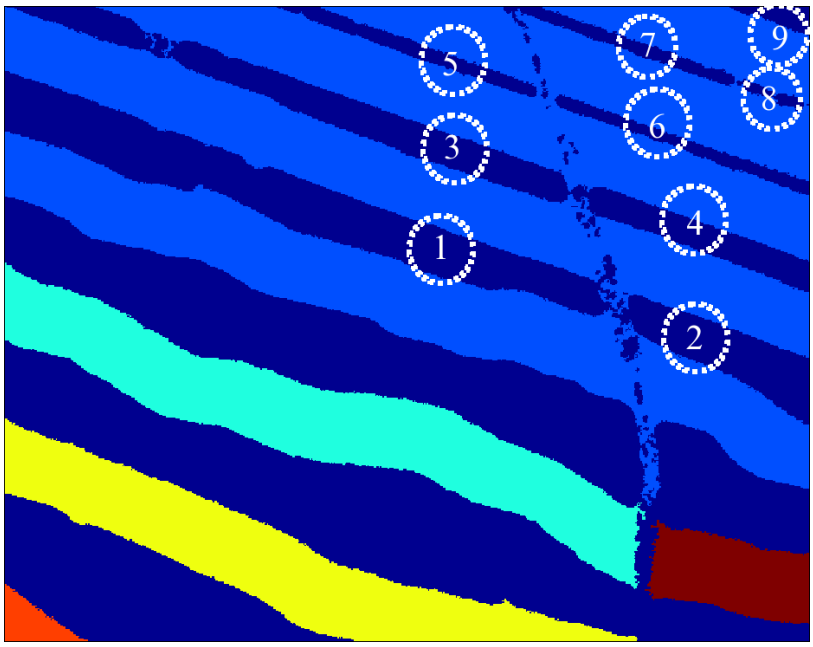


Figure 4.27: The multiple groove ridge components (annotated with indices) contained in the interconnected groove valley CC (the blue one)

A third method is to exploit the fact that in theory, a groove valley contains a single trace of groove bottom; then any groove valley component that contains two or more traces of groove bottom components (upper part of Figure 4.28) is the result of two or more interconnected valleys (upper part of Figure 4.29). After such problematic valleys are detected, they can then be fixed by connecting the ridge CCs that are located between the adjacent traces of groove bottom. These ridge CCs are supposed to belong to the same revolution of the groove ridge and thus need to be reconnected. However, this method needs additional pre-processing before it can be used for detecting the discontinuity in the groove ridges, which is explained as follows. After CCL, the discontinuous groove bottom is recognized as multiple CCs (the upper part of Figure 4.28) contained by the interconnected valley (the upper part of Figure 4.29). They need to be divided into several groups corresponding to the groove revolutions that they belong to. In other words, this method requires that the broken bottom be clustered and reconnected in order to fix the ridges. In practice, the third method proved more effective and robust than the second one. This is because the discontinuities in the groove bottom are usually short breaks

(being a few pixels apart from each other) and are easier to group into tangential groups than some cases of the groove ridges, of which the shapes of the gaps are complicated by the arbitrary shapes of the dust and dirt attached to the ridges. It is therefore more convenient to fix the broken bottom before the broken ridges.

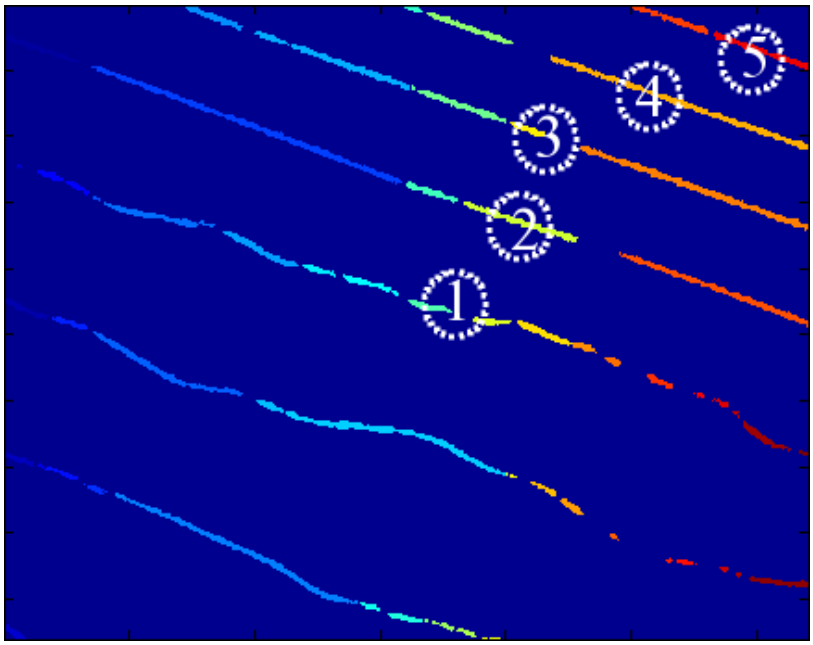


Figure 4.28: The bottom traces (annotated with indices) contained in the interconnected groove valley CCs in Figure 4.18

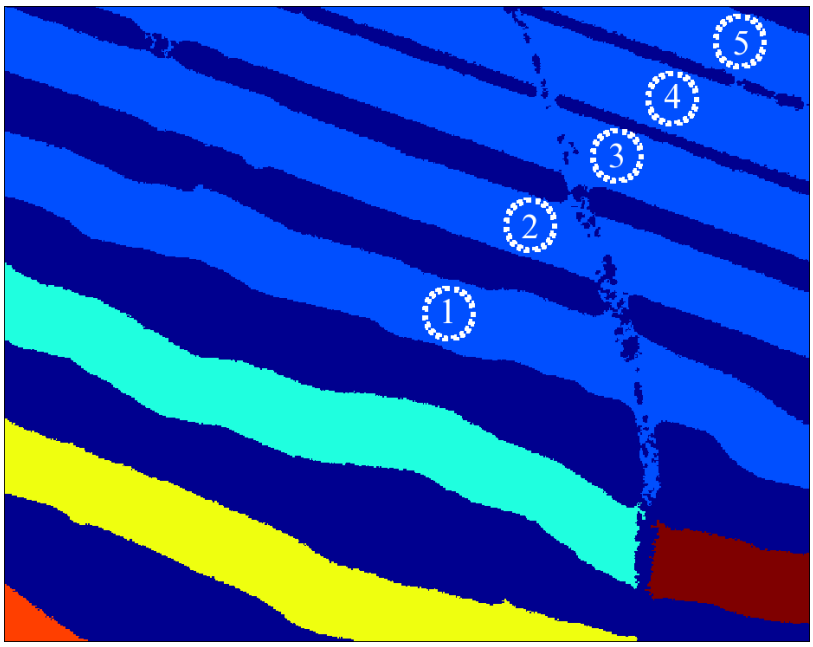


Figure 4.29: The interconnected groove valley components (annotated with indices) that form a single CC after CCL

On the other hand, it is necessary to fix the broken valley before the broken bottom because otherwise the bottom CCs that belong to the same valley would be separately contained in two valley CCs and never be considered in the clustering. For example, as for the lower part of Figure 4.21, the cyan and the red bottom components actually belong to the same groove revolution and need to be reconnected; but because they are contained by two parts of the discontinuous valley components, they would not be considered as in the same group of bottom components for clustering.

According to the above analysis, in our approach, the order in which the three types of discontinuities are fixed is: 1) the broken valleys, 2) the broken bottoms, 3) the broken ridges.

4.5.2.2 Reconnecting the Broken Groove Valleys

In the lower part of Figure 4.19, the cyan and red components are an example of the broken groove valley components.

Based on the tangential continuity property of the groove components mentioned in Section 4.2.2, heuristics are used to detect the broken groove valleys. Using the polar

coordinate system, denote a groove valley CC as $C_V^{(i)}(r,\theta)$. For each $C_V^{(i)}(r,\theta)$, check if there are any $C_V^{(j)}(r,\theta)$ that satisfy:

$$\left\{r_V^{(j)}(\theta)\right\} \cup \left\{r_V^{(i)}(\theta)\right\} \neq \Phi, \forall \theta_V^{(i)}, \theta_V^{(j)}, \theta_V^{(j)} > \theta_V^{(i)} \qquad \text{Equation 4.13}$$

which means $C_V^{(j)}(r,\theta)$ radially overlaps with $C_V^{(i)}(r,\theta)$; if such an overlapping CC exist, then they are considered two broken parts of a single valley; they are then reconnected by tangentially interpolating between their gaps. Note that the restoration here is only for correctly indexing the groove valleys; we do not try to fix the shape of the groove. The result of the reconnection is shown in Figure 4.30.

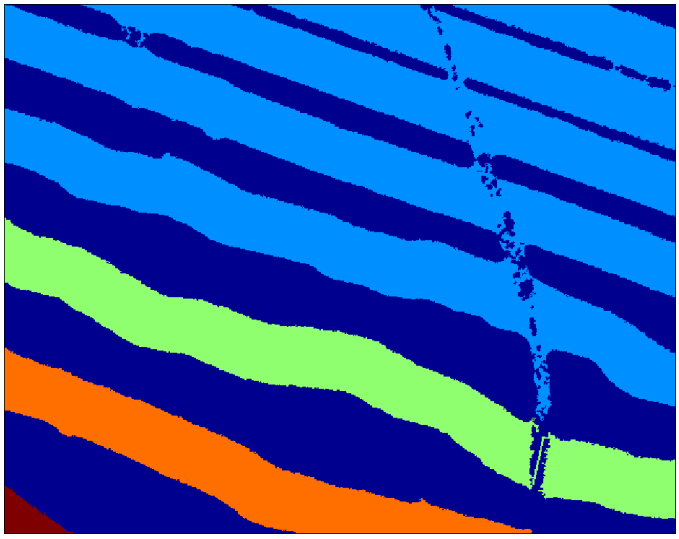


Figure 4.30: The reconnected and re-indexed groove valley component (the green one) from Figure 4.19

4.5.2.3 Reconnecting the Broken Groove Bottom

As introduced previously, the acquired groove bottom components along the same groove revolution are often discontinuous due to the inability of the WLI to resolve steep slopes. Figure 4.21 shows the typical groove bottom after CCL. We assume that the gaps between the adjacent bottom CCs in the same groove revolution are small so that the bottom continuity can be restored by interpolation without losing too much detail of the shape of the bottom. In practice, this was usually true for all our experiments under 10X magnification. If this were not the case, we would have considered the image acquisition quality unacceptable and adjust the image acquisition configuration such as the scan

length to achieve better continuity in the groove bottom instead of improving the algorithm for reconnecting them. The aforementioned situations were under the 10X magnification with the Michelson interferometer objective. On the other hand, according to our experiments, using the Mirau interferometer objective, which has higher magnification with the help of the FOV lenses (27.5X–100X), helps reduce the loss of the groove bottom.

With an acceptable groove bottom continuity, our algorithm for reconnecting the broken bottom is described below. We denote the individual broken pieces of the groove bottom CCs contained in a groove valley $C_V^{(i)}$ as $C_B^{(i,j)}, j = 1,\cdots,N_B$, where N_B is the number of bottom CCs contained in the valley.

First, the bottom CCs that radially overlap each other are grouped in their angular order into one or more groups of CCs. Take Figure 4.21 as an example, the result of this step is shown in Figure 4.31. It can be seen that most broken bottom components were correctly grouped (shown by the colour codes) except the short ones near the right side of the image.

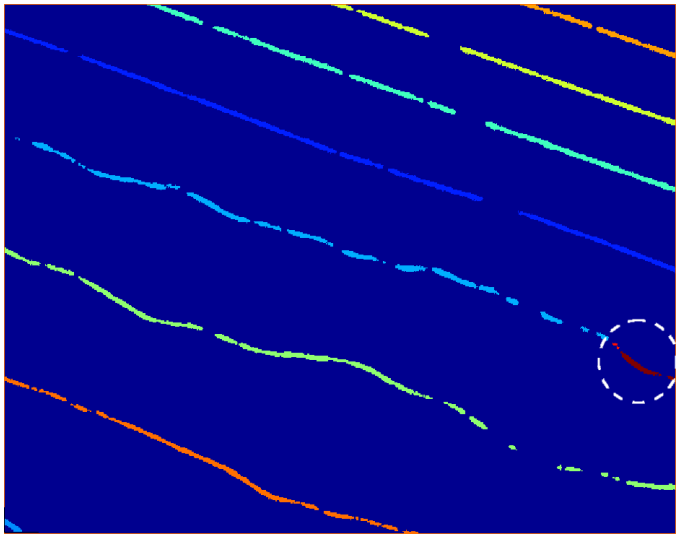


Figure 4.31: Most broken bottom components are grouped by checking radial overlapping between them; the components that belong to the same group are of the same colour; the ones that are not correctly grouped is marked with the dashed circle

Next, if not all the $C_B^{(i,j)}, j = 1,\cdots,N_B$ overlap radially (the components marked with the dashed circle in Figure 4.31), then a vector-based method is used to further group the CC groups obtained from the first step. For convenience of discussion, we still

consider the CC groups obtained from the first step CCs, $C_B^{(i,j)}$. Assume that $C_B^{(i,j)}$ are of linear shapes. Using the first and the last point in the angular direction, $P_1^{(i,j)} = C_B^{(i,j)}(r_1, \theta_1)$ and $P_m^{(i,j)} = C_B^{(i,j)}(r_m, \theta_m)$, as the vector that represents the rough orientation of the CC (see Figure 4.32), the vector of $C_B^{(i,j)}$ is denoted as $\vec{C}_B^{(i,j)} = \overrightarrow{P_1^{(i,j)} P_m^{(i,j)}}$. An inter-vector angle thresholding is then performed to iteratively cluster the angularly adjacent vectors that are aligned in direction as those that belong to the same revolution. The angle between two vectors $\vec{C}_B^{(i,j)}$ and $\vec{C}_B^{(i,k)}$ is represented as:

$$\alpha_B^{(i,j,k)} = \arccos\left(\frac{\vec{C}_B^{(i,j)} \cdot \vec{C}_B^{(i,k)}}{\left\|\vec{C}_B^{(i,j)}\right\| \left\|\vec{C}_B^{(i,k)}\right\|} \right) \qquad \text{Equation 4.14}$$

Then if $\alpha_B^{(i,j,k)}$ satisfies:

$$\alpha_B^{(i,j,k)} < \alpha_{th}^{(i,j,k)} \qquad \text{Equation 4.15}$$

then $\vec{C}_B^{(i,j)}$ and $\vec{C}_B^{(i,k)}$ are treated as aligned with each other. $\alpha_{th}^{(i,j,k)}$ is the threshold for the inter-vector angle, whose value depends on the distance, $\left|P_m^{(i,j)} P_1^{(i,k)}\right|$, between the last point of $\vec{C}_B^{(i,j)}$, $P_m^{(i,j)}$, and the first point of $\vec{C}_B^{(i,k)}$, $P_1^{(i,k)}$. The smaller such distance is, the lower the angle threshold will be. For instance, 30° was used when $\left|P_m^{(i,j)} P_1^{(i,k)}\right|$ was up to 20μm distance. The adjacent vectors that present an inter-vector angle are considered to be aligned and need to be grouped together. At the end of this process, all the individual bottom CCs contained by $C_V^{(i)}$ are grouped into N_b groups, where N_b is the number of the bottom traces found from the clustering. For example, the four vectors in Figure 4.32 are aligned and are grouped together. The result of this step on the bottom components is shown in Figure 4.33. As can be seen in the resulting colour-codes of the components, the group bottom components were correctly grouped.

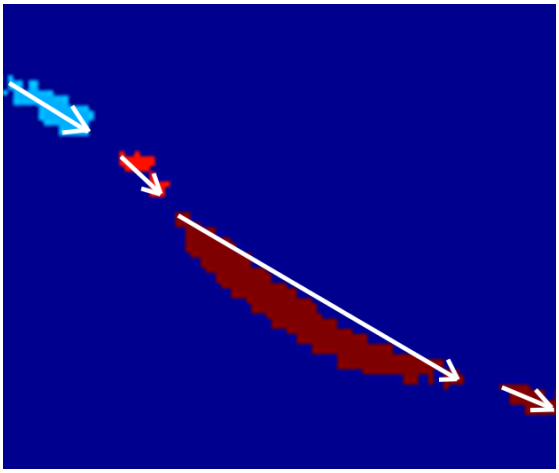


Figure 4.32: The vectors created for grouping the groove bottom components that are unable to be grouped by checking the radial overlapping from Figure 4.31: The white arrows represent the vectors of each bottom components of interest

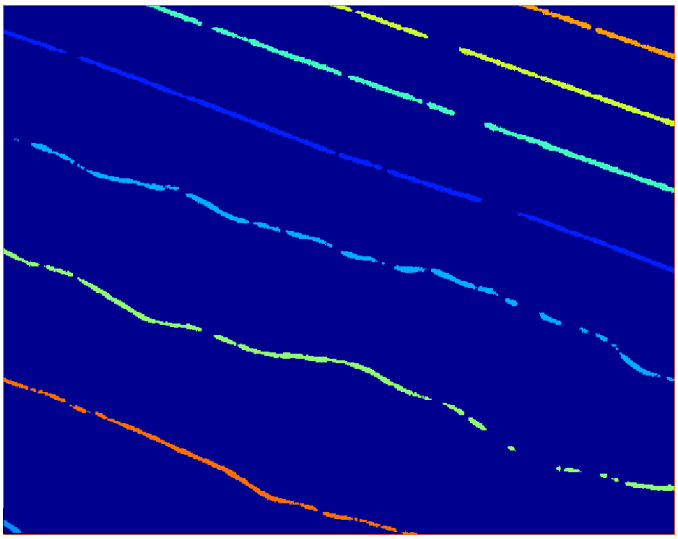


Figure 4.33: The result of the vector-based grouping for the remaining wrongly-grouped bottom components from the radial-overlap checking step

Finally, the bottom CCs within each group are then iteratively reconnected according to their angular order by interpolating between the adjacent CCs, and CCL is performed to result in N_b indexed bottom CCs contained by each $C_V^{(i)}$, $C_B^{(i,j)}, j = 1,\cdots,N_b, N_b \le N_B$, each representing a separate groove-bottom revolution. The reconnection method linearly interpolates between the last point of the CC of interest and the first point of the next CC in the angularly-sorted CCs of the same angular group. Note that they are uniformly labelled as associated with the same container valley component.

Note that if only one bottom component is found in the valley of interest, which corresponds to the case where no discontinuity issues in the bottom and the ridges of the same groove revolution, this whole process is then skipped. Figure 4.34 demonstrates the final result of reconnecting the broken groove bottom components for the images of Figure 4.18.

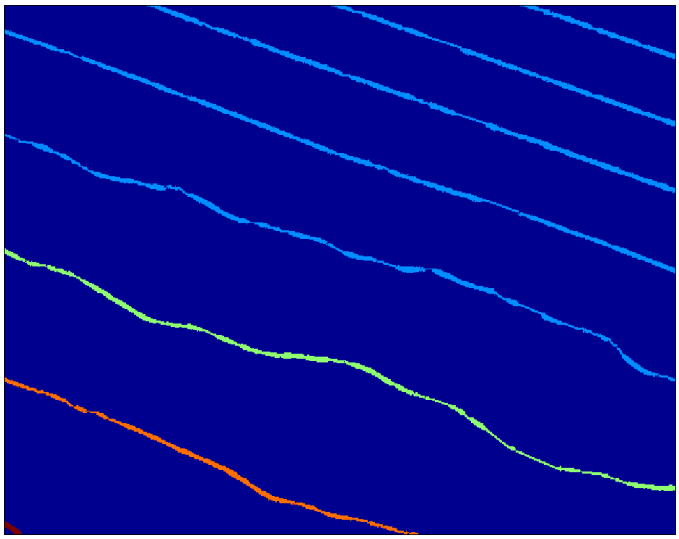


Figure 4.34: The result of reconnecting the broken groove components for the image of Figure 4.18

4.5.2.4 Reconnecting the Broken Groove Ridges

After fixing the discontinuity in the groove valleys and bottom, the broken groove ridges (or the interconnected adjacent groove valleys) are detected by checking for each $C_V^{(v)}$ the number of $C_B^{(v)}$ that it contains. The ones that contains more than one $C_B^{(v)}$ indicate a discontinuity in the groove ridges on the same groove revolution of the $C_V^{(v)}$.

To fix the broken groove ridges, first, the groove ridges that belong to the same revolution are clustered by checking which bottom CCs are to each side of them. Denote the radius of a point at the azimuth θ in the groove component $C_R^{(i)}$, $C_V^{(i)}$, and $C_B^{(i)}$ as $r_R^{(i)}(\theta)$, $r_V^{(i)}(\theta)$, and $r_B^{(i)}(\theta)$, respectively. The geometric relationship of a groove ridge CC being located between two groove bottom CCs is defined as: $C_R^{(i)}(r,\theta)$ is between $C_B^{(j)}(r,\theta)$ and $C_B^{(k)}(r,\theta)$, if $\forall\, r_R^{(i)}(\theta)$,

$$r_B^{(j)}(\theta) < r_R^{(i)}(\theta) < r_B^{(k)}(\theta) \qquad \text{Equation 4.16}$$

Note that this definition can be extended to apply to any combination of the groove component types. The $C_R^{(i)}$ that are between $C_B^{(j)}$ and $C_B^{(k)}$ are then reconnected in their angular order by interpolating between their gaps. CCL is then performed to re-index the ridge CCs.

Recall that we did not remove blob noise contained by the groove valley components because broken groove ridges may be wrongly removed as well (see Section 4.5.1). Since the broken groove ridges are now all correctly reconnected, any ridge CCs that are contained by a groove valley are blob noise. As a post-processing step, such blob noise is removed by labelling them as part of their container groove valleys.

After this step, the groove components are considered to be correctly segmented and indexed. The resulting groove valley, ridge, and bottom components are shown in Figure 4.35–Figure 4.37.

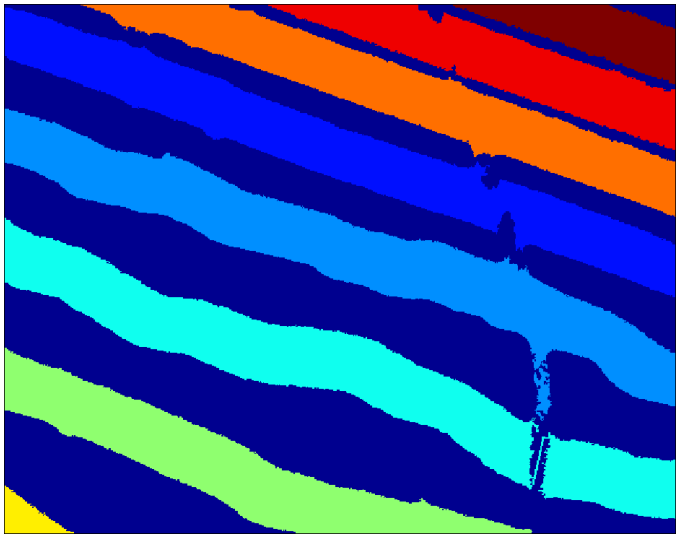


Figure 4.35: The resulting groove valley components after fixing all discontinuities in the groove components in the image of Figure 4.18

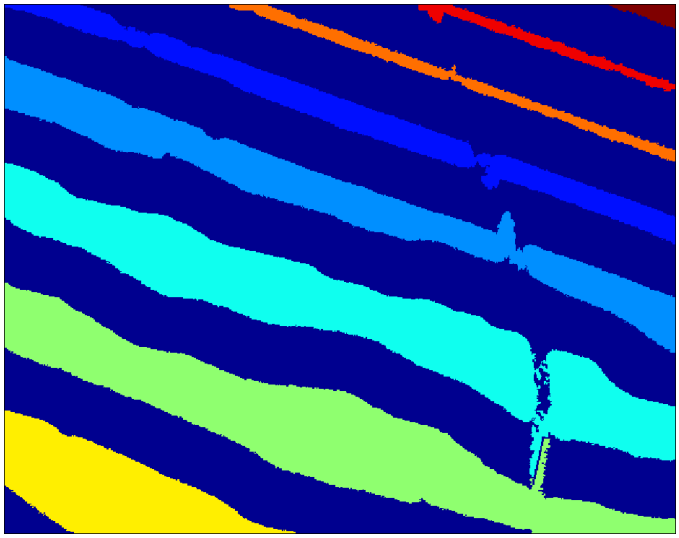


Figure 4.36: The resulting groove ridges after fixing all discontinuities in the groove components in the image of Figure 4.18

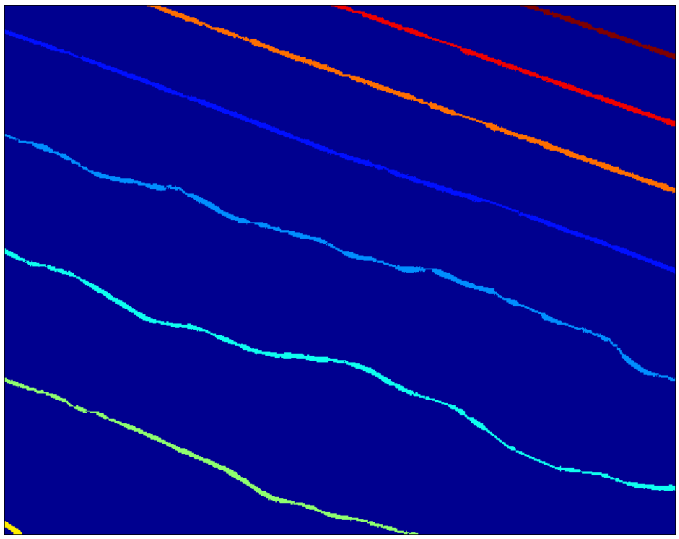


Figure 4.37: The resulting groove bottom components after fixing all discontinuities in the groove components in the image of Figure 4.18

4.5.2.5 Discussion

According to our experiments with signals from three different stereo LPs, each containing about 2000–5000 images in the acquisition results, the average per-image occurrence rates of the discontinuities in the groove valley, ridge, and bottom in the images acquired under 10X magnification are about 0.3%, 20%, and 90%, respectively. The statistics are obtained by considering an image that shows any type of defect as one occurrence of the defect of the type. The first two types of discontinuities suggest that it

would be desirable to perform the record surface profile acquisition in a dust-free environment. The discontinuity in the groove bottom is mostly determined by the chosen magnification used in the record surface image acquisition process.

The algorithms presented in this section strive only to restore the groove component continuity and correctly index the groove components; they, however, do not attempt to restore the actual groove shapes. They therefore leave the type of noise called *stings and cuts* in this dissertation (Figure 4.38). Section 4.6.2 tackles these types of noise.

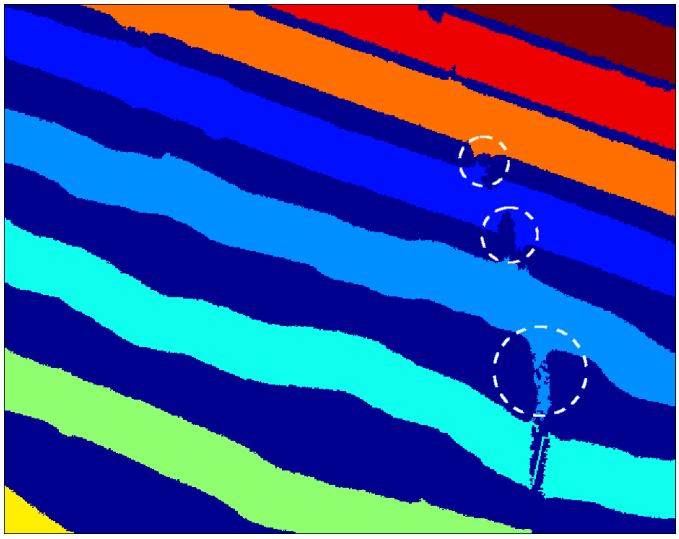


Figure 4.38: The sting and cut noise (marked with dashed circles) that remains as part of the resulting groove components from fixing the component discontinuities

4.6 The Groove Undulation Extraction

After removing various types of noise and correcting the indexing errors for the groove components, the groove undulation needs to be extracted from the groove components. The edge detection technique for 2D images is used for extracting the boundaries between the groove valleys and their adjacent groove ridges. Then the post processing is performed to fix the artefacts, the stings and cuts, which are the debris from reconnecting the discontinuous groove components presented in Section 4.5.

