UNIVERSITY OF SOUTHAMPTON

FACULTY OF ENGINEERING, SCIENCE & MATHEMATICS

SCHOOL OF ENGINEERING SCIENCES

Methods For Signal Reproduction From Mechanical Sound Carriers, Measured Via Non-Contact, Full Surface Mapping

by

Antony James Nascè

Thesis for the degree of Doctor of Philosophy

December 1st 2009

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING, SCIENCE & MATHEMATICS
SCHOOL OF ENGINEERING SCIENCES

Doctor of Philosophy

METHODS FOR SIGNAL REPRODUCTION FROM MECHANICAL SOUND CARRIERS, MEASURED VIA NON-CONTACT, FULL SURFACE MAPPING

by Antony James Nascè

Archivists are often faced with a dilemma when transferring audio from early mechanical sound carriers to digital formats. Traditional contacting stylus methods are not always appropriate for recordings such as wax cylinders or coarse groove discs. The recordings in question may be broken, incompatible with replay systems, or too historically significant to undergo stylus forces, due to the risk of causing irreversible damage.

A non-contact surface metrology technique has been developed which enables the full topology of grooved recordings to be stored as a digital representation of surface heights. From this data, it is possible to reproduce the audio signal encoded in the grooves without any risk of damaging the artefact. This method has great potential for archivists aiming to digitise content from recordings that may otherwise be 'unplayable' with a stylus.

This thesis presents the first detailed study into signal reproduction methods from the non-contact full surface mapping method, developed for cylinder and disc media. The methods developed rely on parameter estimation theory, provided by a Fourier analysis of the measured groove profiles. Savitzky-Golay, polynomial smoothing filters are also used throughout, and are shown to be well suited to enhancing the medial valley structure of the record grooves.

Two methods of stylus trajectory estimation for cylinders have been developed, to enable tracking of specimens in both good and poor surface condition. Trajectory estimation for cylinders in good condition relies on a global estimation of the groove's shift. For damaged recordings, (or for those where the track is poorly defined), a method has been developed based on the tracking of local groove minima.

A number of different methods for estimating the depth signal from a discrete groove cross-section are evaluated. In order to quantify signal reproduction, a test cylinder encoded with sinusoids was produced allowing for signal quality metrics (signal-to-noise ratio and total harmonic distortion) to be calculated and compared with conventional stylus replay.

The sampling and resolution of the measurement system are considered with respect to the requirements for digital archiving of cylinder recordings. Methods are also described for overcoming distortions introduced by the measurement process. Preliminary results for reproduction from discs and tinfoil measurements are also presented. Recommendations for future work into non-contact methods of sound reproductions are given.

Contents

Declaration of Authorship

I, Antony James Nascè, declare that the thesis entitled, **'Methods For Signal Reproduction From Mechanical Sound Carriers, Measured Via Non-Contact, Full Surface Mapping'**, and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- none of this work has been published before submission.

Signed: ..  .....

Date: 30/11/09

Acknowledgements

Firstly, I would like to thank my supervision team - Prof. Martyn Hill and Prof. John W. McBride. They have had to endure my obsession with getting distracted by matters less pressing, and my eventual decision to move to London. For their support and patience I am truly grateful.

Secondly, the reason that I moved to London – my darling fianceé Sophie, whose time management and organisational skills I hope one day will transfer to me by osmosis. You have been loving and supportive throughout and make an excellent banana muffin.

Thirdly, a big thank you to my parents who have supported me throughout the PhD rollercoaster, and have always given me the motivation and guidance to keep on going in times of doubt.

Lastly, a big thank you to the EPSRC, for funding this exciting research. The British Library Sound Archive, the Welsh Sound Archive, and Adrian Tuddenham of Poppy Records, for providing sound artefacts and knowledge.

Nomenclature

$\Delta\theta$ – circumferential grid spacing.

Δx - linescan grid spacing.

Δz – axial resolution of sensor.

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

$z_j'(x)$ – Savitzky-Golay filtered linescan.

$\tilde{z}_j(k)$- Discrete Fourier transform of the j-th linescan sequence.

$Z(x_i, \theta_j)$ – Discrete surface measurement.

$Z_n(x_i^n, \theta_j^n)$ – Discrete surface patch.

$Z'(x_i, \theta_j)$ – Savitzky-Golay filtered surface.

λ_x – groove pitch (inter-groove spacing)

g_k – groove seed vector, containing x-index of the k-th groove along the x-axis

τ – groove shift estimation vector.

A_n – stylus trajectory estimate, composed of x-θ point pairs

M(x) – average groove shape model

$z_a(n)$ – the radial displacement track formed by the cutting tool.

$I(h,w)$ – grayscale image

ρ - centre of mass quantity for grayscale image tracking

Acronyms

SAP – Sound Archive Project

NCFSM – Non-contact, full surface mapping

LBNL – Lawrence Berkeley National Laboratory

BLSA – British Library Sound Archive

SG – Savitzky-Golay

FIR – Finite Impulse Response

IIR – Infinite Impulse Response

DFT – Discrete Fourier Transform

SNR – Signal-to-noise Ratio

THD – Total Harmonic Distortion

STFT – Short-time Fourier Transform

nint – Nearest integer

Chapter 1

Introduction

1.1 Background

Sound recording and reproduction by acousto-mechanical means was first made practicable in 1877, with the advent of Thomas Edison's Phonograph[†] [1]. These recordings were produced by channelling acoustic pressure variations in air towards a diaphragm, which in turn drove a recording stylus. The stylus moved in synchrony with the incident acoustic pressure, embossing 'hill-and-dale' indentations onto a revolving cylinder, wrapped with a layer of tinfoil. Sound reproduction from the recorded surface was achieved by tracing over the physical indentations with a contacting stylus. Upon playback, surface undulations of the embossed tinfoil were then transmitted to a diaphragm-horn arrangement, into re-radiated acoustic pressure.

Edison's vision of being able to "store and reproduce automatically at any future time the human voice perfectly" [2] was initially marred by low quality sound reproduction, provided by the tinfoil. The fidelity and durability of early tinfoil recordings were improved upon by cylinders made of wax, and later celluloid. Phonograph records gained immense popularity with the public during the late 19^{th} and early 20^{th} century. As well as giving birth to the popular music industry, the phonograph was used to record political speeches, and provided a valuable tool for office dictation, linguistics and education.

[†] The term 'phonograph' is often used in modern language to denote any early acoustic recording, including cylinders and disks. In this thesis, the use of 'phonograph' will refer to Edison's original cylinder technology.

By 1929, mass production of cylinders ceased, and was superseded by superior flat disc technologies, boasting longer playback times and eventually, stereophonic sound. Cylinders were however used in scientific environments and for office dictation up until the 1960's [3]. Disc recording technology continued to develop throughout the 20^{th} century and today, the modern LP record remains a popular audio format. Nowadays, phonograph cylinders and pre-stereophonic discs are considered obsolete formats and remain confined to the vaults of museums and sound archives.

Many of these early recordings hold considerable historical and ethnographic importance to our society, but their life expectancy is finite. If improperly stored, the surface condition will deteriorate over time. Factors such as mould growth, creep deformation and cracking can destroy or distort the original audio content. Modern sound archives have therefore sought to preserve these artefacts through careful storage and transferral of the sound onto more stable, digital formats. The British Library Sound Archive [4] for example holds a collection of over 3,000 ethnographic wax cylinders, dating from 1898 to 1915, of which around 200 recordings have been made available for public access via the Internet.

There remains however a vast amount of content worldwide which has not yet been digitised. In some cases, these recordings are considered 'unplayable' via conventional stylus methods. This may be because the recording is warped, or broken, and no longer complies with the playback apparatus. In other cases a recording may be deemed too fragile to withstand the pressure of a contacting stylus.

There is therefore a growing need for alternative sound reproduction methods, which are capable of preserving the surface and audio content of sound carriers which are 'unplayable' via conventional stylus methods. The preservation and playback of these recordings via non-contact, non-destructive methods provides the motivation for this research.

1.2 The Sound Archive Project

The work described in this thesis is carried out under the Sound Archive Project (SAP). The SAP began in March 2005 and is funded by the Engineering and Physical

Sciences Research Council (EPSRC), to carry out research into 'Non-Contact Surface Scanning Systems for the Retrieval and Protection of Archived Sound Recordings' (Ref: EP/C00857X/1). Research has been undertaken in collaboration between the University of Southampton, TaiCaan Technologies Ltd, British Library Sound Archive and the Lawrence Berkeley National Laboratory.

The aim of the project is to investigate methods for the non-contact scanning of archived sound recordings, such that the surface and encoded sound can be preserved in a digital format. The proposal is timely and critical, since at present, many sound carriers such as early wax cylinders, are archivally unstable and at risk of deterioration. A non-contact method of sound reproduction ensures that no pressure is exerted on the recorded surface, meaning that no further damage is caused during playback.

The SAP uses a method of non-contact sound recovery, based on mapping the full surface topology of early mechanical sound carriers, followed by audio signal reproduction in post-processing. Throughout this thesis, this method of measurement shall be referred to as the non-contact full surface mapping (NCFSM) method. The philosophy behind this technique is preservation of the full recorded surface, not real-time playback. The SAP is a multi-disciplinary project focusing on the following research topics:

i. **Non-Contact Surface Measurement** - development of metrology systems for mapping the 3D surface topology of cylinders and flat discs.
ii. **Sensor Development** - design of optical sensors with improved angular tolerance and sensing speed for measurement of grooved surfaces.
iii. **Signal Reproduction** – development of methods for sound reproduction from discrete surface maps of cylinders and discs measured by the NCFSM method.

The research described in this thesis focuses on (iii) - the development of audio signal reproduction methods from the non-contact, full surface mapping (NCFSM) method.

1.3 Research Objectives

The aim of this research is to develop methods for accurate signal reproduction from non-contact surface measurements of early acoustic recordings, including cylinders and flat discs. More formally, the research objectives are:

(i) To develop robust, automated (or semi-automated) sound extraction algorithms for accurate sound reproduction from 3D surface measurements of early acoustic cylinder and flat disc recordings of various type and surface condition.

(ii) To ensure that the techniques developed in (i) do not introduce uncharacteristic noise or distortion, which is misrepresentative of the original encoded sound.

(iii) To develop methods for reproducing sound from partial surface scans (segments), which will enable the full recovery of sound from broken recordings.

(iv) To quantitatively assess and compare the audio quality achievable from both non-contact and stylus-based sound reproduction methods.

1.4 Research Contributions

This thesis presents the first in-depth study for signal reproduction from early mechanical sound recordings, measured by the non-contact, full surface mapping method (NCFSM).

Signal reproduction from a NCFSM method for cylinders in good surface condition was first described in 2005 by Fadeyev et al. [5]. In this thesis, the ideas introduced in [5] are explored in more detail for cylinders of various types and surface condition.

An analysis of the noise generation processes associated with measurement and signal estimation is given. This is carried out in order to understand the limitations of the NCFSM method. Resolution requirements for measurement are also considered with

respect to archival standards for digitisation of early mechanical sound carriers. Conclusions are drawn regarding the advantages and disadvantages of non-contact methods, over conventional stylus methods.

In order to investigate different audio signal estimates derived from the same surface data, a cylinder recording encoded with test signals was specially produced. This represents the first example of quantitatively assessing and comparing non-contact signal reproduction from cylinder recordings.

Two methods of tracking vertically cut grooves from discrete surface data have been developed. These methods are suitable for cylinders in good and poor surface condition. The methods developed for cylinders can also be applied to signal reproduction from tinfoil recordings and discs.

The majority of research described in this thesis concerns cylinder recordings, as this has been the major focus of the Sound Archive Project up until time of writing. Flat disc recordings are considered, but not in as much detail. Recommendations are also made for further research into signal reproduction from the NCFSM method.

1.5 Thesis Outline

The thesis is outlined as follows. In Chapter, 2 a review of related literature is given. This starts with a description of the main subject matter – mechanical sound carriers. Sound reproduction methods are then described, including stylus, and non-contact methods. Next, the sensor technology investigated by the SAP is described. Finally, a summary of relevant techniques for signal reproduction from discrete grooved surfaces is presented.

In Chapter 3, the two systems developed for non-contact, full surface mapping of cylinder and disc recordings is described. The resolution requirements for accurate signal production are considered here. The limitations of the NCFSM method, and typical noise and distortion are described.

Chapter 4 represents the main contributions of this thesis – proposed signal reproduction methods from NCFSM data. A consideration of how the audio signal is encoded on discrete surface data of cylinders and discs is described. A logical

progression of methods developed from raw surface data import to extracted audio signal is then given. A consideration of recordings in varying surface condition is given. The majority of this chapter concerns cylinder recordings, however methods for disc recordings are also considered.

Chapter 5 presents results and evaluation of signal reproduction methods described in the previous chapter. A testing method of signal quality is presented, using a test cylinder recording. Results from cylinders in poor surface condition are also presented and where possible, comparisons are made with stylus replay. Preliminary results are also given from an early tinfoil Phonograph.

Chapter 6 draws conclusions about the work presented in this thesis. In particular, the feasibility of the NCFSM method for surface preservation and signal recovery is described, with reference to archival standards. The final section in this chapter concludes with suggestions for further investigation into signal reproduction for the Sound Archive Project.

Reference cylinder surfaces measured by the NCFSM method are referred to throughout this thesis by reference codes, which can found in Appendix A.

Figure 1.1: The Sound Archive project aims to develop methods for digital preservation of early mechanical sound carriers, such as the wax cylinder depicted here.

Chapter 2

Literature Review

Research into non-contact, *surface-preserving* methods for sound reproduction from early grooved sound carriers began in 2003, with a grayscale imaging method developed for disc records [6]. In 2005, a non-contact full surface mapping (NCFSM) method for signal reproduction was introduced [5], which utilised an optical metrology system to measure the 3-D topology of cylinder recordings. As a relatively recent research topic, literature relating *directly* to signal reproduction from NCFSM methods of sound reproduction is limited. Nevertheless, it is the intention of this chapter to provide a summary of literature relating to the techniques developed in this thesis.

In order to familiarise the reader with mechanical sound carriers, an overview of sound recording technology from the 'acoustic era' of recorded sound (1877-1929) is first presented, including cylinders and discs. Sound reproduction from these recordings is then described, including both contacting and non-contact methods. Non-contact methods are divided into three categories: optical reproduction *in situ*, grayscale imaging methods, and NCFSM – the latter of which is most applicable to this work. One aim for this thesis is to assess the quality of signal reproduction from the NCFSM method; hence we look at how this has been accomplished in the literature for other non-contact methods.

Finally, a review of competing optical sensor technologies for high resolution, 3-D surface measurement is then given. This section details the operating principles, measurement parameters, and limitations of three competing sensor technologies – white light (WL) confocal, laser triangulation (LT) and laser confocal (CL).

Conclusions are drawn as to which sensor is most suitable for measurement of recorded surfaces for the Sound Archive Project.

2.1 Cylinder Recordings

Cylinder recordings represent the earliest mechanical means for storing and reproducing sound. For accurate signal reproduction, it is important to have an understanding of the encoding principles of these early mechanical sound carriers. Recording characteristics and groove parameters must be understood, so that non-contact surface measurements can be carried out to sufficient resolution. This section gives a brief history of cylinder recording technology, including the major historical advances and recording principles. For the interested reader, detailed accounts of the acoustic era of recorded sound are given by Millard [7], or Read & Welch [8].

2.1.1 Tinfoil Phonograph

Thomas Alva Edison produced the first successful sound recording† and reproduction device in 1877 [1], in the form of the tinfoil phonograph (see Fig. 2.1). The original phonograph embossed vertical undulations, produced by acoustic pressure variations in the air, upon a tinfoil sheet wrapped around a rotating cylinder drum (see Fig 2.2). The machine was originally driven by a hand crank, thus maintaining a constant playback speed proved to be problematic. Technical data on tinfoil machines is limited, but typical recording speeds were around 60 rpm [10]. Tinfoil phonograph apparatus varied in size but typically the cylinder drum had a diameter of 4" (100mm) and length 80mm. The pitch of the helical groove, (in which the sound was encoded), was coarse compared with later cylinders, with a nominal pitch of 10 turns per inch (tpi) [10]. This meant that recording duration was generally less than one minute.

Thin tinfoil sheets meant that poor longevity was also one the early phonograph's failings. Tinfoil recordings could only be played back a few times before they became incomprehensible, due to wear. As a result, there are very few archived examples of sound transfers made from Edison's original tinfoil phonograph.

† Édouard-Léon Scott's 'Phonoautograph', patented in 1857 [9] was the first sound recording device. It consisted of a horn-membrane arrangement, with a stylus that made a trace on a lamp-blacked surface. Incident sound caused the membrane to vibrate a stylus, producing a visual representation of the sound. It was originally intended for visualising sound, not playback.

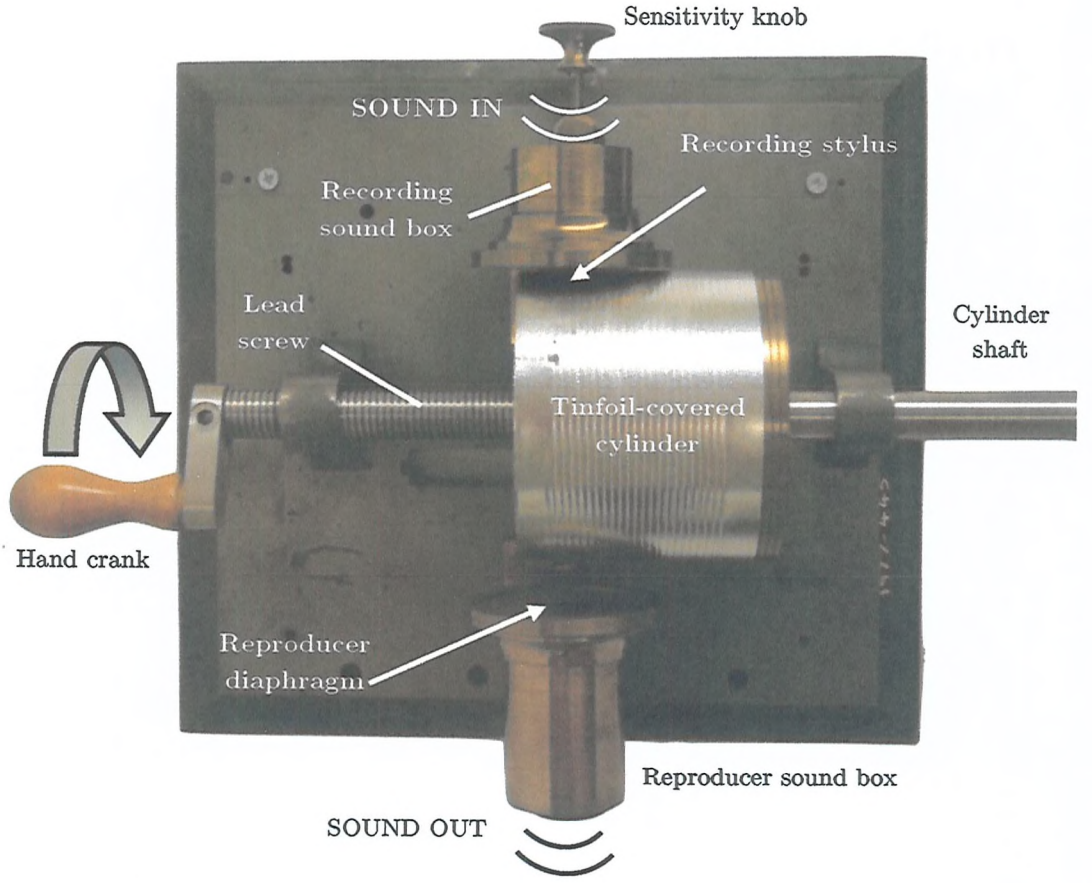

Figure 2.1: Schematic of Edison's original tinfoil phonograph apparatus, invented in 1877 (Image is of a phonograph replica, held at the British Museum).

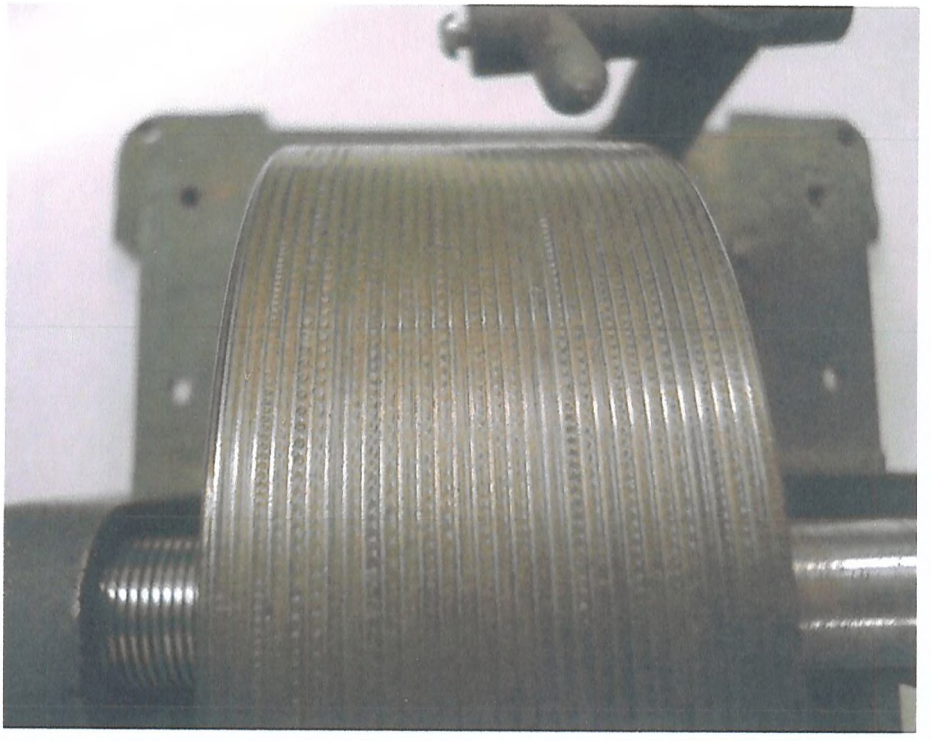

Figure 2.2: Examples of Sound embossed onto a sheet of tinfoil, mounted onto an Edison phonograph produced in 1879. Vertical groove modulations are clearly visible. (Image from [12]).

Examples of tinfoil recordings do exist in modern archives, but attempts to transfer the sound from original artefacts are seldom made, due to a lack of suitable playback apparatus and risk of causing irreversible damage by stylus playback.

2.1.2 Wax Cylinders

In 1881, Charles Sumner Tainter, Alexander Graham Bell, and Chichester Bell developed a cylinder recorder that used wax instead of tinfoil. Wax brought greater malleability and resistance to tearing; hence more playback cycles were possible.

Tainter's Graphophone [13] from 1885, (see Fig. 2.3a) recorded sound onto a paper cylinder, coated with a layer of wax. Playback duration could reach 10 minutes, with a pitch of 120 tpi – a vast improvement over tinfoil recordings. Later Graphophones used smaller cylinders, 6" (150mm) in length, 1 3/8" (35mm) diameter with a finer recording pitch of 150 tpi.

Another development made by Tainter was the speed governor, which ensured that the cylinder rotated at a constant surface velocity. This greatly reduced the effect of pitch modulation – an effect caused by variable surface speed – common with hand-cranked apparatus.

By 1887, advances in durability and longevity of the Graphophone cylinder meant that recordings and playback apparatus were commercially available to the public. The following year, Edison announced that he had abandoned tinfoil, in preference for cylinders composed entirely of wax. With Edison's 'Perfected Phonograph' (see Fig. 2.3b), the same blank wax cylinder could be shaved and re-recorded for "fifteen or twenty successive records", and once recorded, could be replayed "thousands of times with undiminished clearness" [14]. Edison had effectively developed the first 're-writable' sound medium. Hand-cranked motors of his early phonograph were replaced with an electrical motor, which ensured a constant playback speed. Early wax cylinder recording speeds varied; slower speeds such as 90 rpm, used often for spoken material, would yield nearly a 4-minute recording, while faster speeds from 120 to 160 rpm (or more) for music would run for around 2 to 2½ minutes. Recording speeds varied until 1902, when a standard of 160 rpm was adopted.

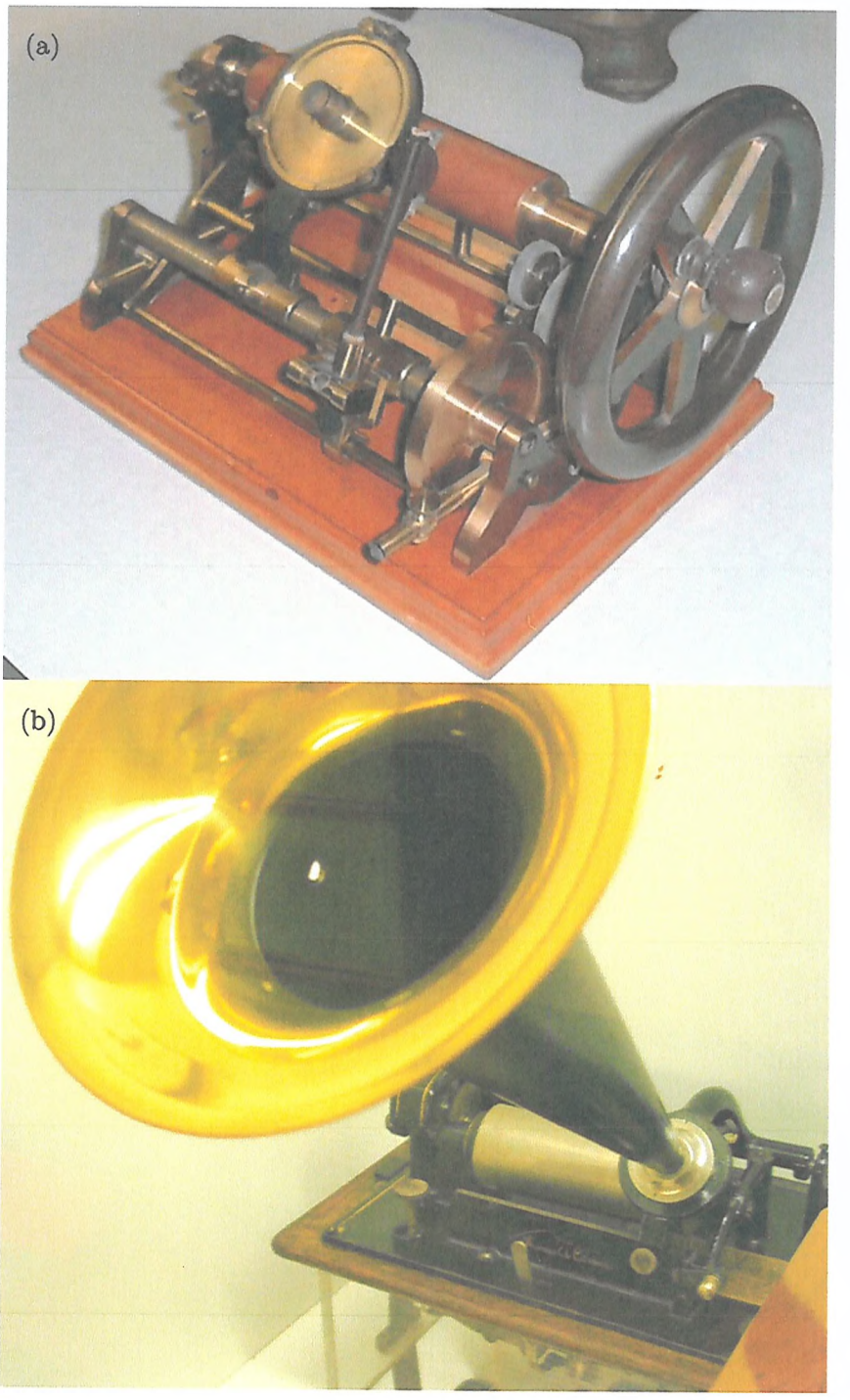

Figure 2.3: Early wax cylinder recording machines: (a) Graphophone apparatus from 1885, shown without horn attachment (image from [12]), and (b) Edison phonograph shown with horn (image taken at British Museum)

Edison's earliest wax cylinders (see Fig. 2.4) were composed of a mixture of beeswax, stearic wax and ceresin. The variability in composition of beeswax was later replaced with a standardised metallic soap composite, often referred to as 'brown wax'. Organic wax cylinders are susceptible to mould growth, which can attack and destroy the recorded groove structure. This means that the immediate transfer and preservation of wax cylinders is a priority for many sound archives.

Figure 2.4: A selection of two-minute wax cylinder recordings, 2" in diameter, 4" in length. In 1888, this became the standard size for Edison cylinders. (Image from [10])

2.1.3 Mass Produced Cylinders

In the late 1890s, Edison experimented with methods of mass-producing cylinders from a single matrix master recording. The introduction of the Edison Gold Moulded Record in 1902, meant that sales of cylinders rose from 4.38 million in 1902 and 7.66 million in 1903 [15]. By 1912, Edison's National Phonograph Company started to replace wax cylinders with celluloid. Celluloid is regarded as one of the first thermoplastics, composed of nitrocellulose and camphor. The main advantage of celluloid over wax is that it is not susceptible to mould growth. It is also more durable, and can be played back more times without risk of causing damage to the grooves. In cases of poor storage, celluloid cylinders can still develop cracks in the surface, caused by internal stresses built up over time, as volatile camphor gradually evaporates from the celluloid material, causing shrinkage [15]. Mass-production of cylinders ceased in 1929, but the

use of cylinders for dictation and archiving in the scientific community continued sparingly until the early 1960s [3].

Summary

A comparison of cylinder geometries and playback parameters is given in Figure 2.5 and a summary of important dates in the history of cylinder recording technology is shown in Figure 2.6.

Tinfoil recordings, dating back to 1877 are in most need of a non-contact method of reproduction, because modern stylus replay systems for tinfoils do not exist. The likelihood of a tinfoil existing in good surface condition is low, unless the recording has remained un-played to this day.

The majority of mass produced cylinders (1912-1929), have been successfully transferred via conventional stylus replay. Multiple copies of mass-produced cylinders exist, meaning that there is usually at least one copy in good condition, suitable for stylus transfer. Non-commercial wax cylinder recordings (i.e. not mass-produced) are most likely to benefit from a non-contact sound reproduction method, due to the risk of causing irreversible damage with a stylus.

Parameter	Tin-foil	Wax Cylinders	Celluloid Cylinders
Date of Production	1877-1880	1885-1911	1912-1929
Material	Tin-foil sheet	Wax	Celluloid
Cylinder Diameter (mm)	~100	35-127	60
Cylinder Length (mm)	~80	152 – 229	102
Playback Speed (rpm)	~60	90-160	160
Pitch (tpi)	10	100 – 150	100 – 200
Playback Duration (mins)	< 1	2 – 10	2 – 4
Recorded track length (m)	~10	66 – 540	75 -150
Groove surface area (m^2)	0.025	0.017 - 0.09	0.0192

Figure 2.5: Summary of Acoustic Cylinder Recordings (Collated from: [7-8,10-12,15]).

1877	Edison files patent [1] for Tinfoil Phonograph. The first successful recording and reproduction of the human voice.
1879	Edison halts further development of Phonograph in favour of the light bulb.
1880	Alexander Graham Bell publishes: "On the Production and Reproduction of Sound by Light", read before the American Association of Advancement of Science.
1881-84	Tainter, A.G Bell & Chichester Bell experiment with wax cylinder and disc technologies, to improve upon Edison's tinfoil phonograph. Speed governor produces recordings with constant recording/playback speed.
1885	C. Bell and Tainter file for first Graphophone patents [13] The Graphophone recorded sound onto a paper cylinder, coated in wax.
1887	Graphophone and cylinder recordings commercially available to public. Edison starts work on improved phonograph, to utilise wax cylinders.
1888	Edison announces his Perfected Phonograph, which uses wax cylinders and an electric motor.
1895	Edison announces spring motor phonograph. First widespread, commercially available cylinders in brown wax. Playback time is approximately 2 minutes.
1899	Edison Concert Phonograph introduced. Concert cylinders were over double the size of standard cylinders, and were suitable for public performance.
1903	Sales of Edison Gold Moulded cylinders reach 7.66 million.
1908	Amberol Record introduced. Cylinder playback time is 4 minutes.
1912	Edison introduces celluloid, Blue Amberol Cylinders offering greater durability.
1929	Edison ceases mass production of Blue Amberol cylinders.
1930 – 1960	Cylinder recordings used for scientific, office dictation purposes.

Figure 2.6: Historical developments in cylinder recording technology
(collated from: [7-8,10-12,15])

2.1.4 Recording Principles

Recording systems and cylinder technology developed between 1877 and 1929, but the basic recording principles remained the same (see Fig. 2.7). In recording mode, a cylinder 'blank' is set to rotate with constant angular velocity, $\dot{\theta}$. A cutting tool (stylus), which is coupled to a diaphragm-horn arrangement, advances along the cylinder axis (x-axis), via a lead screw with constant pitch, λ_x. Prior to recording, the cutting stylus is forced into the outer cylinder surface, producing an unmodulated cut

of fixed radial depth. During recording, incident acoustic pressure $p(t)$, is channelled via a mouthpiece or horn from the surrounding air, towards a thin diaphragm, forcing the stylus to cut undulations which are normal to the surface, (in the radial plane), about its mean cut depth. This radial modulation, $z_a(t)$ is often referred to as "hill-and-dale" or "vertical" (see Fig. 2.8). As the cutting stylus vibrates in the radial plane at greater amplitude, the groove walls open out more (a-a'). As the cutter rises, the groove wall gets narrower (b-b'). The encoded groove follows a helix of pitch λ_w, with cross-section dependent on the stylus tip geometry.

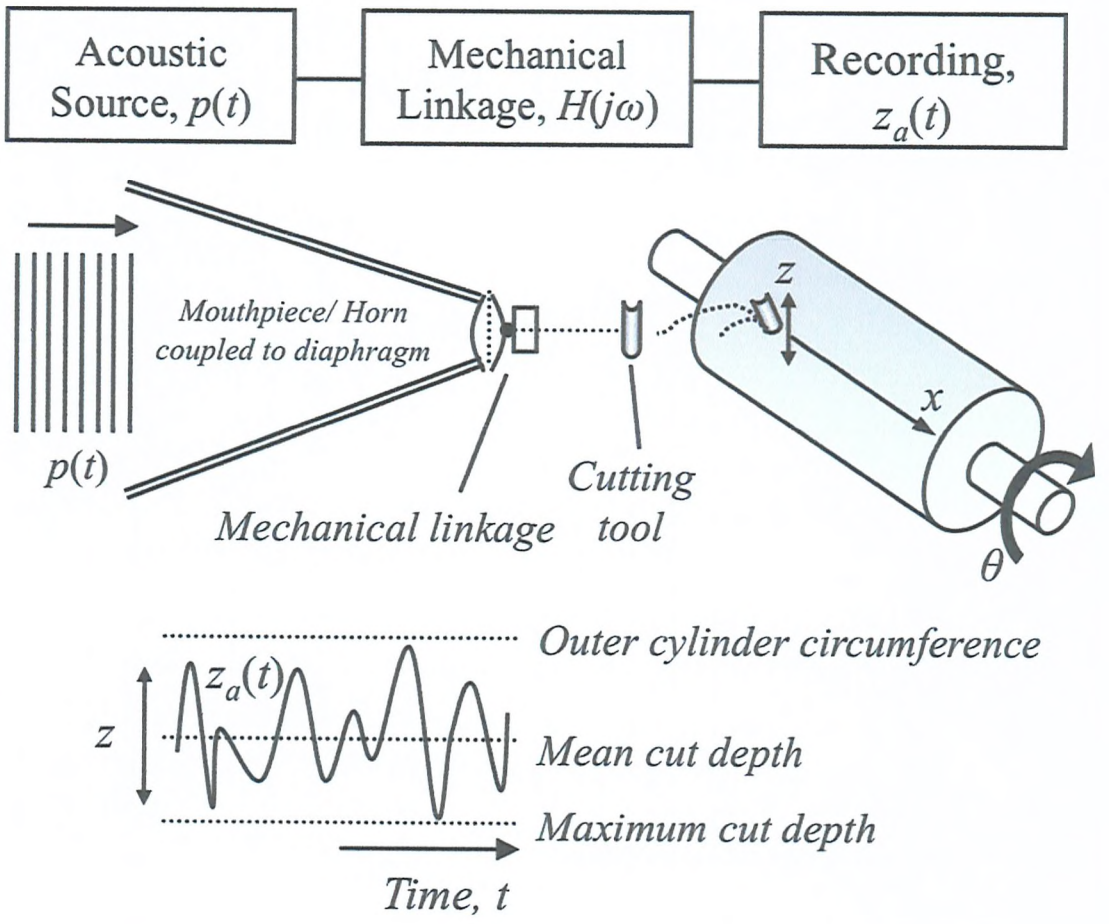

Figure 2.7: Recording principle for vertically cut groove cylinders.