4.6.1 Edge Detection

As introduced in 4.2.2, the inner and the outer edges of the groove valley with the adjacent groove ridges represent the groove undulation. It is thus necessary to extract

these two edges for each groove valley segment in each image. Edge detection for 2D images is a well-documented image processing technique. For the detail of the generic edge detection techniques, readers can refer to literature such as Ziou and Tabbone (1998) and Shapiro and Stockman (2001, 326). However, there are several context differences between our task and the generic 2D edge detection applications.

First of all, the generic edge detection usually tackles images without prior knowledge of the target objects in the images. It seeks the image intensity variations that represent the physical edges (Ziou and Tabbone 1998, 3). In our case, however, the groove components are already segmented and indexed; the problem is thus simplified to detecting the boundaries between the known pixel areas that represent the indexed groove valley components and ridge components. This is similar to extract edges in a binary image.

Secondly, in our groove undulation extraction task, it is also necessary to distinguish between the inner and the outer edges in addition to extracting the raw edges. This requires the knowledge of our task-specific coordinate system.

As can be seen from the above context differences, generic 2D edge detection is not suitable for our tasks. A custom edge detection algorithm is thus used. Our algorithm for detecting edge of a specific groove valley segment is described below.

1) Along its tangential direction, the valley of interest, $C_V^{(i)}$, is divided into N_w two-pixel-wide angularly-overlapping angular windows along its tangential direction, which are denoted as $W_V^{(i)}(\theta_j)$, $j = 1,\cdots,N_w$. Each window corresponds to a narrow group of pixels that covers the groove radial cross-section at the azimuth θ_j. Overlap between the adjacent angular windows is used to ensure the complete coverage of the edge series. The overlaps avoid skipping the edge points due to the round-off errors in the window division
2) For each window, the inner edge is then the point with the shortest radius in $W_V^{(i)}(\theta_j)$, while the outer edge the point with the longest radius in the same window (Figure 4.39). The window is of a narrow width so certain edge points of the window may not be extracted, but the angular overlap with the adjacent ensures that no edge points be left out after all the windows are processed.

3) The possible redundant edge points introduced by the overlap reserved in Step 1) are removed.

An example image, its groove valley components, and the results of the detected inner and outer edges, are shown in Figures 4.40–4.42. In Figure 4.42, the upper edge of each pair of edges is the inner edge and the bottom one the outer edge.

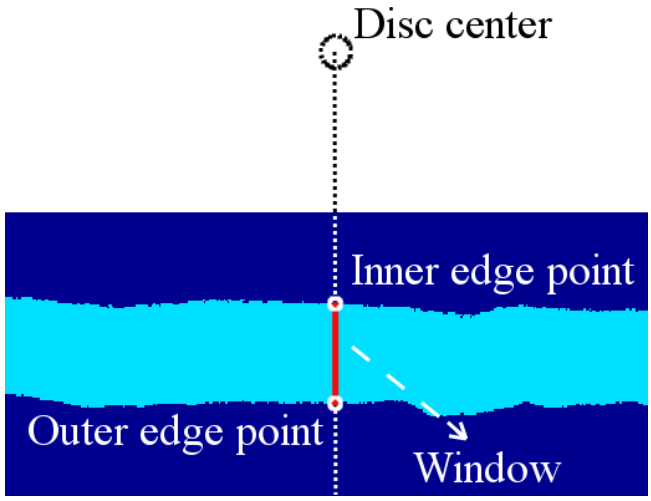


Figure 4.39: The inner and outer edge points of a window along the radius across the groove cross-section

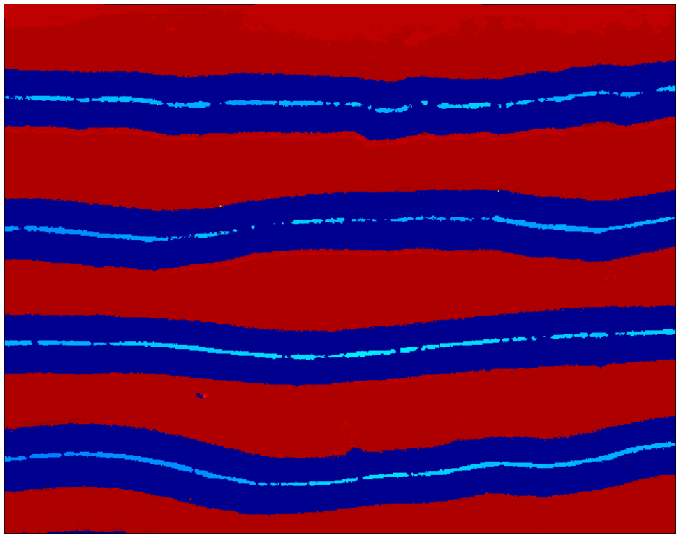


Figure 4.40: An example record surface profile image

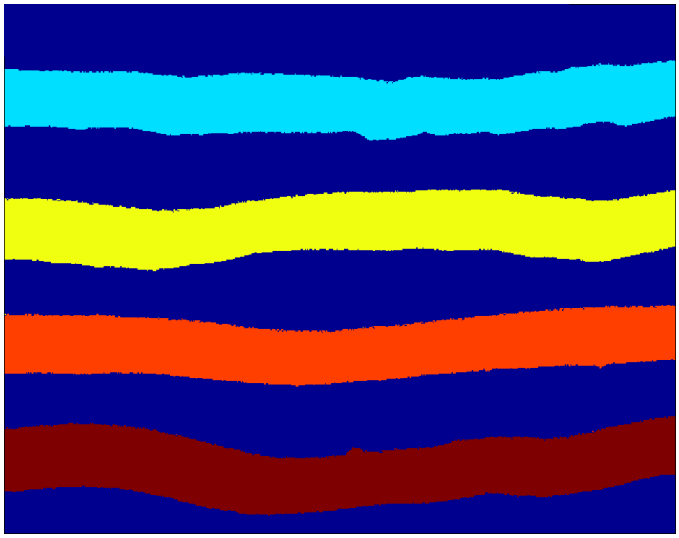


Figure 4.41: The groove valley components extracted from the image of Figure 4.40

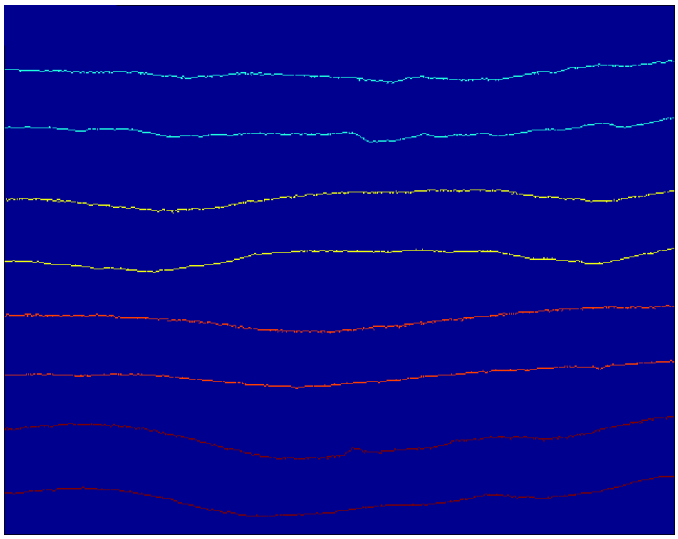


Figure 4.42: The extracted inner and outer groove undulation for the grooves in the image of Figure 4.40: The edges are in the same colour as the corresponding valley components in Figure 4.41

4.6.2 Removing the Sting and Cut Noise

The immediate result of the edge detection, the raw groove undulation segments, may include the sting and cut noise (Figure 4.43); the ones that appear to be convex towards the outside of the groove valley are called *the stings* while the ones that appear to be concave towards inside the groove valley are called *the cuts*. Some are the side effects of fixing the discontinuous groove components with the simple interpolation discussed in Section 4.5.2 (Figure 4.38); others are the possible groove damages. If untreated, the sting

and cut noise would become part of the output audio signal and generate noise. They thus need to be detected and removed or reduced.

No matter how complex their shapes are, there are two basic types of the sting and cut noise, shown in Figure 4.44, where the surrounding horizontal lines represent the true groove undulations (i.e., the extracted edges between the groove valley and ridge components):

1) Type A: The shape of the sting and cut noise is in the manner that no portion of the noise has an angular overlap with the true edges.
2) Type B: The shape of sting and the true edges are at a sharp angle so that they share a common angular range.

One may think the stings and cuts could be treated as equivalent to high-frequency impulse noises in a 1D analog signal and thus can be removed by applying low-pass filtering. However, according to our experiments, the stings are of diverse shapes and often contain broadband frequency content that spans large portion of the human hearing range. It thus proved difficult to be removed without compromising the frequency response of the reproduced audio signal. We thus apply a heuristic smoothing algorithm to reduce the sizes of the stings.

A two-pass heuristic approach is adopted to reduce the sizes of the stings and cuts. In the first pass, for each of the inner and outer edges, the Type B noise are detected by using the angular window introduced in Section 4.6.1 along the tangential direction of the edge to find the angularly overlapping points (radially aligned points). For example, in Figure 4.44, the edge points, R1, R2, and R3, are aligned along the same radius. This indicates a sting or cut at a sharp angle with the groove. In the example, only R3 is the true edge point. Then all but one of these points are removed based on the following simple rules:

1) For the inner edge, only the outermost point (i.e., the one with the greatest radius) among the radially aligned points is preserved.
2) For the outer edge, only the innermost point (i.e., the one with the smallest radius) among the radially aligned points is preserved.

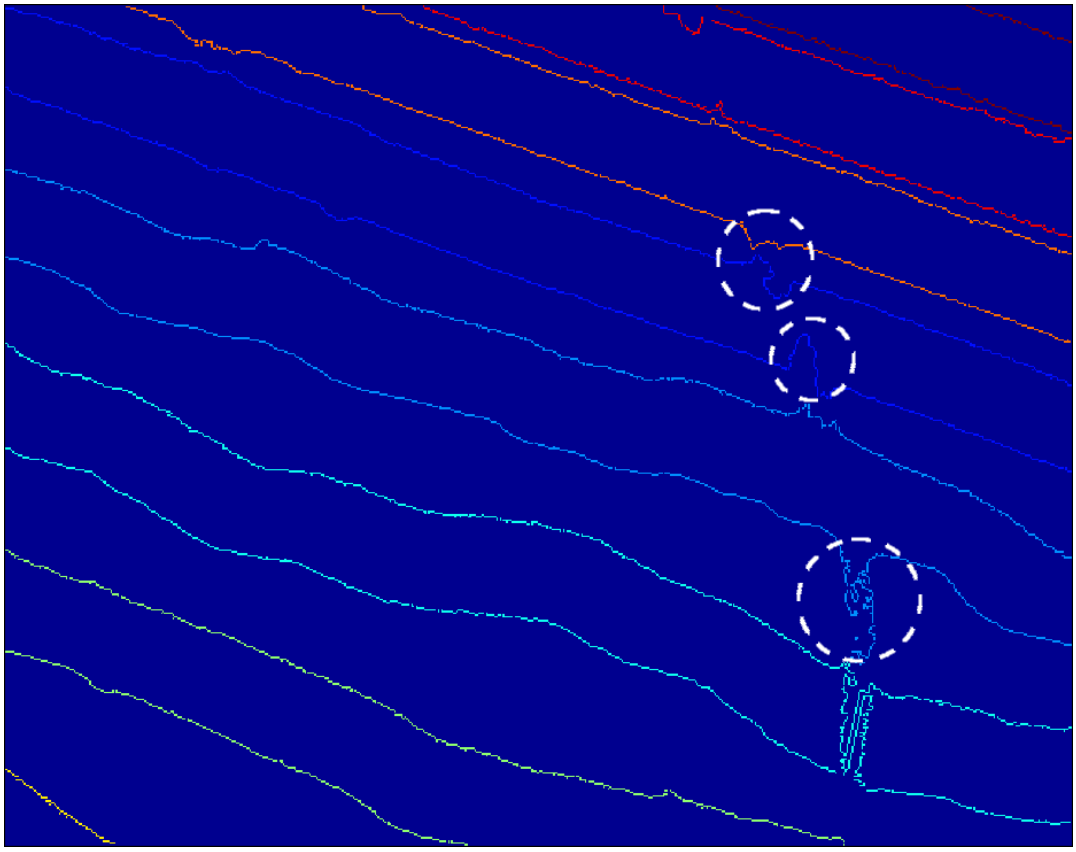


Figure 4.43: The sting noise (marked with dashed circles) present in the raw groove undulations extracted from the valley components in Figure 4.38

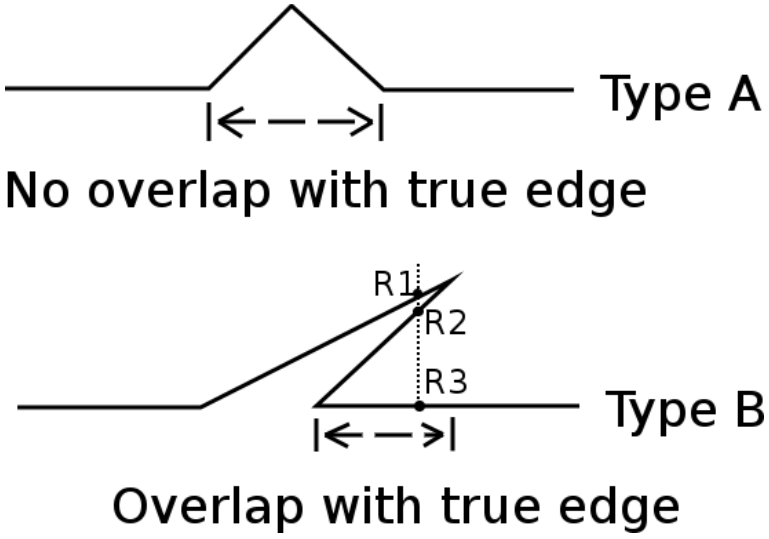
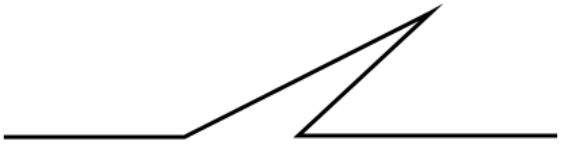


Figure 4.44: The types of stings: Type A does not have angular overlap with the true edge; Type B has a portion that overlaps with the true groove edge

For each acquired image, all azimuths along each inner and outer edge segments are checked to remove any possible radially aligned sting points in the above manner. After the first pass, all Type B noise is reduced to Type A noise (illustrated in Figure 4.45).



Type B noise before the first pass of our algorithm

Type B noise after the first pass of our algorithm

Figure 4.45: The result of the first pass of the algorithm for removing the sting and cut noise

In the second pass, the Type-A noise are detected and removed based on the assumed shape consistency of the groove cross-sections (Figure 4.46). In general, any abrupt changes in the radii along the groove undulation are considered noise. The degrees of such radius changes are measured by exploiting the knowledge about the groove cross-section shape. Refer to Figure 2.15 in Section 2.3.2.4, it is assumed that the groove cross-section can be modeled as a triangle by a shape descriptor; the descriptor remains relatively consistent along a short angular range of the groove. Therefore, if evaluated by short windows along the groove, any dramatic change in the cross-section shape descriptor would indicate an instance of noise. The detail of the strategy is explained as follows.

Because the groove represents the cutting stylus trajectory of the 45/45 system (see Section 2.3.2.4), the groove cross-section remains a consistent triangular shape along the stylus trajectory. For laterally cut disc records, the shape and the size of the triangle theoretically remain identical at any angular position along the groove. In the case of the stereo disc record, although the groove width and depth may vary along the stylus trajectory, some basic parameters are assumed to be stable within a short angular range.

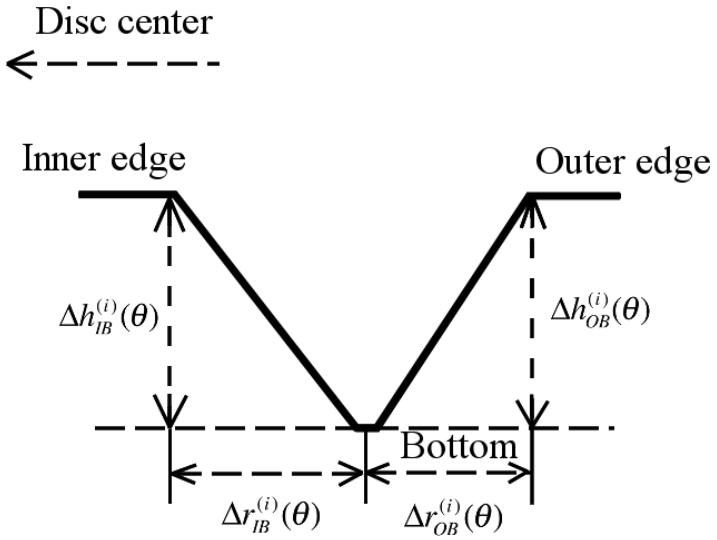


Figure 4.46: The groove shape parameters used for the second pass of our algorithm to remove the Type A sting and cut noise

As shown in the groove cross-section in Figure 4.46, the consistent groove cross-section shape parameters that are considered here include:

1) The ratio between the inner-sidewall width $\Delta r_{IB}^{(i)}$, and the outer-sidewall width $\Delta r_{OB}^{(i)}$, at the azimuth θ:

$$R_{IO}^{(i)}(\theta) = \frac{\Delta r_{IB}^{(i)}(\theta)}{\Delta r_{OB}^{(i)}(\theta)} \qquad \text{Equation 4.17}$$

2) The ratio of the inner-sidewall width over its depth, which is the height difference between the inner edge point and the median position of the groove bottom at the azimuth θ (Note that this ratio does not exist and is denoted as NaN if the bottom is missing at the position):

$$R_{IB}^{(i)}(\theta) = \frac{\Delta r_{IB}^{(i)}(\theta)}{\Delta h_{IB}^{(i)}(\theta)} \qquad \text{Equation 4.18}$$

3) The ratio of the outer-sidewall width over its depth, similar to 2) which is the height difference between the inner edge point and the median position of the groove bottom at the at the azimuth θ:

$$R_{OB}^{(i)}(\theta) = \frac{\Delta r_{OB}^{(i)}(\theta)}{\Delta h_{OB}^{(i)}(\theta)} \qquad \text{Equation 4.19}$$

We assume the above three parameters vary little within a short tangential angle range (temporal range):

$$\left|R_{IO}^{(i)}(\theta) - R_{IO}^{(i)}(\theta + \Delta\theta)\right| < \varepsilon_{IO} \qquad \text{Equation 4.20}$$

$$\left|R_{IB}^{(i)}(\theta) - R_{IB}^{(i)}(\theta + \Delta\theta)\right| < \varepsilon_{IB} \qquad \text{Equation 4.21}$$

$$\left|R_{OB}^{(i)}(\theta) - R_{OB}^{(i)}(\theta + \Delta\theta)\right| < \varepsilon_{OB} \qquad \text{Equation 4.22}$$

where $\theta - \delta < \Delta\theta < \theta + \delta$, δ, ε_{IO}, ε_{IB}, and ε_{OB} are the adaptive thresholds; they are derived from the groove valley of interest along its entire angular range. Then any edge point, at which Ratio 1), 2), and 3) fail to conform to Equation 4.20–Equation 4.22, respectively, is considered part of a sting on either the inner or the outer edge. Such points are denoted as $S_I^{(i)}$ and $S_O^{(i)}$ for the ones on the inner edge and the outer edge, respectively.

A heuristic smoothing algorithm is applied to reduce the sizes of the stings. The algorithm treats the $S_I^{(i)}$ and $S_O^{(i)}$ as the central nodes of the stings and iteratively taper off the radii of these edge points and their surrounding points until Equation 4.20–Equation 4.22 are satisfied for these points and no abrupt radius fluctuation can be found among them. An example of the result of the sting removal for Figure 4.43 is shown in Figure 4.47.

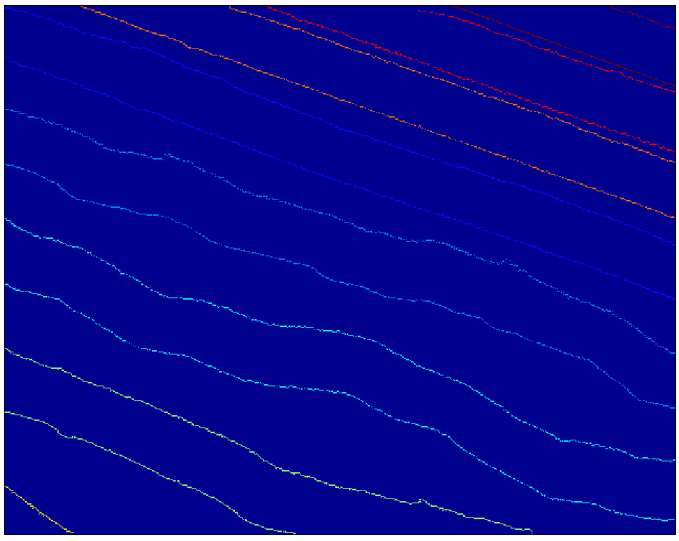


Figure 4.47: The result of removing the sting and cut noise in the groove undulation segments in Figure 4.43

Note that by adjusting the thresholds ε_{IO}, ε_{IB}, and ε_{OB}, the sizes of the stings and cuts to ignore can be controlled to minimize false alarms.

4.7 The Groove Undulation Tracking and Unwrapping

After the inner and the outer undulations of the groove segments in each record surface profile image are extracted and the stings in them removed, they are tracked and stitched together into two 1D time series, denoted as $U_I(\theta)$ and $U_O(\theta)$. Taking the inner groove undulation as an example, the tracking and stitching workflow is described below:

1) *Selecting a starting image and starting groove revolution.* Among all the acquired images, manually select an image that contains the outermost groove revolution of the signal of interest; use the groove inner edge of the outermost groove revolution as the initial undulation segment.
2) *Finding successor undulation segment.* Given an undulation segment of interest, denoted as $E_I^{(i)}$, in the cell-image $I(r,c)$, consider all the undulation segments from the 8-neighbor cell-images in the grid of the GABM (grid automatic batch measurement) a candidate successor undulation segment set, $E_{IS}^{(j)}$, $j=1,\cdots,N_s$, where N_s is the number of the candidates, then for a matching successor undulation segment $E_{IS}^{(s)}$, $s\in\{1,\cdots,N_s\}$, satisfies:

$$s=\arg\min_{j\in\{1,\cdots,N_s\}}\sum_{\theta_{om}^{(j)}<\theta<\theta_{oM}^{(j)}}\left|r_I^{(i)}(\theta)-r_{IS}^{(j)}(\theta)\right| \qquad \text{Equation 4.23}$$

$$\theta_I^{(i)}(N_I)<\theta_{IS}^{(j)}(N_{IS}) \qquad \text{Equation 4.24}$$

where $r_I^{(i)}(\theta)$ and $r_{IS}^{(j)}(\theta)$ represent the radius of a point of $E_I^{(i)}$ and that of $E_{IS}^{(j)}$ within the overlap region of the two adjacent images; $[\theta_{om}^{(j)},\theta_{\mathrm{oM}}^{(j)}]$ is the angular range of the overlap region; $\theta_I^{(i)}(N_I)$ and $\theta_{IS}^{(j)}(N_{IS})$ are the azimuths of the end points of $E_I^{(i)}$ and $E_{IS}^{(j)}$, N_I and N_{IS} the number of points of both segments, respectively. When multiple successor candidates are available, the one with the largest angular range is selected.

3) *Unwrapping the azimuths.* As we follow the groove undulations, the azimuths that were initially between 0 and 2π are unwrapped according to the groove revolution to which the undulation points belong.

4) *Stitching.* Although no actual stitching was performed during the image alignment step (see Section 3.5), that is, the images were not stitched together to create a single giant image, it has to be done for the groove undulation segments to have the complete time-series that are eventually converted the output audio signals. Therefore, as the new undulation segment is found to be the successor of the current undulation segment of interest, they are stitched together. Because the stitching in Section 3.5 was only performed to remove the 2D image misalignment, the possible discrepancies in heights are left unresolved. In this stitching step, any such discrepancies in heights are resolved by linearly shifting the successor segment height to minimize the height difference between both segments in the overlap region.

The workflow for tracking the outer groove undulation is the same as the above. Note that one may argue that it could be more convenient to track the groove valley component as a whole instead of tracking the two edges separately. However, this proved problematic in practice. Because in each image, the inner and the outer undulation segments of the same groove revolution may end at different edge of the FOV, their successors would thus be in different adjacent images. As discussed in Section 3.5, it is too expensive to stitch all the images together before extracting the groove undulations. As a result, the list of images that comprises the inner undulation time series would usually be different from that for the outer undulation time series. Therefore, the inner and outer groove undulations need to be tracked and stitched separately (Figure 4.48).

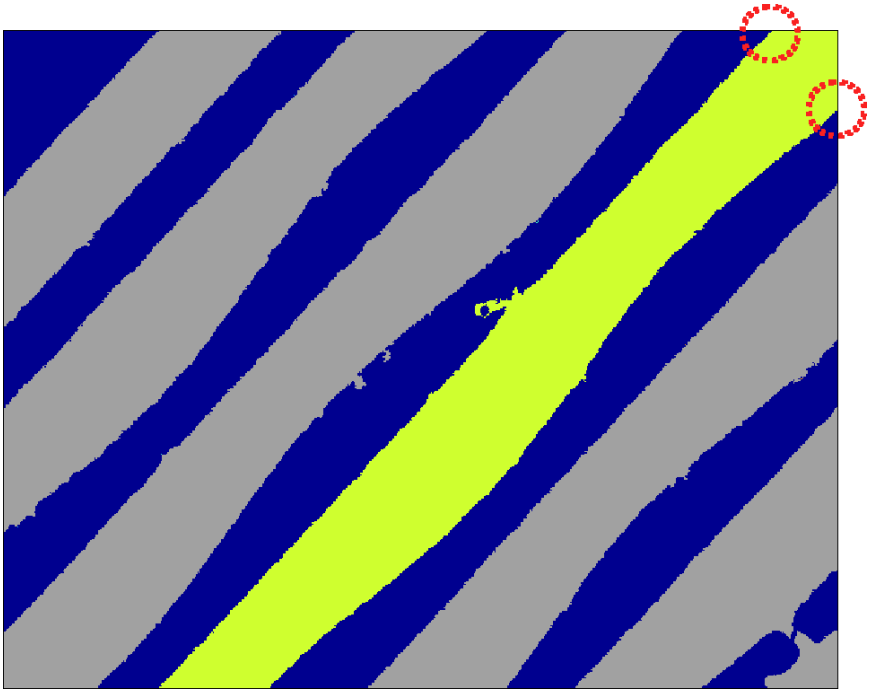


Figure 4.48: The inner and outer undulation of a groove have different successor FOVs: The successors of the inner undulation is in the upper neighbour FOV while that of the outer undulation is in the right neighbour FOV

The results of the groove tracking and stitching are two 1D time series that contain the stereo channel audio information. They will be further processed to reproduce the stereo audio. These processes will be explained in the next chapter.

4.8 Summary

In this chapter, the disc record groove model in the acquired record surface images and the coordinate system for locating the grooves were defined. The groove was decomposed into several components by using height thresholding, and these components were indexed by using the connected-component labelling. The noise and defects were present in the raw groove components. They were then removed by using heuristic algorithms and the groove components are then correctly re-indexed. The groove undulations were defined as the boundary between the groove valley and ridge components, which were extracted by using custom edge-detection algorithms. The noise in the undulations was further filtered to avoid audio quality degradation. Finally, the groove undulations were tracked across the acquired record surface profile images and stitched together into two 1D time series that encode the stereo channel audio information.

Chapter 5 The Audio Reproduction

5.1 About the Chapter

After extracting the groove undulations and stitching them together into two time series (the left and right channels), in this chapter we then discuss the workflow for reproducing the stereo audio from the time series.

The basic audio reproduction workflow is introduced in Section 5.2, which consists of three steps: First, the raw groove undulation time series are resampled at a specific audio sampling rate (Section 5.3). The resampled time series are then differentiated to reproduce the raw stereo audio (Section 5.4), which are further filtered to revert the RIAA (Recording Industry Association of America) emphasis equalization (Section 5.5).

Finally, the pitch fluctuation defect usually present in the raw reproduced audio as a result of inaccurate record center estimation is introduced, and the solution to remove the defect is discussed (Section 5.6). It should be noted that this correction is, in fact, performed at the beginning of the audio reproduction workflow.

5.2 The Basic Workflow for Stereo Audio Reproduction

The left and right channel signals, $A_l(t)$ and $A_r(i)$, respectively, of a stereo audio signal of an LP are represented by:

$$A_l(t) = \frac{dU_i(t)}{dt} \qquad \text{Equation 5.1}$$

$$A_r(t) = \frac{dU_o(t)}{dt} \qquad \text{Equation 5.2}$$

where $U_i(t)$ and $U_o(t)$ are the inner and outer groove undulation radius time series converted to a time series at a specific audio sampling rate. Because the grooves of the stereo LPs are cut with the constant velocity schemes (see Section 2.3.2.6), the audio signal is encoded in the radial velocity of the cutting stylus. Numerical differentiation (derivative) is thus required to be performed to the groove undulation time series to decode the stereo channel signals.

Note that although the stereo audio signals are actually encoded in the groove sidewall displacements (see Section 2.3.2.3.2), they still can be represented by the groove undulation time series because of the 45/45 system used to cut the grooves (see Section 2.3.2.3). We have

$$\frac{dU_i(t)}{dt} \propto \sqrt{2} \cdot \Delta W_i(t) \qquad \text{Equation 5.3}$$

$$\frac{dU_o(t)}{dt} \propto \sqrt{2} \cdot \Delta W_o(t) \qquad \text{Equation 5.4}$$

where $\Delta W_i(t)$ and $\Delta W_o(t)$ are the groove sidewall displacement time series. For simplicity the scalar $\sqrt{2}$ is ignored in Equation 5.1 and Equation 5.2.

From the above analysis, and also taking into account the fact that the disc record grooves were cut with the RIAA emphasis equalization (see Section 2.3.2.6), the basic stereo audio reproduction workflow is as follows:

1) Resample the raw groove undulation time series to obtain the digital time series at a fixed audio sampling rate.
2) Differentiate the resampled time series.
3) Apply RIAA de-emphasis equalization filtering to restore the frequency content.

5.3 Resampling the Groove Undulations

5.3.1 The Purpose of the Resampling

The raw groove undulation (the inner or the outer one) extracted from the acquired record surface profile images is a discrete time series of variable temporal resolution. Here the temporal resolution refers to the time interval between arbitrary adjacent data points in the time series. The variable temporal resolution can be explained by the following factors:

1) The outer revolutions of a spiralling disc record groove are longer than the inner ones. They nevertheless are of the same time duration because of the constant rotation speed. For example, this duration is about 1.8 seconds for a 33rpm stereo LP (long-playing) record. The temporal resolution in terms of the spatial distance

along the groove is therefore variable (increasing from the outer to inner revolutions)

2) With our acquisition system, all the record surface profile images are rasterized with the uniform pixel resolution when using the same overall microscope magnification.
3) The pixels of an image are square in shape and tiled according to the square pixel tessellation. Trajectories of the same number of pixels but different orientations may span different spatial lengths (Figure 5.1).
4) Unlike the image acquisition systems of Stotzer (2006) and Tian (2006), our acquisition system does not unwrap the groove during the scanning. The groove segments in the individual images are therefore not straight but of various orientations.

Due to 1) and 2), the temporal resolution of the extracted groove undulation time series is variable between different grooves. From 3) and 4), the temporal resolution of the extracted groove undulation time series also varies on a sub-revolution scale. The raw groove undulation time series thus need to be converted into a time series with a constant audio sampling rate.

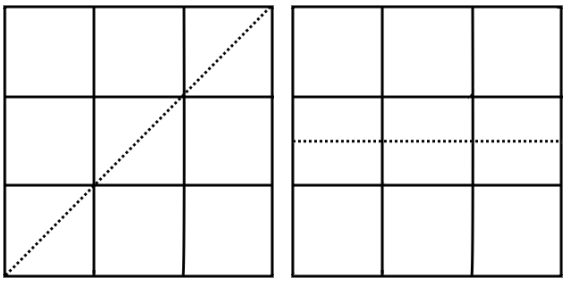


Figure 5.1: The square pixel tessellation: The pixels along the diagonal of the area (the left figure) span a longer distance (as of the dotted line) than the horizontal pixel groups of the same number of pixels (the right figure)

5.3.2 The Equivalent Audio Sampling Rate

To be able to resample the groove undulation time series, it is necessary to determine the equivalent audio sampling rate that the extracted groove undulation time series can represent.

It is discussed in 5.3.1 that the temporal resolution along the groove undulation temporal series varies from groove to groove and also fluctuates on a sub-revolution spatial scale; it is thus difficult to derive the equivalent sampling rate that applies to arbitrary portions of the time series. However, it is still useful to calculate the lowest and highest average sampling rates to determine whether the groove undulations are over-sampled or under-sampled before resampling.

For a rough estimation, if we treat the groove undulation time series as the equivalent output audio samples, the average sampling rate of an arbitrary groove revolution S_R is:

$$S_R = \frac{N_R}{T_R} = \frac{L_R \cdot RPM}{60 \cdot l_p} \qquad \text{Equation 5.5}$$

where N_R and L_R the number of points and the spatial length of the groove of interest, respectively, T_R the temporal duration of any groove revolution, and l_p the lateral pixel resolution of the image. For a 33rpm record, assuming each revolution of groove is a perfect circle, S_R becomes

$$S_R = \frac{2\pi \cdot \left(33\frac{1}{3}\right)}{60 l_p} R \approx \frac{3.49R}{l_p} \qquad \text{Equation 5.6}$$

where R represents the radius of the groove of interest. Because l_p is constant for the record surface profile images obtained by a given overall magnification of our acquisition system, S_R is then proportional to the length of the groove revolution of interest. The innermost and the outermost grooves represent the revolutions that provide the lowest and the highest equivalent audio sampling rate. According to RIAA (1963), the innermost radius is about 4.75inch (12.07cm) and the outermost 9.50inch (24.13cm) for a 12-inch (30-cm) 33rpm stereo disc record. Since the groove length decreases from the outer grooves to the inner ones, it thus requires that the resampling frequency be no greater than S_R of the innermost groove of the groove undulation time series of interest.

Table 5.1 shows the average sampling rates provided by our image acquisition systems under various overall magnifications for the innermost and outermost grooves of the 12-inch (30-cm) 33rpm stereo disc records.

	2.75X	5X	10X	27.5X	50X	100X
Lateral image resolution (μm/pixel)	3.95	2.00	1.01	0.39	0.20	0.10
Innermost groove sampling rates (kHz)	53.27	106.12	213.92	539.10	1036.70	2246.10
Outermost groove sampling rates (kHz)	128.96	256.91	517.90	1305.10	2509.80	5438.00

Table 5.1: The various average equivalent audio sampling rates provided by our acquisition system (Equation 5.5) for the innermost and outermost grooves of the stereo disc records

It can be seen in the above table that when using an overall magnification of 10X or above, the resulting signal is oversampled compared to the CD (compact disc) sampling rate (44.1kHz).

5.3.3 The Resampling Method

The resampling process aims to convert the raw groove undulation time series to a discrete series at a constant time interval with the least distortion to the original time series. Because a raw groove undulation has variable equivalent audio sampling rates (see 5.3.1), the resampling here is different from the sampling-rate conversion for digital audio signals. The input and output signals in the sampling-rate conversion are both digital signals at two different constant sampling rates. In comparison, our resampling task is to fit a curve to the raw groove undulation time series with a new equally-spaced time series.

The extracted raw groove undulation time series are usually noisy due to the roughness of the record groove. If the time series are not smooth enough, the later differentiation step will tend to enlarge such roughness, resulting in noisy audio signals. The goal of the resampling process is thus to create a curve that is as smooth as possible without distorting the audio information encoded in the time series.

Several general approaches were examined:

1) Global linear interpolation: Treat the groove undulation time series as an entire curve and then interpolate the radii of each data point to an evenly-spaced time series at a specified sampling rate.
2) Global linear interpolation with moving-average filtering: Use 1) for resampling, but pre-process the time series with moving-average filtering.
3) Window-based curve fitting: Divide the groove undulation time series into a number of adjacent windows, fit the radii of each data point of the undulation segment in each window to an evenly-spaced time series at the same specified sampling rate, and finally combine the results from each window into the complete resampled series.

Figure 5.2–Figure 5.5 show a groove undulation segment from the original time series before and after being processed by the resampling methods that were investigated. The resampling methods considered here include global linear fitting, global polynomial fitting, and window-based polynomial fitting. The signal was from a 1kHz sine wave of our test signal.

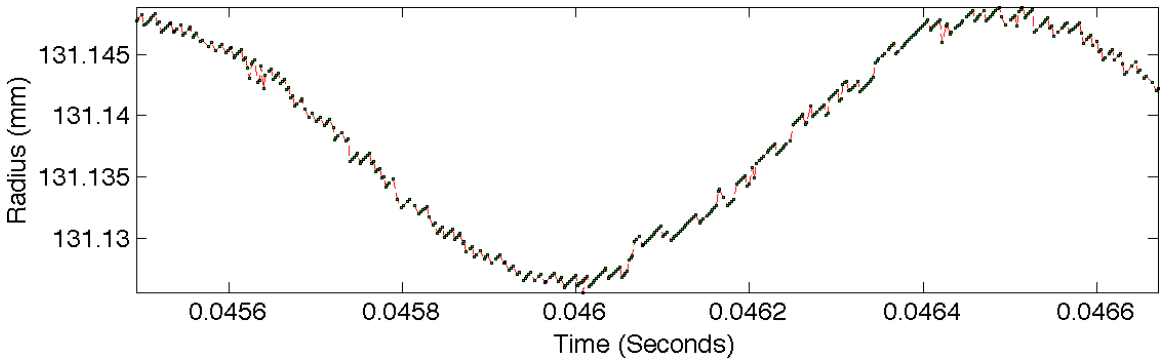


Figure 5.2: A segment of the raw groove undulation time series of a 1kHz sine signal (about one period)

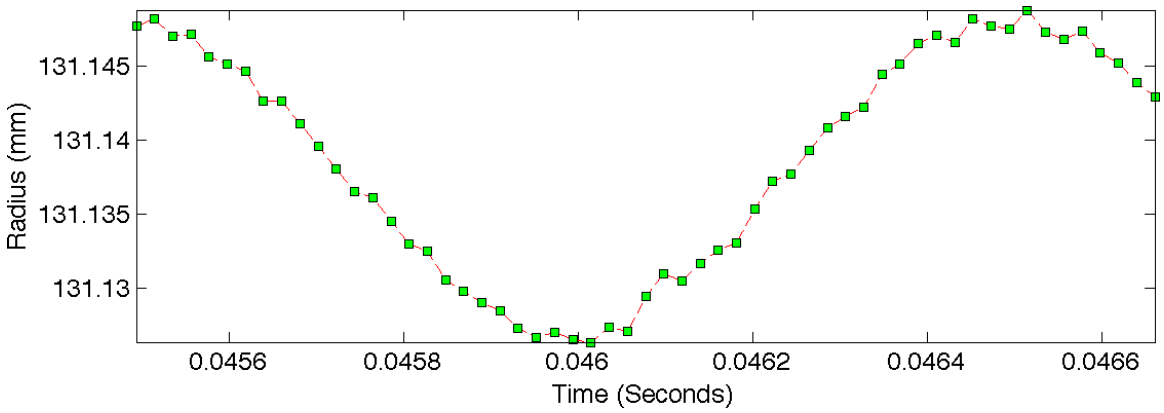


Figure 5.3: The resampled time series from the signal in Figure 5.2, using global linear fitting

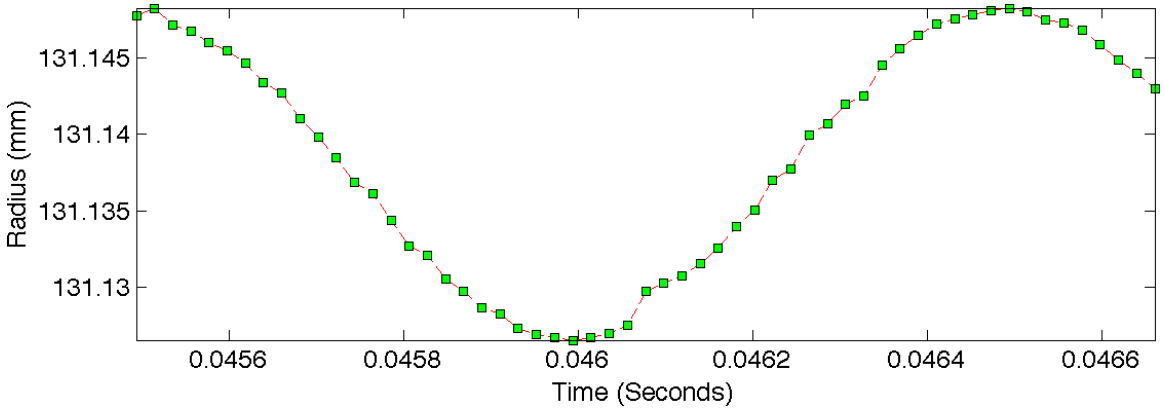


Figure 5.4: The resampled time series from the signal in Figure 5.2, using moving-average smoothing and global linear fitting (with window size 0.08ms)

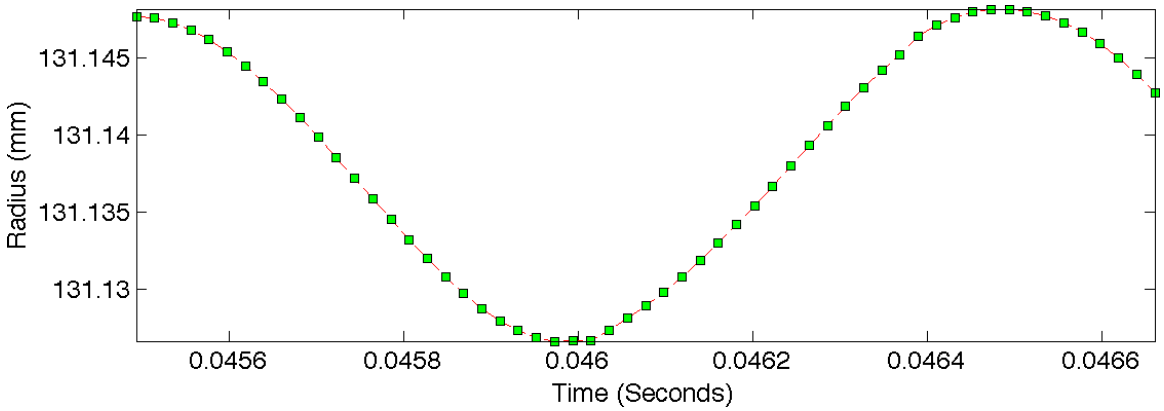


Figure 5.5: The resampled time series from the signal in Figure 5.2, using window-based polynomial fitting, with window size 0.5ms and 4th-order polynomial

The final adopted method is the window-based polynomial fitting. According to our experiments with the 1kHz sine signal track from the test record, this method provided the smoothest curve among all the methods tested. To further improve the

resampling quality, we adopted the oversampling and decimation technique (Hauser 1990), which uses an out sampling rate higher than the output during the polynomial fitting, and then downsamples (decimates) the oversampled signal at the lower output sampling rate. Such a technique is widely used for improving the signal-to-noise ratio of an analog-to-digital conversion system, which is discussed in detail in Section 6.3.2. A 10th-order digital Butterworth filter is used as the anti-aliasing low-pass filter.

The detailed evaluation of the audio quality resulting from the investigated resampling methods is presented in Chapter 6.

5.4 The Differentiation

5.4.1 The Challenges of Numerical Differentiation

After the resampling procedure, the groove undulations are now a time series at a fixed sampling rate. The time series then needs to be differentiated to derive the audio signal because the audio is encoded as the radial velocities of the cutting stylus along the spiralling groove. This is done by using numerical differentiation. Because the groove undulation time series is noisy even after the resampling with smoothing techniques, the input data to be handled by the numerical differentiation still contain some noise. Our differentiation step borrows the techniques from the field of numerical differentiation of non-exact (or observational) data (Anderssen et al. 1998), which is defined below. According to Anderssen et al. (1998), assume that we have an observational, evenly-sampled data series:

$$y := \left\{ y_j = f(t_j) + \varepsilon_j;\ t_j = jT, T = \frac{1}{f_s}, j = 0,1,\cdots,n \right\}$$

where $f(t)$ represents the underlying but unknown model of the observed data series, ε_j the observational errors (additive Gaussian noise), f_s the data sampling rate and T the sampling period. The goal of the numerical differentiation for such non-exact data is to recover from the data series an estimate of its derivatives:

$$f^{(p)}(t) = \frac{d^p f(t)}{dt^p}, p = 1,2,\cdots,n$$

In our OAR application, only the first derivative ($p = 1$) is needed.

In practice, the first derivative of a time series at time sample t_j can be estimated by the backward difference as below:

$$f^{(1)}(t_j) \approx \frac{\Delta y}{\Delta t} = \frac{y(t_j) - y(t_j - T)}{T} \qquad \text{Equation 5.7}$$

It can also be represented by the central difference scheme:

$$f^{(1)}_{j,m}(t_j) = \frac{y(t_{j+m}) - y(t_{j-m})}{2mT} \qquad \text{Equation 5.8}$$

where $m = 1,2,\cdots,M, M < n$, $j = m, m+1, \cdots, n-m$. The central difference scheme has the advantage of higher order of convergence and thus is used more often than the backward difference scheme (Anderssen et al. 1998).

It is well known that numerical differentiation is an ill-posed problem; that is, small measurement errors in the observed data may result in large errors in the derivatives (Diop et al. 2003). Figure 5.6 and Figure 5.7 show a segment of the resampled groove undulation time series and the result of numerically differentiating the time series with the backward differentiation. Because the time series correspond to our sine wave test signal, the derivatives of the signal should be an equally smooth sine signal; the differentiation result by using the backward differentiation, however, shows a noisy sinusoidal signal. To reduce the noise in the differentiation results, regularization techniques are widely used to reduce the differentiation errors by smoothing the raw signal (Diop et al. 2003). Next, we discuss several standard regularization approaches and our choice of such a regularization-based differentiator.

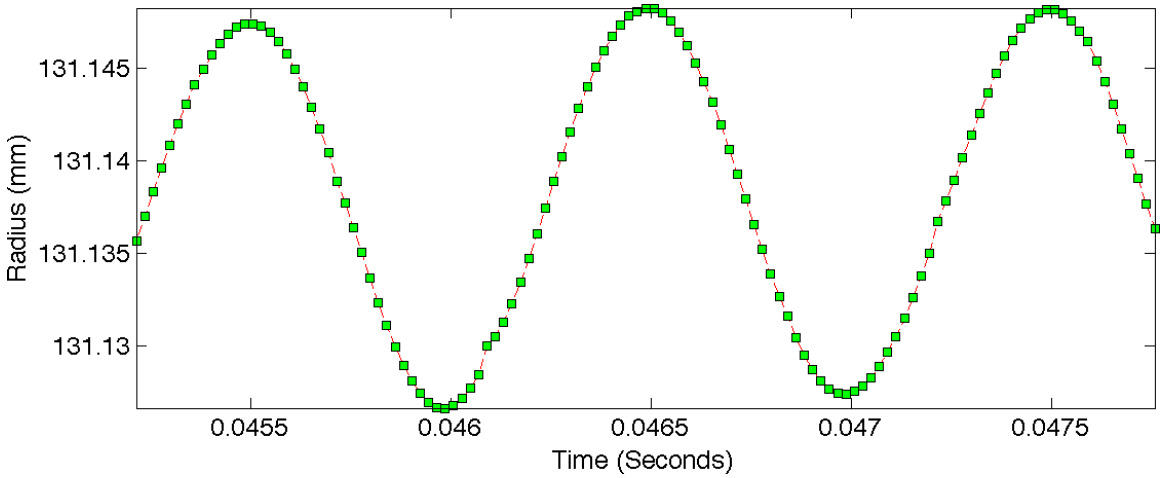


Figure 5.6: A segment of the resampled groove undulation time series at 48kHz sampling rate

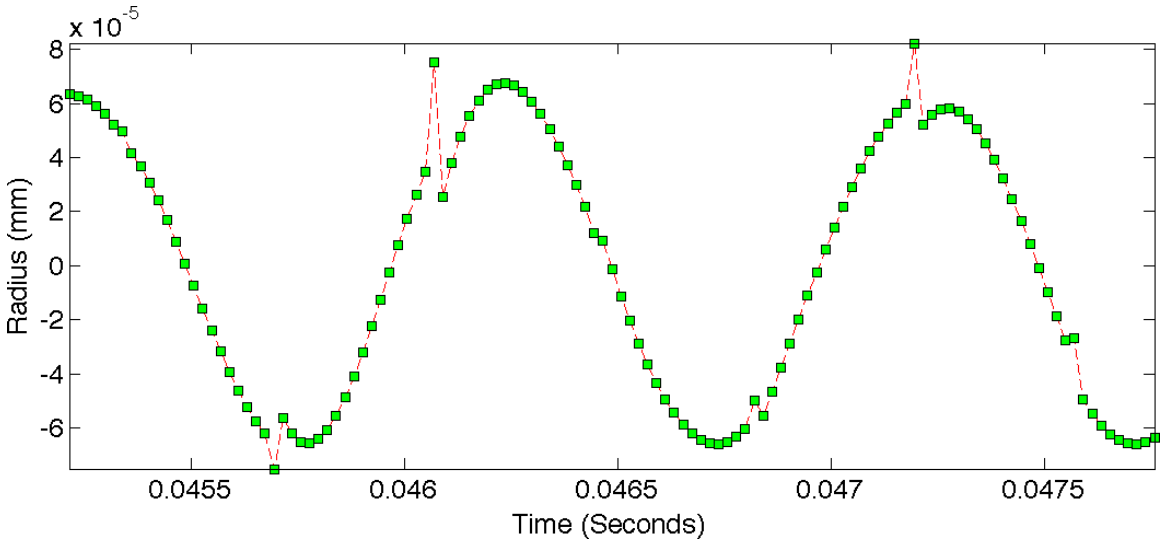


Figure 5.7: The result of differentiating the time series in Figure 5.6 by using backward differentiation

5.4.2 The Choice of the Differentiation Method

Diop et al. (2003) presents a good review of regularization-based numerical differentiation techniques. These techniques strive to achieve a trade-off between two conflicting goals: maximizing the accuracy of the differentiation approximation while smoothing out the noise in the data. Diop et al. reviewed three important regularization-based differentiation techniques, namely the averaged finite difference scheme (Anderssen et al. 1998), the Savitzky-Golay differentiation (Savitzky and Golay 1964), and differentiation based on the wavelet transform (Donoho and Johnstone 1994). These

techniques are briefly introduced below. Readers can refer to Diop et al. (2003) for detailed discussions on these techniques.

For each data point in the input time series, the averaged finite difference scheme specifies a window centered at the time point of interest, computes the derivative for each time point in the window, and takes the average of these results as the derivative of the time point of interest. This method tries to smooth the derivatives of the signals directly. Note that this scheme is usually based on the central difference scheme (Equation 5.8). The control parameters of this approach are the window size and the central difference parameter *m*. The effects of these parameters on the resulting audio are discussed in Chapter 6.

The Savitzky-Golay differentiation divides the signal of interest into windows, fits a polynomial to each window by using the least-squares convergence criteria, and then computes the analytical differentiation of the approximate polynomial. This method aims to smooth the input signal before differentiation. The control parameters of this method include the window size, the order of polynomials used, and an additional manually-specified weighting coefficient array that can be applied to the window. The effects of these parameters on the resulting audio are discussed in Chapter 6.

To introduce the third approach, the wavelet-transform-based differentiation, the wavelet transform is briefly introduced. Readers can refer to Mallat (2009) and Persival and Walden (2000) for the detail of the wavelet transform. The wavelet transform is a multi-resolution signal analysis method, which provides a sparse representation of finite-energy signals that may contain transients and singularities (Mallat 2009, 3). With the wavelet transform, a time series can be represented by the linear combination of its wavelet inner-product coefficients $\langle f, \psi_{\tau,s} \rangle$ and the wavelet basis $\psi_{\tau,s}(t)$ on the square-integrable real-value space $L^2(R)$ (Mallat 2009, 3):

$$f(t) = \sum_{l} \langle f, \psi_{\tau,s} \rangle \psi_{\tau,s}(t) \qquad \text{Equation 5.9}$$

The wavelet coefficients are derived by:

$$\langle f, \psi_{\tau,s} \rangle = \int_{-\infty}^{+\infty} f(t) \psi_{\tau,s}(t) dt \qquad \text{Equation 5.10}$$

The general form of the wavelet basis is shown below (Persival and Walden 2000, 10):

$$\psi_{\tau,s}(t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-\tau}{s}\right) \qquad \text{Equation 5.11}$$

where τ and s, $\tau \in R$, $s > 0$, are the parameters of the translation and dilation operations defined for the specific type of mother wavelet $\psi(t)$, respectively. The mother wavelet satisfies:

$$\int_{-\infty}^{+\infty} \psi(t)dt = 0 \qquad \text{Equation 5.12}$$

With this representation, multi-resolution time-frequency analysis can be performed on the signal by controlling the translation and the dilation operations.

With the wavelet transform, the problem of differentiating the original signal becomes one of computing the derivatives of its wavelet-based sparse representation. According to Equation 5.10 and Equation 5.11, large wavelet coefficients generally correspond to the sharp transient portions of the signal (i.e., the high-frequency components) and small wavelet coefficients the regular portions (i.e., the low-frequency components) (Mallat 2009, 3). Therefore, by down-scaling the computed wavelet coefficients, in which only the irregular transients present large coefficients, the irregularity of the differentiation result signal can be smoothed out.

The differentiation results of the above methods on the time series in Figure 5.6 are shown in Figure 5.8–Figure 5.10. These approaches are all able to provide smoother results than by directly using the backward differentiation. In practice, it is found that wavelet-transform-based differentiation produced the best audio quality among the three methods, and it is adopted as the differentiator in our audio reproduction workflow. The detailed evaluation of these methods in terms of the achieved audio quality is discussed in Chapter 6.