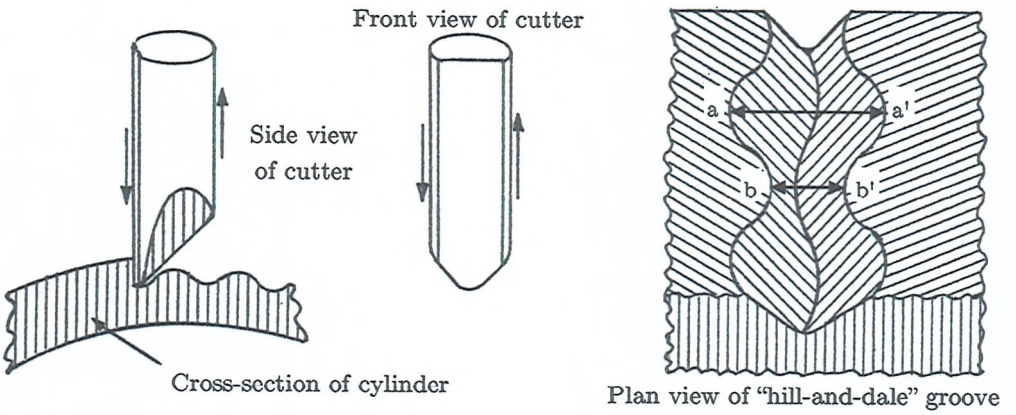

Figure 2.8: The 'hill-and-dale' recording principle used for cylinder recordings (Image from [16]).

2.1.5 Recording Characteristics

Modulation Depth

For vertical cut grooves, there is a limit to how deep the cutting tool can travel into the surface before affecting the adjacent groove ('inter-cutting'). The maximum depth of the modulated groove, $d_{\max}$ prior to inter-cutting is limited by groove pitch, λ_x, and the radius of curvature of the cutting tool. For a vertically modulated groove, this relationship can be calculated if one assumes that the cutting tool cross-section is a circle segment, with radius r_{stylus} (see Fig. 2.9), by:

$$d_{\max} = r_{stylus} - \sqrt{r_{stylus}^{\,2} + \left(\frac{\lambda_x}{2}\right)^2}$$

Figure 2.10 gives example calculations of maximum cut depths for two typical stylus tip radii [17], at two groove pitches. These maximum allowable cut depths range from 6.8 to 102.2 μm and represent the absolute maximum range of groove modulation, from peak-to-trough (See Figure 2.11). From these calculations, it can be seen that cylinder groove depth is typically on a micron scale.

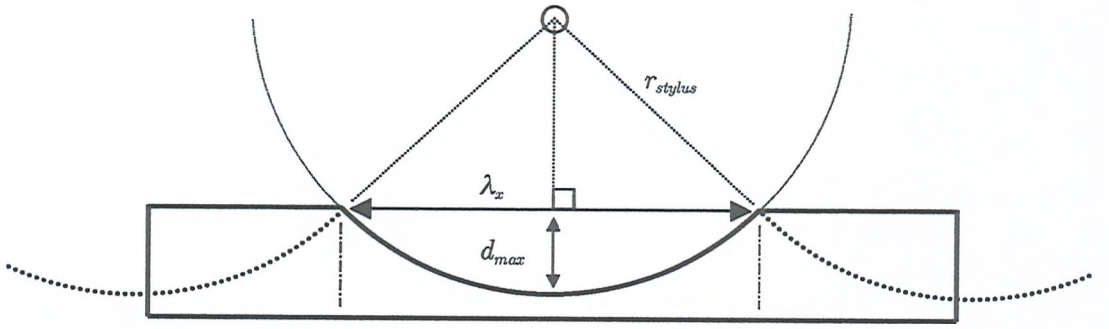

Figure 2.9: Consideration of groove cross-section cut by spherical stylus tip with radius, r_{stylus}, for calculating maximum allowable groove depth, d_{max}, for vertical cut grooves.

Stylus tip radius, r_{stylus}	d_{max}, depth limit before inter-cutting occurs (μm)	
	100 tpi	**200 tpi**
130 μm	102.2	**16.6**
300 μm	28.2	**6.8**

Figure 2.10: Comparison of maximum intercutting depth for two stylus tip radii and groove pitches.

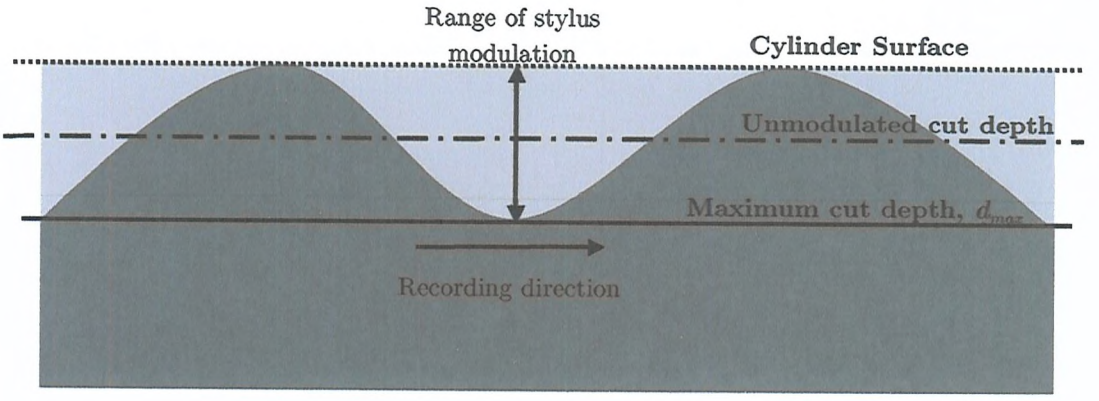

Figure 2.11: Longitudinal section of modulated cylinder groove, demonstrating the maximum range of stylus motion, given by d_{max}. This distance ranges between 6.8 to 102 μm, for cylinder recordings.

Frequency Response

As cylinder recording technology developed, it became apparent that the recording horn played a vital role in improving the frequency response. The radius of the recording diaphragm r_D, is typically very small, and for acoustic wavelengths of $\lambda > 2\pi r_D$, the diaphragm is not an efficient transducer. There is said to be an impedance 'mismatch' between the large volume of surrounding air (low impedance), and the small cross-section of the stiff diaphragm (high impedance). The horn acts as an 'impedance matcher', matching the low impedance of surrounding air at the mouth-end, to the high impedance of the diaphragm, at the throat of the horn. A smooth reduction in cross-sectional area towards the horn's throat helps to couple the diaphragm to the surrounding air, thus increasing the effectiveness of sound transmission and radiation in recording and playback. A detailed description of acoustical impedance and horn theory, with reference to the phonograph is given by Webster [18] and Wilson [19].

The characteristic impedance of the horn and mechanical linkage introduces a complex transfer function, $H(j\omega)$ to the signal chain, between acoustic pressure $p(t)$ and the encoded radial modulation, $z_a(t)$. The result of which meant that recordings produced by acoustic means were typically band limited between 150 Hz and 5 kHz. Figure 2.12 shows a typical frequency response of a mechanical phonograph discussed and analysed by Maxfield and Harrison [20].

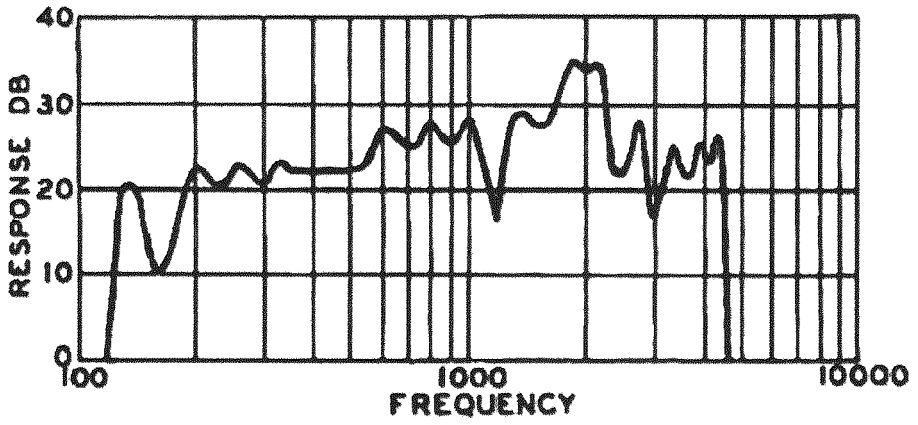

Figure 2.12: Typical frequency response of a mechanical phonograph, showing how the range of frequencies for acoustic recordings was limited between 100 Hz and 5 kHz. (Discussed by Maxfield and Harrison, [20] and re-plotted here by Olson [21])

Equalisation

There are essentially two different recording modes for storing a signal in a modulated groove: constant displacement amplitude and constant velocity. For a recording made with constant signal level, a constant amplitude recording encodes the same groove amplitude over the full frequency range, thus stylus velocity increases with frequency. A signal encoded with constant velocity maintains a constant stylus velocity at all frequencies, thus a doubling of frequency means that the peak groove displacement of the recorded tone is halved, at the rate of 6dB per octave (see Fig. 2.13).

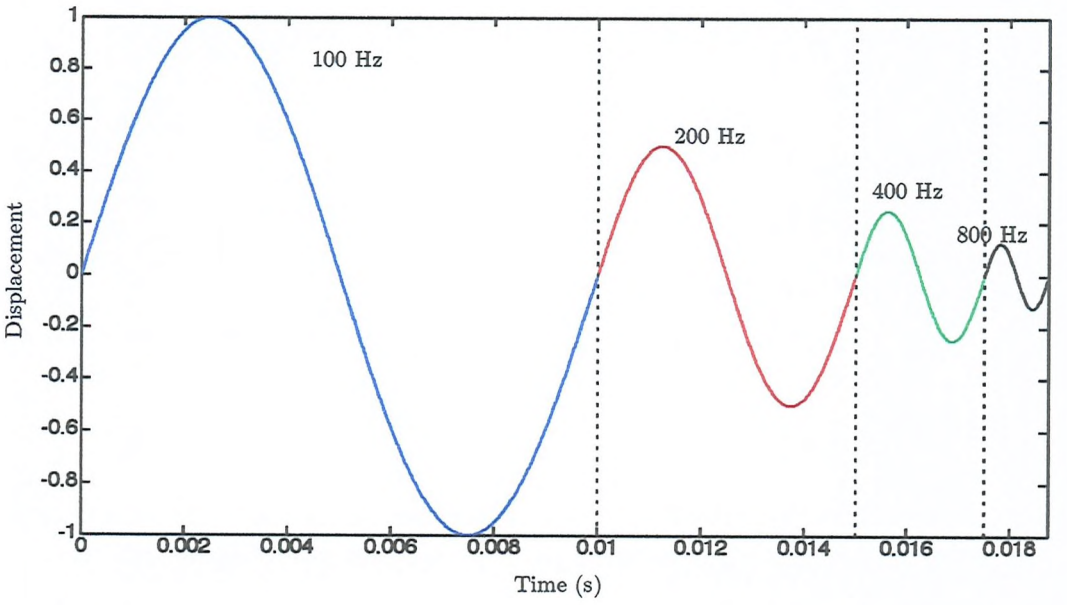

Figure 2.13: Example of a constant-velocity waveform. A doubling in frequency produces a halving in displacement.

For grooved recordings, the audio signal can therefore be considered as the groove position (constant amplitude), or the first time-derivative of groove position (constant velocity). Each recording characteristic has its limitations, for example in constant amplitude mode at high frequency, the radius of groove curvature decreases. There is therefore an upper frequency limit during recording, above which the radius of the groove will become smaller than the cutting tip, which can lead to distortion on playback. Similarly for constant velocity recordings, at low frequencies (larger groove amplitudes), there is a limit at which groove excursions will impede adjacent grooves

('inter-cutting'), which adds a constraint to the minimum allowable groove pitch of a recording.

Over a limited frequency range (~150 Hz - 4kHz) acoustic recordings are said to exhibit a constant-velocity characteristic [22]. The implication of this is that higher frequencies have very low displacement amplitudes, and can become masked by surface noise.

Through a non-contact depth measurement of the recorded surface, the radial displacement position, $z_a(t)$ produced by the cutting stylus can be estimated. To obtain a signal that is proportional to the acoustic pressure recorded, the first derivative of this vector should be taken to obtain a signal that is equivalent to the radial velocity of the stylus. For a continuous harmonic signal of the form: $z_a = z_a e^{j\omega t}$, the first time derivative can be obtained, simply by multiplying by $j\omega$. For discrete signals, the numerical gradient at a point can be obtained via a finite difference operation. Different numerical differentiation methods (e.g. forward/backward/central difference), will introduce a unique frequency response, based on its z-transform. It is therefore suggested that some discrete filtering operation, incorporating both the differentiation stage and desired equalisation is applied to z_a. This approach has been implemented in previous work by Fadeyev et al. via FFT filtering [5].

Due to the complex transfer function of the horn, mechanical linkage, and surface roughness, an optimal equalisation will vary from cylinder to cylinder. It is therefore suggested here, that any final equalisation is left to the expertise of the transfer engineer or sound archivist. Regardless of which filtering method is used (prior to public access), the most important stage in the playback chain following surface measurement, is to obtain the best estimate for the displacement track, z_a.

2.1.6 Noise and Distortion

Cylinder records are subject to many sources of noise and distortion, during the recording and playback process. Noise, in the context of audio signals can be defined as any unwanted signal that interferes with the communication, measurement or processing of an information-bearing signal [23]. This section considers some of the

inherent sources of noise and distortion, relating to cylinders. A consideration of these problems is important for development of a scanning strategy for the NCFSM method and signal reproduction.

Surface Noise

Hunt made a detailed study [24] into phonograph surface noise from stylus reproduction. He states two mechanisms for surface noise, both of which had origins in the playback process itself. One stems from the stick-slip nature of sliding friction between the groove wall and reproducing stylus. The second comes from the microscopic imperfections of the surface (micro-roughness), and the resultant displacement of the stylus upon it. The first condition does not affect the non-contact method, but the issue of surface roughness does. Micro-roughness of the surface will be measured by the NCFSM method; hence signal reproduction will still be affected by characteristic surface noise. Surface noise is random, broadband noise, with a relatively flat frequency spectrum up to the highest audible frequency. It is commonly referred to as "hiss", and is analogous to the prolonged sibilant sounds heard in speech, for example, with the word "yes".

Surface Eccentricity

Imperfect manufacturing processes and long-term storage effects mean that cylinders are generally non-concentric, i.e. the cylinder radius is variable along its length. This has implications for both measurement, and signal reproduction. For playback, surface eccentricity can result in a perceived fluctuation in pitch, because the surface speed is no longer constant upon playback. This is sometimes referred to as "wow". For non-contact surface measurement, greater cylinder eccentricity means that the sensor must operate over a larger range (sensor gauge range is discussed more in Section 2.5.4).

Deformation, debris and mould growth

Early cylinder recordings can only be played back a finite number of times before the signal degrades. With every successive stylus playback, the grooves undergo a certain degree of plastic deformation, caused by contacting forces. If these forces are too high,

groove deformation occurs. An NCFSM surface measurement in Figure 2.14 shows an example of wear, caused by a stylus scratching across the groove. Larger scale imperfections of the grooves, such as grit, pits or scratches result in impulsive noise upon playback. These impulses in the playback signal are often referred to as 'clicks' or 'pops'. Signal processing techniques developed by Godsill et al. [25-27] have been devised for the detection and removal of scratch noise and transients, typically produced by surface wear and debris.

Wax cylinders are also susceptible to mould and fungus growth, which causes pitting of the surface (see Fig. 2.15). Mould growth is a destructive process, which results in a loss of signal fidelity.

Groove Loss

Groove loss occurs when a sustained portion of the groove has been removed or worn, to a point where the audio signal is no longer retrievable. This is very likely to be encountered with tinfoil records for example, which could only be played back a few times before becoming incomprehensible. Broken cylinders that have been reconstructed (see Fig. 2.16b for example), will exhibit groove loss if surface regions are not fully reconstructed.

In addition to these damage instances, vertical-cut grooves can exhibit discontinuities (see Fig. 2.17) when the cutting stylus temporarily leaves the surface of the cylinder, during extreme excursions of the recorder diaphragm [17]. To counteract the effect of groove loss, an interpolation of the audio signal will be required, to ensure that the track appears to be continuous.

Temporal Distortion (Speed Effects)

Recordings which have become warped or deformed, (through poor storage or from creep for example), may result in fluctuations in playback pitch, an effect often referred to as 'wow'. This defect occurs because the surface speed of the cylinder is not constant. Pitch defects can also be introduced at the time of recording, if the rotational speed of cylinder apparatus is not held constant.

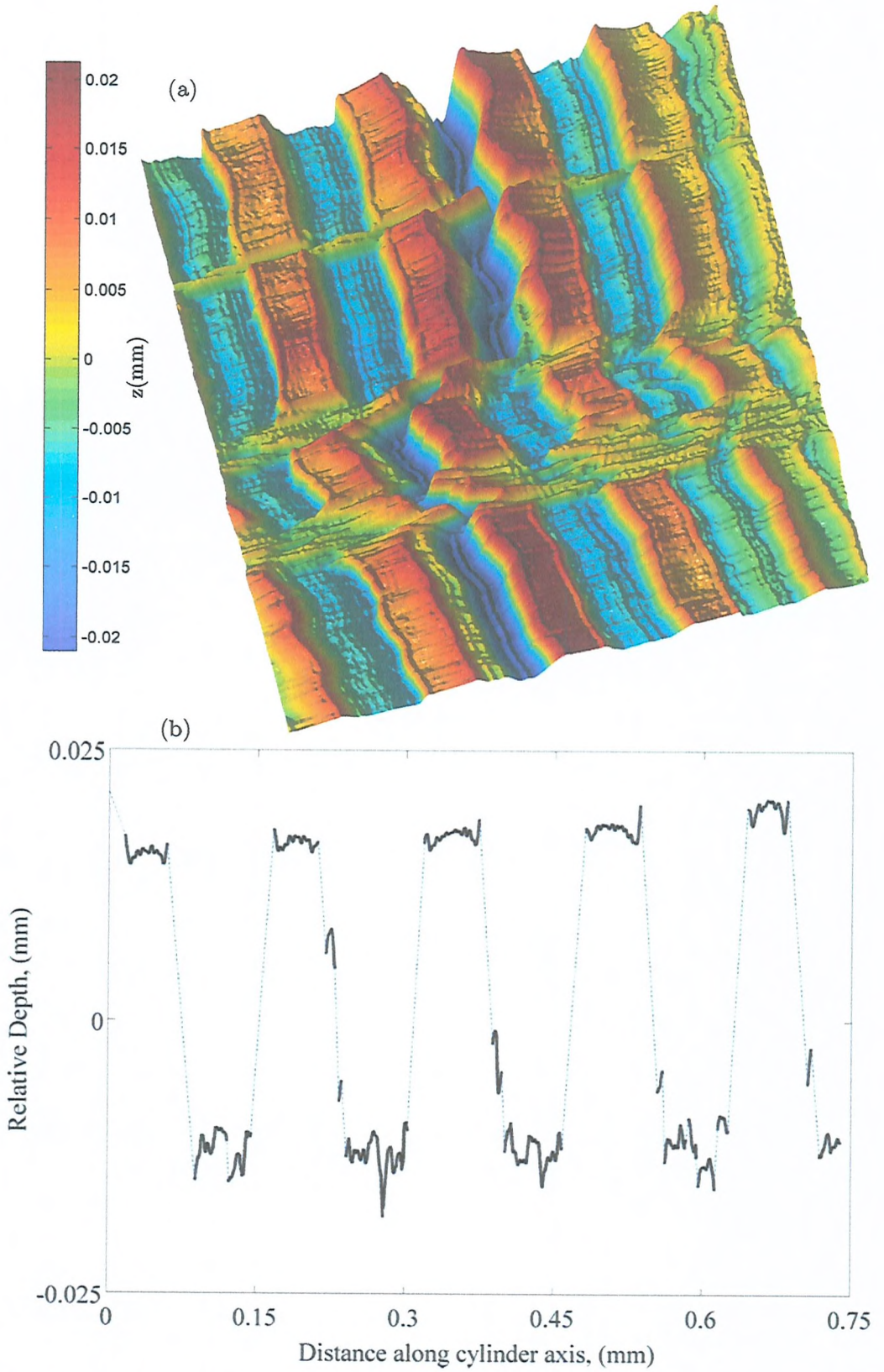

Figure 2.14: (a) 3D surface measurement of Graphophone cylinder grooves (GRAPH) showing damage caused by stylus replay. The deeper region (dark blue) is seen to cut across the adjacent groove. This surface plot measured by the Sound Archive Project system is produced from a series of depth profile measurements, as shown in (b). Note that dotted-line data in (b) was not measured by the sensor and has been interpolated for visualisation purposes.

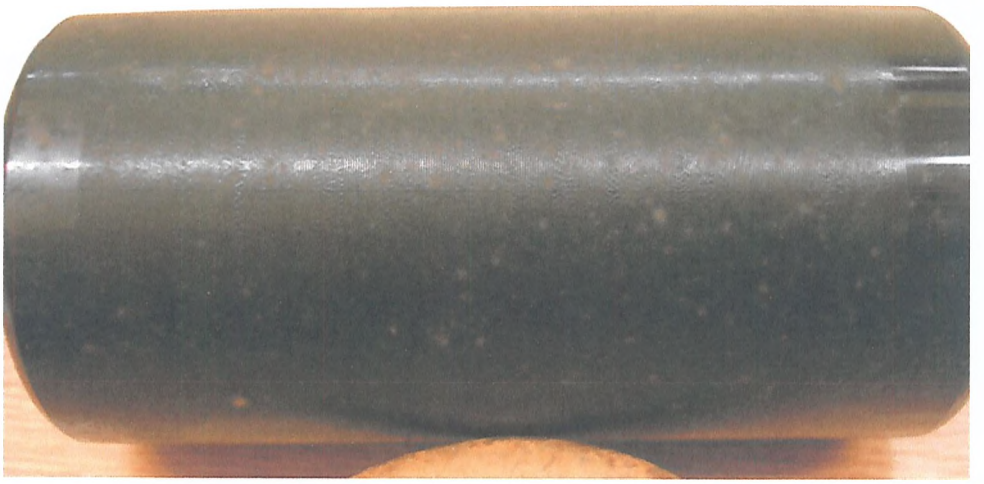

Figure 2.15: The organic composition of wax cylinders makes them susceptible to mould growth, which can be seen with this specimen (BIRDS) by the lighter 'speckled' regions.

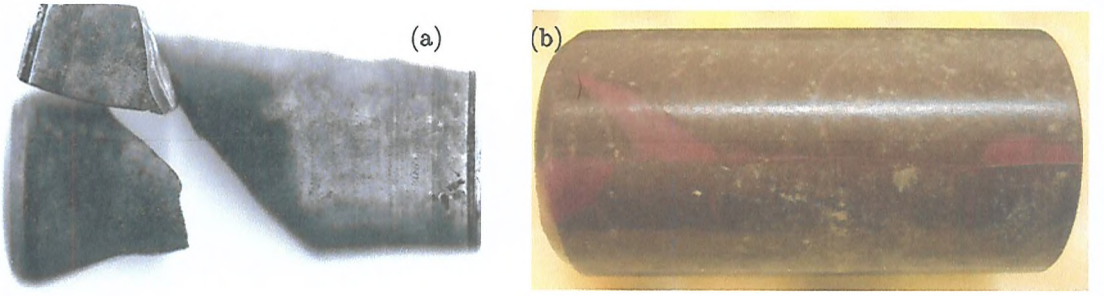

Figure 2.16: Examples of (a) wax cylinder breakage and (b) efforts to reconstruct a broken cylinder (EVANS). Red regions in (b) are areas of wax used to re-attach cylinder fragments.

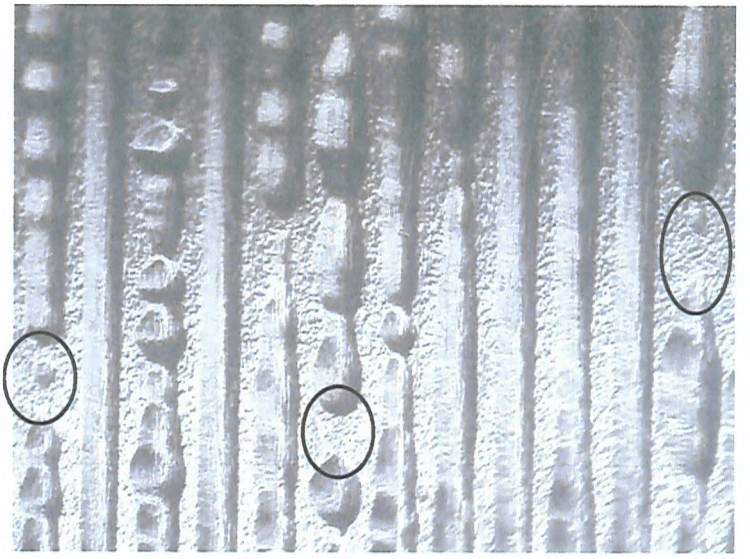

Figure 2.17: Photomicrograph showing discontinuous cylinder record groove. Ovals indicate regions where the cutting stylus has momentarily left the surface, resulting in signal loss. (Image from [28]).

2.2 Disc Recordings

Although the focus of this thesis concerns signal reproduction from cylinder recordings, a brief consideration of monophonic disc recording technology is now given. For the interested reader, a collection of technical papers regarding disc recording and reproduction can be found in [29], also [30]. For historical developments, the reader is again referred to [7-8].

2.2.1 Gramophone

In 1887, ten years after Edison's first cylinder system, Emile Berliner produced the Gramophone [31], which was the first system to record and reproduce sound onto disc. Discs produced on Berliner's Gramophone had a laterally modulated ("side-to-side") groove as opposed to the vertical, "hill-and-dale" cylinder system. The lateral recording process was similar to that of the Phonautograph, produced by Leon Scott in 1857 [9], which used a hog bristle to produce a trace of the incident sound onto lamp-blacked paper (although Scott's machine was never intended as a system for reproducing sound, merely recording). Berliner initially used lamp-blacked paper as the recording medium, but combined this with an etching process that allowed for transfer of the original traces to a copper or nickel layer [30]. Sound could then be reproduced from this metal layer and duplicates could be made from a master. By 1895, Berliner favoured discs that were coated with wax.

Disc materials

Various kinds of disc materials were experimented with in later years. Three of the most commonly used materials were Shellac, Acetate, and Polyvinyl Chloride (PVC, or more commonly 'Vinyl'). Shellac discs contain organic materials and hence like wax, were susceptible to mould growth. Acetate records use a layer of cellulose nitrate lacquer with a finer grain than shellac, which increases the fidelity of the sound. Acetates are however most susceptible to breakage, due to stresses which form in the lacquer layers over time. Most modern pressed records are vinyl records. PVC has a finer grain still over acetate, providing even greater fidelity. Vinyl records are also the most stable material and are resistant to mould growth.

Recording Speed

Disc recording speeds varied throughout the 20th century, but eventually 3 standard playback rates were developed: 78 rpm, 33 rpm, and later 45 rpm. As with early tinfoil and cylinder recordings, the earliest disc records did not have a standardised playback speed.

Summary of Disc Recording

Whereas research into cylinder technology ceased after 1929, disc recording and reproduction technology continued to advance throughout the 20th Century. Electrical recording and reproduction methods improved the frequency response and fidelity of the sound. Other advances included double-sided disc recording, long-play (LP) records, (33 rpm discs), increased playback duration (up to 30 minutes per side for a 12" LP record for example), and the invention of Stereophonic (two-channel) recording. In general, discs have a greater surface area and longer playback duration than cylinders, hence the measurement requirements are much more demanding. A summary of parameters for two common disc standards is given in Figure 2.18.

Parameter	**78 rpm (10" diameter)**	**$33^{1/3}$ rpm (12" diameter)**
Groove Width (μm)	150 – 200	25 – 75
Groove Depth (μm)	~74	15
Bottom Radius (μm)	38 - 58	3.8
Groove Angle	82 – 98°	87 - 92°
Grooves per mm	3.78 – 5.35	7.87 – 11.81
Inter-groove spacing (μm)	175 – 250	38 – 50
Max. groove amplitude before intercutting (μm)	100 – 125	38 – 50
Recorded track length (m)	152	437
Grooved surface area (m^2)	0.0386	0.055

Figure 2.18: Basic mechanical parameters of lateral groove recordings (from [6]).

2.2.2 Recording Principles

The recording principle for acoustically recorded discs is similar to that shown in Figure 2.7, except that a rotating flat disc replaces the cylinder, and the cutting stylus propagates inwards, in a concentric spiral from the outer rim to the disc centre. Monaurally recorded discs employed a lateral, 'side-to-side' cutting mode, as shown in Figure 2.19. The lateral cutting action is analogous to vibrating a pen upon a piece of paper, whilst sliding the paper orthogonally to the direction of vibration. Unlike cylinder records, the groove width *a-a'* is constant throughout the duration of the recording. The encoded groove (see Figure 2.20) consists of two walls, with a curved bottom, with curvature dependent on the cutting tool radius. The surface between adjacent grooves is referred to as the land. The resulting groove walls produced by the cutting tool, which are generally 'V-shaped' (for 78 rpm), or 'U-shaped' (for 33rpm) run parallel, along a spiral towards the record centre.

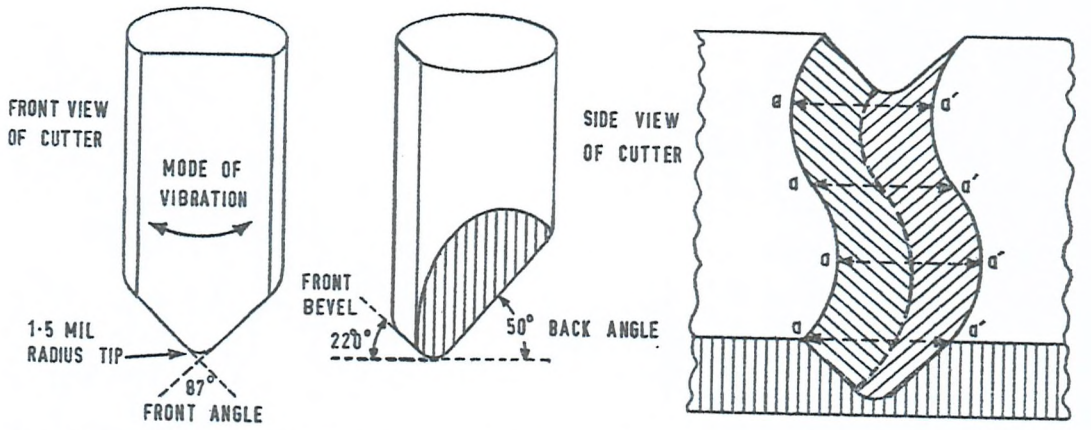

Figure 2.19: The lateral cut groove system ("side-to-side") of recording. The cutting stylus moves in the horizontal plane, producing a groove of constant depth. The distance a-a' is constant. (Image from [16]).

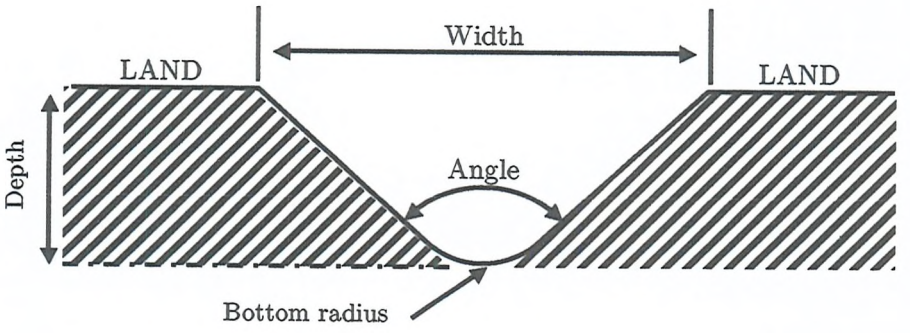

Figure 2.20: Disc groove cross-section, typical of 78 / 33 rpm disc.

2.3 Sound Reproduction Methods

In this section, reproduction methods from mechanical sound recordings are described. This begins with a brief look at stylus reproduction, and some of the issues associated with contacting methods. Next, non-contact methods of reproduction are introduced. The development of optical reproduction systems for both cylinders and discs has been the subject of much research in recent years. These are defined in three categories: (i) optical reproduction in situ, (ii) gray-scale imaging methods, and (iii) non-contact, full surface mapping (NCFSM) with post-processing.

2.3.1 Stylus Reproduction

Today, it is uncommon for sound archives to transfer recordings with original phonograph horn reproducers. Instead, direct, electrical transduction methods are preferred. By directly contacting a stylus with the groove and transmitting the modulations via an electrical transducer, the transfer function of the horn and mechanical linkage is bypassed, which results in an extended frequency response. Figure 2.21 shows the measured frequency spectra of pure tones recorded onto a blank Edison cylinder, and transferred using both original phonograph apparatus and electrical pickups. The response produced by the modern pickup has an extended frequency range when compared with the horn-microphone arrangement.

Stylus Choice

It is important for archivists to ensure the correct stylus is used, in order to measure the groove appropriately. Knowledge of the groove geometry is therefore essential, prior to transfer. The British Library Sound Archive (BLSA) for example, use modern stylus tips made of diamond or sapphire in various shapes and sizes. Tip radius ranges from 3.7 thou (94 μm) for 200 tpi, to 8 thou (203 μm) for 100 tpi cylinders. Styli of different shape are often experimented with, until the desired response is achieved.

Although sound is encoded onto a cylinder monophonically (one channel), it is possible for cylinder replay systems to utilise stereo pickups, (measuring two channels). Figure 2.8 showed how the vertical ('hill-and-dale') encoding of a cylinder also has a transverse ('side-to-side') component. Stereo pickups at the BLSA for instance have

been configured such that they respond to both the vertical and transverse modulation of the grooves, thus providing two channels of audio (stereo). The two channels are fed into a signal conditioning system and if necessary, the transverse signal can be subtracted from the vertical signal. The result of this subtraction is that surface noise (which is uncorrelated with the signal), is attenuated.