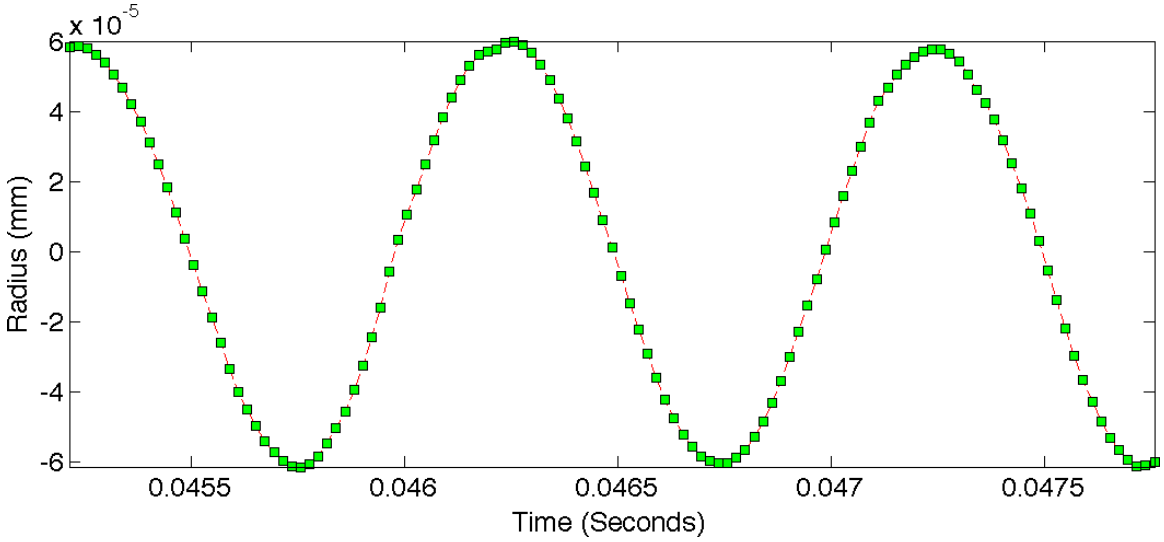


Figure 5.8: The result of differentiating the time series in Figure 5.6 by using the average finite difference scheme, with a window size of 11 samples and the central difference parameter $m = 1$

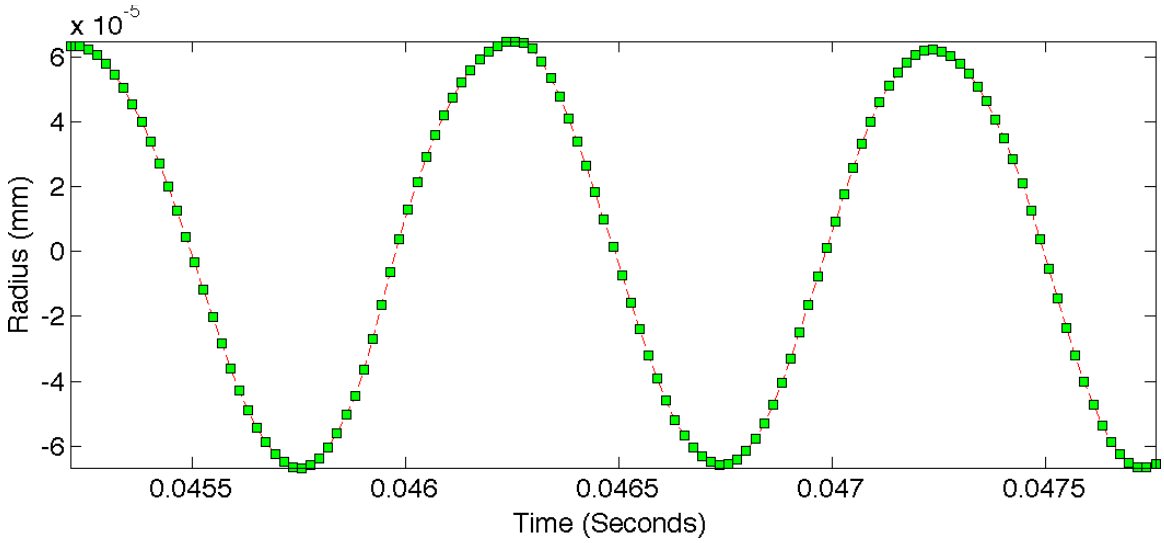


Figure 5.9: The result of differentiating the time series in Figure 5.6 by using the Savitzky-Golay differentiation filter with a window size of 21 samples, 4th-order polynomials, and no additional weighting

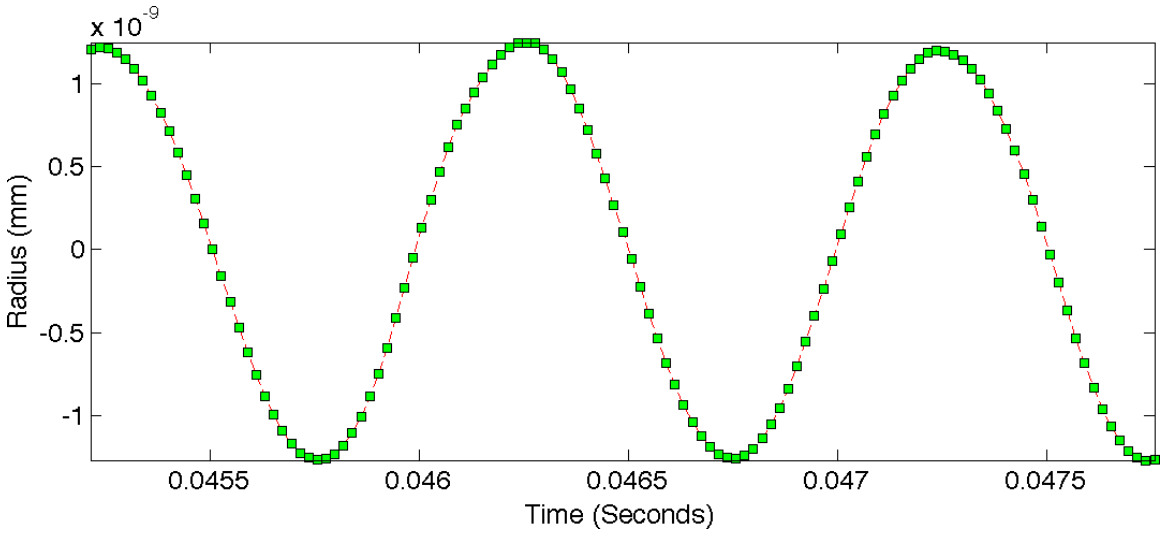


Figure 5.10: The result of differentiating the time series in Figure 5.6 by using the wavelet-transform based differentiation with the Haar wavelets (Perseval and Walden 2000, 57)

5.5 RIAA De-Emphasis Equalization

The result of the differentiation on the resampled time series is the audio signal pre-processed with the RIAA emphasis equalization, which was discussed in Section 2.3.2.6. The audio signal then needs to be filtered with the RIAA de-emphasis equalization to restore the original frequency response of the recorded audio. This de-emphasis filtering procedure has been well documented in Stozer (2006, Chapter 7) and Tian (2006, Chapter 9). We briefly present the procedure below.

The original RIAA emphasis and de-emphasis equalization filters are analog filters with the following transfer functions, H_{pre} and H_{post}, respectively (Tian 2006, 55):

$$H_{pre} = \frac{1+\tau_1 s}{(1+\tau_2 s)(1+\tau_3 s)} \qquad \text{Equation 5.13}$$

$$H_{post} = \frac{(1+\tau_1 s)(1+\tau_2 s)}{(1+\tau_3 s)(1+\tau_4 s)} \qquad \text{Equation 5.14}$$

where

$$s = j\omega$$

$$\tau_i = \omega_i^{-1},\ i = 1,\cdots,4$$

$$\omega_i = 2\pi f_i,\ i = 1,\cdots,4$$

$f_i,\ i = 1,\cdots,4$, are the turnover frequencies of the RIAA equalization curves. See the table below for their values:

	f_1	f_2	f_3	f_4
Frequency (Hz)	500.5	50.05	2122.06	50000

Table 5.2: The turnover frequencies of the RIAA equalization curves: f_4 is added as the upper limit of the frequency response (Tian 2006, 55)

Then the analog filter is converted to a digital Infinite-Impulse-Response (IIR) filter by using the bilinear transform to warp the continuous frequencies in the *s*-plane to the discrete frequencies in the *z*-plane (Diniz et al. 2010, 372):

$$z = e^{j\omega T} = \frac{e^{j\omega T/2}}{e^{-j\omega T/2}} \approx \frac{1 + sT/2}{1 - sT/2} \qquad \text{Equation 5.15}$$

$$\omega_z = \frac{2}{T}\arctan\left(\omega\frac{T}{2}\right)$$

where T is the sampling period, and ω and ω_z are the continuous and discrete frequencies, respectively.

Note that according to Stotzer (2006, 138), the warping from continuous frequencies to the discrete *z*-plane frequencies usually introduces high-frequency distortions above f_3 compared to the analog version (Figure 5.11); an additional turnover f_4 is thus used to avoid this problem. The detailed derivation of the digital de-emphasis filter coefficients can be found in Stozer (2006, 139) and Tian (2006, 55–6).

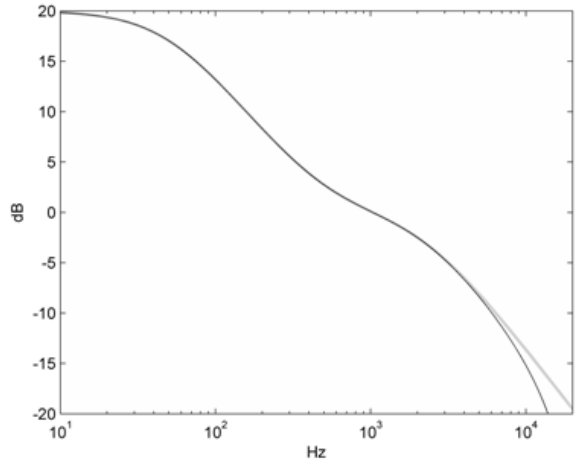


Figure 5.11: The analog (black) and digital (grey) RIAA de-emphasis filter responses (image from Stotzer 2006, 138)

The results of this step on the differentiated inner and outer undulation time series of are the left and right channel signals of the stereo audio encoded in the LP, respectively.

5.6 Removing the Pitch Fluctuation

Figure 5.12 shows the immediate result of the optically reproduced audio signal, which corresponds to a stereo channel encoded in the groove of about two revolutions long. It can be seen that the signal appears to have been shifted by a low-frequency sinusoid, which causes the audible pitch fluctuation similar to the one discussed in Section 2.3.3. This low-frequency component can be seen in the frequency spectrum of the output audio signal (Figure 5.13). This defect thus needs to be fixed.

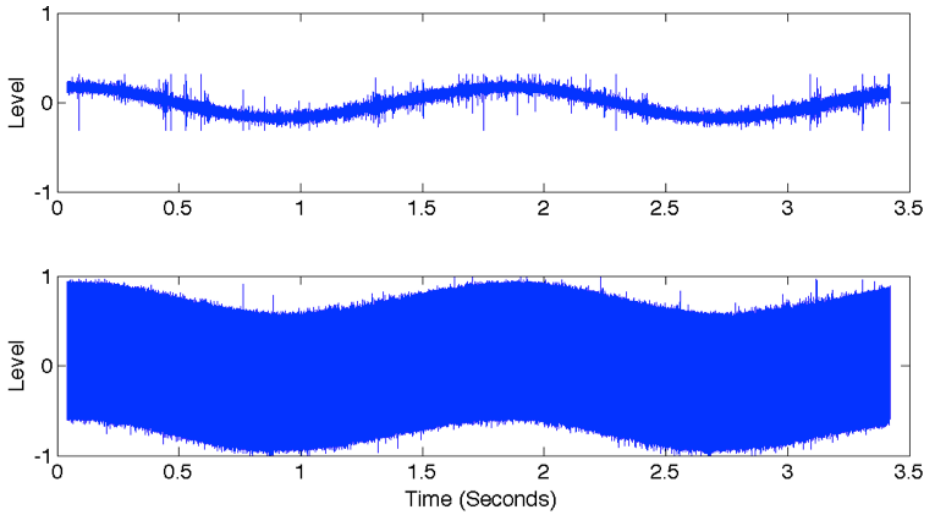


Figure 5.12: The waveform of the output audio that is affected by the pitch fluctuation defect

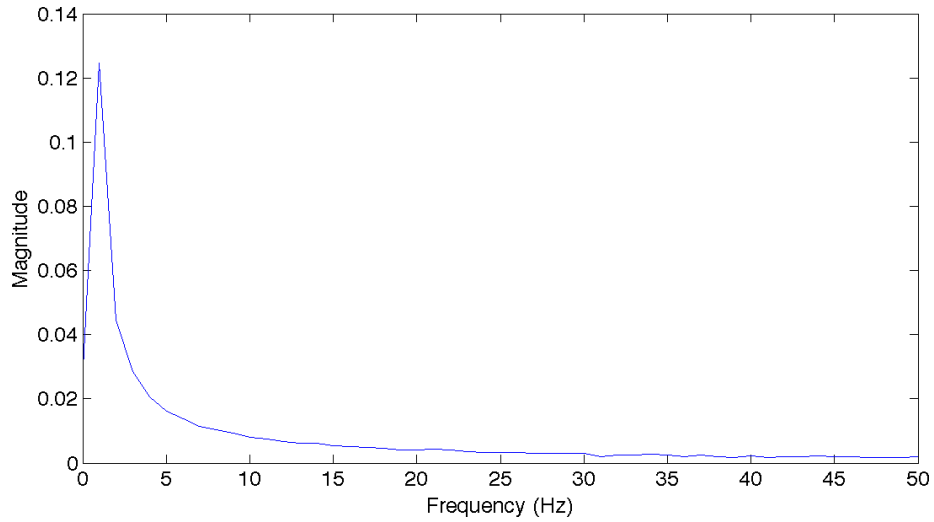


Figure 5.13: The low-frequency peak in the output audio signal due to the pitch fluctuation problem shown on the partial frequency spectrum of the signal

5.6.1 Related Research on Fixing the Pitch Fluctuation Problem

This pitch fluctuation problem was also reported by Fadeyev and Haber (2003) and Stotzer (2006, 110) in their OAR (optical audio reproduction) approaches, respectively. The causes of the pitch fluctuations observed in Fadeyev and Haber (2003) and Stotzer (2006, 110) were both reported to be the errors in estimating the disc center positions. In both of their approaches, the distances from the extracted groove ridge-valley edges to the estimated disc center were interpreted as the raw groove undulation time series. They

provided different solutions to correcting the disc center estimation to fix the pitch fluctuation problems. Their solutions are reviewed in this section.

Without discussing the actual source of the disc center estimation error, Fadeyev and Haber (2003) proposed a time-domain method to improve the accuracy in estimating the disc center by fitting the undulation time series to a function that contains an additive periodic perturbation term:

$$r = a + b \cdot \theta + k \cdot \sin^2(\theta + \theta_0) \qquad \text{Equation 5.16}$$

where r is the radius of the data point of the raw groove undulation time series at the azimuth θ, a, b, k, θ_0 are the parameters of the curve.

In Fadeyev and Haber (2003), the reasons for choosing the particular form of the target function were not explained and no evaluation of the fitting result was reported. However, it can be conjectured that this curve equation is given based on the assumptions about the mono (78rpm) disc recordings that

1) The disc recording groove undulation that encodes a pure silence signal (78rpm records) is in the form of an Archimedean spiral (Equation 5.17).
2) The general audio signal encoded in the disc groove undulations can be represented by an additive periodic perturbation on top of the pure silence signal.

$$r = a + b \cdot \theta \qquad \text{Equation 5.17}$$

Stotzer (2006, 110–4) treated the pitch fluctuation as a type of resampling error when converting the analog record image (i.e., the film) to the digital one by using rotary scanning. During the scanning, as the stage rotates, the CCD (Charge-Coupled Detector) camera mounted on the microscope resamples the film at a constant shutter rate. If the disc record center accurately matched the center of the rotary stage, then the constant sampling period of the camera would correspond to the constant spatial (angular) sampling interval. In practice, however, according to Stotzer (2006, 110), a shift often exists between the two centers. The angular sampling interval then becomes uneven along the groove and a time warp is introduced into the output audio signal. Stotzer took a frequency-domain approach to fix the resampling intervals with the following procedure:

1) Assuming that the pitch fluctuation corresponds to a low frequency component, of which the period is close to one groove revolution, obtain the spectrum of the

undulation time series for one revolution with the Discrete Fourier Transform (DFT). The frequency resolution (i.e., the size of the frequency bin) of the DFT is $\frac{f_s}{2N}$, where f_s is the sampling rate of the output audio and N the number of data points in the groove undulation time series of one revolution.

2) Find the frequency component in the spectrum that accounts for the pitch fluctuation, denoted as f_L. This is done by defining an objective function, which represents the low-frequency component corresponding to the pitch fluctuation in the output audio, using the coefficients of the first sinusoidal component of the DFT spectrum. Then, search the spectrum to find the frequency component that minimizes the objective function.
3) Compute the correct angular sampling intervals at the specified output audio sampling rate when resampling along the groove undulation time series, by using f_L. This is done by fitting the groove undulation time series to a dynamic sampling interval function defined with a f_L-base term to compensate for the time warp due to center shift.

According to Stotzer (2006, 114), this method performs reasonably well when the groove of interest is the innermost groove. On the other hand, he suggested that the method could become inaccurate when the pitch fluctuation is due to causes other than the record center estimation errors. No analysis of these possible alternative cases was given.

5.6.2 Our Solution for Removing the Pitch Fluctuations

5.6.2.1 The Monte-Carlo Optimization Framework

Similar to Stozer (2006) and Fadeyev and Haber (2003), we consider the error in record center estimation as the main cause of the pitch fluctuation problem. Although other factors such as the record surface warping may contribute to the problem, they are not considered in our approach. We propose a Monte-Carlo optimization framework (Rubinstein and Kroese 2008, 201) to iteratively correct the disc record center estimation error. This framework assumes that the true disc center is located in the neighbouring area of the estimated center. Then a grid of candidate locations of the true center is created and each of them is tested by updating the polar coordinates of the extracted groove

undulation time series. The groove undulation time series resulting from each candidate center is then tested against a selected objective function. The candidate center that generates the optimal value for the objective function is finally determined to be the best-estimated disc center of the current iteration, and the coordinates of the groove undulation time series that were derived from the polar coordinates of the pixels in the images are finalized according to this center location. In the next iteration, a new grid of candidate locations is then created around the best-estimated disc center. This procedure is repeated until the distance between the most recent estimate and the previous estimate is below a pre-defined threshold or the pre-defined maximum number of iterations is reached. This iterative procedure of the framework is described below.

First, in the initial iteration, assume the true record center O is located in the surrounding area $A^{(0)}$ centered at the initial estimated record center, $\hat{O}^{(0)} = (\hat{X}_O^{(0)}, \hat{Y}_O^{(0)})$:

$$O \in A^{(0)} = \left\{ (x,y) \middle| \left| x - \hat{X}_O^{(0)} \right| < R_x^{(1)}, \left| y - \hat{Y}_O^{(0)} \right| < R_y^{(1)}, R_x^{(1)} > 0, R_y^{(1)} > 0 \right\}$$

The 2D translation between O and $\hat{O}^{(0)}$ is denoted as $(\Delta X, \Delta Y)$, $|\Delta X| < R_x^{(1)}$, $|\Delta Y| < R_Y^{(1)}$ (Figure 5.14).

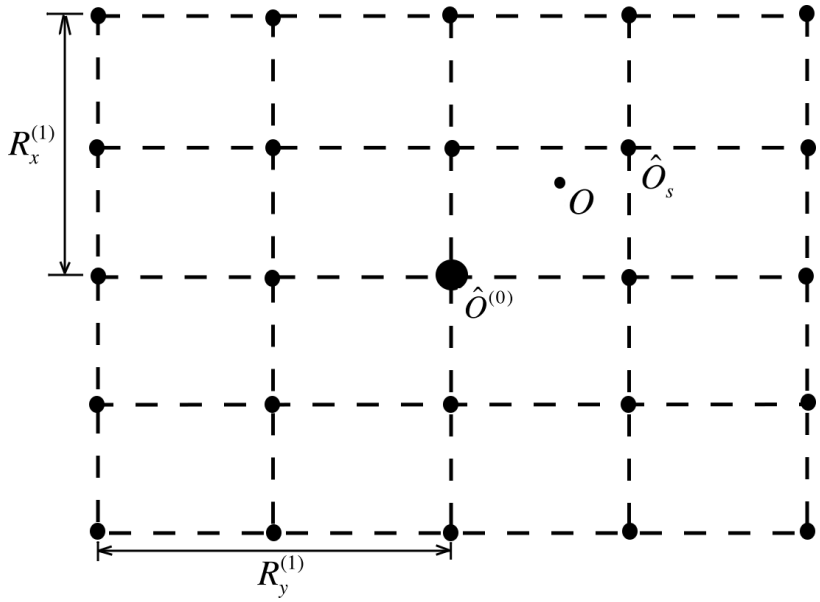


Figure 5.14: A 5x5 node grid of our Monte-Carlo optimization framework for the first iteration

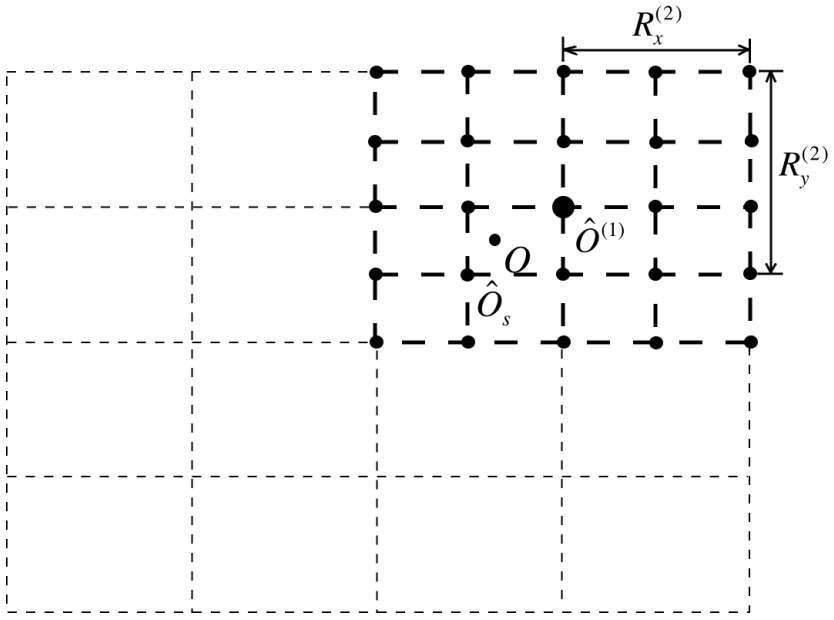


Figure 5.15: The node grid of our Monte-Carlo optimization framework for the second iteration: The initial estimated record center $\hat{O}^{(1)}$ was the optimal node from the previous iteration; the dashed-line region is the node grid from the previous iteration

Second, the area $A^{(1)}$ is then evenly sampled with a $(2N+1)\times(2M+1)$ node grid: $O_s \in S^{(1)} = \left\{(x,y)\middle| x = \hat{X}_o^{(1)} + n\cdot\frac{R_x^{(1)}}{N}, y = \hat{Y}_o^{(1)} + m\cdot\frac{R_y^{(1)}}{M}, n\in[-N,N], m\in[-M,M]\right\}$. Each node in $S^{(1)}$ is a candidate position that may be closer to the true record center than $\hat{O}^{(0)}$.

Next, we search $S^{(1)}$ for the node $\hat{O}_s$ that optimizes a specified objective function, $F(O_s)$. The objective function defines the criteria that minimize the pitch fluctuation in the output audio. $\hat{O}_s$ is then used as the estimated record center position of the current iteration $\hat{O}^{(i)} = (\hat{X}_O^{(i)}, \hat{Y}_O^{(i)})$, $i = 1, 2, \cdots, I_{\max}$, which is supposed to be closer to the true record center than the initial estimation. The iterative optimization terminates when one of the following conditions is satisfied:

$$\left|\hat{O}^{(i)}\hat{O}^{(i-1)}\right| < E_O \qquad \text{Equation 5.18}$$

$$i = I_{max} \qquad \text{Equation 5.19}$$

where $i \in Z$ is the current iteration number, $I_{max} \in Z$ the maximum number of iterations, and E_O the specified minimum translation between the current estimated record center

and the previous one. The objective function is then iteratively optimized until the termination conditions are met. As illustrated in Figure 5.15, for the current iteration i, the new search area for estimating the next iteration is defined as:

$$A^{(i+1)} = \left\{ (x,y) \middle| \left| x - \hat{X}_O^{(i)} \right| < R_x^{(i+1)}, \left| y - \hat{Y}_O^{(i)} \right| < R_y^{(i+1)}, R_x^{(i+1)} > 0, R_y^{(i+1)} > 0 \right\}$$

where $R_x^{(i+1)} = \frac{R_x^{(i)}}{N}$, $R_y^{(i+1)} = \frac{R_y^{(i)}}{M}$.

$A^{(i+1)}$ is then again sampled with a $(2N+1) \times (2M+1)$ node grid, resulting in the new candidate estimated record center point set:

$$S^{(i+1)} = \left\{ (x,y) \middle| x = \hat{X}_O^{(i)} + n \cdot \frac{R_x^{(i+1)}}{N}, y = \hat{Y}_O^{(i)} + m \cdot \frac{R_y^{(i+1)}}{M}, n \in [-N, N], m \in [-M, M] \right\}$$

When the optimization terminates after Iteration i, $\hat{O}^{(i)}$ is then the final corrected record center. The polar coordinates of the groove undulation time series are then re-computed based on the correct record center.

5.6.2.2 The Objective Functions

We experimented with several time-domain and frequency-domain objective functions with our Monte-Carlo optimization framework, which are summarized below.

The first method is performed in the frequency domain. Similar to Stotzer's method (see Section 5.6.1), it is assumed that the pitch fluctuation leads to certain emphasized low-frequency components. Because the period of the pitch fluctuation is usually close to one revolution, the range of the low-frequency components that includes the offensive components can be roughly computed as the inverse of the duration of one revolution. For instance, for a stereo LP, the offensive frequency component is around 0.56Hz. The objective function is thus specified to minimize the low-frequency component power around such a frequency. In practice, we iteratively minimize the energy for the components within [0, 20Hz].

Two time-domain objective functions are proposed based on the simple geometrical property of the record groove. The groove spirals from the outer revolution to the inner ones. The radii of the raw groove undulation thus should display a global linearly decreasing envelope with little magnitude fluctuation around the line. However, the

center estimation errors introduce a modulation to the envelope, and the magnitude of the modulation is usually much larger than the groove undulation magnitudes (Figure 5.12). Two optimization objectives are thus experimented with, denoted TDO1 and TDO2, respectively:

1) TDO1: Simply minimize the average revolution-wise magnitude of the raw groove undulation time series.
2) TDO2: For each groove revolution, that is, the groove undulation time series segment within $\left[2i\pi, 2(i+1)\pi\right], i = 0, \cdots, N_r$, where N_r is the total number of revolutions that the groove undulation time series of interest spans, minimize the deviation around the globally decreasing linear envelope.