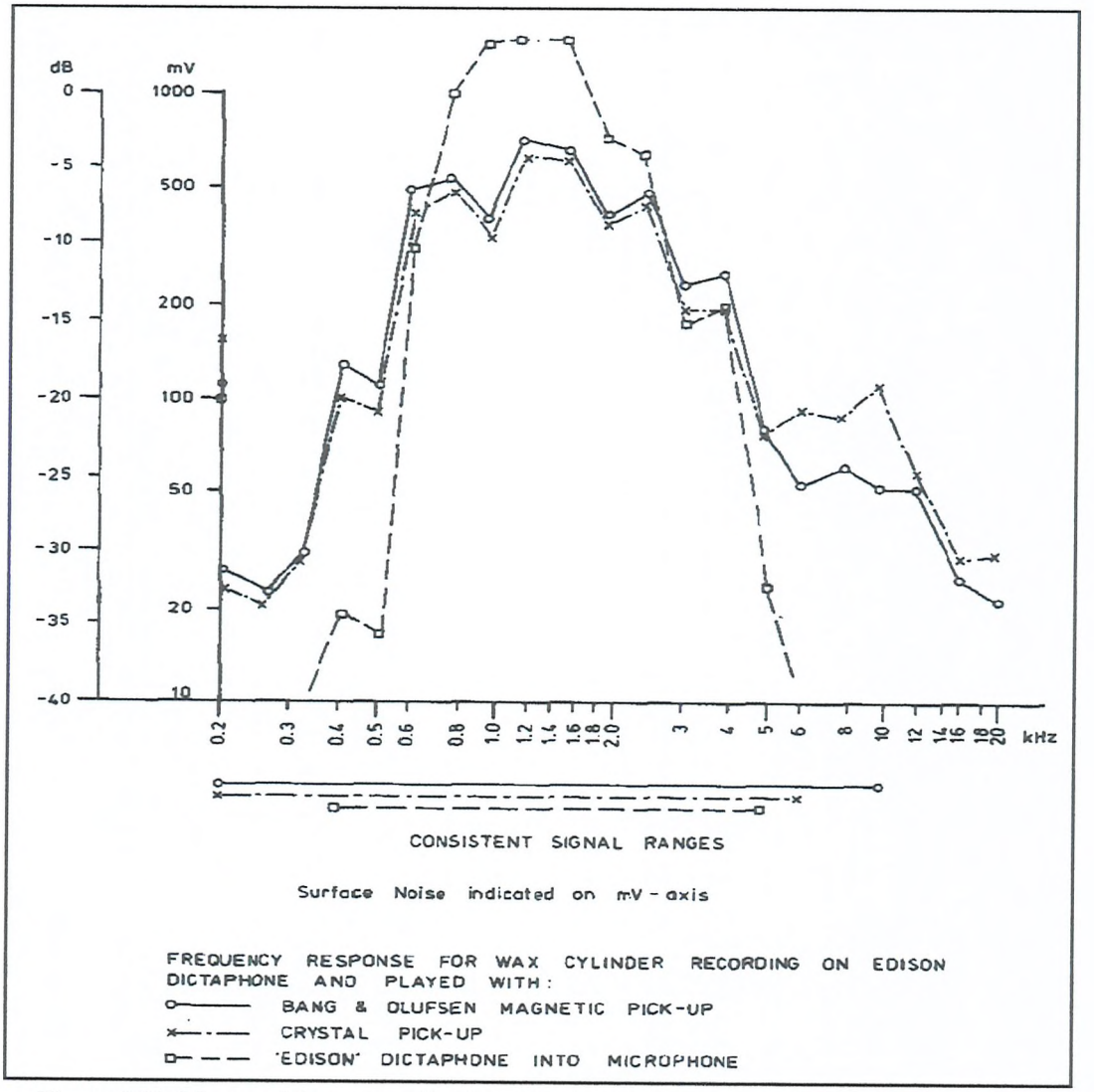

Figure 2.21: Measured frequency distribution of pure tones, recorded onto a blank Edison cylinder [32]. The original acoustic horn machine (Edison Dictaphone, recorded with a microphone) has a limited response above 5 kHz, whereas the other two electrical transduction methods (direct pick-ups) give an extended frequency response up to around 10 kHz.

Archival Standards for Digitisation

With regards to digital principles and standards of transfers, IASA-TC 04 [33] states that a minimum sampling rate of 48 kHz, and a preferred sampling rate of 96 kHz is required for the digital preservation of all recorded artefacts. This standard also recommends that archivists use A/D converters with a bit-depth of 24-bit. Modern digital audio equipment typically allows sampling at 192 kHz / 24-bit resolution, (though it is questionable whether a 24-bit A/D converter practically allows for a dynamic range of 144 dB for real signals). These recommended sampling practices might be considered 'over sampling', for acoustic recordings which have an upper frequency response of around 5 kHz. IASA-TC 03 [34] suggests however that high resolution transfers and over sampling of the 'unintended parts' of a sound document (i.e. noise and transients) may make the future removal of these artefacts by digital signal processing easier, for purposes of public access.

Factors affecting Stylus Reproduction

In order to compare stylus and non-contact methods of sound reproduction, it is necessary to have an understanding of problems associated with contacting the grooves with a stylus. Despite advances in modern reproduction systems, there are still cases where stylus replay is problematic, and the effect of contacting the grooves introduces distortion to the recovered signal. A detailed analysis of stylus wear, sources of surface noise, stylus/groove relationships and stylus geometry can be found in the literature [24,35-40]. These analyses highlight the geometric and dynamic factors of stylus reproduction systems, which can introduce distortion into signal chain. Examples of distortion include:

i) **Tone-arm / Cartridge Resonance** – caused by the dynamic moving mass of the stylus and tone-arm. The tone-arm essentially acts as a cantilever beam, with a mass, stiffness, and an associated resonant frequency. If the stylus becomes excited at resonance during playback, the recovered signal will become distorted.

ii) **Tracing Distortion** – caused by the stylus-groove relationship and arising from the fact that the stylus tip is of finite size [35]. The motion of a reproducer cannot always accurately follow the contour of the recorded surface, which results in distortion of the original signal. The stylus deviation from originally encoded signal depends upon the curvature of the groove and the size and shape of the stylus tip. Tracing distortion typically occurs when the radius of curvature of the recorded groove becomes smaller than the radius of the reproduction stylus, and the stylus is no longer able to trace the groove correctly. In the sinusoidal case shown in Figure 2.22b, the stylus no longer follows the original groove, and the path traced by the centre of the tip of a playback stylus over a perfect sinusoidal undulation is not purely sinusoidal. In general, higher frequency modulations are more affected by tracing distortion, due to increased groove curvature and smaller amplitudes. To reduce this effect, the size of the playback tip must be reduced, which can become impractical.

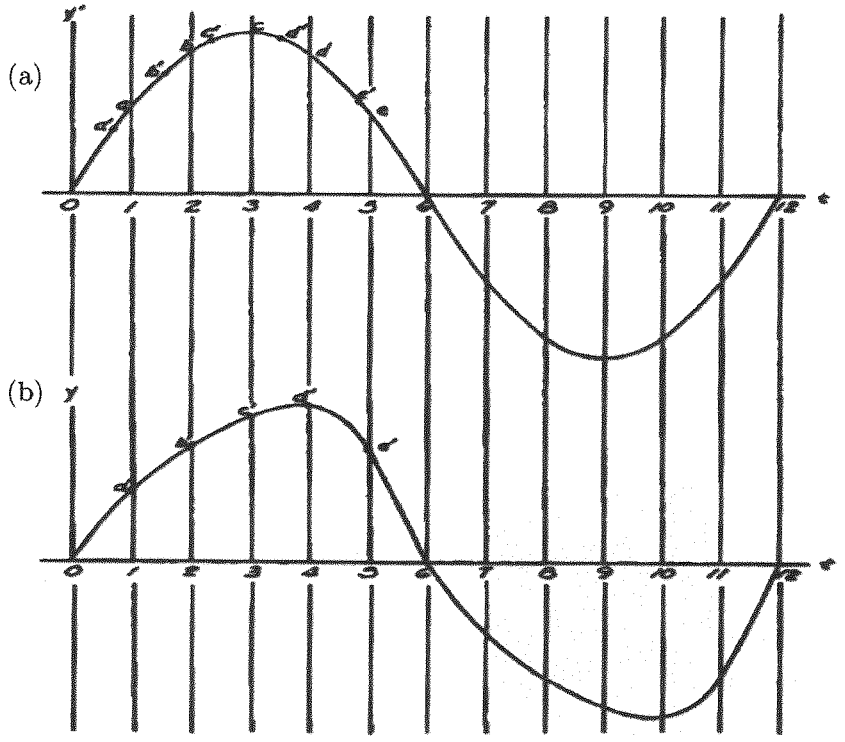

Figure 2.22: Signal trace of (a) ideal pick-up, tracing over a sinusoidal groove and (b) groove trace exhibiting tracing distortion, which no longer appears purely sinusoidal (Image from [38]).

iii) **Tracking issues** – when grooves are shallow, or if the tracking force is not set correctly, a stylus can temporarily leave the groove, causing signal drop out, groove skipping and impulsive noise. Additionally, if the helical path of the groove has been severely damaged, a stylus will not be able to follow the intended signal path in situ.

Summary of Stylus Reproduction

Modern stylus replay systems use direct, electrical transduction of the groove, in order to bypass the complex transfer function of a reproducing horn, which extends the frequency response. Although stylus reproduction may cause damage in extreme cases, a contacting stylus does have some advantages over optical methods. For instance, the stylus has inherent noise-suppressing properties, in terms of rejecting the micro-roughness of the surface, which optical methods are more sensitive to. In addition, the inertia of a stylus means that in some cases, loose debris can be dislodged from the groove, making subsequent playback less prone to impulsive noise.

It was also shown that stylus reproduction has some limitations. For instance, the finite size of the stylus tip limits the reproduction of higher frequencies, when tracing distortion occurs. In addition, stylus reproduction for damaged recordings is problematic, due to errors in tracking the groove correctly in situ.

2.3.2 Optical Reproduction in situ (ORIS)

Optical reproduction in situ is based upon the idea of an 'optical pickup', which translates the physical groove modulations into a digital audio signal. With optical reproduction in situ (ORIS), the full recorded surface is not preserved, but the encoded audio signal is recovered without causing damage to the grooves. Signal reproduction from ORIS methods relies on a groove tracking control system, which ensures that the sensor tracks the groove correctly.

ORIS systems for cylinders include interferometric methods developed by the Institute for Information Recording Problems, Ukraine, [41-42], and Syracuse

University [43], and laser beam reflection methods developed by Hokkaido University group in Japan [44-47].

For disc reproduction, examples of ORIS systems include laser beam reflection methods such as the Laserphone [48], a real-time playback system, known as the 'Laser Turntable' [49-50], and a laser diffraction method developed by Hokkaido University [47,51]. A selection of these ORIS methods and problems associated with playback in situ are now discussed.

Laser Beam Reflection Methods

The use of laser beams as a means of sound reproduction from disc grooves was first described in 1977 by Heine's 'Laser Scanning Phonograph Record player' [48]. Heine was inspired by high-resolution images of grooved records, taken via electron microscope. He postulated that if such fine groove detail was visible from these images, then an electro-optical system means of reproduction was possible. Heine's 'Laserphone' operated by focusing a Helium-Neon gas laser onto a 45/45 phonograph record groove and measuring the reflections. The position of the reflected light pattern (image) varied in response to the modulated groove. Heine was able to detect these variations in position via photo-detectors, which produced an output that was proportional to the magnitude and direction of image displacement. This system incorporated a tracking system with feedback, to ensure that the beam was positioned appropriately in the groove. Heine makes no comment on the success of the Laserphone reproduction, but gives an important observation, noting that for warped records, a long focal length is preferable, to allow for 'warp tolerance'.

A more advanced laser-beam reflection system for cylinder reproduction was developed by the Hokkaido University group. Figure 2.23 shows the principle of the laser-beam reflection method developed in [44-47]. A light source (provided by a He-Ne laser or laser diode), is focused by an objective lens upon the grooves and is reflected with angle θ', obeying the law of reflection. The reflected ray reaches a detecting plane, placed perpendicular to the optical axis, with origin O, centred at the point of intersection with the optical axis.

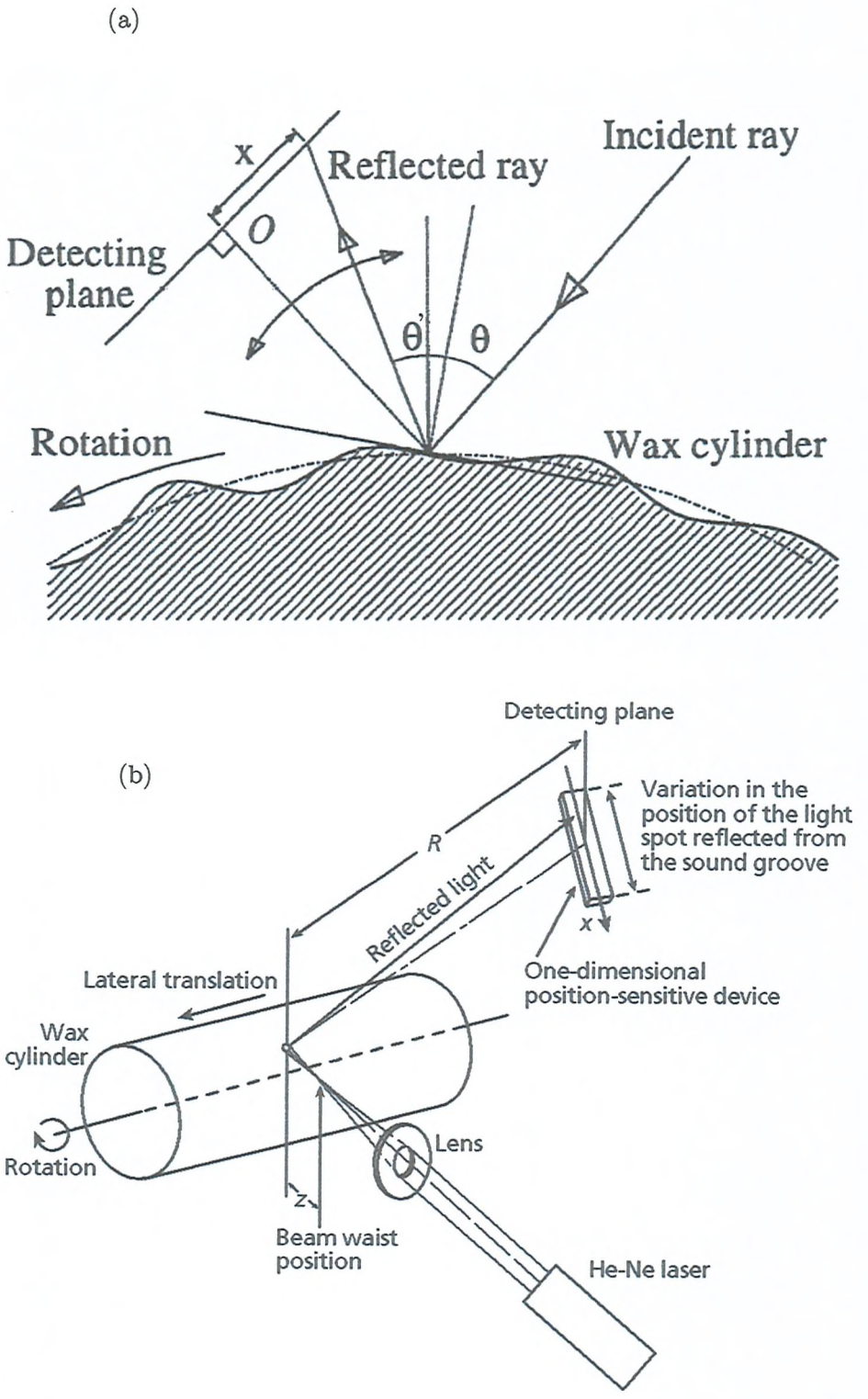

Figure 2.23: Optical sound reproduction in situ method for cylinders, developed at Hokkaido University, showing (a) 2-D schematic of laser beam reflection method with cylinder intersection (Image from [44]) and (b) overview of measurement principle (Image from [47]).

At the detecting plane, the point of intersection of the reflected ray is separated by a distance which is proportional to θ', and the distance between O and the illuminating point on the wax cylinder. When the cylinder rotates, the intersection point at the detecting plane varies temporally. This temporal variation at the detecting plane is measured by a position-sensitive device (PSD), which is later converted to an audio signal and recorded by a tape recorder. The detected position signal corresponds to the time-differentiated sound signal, since the angle θ' depends on the gradient of the groove at the illuminated spot position. The spatial resolution of the PSD, with an effective detection area of 34 x 2.5 mm, was 7μm, which was considered sufficient for sound reproduction.

Iwai et al. [44] found that the 'quality', 'timbre' and 'noise' of the reproduced signal varied as a function of beam width. The 'fidelity' and 'loudness of noise' increased as beam width was decreased; with increasing beam width, the level of high frequency components reduced and fidelity decreased. When the beam width was large compared with the groove dimensions, an 'echo' became audible. In Figure. 2.24(b), the diameter of the incident beam spot, is large compared with the groove width, and illuminates neighbouring grooves. The result for the reproduced signal is a superposition of groove displacements at times t_0, $t_0 + T$ and $t_0 - T$, (where T is the time taken for one cylinder rotation). This results in temporal distortion, or 'echo'. The Hokkaido group concluded that a beam width of 30-130 μm was suitable for cylinders, in order to avoid this echo. This is an expected order of magnitude for wax cylinders, which typically have a groove pitch of 127 – 254 μm.

Another issue with the laser-beam reflection method is that of tracking error, which occurs when the illuminating beam no longer images the grooves appropriately (see Fig. 2.24(c)). Tracking error is reduced by using a 2-D position-sensitive device (2D-PSD), to obtain a tracking error signal, which is fed back to the motion control system. This feedback control system reduces tracking error, but for warped cylinders, tracking error still occurs

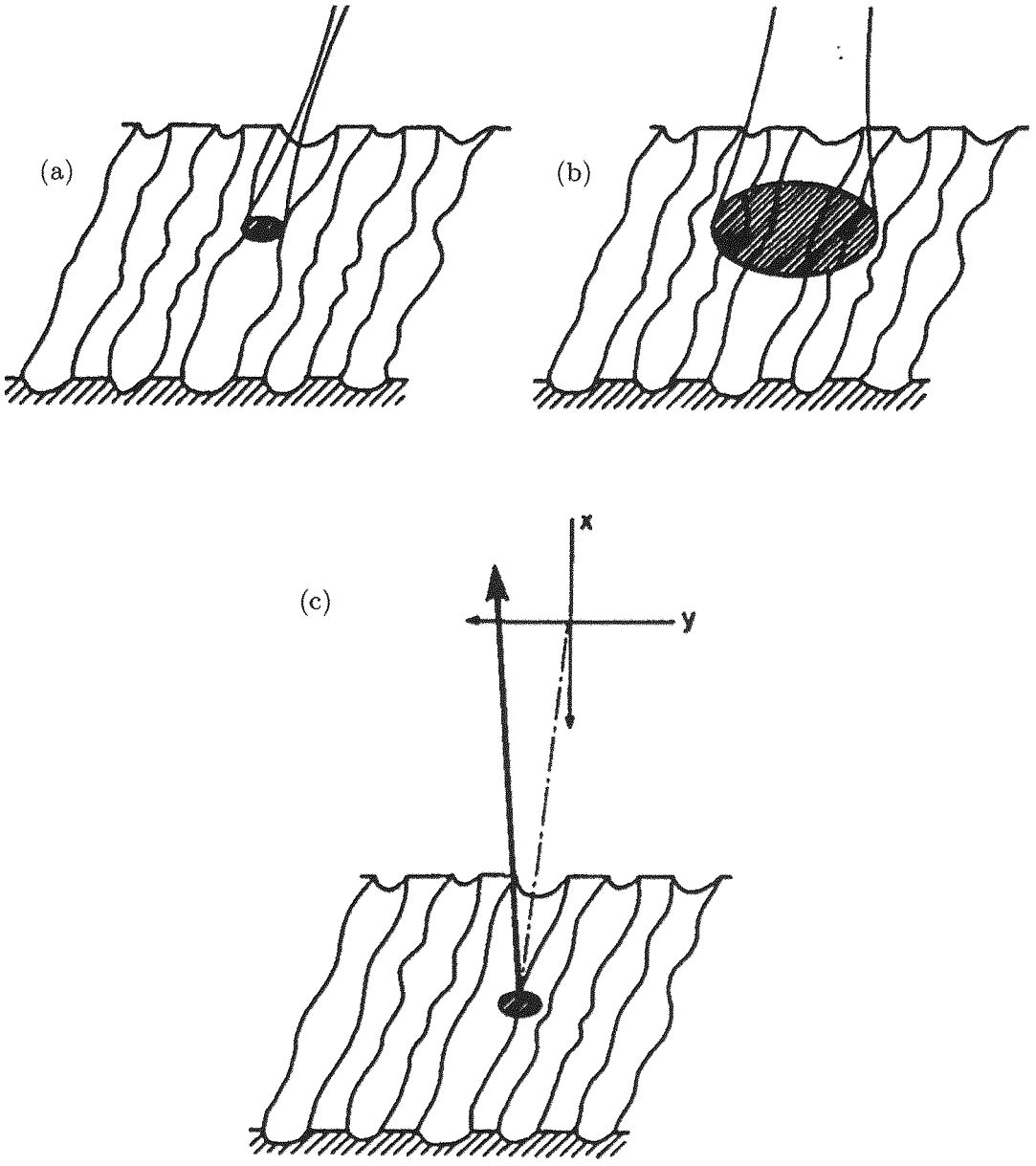

Figure 2.24: Considerations of laser beam reflection methods developed by Hokkaido University. In (a) appropriate beam width is focused correctly in groove valley. In (b) large beam width images multiple grooves, resulting in a lack of fidelity. In (c), the beam is focused incorrectly at the groove ridge resulting in an echo effect due to the superposition of grooves at two different time instances. (Images from [44])

Summary of ORIS methods

The philosophy behind ORIS systems is for sound reproduction only and not for purposes of surface preservation. ORIS methods rely heavily on a real-time tracking system, which must be accurate at time of measurement. Tracking groove features in this way can result in unsatisfactory signal reproduction, especially in the case of damaged or deformed surfaces, with complex groove structures. With regards to the laser beam reflection method the Hokkaido University group note the following problems with ORIS methods:

i) Fidelity of sound is poor,

ii) Noise artefacts are introduced through the measurement process,

iii) Echo (temporal distortion) is introduced,

iv) Accurate tracking of grooves is difficult to achieve in situ.

Comments about signal quality from ORIS methods have been made subjectively, and no attempts to quantify the signal to noise ratio (SNR) have been described in the literature. To the Author's knowledge, ORIS methods have not been used for reproduction from tinfoil recordings.

2.3.3 Grayscale imaging methods for discs

Grayscale imaging methods provided the first example of non-contact methods used for preserving *both the sound and surface topology* of discs. A grayscale image can be represented by a 2-D discrete matrix, $I(h,w)$, of dimensions $h \times w$:

$$I(h,w) = \begin{pmatrix} I(1,1) & \cdots & I(1,w) \\ \vdots & \ddots & \vdots \\ I(h,1) & \cdots & I(h,w) \end{pmatrix} \qquad I(h,w) \in \{0,1,\ldots,N\}$$

Where h and w correspond to height (rows) and width (columns) of the matrix. Each element of the matrix represents an image pixel, and is assigned an intensity, or gray integer value, n, $0 \leq n \leq N$. An image n that can only take 0 and 1 as values is said to

be a *binary image*. It is important to note that values in $I(h,w)$ relate to the reflected light intensity and *not the measured surface depth*. An example grayscale image from a 78 rpm disc acquired by a flat bed scanner is shown in Figure 2.25.

In general, a full 2-D grayscale image of the disc is acquired; with the groove topology stored as grayscale image intensity values. Signal reproduction then occurs in post-processing, which allows for more accurate tracking of the groove than with ORIS methods. The sound from monaurally recorded flat discs can be extracted by obtaining high resolution, grayscale images of the surface. This is because sound encoded on monaural discs is stored in the lateral modulation (transverse), as opposed to the vertical ('hill-and-dale') mode for cylinders.

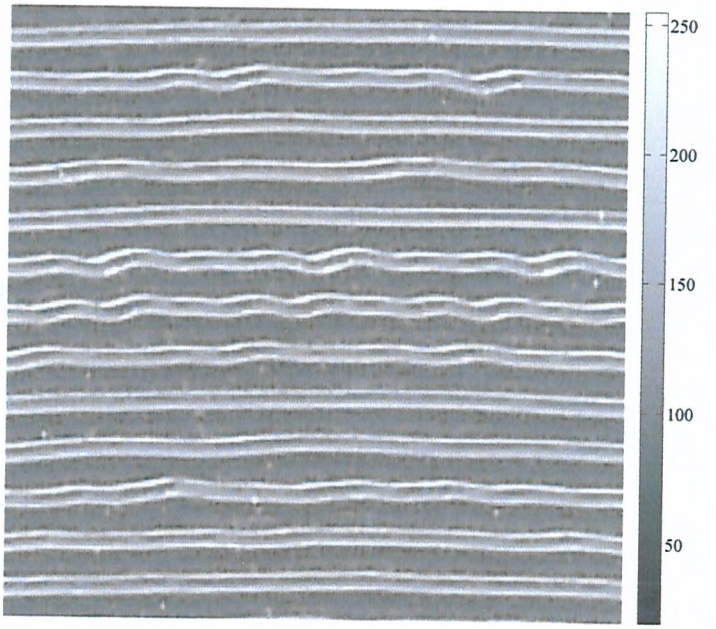

Figure 2.25: Grayscale image of 78rpm disc grooves, measured by Digital Needle flat bed scanner. Grayscale intensities represent image reflectance intensity, not surface depth. (Image from [52])

Several research groups have investigated grayscale imaging techniques for non-contact sound reproduction of disc recordings. These methods are now described.

Digital Needle

Two student projects at the Royal Institute of Technology, Stockholm, (Project Light-Blue [52] and Light Green, [53]) investigated methods of extracting sound information

from images of 78 rpm disc records, measured by a retail flat bed scanner. The scanner used was a CanoScan 5000F, capable of 2400 dpi (dots per inch). The sound extraction algorithms developed by Project Light-Blue are as follows:

Groove Localisation

In this step, the number of tracks and starting position of grooves on the record is determined. This is accomplished by firstly determining the groove spacing via FFT methods, and then detecting peaks in grayscale image intensity profiles (see Fig. 2.26). With the starting positions of grooves determined (See Fig 2.27), each groove can then be followed, in order to form a coherent track.

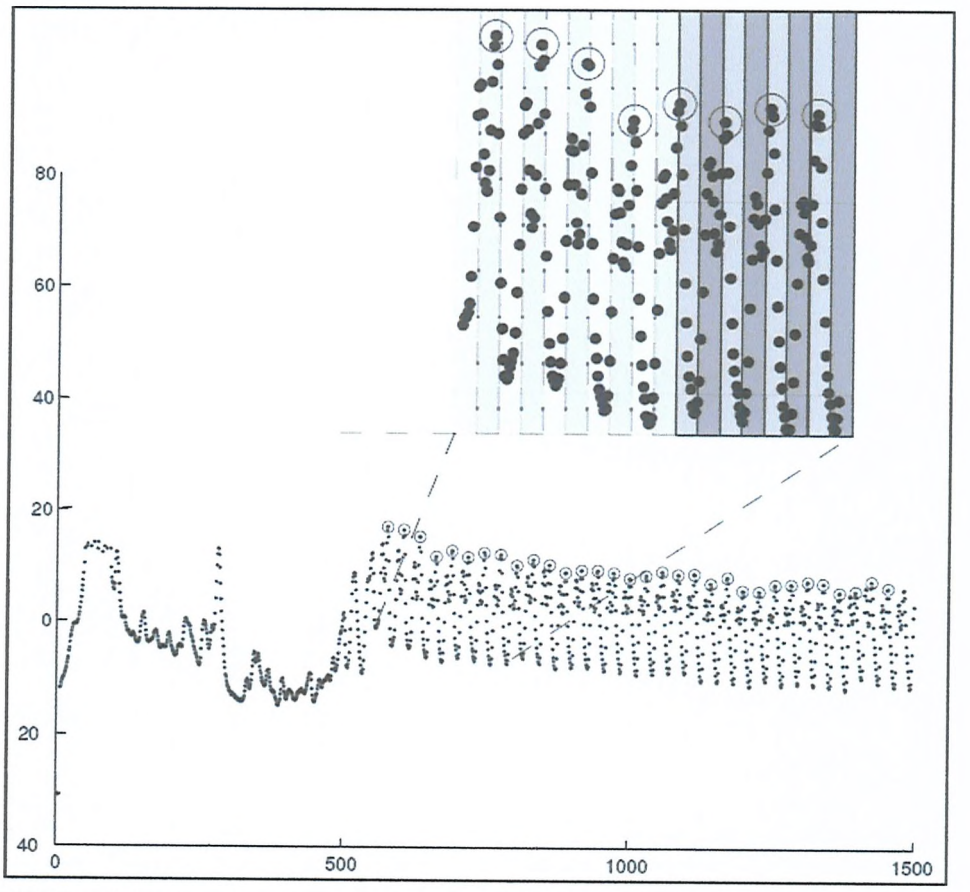

Figure 2.26: The detection of peaks values (indicated by circles) in image intensity profiles forms the initialisation stage for Digital Needle (Light Blue). This stage is required to ensure that all groove information is retrieved (Image from [52]).

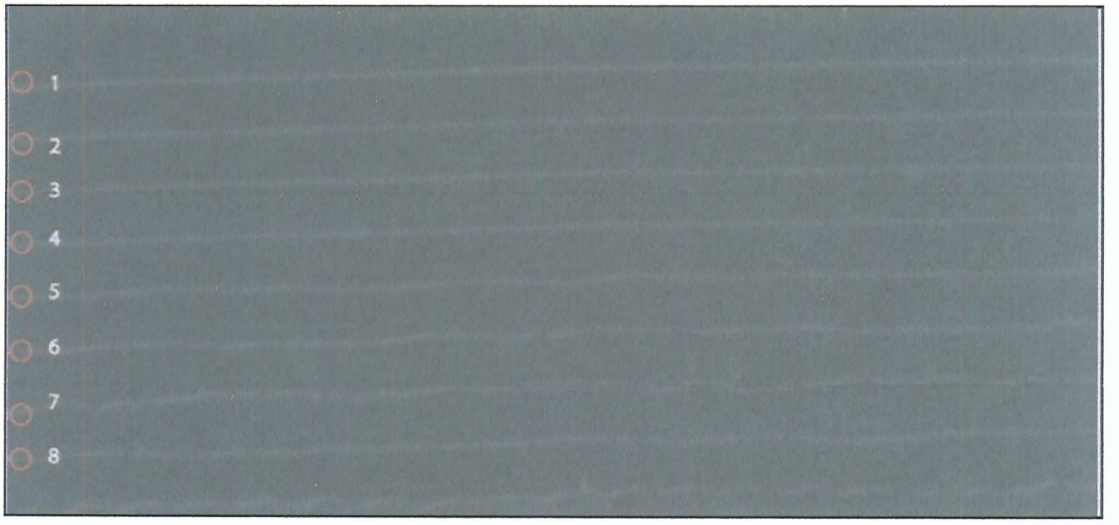

Figure 2.27: Track localisation for Digital Needle, Light blue group. Red circles represent the starting position from tracking grooves. (Image from [52])

Groove Tracking

In this stage, the grooves are tracked from the starting positions determined in the previous step. Grooves are tracked using a moving window method. The window P is a column vector of five pixels, centred on the starting position, y_1 given by (see Figure. 2.28):

$$P = \begin{pmatrix} I(y_1 + r(0),c) \\ I(y_1 + r(1),c) \\ I(y_1 + r(2),c) \\ I(y_1 + r(3),c) \\ I(y_1 + r(4),c) \end{pmatrix} = \begin{pmatrix} I(y_1 - 2,1) \\ I(y_1 - 1,1) \\ I(y_1,1) \\ I(y_1 + 1,1) \\ I(y_1 + 2,1) \end{pmatrix} = \begin{pmatrix} 0.8 \\ 0.8 \\ 0.9 \\ 0.9 \\ 0.8 \end{pmatrix}$$

Where c is the column index. The intensity values in P are then used as tracking weights, and a centre of mass quantity, $\rho(n)$ is formed, based on a weighted average of the neighbouring pixel intensities:

$$\rho(n) = \frac{\sum_{i=1}^{5} P(i) \cdot r(i)}{\sum_{i=1}^{5} P(i)}$$

With $r(i)$ = [-2, -1, 0, 1, 2].

In the next tracking step n+1, the column index c is increased by 1 and a new column vector of size [5×1], P is formed. This time, the central row index of the window is updated with index $y'_1 = y_1 + nint(\rho(n))$ obtained in the previous step (the operator '*nint*' stands for nearest integer function). By using the value $y_1 + nint(\rho(n))$, the search region for the updated vector's centre of mass is confined to a neighbourhood, in the vicinity of the previous step. The tracking procedure (shown in Fig. 2.28) continues until the end of image is reached, and then restarted for the next groove starting position, y_2, $y_3 \ldots y_n$. This technique is an example of image feature tracking using window neighbourhood methods. The size of neighbourhood used to calculate the centre of mass value is clearly of importance here, and must be determined from knowledge of groove spacing. The resultant track data given by the centre of mass quantity $\rho(n)$, gives an estimate of the radial position of the groove, corresponds to the audio signal. With regards to quality of signal reproduction from the Digital Needle Projects, no conclusions were given.

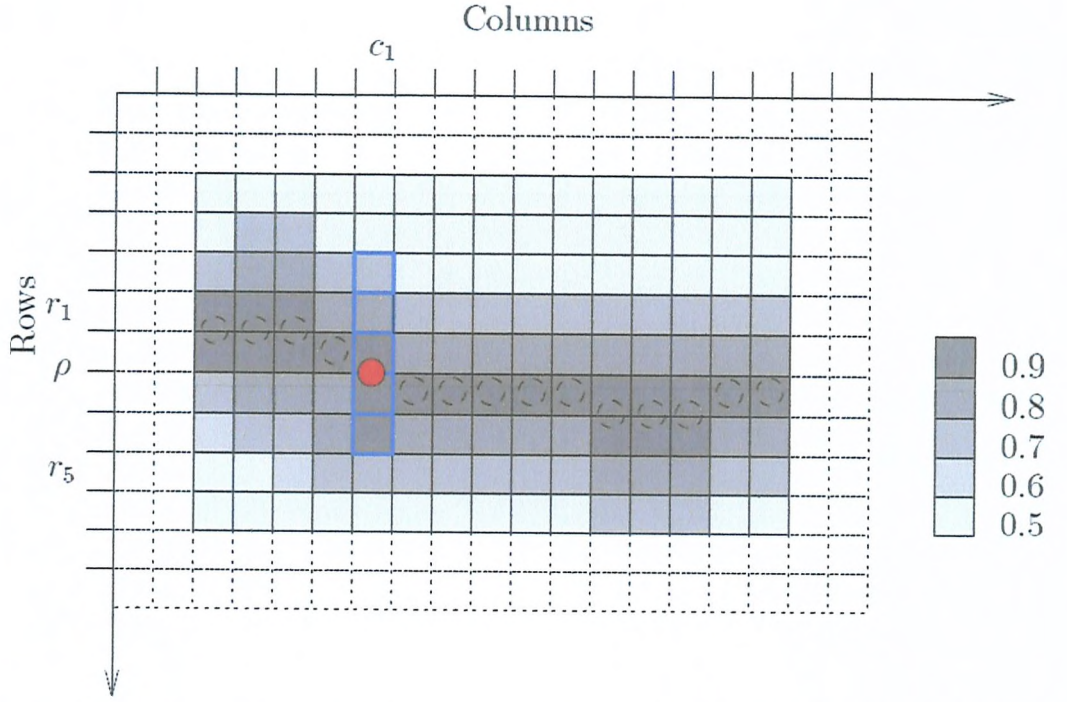

Figure 2.28: Method used by Digital Needle (Light-blue group) for tracking grooves in grayscale image, based on a moving-window method. The window, shown by the blue outline propagates across the column dimension, based on the 'centre of mass' value of the grayscale pixels in the current frame. Dashed circles represent the tracking sequence for this groove segment. (Image from [52])

VisualAudio

Stotzer et al. [54-59] developed a two-stage, grayscale imaging approach for sound recovery of laterally modulated disc grooves (see Fig. 2.29). Firstly, an analogue photographic process records an image of the disc onto analogue film. This film is then digitised using a rotating scanner (CCD camera) for processing in the computer. The author's claim that the analogue photography step allows for a larger depth of field (DOF), while imaging the disc, and ensures that the image to be digitised (the analogue film), fits into the reduced DOF required by the microscope's optics. In addition, an analogue copy of the record is stored, for long-term archival purposes. The VisualAudio (VA) method has been applied to 78 rpm and 33 1/3 rpm disc records with round and triangular groove shapes respectively, examples of which are shown in Figure 2.30.