We also experimented with the curve-fitting-based approach with a simple Archimedean spiral (Equation 5.17) and the curve that Fadeyev and Haber (2003) proposed (Equation 5.16), which adds a periodic perturbation to the Archimedean spiral.

According to our comparison, the TDO2 achieved the best results for the record center correction and was adopted as the solution to remove the pitch fluctuation defects. The pitch fluctuations became inaudible in the output audio as the result of using our record center correction technique. The detailed evaluation of the aforementioned objective functions is presented in Chapter 6.

The resulting audio signal and the low-frequency components of its spectrum from the corrected disc center estimate are shown in Figure 5.16 and Figure 5.17, respectively.

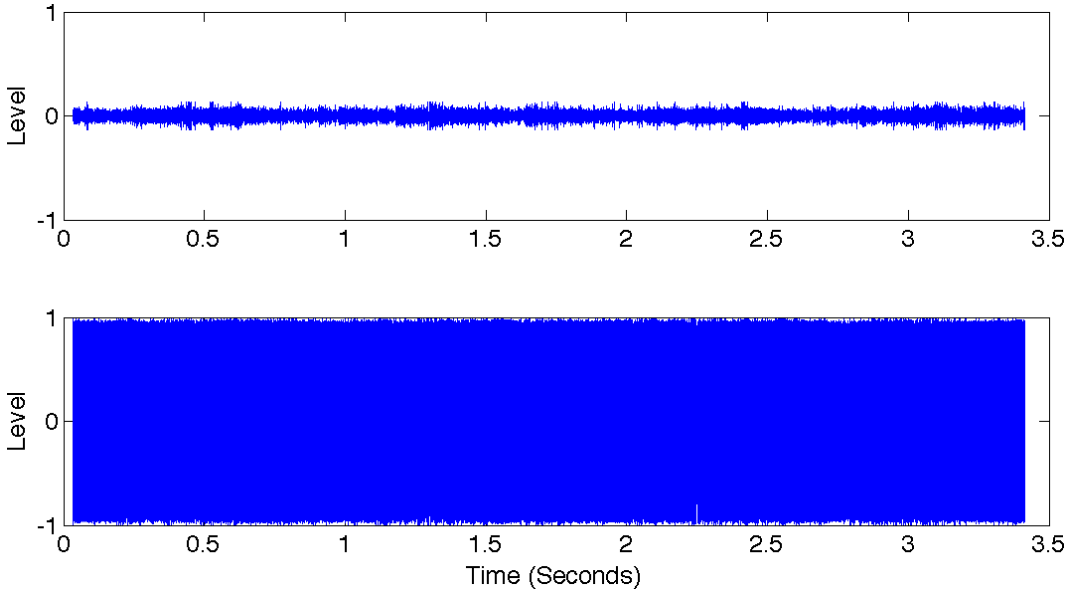


Figure 5.16: waveform of the output audio after removing the pitch fluctuation

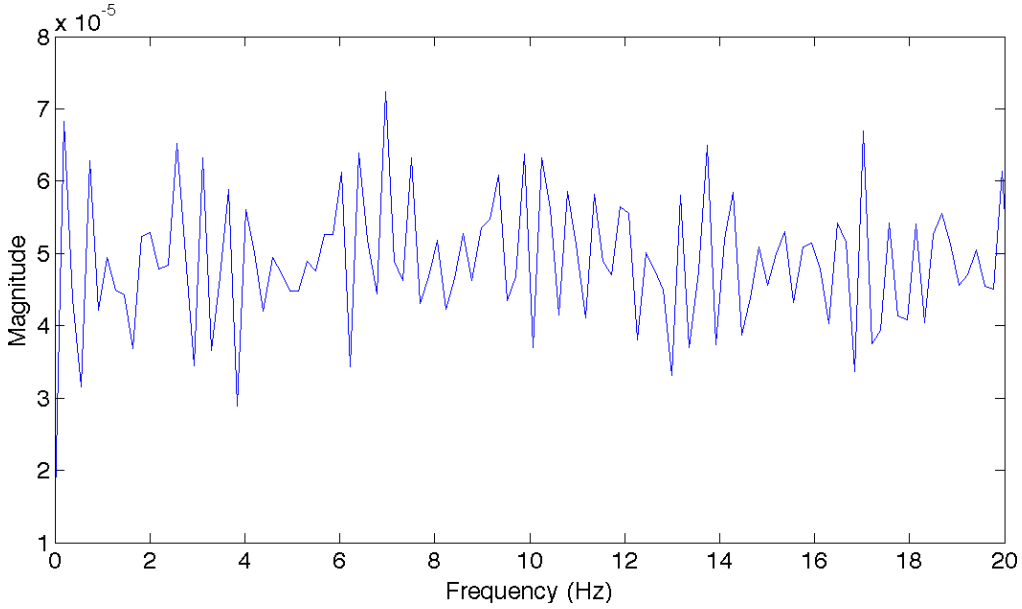


Figure 5.17: The lower-band of the frequency spectrum in Figure 6.5 that includes the frequency range related to the pitch fluctuation defect: The magnitudes of these frequency components are much lower than the corresponding ones in Figure 5.13

Note that the record center correction for removing pitch fluctuation in the audio is usually used as the initial step of our practical audio reproduction workflow for the extracted raw groove undulation time series.

5.7 Summary

This chapter presented the audio reproduction workflow for extracting the groove undulation time series from the record surface profile images. The basic workflow first fixes the pitch fluctuation problem by correcting the estimated disc center and the image polar coordinates based on the center. The raw time series, which is equivalent to unevenly sampled discrete data points from the analog signals, is then resampled to evenly-spaced data points at a specified output audio sampling rate. The global and window-based curve fitting methods were examined and the window-based polynomial fitting was adopted as the resampling method. Then, the resampled undulation time series were differentiated to produce the stereo channel audio signals. Three numerical differentiation methods were evaluated, and the wavelet-transform-based method was chosen as our differentiation method. Finally, the RIAA de-emphasis equalization filtering was applied to restore the original audio signal frequency content, which was modified by the RIAA emphasis equalization.

The pitch fluctuation is a defect that usually exists in the output audio when using the basic audio reproduction workflow. This defect is mainly due to the errors in estimating the record center, which is key to deriving the polar coordinates of the groove undulation time series. The record center correction methods in the exiting OAR approaches were reviewed and a Monte-Carlo optimization framework was proposed to remove the pitch fluctuation by correcting the record center estimation errors; various optimization objective functions and the results of the pitch fluctuation removal based on optimizing these functions were discussed. This process was thus integrated with the basic audio reproduction workflow as the initial step. The result of this correction will be shown in the next chapter as well as the result of applying the entire audio reproduction workflow.

Chapter 6 Evaluation

6.1 About the Chapter

Having presented our optical audio reproduction (OAR) workflow in the previous chapters, in this chapter we evaluate our workflow and the output audio quality based on the experiments and analysis with our stereo long-play (LP) test records (Section 6.2). First, the theoretically-achievable audio quality of our OAR system is analyzed (Section 6.3); our OAR workflow is then evaluated in terms of the output audio quality to show the best practice among the available choices when performing certain steps (Section 6.4); next we present the practically-achievable audio quality of the test sine signals that we use via the best practice of our OAR under several resolutions of the record surface profile images and the results are compared with those reproduced by the turntable (Section 6.5); finally we qualitatively compare a musical audio signal reproduced optically with that from the turntable and the CD version (Section 6.6).

6.2 The Test Records and Signals

We use an audio system test LP (Williams and Wolifson) for quantitative audio quality evaluation. A music LP recording (Bernstein CBS IM 77215) and its corresponding CD (compact disc) version (Bernstein SMK 60208–AC) are used for qualitative comparison of the optically-reproduced audio with the other two versions.

In the experiments for the quantitative evaluation, we mainly used a stereo signal in the test LP, of which the left channel is silence and the right channel is a standard 1kHz sine signal produced at the 7cm/sec peak velocity (i.e., the radial velocity of the cutting stylus at the peak of the groove undulation), which corresponds to the 0dB reference level. The signal acquired is of 3.4-second duration, approximately two groove revolutions. This signal was used for evaluating the theoretical audio dynamic range and our OAR workflow. This stereo signal was also used for testing the relationship between the image resolution and the output audio quality.

The music recording used was an about 13-second excerpt from Mahler's Symphony No.6 (Bernstein CBS IM 77215).

6.3 The Theoretically-Achievable Audio Quality of our OAR System

In this section, we discuss the equivalent audio sampling rates and the dynamic ranges that our OAR system can provide with its various image resolutions (magnifications).

6.3.1 The Equivalent Sampling Rates

In Section 5.3.1 and 5.3.2, we have explained that the raw groove undulation can be seen as time series of uneven sampling rates, which are determined by both the lengths of the groove revolutions that we examine and the lateral image resolution used by the microscope to acquire the record surface profile images. Recall from Chapter 5 that the equivalent sampling rate S_R can be represented as

$$S_R = \frac{L_R \cdot RPM}{60 \cdot d_{pixel}} \qquad \text{Equation 6.1}$$

where L_R is the length of the groove revolution of interest, RPM the rotation speed, and d_{pixel} the lateral image resolution.

The estimated equivalent sampling rates corresponding to the available lateral image resolution options of our image acquisition system are listed in Table 6.1.

Magnification	**2.75X**	**5X**	**10X**	**27.5X**	**50X**	**100X**
Image resolution (µm/pixel)	3.95	2.00	1.01	0.39	0.20	0.10
Minimum average sampling rate (kHz)	53.27	106.12	213.92	539.10	1036.70	2246.10
Maximum average sampling rate (kHz)	128.96	256.91	517.90	1305.10	2509.80	5438.00

Table 6.1: The various average audio sampling rates provided by our acquisition system for the innermost (minimum) and the outermost grooves (maximum) of the stereo disc records

It can be seen from Table 6.1 that the equivalent sampling rates that our image acquisition system can provide are all above the sampling rate of the CD recordings (44.1kHz).

6.3.2 The Equivalent Signal-to-Noise Ratio

In this section, we derive the relationship between achievable audio signal-to-noise ratios (SNRs) and the lateral image resolution of our image acquisition system. We then derive the required lateral image resolution for achieving specified output audio bit depth. Finally, we present the achievable SNRs and bit depths by using the available lateral image resolutions of our image acquisition system, and compare them with the SNRs specified by the NAB audio recording and reproducing standards for disc recordings (National Association of Broadcasters 1964).

Our OAR system is considered as an analog-to-digital converter (ADC) for audio. We therefore first introduce the SNR for an ideal N-bit ADC and then derive the SNR for our OAR system based on the definition of SNR for an ADC.

The SNR of an ADC is defined as:

$$SNR = 10 \cdot \log_{10}\left(\frac{RMS_{signal}}{RMS_{noise}}\right) \qquad \text{Equation 6.2}$$

where RMS_{signal} and RMS_{noise} are the root-mean-square (RMS) signal powers of the input analog signal and the quantization noise over the duration of the signal, respectively. The quantization noise of an ideal N-bit ADC is usually assumed as being a white noise independent of the original signal (Bennett 1948, Kollár 1986), which can be represented as a sawtooth error signal with its peak-to-dip amplitude being the least-significant bit (LSB), q (see Figure 6.1). The error signal can be represented as:

$$e(t) = s \cdot t,\ t \in \left[-q/2,\ q/2\right]$$

Its RMS power is thus:

$$RMS_{noise} = \sqrt{\overline{e^2(t)}} = \sqrt{\frac{s}{q}\int_{-q/2}^{q/2}\left(s \cdot t\right)^2 dt} = \frac{q}{\sqrt{12}} \qquad \text{Equation 6.3}$$

The input analog signal is assumed to be a full-scale (FS) sine wave:

$$v(t) = \frac{q \cdot 2^N}{2} \cdot \sin(2\pi \cdot f \cdot t)$$

where N is the number of bits, and f the frequency of the sine signal. Its RMS power is thus:

$$RMS_{signal} = \frac{q \cdot 2^N}{2\sqrt{2}} \qquad \text{Equation 6.4}$$

From Equation 6.2–Equation 6.4, the SNR of an ideal N-bit ADC can be derived as:

$$SNR = 20 \cdot \log_{10} 2^N + 20 \cdot \log_{10} \sqrt{\frac{3}{2}} \approx 6.02N + 1.76 \qquad \text{Equation 6.5}$$

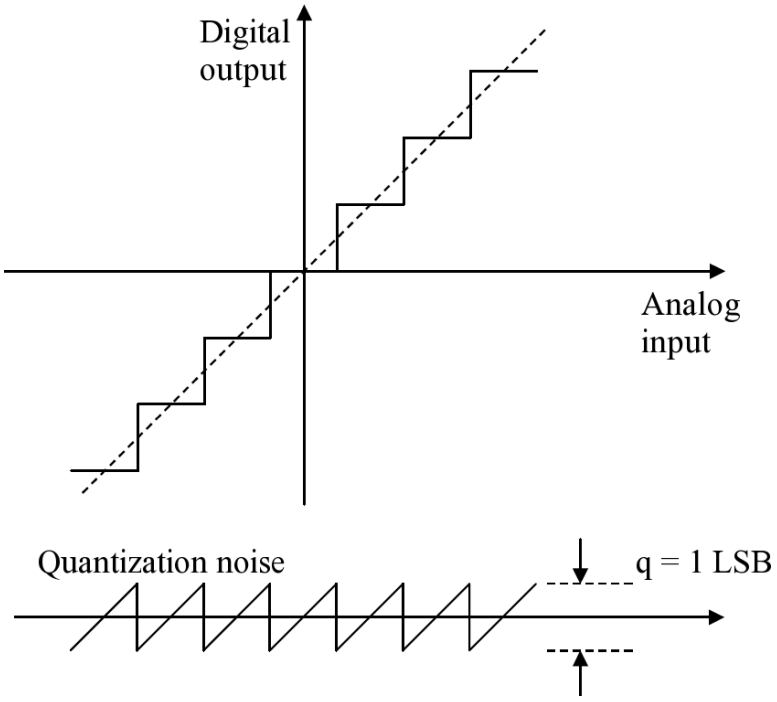


Figure 6.1: The quantization noise of an ADC

The above theory is derived without taking into consideration the oversampling and decimation techniques, which have been widely used for improving the SNR of digital audio systems. Hauser (1990) presents a good review of these techniques. By increasing the sampling rate, more data are used to depict the signal curve of limited value range and the sampling quantization noise becomes spread out to wider frequency range; then the noise contained in the high-frequency band is removed by decimation (i.e., low-pass filtering and downsampling). The quantization noise power is thus reduced in the resulting digital signal. Given the basic SNR in Equation 6.12, and the oversampling and decimation factor D, that is, the ratio between the oversampling

sampling rate and the output sampling rate of the digital signal), the improved SNR (Hauser 1990, 6) is thus

$$SNR' = SNR + 20 \cdot \log_{10} D \qquad \text{Equation 6.6}$$

As listed in Table 6.1, the average equivalent sampling rates of the groove undulation time series are usually much greater than the output audio sampling rates of our OAR workflow. The SNRs can thus be improved by exploiting the oversampling and decimation techniques. Given the equivalent sampling-rate of a specific groove revolution S_R, and the output audio sampling rate S_{out}, D can be represented as

$$D = \frac{S_R}{S_{out}} \qquad \text{Equation 6.7}$$

In our case, S_{out} is the CD sampling-rate (44.1kHz).

We then derive the equivalent audio SNRs of our image acquisition system. In our OAR application, the number of steps in the output digital signal to represent the groove undulation range can be represented by the ratio between the undulation peak-to-dip amplitude and the lateral pixel resolution. Without considering the oversampling, the SNR of the audio signal on a disc record, based on Equation 6.6, can be defined as:

$$SNR = 20 \cdot \log_{10}\left(\frac{2A_{peak}}{d_{pixel}}\right) + 1.76 \qquad \text{Equation 6.8}$$

where A_{peak} is the spatial displacement corresponding to the peak amplitude of the groove undulation of the signal of interest, and d_{pixel} is the lateral image resolution. The derivation of the peak amplitude of our test sine signal is presented below.

The groove undulation for a sine signal of frequency f can be represented as

$$x(t) = A \cdot \sin(2\pi \cdot f \cdot t) \qquad \text{Equation 6.9}$$

Due to the constant velocity groove-cutting scheme (see Section 2.3.2.6), the audio signal is encoded in the stylus velocity. According to Equation 6.9, the stylus velocity is then

$$v(t) = 2\pi \cdot f \cdot A \cdot \sin(2\pi \cdot f \cdot t) \qquad \text{Equation 6.10}$$

The spatial peak amplitude A_{peak} is thus represented by

$$A_{peak} = \frac{v_{peak}}{2\pi \cdot f} \qquad \text{Equation 6.11}$$

Based on Equation 6.8 and Equation 6.11, the SNR of the test sine signal without considering oversampling is therefore:

$$SNR = 20 \cdot \log_{10}\left(\frac{v_{peak}}{\pi \cdot f \cdot d_{pixel}}\right) + 1.76 \qquad \text{Equation 6.12}$$

If we consider the SNR gain from oversampling, from Equation 6.6, Equation 6.7, and Equation 6.12, the SNR of the test sine signal by considering oversampling and decimation is thus:

$$SNR = 20 \cdot \log_{10}\left(\frac{v_{peak} \cdot S_R}{\pi \cdot f \cdot d_{pixel} \cdot S_{out}}\right) + 1.76 \qquad \text{Equation 6.13}$$

By combining Equation 6.1 and Equation 6.13, the minimum required lateral image resolution for achieving a desired N-bit quantization bit depth can be represented as:

$$d_{\min} = \sqrt{\frac{v_{peak} \cdot L_R \cdot RPM}{60\pi \cdot f \cdot S_{out} \cdot 2^N}} \qquad \text{Equation 6.14}$$

According to Equation 6.14, in order to achieve the CD-comparable output audio quality, that is, N = 16 bits, and S_{out} = 44.1kHz output sampling rate, for all groove revolutions, the minimum required lateral image resolution is about 40.3nm/pixel for 1kHz sine signal recorded at 0dB. From the available lateral image resolutions of our image acquisition system (Table 6.1), none of the lateral image resolutions available in our image acquisition system can provide an equivalent 16-bit bit depth under 44.1kHz output sampling rate even with the help of oversampling and decimation.

According to the NAB audio recording and reproducing standards for disc recording and reproducing (National Association of Broadcasters 1964, 5), the disc phonograph recording can reach a signal-to-noise ratio (SNR) of 60dB for 1kHz signal without the help of a pre-amplifier. According to Equation 6.14, in order to achieve the SNR specified in the NAB standard for a 1kHz sine signal over the entire groove of a 12-inch (30cm) stereo LP record, the minimum required lateral image resolution is 0.23μm/pixel for the 1kHz sine signal. As for our image acquisition system, only the 50X and 100X magnification can achieve the aforementioned NAB standard SNR for the 1kHz sine signal along the entire groove of a 12-inch stereo LP record.

Based on Equation 6.12, the theoretical SNRs and bit depths for our 1kHz test sine signal without considering oversampling are listed in Table 6.2.

Magnification	2.75X	5X	10X	27.5X	50X	100X
Lateral image resolution (μm/pixel)	3.95	2.00	1.01	0.39	0.20	0.10
SNR (dB)	16.78	22.68	28.66	36.92	42.83	48.86
Bit depth (bits)	2.5	3.48	4.47	5.84	6.82	7.82

Table 6.2: The theoretical SNRs and bit depths for the 1kHz sine signal with 7cm/sec peak velocity, derived by the available lateral image resolution options of our image acquisition system without oversampling

Based on Equation 6.13 and the data from Table 6.1 and Table 6.2, the theoretical SNRs and bit depths for our 1kHz test sine signals with oversampling as compared to the CD sampling rate are listed in Table 6.3.

Magnification	2.75X	5X	10X	27.5X	50X	100X
Lateral image resolution (μm/pixel)	3.95	2.00	1.01	0.39	0.20	0.10
Average sampling rate range (kHz)	53.27 – 128.96	106.12 – 256.91	213.92 – 517.90	539.10 – 1305.10	1036.70 – 2509.80	2246.10 – 5438.00
Oversampling factor	1.21 – 2.92	2.41 – 5.83	4.85 – 11.74	12.22 – 29.59	23.49 – 56.91	50.93 – 123.31
SNR (dB)	18.42 – 26.10	30.30 – 37.98	42.38 – 50.06	58.66 – 66.33	70.26 – 77.94	83.00 – 90.68
Bit depths (bits)	2.77 – 4.04	4.74 – 6.02	6.75 – 8.02	9.45 – 10.73	11.38 – 12.65	10.66 – 14.77

Table 6.3: The theoretical SNRs and bit depths for the 1kHz sine signal with 7cm/sec peak velocity, derived by the available lateral image resolution options of our image acquisition system with oversampling and decimation

The above analytical results present the ranges of the average equivalent sampling rates and SNRs over the entire groove. The selected test signal for our subsequent experiments, however, belongs to specific groove revolutions. The average equivalent

sampling rates and the SNRs for the groove revolutions corresponding to this particular test signal are thus shown in Table 6.4.

Magnification	2.75X	5X	10X	27.5X	50X	100X
Lateral image resolution (μm/pixel)	3.95	2.00	1.01	0.39	0.20	0.10
Average sampling rate (kHz)	115.14	227.06	452.35	1169.90	2312.00	4624.10
Oversampling factor	2.61	5.15	10.26	26.53	52.43	104.85
SNR (dB)	19.10	36.91	48.89	65.39	77.23	89.27
Bit depth (bits)	3.88	5.84	7.83	10.57	12.54	14.54

Table 6.4: The theoretical SNRs and bit depths for the 1kHz sine signal with 7cm/sec peak velocity at the groove revolution with an average radius of 13.04cm of our image acquisition system with oversampling and decimation

From Table 6.4, for the groove revolutions corresponding to the particular test signal, the 27.5X, 50X, and the 100X magnification of our image acquisition system can provide higher theoretical SNRs than that of the NAB standard.

6.4 The Evaluation of our OAR Workflow

Our complete OAR workflow has been presented in Chapters 3–5. Several strategies or algorithm configurations are available for certain steps of the workflow. In this section, these options are evaluated to justify the final adopted solution for each related step. We first introduce our evaluation protocol (Section 6.4.1); then briefly report the adopted image acquisition parameter configurations; finally, the optional strategies for each step in reproducing audio from the groove undulations are evaluated (Sections 6.4.3–6.4.5).

6.4.1 The Evaluation Protocols

The signal used in OAR workflow is the stereo signal with a silent left channel and a 1kHz sine wave as the right channel (see Section 6.2). The evaluations are performed on the images acquired at a specific magnification of the acquisition system (i.e., of a specific lateral image resolution).

The signal-to-noise ratio (SNR) and the total harmonic distortion (THD) are used as the audio quality measures for quantitatively evaluating most steps of our workflow. The SNR is defined in Equation 6.2. In practice, the RMS power of the right channel of our test stereo signal that contains the 1kHz sine signal was treated as the signal RMS power. As for the noise power, we experimented with two different methods. Taking advantage of the fact that the left channel of our test stereo signal is silence, the first method uses the RMS power of the silence track (the left channel) as the noise RMS power. This method proved problematic because the left channel turned out to contain noise correlated to the right channel signal according to our experiments. This was caused by the inter-channel crosstalk found in our test stereo signal (see Section 6.4.3). The second method was to scan and reproduce the audio for a segment of the silent lead-out groove by using our OAR workflow, and use the RMS power of the signal extracted from the lead-out groove as the noise power, which was adopted in our evaluations.
The THD is defined as:

$$THD = 10 \cdot \log_{10}\left(\frac{P_{harmonics}}{P_{signal}}\right) \qquad \text{Equation 6.15}$$

where P_{signal} and $P_{harmonics}$ are the signal power of the sine signal at its reference frequency and the power of the harmonics of the reference frequency up to the Nyquist frequency. Each frequency component under comparison is with a ±5Hz margin.
It is expected that the higher quality audio should present high SNR and low THD.
The output sampling-rate was chosen to be 44.1kHz.

6.4.2 The Image Acquisition

In Chapter 3, various image acquisition parameters for the Vertical Scanning Interferometry (VSI) were introduced. After various experiments, an optimal VSI parameter configuration is adopted, which are listed in Table 6.6.

VSI parameters	Light intensity	Scan range	Scan speed	Modulation threshold
Adopted values	Ramp from 13–17%	100μm	3X (15μm/sec)	2%

Table 6.5: The adopted VSI parameter configuration

For the grid-based automatic measurement and stitching, the groove area of interest is divided into eight grids. A 20% overlap is used between adjacent cells. According to Equation 3.2 and 3.13, the scan time and storage space required for acquiring the surface profile images for one side of a stereo LP disc record under the available image resolutions are listed in Table 6.6.

Magnification	**2.75X**	**5X**	**10X**	**27.5X**	**50X**	**100X**
Lateral image resolution (μm/pixel)	3.95	2.00	1.01	0.39	0.20	0.10
Number of Images	$1.79x10^4$	$7.13x10^4$	$2.85x10^5$	$1.81x10^6$	$6.62x10^6$	$2.87x10^7$
Scan time (days)	1.38	5.50	21.93	139.60	510.08	2214.50
Storage space (gigabytes)	26.82	106.84	426.28	2714.80	9915.90	$4.30x10^4$

Table 6.6: The scan time (in days) and storage space (in gigabytes) required for acquiring the surface profile images for one side of a stereo LP disc record under the available image resolutions, with the parameter configuration in Table 6.5 and 20% inter-cell overlap for an 8-grid automatic measurement

In the subsequent workflow evaluation, if there is no special indication, the experiments use the images acquired under the 10X magnification, as higher magnifications are impractical. The scan time and storage space information for the test signal and the musical signal in our experiments is listed in Table 6.7.

	Duration (seconds)	**Average groove radius (cm)**	**Number of images**	**Scan time (hours)**	**Storage space (gigabytes)**
Test signal	3.4	13.17	3223	5.97	4.98
Music	12.9	6.79	3897	7.22	6.02

Table 6.7: The scan time and storage space spent acquiring the surface profile images for the 1kHz test sine signal and the music signal

6.4.3 The Pitch Fluctuation Removal

In Section 5.6, we presented a Monte-Carlo optimization framework for removing the pitch fluctuations in our optically-reproduced audio. The pitch fluctuations were caused by the inaccuracies in estimating the disc center before converting the image pixel coordinates to the polar coordinate system relative to the disc center. Our Monte-Carlo framework thus optimizes an objective function to iteratively correct the disc center estimation error. The choice of the objective function is the key to the removal of the pitch fluctuations. Several objective functions (Section 5.6.2.2) were experimented with and are the keys to the evaluated in this section in terms of the output audio quality. Figure 6.1–6.3 show the output stereo audio extracted without correcting the pitch fluctuations. The spectrum is of the right channel signal only. It can be seen that the overall envelope of the waveform appears to be modulated by a low frequency signal (Figure 6.2); also, a low frequency peak can be seen in the frequency spectrum when zoomed in to the frequency band lower than the 1kHz reference frequency (Figure 6.4).

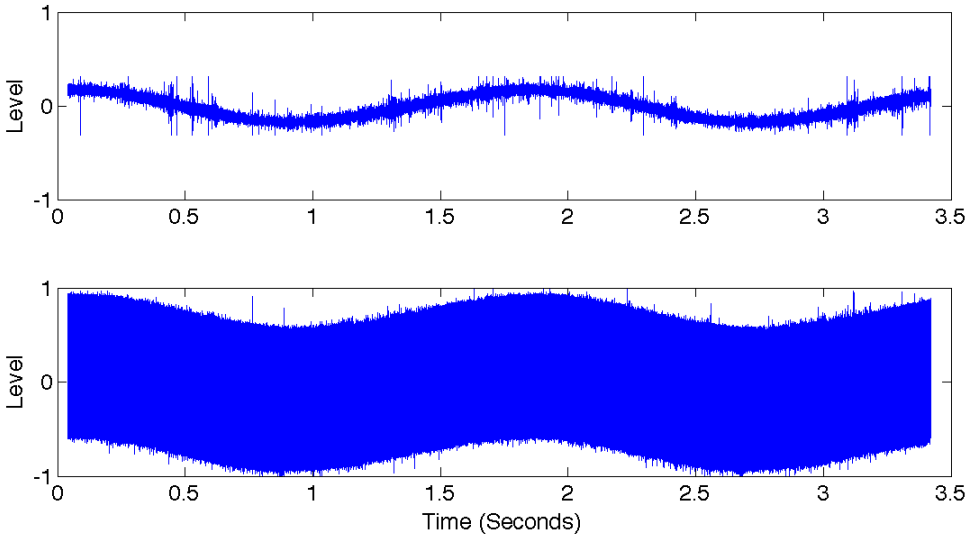


Figure 6.2: The left channel (the upper plot) and right channel (the lower plot) signal waveforms of the output stereo audio with pitch fluctuation

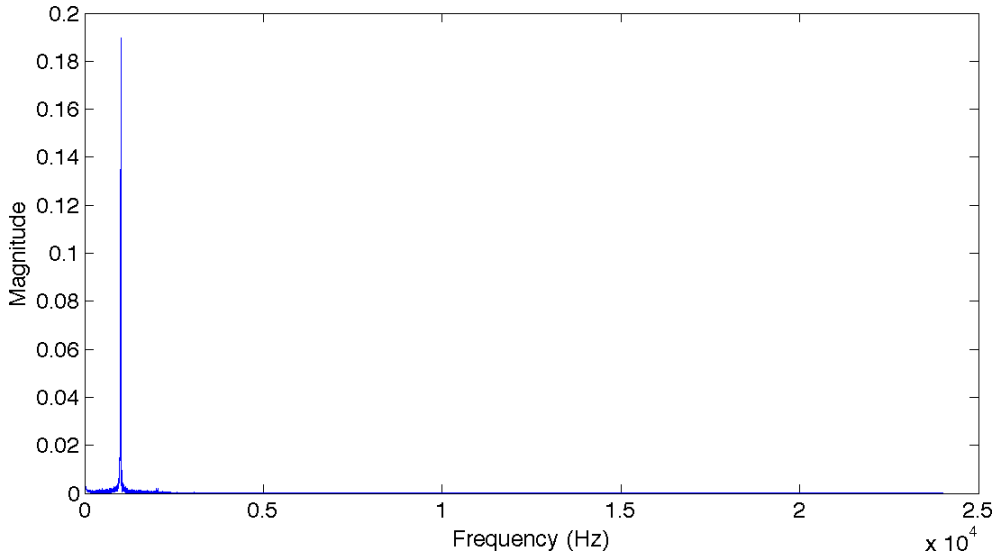


Figure 6.3: The frequency spectrum of the resulting right-channel audio signal up to the Nyquist frequency of the chosen sampling rate

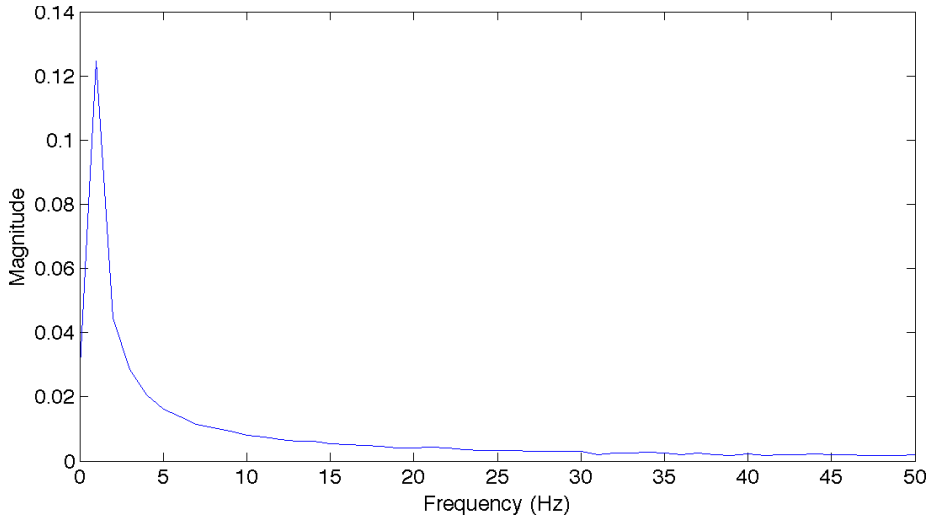


Figure 6.4: The low-frequency peak due to the pitch fluctuation problem from the right channel audio signal (1kHz sine wave)

The SNR and THD results of the output audio without removing the pitch fluctuation and with pitch fluctuation removal by optimizing the available objective functions are presented in Table 6.8. The objective functions are denoted as follows:

1) Minimizing the low-frequency peak below 20Hz
2) Minimizing the value range of raw groove undulation for each revolution
3) Minimizing the deviation around the linearly decreasing envelope for each revolution of the raw groove undulation.
4) Fitting the raw groove undulation to an Archimedean spiral

5) Fitting the raw groove undulation to a spiral curve specified in Fadeyev and Haber (2003)

Objective functions	**Original**	**function 1**	**function 2**	**function 3**	**function 4**	**function 5**
SNR (dB)	12.79	23.59	23.59	23.60	23.59	23.60
THD (dB)	-40.80	-64.20	-64.31	-71.40	-59.22	-59.34

Table 6.8: The SNR and THD of the output audio without removing the pitch fluctuation (Original) and with pitch fluctuation removal by optimizing the available objective functions

It can be seen from Table 6.8 that all objective functions are capable of increasing the SNR and reducing the THD of the output audio compared to the result without pitch fluctuation removal. The objective function 3) leads to the highest SNR and the lowest THD among all. It is thus adopted as the objective function for our Monte-Carlo optimization framework for removing the pitch fluctuations. The waveform and spectrum of the audio with the adopted pitch fluctuation removal approach is shown in Figure 6.5– Figure 6.8, respectively. Figure 6.4 shows that the low-frequency modulation on the signal envelope is removed compared to that of Figure 6.2. Also, it can be seen in Figure 6.7 and Figure 6.8 that the low-frequency peak is largely attenuated compared to Figure 6.4. However, it was also found that none of the available object functions can completely remove the flutter noise.

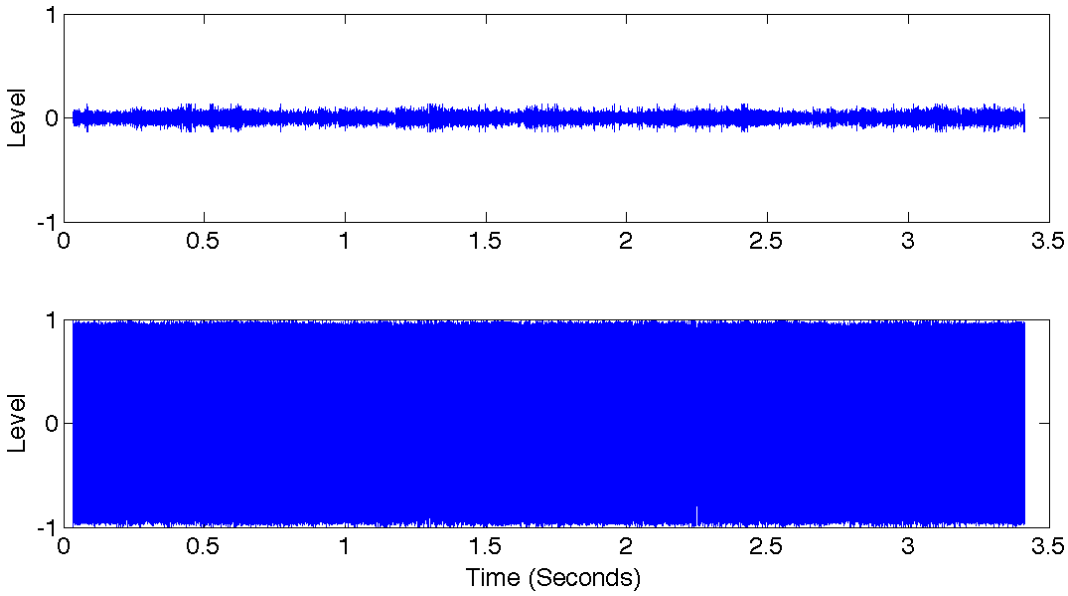


Figure 6.5: The waveform of the output stereo audio after removing the pitch fluctuation by using the objective function 3)

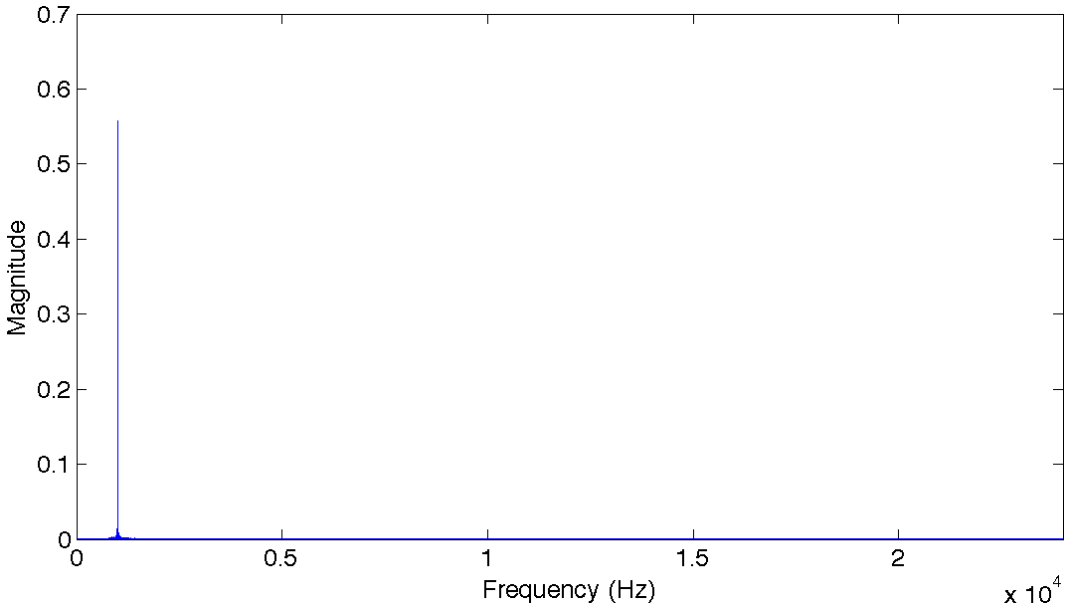


Figure 6.6: The frequency spectrum of the right channel of the same audio signal in Figure 6.5

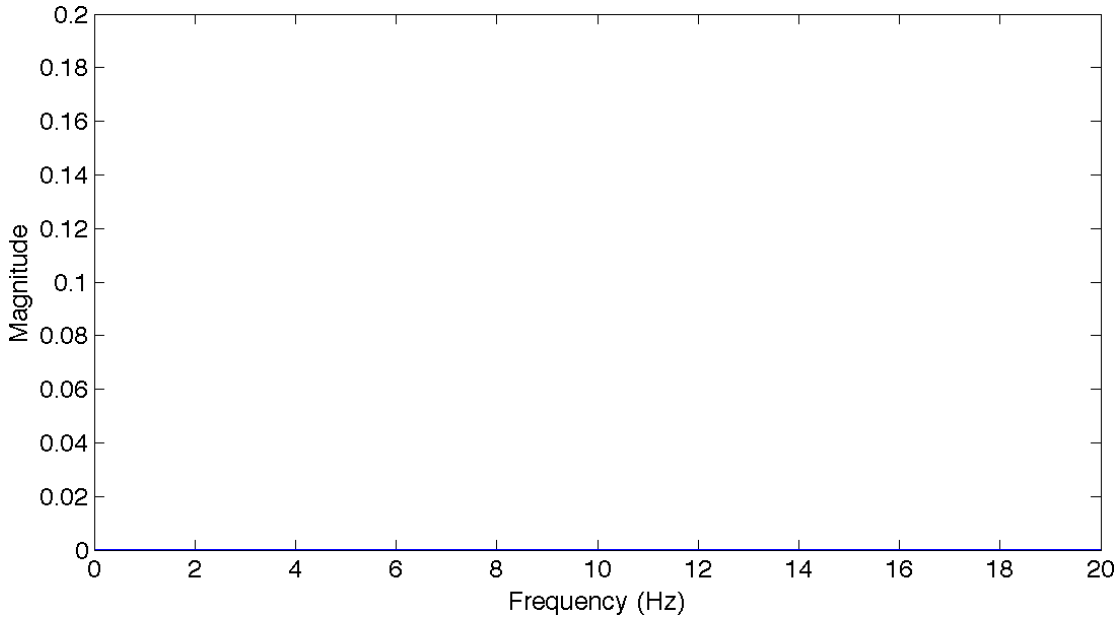


Figure 6.7: The frequency spectrum in Figure 6.6 zoomed in to between 0 and 20Hz. Note that the frequency components affected by pitch fluctuation are greatly attenuated, to the degree that their magnitudes are almost invisible in the figure

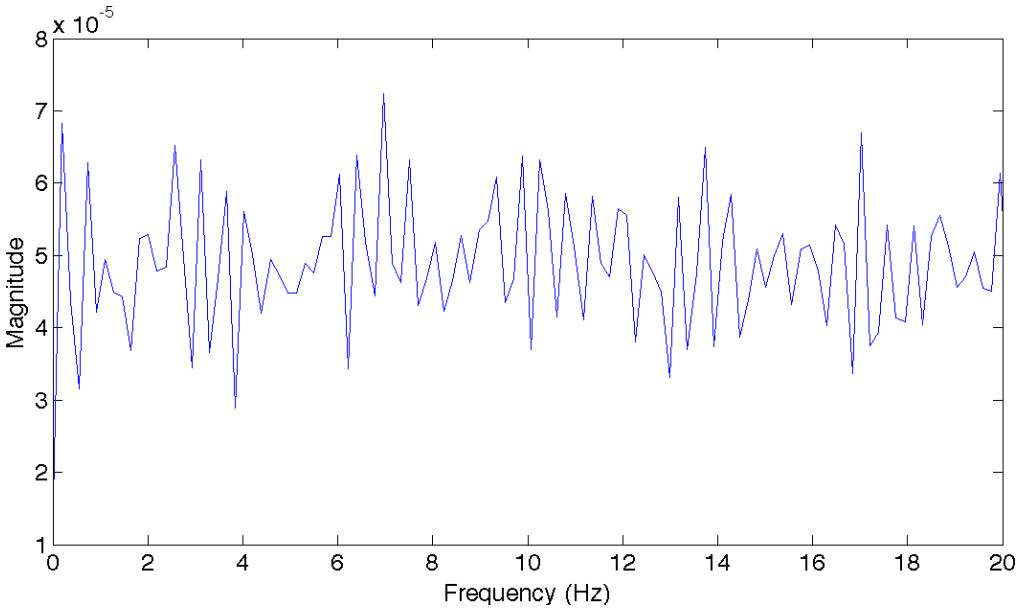


Figure 6.8: The low-frequency spectrum as in Figure 6.7 with a magnified view of its magnitudes

In the experiments for deriving the SNRs and THDs for this section, the resampling algorithm was the piece-wise polynomial fit with 0.5ms window length and 1/4 overlap between windows. The differentiation algorithm was the wavelet differentiator using the discrete approximation of the Meyer wavelet (See Section 6.4.5).

6.4.4 The Resampling Methods

In Section 5.3.3, three available resampling algorithms, namely the global linear interpolation, the global linear interpolation with moving-average filtering, and window-based polynomial curve fitting were introduced, which were used to fit the raw groove undulation time series to an evenly sampled time series for reproducing the digital audio. The approximate input sampling-rate of the extracted raw groove undulation of the test stereo signal was 452.35kHz, which is the theoretical equivalent average audio sampling-rate achieved under 10X magnification for the specific groove revolutions (see Table 6.4). The output sampling-rate was 44.1kHz. The input data was the groove undulation after removing the pitch fluctuation by using our Monte-Carlo optimization framework with the objective function 3 in the previous section.

All of the differentiation methods presented in the next section can have smoothing effect on the results, which will make it difficult to determine the optimal resampling methods. In order to better present the different performances of the resampling methods, we chose to sacrifice the best achievable SNR and THD by our overall OAR workflow by using a differentiator that has only moderate smoothing effect, namely a fourth-order Savitzky-Golay differentiator (Savitzky and Golay 1964) with the window size 49 points, about 1.1ms under the output sampling-rate.

The best SNR and THD results from each of the tested resampling methods are listed in Table 6.9.

	Global linear interpolation	**Global linear interpolation with moving-average filtering**	**Window-based polynomial curve fitting**
SNR (dB)	3.12	19.07	20.18
THD (dB)	-22.68	-25.90	-32.01

Table 6.9: The best SNR and THD achieved by using the optimal parameters of each candidate resampling method

With the global linear interpolation with moving-average filtering, we experimented with various moving-average window sizes, which are from 3 to 113. The maximum window size (113) was about $T_0/4$ where $T_0 =$ 1ms, the period of the test sine

signal of the right channel of the stereo signal. It was discovered that, as the window size increased, the SNR would increase because of the smoothing effect, but the THD would decrease. The best THD was achieved by using the window size 45, about 0.10ms under the input sampling-rate. The best SNR was obtained by using the window size 113.

With the window-based polynomial curve fitting, various polynomial orders and windows were tested. The best SNR and THD results were obtained from fourth-order polynomial fitting with a window size of 0.5ms.

It can be seen that the polynomial curve fitting produces the best audio quality among the three candidate methods. It is thus adopted by our audio reproduction workflow as the resampling method.

6.4.5 The Differentiation Methods

Three differentiation methods were discussed in Section 5.4.2, namely the averaged finite difference scheme (Anderssen et al. 1998), the Savitzky-Golay differentiator (Savitzky and Golay 1964), and the wavelet-transform-based differentiator (Donoho and Johnstone 1994). They are compared in this section with our experimental results. The input data was the groove undulation of the test stereo signal resampled at 44.1kHz by using the fourth-order polynomial fitting with a window size of 0.5ms.

The best SNR and THD achieved by each differentiation method are shown in Table 6.10.

	Averaged finite difference scheme	**Savitzky-Golay differentiator**	**Wavelet-transform-based differentiator**
SNR (dB)	22.43	24.24	37.07
THD (dB)	-59.73	-68.10	-73.91

Table 6.10: The best SNR and THD achieved by using the optimal parameters of each differentiation method

For the averaged finite difference scheme, various window sizes and central difference parameters were tested. The best results were achieved with the window size 0.52ms (23 samples under 44.1kHz) with the central difference parameter 9.

For the Savitzky-Golay differentiator, various polynomial orders and window sizes were experimented with. The best results were obtained with the ninth-order polynomial and the window size 0.88ms (39 samples under 44.1kHz).

For the wavelet-transform-based differentiator, various wavelets and scaling factors were tested. The best results were obtained with the discrete approximation of the Meyer wavelet, which is an FIR filter network implemented to approximate the Meyer wavelet (Mallat 2009, 289–90). It can be seen from Table 6.10 that this wavelet-transform-based differentiator produces the best audio quality among the three differentiators.

6.5 The Output Audio Quality Results from the Available Lateral Image Resolutions

It was shown in Section 6.3 that the lateral image resolution is the key factor for evaluating the output audio quality. It is relevant to the theoretical equivalent audio sampling-rates and dynamic ranges. By adopting the best practice of our OAR workflow shown in Section 6.4, the SNRs and THDs for the stereo signal under various image resolutions are shown in Table 6.11.

Magnification	**2.75X**	**5X**	**10X**
Lateral image resolution (μm/pixel)	3.95	2.00	1.01
SNR (dB)	20.87	24.92	37.07
THD (dB)	6.76	-9.57	-73.91

Table 6.11: The audio quality results of the test stereo signal under the available lateral image resolution options of our image acquisition system

The frequency spectra of both channels of the stereo signal optically reproduced by using 10X magnification, in both normal scale and decibel scale, are shown in Figure 6.9 and Figure 6.10, respectively.

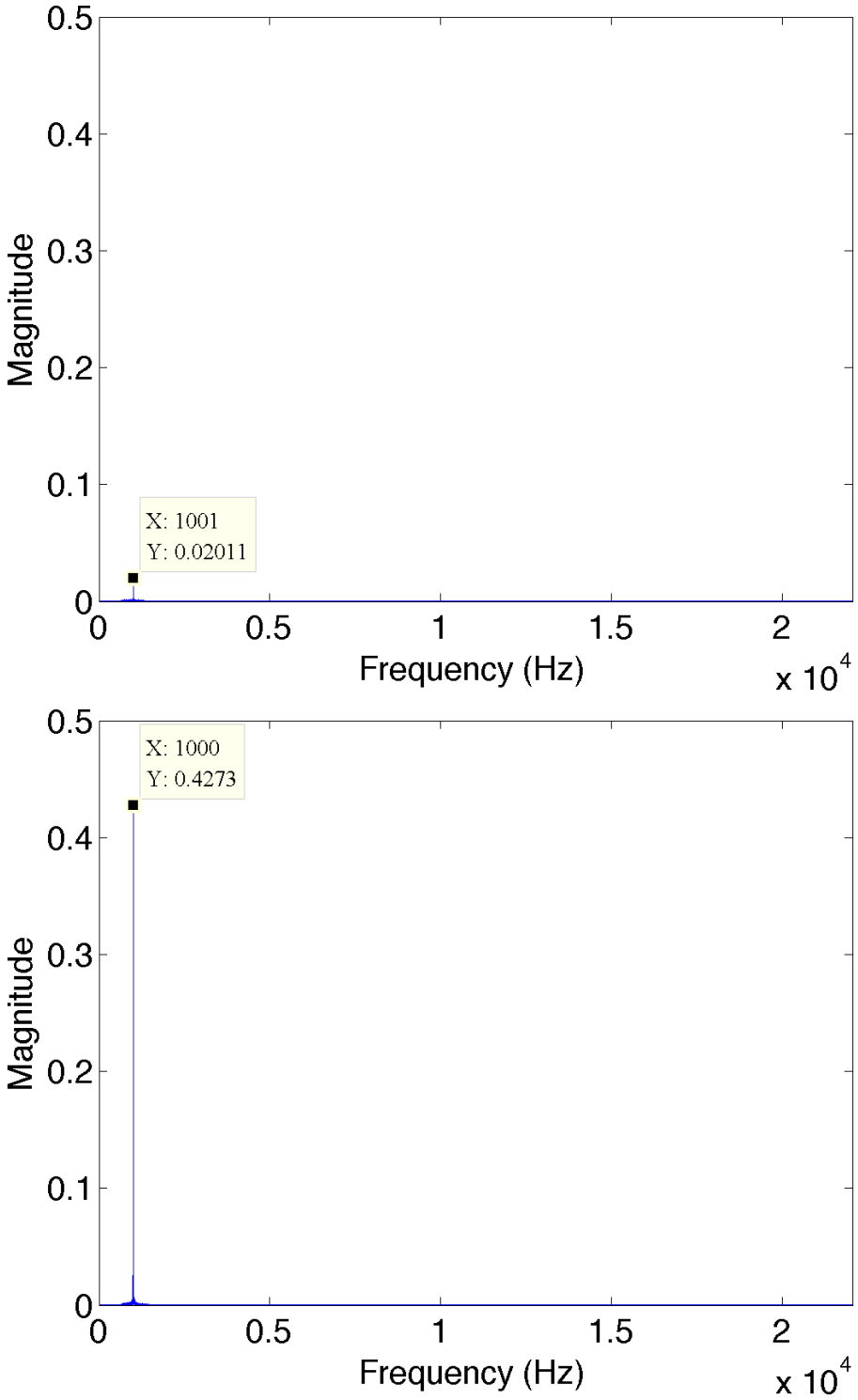


Figure 6.9: The reproduced audio frequency spectrum of the left channel (the upper figure) and right channel (the lower figure) of our test stereo signal in normal scale: The reference frequency and its magnitude are shown in the data tip

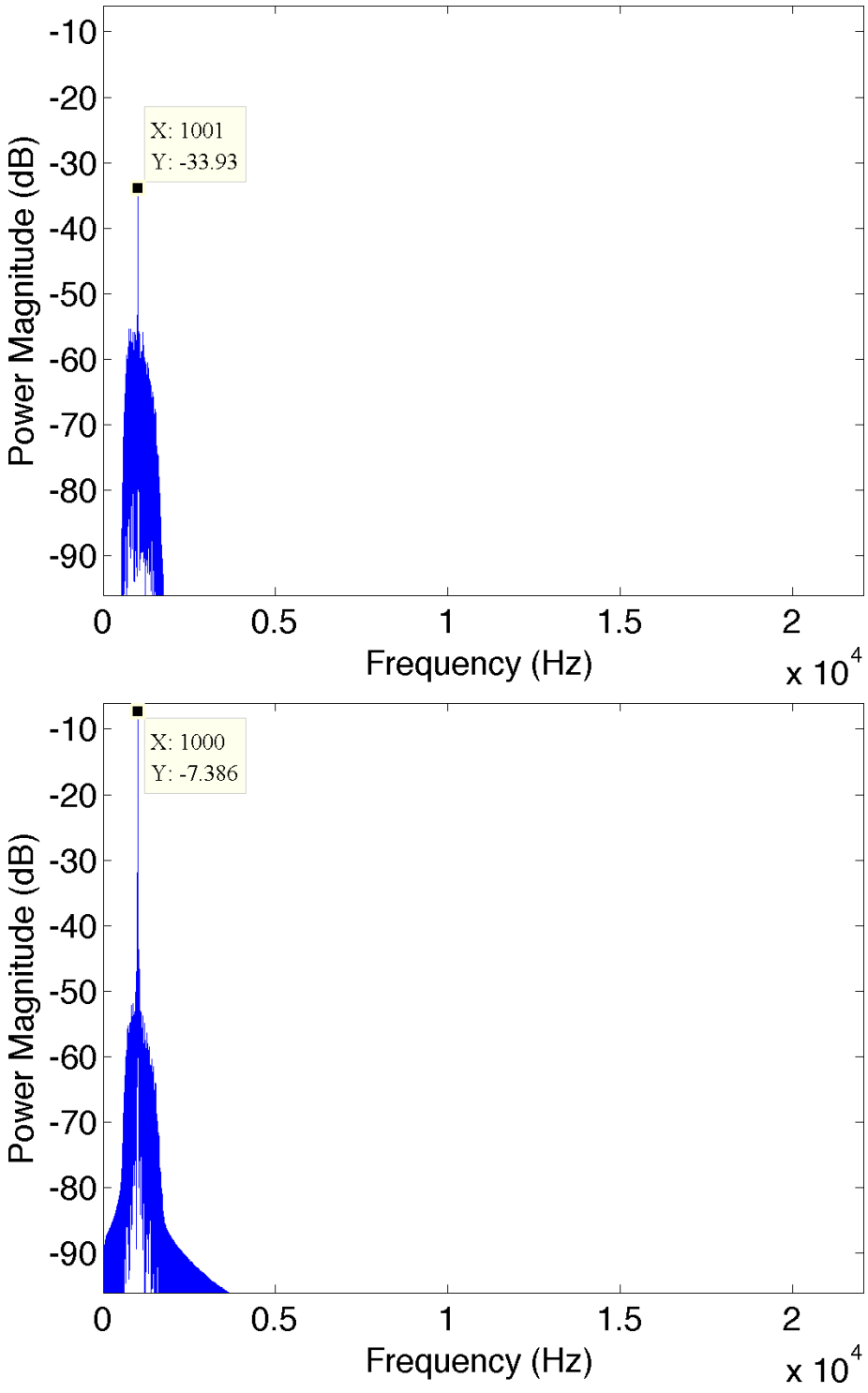


Figure 6.10: The reproduced audio frequency power spectrum of the left channel (the upper figure) and right channel (the lower figure) of our test stereo signal in decibel scale: The reference frequency and its power magnitude are shown with the data tip

By using a turntable digitization system to digitize the same signal without the use of preamp, the resulting SNR and THD are 22.16dB and -46.61dB. The frequency spectra of both channels in normal scale and decibel scale are shown in Figure 6.11 and Figure 6.12, respectively.