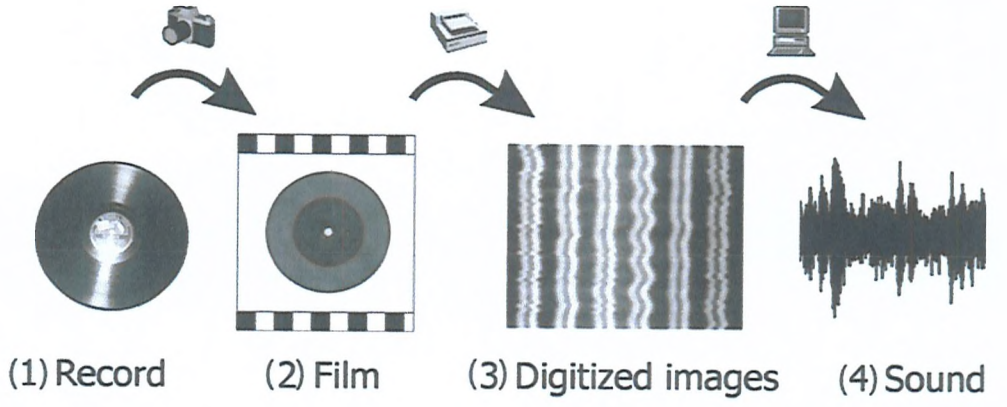

Figure 2.29: The VisualAudio concept of audio signal reproduction from disk records. (image from [54]).

Measurement Principles

With the VA method, microscopic photographs of the record are stored on high-grade, (grain size: 0.2 – 2 μm) black and white analogue film. The disc record is illuminated to ensure homogeneous illumination, and a high-resolution image of the record is taken. Illuminated flat sections of the groove (tops and bottoms) producing strong specular reflections appear black on the negative film. With the use of a circular scanner, the analogue film is then digitized via a 2048-sensor CCD-linear camera, mounted onto microscope optics. The record lies on a glass turntable and with each rotation, a radial

measurement is recorded. Each sensor in the CCD integrates the light of a 1 x 1 μm area of the surface. Adjacent sections are scanned in order to digitise the whole record surface.

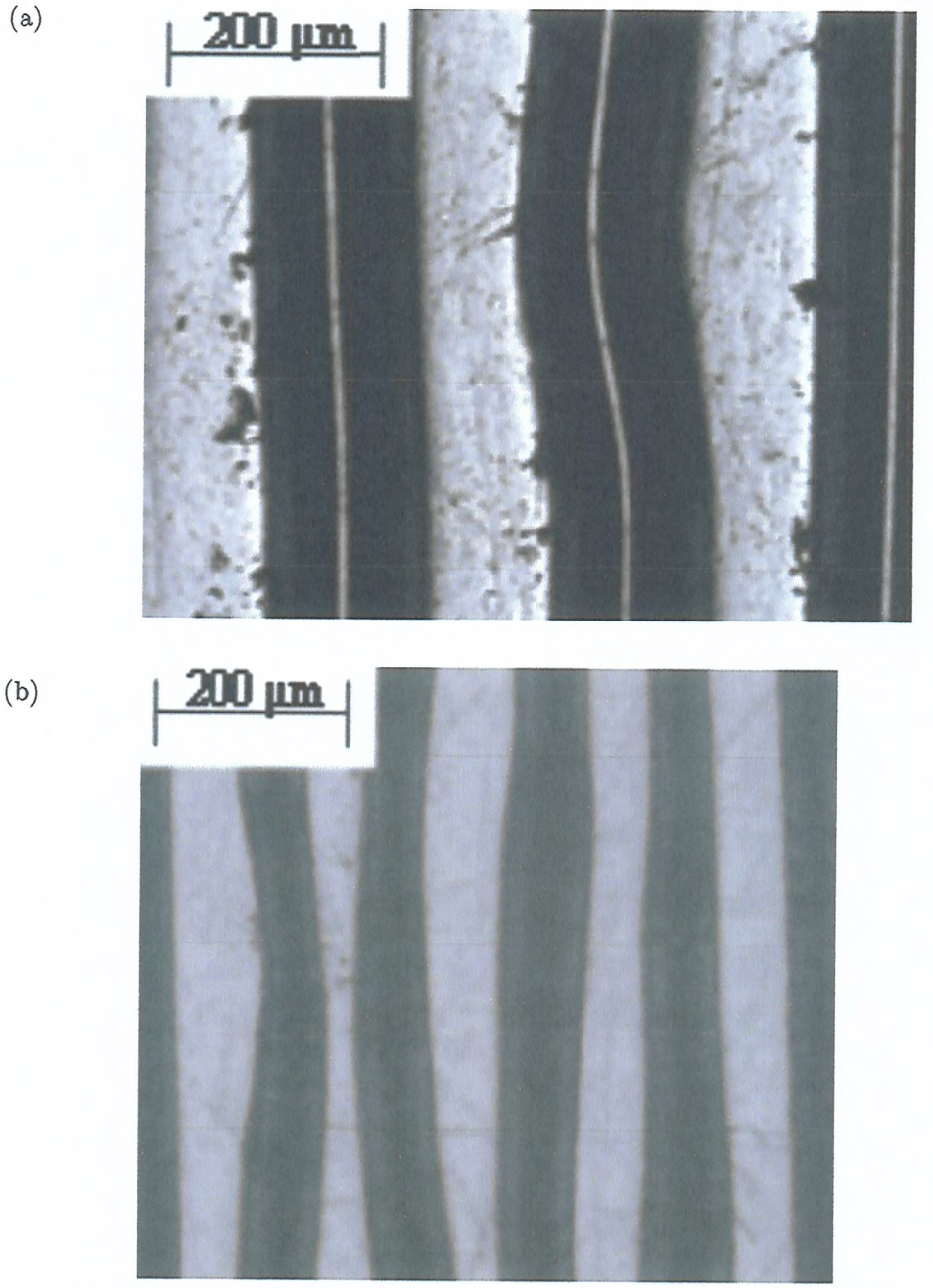

Figure 2.30: Grayscale groove images produced by the VisualAudio group of (a) 78 rpm disc with triangular groove cross-section and (b) 33 1/3 rpm disc with rounded groove cross-section. (Image from [59])

Groove Extraction from VisualAudio Method

Following image acquisition, a groove extraction stage is required to estimate the radial groove position, which corresponds to the instantaneous sound signal [54]. The image is processed by examining 1-D intensity profiles, taken across the grooves (See Figure 2.31a-b). Groove extraction algorithms, are summarised in three stages:

1. **Coarse Edge Detection** – achieved by convolving intensity profiles with a double box filter kernel $b(x)$ of size $\lambda \times 1$:

$$b = [-1 \ldots -1\ 0\ 1 \ldots 1]$$

The choice of λ, is based on the groove width, typically 10 to 20 pixels. The result of this convolution process is a smooth approximation of the derivative of the intensity profile (see Figure 2.31c). The position of extrema (maxima and minima) of this derivative profile forms a coarse approximation to the groove edge positions.

2. **Fine Edge Detection** – achieved by use of an adaptive pixel level threshold and linear fitting method (LMS). Edge candidate pairs, cr_k and cf_k, correspond to the 'raising' and 'falling' edges of the grooves, are returned. The detected edges for a single intensity profile are shown in Figure 2.31d)
3. **Trace following and sound** - This stage uses the sequence of fine edge pixel candidates, to iteratively reconstruct groove traces. Each trace T_i, is comprised of a rising edge r_i , and falling edge sequence f_i (see Figure 2.32). The groove trace following algorithm is as follows:

I) Initialise the traces T_i

II) Define raising and falling edge positions, $\hat{r}_i(j)$ and $\hat{f}_i(j)$ for the j-th profile.

III) A pixel neighbourhood m, is defined around each $\hat{r}_i(j)$ and $\hat{f}_i(j)$ pixel.

IV) Best edge point candidate c_k, is chosen for each edge sequence r_i and f_i.
If multiple edge candidates exist, best edge pixels are selected among these candidates $[cr_k, cf_k]$ by minimising $|\hat{r}_i(j) - cr_k|$ and $|\hat{f}_i(j) - cf_k|$ respectively.

V) Increment j and go to step (II).

The audio signal is then obtained by extracting the radial position of the groove traces.

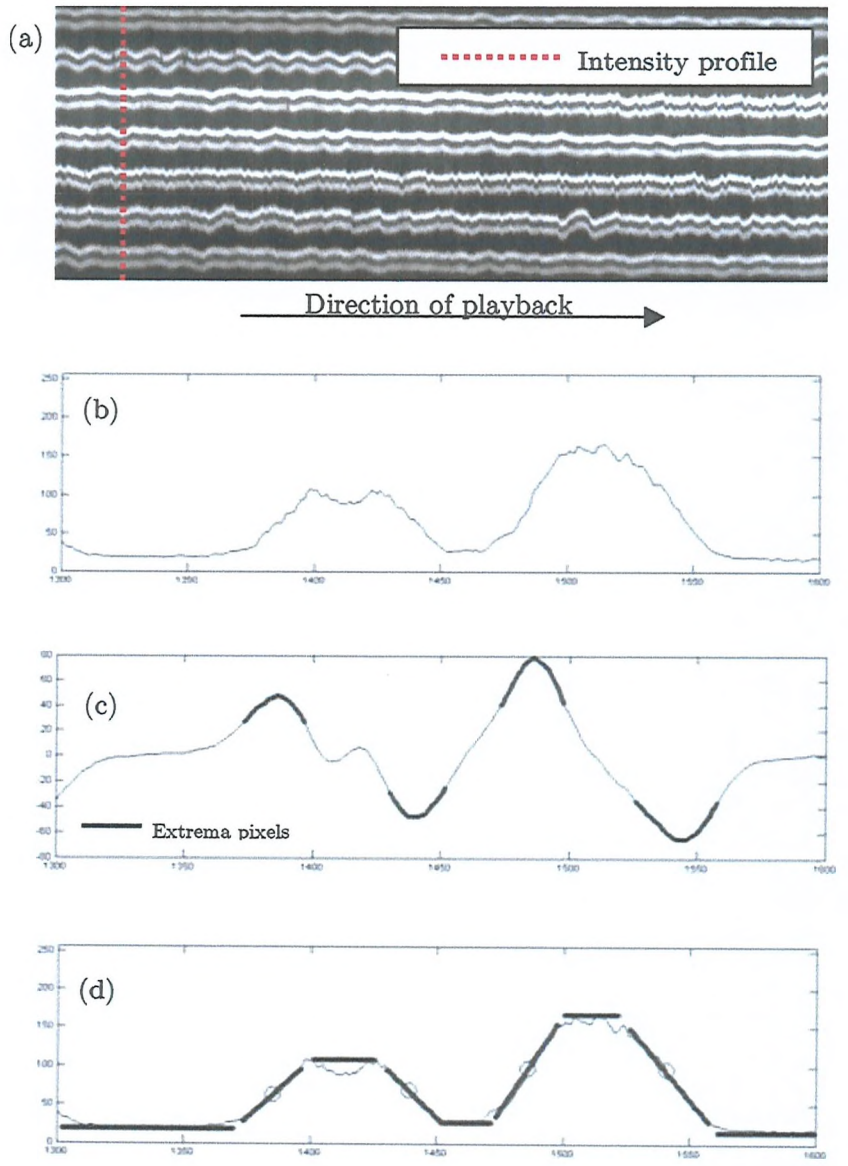

Figure 2.31: VisualAudio [55] groove edge extraction from using (a) grayscale image of 78rpm disc (b) intensity profile, (c) 1st derivative profile and (d) edges fitted via least squares fitting to the extrema pixels shown in the derivative profile (circles indicate detected edge centres)

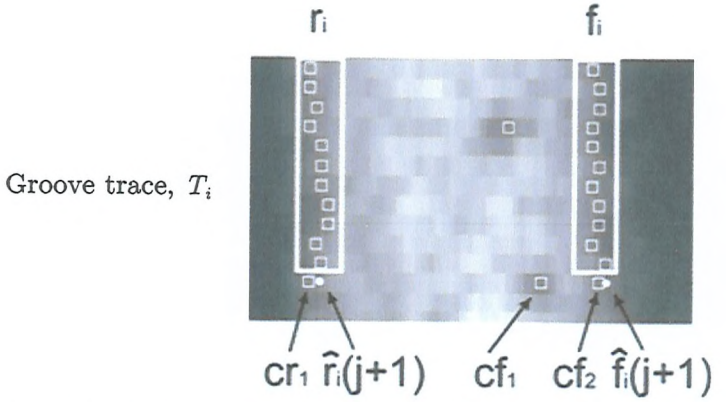

Figure 2.32: VisualAudio [57] groove traces T_i are built up by selecting edge pixel candidates $\hat{r}_i(j)$ and $\hat{f}_i(j)$ corresponding to rising (r_i) and falling edges (f_i) of the intensity profile.

Playback Sample Rate

Johnsen et al. [55] describes a maximum audio playback sampling frequency from the VA method of f_s = 72.8 kHz for 33rpm, and f_s = 170.4 kHz for 78rpm records. These values are determined from the number of radial measurements taken by the scanner, with a maximum of 1.31 x 10^5 radial measurements per disk rotation. Johnsen [55] notes subjectively however, that an image size, based on sampling of 6.5 x 10^4 radial measurements per rotation, does not significantly reduce the quality of the signal.

The aim for the VisualAudio group is to produce a signal that is 'comparable to stylus playback for records in good condition' [54]. This means a signal-to-noise ratio of up to 40 dB for 78 rpm records. Stotzer [54,57] gave an analytical consideration of the maximum allowable noise of the groove edge, in order to achieve a SNR that was comparable to stylus replay, concluding that the system had 'sufficient resolution' for 78rpm discs. A quantitative analysis of test records in [57] gave SNR results of 19dB for 78s and 16 dB for 33 rpm discs. Stotzer concluded that VisualAudio sound extraction resulted in lower sound quality than standard stylus playback, but does allow playback of records considered unplayable on turntables.

Local Damage Correction of Disk Image

Methods were also developed [55,57] for dealing with damaged recordings (see Fig. 2.33 for example). The assumption made is that corrupted pixels, caused by debris/scratch sites, produce replacement noise, which is uncorrelated with the original groove pixels. A 'corrupted pixel map' is used to detect and correct damage sites. Two gradient masks are applied to pixel data to detect outliers. Those pixels exceeding a gradient threshold are considered in the 'corrupted pixel map'. Stotzer designed these gradient masks to separate degradations into two different scales:

- Small, localised degradations, due to film grain or small dust.
- Larger scale degradations (due to scratches for example).

Pixels identified in the corrupted map were then replaced either via linear interpolation, least mean squares (LMS) fitting of neighbouring samples, or by an edge copy method, (the preferred correction method was not specified in [57]).

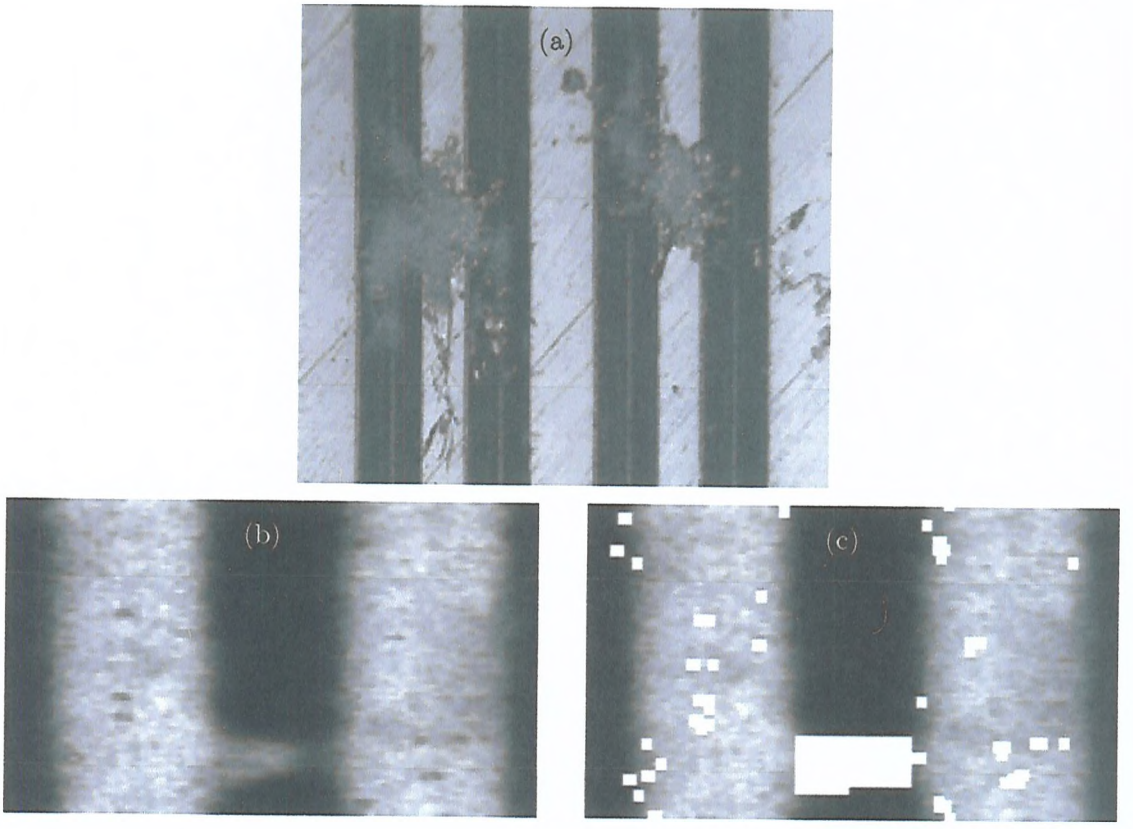

Figure 2.33: VisualAudio considerations of image degradations, showing (a) microscopic view of record grooves with fungus growth, (b) groove image showing local degradation and (c) corrupted pixel map, with corrupted pixels marked in white. (Images from [55]).

IRENE

A gray-scale imaging process was also developed for sound reproduction of discs at the Lawrence Berkeley National Laboratory. Their system, known as IRENE (Image, Reconstruct, Erase Noise, etc.) combines image processing with a general-purpose optical metrology system. In their earliest engineering report, [6], a large number of appropriately magnified, overlapping digital images of grooves from a 78 rpm disc are measured. The record is illuminated coaxially with a narrow beam of light. Flat regions of the groove (bottom and top) appear white, while the groove walls, scratches, and debris appear black (see Figure 2.34). The 2-D imaging process focuses on the bottom of the groove and sound is reconstructed by deducing a stylus trajectory. The device used to capture images consists of a video zoom microscope (with CCD image sensor)

and a precision X-Y table. With chosen magnification, the field of view is 700 x 540 μm^2, which meant that each pixel of the CCD camera corresponded to 0.91 x 1.90 μm area on the record surface. This gave a point resolution of 0.26 x 0.29 μm (compared with 2.5 μm^2 for the VisualAudio method).

Only a brief description of signal reproduction methods from the IRENE system is available in current literature, summarised in [6] as follows:

- Data from each image sequence frame is merged into a global polar (R,ϕ) coordinate system, where R is the distance from disc centre, and ϕ corresponds to azimuth angle.
- Using width information of points either side of the groove bottom (the groove width is approximately constant for 78 rpm discs), certain points are filtered out from the image.
- Small mechanical shifts due to movement of the mechanical stages were removed using correlation of the overlapping image segments.
- The edge point variation in R (radius) with change in ϕ (circumferential) results in the audio signal.
- 15-point edge sequences are fitted to a polynomial function, which is then differentiated analytically to approximate the velocity response of a magnetic cartridge.

Sound from this method was judged to be 'comparable' with stylus reproduction of the same recording and the IRENE system is currently under evaluation by the Library of Congress for digitisation of 78 rpm discs. Scanning times for discs are approaching real-time, with 4 seconds required to recover 1 second of audio.

The IRENE system has also been used to extract sound from a phonautograph recording, which predates the Tinfoil phonograph (c. 1857). The phonautograph was never produced with the intention of playback, and was used as a means of visualising sound for scientific purposes [63]. The phonautograph groove shown in Figure 2.35 is visually similar to a 78 rpm disc scan, thus signal reproduction was analysed in the same way, although the grooves had much lower fidelity.

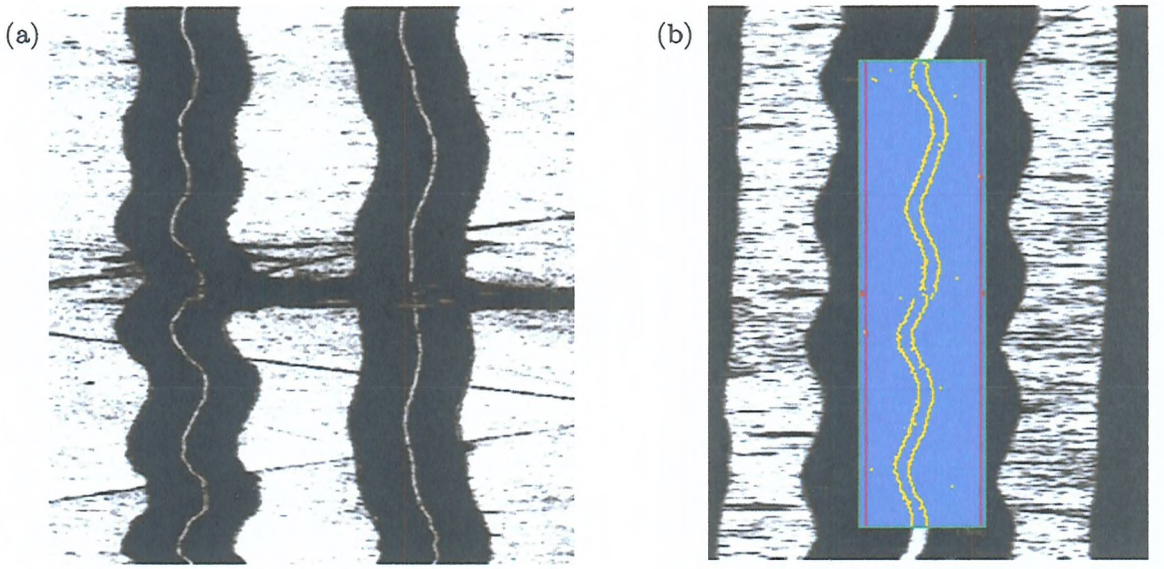

Figure 2.34: Grayscale images from IRENE system showing (a) scratch region across two 78 rpm disc grooves and (b) close-up highlighting the traced groove.
(Images from Fermilab presentation [64]).

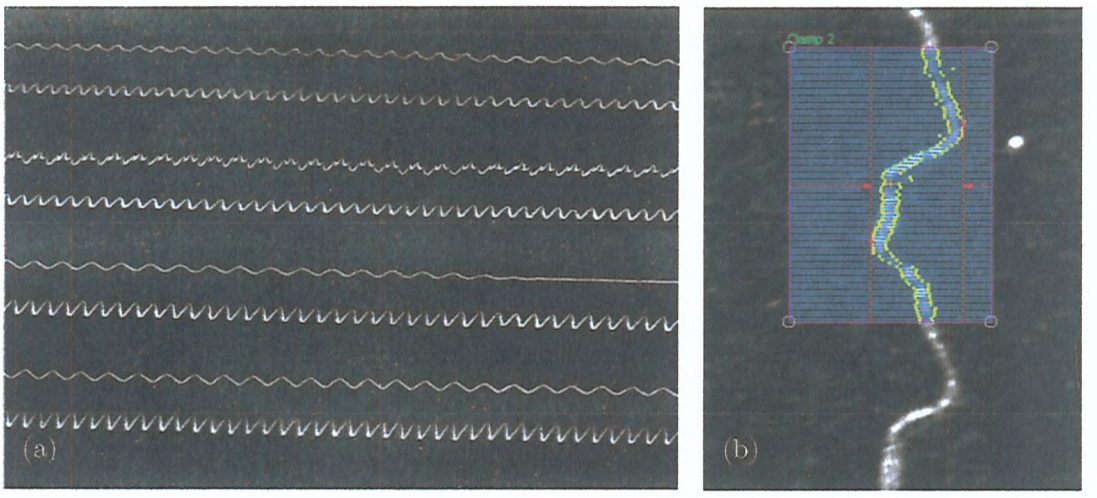

Figure 2.35: Phonautograph grooves produced using the IRENE system showing (a) eight groove traces and (b) close up of groove trace.
(Images from Fermilab presentation [64]).

Summary of Grayscale Imaging Methods

2-D imaging methods have been developed for non-contact sound reproduction from laterally recorded disk recordings. Grayscale images do not provide accurate depth information, which means that only lateral information from grooves can be retrieved. This makes grayscale imaging unsuitable for cylinder reproduction. In contrast to optical reproduction in situ, grayscale imaging methods do not suffer from issues due to tracking. This highlights one of the main advantages of full surface measurement, combined with post-processing.

In general, grayscale imaging methods provide a high playback sampling frequency of up to 170 kHz. Areal measurements can be acquired relatively quickly, when compared to single linescan profiles. The highest available pixel resolution quoted for the IRENE system is sub-micron, (0.25μm). From a preservation perspective, the VA method has some advantages over Digital Needle and IRENE systems, because an analogue film copy of the surface is stored, however film grain size is the limiting factor in terms of surface resolution.

Various post-processing methods have been devised to retrieve audio from grayscale images of monaural discs. These methods mainly rely on edge detection and knowledge of the groove width and encoding principles.

The use of test records with the VA group enables signal-to-noise ratio and total harmonic distortion to be calculated. Comparisons of these signal quality metrics can then be made with stylus reproduction. Without test signals, audio signal quality can only be judged subjectively, via listening tests, which is not satisfactory from a scientific perspective.

2.3.4 Non-contact, full surface mapping for cylinders

Cylinder recordings store sound primarily in the vertical modulation of the groove, which means that an accurate measurement of surface depth is required. Methods of estimating depth fields from grayscale images are possible, but they do not give sufficient resolution for sound reproduction. Moreover, the cylinder geometry cannot be

measured in a flat bed scanner or photographed to the same resolution, as per the methods described in [52-59].

An alternative reproduction strategy for cylinders introduced by Lawrence Berkeley National Laboratory (LBNL) in collaboration with the University of Southampton and TaiCaan Technologies is based on mapping the full 3-D surface topology of the recorded surface via optical sensor. The philosophy behind non-contact, full surface mapping (NCFSM) methods is preservation of the recorded surface, which allows access to the audio signal in post-processing. The feasibility and advantages of this scanning strategy were first investigated for a Blue Amberol cylinder in 2005 [5].

With the NCFSM method, a cylinder surface is mapped via a single-point optical displacement sensor (sensor technology is described in next section) combined with rotary and linear stage motion systems. By measuring a series of groove depth profile measurements (see Figure 2.36) at fixed azimuth angle, a discrete surface map $z(x_i,\vartheta_j)$, is formed (see Figure 2.37).

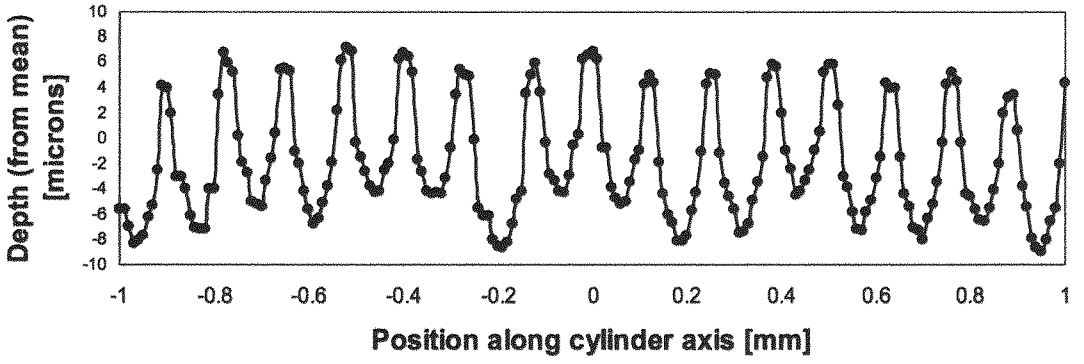

Figure 2.36: A 2mm linescan segment taken at fixed azimuth from a Blue Amberol cylinder, demonstrating the groove shape profile. Linescan grid spacing, Δx is 10μm. (Image from [5]).

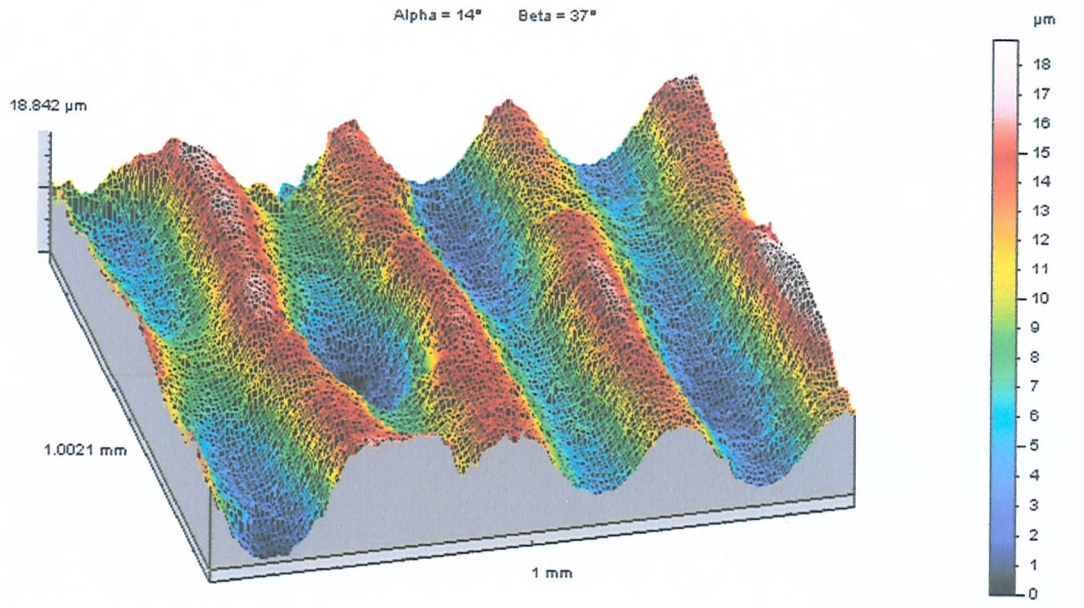

Figure 2.37: 1mm x 1mm segment of an Edison cylinder surface formed by a series of linear scans, acquired using the STIL CHR150 confocal probe. (Image from [5]).

This scanning strategy is the same used for methods developed in this thesis. A thorough description of surface measurement for cylinders is given in Chapter 3. In this chapter, we consider the signal reproduction methods initially developed by Fadeyev et al. A summary of signal reproduction methods from the Blue Amberol cylinder, as described in [5] is now given.

Groove Extrema (valleys and troughs) are located for each linescan. These points are identified by being data points that were higher or lower than the surrounding data, along the linescan.

I. The correct groove extrema (ridges and valleys) form a grid with fixed distance between neighbours, determined by the groove pitch λ_x. A relative phase or 'offset' of this grid along the cylinder axis is fitted to the list of extrema. Extrema candidates were compared to this grid and those that exceeded a distance threshold from the ideal fitted grid were considered outliers and removed from the list of extrema.

II. Groove ridge heights were taken from the list of valid extrema found in (II). For the k-th valley (groove bottom), the minimum height of the groove (valley) is determined by fitting a parabola with fixed quadratic term to the 10 surrounding points of the valley. This set of points h_{ij}, is fitted to a parabolic function, $H_k(x)$, with fixed quadratic term using a χ^2 minimisation:

$$H_k(x) = A_k x^2 + B_k x + C_k$$

with A_k held constant and:

$$\chi_k^2 = \frac{1}{\sigma^2} \sum_{i=l-5}^{l+5} \left[h_{ij} - H_k(x_{ij}) \right]^2$$

where l is the raw data index of the k-th valley, j is the linescan angular index. The parabolic shape is an approximation to a circular segment, and the fixed quadratic term, A_k is chosen to match a cutting tool of radius 130 μm, which was found to agree well with historical data [17]. Outliers were iteratively removed from this fit if they deviated by more than 3σ, and the fit was reiterated. Fit iterations stopped if no points deviated by more than 3σ, or only 5 points remained. The procedure of discarding outliers handles cases of debris or groove damage. As a result of the fit, the groove depth is found by:

$$d = C_k - B_k^2 / (4A).$$

III. With a set of groove extrema at fixed azimuth, the data is appropriately reordered into a time series that follows the spiral trajectory of the groove bottom. This is achieved by establishing the "seed" groove bottom positions from the first slice. For each position, the groove bottom (ridge) positions in the next slice are searched for within a neighbourhood along the cylinder axis. If this distance is less than 30 μm, then the position was attributed to the same groove,

and the seed value is updated to that of the next slice. When fitted and re-ordered, the data set contains 36,000 points per cylinder rotation. Assuming that the cylinder is perfectly concentric, a linear playback sample rate can be calculated as follows:

$$f_s = N_\theta \frac{\omega_{RPM}}{60}$$

Where N_θ is the number of linescans covering the cylinder surface and ω_{RPM} is the speed of cylinder rotation, in revolutions per minute. For N_θ = 36,000 and ω_{RPM} = 160, this equates to a linear playback sample rate of 96 kHz.

IV. Any errors in the data found during the previous fitting stage are linearly interpolated from the neighbouring points to form a displacement track. The authors note that three distinct data streams can be derived from the groove minima and maxima. These data streams are:

- Groove bottoms (valleys)
- Groove tops (ridges)
- Groove top minus groove valley.

The prescence of an audio signal contained within the groove top stream means that top-bottom subtraction method is biased and has the potential to introduce echo, due to the superposition of adjacent grooves. The authors postulate however that due to the logarithmic perception of sound amplitude, the samples would have to be very close for degradation to occur, thus the effect of top-bottom subtraction has little effect on the output signal.

V. The authors note that over a given frequency range, cylinder recordings exhibit constant velocity characteristics [5]. In order to obtain a signal with a flat response over a given frequency range, the extracted displacement data is differentiated numerically and digitally filtered in one step. This is implemented

via a discrete Fourier transform (DFT) of the displacment signal, and a multiplication with filter coefficeints in the frequency domain. The filter coefficents are set to match the operating frequency range of the recording system, which the authors state as being 20 Hz to 4.8 kHz. The audio signal is then resampled to 22.05 kHz sample rate, and scaled to have a 16-bit dynamic range.

In order to make subjective comparisons with stylus playback, the same cylinder was transferred using the Archeophone. It was noted in [5], that the similarity of the waveforms in Figure 2.38 demonstrates the accuracy of sound reproduction, when compared with a mechanical transfer, although no attempt was made to quantify the degree of similarity, or comparative audio quality of the two methods. Examination of the spectra in Figure 2.39 shows how the stylus and non-contact methods exhibit very different low frequency responses below 90 Hz. The sharp cut-off at 4.8 kHz in Figure 2.39 is due to the filtering operation, described in step (V).

One issue to note is that due to the finite gauge range over which the sensor can resolve surface displacements, only 30 seconds of the full four minute recording was successfully extracted. In order for this method to become practical for archival purposes, the full surface of the recording must be mapped. This requires advances to be made in optics, and sensor technology, to ensure that the sensor can operate over a greater gauge range. Nevertheless, the ability to map and preserve the full surface topography has great potential for archivists and has great advantages, over in situ methods, which rely on accurate tracking in real-time, as opposed to reconstructing the audio signal in post-processing.