It can be seen from Figure 6.9 and Figure 6.10 that the 1kHz sine signal on the right channel was successfully reproduced. The powers of the non-reference frequency components are more than 50dB below the reference frequency. However, although the left channel signal is supposed to be silence, the spectrum of the left channel signal contains a 1kHz peak that is about 26dB below the power of the 1kHz reference frequency component on the right channel. This appears to be an inter-channel crosstalk. It can be seen from the result that the SNR is lower than theoretically expected (see Table 6.4). Possible reasons include the fact that the pitch fluctuation removal cannot fully remove the low-frequency components introduced by the pitch fluctuation, which contributes to the noise power.

Figure 6.11 and Figure 6.12 show the spectra of the test stereo signal reproduced from the turntable digitization system. In Figure 6.12, harmonic distortion appears to be more significant than the optically-reproduced audio (see Figure 6.10). Also, the inter-channel crosstalk is also present in the signal.

In comparison with the turntable-digitized version, the OAR version achieved better SNR and THD.

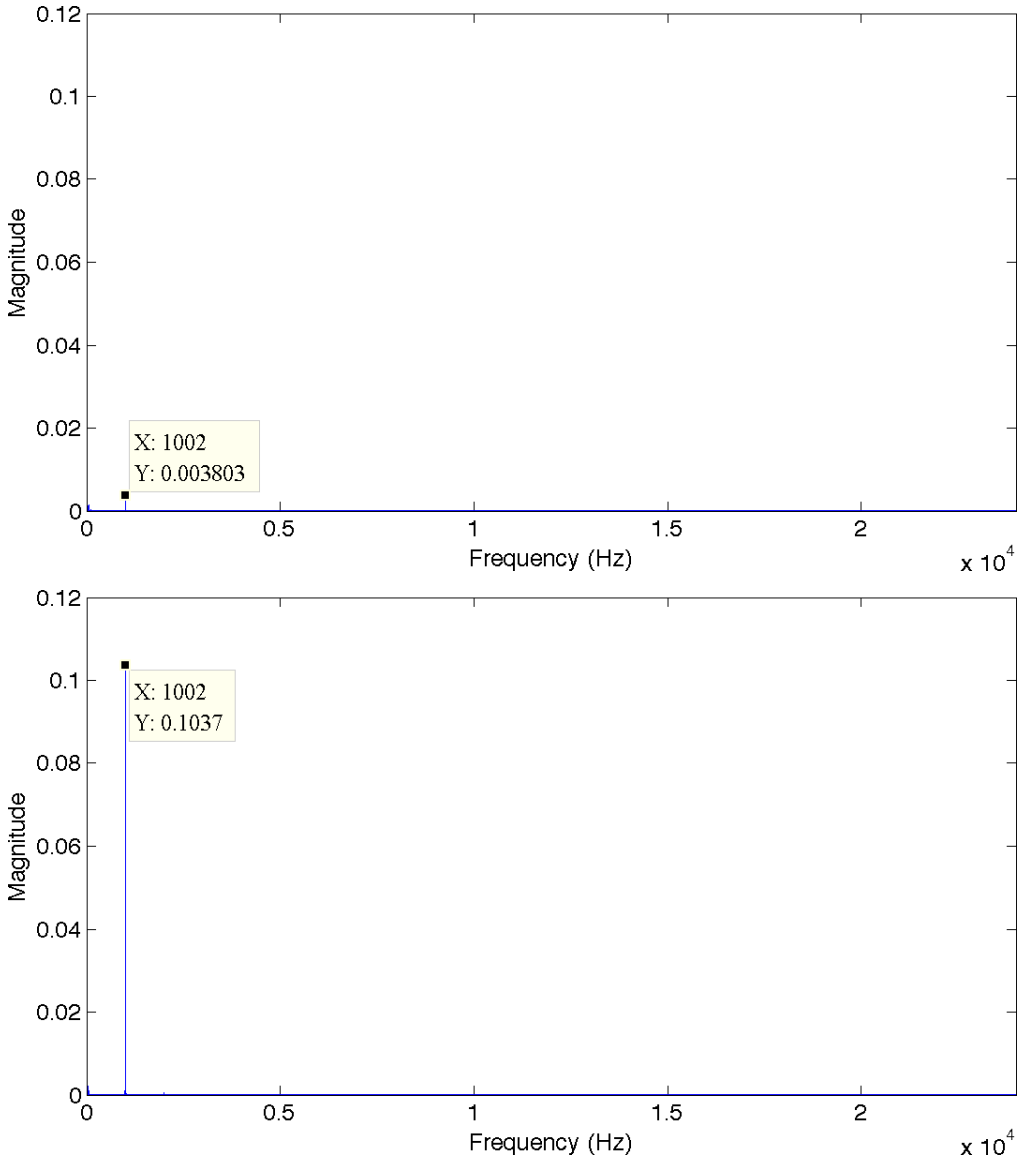


Figure 6.11: The frequency spectra of the left channel (the upper figure) and right channel (the lower figure) of the test stereo signal in normal scale reproduced by using a turntable digitization system: The reference frequency and its magnitude are shown in the data tip

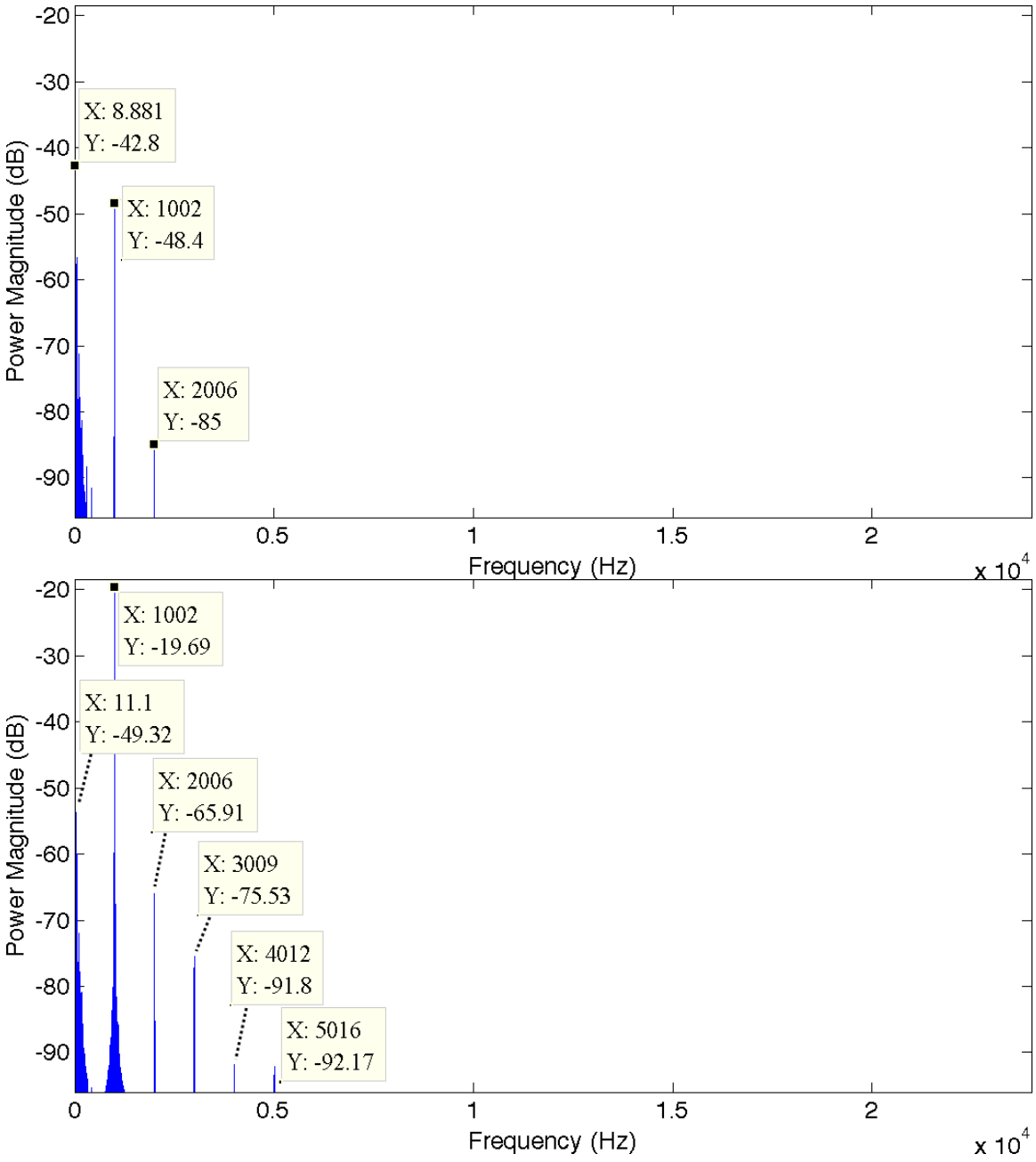


Figure 6.12: The same frequency spectra as in Figure 6.11 in decibel scale with the data tips showing the frequencies and power magnitudes of the reference frequency component and the harmonics

The specifications of the turntable digitization system for producing the turntable version of the test signal are listed below:

- Turntable: VPI HRX with JMW 12.7 Tonearm (VPI Industries, Inc. 2011)
- Cartridge: Ortofon Kontrapunkt B (Ortofon A/S 2011)

- Preamplifier: Manley Steelhead with INPUT: MC1, Gain: 55 dB, Load 50 Ohms, Load cap. pF: x100: 0, x10: 0 (Manley Laboratories, Inc. 2011)
- ADC: PrismSound ADA8XR, with 44.1kHz output sampling rate, and 16bit bit-depth (Prism Sound Limited 2011)
- Interconnection cables: Kimber Kable KCAG (Kimber Kable 2011)
- Software: Logic stereo audio 7 (Apple Inc. 2011)

6.6 The Comparison of the Musical Audio Quality with the Turntable and the CD Versions

A stereo musical excerpt of about 13 seconds (see Section 6.2) is optically reproduced with our OAR workflow under the 10X magnification. The waveforms and the spectrograms of the left and right channel signals of the optically-reproduced version, the turntable-digitized version (from Bernstein SMK 60208–AC), and the CD version (Bernstein SMK 60208–AC) are shown in Figure 6.13–Figure 6.15 and Figure 6.16–Figure 6.18, respectively. In these figures, the upper plot represents the left channel and the lower the right channel. The sampling-rate is 44.1kHz. The short-time Fourier transform for creating the spectrogram uses the Hamming window of size 2048 samples and 50% overlap; the length of the Fast Fourier Transform is 256.

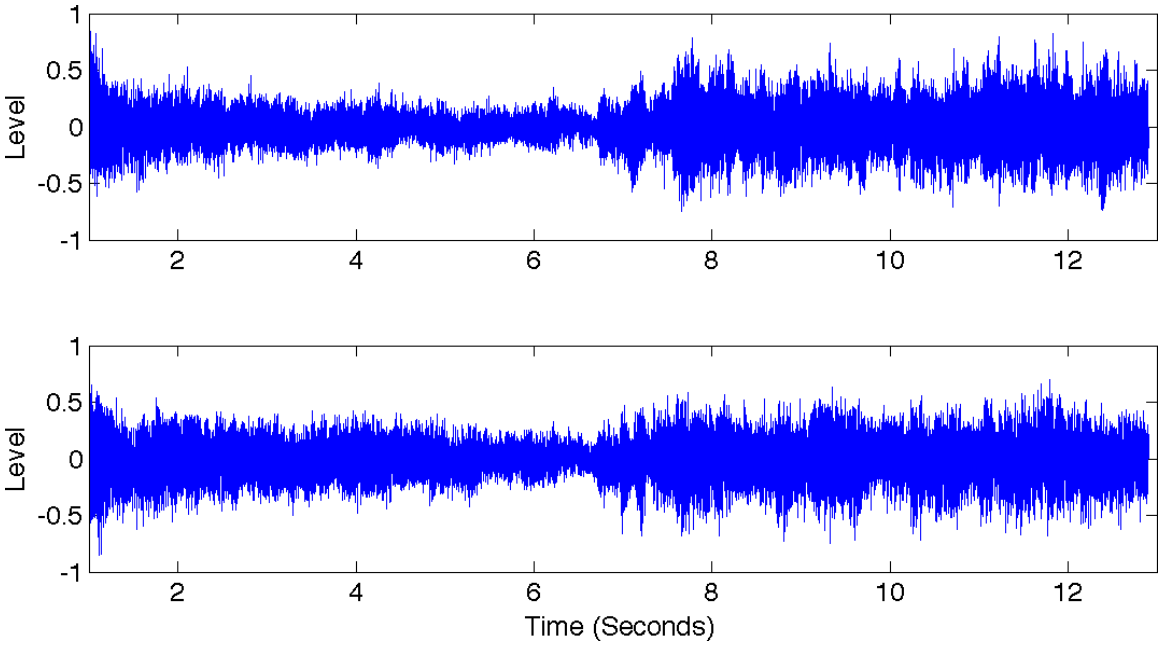


Figure 6.13: The waveform of the optically-reproduced stereo musical signal

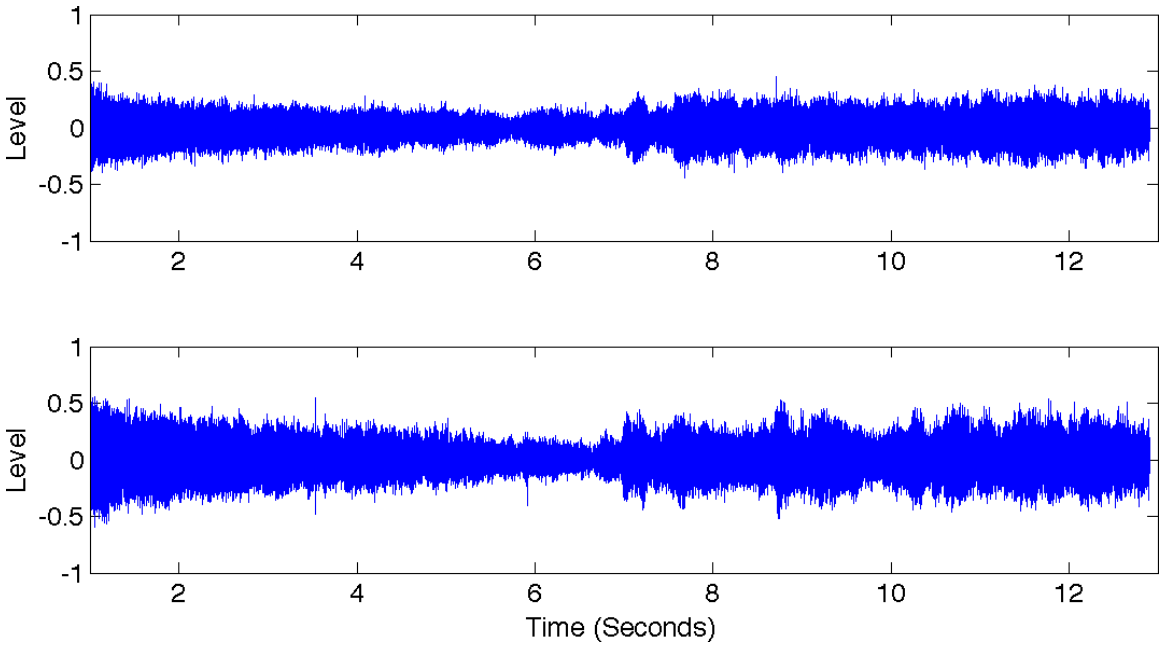


Figure 6.14: The waveform of the same music signal as in Figure 6.13 digitized through a turntable

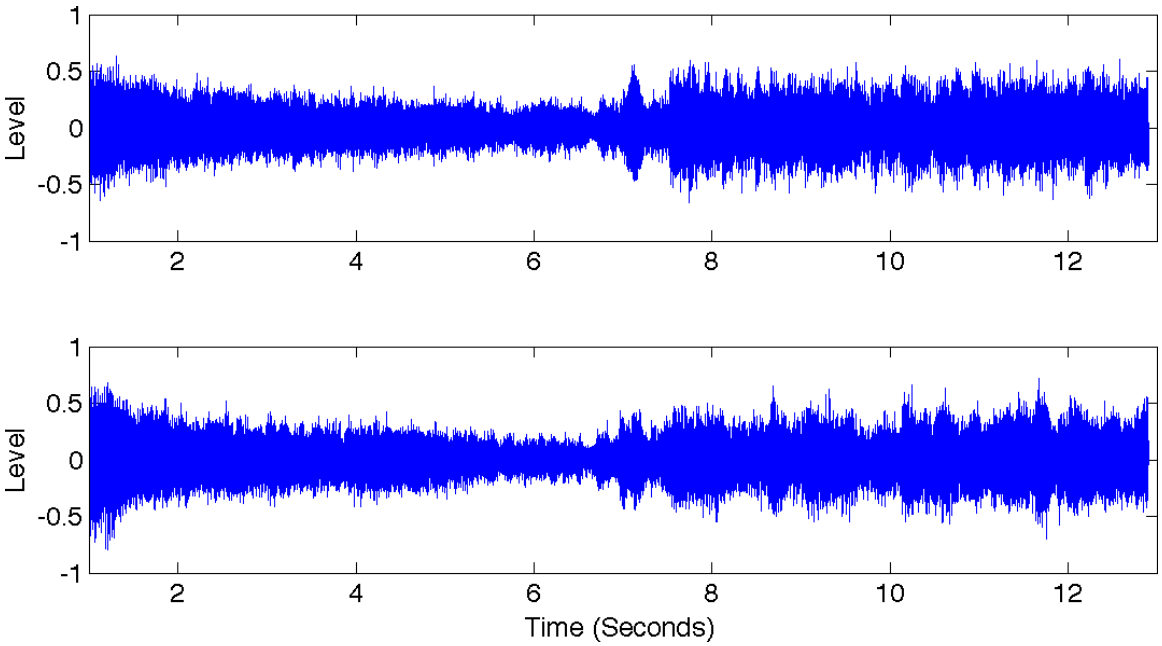


Figure 6.15: The waveform of the CD version of the same music signal as in Figure 6.13

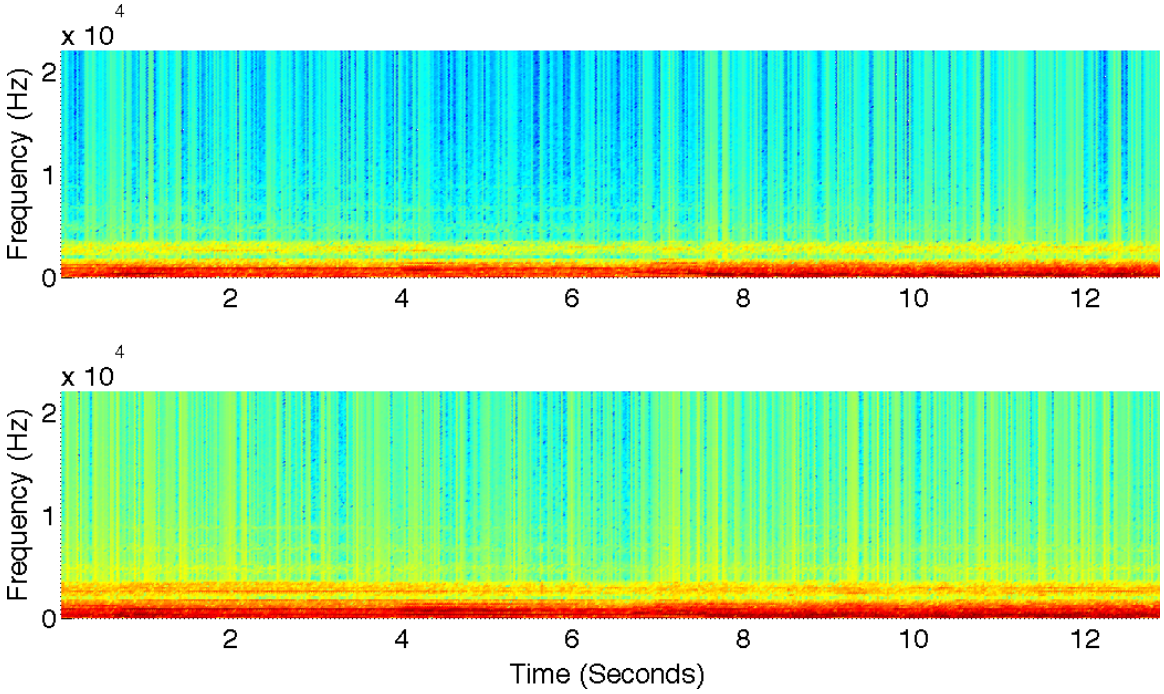


Figure 6.16: The spectrogram of the optically-reproduced stereo musical signal

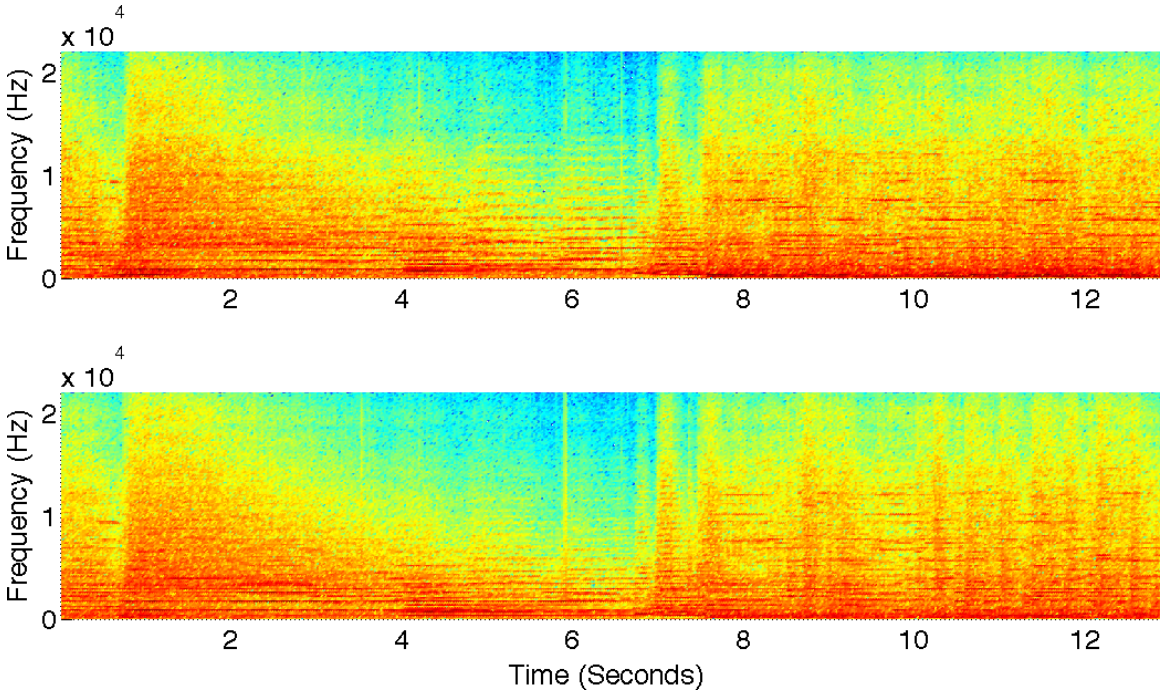


Figure 6.17: The spectrogram of the stereo musical signal digitized by a turntable system as in Figure 6.16

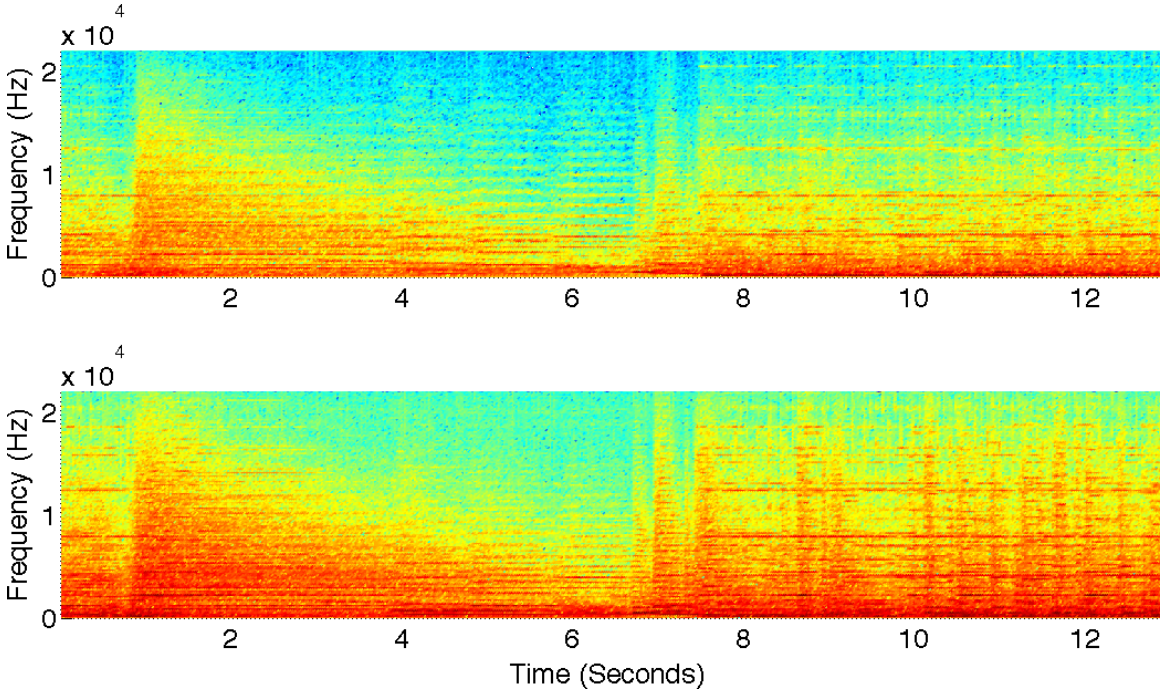


Figure 6.18: The spectrogram of the CD version of the stereo musical signal

It can be seen in Figure 6.13–Figure 6.15 that the waveforms from these different versions are similar to each other. From Figure 6.16–Figure 6.18, we notice that the mid-range and high-frequency components of the OAR version are largely attenuated compared with the CD version and the LP version. This was due to the smoothing effect of the wavelet differentiator. By decreasing the scaling factor of the wavelet differentiator, we have limited ability to enhance the high-frequency components, at the expense of greater noise. For example, by using the scaling factor 8 and the same mother wavelet as in Figure 6.16, the mid and high frequency components of resulting signal in Figure 6.19 are enhanced compared to those in Figure 6.16.