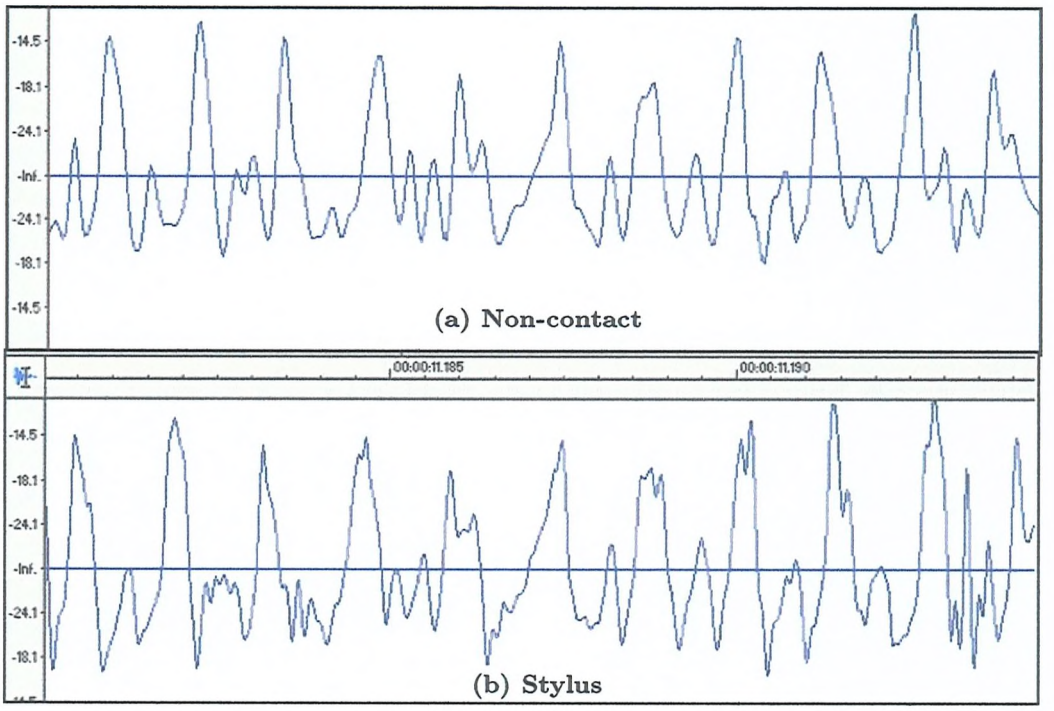

Figure 2.38: Comparison of audio waveforms (14ms) from (a) non-contact reproduction and (b) corresponding section from the stylus transfer. (Images from [5])

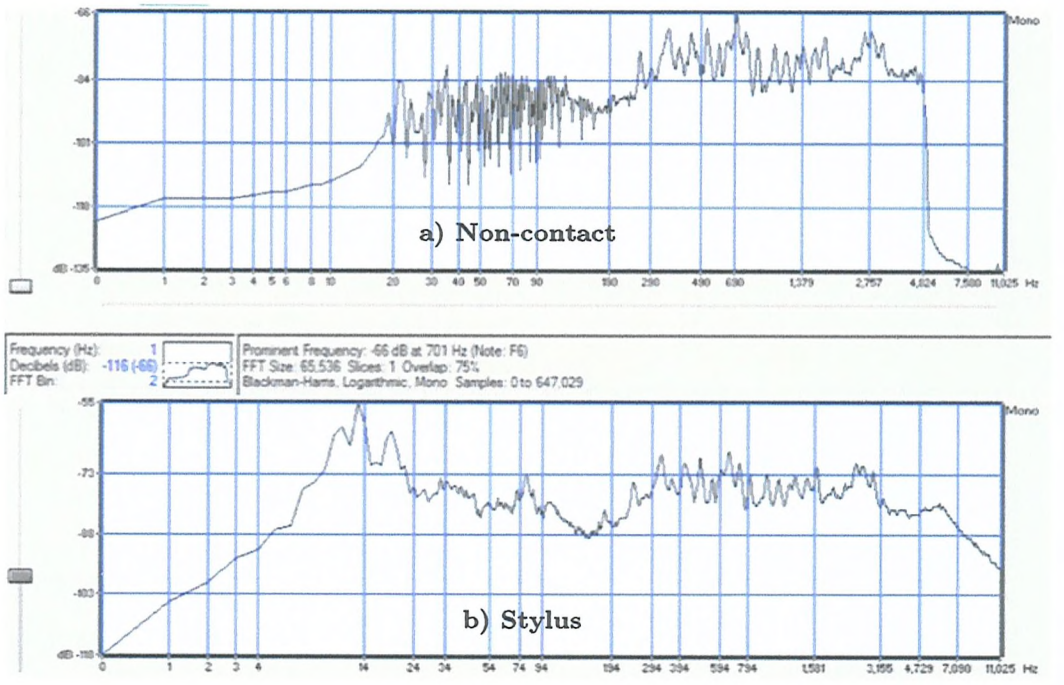

Figure 2.39: Comparison of frequency spectrums with (a) the non-contact method and (b) stylus transfer using Archeophone. Note the difference in low frequency response below 90 Hz. (Images from [5])

Summary of NCFSM Methods

The primary advantage of NCFSM methods is the ability to resolve surface depth to high resolution. By measuring true surface depth as (opposed to image reflectance as with grayscale imaging), NCFSM methods can potentially be used for modulated groove media, either vertical or laterally cut. Mapping the full surface topology have distinct advantages for recordings that are in poor surface condition, for which tracking the grooves in situ can be problematic. The NCFSM method uses a single-point measurement sensor to construct a surface composed of linescan profiles. This in contrast to grayscale imaging methods, which utilise CCD arrays, to take areal measurements. As a result, scan times for the NCFSM method are greater than with grayscale imaging methods.

Post-processing ensures accurate tracking of the grooves, and allows for damage sites to be identified and filtered. High resolution mapping of the surface also has the potential for sound recovery from recordings that are broken, by scanning the fragments individually, and reconstructing the full surface in the computer.

The majority of stylus replay systems use a magnetic cartridge, which results in a velocity signal, (not a displacement signal), hence to approximate the response of a stylus from the raw displacement track, a differentiation process is required. Haber accomplishes this by differentiating via an FFT method, which incorporates additional equalisation. It is unclear whether this method offers archivists a signal that most closely represents the original encoded sound. Ideally, for archival purposes the resultant audio signal should be presented in a raw, unfiltered form, and equalised later for purposes of public access.

Fitting an ideal shape primitive to the groove cross-section has advantages in terms of rejecting noise caused by outlying debris. The method developed by Fadeyev and Haber, which incorporate quadratic fitting of the groove cross-section, is well suited to Blue Amberol cylinders whose groove resembles a parabola. For earlier cylinders (such as the Graphophone groove profile shown in Figure 2.14b), a different fitting model is required. The issue of the 'ideal groove shape model' provides a topic for investigation in this thesis.

2.4 Related Signal Processing

This section describes techniques for quantifying signal reproduction from the NCFSM method. Ridge-valley detection, and smoothing filters, (in particular the Savitzky-Golay filter), are then described, with an emphasis on their importance to the methodology developed in this thesis.

2.4.1 Audio Signal Quality Assessment

Many methods for audio signal quality assessment based on the modelling of auditory perception exist in the literature, such as PEAQ (perceptual evaluation of audio quality), [60], or PEMO-Q (Perceptual model, Quality assessment), [61]. These metrics attempt to characterise the effective signal processing performed by the ear, taking into account psychoacoustic effects of the auditory system. Modelling such processes may introduce bias, due to the fact that the auditory system varies from one subject to the next. Ideally one should seek a quantitative measure of audio quality, which can be calculated numerically without bias.

For musical recordings (typically encountered with cylinder recordings for instance), the detection of a signal present in noise ultimately reduces to a subjective judgement [62]. For recordings with a known signal, such as a sinusoid embedded in Gaussian noise, the signal and noise models are more easily defined. In [57], several metrics for calculating the signal-to-noise ratio (SNR) were examined for signals reproduced from the scanning of a disc recording. The disc contained pure tones (sinusoids) and the following SNR metrics were defined:

Signal-to-Noise Ratio (SNR)

The SNR gives the ratio of the power between a known signal and unwanted noise. SNR is expressed in dB and is defined as:

$$SNR = 10\log_{10}\left(\frac{P_{signal}}{P_{noise}}\right)$$

Where P_{signal} is the power of the signal at frequency f_0 and P_{noise} is the power of the noise floor. For purposes of integration, P_{signal} is calculated on a frequency range of

$\pm f_{margin}$, around the peak frequency, f_0. Stotzer chooses a frequency range f_{margin} of $\pm$ 5 Hz in [57]. The noise floor power P_{noise} is then calculated over the remaining frequency range, from 100 Hz to 10 kHz.

Total Harmonic Distortion (THD)

The THD of an audio signal is a measurement of the harmonic distortion present and is defined as the ratio of the sum of the powers of all harmonic components to the power of the fundamental. It is defined as the ratio of the sum of the powers of all harmonic frequencies, f_1, f_2 f_3... f_n , above the fundamental frequency f_0, to the power of f_0 and is computed as follows:

$$THD = 10\log_{10}\frac{P_{harmonics}}{P_{sig}}$$

Where $P_{harmonics}$ is the sum of the powers of the first three harmonics above f_0. The power of each harmonic is calculated in a range of $\pm$ 5 Hz, of the harmonic peak. Stotzer rejects harmonic components above 10 kHz.

A high THD implies that the recovered signal is not purely sinusoidal, and is distorted by higher harmonic components. This implies that distortion of the sinusoid has occurred at either the recording or reproduction stage, possibly by the tracing distortion mechanism.

2.4.2 Ridge-Valley Detection

The purpose of ridge-valley detection is usually to capture the major axis of symmetry of thin image features and is a common task in many image processing applications [65-69]. This differs from edge detection, where the task is typically to estimate the boundary of an object. With grayscale imaging for example, intensity ridges and valleys form medial structures for thin objects [66] such as blood vessels, road networks or fingerprints images.

The topography of grooved media can similarly be thought of as a succession of ridges and valleys. For grooved media, the location of valley structures form the medial

axis of the recorded path, traced by a stylus. The detection and enhancement of these medial lines is therefore important for signal reproduction from data acquired by the NCFSM method.

Haralick [65] describes ridge/valley detection in two ways:

I. The location of zero crossings of the first directional derivative, taken in the direction which extremises the second directional derivative.
II. The position of a ridge (or valley) occurs when there is a simply connected sequence of pixels that have intensities that are significantly higher (or lower) in the sequence than those in the neighbouring sequence.

A comprehensive evaluation of methods for ridge and valley detection can be found in [67].

2.4.3 Smoothing Filters

Data acquired from the NCFSM method typically contains high frequency noise, due to the micro-roughness of the material. To help detect ridge-valley features, a filter can be applied to smooth the data. Data smoothing can be thought of as applying a suitably designed, low-pass filter to a time-series [70]. When smoothing data, it is often useful to replace data points with some local average of the surrounding data. In general, neighbouring points measure closely to the same underlying value; hence appropriate averaging can reduce the degree of noise, without much loss of the overall form of the signal.

Moving Average Filter

The simplest type of smoothing filters is the moving average filter (MA). MA filters replace each data value f_i by a linear combination g_i, of itself and some number of nearby neighbouring points,

$$g_i = \sum_{n=-n_L}^{n_R} c_n f_{i+n}$$

Where: n_L : number of points used 'to the left' of a data point i, (earlier abscissa points)

n_R : number of points 'to the right' (later abscissa points)

c_n : filter coefficient(s)

For an un-weighted MA filter with symmetric window ($n_L = n_R$) the replaced value, g_i is the average of the data points from f_{i-nL} to f_{i-nR} with constant weighting, $c_n = 1/(n_L + n_R + 1)$.

If the underlying function f_i , changes linearly with time, then no bias is introduced as a result of this filtering process because high points at one end of the averaging interval are balanced by lower points at the other end. For functions with non-zero second derivative, a bias *is* introduced, for example, at a local maximum; a moving window average will always reduce the function value.

In the case of impulse (or sharp fluctuation) present in a time-varying signal, the moving window average preserves the area underneath the impulse and its position in time, but the width of the impulse is corrupted. In cases where the width of an impulse is important, simple moving averaging techniques are sub-optimal for use as smoothing filters [70].

Savitzky-Golay Filter

The Savitzky-Golay FIR smoothing filter (SG) or polynomial smoothing filter is useful in applications where it is important to preserve the width of an impulse, such as spectroscopy [71-73]. The idea of SG filtering is to find filter coefficients c_n , that preserve higher moments of the signal, by approximating the underlying function by a polynomial of some order. Filter coefficients are found by a least-square fit, which selects c_n that minimises the total mean-square error.

An SG filter can be applied to a series of uniformly spaced data values $f_i \equiv f(t_i)$. For each point f_i , a polynomial is least-squares fitted to all $n_L + n_R + 1$ points in the window, then g_i is set to be the value of that polynomial at position i (ordinate

value at the centre of the polynomial fit). This process is shown graphically in Figure 2.40, with $n_L = n_R = 3$.

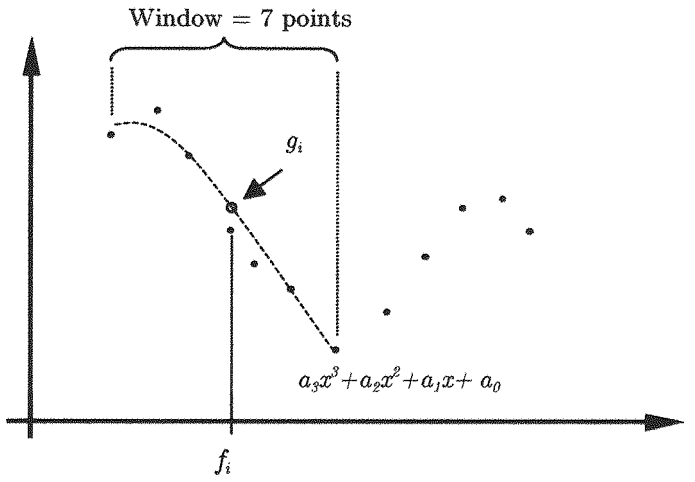

Figure 2.40: Demonstrating principle of SG polynomial filter.

The 'preservation of moments' property of SG filters is useful for applications such as ECG processing [74], where the desired signal has one or more peaks, whose widths must be preserved after smoothing. Equally, they are desirable for enhancing ridge-valley structures, commonly found in linescan profile measurements of phonograph grooves. The ridge-valley structure of linescan data can be thought of as a series of piece-wise cubic polynomials. A cubic polynomial with form $P(x) = a_3x^3+a_2x^2+a_1x+a_0$ can have turning points, when $P'(x) = 0$, corresponding to the position of a local maxima and minima. For linescan data, cubic SG filters are well suited for enhancing of ridges (maxima) and valley (minima) structures.

SG filters derive directly from a particular formulation of the data-smoothing problem in the time or spatial domain only and do not introduce phase distortion. A thorough consideration of SG Filters for signal enhancement can be found in the book by Orfanidis [74]. Some other points to note regarding SG filtering are:

- The value of polynomial is not used at any other point other than at i.
- The frame length must be odd.

- When moving to the next point f_{i+1} , a whole new least-squares fit using the shifted window is calculated.
- All the least-squares fits do not have to be laboriously calculated in turn as described. Since the process of least-squares fitting involves only a linear matrix inversion, the coefficients of a fitted polynomial are themselves linear in the values of the data, and the polynomial coefficients can be computed with a linear filter.
- It was been shown by Bromba et al. [75] that SG filters are approximately optimal smoothing filters for signals that exhibit polynomial characteristics, (such as the ridge-valley pattern of the linescan profile).

SG filters are attractive in the application of groove cross-section enhancement from NCFSM data because their design is determined by geometric heuristics, derived from knowledge of the spacing and complexity of the groove shape.

2.5 Optical Sensing Technology for Surface Measurement

The NCFSM method relies on the ability to measure grooved surfaces to high precision, over a large area. Atomic force microscopy provides very high spatial resolution [76], but is impractical for large specimens, such as phonograph recordings. Instead, the Sound Archive Project has favoured single-point displacement measuring sensors, combined with a raster scanning system to map the full surface topology of a grooved surface. Three optical sensing technologies have been investigated and are now described.

2.5.1 Triangulation Laser (TL)

The triangulation laser measurement principle is as follows: (see Figure 2.41)

- A laser beam is focused onto a target surface at a normal angle to the surface profile of interest.
- The beam reflects off the target surface and passes through a custom designed receiver lens system, positioned off-axis, which collects the diffuse reflections and focuses a beam onto a charge-coupled device (CCD) sensing array, held within the sensor.
- The CCD detects the position of maximum light intensity of the reflected beam, which is representative of the surface height displacement.
- Surface displacements can typically be resolved to less than 1 μm.
- The reflected light used by the sensor requires unobstructed light transmission towards the off-axis receiver lens and detector. This means that measurement of steep surface inclines is problematic, because the surface form can obstruct the path back to the detector.

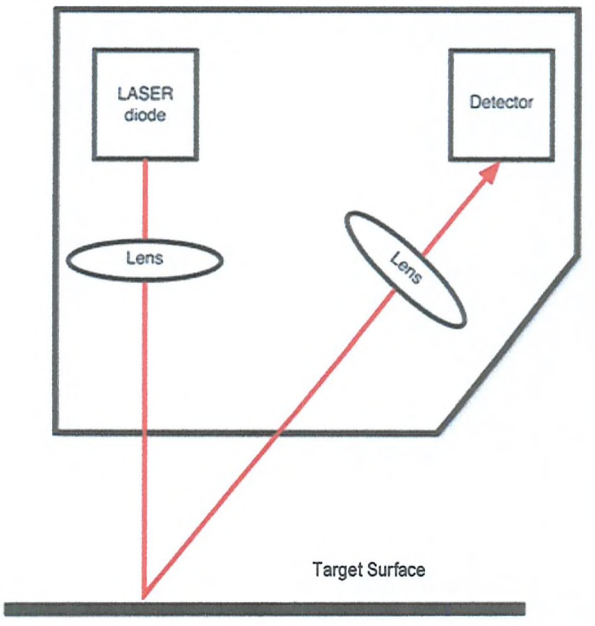

Figure 2.41: The Triangulation Laser principle.

2.5.2 Confocal laser (CL)

The confocal laser principle is shown in Figure 2.42 and is now described:

- A continuous wave red light (λ=670 nm) laser source is passed through a collimator lens to limit divergence of the beam.
- An objective lens focuses this light into a small spot on the target surface.
- The reflected light passes back though the focusing lens and is diverted towards a pinhole using a beam splitter.
- The focusing lens is oscillated at high frequency using a tuning fork arrangement, such that the local point for the light spot caries as a function of the instantaneous tuning fork position.
- Surface height is then inferred by detecting the lens position at the point of maximum received light intensity, which is cross-referenced with a calibrated height measurement.

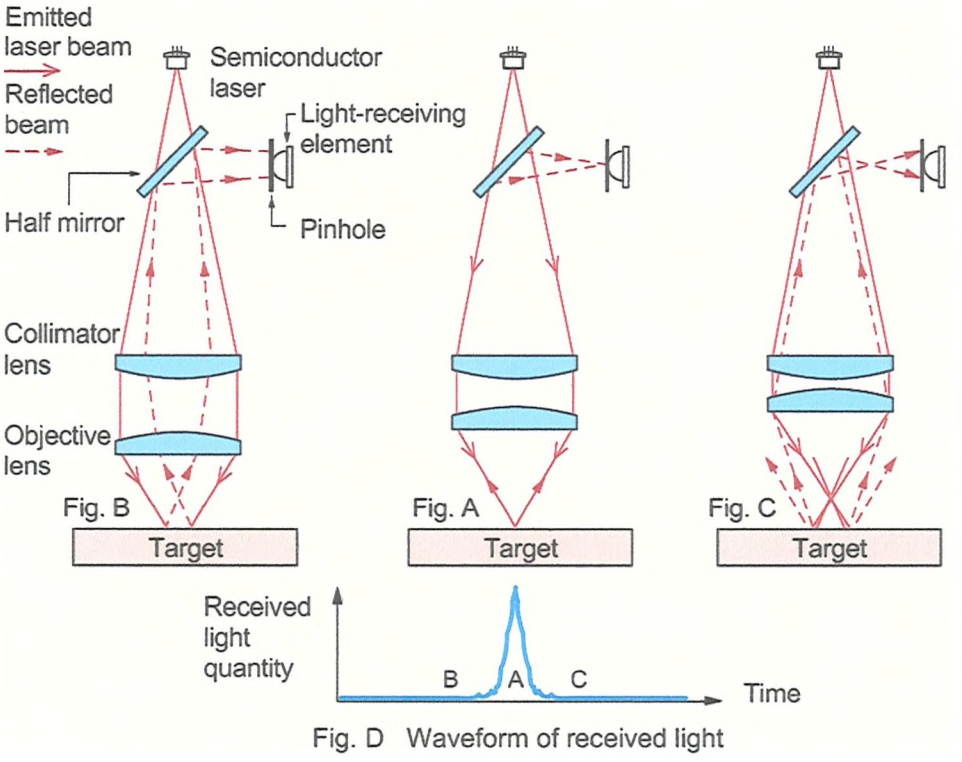

Figure 2.42: The confocal laser principle (Image from [77]).

2.5.3 White Light Confocal Probe (WL)

The WL sensor also operates on the confocal principle (see Figure 2.43), but uses polychromatic (white) light and unlike the CL sensor, does not contain moving optical components. This WL sensor operates as follows:

- A polychromatic light source is focused onto the target surface.
- Lens-induced chromatic aberration causes the light to be split into a continuous distribution by wavelength, whose individual focal points vary with wavelength.
- Reflections from the surface are transmitted through the confocal system to a spectrometer system via fibre optic cable.
- The wavelength corresponding to the highest detected intensity at wavelength λ_M is representative of the distance between the sensor lens and target surface, which is calibrated to relate wavelength to a given surface height.

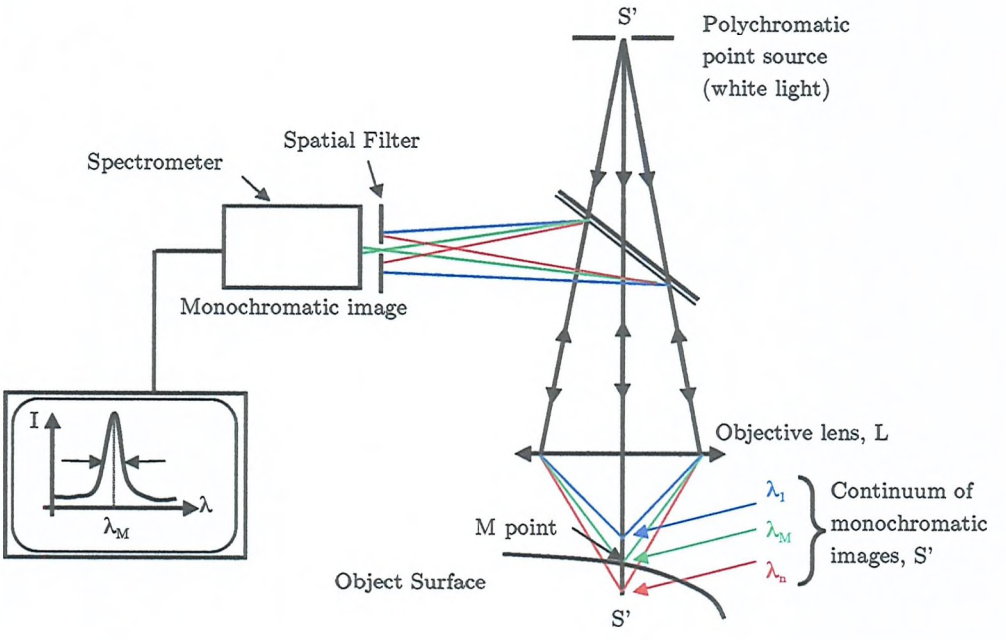

Figure 2.43: White Light confocal measurement principle. (Image adapted from [78])

2.5.4 Benchmarking of Optical Displacement Sensors

In order to assess the quality of surface reproduction and suitability of these sensors (LT, CL, WL), for the Sound Archive Project's NCFSM system, a benchmarking exercise was carried out [76]. Three sensors were compared by taking surface measurements of a machined reference cylinder and a Blue Amberol Edison cylinder. Figure 2.44 gives a comparison of various sensor parameters for the three optical sensors. These parameters are now described:

Spot size

Each sensor has a finite spot size parameter, thus each measurement results from an averaging process of contributions from throughout of the light spot. The displacement measurement is taken for a finite period of time as the sensor moves past the required grid position. The surface area addressed by the sensor during a single measurement is therefore elongated through the movement of the linear stage system.

Sampling Frequency

This is the number of displacement measurements the sensor is capable of taking per second and has a direct bearing on the scan time; hence a higher sampling frequency is preferable.

Gauge range

This is the range over which the sensor can record displacement data, within some fixed standoff distance. Many cylinders recordings are non-concentric due to imperfections in the manufacturing processes and factors such as creep, caused by long-term storage. As a result, the sensor may go out of range, and data points may be missed.

Axial resolution

This is the ability of the sensor to resolve the smallest displacement change. It is equivalent in signal processing terms to the 'bit' resolution of an analogue to digital converter (e.g. 16-bit ADC can resolve an analogue signal at 2^{16} , i.e. 65536 discrete values).

Angular tolerance

The ability of the sensor to measure surfaces incline is defined by the limit of angular tolerance. If the slope of the grooves on a cylinder exceeds the angular tolerance, the sensor is unable to produce a displacement measurement. This poses problems for groove shapes with steep gradients, for example, 78 rpm discs, which have 45°/45° groove walls.

Scan time

The time taken to acquire and store data. This is determined by the sampling increments in x and θ, the speed of the linear stages and sampling frequency of the sensor. Scan times shown in Figure 2.44 correspond to the values using a sampling frequency in this particular study [76], shown in parentheses, with a grid sampling of $\Delta x = 10\mu m$ and $\Delta\theta = 0.1°$.

Sensor	Spot size (μm)	Sample frequency (Hz)	Gauge range (mm)	Axial resolution (nm)	Angular tolerance (degrees)	Scan time for 10mm length (hours)
WL	7	1000 (1000)	0.35	10	27	22
CL	2	1400 (82)	0.6	10	17	160
TL	30	2000 (2000)	10	1000	17	19.5

Figure 2.44: Comparison of sensor parameters used in benchmarking exercise. Sampling frequencies in parentheses represent the effective data acquisition rates by the scanning system, which is affected by sensor averaging processes.

Sensor Choice for NCFSM Method

Figure 2.45 shows linescan and surface measurements taken from the same region of the Edison cylinder examined in [76]. Due to the TL sensor's large spot size and insufficient axial resolution the linescan shows a noisy groove structure. The TL sensor was found to have neither the resolution nor the accuracy required for sound reproduction and was disregarded as a viable sensor. The WL and CL measurements are very similar, showing a groove structure that is smooth, and periodic, from which sound reproduction was possible. The CL sensor was able to resolve the groove structure sufficiently, but only when operated at very low sampling rates and was hence disregarded due to longer scan times. The WL sensor offered the highest angular tolerance (27°), best axial resolution (10 nm) at high sample rates (1 kHz) and was deemed subjectively to produce the highest quality sound reproduction over the CL

sensor. The benchmarking exercise concluded that for typical cylinder recordings, the confocal white light probe (WL) was the preferred sensor of choice.

The recovered audio signal from a Blue Amberol cylinder was also examined in order to determine the axial resolution requirements for resolving the smallest amplitude modulations, likely to be encountered on typical cylinder recordings. Frequency analysis (see Fig. 2.46) of a quiet section from a Blue Amberol cylinder measured by the WL sensor showed that groove displacement amplitudes attributed to vocal sounds in the 2.4-3.2 kHz range were less than 50 nm in places. The 10 nm axial resolution offered by commercial WL sensors can reach this requirement for resolving the sub-50nm amplitude spectra observed for a Blue Amberol cylinder [79].

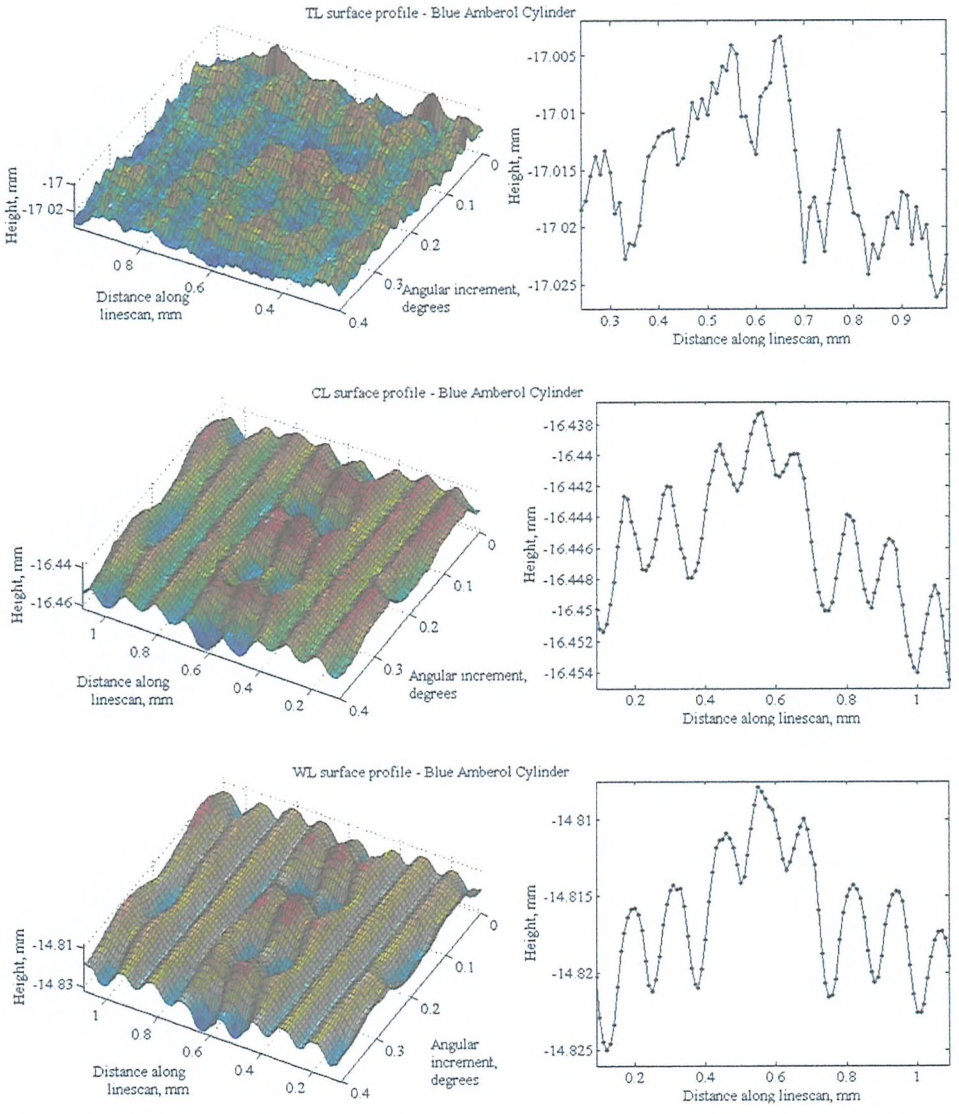

Figure 2.45: Surface and Linescan measurements from Edison cylinder for the three investigated sensors (Image from [76]).

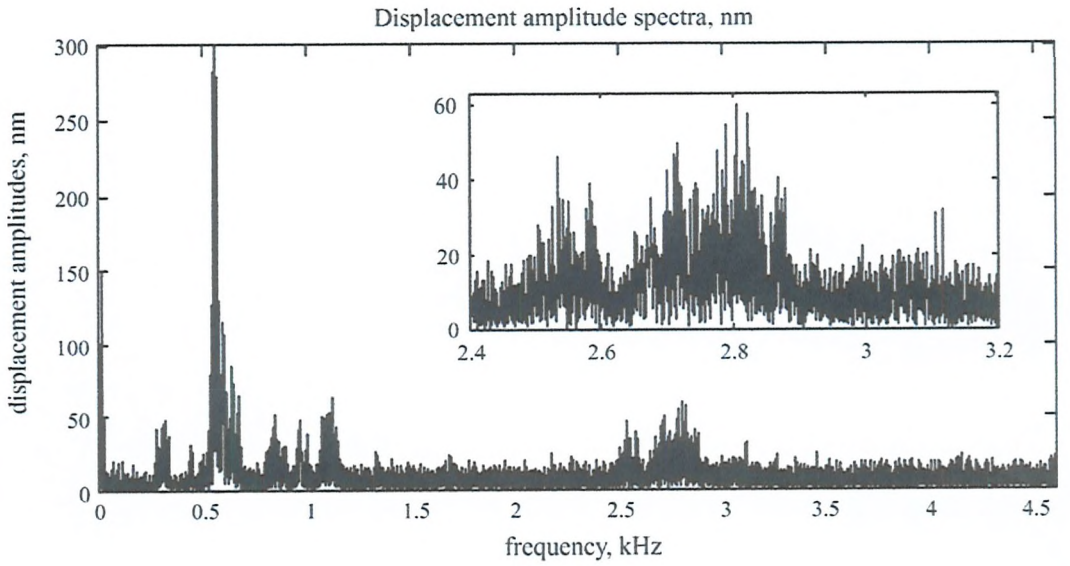

Figure 2.46: FFT analysis of sound recovered from a Blue Amberol cylinder, measured by WL sensor. Insert plot shows a zoomed view of the 2.4-3.2 kHz range, attributed to vocal content. Observed displacement amplitudes in this range are around 50nm (Image from [76]).

2.6 Chapter Summary

In this chapter, the technology required to record and reproduce sound on grooved recordings has been discussed. Modern stylus playback systems are capable of archiving the majority of mass-produced recordings with very high audio resolution (192 kHz/ 24-bit). There are cases however where stylus playback is not an option, for example, playback of broken cylinders/discs, and early tinfoil recordings. An optical playback system for archival purposes is therefore highly desirable.

Several research groups have investigated methods for reproducing sound optically. Methods of optical reproduction in situ (ORIS) are not suitable for archiving damaged recordings, due to the lack of control when tracking complicated groove structures. The advantage of the non-contact full surface mapping (NCFSM) method with signal reproduction in post-processing, is the ability to control more precisely where one places the 'virtual stylus'. This means that in theory, signal reproduction from the NCFSM method can be far more accurate, and can be applied to more recording specimens, than ORIS methods.

Fadeyev et al. [5] presented a signal reproduction method from cylinder recordings using the NCFSM method, but only for specimens in 'good' surface condition. The use of their quadratic fitting method is unlikely to be optimal for cylinders with grooves in poorer surface condition, or those that are cut with non-parabolic shaped cutting tools. **One of the objectives in this thesis is therefore to develop robust techniques for tracking of grooves in poor surface condition, for specimens any groove shape type.**

In this chapter, grayscale imaging methods were also described. The VisualAudio and IRENE systems described are capable of reproducing the signals from laterally cut discs with a playback sample rate greater than 96 kHz (which is the recommended sample rate for archival purposes by IASA). The surface area occupied by 1 pixel with the grayscale imaging systems discussed in this chapter is typically 0.25 – 1 μm in size. Grayscale pixel values relate to the reflected light intensity and not absolute surface depth measurements. As such, grayscale imaging methods are not suitable for reproducing signals from vertically cut groove recordings, such as wax cylinders or tinfoils. This limitation is the main reason why the Sound Archive Project has chosen to develop an absolute depth measuring system. Such a system has the potential for archiving *any* grooved recording - be it vertically or laterally recorded. **Another objective of this thesis shall therefore be to develop signal reproduction methods from discrete surface maps of both vertically and laterally encoded recordings, measured by the NCFSM system.**

Stotzer et al. [54,57] described methods for quantifying the signal output of their grayscale imaging method for disc reproduction, in terms of signal-to-noise ratio (SNR) and Total Harmonic Distortion (THD). To date, there have been no attempts to quantify the output of a NCFSM system for cylinders. **The ability to quantify signal reproduction from data produced by the NCFSM system developed by the Sound Archive Project provides another objective for this thesis.**

In general, grooved recordings are made up of a series of ridges and valleys, hence the detection of these features provides one of the primary tasks in post-processing. The Savitzky-Golay (SG) polynomial-smoothing filter has been identified as

a useful tool for enhancing and detecting ridge-valley structures, due to its ability to preserve the width and position of impulsive data. In the case of grooved data, these 'impulses' can be considered as the positions of ridges and valleys.

In terms of frequency response and signal fidelity, early acoustic recordings are considered 'low-fidelity', by today's standards. Despite this, the resolution requirements are very demanding, even for today's optical systems, with cylinder recordings requiring groove displacements to be resolved to 10 nm or less. For flat disc recordings, the measurement requirements are even more demanding than for cylinders, due to larger surface area, lower playback speeds and greater frequency response. In this chapter, three competing sensor technologies were compared. It was concluded that the White light (WL) sensor provided the most suitable sensing technology for the Sound Archive Project. **A final objective for this thesis is to consider noise or distortion introduced by the measurement process, and to incorporate methods for overcoming them in post-processing.**

Chapter 3

Surface Measurement

3.1 Introduction

Prior to audio signal reproduction, the recording specimen is measured by a non-contact, full surface mapping method (NCFSM). The full surface topology of the specimen (be it cylinder, or disk), is mapped via single-point sensor, combined with a raster scanning system, which results in a discrete representation of the continuous surface. It is the role of this chapter to present details of the NCFSM method, and its implications for audio signal reproduction.

The sensor technology available for surface measurement was discussed in Chapter 2. It was concluded that the white light (WL) sensor was most suitable for the NCFSM system, due to its ability to resolve sub-micron groove modulations typically associated with the sound carrying groove. In this chapter, the two WL sensor systems for measurement of cylinders and flat media are described. Particular attention is given to the measurement and resolution requirements for sound reproduction. The naming conventions of data sets described throughout this thesis are introduced. Sensor parameters such as axial resolution, gauge range and data acquisition rates are considered. Some inherent implications and limitations of the NCFSM method are also discussed. All measurement systems introduce some form of unwanted noise or distortion to the desired signal. For this reason, sources of noise and distortion associated with the NCFSM method are discussed and quantified, again, with respect to audio signal reproduction.

3.2 Cylinder Measurement System

3.2.1 Methodology

Research carried out in collaboration with TaiCaan Technologies and the University of Southampton, resulted in the development of a measurement system for discrete surface mapping of cylinder recordings. Figure 3.1 shows a Black Edison cylinder mounted onto the measurement system, with the three measurement axes depicted. A schematic of the cylinder system is shown in Figure 3.2.

The cylinder specimen is mounted onto a mandrel, coupled to a rotary stage, which allows for rotation in the θ-axis. A second, two-axis (x-z) linear stage system is used to position the sensor. The x-stage traverses the WL sensor parallel to the cylinder's x-axis, recording z-values (depth) at discrete x-positions, triggered by the motion controller. The z-stage, is fixed orthogonally to the x-axis and controls the sensor's stand-off distance from the specimen. This can be adjusted to ensure that the surface remains within the sensor's gauge range. The sensor maps the continuous cylinder surface $Z(x,\theta)$, via a succession of linear scan measurements (linescan), at incremental azimuth angles. Each measurement in $Z(x_i,\theta_j)$ results from an averaging process of contributions from throughout the white light spot, which has a finite size 7 μm diameter.

Due to the number of samples required for the surface scans, it is impractical for the linear stages to stop at each raster point so the displacement measurement is triggered for a finite period of time as the sensor moves past the required grid position. The surface area addressed by the sensor during a single measurement is therefore elongated through the stage's movement. Appendix B gives the specifications of the cylinder measurement system developed for the Sound Archive Project.

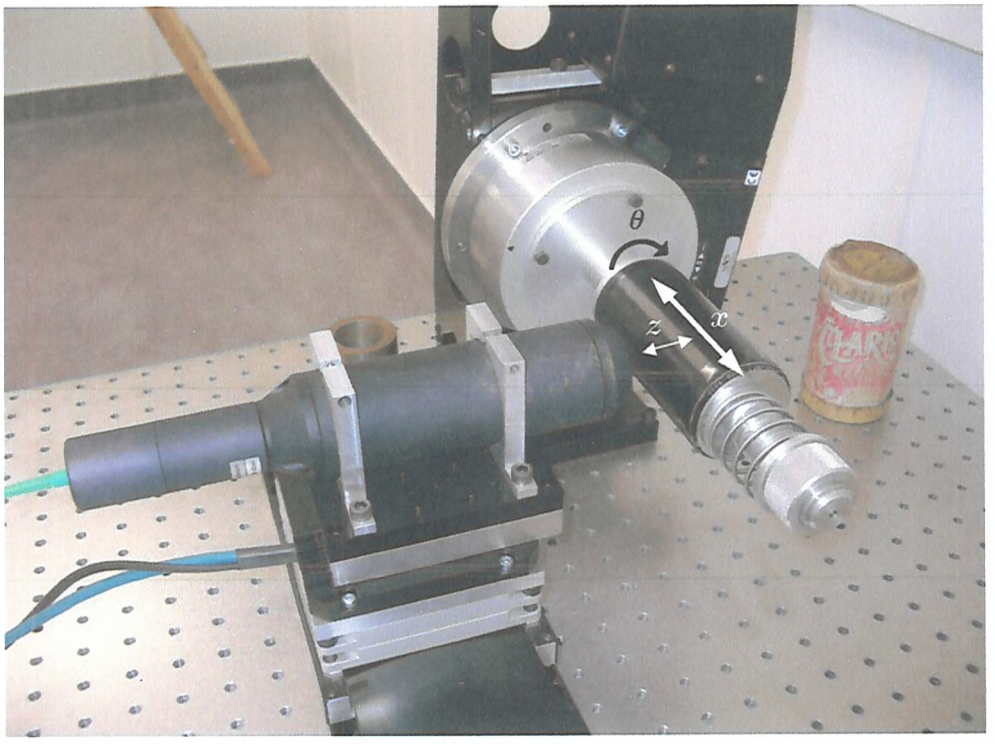

Figure 3.1: Cylinder mounted on mandrel showing three measurement axes (x, θ and z). The WL sensor is mounted upon x-y translation stages. Linescan measurements are recorded by traversing the linear stage parallel to the cylinder along the x-axis. Sensor stand-off distance can be adjusted by advancing the sensor platform in the radial (z) direction. The rotary stage increments the cylinder's rotational position (θ) between linescans.

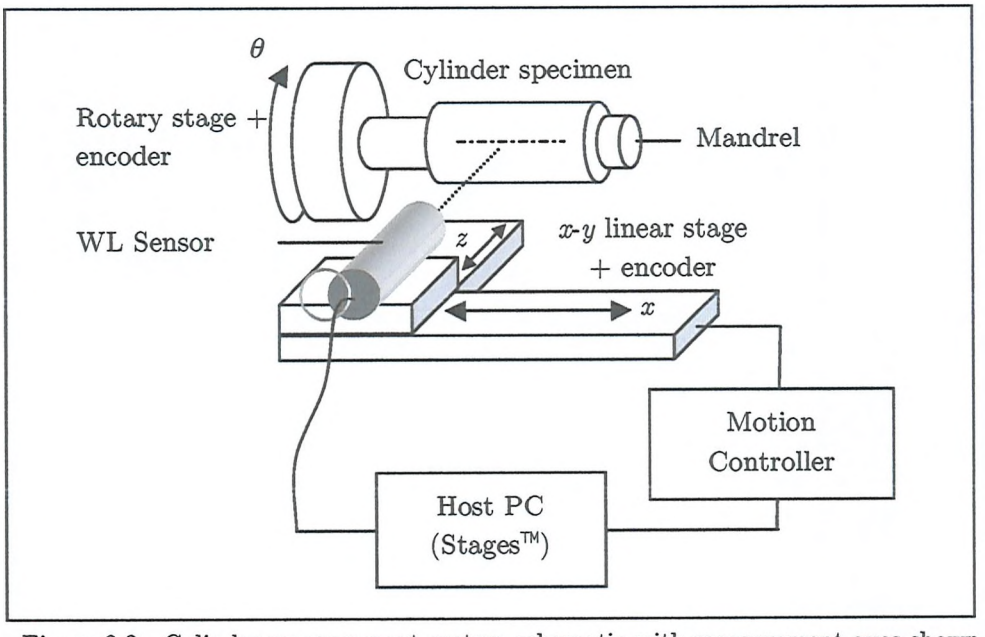

Figure 3.2: Cylinder measurement system schematic with measurement axes shown.

3.2.2 Data Definitions

The terminology and definitions of data acquired by the NCFSM methods and described throughout this thesis is now given:

Linescan, $z_j(x)$

The j-th *linescan*, denoted by $z_j(x)$, is the linear scan of relative surface heights (mm) measured along the x-axis at fixed rotational index: $j = [0, 360°]$. Figure 3.3(a-b) show example linescan measurements from two different cylinder specimens.

Measurement grid increments, Δx, $\Delta \theta$

The discrete grid over which the cylinder is mapped is defined by the two spatial increments, Δx (in microns) and $\Delta\,\theta$ (in degrees), which are the linescan and rotational increments of the motion system respectively. This measurement grid is shown in Figure 3.4.

Discrete Surface, $Z(x_i,\theta_j)$

We denote the discrete surface by $Z(x_i,\theta_j)$, which is a matrix of z-values (surface heights), measured at grid positions, x_i and θ_j. The discrete surface is constructed by concatenating a succession of linescans. Figure 3.5 shows a discrete surface measured from a Blue Amberol cylinder.

Discrete Surface Patch, $Z_N(x_i^N,\theta_j^N)$

The full cylinder surface is typically measured in N surface patches (discussed in Section 3.2.4). A scanning segmentation script produced prior to full surface measurement gives the positional grid indices, x_i^N and θ_j^N for the N surface patches. The N-th surface patch is therefore denoted by $Z_N(x_i^N,\theta_j^N)$.

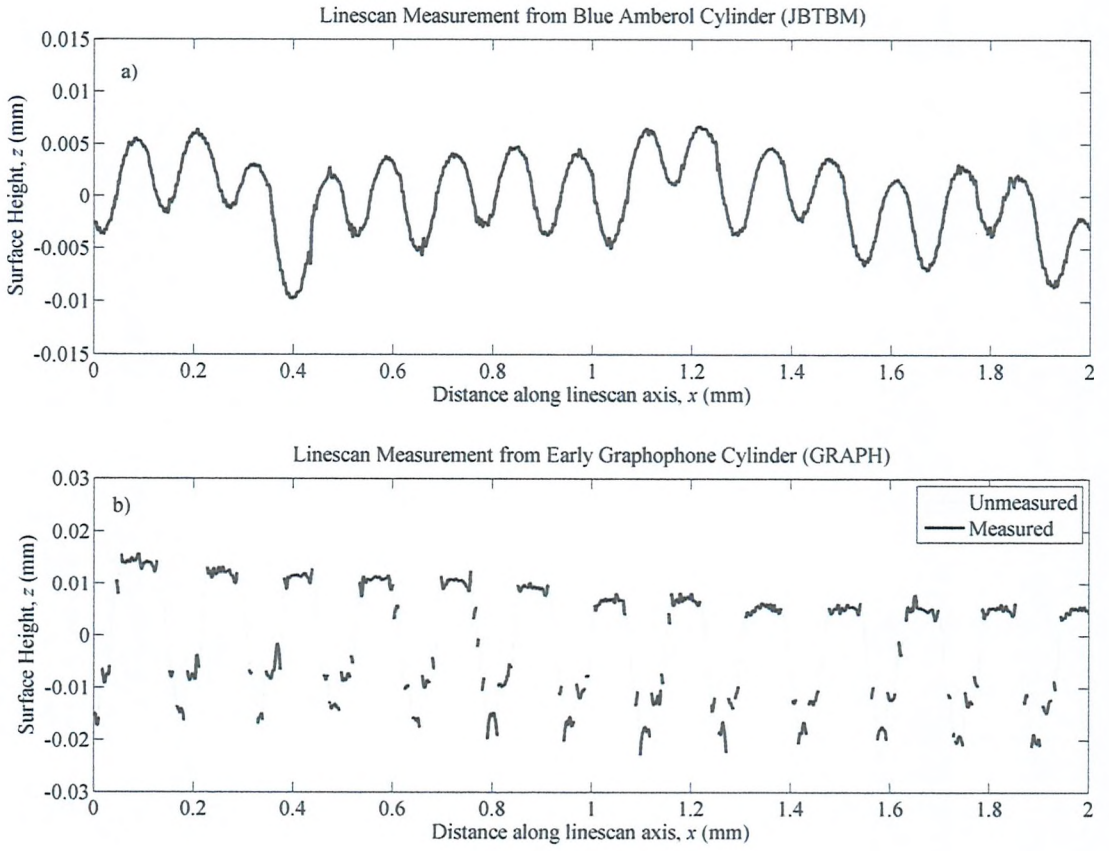

Figure 3.3: Linescan measurements from a) Blue Amberol (JBTBM) and b) early Graphophone cylinder (GRAPH). Profile length in both examples is 2mm. Note that due to the steep incline of the groove cross-section in b), much of the data at the side wall is unmeasured by the sensor. Unmeasured data has be linearly interpolated for visualisation purposes.

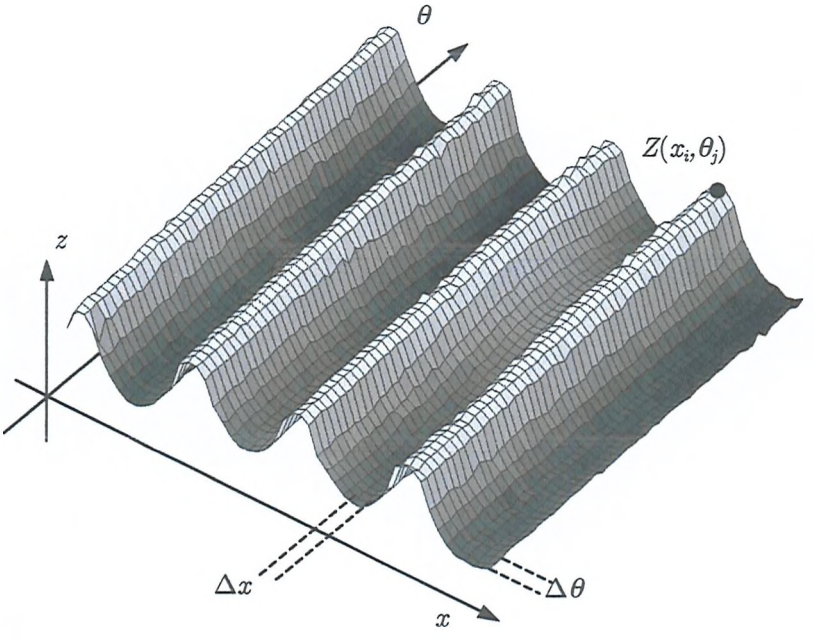

Figure 3.4: The discrete measurement grid, defined by the increments, Δx and $\Delta\theta$, over which the continuous surface is discretised. Four cylinder grooves are shown here.

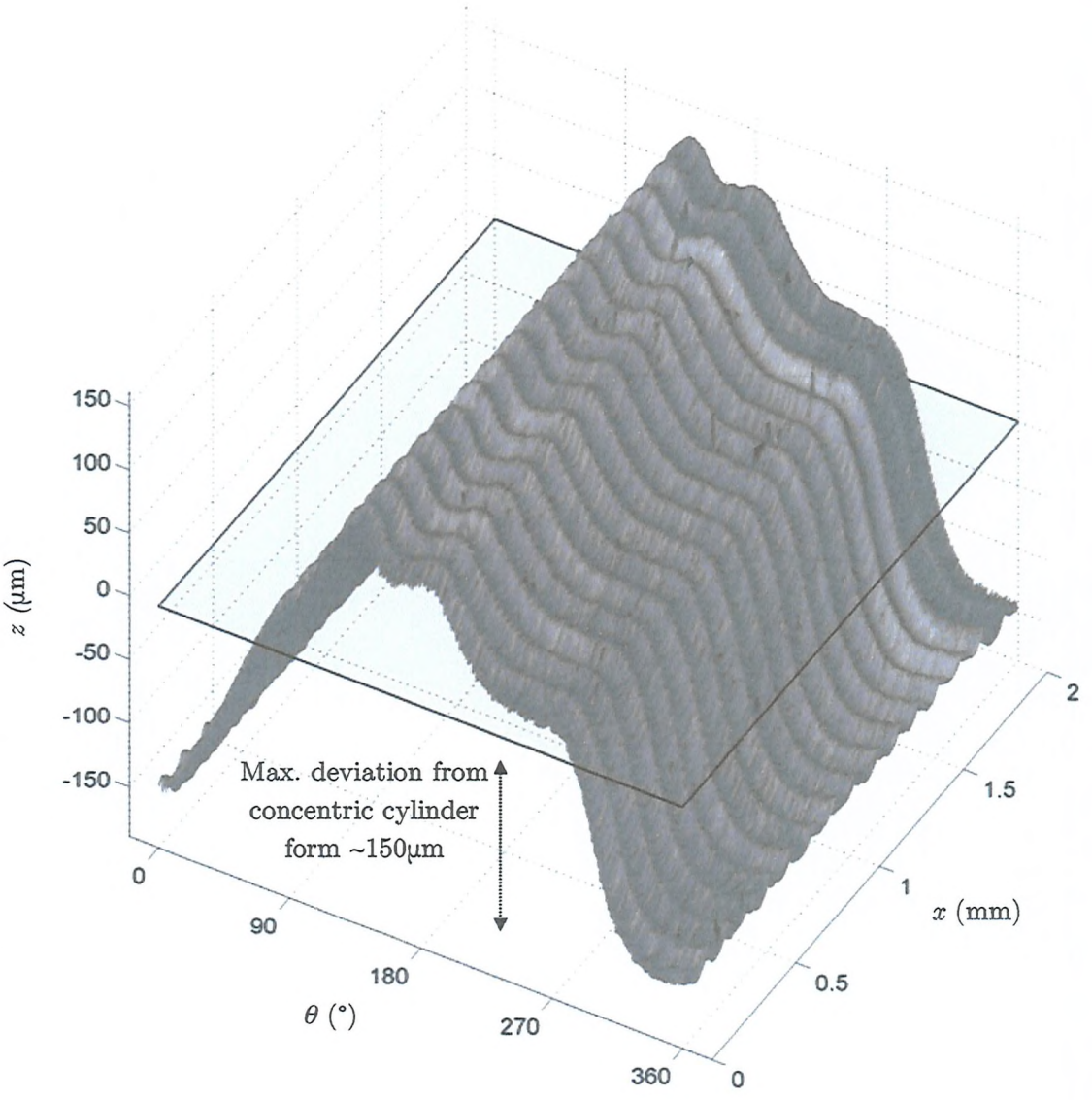

Figure 3.5: Discrete surface plot of 2 mm segment from a Blue Amberol cylinder (JBTBM). This particular cylinder is non-concentric (a perfect cylinder would assume a flat form at fixed z-offset).

3.2.3 Measurement requirements for Cylinder Reproduction

To ensure that sound can be accurately recovered from a cylinder, the sampling resolution of the measurement system must be considered, to avoid aliasing and to optimise scanning time. We consider the linescan, circumferential and axial resolution of the measurement system, and their effect on audio signal reproduction.

Linescan Sampling, Δx

The choice of Δx determines the number of samples over which the groove cross-section is resolved along the linescan dimension. This is dependent upon the groove pitch λ_x , and complexity of the groove cross-section. Strictly, to avoid aliasing of the fundamental groove structure, with spatial wavelength λ_x, the following equality (Nyquist limit) must be satisfied:

$$\Delta x < \frac{\lambda_x}{2} \tag{1}$$

In order to resolve spatial wavelengths less than λ_x (damage structures for instance), the choice of Δx must be much lower than (1) suggests. For a cylinder recorded with 200 tpi, a linescan grid spacing of $\Delta x = 10$ µm, gives approximately 13 samples per groove cross-section. This sampling has been previously judged to provide sufficient resolution for sound reproduction [5].

Circumferential Sampling, $\Delta\theta$

The choice of $\Delta\theta$ (in degrees) determines the playback sample rate f_s , for the recovered audio signal. Assuming that the cylinder was recorded with a constant surface speed of ω_{rpm} (revolutions per minute), and the full cylinder circumference is sampled with $\Delta\theta$ (°) the estimated playback sample rate (in Hertz) is given by:

$$f_s(Hz) = \frac{6\omega_{rpm}}{\Delta\theta} \tag{2}$$

Strictly, to avoid aliasing of the audio signal by the Nyquist criterion, we have:

$$f_s > 2f_{max} \tag{3}$$

where f_{max} is the maximum frequency (Hz) of recorded information on the recording. For cylinder recordings, this upper limit is around 5 kHz. For 160 rpm cylinders, an angular sampling of $\Delta\theta = 0.1°$ gives a playback sample rate of 9.6 kHz, thus by the Nyquist criterion, this spatial sampling is insufficient for reproduction of frequencies

above 4.8 kHz. The technical committee of the International Association of Sound and Audiovisual Archives (IASA) recommend a minimum sample rate of 48 kHz, at 24-bit word length for digitizing analogue recordings [33]. With the non-contact system, a playback sample rate of 96 kHz is achievable with an angular sampling of $\Delta\theta = 0.01°$ (for cylinders recorded at 160 rpm). Cylinders recorded at lower speeds (e.g. Tinfoil, early Graphophone), of 60 – 100 rpm will require $\Delta\theta < 0.01°$ to achieve the same playback sampling frequency. A table of standard playback sample rates and values for $\Delta\theta$ is shown in Figure 3.6.

Sampling Scheme $\Delta\theta$ / N **linescans**	**Cylinder recording speed and estimated playback sampling frequency for given sampling scheme**		
	60 rpm	**100 rpm**	**160 rpm**
0.1° / $N = 3601$	3.6 kHz	6 kHz	9.6 kHz
0.05° / $N = 7201$	7.2 kHz	12 kHz	19.2 kHz
0.02° / $N = 18001$	18 kHz	30 kHz	48 kHz
0.01° / $N = 36001$	36 kHz	60 kHz	96 kHz

Figure 3.6: Table of estimated playback sample rates for four standardised circumferential sampling grids, for various cylinder recording speeds.

Dual consideration of Δx and $\Delta\theta$

For one complete cylinder rotation, the x-position of a single groove moves through a distance of λ_x. Assuming a perfect helical trajectory it is possible to estimate the x-distance ε_x, a groove will have shifted along the x-axis between consecutive linescans, based on the selection of circumferential increment, $\Delta\theta$ (in degrees) by:

$$\varepsilon_x = \frac{\Delta\theta \cdot \lambda_x}{360} \qquad (4)$$

For a groove pitch of λ_x =127 µm [λ_x = 254 µm] this shift is ε_x = 35 nm [70 nm] for $\Delta\theta = 0.1°$ and ε_x = 3.5 nm [7 nm] for $\Delta\theta = 0.01°$. These values are small compared to the x-sampling grid typically used (10 µm).

Axial resolution of Sensor, Δz

The third sampling dimension is the axial resolution of the sensor, Δz which is the ability of the sensor to resolve displacements in the z-axis. It was shown previously in Chapter 2.5.4, that in order to resolve the smallest displacement amplitudes (high frequency, low amplitude) for typical cylinder recordings, an axial resolution in the order of 10 nm is required. For vertically modulated grooves, Δz determines the dynamic range of the digital audio signal. Current hardware permits a measurement dynamic range of ~15-Bit (3.5 x 10^4 discrete z-values), over the sensor's full gauge range (0.35 mm), at a fixed offset distance (note that this offset distance is generally adjusted during measurement, in order to account from surface eccentricity). In order to relate Δz to the dynamic range of the digital audio signal, the range over which the groove depth modulates in the z-axis, must be known. A typical range value for a cylinder described later in this thesis is 10 µm for a 1 kHz tone, measured from peak-to-trough. The WL sensor has an axial resolution of Δz = 10 nm, which in this case gives an equivalent digital audio bit-depth of just under 10-bit (1,000 discrete z-values).

3.2.4 Surface Scanning Segmentation Strategy

In many cases of high resolution imaging of large areas, it is common for a complete scene to be composed of a number of different measurements, taken by multiple sensors and/or at multiple locations. Multiple scene reconstruction introduces an extra complexity in post-processing, but does improve data handling and saving of large data sets. The main reason for requiring multiple scene reconstruction with the NCFSM method is due to a limitation of the optics – namely the sensor's gauge range.

At a fixed standoff distance, there is a finite range (gauge range) over which the sensor can resolve displacements. The groove depth of cylinder recordings is typically 15 – 30 μm, which is well within the operating gauge range of the chosen white light sensor, (350 μm, see Appendix B). However, due to the effects of poor storage, warping or creep, the macroscopic form of cylinder recordings is generally non-concentric. This lack of concentricity has implications for the scanning process.

In order to measure the full surface topology of non-concentric cylinders, the stand-off distance must be adjusted throughout the scanning process, in order to keep the surface within the sensor's gauge range. Three scanning strategies have been considered [79]:

i. Optimum sensor standoff distance adjustment based on preceding linescan z-range,
ii. Adaptive control of the sensor stage to track the surface, ensuring that the surface remains in operating range of the sensor,
iii. Scan surface in a series of segments, which can stitched together in post-processing, with sensor standoff distance fixed for a given segment.

Strategy (i) was shown to be successful in some cases, however for certain cylinders (e.g. Blue Amberol), the variation in macroscopic form along a given linescan exceeds the sensor's gauge range, which makes it sub-optimal for full surface preservation. Misalignment of the cylinder on the mandrel can also cause the specimen to out of range of the sensor.

Strategy (ii), was also avoided in the system design because the movement of the z-axis stage has the potential to influence measurement accuracy, as a result of stage position uncertainties.

Instead, strategy (iii) was chosen, whereby the cylinder scan is divided into a number of manageable segments, with the sensor fixed at a pre-determined standoff distance. An initial, low resolution 'preview scan' is obtained prior to the high resolution measurement, in order to record the macroscopic form of the

specimen. An optimisation routine then determines the minimum number of segments scans (standoff adjustments) required to map the full surface.

For a cylinder of length L, the optimisation for selecting the minimum number of stand-off adjustments N_s for surface segmentation can be summarised by the following list of constraints:

Minimise N_s, subject to the following constraints [79]:

1. Range of the surface height within a segment all within sensor's gauge range
2. Rectangular segments
3. Equal angular range for a given set of linescan axis subdivisions, to simplify data stitching
4. Each segment must exceed a minimum linescan length, l
5. Segments must exceed a minimum area, to avoid overly complex solution
6. Segments must not exceed a maximum area, for data handling, data saving and to minimise data loss in the event of system failure.

The optimisation procedure operates by dividing the linescan axis at rotational angle θ_a into an integer number of subdivisions m, where m = 1,2,3...$nint(L/l)$. For each subdivision scheme, the angular range is incrementally increased until the constraints are contravened at $\theta_b(m)$, presenting a (1 × m) vector of possible segment widths. The linescan division scheme m that maximises the area per segment for the current nominal angular range is adopted, and thus determines the segment's end point θ_b. The process repeats for rotation angles exceeding θ_b. For a given surface segment, the sensor standoff distance is specified such that the midpoint of the sensor's working range coincides with the mean height of the surface segment.

The likelihood of the surface going out of range is minimised by specifying a conservative gauge range parameter for the optimisation program, for example, by specifying 90% of the actual sensor's gauge range. Figure 3.7 demonstrates the

preview scan used to determine the optimal sensor standoff distance adjustments for each discrete surface patch.

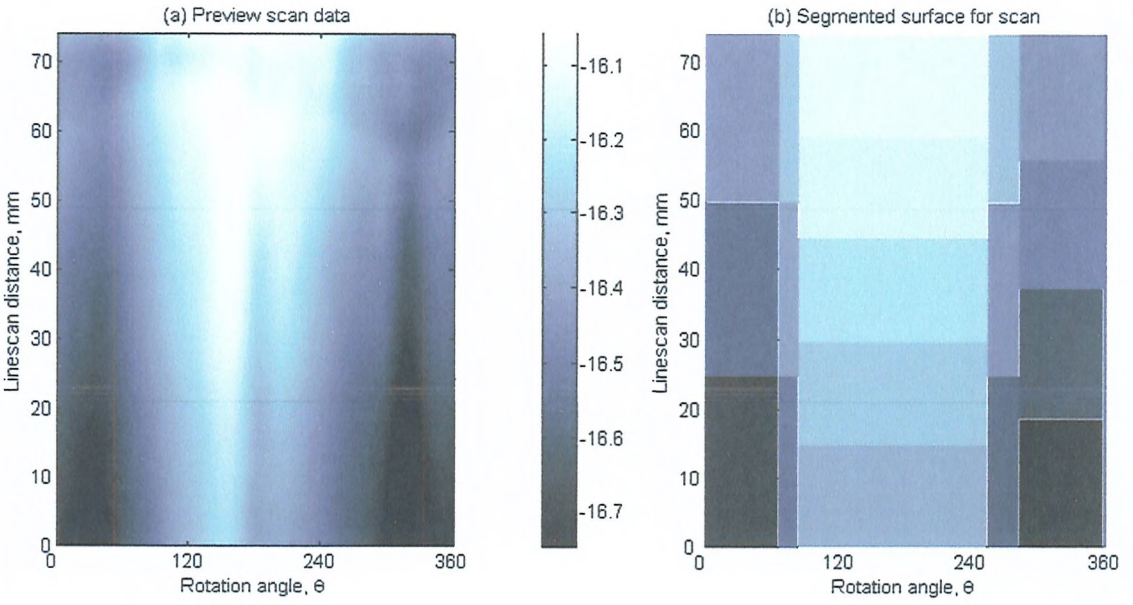

Figure 3.7: (a) Low spatial resolution preview scan of cylinder used to produce segmented scanning shown in (b), optimised based on sensor's gauge range and minimising the number of sensor standoff distance adjustments. Each rectangle in (b) represents a discrete surface patch area.
Note: This figure (from [79]) was taken at an early stage of the segmentation strategy. The final strategy developed ensured strips of equal x-width around the circumference, as shown in Fig. 3.8.

Errors with Segmentation strategy

The scanning optimisation routine generally ensures that the sensor remains within operating distance of the segment being measured. However, in some cases, there is noticeable data loss towards the edges of the segment boundaries (see Fig. 3.8 and 3.9). This data loss occurs when a surface is locally more eccentric, and the range threshold set following the preview scan of the optimisation routine is no longer sufficient. The current measurement system does not include a data integrity check to mitigate for such circumstances. In these cases, data will need to be re-scanned using a new stand-off distance, or new segmentation strategy.

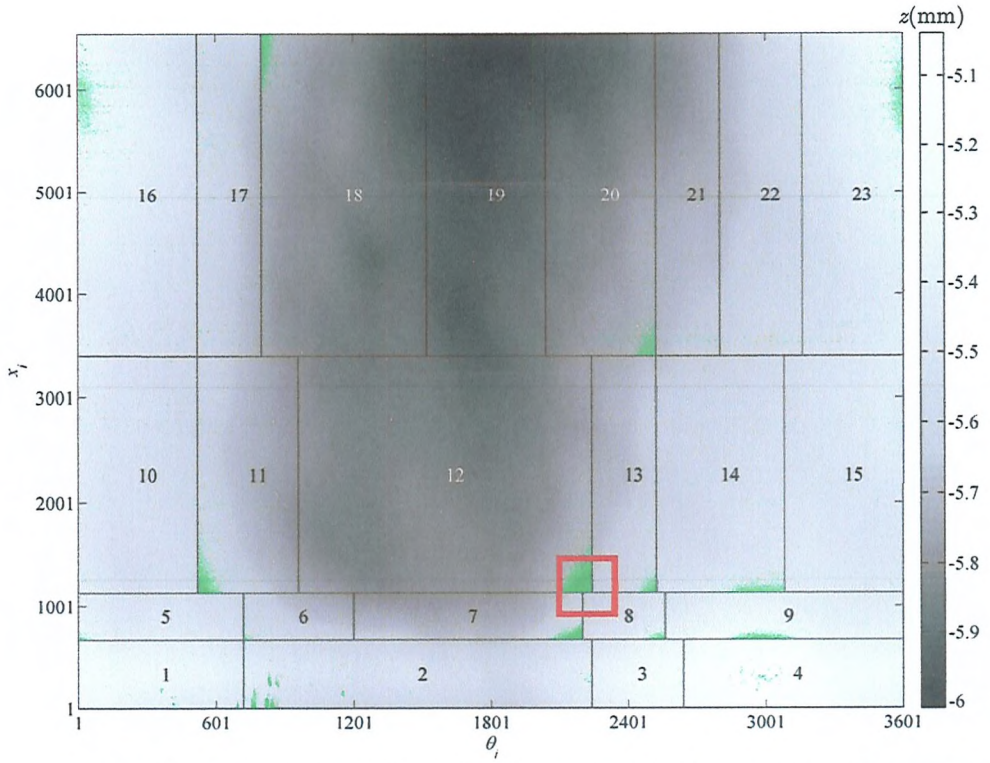

Figure 3.8: Wax cylinder surface (IRISH) acquired in 23 segments, with segmentation map overlaid. Green regions denote positions of missing data, where the sensor has gone out of range. Data loss generally occurs towards the segment boundaries. The red box highlights a large amount of data loss, at the joins between segments 12-13.

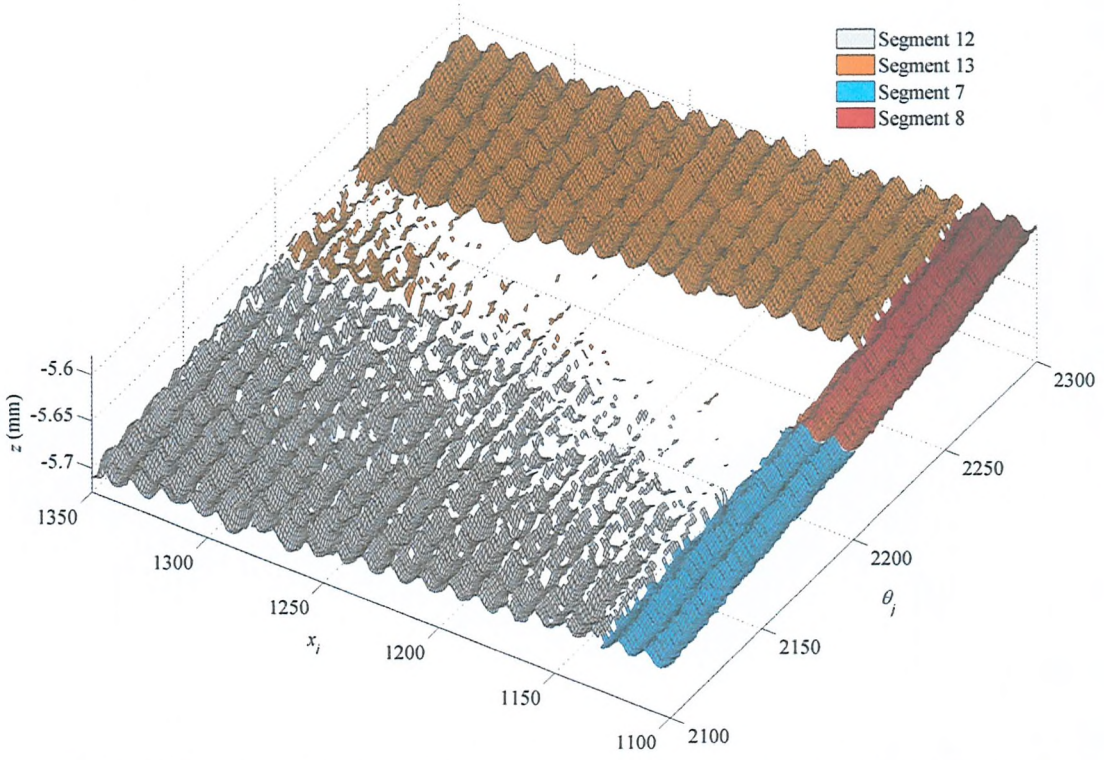

Figure 3.9: 3D surface plot of red boxed region, shown in Fig. 3.8, showing increased data loss at segment joins, due to sensor out of range issues.

3.2.5 Discussion of Scanning Method

Choice of Linescan Strategy

Scanning across the grooves in a linescan fashion, as opposed to mapping the cylinder via a succession of circumferential sweeps, may seem counter-intuitive, because the sound is primarily encoded in the θ-direction. However, a linescan strategy has the following advantages for reducing scan times:

I. **Sensor gauge range** – it has been found that the z-range over which the surface deviates from a perfect cylinder is less along the linescan (x-axis), than for a circumferential sweep. This means that the sensor stand-off distance does not have to be updated as frequently for linescans measurements when compared to circumferential sweeps, to ensure the sensor remains in range. This results in reduced scan times.

II. **Speed of linear stages** - The x-y stage can operate at a higher speed when compared with the rotary stage, which reduces scanning time.

Constant Measurement Grid Assumption

Throughout measurement, a constant grid spacing is assumed in both x and θ axes. This means that only the z-matrix (depth) and measurement grid *vectors*, x_i and θ_j need to be stored. This reduces data rates by two thirds, and allows for operations which require a constant grid spacing, without the need for interpolation.

3.3 Air Bearing System for Flat Media

A second measurement system was developed for full surface mapping of flat media, such as discs recordings and unravelled tinfoil sheets. The air-bearing system which was custom-made for the needs of the Sound Archive Project, is similar to that used for silicon wafer inspection in the semiconductor industry. The sensor is mounted on an overhead gantry system pointing downwards onto the disc, with vertical standoff

distance controlled using a high resolution linear z-stage. The air-bearing system specification is given in Appendix C. Figure 3.10 shows this air-bearing system with its measurement axes depicted.