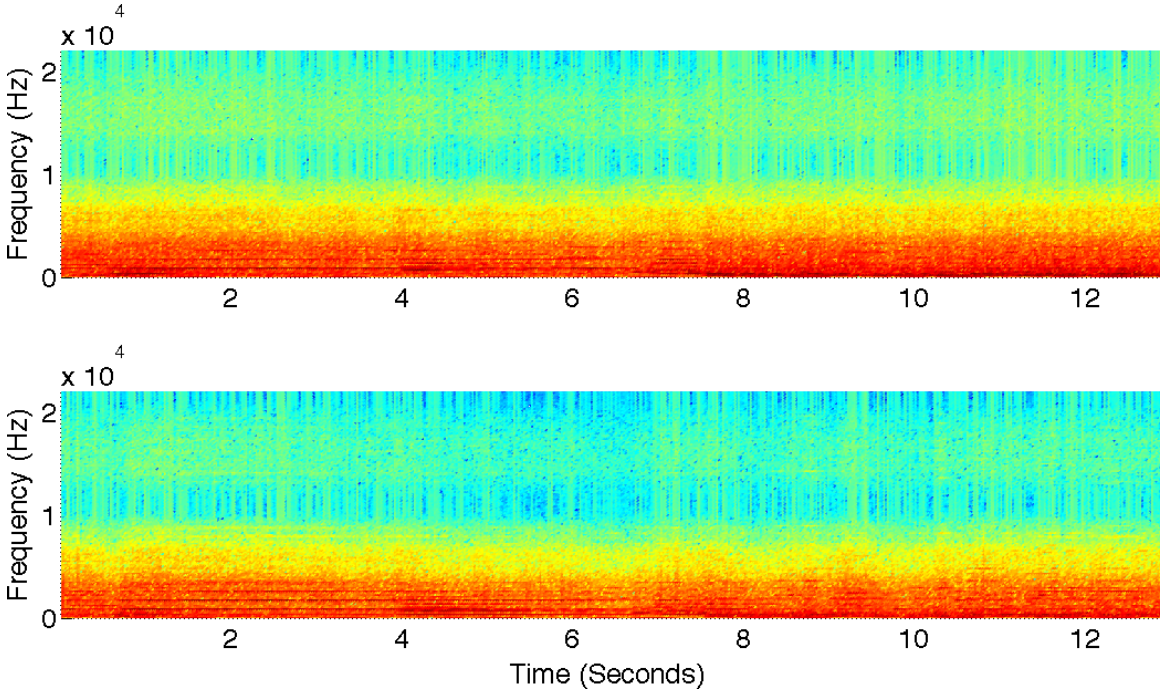


Figure 6.19: The spectrogram of the optically-reproduced musical signal by using a small scaling factor for the wavelet differentiator

By listening to the three versions of the music excerpt, it was found that the turntable-digitized version contains occasional clicks, which was not heard in the CD version or the optically reproduced version. The optically reproduced version sounds muffled compared to the other two versions. This was due to the smoothing effect of the differentiation, similar to low-pass filtering. The sounds of CD version and the turntable-digitized version sound much brighter, that is to say, rich in high frequency content.

It is worth mentioning that we did not compare the sound quality of our optically reproduced music signal with the master tape of the same recording, if it existed, because it was difficult for us to find it at the time. Due to the same reason, the equipment used for creating the master CD, such as the pre-amp, was also unavailable.

6.7 Summary

In this chapter, the theoretical sampling rates and dynamic ranges of our OAR system were analyzed. The workflow for optically reproducing the stereo audio was evaluated to ensure that the best practice be used when multiple candidate solutions are available at the certain steps. The optically-reproduced audio quality was evaluated and compared with

the turntable and CD versions. Our OAR approach can successfully reproduce the stereo audio information from the acquired record surface profile images.

Chapter 7 Conclusions and Future Work

7.1 About the Chapter

This dissertation focused on optically reproducing stereo audio signals from the stereo LP (long-play) phonograph records by using a specific non-contact surface metrology technique, the white-light interferometry (WLI). We were able to reproduce the stereo audio by dividing the workflow into three major steps. First the record surface profile images were acquired by scanning the record surface with a WLI microscope; then by using image processing techniques, the record groove undulations were extracted; finally the audio signal is reproduced by processing the extracted groove undulation time-series with signal processing techniques.

In this concluding chapter, we first summarize the research of this dissertation in Section 7.2 and present some of the original contributions in 7.3. Finally the possible future work is discussed in Section 7.4.

7.2 Summary and Conclusions

The digital preservation of phonograph records has been acknowledged as a government decision in the U.S. (*US National Recording Preservation Act of 2000* 2000). As an alternative approach to the turntable-based digitization, the methods of optically reproducing the audio content from the phonograph records have also received sponsorships from the government (Fadeyev and Haber 2003). Previous optical audio reproduction (OAR) approaches have demonstrated their abilities to reproduce the audio from the cylinder recordings (Fadeyev and Haber 2004, Nascé 2009) and the mono 78rpm recordings (Fadeyev and Haber 2003, Stotzer 2006, Tian 2006).

This dissertation presents a new approach to reproducing the stereo audio from the stereo LP records without contacting the record surface. This approach uses a general-purpose white-light interferometry (WLI) microscope to acquire the 3D record surface profile images. The undulations of the record grooves are then extracted and certain noise and distortions present in the images are removed by using image processing techniques. The resulting groove undulations, as time series, are further processed to output the stereo audio signals. The entire OAR workflow for reproducing the stereo audio signal was

examined to ensure that the best practice is adopted in each step of the procedure. The optimal OAR workflow is performed on a test LP record that contains standard sine signals and a regular LP of musical signals. The experimental results of the output audio quality are quantitatively evaluated and qualitatively compared with the qualities of the turntable-digitized version and the CD (compact disc) version.

7.3 Research Contributions

This dissertation presents the first optical audio reproduction approach in the world for extracting stereo audio from the stereo LP phonograph records. The original contributions of this dissertation includes applying the WLI to measure the 3D profile of the record grooves that encode stereo audio, exploiting the acquired 3D profile and existing digital image processing techniques to extract groove geometry and restore damaged grooves, and unwrapping grooves into continuous audio signals via software methods. To the best of our knowledge, the WLI has never been used for the audio reproduction for the stereo LP records before the research presented in this dissertation. Different from many previous OAR approaches for disc recordings (Fadeyev and Haber 2003, Stotzer 2006, Tian 2007), the height information of the record surface is used for extracting the groove undulations from the record surface profile images during the image analysis phase of our approach. According to our experiments, it is proved that without being able to obtain the groove sidewall information, the stereo audio signals can still be reproduced with our OAR approach. We also presented the analysis of the noise and distortion in the acquired record surface profile images, their causes, and the detailed solutions for removing them by using image processing methods. The theoretically achievable audio sampling rates and dynamic ranges corresponding to the available image resolution options of our microscope and the relationship between the image resolution and the audio quality of our WLI system were analyzed, which is an unprecedented effort according to our knowledge.

7.4 Future Work

It has been shown in this dissertation that our OAR approach can successfully reproduce the stereo audio information from the record surface profile images. There are nevertheless many aspects in which future efforts can be invested to either improve our

current practice or present new directions for optically reproducing stereo audio from LP recordings in a more efficient manner.

In Section 3.4, it was discussed that the height fluctuations present across the record surface would require extra vertical scanning time for acquiring the groove 3D information over a large area. This is because our microscope is unable to dynamically configure the vertical scan range to adapt to the surface contour of interest. A specially designed multi-grid automatic measurement was used to reduce the scanning time of our microscope for resolving depth acquiring large area of the disc. This, however, requires manually focusing onto the groove surface for each grid of the multi-grid measurement. In the future, it is worth exploring the possibility of dynamic scan range control so that the vertical scan time can be reduced by following closely to the record surface without damaging the lens.

In Section 4.5, considerable efforts were made to remove the noise and distortions caused by dusts that accumulate on the record surface during the long course of the acquisition and processing. In the future, it is thus reasonable to exploit the possibility of conducting the acquisition of the record surface profile images in a dust-free environment and evaluate the improvement in the noise conditions of the images that the adoption of such an environment could possibly achieve.

More investigation can be made to determine the relationship between the pitch fluctuation removal process and the inter-channel crosstalk defect discussed in Section 6.4. So far it has not been clearly understood.

In our approach, although 3D information of the record surface is partially obtained (without the groove sidewalls) and used in the extraction of the groove undulations from the record images (Section 4.3), the groove depth information was not exploited in the decoding of the stereo audio information. This was mainly due to following two reasons:

1) The complete groove bottom information was often difficult to obtain by using our image acquisition system (Section 4.5).

2) The depth information is unnecessary to obtain the stereo audio information. The 45/45 groove-cutting scheme can be utilized to infer the groove sidewall

orientation and the depth from the groove width information, which is available in the lateral information of the acquired record surface profile images.

Despite these reasons, it would nevertheless be interesting to further investigate how the groove depth information can be used to improve the groove extraction and audio reproduction workflow.

It would also be interesting to investigate alternative image acquisition schemes with the same WLI microscope. Our image acquisition was performed with an X-Y stage, which leads to two consequences that determine the later data analysis processes:

1) The groove orientations in the resulting digital images are arbitrary so that no simple correspondence can be found between the pixels and the eventual audio samples. Considering the square pixel tessellation scheme of the digital images (Section 5.3.1), the extracted groove undulation time series are therefore unevenly spaced in time. Moreover, because the groove undulation time series are usually oversampled compared to the output audio sampling rate, interpolation is thus required to resample the time-series. This process could introduce errors in estimating the contour of the downsampled groove time series.

2) As a side effect of 1), the image stitching, which is required for the scanning of large area of the record surface, can introduce 2D random sub-pixel errors. These errors are further converted into more complex errors in the extracted 1D groove undulation time-series during the software groove unwrapping. This is due to the non-uniform orientation of the grooves in the large number of images so that the 2D pixel errors cannot be easily translated into the errors in time alignment and amplitude mismatching in the time-series resulting from the groove undulation extraction and unwrapping. This situation could distort the extracted groove undulations and increase the noise level.

3) A process separate from the image acquisition is required to estimate the center of the disc record when deriving the polar coordinates of the acquired images. Furthermore, this initial estimate is usually inaccurate and needs to be corrected to avoid the pitch fluctuations in the output audio.

If a rotary disc-hosting device and the rotary scan as used in many previous OAR approaches (Stotzer 2006, Tian 2006, Nascé 2009) were used during the scanning, several advantages could be achieved to solve the above problems in our current workflow more conveniently. First of all, the rotary device can unwrap the grooves during the image acquisition so that grooves are uniformly straight in the obtained images. As a result, the later downsampling for the groove undulations becomes more convenient because the pixels that compose the groove undulation time-series would be evenly spaced in pixels in the unwrapped grooves in the images. Secondly, if the grooves are already straightened in the acquired images, the stitching errors between the adjacent images in an automatic scan would become easier to predict and compensate because the errors can be decomposed directly into 1D pixel shifts in time (the tangential direction of the groove) or the amplitudes (radii) of the extracted groove undulation time-series. Finally, a mechanical rotary device usually provides a native center that can be utilized to match the disc center hole during the scanning, which could reduce the center estimation problem in the above Problem 3). However, we must also realize that having a mechanical rotary device during the image acquisition can introduce other defects that previous researches have reported (Stotzer 2006, 67). Much work can be expected for investigating the usage of the rotary scan.

Bibliography

Alexanderson, E. 1934. Sound reproducing apparatus. *US Patent* 1,978,183, filed 27 January 1934, and issued 23 October 1934.

Anderssen, R., F. de Hoog, and M. Hegland. 1998. A stable finite difference Ansatz for higher order differentiation of non-exact data. *Bulletin of Australian Mathematical Society* 58: 223–32.

Apple Inc. 2011. Logic Studio. http://www.apple.com/logicstudio/ (accessed 6 February 2011).

Audio Junkies, Inc. 2008. Turntable basics: A beginner's guide to turntables and vinyl records. http://media.audiojunkies.com/turntable-parts-record-player-technics-sl1200-sl-1200.jpg (accessed 6 February 2011).

Ballard, D. 1982. *Computer Vision*. Englewood, NJ: Prentice-Hall.

Balsubramanian, N. 1980. Optical system for surface topography measurement. *US patent* 4,340,306, filed 4 February 1980; issued 20 July 1982.

Bennett, W. 1948. Spectra of quantized signals. *Bell System Technical Journal* 27 (4): 446–72.

Berliner, E. 1887. Gramaphone. *US Patent* 372,786, filed 4 May 1887, and issued 8 November 1887.

Bhushan, B., J. Wyant, and C. Koliopoulos. 2002. Measurement of surface topography of magnetic tapes by Mirau interferometry. *Applied Optics* 24 (10): 1489–97.

Blumlein, A. 1931. Improvements in and relating to sound-transmission, sound-recording, and sound-reproducing systems. *US Patent* 394,325, filed 14 December 1931, and issued 14 June 1933.

Blunt, R. 2006. White Light Interferometry: A production worthy technique for measuring surface roughness on semiconductor wafers. *Proceedings of the CS MANTECH Conference*: 59–62.

Boltryk, P., M. Hill, J. McBride, A. Nascè, N. Bewley, and W. Prentice. 2007. Non-contact surface metrology for preservation and sound recovery from mechanical sound recordings. *Proceedings of Joint Technical Symposium: Audio Visual Heritage and the Digital Universe*: 28–30.

Brock-Nanestad, G. 2001. The attraction of optical replay of mechanical recordings. *Proceedings of International Conference of Archiving, Restoration, and New Methods of Recording*: 1935–9.

Bruderhofer, N. 2010. The Cylinder Archive: All about Edison phonograph cylinder records. http://www.cylinder.de/guide_black-wax-cylinders.html (accessed 6 February 2011).

Cornell, E., V. Fadeyev, C. Haber, J. Jin, R. Nordmeyer, and M. Golden. 2007. Using optical metrology to reconstruct sound recordings. *Nuclear Instruments and Methods in Physics Research* A (579): 901–4.

Dahlen, P. 1965. Optical phonograph apparatus with polarized light. *US Patent* 3,452,163, filed 8 December 1965, and issued 24 June 1969.

Diniz, P., E. da Silva, and S. Netto. 2010. *Digital Signal Processing System Analysis and Design*. 2nd ed. New York: Cambridge University Press.

Diop, S., J. Grizzle, and F. Chaplais. 2003. On numerical differentiation algorithms for nonlinear estimation. *Proceedings of the IEEE Conference on Decision and Control* 3: 2896–901.

Edison, T. 1877. An improvement in Phonograph, or speaking machine. *US Patent* 200,521, filed 15 December 1877, and issued 19 February 1878.

ELP Corporation. ELP Laser Turntable: plays vinyl records without a needle. http://www.elpj.com/main.html (accessed 6 February 2011).

Eugenio, R. 2010. Garrard 401, Dennesen air bearing linear tracking tonearm. http://www.theanalogdept.com/images/spp6_pics/Garrard/Rodney%20Eugenio/re024.jpg (accessed 5 May 2011).

Fadeyev, V., and C. Haber. 2003. Reconstruction of mechanically recorded sound by image processing. *LBNL Report* 51983.

Fadeyev, V., C. Haber, C. Maul, J. McBride, and M. Golden. 2005. Reconstruction of recorded sound from an Edison cylinder using three-dimensional non-contact optical surface metrology. *Journal of Audio Engineering Society* 53 (6): 485–508.

Finial Technology. 1986. *Finial LT-1 optical turntable*.

Friebus, R. 1929. Sound reproducing means and method. *US Patent* 1,916,973, filed May 21, 1929, and issued July 4, 1933.

Fujinaga, I. 1997. *Adaptive optical music recognition*. PhD diss., McGill University.

Galo, G. 1996. Disc recording equalization demystified. *ARSC Journal* 17 (2): 188–211.

Glackin, D, and S. Ricker. 2002. Mastering Maven Mills microscopically minute miracle. http://www.positive-feedback.com/Issue2/mastering.htm (accessed 6 February 2011).

Gotoh, T., Y. Ohta, M. Yoshida, and Y. Shirai. 1987. Component labeling algorithm for video rate processing. *Proceedings of SPIE: Advances in Image Processing* 804: 217–24.

Gray, K. 1997. Producing great sounding phonograph records. http://www.recordtech.com/prodsounds.htm (accessed 6 February 2011).

Haber, C. 2008. Imaging historical voices. *International Preservation News* 46: 23–8.

Haralick, R., and L. Shapiro. 1992. *Computer and Robot Vision, Volume I*. Boston, MA: Addison-Wesley.

Harasaki, A., J. Schmit, and J. Wyant. 2000. Improved vertical-scanning interferometry. *Applied Optics* 39 (13): 2107–15.

Hariharan, P. 1992. *Basics of Interferometry*. Boston: Academic Press.

Hauser M. 1990. Overview of oversampling A/D conversion. *The Audio Engineering Society Convention*. 2973–3005.

Heine, W. 1974. Disc phonograph record playback by laser generated diffraction pattern. *US Patent* 3,992,593, filed 22 August 1974, and issued 16 November 1976.

Hodges, R. 1989. The optical turntable. *Stereo Review* 54 (March): 136.

Hoffmann, F., and H. Ferstler. 2005. *Encyclopedia of Recorded Sound*. 2nd ed. New York: Routledge.

Huber, D., and R. Runstein. 2005. *Modern Recording Techniques*. 6th ed. Burlington, MA: Elsevier Inc.

IASA Technical Committee. 2004. Guidelines on the Production and Preservation of Digital Audio Objects, ed. by Bradley, K. IASA-TC-04.

Inoue, T., N. Takahashi, and I. Owaki. 1970. A discrete four-channel disc and its reproducing system (CD-4 system). *Journal of the Audio Engineering Society* 19 (7): 576–83.

Iwai, T., T. Asakura, T. Ifubuke, and T. Kawashima. 1986. Reproduction of sound from old wax phonograph cylinders using the laser-beam reflection method. *Applied Optics* 25 (5): 597−604.

Kimber Kable 2011. The KCAG interconnection cables. http://www.kimber.com/products/interconnects/analog/kcag/ (accessed 6 February 2011).

Kino, G., and S. Chim. 1990. Mirau correlation microscope. *Applied Optics* 29 (26): 3775–83.

Kollár, I. 1986. The noise model of quantization. *Proceedings of IMEKO TC4 Symposium of Noise in Electrical Measurements*: 125–9.

Li, B., S. de Leon, and I. Fujinaga. 2007. Alternative digitization approach for stereo phonograph records using optical audio reconstruction. *Proceedings of International Conference on Music Information Retrieval*. Vienna. 165–6.

Library of Congress. 1930. Emile Berliner, with the model of the first phonograph machine which he invented. http://lcweb2.loc.gov/service/pnp/cph/3c20000/3c24000/3c24100/3c24124r.jpg (accessed 6 February 2011).

Mallat S. 2009. *A Wavelet Tour of Signal Processing*. 3rd ed. Burlington, MA: Elsevier Inc.

Manley Laboratories, Inc. 2011. The Manley Steelhead RC Phono Stage. http://www.manley.com/mshlr.php (accessed 6 February 2011).

Maxfield, J., and H. Harrison. 1925. Vibratory system. *US Patent* 1,535538, filed 3 January 1923, and issued 28 April 1925.

Maxfield, J., and H. Harrison. 1926. Methods of high quality recording and reproduction of music and speech based on telephone research. *The Bell System Technical Journal* 3: 493–523.

Morgan, J. 1953. *Introduction to Geometrical and Physical Optics*. New York: McGraw-Hill Book Company, Inc.

Nakamura, T., T. Ushizaka, J. Uozumi, and A. Toshimitsu. 1997. Optical reproduction of sounds from old phonographic wax cylinders. *Proceedings of SPIE* 3190: 304–13.

Nascè, A. 2009. Methods for signal reproduction from mechanical sound carriers, measured via non-contact, full surface mapping. PhD diss., University of Southampton.

National Association of Broadcasters. 1964. *NAB Recording and Reproducing Standards for Disc Recording and Reproducing*. Washington, DC: National Association of Broadcasters.

Newville, L. 2009. Development of the phonograph at Alexander Graham Bell's Volta Laboratory. *United States National Museum Bulletin* 218, *Paper* 5: 69–79.

Olszak A. 2000. Lateral scanning white-light interferometer. *Applied Optics* 39 (22): 3906–13.

Ortofon A/S. 2011. The Kontrapunkt B Phonograph Cartridge. http://www.ortofon.com/index.php?option=com_content&view=article&id=63&Itemid=97 (accessed 6 February 2011).

Persival, D., and A. Walden. 2000. *Wavelet Methods for Time Series Analysis*. Cambridge, UK: Cambridge University Press.

Petrov, V., A. Kryuchin, S. Shanoylo, I. Ryabokon, and O. Onyshchenko. 1997. Optomechanical method of Edison cylinders sound reproduction. *Proceedings of Convention of Audio Engineering Society*: Preprint 4491.

Poliak, J., P. Robert, and J. Goy. 1992. Optical fibre turntable for archive records. *Proceedings of Convention of the Audio Engineering Society*: Preprint 3239.

Prism Sound Limited. 2011. The ADA8XR multi-channel modular AD/DA converter. http://www.prismsound.com/music_recording/products_subs/ada8xr/ada8xr_home.php (accessed 6 February 2011).

Rabinow, J., and A. Morse. 1961. Record player using light transducer and servo. *US Patent* 3,138,669, filed 6 June 1961, and issued 23 June 1964.

Rabinow, J., and W. Fischer. 1963. Stereo record player using optical means separate from the tone arm. *US Patent* 3,325,603, filed 20 November 1963, and issued 13 June 1967.

Rubinstein, R., and D. Kroese. 2008. *Simulation and the Monte Carlo Method*. 2nd ed. Hoboken, NJ: John Wiley & Sons, Inc.

Sage, G. 2010. Early recorded sounds and wax cylinders. http://www.tinfoil.com/cy-b.jpg (accessed 6 February 2011).

Savitzky, A., and M. Golay. 1964. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry* 36 (8): 1627–39.

Shapiro, L., and G. Stockman. 2001. *Computer Vision*. Upper Saddle River, NJ: Prentice-Hall.

Smithsonian National Museum of American History. 2010. Recording devices in Smithsonian National Museum of American History. http://history.sandiego.edu/gen/recording/nonelectrical (accessed 6 February 2011).

Stauderman, S. 2003. Pictorial guide to sound recording media. *Sound Savings: Preserving Audio Collections*: 29–42.

STIL. Line sensor. http://www.stilsa.com/EN/prod/chr/mpls_line.htm (accessed 5 May 2011).

Stoddard, R. 1986. Optical turntable system with reflected spot position detection. *US Patent* 4,870,631, filed 30 May 1986, and issued 26 September 1989.

Stoddard, R., and R. Stark. 1988. Dual beam optical turntable. *US Patent* 4,972,344, filed 4 November 1988, and issued 20 November 1990.

Stotzer, S., O. Johnsen, F. Bapst, C. Sudan, and R. Ingold. 2004. Phonographic sound extraction using image and signal processing. *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing* 4: 289–92.

Stotzer, S. 2006. Phonographic record sound extraction by image processing. PhD diss., University of Fribourg.

Szeliski, R. 2006. Image alignment and stitching: A tutorial. *Microsoft Research Technical Report* MSR-TR-2004-92.

Tanaka, S., I. Matsuda, N. Hirose, and K. Kobayashi. 1980. Optical signal reproducing apparatus. *US Patent* 4,504,934, filed 2 May 1980, and issued 12 March 1985.

Tian, B. 2006. *Reproduction of sound signal from gramophone records using 3D scene reconstruction*. PhD diss., The University of Western Ontario.

Tian, B., and J. Barron. 2007. Sound from gramophone record groove surface orientation. *Proceedings of IEEE International Conference on Image Processing* 6: 541–4.

U.S. Congress. House. *US National Recording Preservation Act of 2000*. HR4846. Public Law 106–474, 106th Congress (9 November 2000).

Veeco Instruments Inc. 2006. *Wyko NT8000 optical profiling system data sheet*. Tucson, AZ: Veeco Instruments, Inc.

Veeco Metrology Group. 2003. *WYKO NT8000 setup and operation guide*. Tucson, AZ: Veeco Instruments, Inc.

Vinyl Records Collector. 2010. Frequency Response of Acetate Records. http://www.vinylrecordscollector.co.uk/text/vinylhist10a.html (accessed 6 February 2011).

VPI Industries, Inc. 2011. The HRX Turntable. http://www.vpiindustries.com/static.php?page=HRX_Turntable (accessed 6 February 2011).

Wikipedia. 2011. 4-connectivity. http://upload.wikimedia.org/wikipedia/commons/4/49/Square_4_connectivity.png (accessed 5 May 2011).

Wikipedia. 2011. 8-connectivity. http://upload.wikimedia.org/wikipedia/commons/5/5a/Square_8_connectivity.png (accessed 5 May 2011).

Wikipedia. 2011. A Polish made Unitra turntable atop an Electromureş (Unitra-Diora) receiver, circa 1979. http://upload.wikimedia.org/wikipedia/commons/8/8f/Romanian_pickup1.jpg (accessed 5 May 2011)

Wikipedia. 2011. A typical LP, showing its center label. http://upload.wikimedia.org/wikipedia/commons/8/8e/Vynil_vinil_92837841.png (accessed 5 May 2011)

Wikipedia. 2011. Analog recording vs. digital recording. http://en.wikipedia.org/wiki/Analog_recording_vs._digital_recording (accessed 6 February 2011).

Wikipedia. 2011. Connected region extraction. http://upload.wikimedia.org/wikipedia/commons/a/ae/Screenshot-Figure_1.png (accessed 5 May 2011).

Wikipedia. 2011. Edison and phonograph. http://upload.wikimedia.org/wikipedia/commons/0/03/Edison_and_phonograph_edit1.jpg (accessed 6 February 2011).

Wikipedia. 2011. Edison Home Phonograph, Suitcase-Modell. http://upload.wikimedia.org/wikipedia/commons/a/a0/EdisonPhonograph.jpg (accessed 5 May 2011)

Wikipedia. 2011. Interference (wave propagation). http://upload.wikimedia.org/wikipedia/commons/5/5d/Michelson_Interferometer_Green_Laser_Interference.jpg (accessed 6 February 2011).

Wikipedia. 2011. Interferometer. http://upload.wikimedia.org/wikipedia/commons/d/d7/Interferometer.png (accessed 6 February 2011).

Wikipedia. 2011. Pixel region. http://upload.wikimedia.org/wikipedia/commons/5/58/Screenshot-Pixel_Region_%28Figure_3%29.png (accessed 5 May 2011).

Wikipedia. 2011. Principle of confocal microscope. http://upload.wikimedia.org/wikipedia/commons/d/dc/Confocalprinciple_in_English.svg (accessed 5 May 2011).

Wikipedia. 2011. Typical phonograph tonearm. http://upload.wikimedia.org/wikipedia/commons/9/9c/Romanian_pickup2.jpg. (accessed 5 May 2011).

Williams, W. 1930. Reproduction of sound records. *US Patent* 1,917,003, filed 20 January 1930, and issued 4 July 1934.

Wyant, J. 1974. White light extended source shearing interferometer. *Applied Optics* 13: 200–2.

Ziou, D., and S. Tabbone. 1998. Edge detection techniques: An overview. *International Journal of Pattern Recognition and Image Analysis* 8 (4): 537–59.

Discography

Bernstein, L., dir. *Symphony no. 6 in A minor*, by Gustav Mahler. New York Philharmonic. CBS IM 77215.

Bernstein, L., dir. *Symphony no. 6 in A minor*, by Gustav Mahler. New York Philharmonic. SMK 60208–AC.

Williams C., and B. Wolifson. *Ultimate analog test LP*. Analogue Productions. Disc phonograph record.