Figure 3.10: Air-bearing system shown here measuring a 78 rpm disc.

3.3.1 Methodology

The principle of measurement and use of WL single-point sensor is similar to that of the cylinder system. For discs, a discrete surface is built up by a succession of radial linescans measurements, taken from the disc centre toward its outer circumference. The x-axis is taken from the disc origin (centre) along its radius. The disc is rotated via a rotational stage, in increments of $\Delta\theta$.

Example linescan data from a 78 rpm disc is shown in Figure 3.11 and a surface measurement of a 5mm ring is shown in Figure 3.12 and 3.13. Due to the 45° gradient of the groove wall, the WL sensor is unable to measure the full groove cross-section. Instead, only the groove bottom and tops (land) are measured. A new scanning strategy is proposed for future work, based on orienting the sensor at 45°, (orthogonal to a single groove wall), in order to resolve these features.

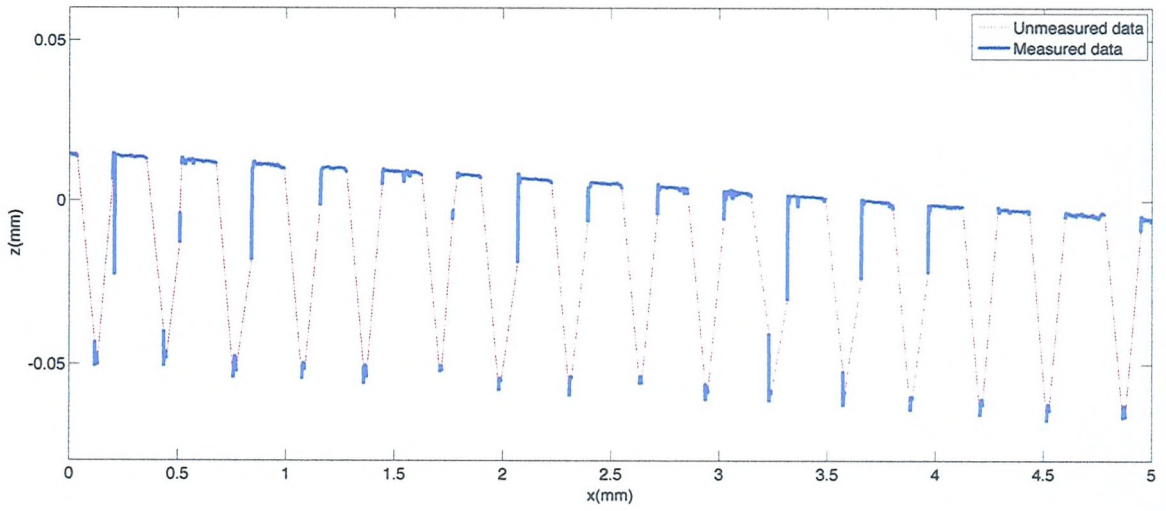

Figure 3.11: Preliminary linescan measurement of a 78 rpm disc. Groove wall data is largely unmeasured by the sensor due to the 45° incline, which exceeds the angular tolerance of the WL sensor.

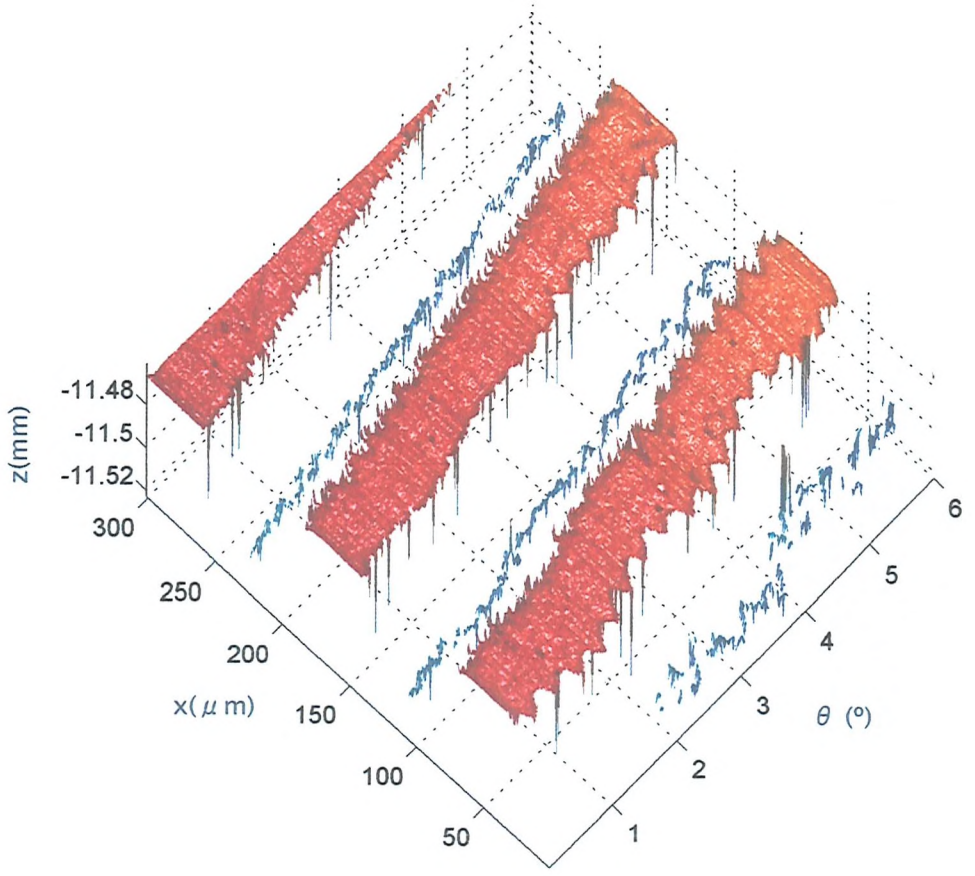

Figure 3.12: Uninterpolated 3D surface plot of 78 rpm disc grooves measured by the WL sensor. In general, the groove tops (land) is well imaged by the sensor. Lateral modulations corresponding to the sound are visible.

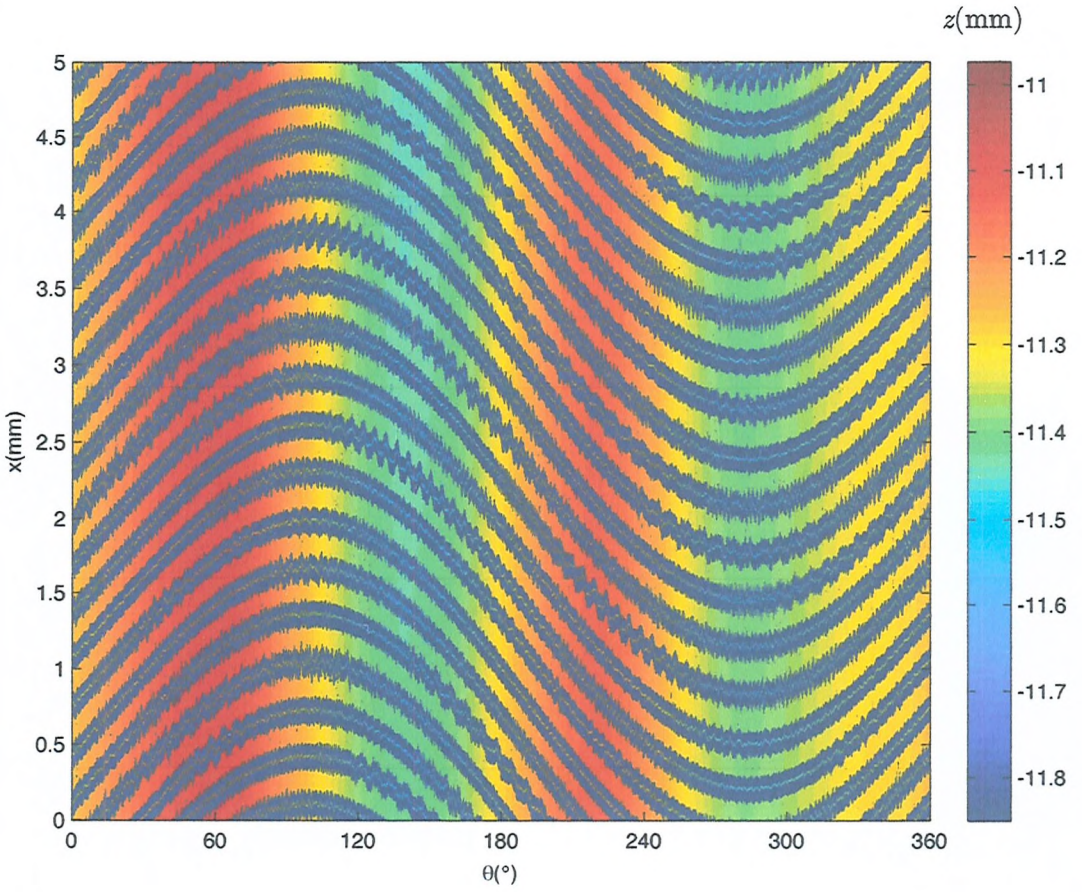

Figure 3.13: Surface intensity image plot of 78 rpm disc. Dark blue data found at the groove wall is unmeasured by the sensor.

3.3.2 Measurement requirements for Disc Reproduction

As with cylinders, it is possible to relate the measurement sampling grid of monaurally recorded discs to the resolution of the output audio signal. Consider a single groove channel modulated only in the radial plane (x-axis). To avoid intercutting between adjacent grooves, there is a maximum groove displacement limit, which is dependent upon the pitch of the groove spiral and width of the cutting tool facet. This pitch for 78 rpm discs, the inter-groove spacing for an unmodulated groove spiral typically 211 – 357 µm.

Fadeyev et al. [6] noted that the maximum groove displacement (maximum displacement from an unmodulated path) for 78rpm flat disc is approximately 100-125 µm. Taking the largest case, the maximum groove displacement (from peak to trough) for a single groove $d_{\max}$ is 250 µm. Assuming that the audio signal is defined purely by the groove modulation in x, the choice of linescan sampling Δx

determines the maximum dynamic range (bit-depth), available from the discrete surface measurement. In order to obtain a 16-bit dynamic range (2^{16} discrete x-values), for a disc with $d_{\max} = 250$ µm, the linescan sampling required is ~$\Delta x = 3.8$ nm.

Appendix C shows that the air-bearing linear stages are capable of providing the sensor with a positional resolution of $\Delta x = 2$ nm, which in theory means that a 16-bit dynamic range is achievable. In reality however, the system resolution is limited by the size of the sensor spot, which for the SAP system developed is 7µm. Attempts to reduce the WL sensor spots size resulted in an increase in missing data, due to a reduction in the reflected light intensity. Figure 3.14 gives a summary of the best dynamic range achievable for monaural discs with $D_{\max} = 250$µm.

Linescan Sampling, Δx (µm)	Number of discrete values over d_{max}	Equivalent dynamic Range
10	25	< 5-bit
5	50	< 6-bit
1	250	< 8-bit
0.1	2500	< 11-bit
0.01	25000	< 15-bit

Figure 3.14: Linescan sampling and equivalent dynamic range values for a 78 rpm flat disc recording, measured using the air-bearing system.

The measurement requirements for flat disc records are more demanding than for cylinders. This is largely due to the fact that the rotational playback speed of flat discs is slower than with cylinders, (33-78 rpm for discs, as opposed to 160rpm for cylinders). A longer duration for one disc revolution means that the angular sampling grid spacing, $\Delta\theta$ must be increased. Figure 3.15 shows a comparison of sampling required to achieve standard playback sample rates.

The upper frequency limit of recorded sound for discs is typically higher than for cylinders (>8 kHz) meaning that a finer angular sampling grid is required.

In addition, the larger grooved surface area of discs (12", 33rpm disc: 0.055 m^2, as opposed to 0.017 m^2 for a typical 4" Amberol cylinder), means that a larger measurement grid is required. These factors combined mean that the memory requirements for storage of disc surface data are much greater than for cylinders.

Desired Playback	Sampling for 78 rpm disc		Sampling for 160 rpm cylinder	
Sample rate, f_s	N linescans	$\Delta\theta$ (°)	N linescans	$\Delta\theta$ (°)
12 kHz	9231	0.039	4500	0.08
24 kHz	18,462	0.0195	9000	0.04
48 kHz	36,923	0.00975	18000	0.02
96 kHz	73,846	0.004875	36000	0.01

Figure 3.15: Comparison of angular sampling requirements for 78rpm flat disc and 160rpm cylinder recording. (N linescans is the number of linescans required for full 360° measurement).

3.4 Noise & Distortion caused by Surface Measurement

This section describes the sources of noise and distortion which can be introduced by the NCFSM measurement system.

General Acquisition Noise

All measurement systems are prone to adding some unwanted noise. Sources of noise include:

- **Thermal noise** – generated by the random movements of thermally energised particles. This noise is associated with temperature-dependent random movements of free particles. Although the average movement of random particles is zero, the fluctuations about this average constitute thermal noise. Thermal noise is generally considered to have a flat spectrum, i.e. white noise. Differences in temperature and air flow in the laboratory can also cause the specimen under test to deform over time.

- **Electromagnetic noise** – any electrical device which generates, consumes or transmits electrical signals are potential sources of electromagnetic noise. Common sources of electromagnetic noise are transformers, A/C power lines and fluorescent lamps.
- **Acquisition Error** – when the sensor is unable to obtain an accurate measurement, the sensor generally returns 'NaN' (not a number), to denote that measurement has failed. On some occasions however, it has been observed (especially at points where surface gradient is close to the angular tolerance of the sensor) that 'false' measurements are returned, as opposed to 'failed' measurements (NaN). These 'false positives' manifest themselves as depth values which are must greater or lower than the surrounding data points.

Imaging of Debris

The sensor measures the 3D surface with respect to a 2-D image plane. Any objects, (e.g. cotton fibres, grit, dust), lying above the image plane will obstruct the underlying groove structure beneath (see Fig. 3.16 for example). This results in a loss of information, as it is impossible to know exactly the nature of the groove below the obstruction. The effect of this for the extracted audio signal will be heard as impulsive noise. It is therefore advisable that prior to measurement, any loose debris lying on top of the grooves is removed via compressed air gun (or otherwise), to ensure that only the groove structure and originally recorded sound is measured by the sensor.

Measurement Drift in z-axis

As previously stated, fluctuations in temperature in the measurement lab can cause the surface to deform slowly with time, causing a drift in the absolute z-measurement. Measurement drift has implications for audio signal reproduction, especially at segment joins. Figure 3.17 shows an example of a cylinder surface

measurement which exhibits measurement drift. In this case, the reproduced signal will experience impulsive noise with every cylinder rotation.

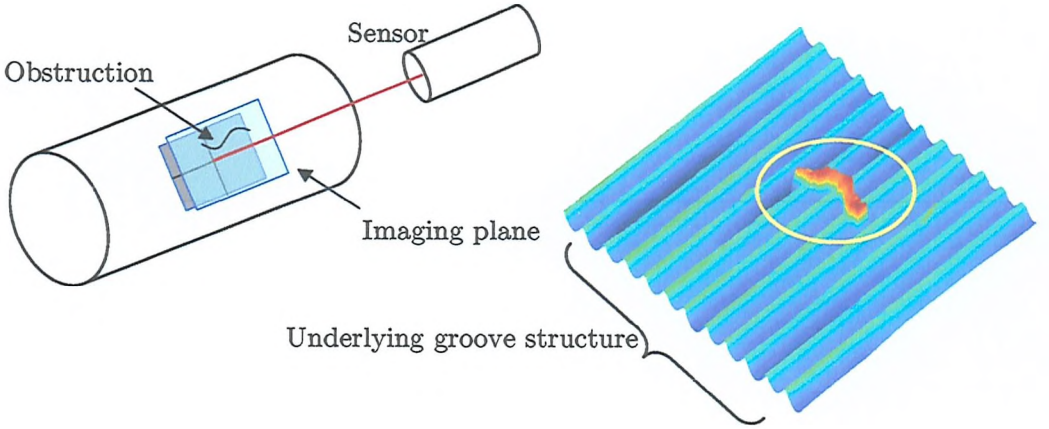

Figure 3.16: Debris lying on the cylinder surface is obstructing four of the underlying grooves.

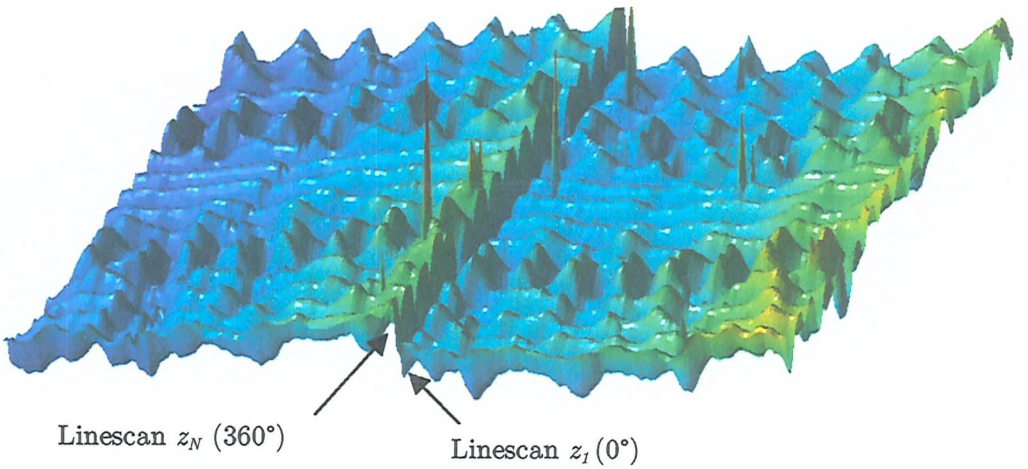

Figure 3.17: Example of measurement drift. A z-offset exists between linescans at 0 and 360°.

Linescan x-offset error

During the development of the measurement systems for the SAP, x-offset errors have been observed between consecutive lines. This is caused by errors in the acquisition software when logging the x-position of the sensor. An x-offset error is seen as a misalignment between linescans (see Figure 3.18b for example). To test for typical linescan offset error of the cylinder measurement system, a repeatability test was carried out. Repeated linescan measurements of a test cylinder (TC2) were recorded over a 12 hour period with a linescan grid spacing of $\Delta x = 1\mu m$. Relative x-offsets were calculated by cross-correlating the 1st linescan with subsequent linescan measurements and detecting the peak lag in x. Figure 3.19 shows the linescan offset error in x, recorded over a 12-hour period. In this particular example with the cylinder measurement system, the cumulative x-offset error after 12 hours was 4 μm.

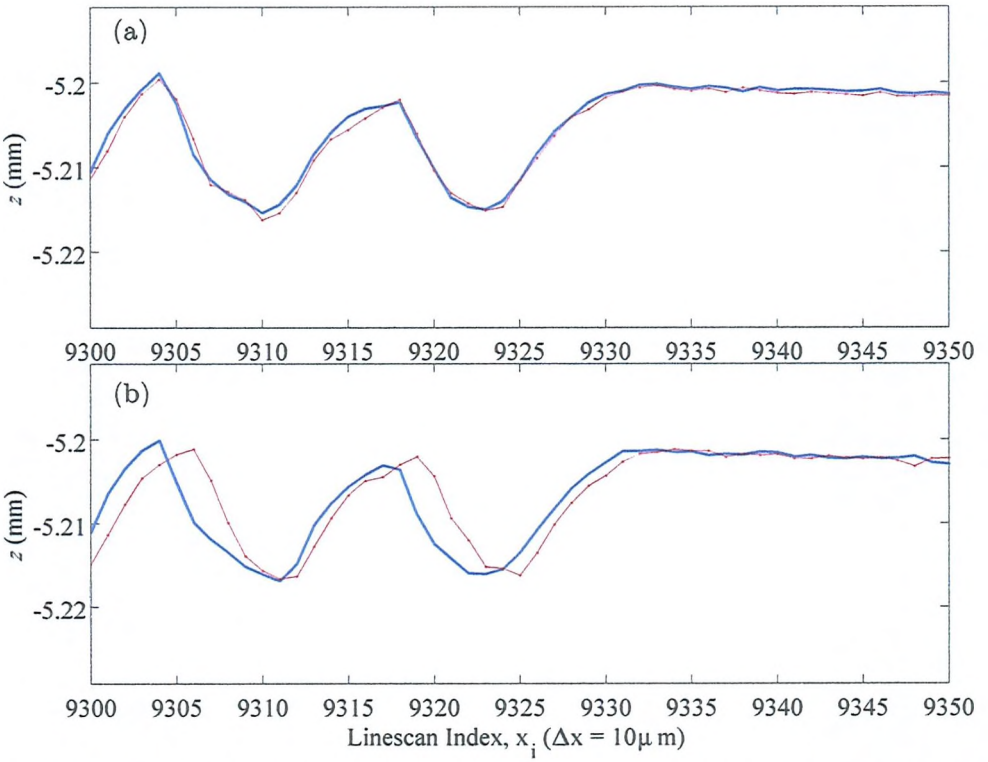

Figure 3.18: Linescan X-Offset error. Consecutive linescans are shown here measured from a cylinder surface (PBEAR), (a) well aligned, and (b) offset by approximately 2 measurement points, equivalent to 20μm.

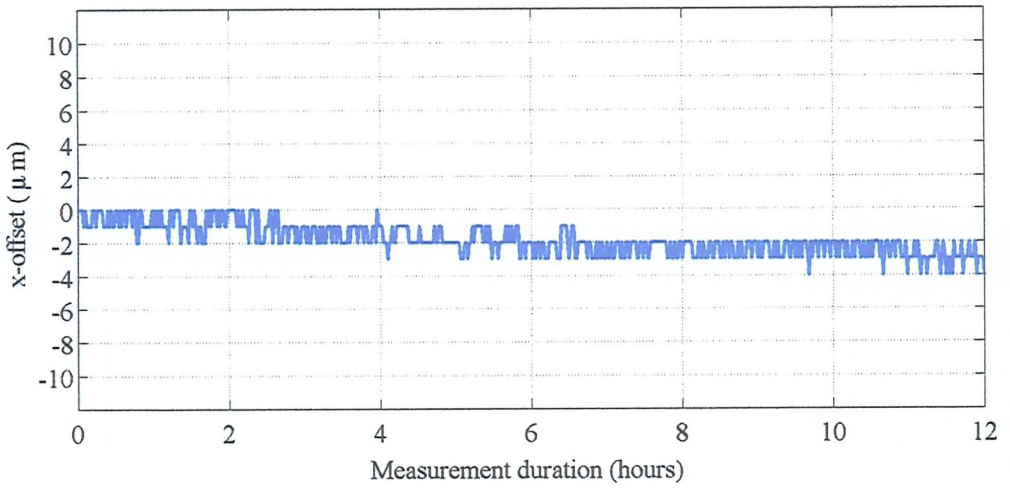

Figure 3.19: Repeatability measurement test ($\Delta x = 1$ μm) to show typical linescan offset error of the cylinder measurement system. Total drift in this is example (repeated linescan measurements of TC2 cylinder) after 12 hours of measurement is ≈ 3 μm.

3.5 Chapter Summary

In this chapter, the NCFSM method has been described. Definitions of the data, including the linescan, discrete surface and measurement grid have been given. The measurement parameters have been considered with regards to the minimum requirements for sound reproduction. Due to the limited gauge range of the sensor, a scanning segmentation strategy is required for specimens with high eccentricity.

With all discrete measurements, the grid resolution over which the surface is measured is particularly important. There is a trade off between data redundancy, scan time, and loss of audio signal quality. Archivists are likely to want as much data as possible, but care must be taken to ensure that data sets don't become unmanageable with modern computing systems.

Overcoming measurement error in the sound reproduction step is key. Errors introduced by the measurement process include drift in both the x and z axes. Problems associated with the NCFSM method include the introduction of outliers due to debris, or foreign material lying above the imaging plane. Under this condition, NCFSM is less advantageous when compared with stylus reproduction, which has the advantage of 'dislodging' debris during playback. Despite this, the

NCFSM method will allow for reproduction of surfaces which are considered unplayable via stylus.

The measurement requirements for disc recordings are more demanding than for cylinders, in terms of data storage, scan time and the ability to resolve the full groove cross-section. In its current design, the air-bearing system is not capable of resolving the groove cross-section of a 45/45° disc sufficiently for accurate signal reproduction, due to the angular tolerance of the sensor. In order to overcome this issue, it may be necessary to orient the sensor(s) orthogonally to the groove walls, however this remains an exercise for future research.

Chapter 4

Signal Reproduction Methods

4.1 Introduction

The NCFSM method described in Chapter 3 results in a discrete representation of the continuous grooved surface. This chapter outlines the methods and algorithm development for audio signal reproduction from data acquired by the NCFSM method. The majority of this chapter regards cylinder recordings, as this has been the main focus of the Sound Archive Project to date. Signal reproduction from non-finalised disc data is also described.

The primary aim for signal reproduction is to obtain an accurate transcription of the original encoded sound. The task of signal reproduction can be thought of as simulating the action of a reproduction stylus, as it traces over the groove. Alternatively, we can consider signal reproduction as an estimate of the trace encoded by the original cutting tool. Accuracy of signal reproduction is dependent on both the measurement and sound recovery stages. Noise and distortion introduced by the measurement system must be taken into account when designing signal recovery methods. Every effort must also be made to ensure that the algorithms themselves do not introduce uncharacteristic noise or distortion.

This chapter is laid out as follows. To start, a consideration of how the signal is encoded on the discrete surface for cylinder and disc data is given. Methods developed for signal reproduction from cylinder and disc surface measurement are then described. In the case of cylinders, methods are described for good and poor surface condition. Throughout this chapter, the rationale for the methods chosen are considered and compared with prior methods developed by Fadeyev and Haber [5] for signal reproduction from the NCFSM method.

4.2 Cylinder Signal Encoding

This section outlines the signal encoding principles for discrete surface measurements of cylinder recordings. It should be noted that these considerations also apply to reproduction from tinfoils, which also employ the hill-and-dale mode of recording.

4.2.1 Signal Traces

For cylinder (or tinfoil) recordings, the audio signal is encoded primarily in the z-plane (depth) denoted by z_a. This signal is produced by the radial motion (modulation in z) of the cutting tool in response to the incident acoustic pressure. Negative and positive pressures produce shallower and deeper grooves respectively (see the linescan model in Figure 4.1).

Sound is recorded in a single channel (monaurally), and can be extracted by estimating the vertical modulation of the groove cross-section along a trajectory formed by the cutting tool. In addition to the primary z-modulation z_a , two secondary signals, x_L and x_R, can be observed as lateral variations in x, towards the ridge (see Fig. 4.2). Assuming that the cutting facet is symmetric and tapered, then x_L and x_R are 180° out of phase with each other. Figure 4.2 shows the three signal traces z_a , x_L and x_R for a vertically cut groove cylinder.

In terms of signal resolution, the limiting factor for the primary depth signal, z_a is the vertical resolution Δz, of the sensor. For signals, x_L and x_R the resolution is determined by the linescan grid spacing Δx, used by the linear stage system. In general, for the WL measurement system described in Chapter 3, $\Delta x >> \Delta z$ (typically $\Delta x = 10$

μm and $\Delta z = 10$ nm). Under these typical measurement conditions, z_a can therefore be resolved to 1000 times greater resolution than, x_L and x_R. As a result, the signal traces x_L and x_R are disregarded in the sound extraction procedure for cylinders.

4.2.2 Stylus Trajectory

The recording stylus encodes a continuous helical groove of pitch, λ_x which propagates along the cylinder x-axis. When the cylinder surface is measured, via the NCFSM method, this trajectory is mapped to a 2-D matrix of surface heights, $z(x_i,\theta_j)$ as shown in Figure 4.3. In this representation, the discrete stylus trajectory is considered as the set of trajectories: $\{a_1 + a_2 + ... a_n\}$, which when appropriately ordered and concatenated, forms the complete stylus trajectory. The subscript n denotes the positional index of the n-th groove found along the x-axis. This index is referred to subsequently in this chapter as the *groove seed.*

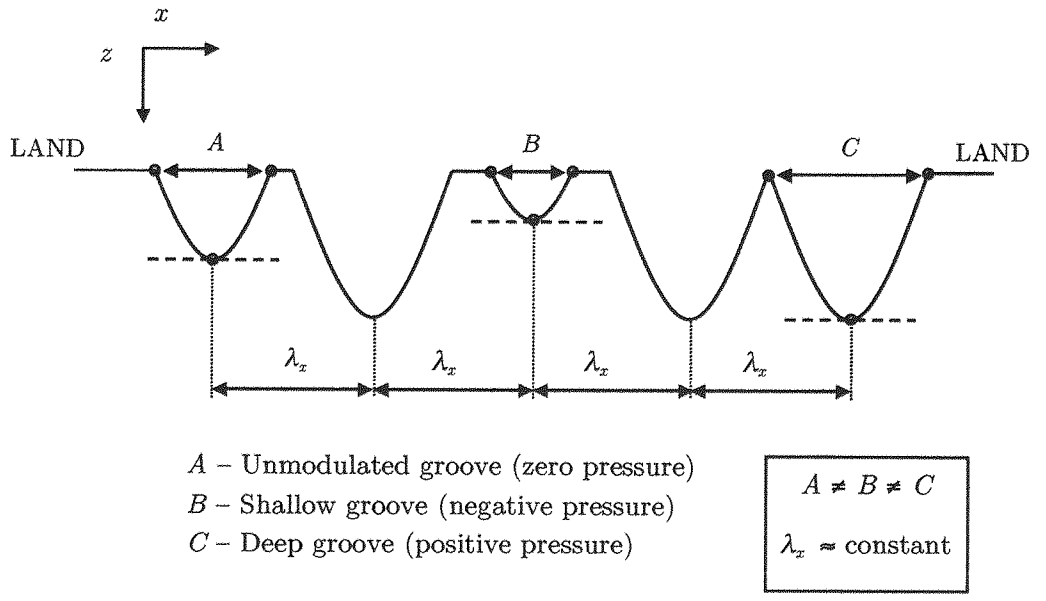

Figure 4.1: Idealised linescan model for vertically cut recording. The groove pitch λ_x remains constant from groove to groove, but the distance between adjacent groove ridges is variable and is dependent upon the depth of cut.

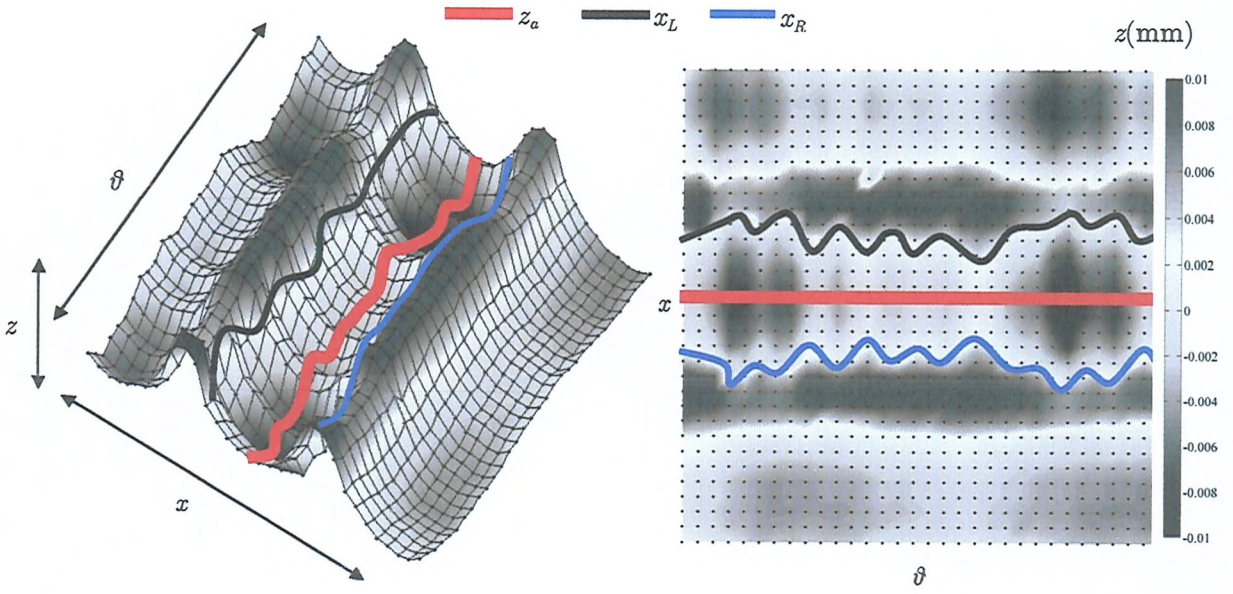

Figure 4.2: Consideration of signal traces for vertically cut groove recordings. The primary depth modulation, z_a and lateral variations in x, (x_L and x_R) are shown.

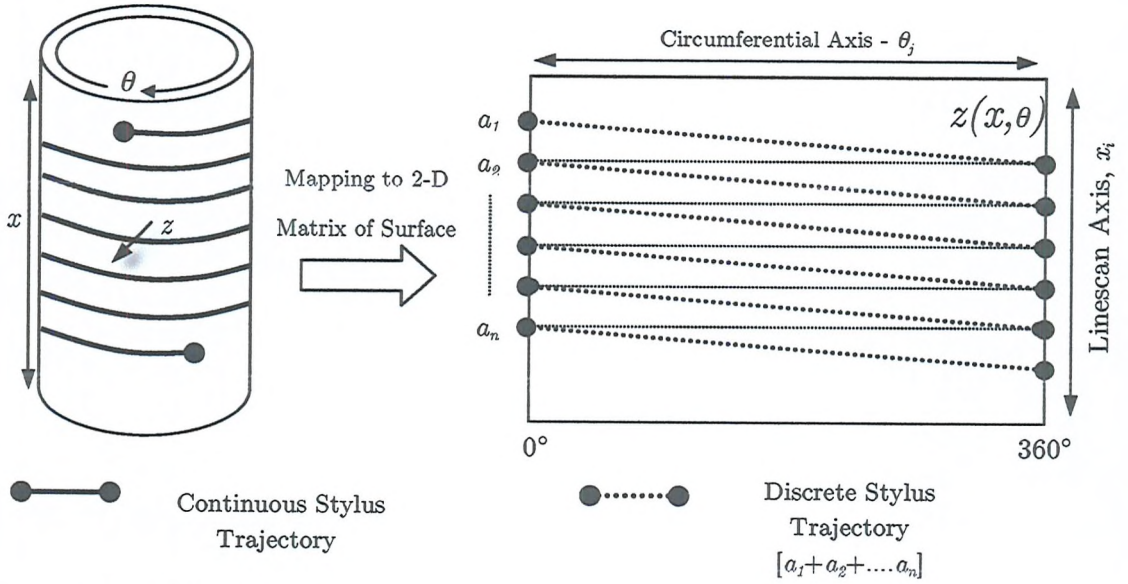

Figure 4.3: Figure to demonstrate the consideration of a discrete stylus trajectory when the continuous cylinder surface is mapped to a 2-D matrix of Surface Heights.

4.3 Flat Disc Signal Encoding

This section considers the encoding principles for monophonic disc records, produced using the lateral (side-to-side) recording mode. To date, stereophonic records have not been investigated by the Sound Archive Project. It should be noted that current measurements from the NCFSM method do not include cross-sectional data at the groove walls (as was shown in Chapter 3, Fig. 3.11-3.13), where the surface incline exceeds the angular tolerance of the sensor. Hence to obtain an audio signal, a special consideration of the signal traces is described. Developments in the disc measurement system will hopefully ensure that the full groove cross-section is measured.

4.3.1 Signal Traces

The audio signal is encoded in radial variations of the groove cross-section found along a spiral trajectory. Radial modulations for discs measured by the NCFSM system correspond to a variation of the groove cross-section position in the x-axis, (as opposed to z-axis for cylinders). The 'v-' or 'u-shaped' groove cross-section is composed of two groove walls W_L and W_R, (left and right) which are angled at $\pm$ 45° from the disc surface normal. Each groove wall has two termination points at: (i) the interface between wall and the land (surface between adjacent grooves) and (ii) the interface between W_L and W_R (at the groove bottom). For preliminary data acquired from the NCFSM method, we can extract the audio signal from x-positional data of three signal traces, A_L, A_R and A_B, (shown in Fig. 4.4), which are the termination points at the left groove wall, right groove wall, and groove bottom.

For an ideal disc, the groove width, w_x (x-distance between termination points A_L and A_R) and depth, d_z (z-distance from the land to the groove bottom) are constant throughout the recording. These constant features can therefore be used to obtain the signal traces, by constructing rules based on these ideal distances.

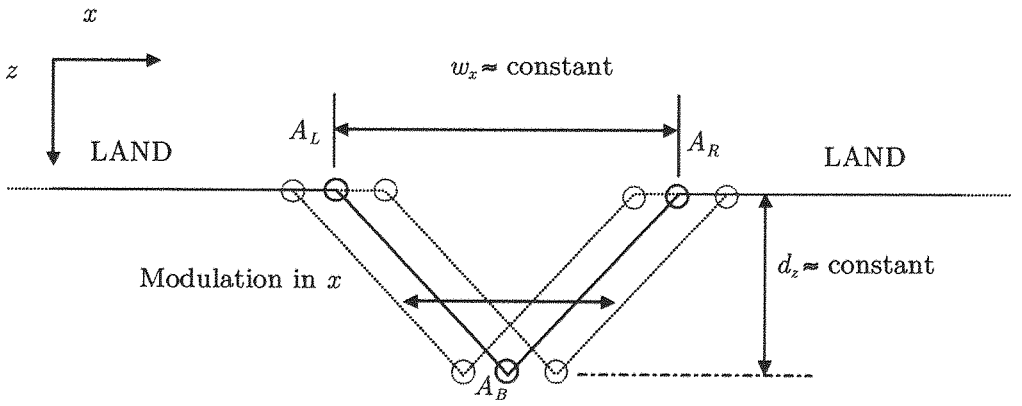

Figure 4.4: Idealised groove cross-section of laterally encoded (monophonic) disc. The audio signal can be obtained from preliminary data measured by the NCFSM system by detecting the three termination points of the groove walls, A_L, A_R, and A_0.

4.3.2 Stylus trajectory

An unmodulated disc groove follows a concentric spiral of constant pitch, λ_x, which starts from the outer disc radius, and propagates towards the disc origin. Figure 4.5 shows the mapping of the unmodulated stylus trajectory for discs measured by the NCFSM method. As with cylinders, the discrete stylus trajectory is considered as the set of trajectories: $\{a_1 + a_2 + ... a_n\}$, which when appropriately ordered and concatenated, forms the complete stylus trajectory.

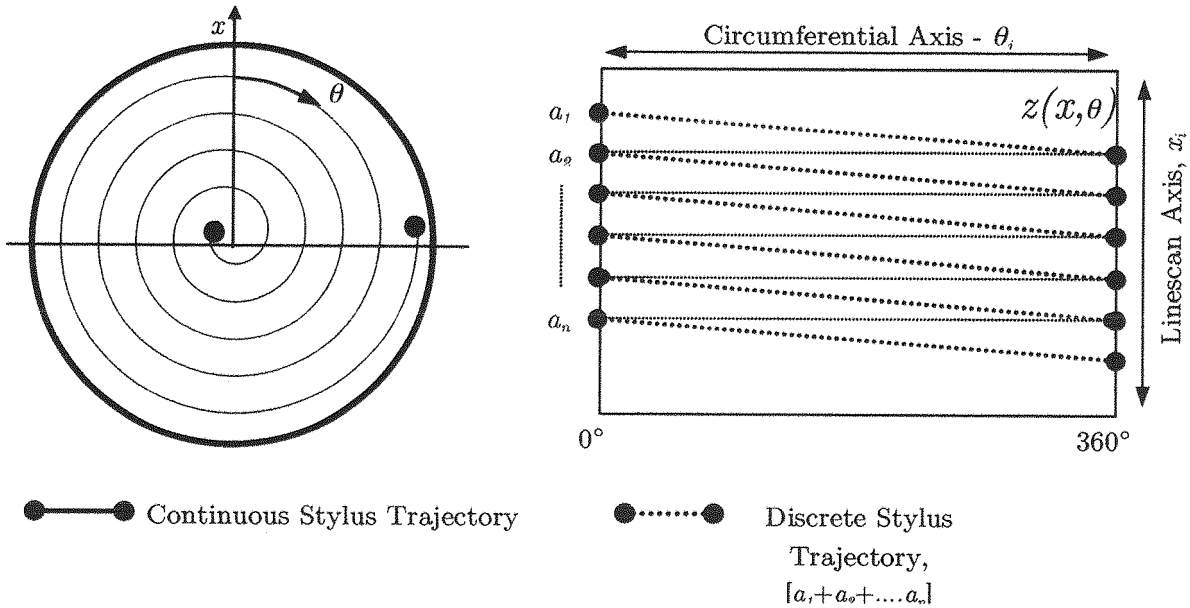

Figure 4.5: Continuous and Discrete stylus trajectory considerations for laterally encoded discs, measured by the NCFSM method.

4.4 Methods Developed for Cylinder Recordings

This section describes the methods developed for signal reproduction from cylinder recordings measured by the NCFSM method.

4.4.1 Overview of Procedure

Signal reproduction from cylinders is decomposed into the following stages:

1. **Data Import and Pre-Processing**: data acquired by the measurement system is imported into MATLAB. The full surface is constructed from surface segments, (as per the method described in Section 3.2.4). This stage is discussed in **Section 4.4.2.**
2. **Groove Parameter Estimation**: the groove pitch, λ_x and nominal height, H_z (from peak-to-trough) of the cylinder grooves is automatically estimated. This stage is discussed in **Section 4.4.3.**
3. **Outlier Detection**: data which may be attributed to damage and/or debris is identified to suppress noise introduced by the measurement system. This step is discussed in **Section 4.4.4.**
4. **Groove Seed Detection:** the groove seed, $g_n(x_i)$ gives the position of each groove valley along the linescan axis. This ensures that all sound carrying grooves are included in the final track. This is discussed in **Section 4.4.5.**
5. **Stylus Trajectory Estimation:** an approximation to the x-θ path traced by a stylus, is formed. Two methods of stylus trajectory estimation are described for dealing with cylinders in good and poor surface condition. These methods are discussed in **Section 4.4.6.**
6. **Depth Signal estimation:** a local approximation of the discrete groove depth found along the stylus trajectory estimate in Step 5) is formed. This depth signal corresponds to the primary signal trace, z_a, in which the sound is encoded. Four different depth estimation schemes are considered in this chapter. These methods are discussed in **Section 4.4.7.**
7. **Signal Correction and Equalisation:** errors introduced by the measurement system such as measurement drift at segment boundaries are accounted for. Equalisation can also be applied if necessary to approximate the velocity response of stylus replay. This is discussed in **Section 4.4.8.**

4.4.2 Data Import and Pre-Processing

The purpose of this stage is to ensure that measured data can be manipulated in the MATLAB workspace environment and that the full surface is reconstructed

appropriately. Recordings are typically measured in a number of segment scans, from which the full surface is constructed, by stitching segments together.

Data Import

Data recorded by the measurement system is stored in an ASCII text file, known as a 'tai' file. The tai-file header gives information about the number of scans point in x and θ, the measurement grid spacing, stage position offsets, as well as scanning times, and precision mode. Data in the tai-file is stored in the proceeding rows, with each row containing three floating-point values that correspond to x, θ, and z measurement points respectively (note that 'θ' is referred to as 'y' in the tai file). Assuming a constant grid spacing for the measurement vectors, x_i and θ, only the z-data and measurement grid indices are required for processing. This reduces the amount of data stored by 66%, and is necessary for processing on a standard desktop PC, due to memory constraints. Tai-file data is imported into MATLAB, and z-values are reshaped to form a 2-D matrix of size, $I \times J$, where I and J denote the number of measurement points in x- and θ-axes respectively.

Data Stitching

In Section 3.2.4, a scanning strategy was described for measurement of non-concentric cylinders, in which the full surface is mapped in a number of segments at fixed z-standoff positions. The z-offset of the measurement stage is recorded for each segment scan and is accounted for when saving z-values to file. Despite this, there often remains a z-offset error between segment joins. In Chapter 3, it was stated that the effect of misaligned surface segments would result in impulsive noise for the output audio signal, due to a mismatch in z-position at the segment joins. To avoid such distortions, the segments must be are accurately positioned and aligned prior to signal extraction.

A cylinder surface is generally scanned in a number of surface patches, with the n-th surface patch denoted by $Z_n(x_i^n, \theta_j^n)$. The grid positions, $[x_i^n, \theta_j^n]$, of each surface patch are stored in a script determined by the scanning optimisation routine, as

described in Section 3.2.4. From this script, a segment map is generated which includes the positional information of each segment (see Figure 4.6).

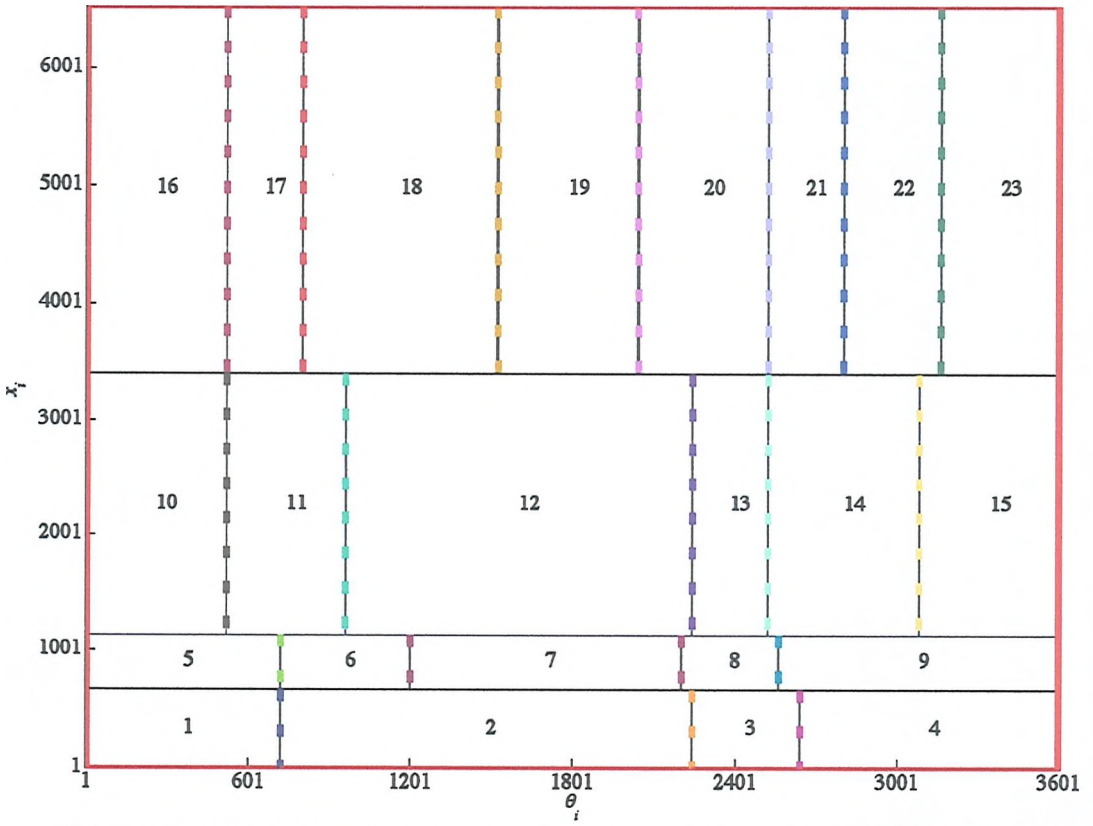

Figure 4.6: Segment map for a cylinder recording, showing positional information of 23 surface patches. This segmentation scheme covers an area of 65mm by 360° (grid spacing: $\Delta\theta = 0.1°$, $\Delta x = 10$ µm). Dashed lines denote the θ-joins between segments.

Unlike some image registration applications, the NCFSM scanning strategy does not typically include any overlap between adjacent surface patches. In light of this, surface stitching is performed by examining the interface (joins) between adjacent segments.

One of the constraints imposed by the scanning routine ensures that cylinder bands of constant width (along the x-axis) are measured, i.e. in the case of Figure 4.6, the patches: $Z_1 - Z_4$ constitute a cylinder band of constant width. This simplifies surface stitching, in that segment registration need only occur at theta joins, e.g. the boundary which lies between Z_1 and Z_2.

An initial search for θ-joins between segments is carried out from the positional information of the scanning scheme script. It is then possible to adjust for any z-offset errors, by shifting each surface patch in z. This routine is carried out iteratively, until

all surface patches are aligned in z. The z-corrections are determined by examining the surface boundary linescan profiles, (e.g. $Z_1(x,\theta_{end})$ and $Z_2(x,\theta_1)$), and minimising the mean error of the two profiles. When all surface patches have been aligned in this way, the full surface, $Z(x_i,\theta_j)$ is formed for subsequent processing.

4.4.3 Groove Parameter Estimation

Each recorded specimen has unique groove characteristics. In order to ensure that the sound extraction procedure is automatic and robust, some parameters of the grooved surface must be estimated. These features, (shown in Fig. 4.7) can be used as input parameters for later stages in the sound extraction process, such as outlier removal, groove seed detection, and filtering operations. The features we wish to detect are:

- Groove pitch, λ_x : the average distance in x between adjacent groove valleys.
- Groove Height, H_z : the average distance in z from peak to trough of the groove.

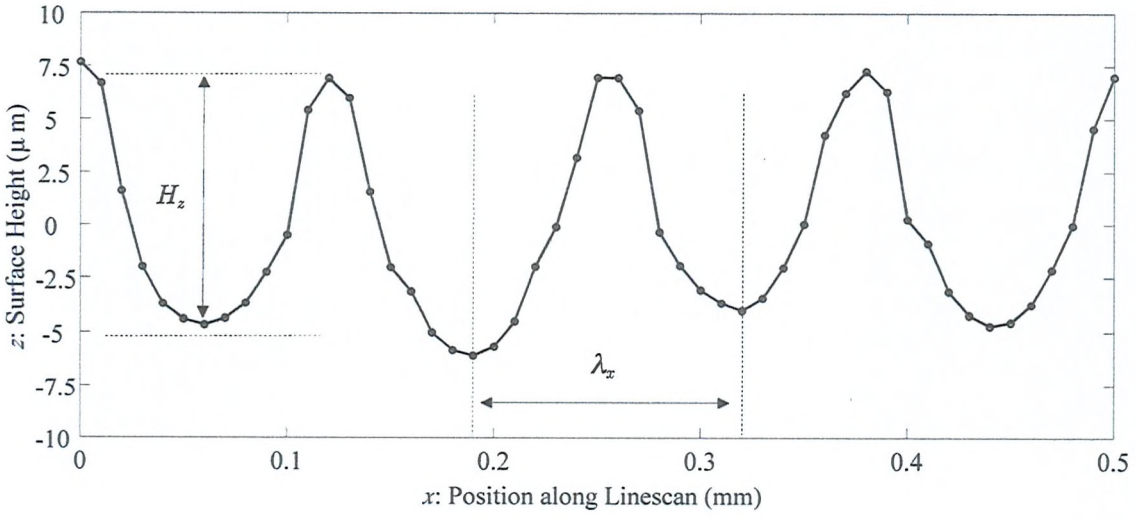

Figure 4.7: Linescan from Blue Amberol cylinder, demonstrating the groove pitch, λ_x and groove height H_z parameters.

Groove Pitch Estimation

Through the recording process, the linescan has a periodic structure, with adjacent grooves separated by a constant distance λ_x, corresponding to the groove pitch. In most cases, λ_x can be derived from historical records or by observation, but for completeness and robustness, it is estimated by spectral analysis of the linescan. The periodic structure of the linescan can be modelled as a single complex sinusoid plus noise term:

$$f(x) = A_0\cos(2\pi k_0 x+\phi) + n(x) \tag{1}$$

- The amplitude A_0 of the model gives an approximation of the peak to trough distance H_z, of the linescan and can be used as a threshold for outlier detection.
- The frequency term k_0, gives the fundamental spatial frequency of the model, from which the groove spacing, λ_x is derived,
- The phase term ϕ, gives information as to the relative position of the grooves along the x-axis.
- The noise term, $n(x)$, includes a low frequency component (surface form), and Gaussian noise (surface roughness).

The maximum likelihood estimate for the frequency of a single sinusoid in white Gaussian noise can be found by locating the frequency at which the Fourier transform of the signal attains its maximum [80]. The discrete Fourier transform (DFT) of the j-th linescan, is given by the complex sequence $\tilde{z}_j(k)$, where k denotes the k-th spatial frequency bin. The fundamental spatial frequency of the linescan, k_0 is related to λ_x via the reciprocal relationship of period and frequency and is found by calculating the expectation, E, for all linescans:

$$k_0 = E\,[\,max\mid \tilde{z}_j(k_p)\mid\,], \qquad j \in \theta : [0°,360°] \tag{2}$$

where k_p is the range of possible frequency bins in which k_0 is valid, which for cylinder recordings of the era is between 3.9 and 7.9 mm^{-1} (for 200 and 100 tpi respectively). The groove spacing rounded to the nearest integer ($NINT$) sample:

$$\lambda_x = nint(\ k_0^{-1}\ ) \tag{3}$$

Average Groove Height Estimation

An approximation for H_z can be obtained by observing the dominant amplitude at frequency k_0 for an ensemble of linescan measurements:

$$H_z = E\left[\ |\tilde{z}_j(k_0)|\ \right] \tag{4}$$

Due the vertical encoding process, the average groove height H_z is not strictly constant for all groove cross-sections, and is dependent upon the depth of cut. The variation in H_z is however small when compared with macroscopic debris features, such as dust. For this reason, it is a useful parameter for identifying outlying data points on the basis of a z-threshold envelope.

4.4.4 Outlier Removal

In order to minimise the introduction of impulsive noise to the recovered signal, samples on the surface $Z(x_i,\theta_j)$, which are attributed to surface damage or debris, must be removed. In addition to these 'naturally occurring outliers', artificial outliers introduced by errors in the acquisition system must also be removed. Impulsive noise reduction is typically carried out in the time-domain for phonograph recordings, using audio signal processing techniques [25-27]. With the NCFSM method, impulsive noise can be removed 'at source', prior to audio signal extraction, by examining samples in the spatial domain.

Hawkins [81] defines an outlier as 'an observation which deviates so much from other observations as to arouse suspicions that it was generated by a different mechanism'. In the case of NCFSM data, the following may cause these 'different mechanisms':

1. **Additive Debris** – caused by foreign objects that lie above the imaging plane of the underlying groove structure. Examples of additive debris are particulates (e.g. dust), or fibres (e.g. hair).
2. **Subtractive Damage** – caused by the removal of material from the grooved surface. Examples of subtractive damage include scratches caused by poor handling / stylus replay, or 'pitting' caused by mould growth.
3. **Acquisition Error** – errors introduced by the NCFSM acquisition system that are not inherent to the specimen. Outliers introduced by measurement error can be introduced at positions of the surface that exceed the angular tolerance of the sensor or at positions on the surface which exceed the sensor's operating range. Additionally, random outliers have been observed due to errors in the acquisition system.

In previous work by Fadeyev et al. [5], outlier rejection from NCFSM cylinder data was achieved by fitting a quadratic to the groove shape and observing samples whose residual error exceeded a threshold. This method is well suited to a Blue Amberol cylinder, whose groove shape resembles a parabola. To achieve a method of outlier removal, which is robust for all specimens, (regardless of groove shape), an outlier removal stage based on threshold filters and geometric considerations of the ideal groove structure is now described. This process relies on the groove pitch, λ_x and average peak-to-trough height, H_z , as defined in section 4.4.3. Two outlier removal stages have been developed:

Stage 1: Outlier removal via Median and SG Filter

The first stage is based on a similarity measure from a median filtered linescan, denoted by z_{med}. The median statistic is the centre of a rank-ordered distribution and has a well-known ability to remove isolated outliers (referred to as 'salt and pepper' noise in image processing [82]). Isolated outlier samples may be caused be random errors introduced by the acquisition system for instance, or small regions of isolated debris.

When a suitably designed median filter is applied to the linescan, isolated outlier samples will appear at either end of a rank ordered list and are removed by the median process. By comparing raw linescan data $z_j(x)$ with the median filtered linescan, outliers can be identified by observing samples whose absolute difference from $z_{med}(x)$ exceeds a threshold. The difference threshold is set to $0.5H_z$. The median filter operates on a block length $L = \lambda_x/4$ (rounded to the nearest integer sample), i.e. a quarter of the groove pitch period. Over this block length, L the median filter was found to discriminate well against isolated outliers.

A second similarity measure from a Savitzky-Golay (SG) smoothed linescan, z_{sg} is also carried out at this stage. In this case, the order of the filter P=4 (quartic), and the frame length is set to λ_x. In this way, the SG filter operates over a frame range equivalent to the groove pitch, and in this limit, a maximum of 3 turning points are assumed, the two ridges (maxima) and the valley (minima).

The procedure for outlier removal based on these two filtered linescans is as follows:

For the j-th linescan, $z_j(x)$:

Compute the Median (z_{med}) and SG (z_{sg}) filtered linescans
Compute the absolute differences:

$$e_{med}(x) = \left| z_j(x) - z_{med}(x) \right|$$

and:

$$e_{sg}(x) = \left| z_j(x) - z_{sg}(x) \right|$$

Outliers in $z_j(x)$ are considered at x_i if:

$$e_{med}(x_i) > 0.5H_z$$

or:

$$e_{sg}(x_i) > 0.5H_z$$

Values at $z_j(x_i)$ are replaced with NaN ('not a number')

The procedure is carried out iteratively with median and SG filters for each linescan until no outliers are detected. Figure 4.8 shows an example of outliers identified by via the median filtered linescan test.

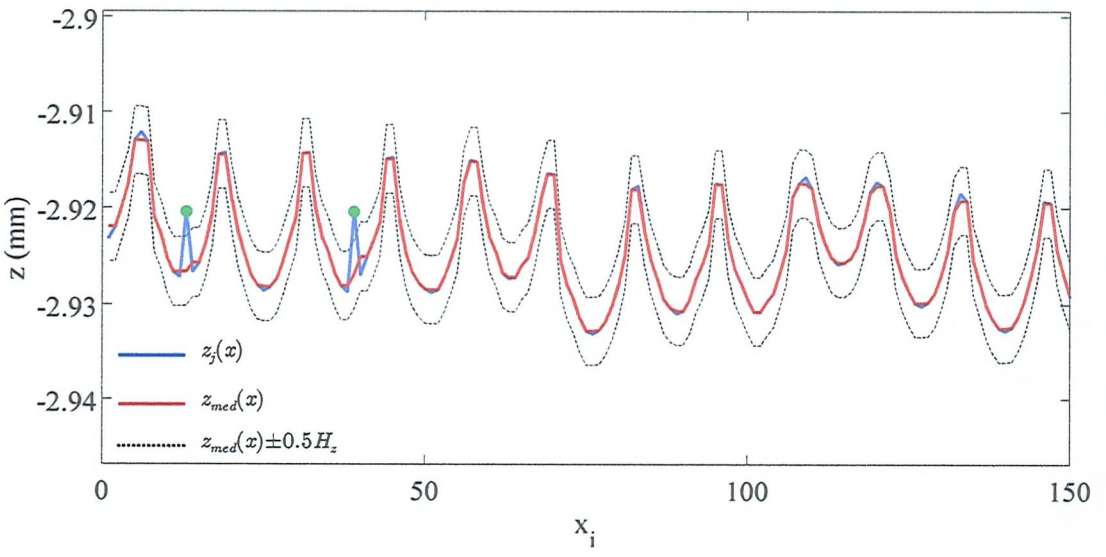

Figure 4.8: Example of median filtered linescan for outlier removal with Blue Amberol cylinder linescan, (outliers identified are highlighted by green markers).

Stage 2: Groove Envelope Function Test

Outlier Stage 1 may not detect all outlying samples, if there are sustained (as opposed to isolated) regions of damage. The purpose of Outlier Test 2 is to detect regions of outliers, which may be caused by fibres or scratches, lying across the groove structure for instance. To do this, two envelope functions are formed: z_{low} and z_{up}, which are the lower and upper envelope functions respectively. These envelope functions are formed by detecting ridge and valley points of each groove. Sets of ridge points $R(x_N)$ / valley points $V(x_N)$ are determined by locating samples in $z_j(x_i)$, which are higher / lower than a neighbourhood of $x_i \pm \lambda_x/2$ samples. If ridge / valley candidates happen to be located at outlier positions, they are iteratively removed from the set of $R(x_N)$ and $V(x_N)$, if they exceed a z-distance threshold of $\pm 0.5H_z$ from neighbouring candidates.

Linescan envelopes functions are generated by linearly interpolating at node positions specified at $R(x_N)$ and $V(x_N)$. An example of these linescan envelopes functions are shown in Figure 4.10. Outlying samples are then removed if they deviate by more

than $0.25H_z$ above/below the two envelope functions z_{low} and z_{up}. Outliers identified by this procedure are shown in Figure 4.9. In this example, the expected ridge node candidate between x_i = 250 – 300, is removed from the list of $R(x_N)$, as it exceeds a z-distance of 0.5Hz from surrounding candidates. The procedure for outlier removal via the groove envelope test is carried out as follows:

FOR the j-th linescan, $z_j(x)$:
Determine a list of ridge, $R(x_N)$ and valley candidates $V(x_N)$
A ridge candidate $R(x_i)$ is found at x_i IF:

$z_j(x_i) > z_j(x_{i-1/2}: x_{i+1/2})$

A valley candidate $V(x_i)$ is found at x_i IF:

$z_j(x_i) < z_j(x_{i-1/2}: x_{i+1/2})$

Remove candidates from $R(x_N)$ and $V(x_N)$ IF :

$|R(x_N) - R(x_{N-1})| > 0.5H_z$
$|V(x_N) - V(x_{N-1})| > 0.5H_z$

Form upper (z_{upp}) and lower (z_{low}) envelope functions, using valid $R(x_N)$ and $V(x_N)$ nodes. Outliers in $z_j(x)$ are considered at x_i if:

$z_j(x_i) > z_{upp} + 0.25H_z$
or:
$z_j(x_i) < z_{low} - 0.25H_z$

Values at $z_j(x_i)$ are replaced with NaN ('not a number')

Figure 4.10 shows a Blue Amberol cylinder surface (JBTBM) with outliers identified via the outlier tests described above. It should be noted, that no attempt has been made to replace the regions caused by damage or debris sites and such reconstruction operations lie beyond the scope of this work. The fact that outliers have been identified and stored as *NaN* means that when an audio signal is created, *NaN* samples in the time series can be interpolated. This process ensures that impulsive noise caused by debris or measurement error is minimised.

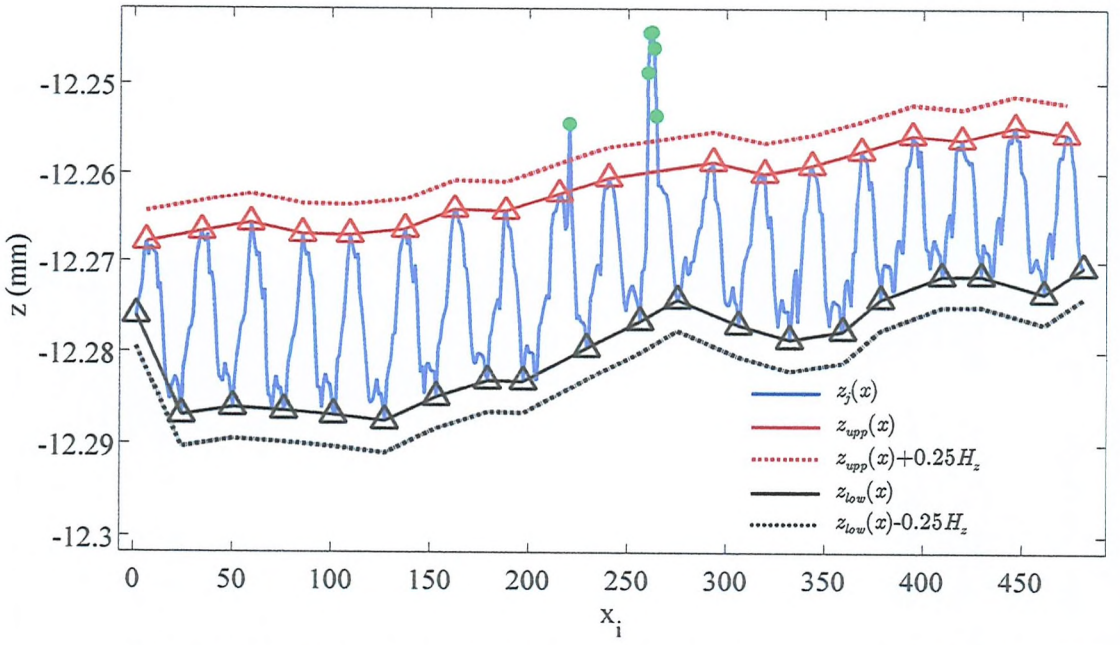

Figure 4.9: Example of linescan envelopes used for outlier removal with TC1 linescan (outliers identified denoted by green markers). Triangles denote ridge, $R(x_N)$ and valley, $V(x_N)$ nodes used to form the linescan envelopes.

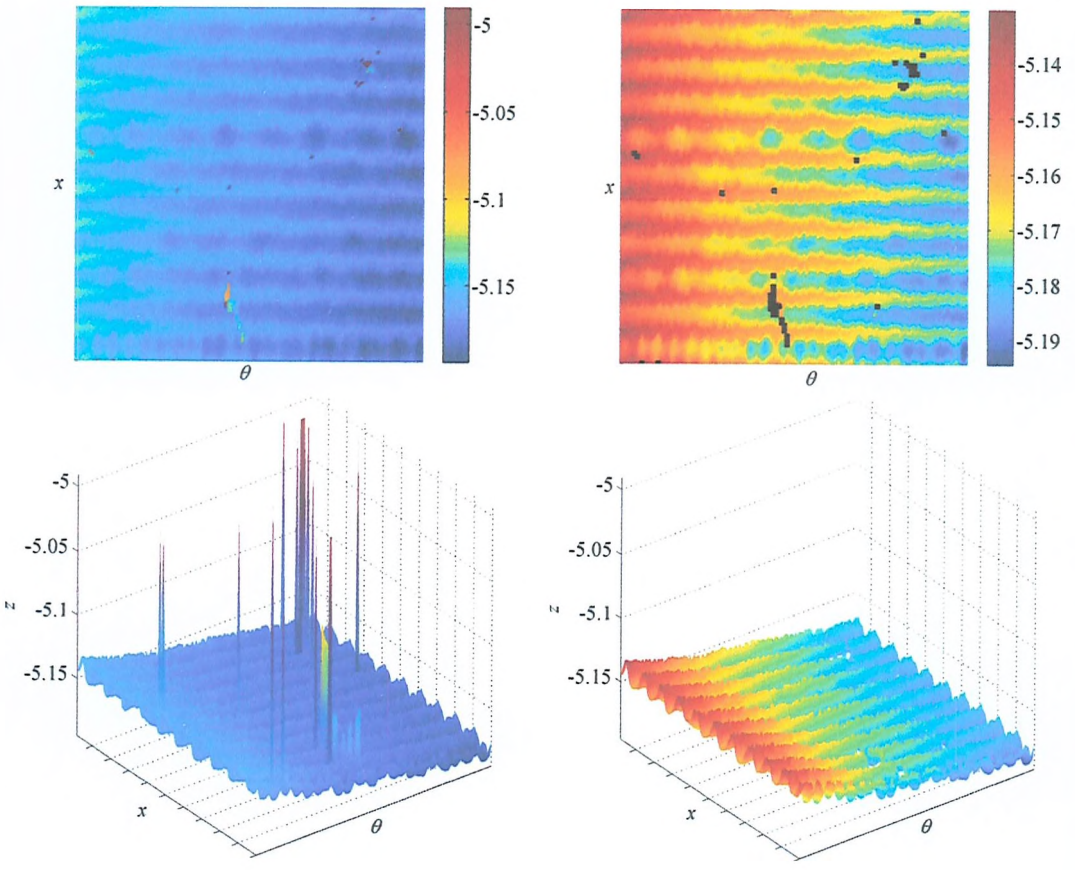

Figure 4.10: 3D surface and 2D intensity plots of a Blue Amberol cylinder (PBEAR) to demonstrating outlier removal. Outliers identified are highlighted by black markers.

4.4.5 Groove Seed Detection

The purpose of this stage is to determine the x-index of grooves along the surface at which the recorded track exists, which involves valley detection. This task is aided by prior knowledge of the groove spacing λ_x determined by Eq. (3).

The k-th groove seed, g_k is found by locating the medial axis of each groove valley along the linescan at $\theta = 0°$. For smooth, symmetric groove cross-sections, the medial axis, is indexed by the local minimum point of the groove and can be found via a simple neighbourhood search. In cases where the groove cross-section is rough and/or asymmetric the medial axis does not always lie at the local minimum point. To ensure that each groove is indexed by its medial axis, an appropriately designed Savitzky-Golay (SG) polynomial smoothing filter is applied to the linescan $z_j(x)$ to form the filtered linescan $z_j'(x)$.

In Chapter 2, Savitzky-Golay (SG) polynomial smoothing filters were described. SG filters were initially developed to identify the relative widths and heights of spectral peaks in noisy spectrometric data [71]. In this work, the SG filter can be used in a similar way, to locate the medial axis of a groove valley.

The frame length L, of the SG filter is chosen to match the groove pitch, λ_x. In this way, the SG filter operates locally on each groove valley, replacing the rough groove cross-section with a smoothed version. Figure 4.11 shows a noisy linescan (data from TC1 cylinder), filtered with a cubic SG filter ($P = 3$) with frame length $L = 25$, (the approximate number of samples across one groove cross-section for a 100 tpi cylinder, measured with $\Delta x = 10$ µm). It can be seen that the position of minima in the filtered linescan, $z_j'(x)$ are located closer to the middle (medial axis) of the groove valley, than minima found using the raw linescan $z_j(x)$.

The local minima of this SG filtered linescan are used to determine the groove seed. The k-th groove seed, g_k is found by locating the minimum point in a linescan neighbourhood: $z_j'(x_{i-(\lambda x/4)}, x_{i+(\lambda x/4)})$, looping over the index i. The search neighbourhood is set to $\pm\lambda_x/4$ linescan samples to ensure that minima are not falsely located from adjacent grooves. Figure 4.12 shows the groove seed positions from a cylinder surface (TC1) overlaid onto the surface formed by the noisy linescans shown in Figure 4.11.

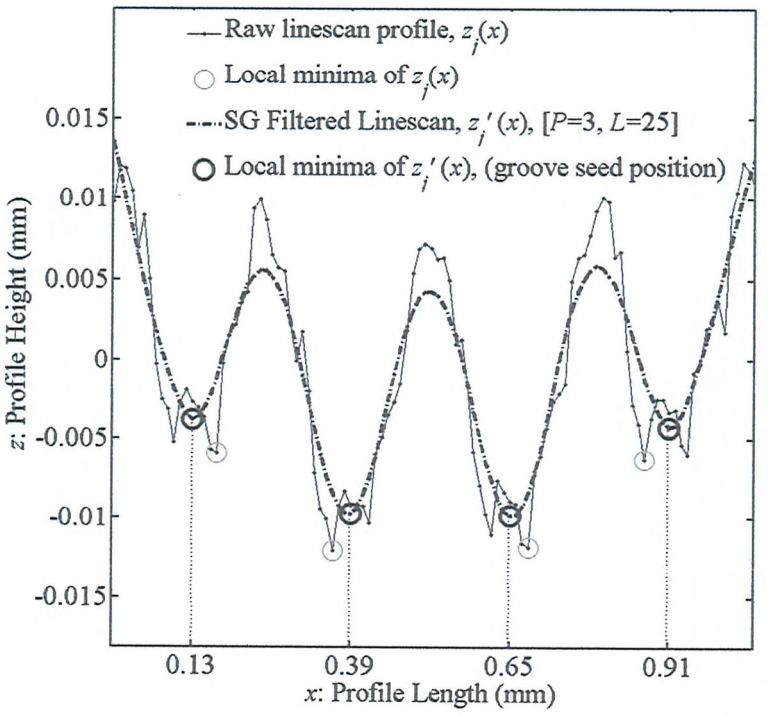

Figure 4.11: Demonstration of groove seed detection. The abscissa of local groove minima from the raw linescan profile (TC1) $z_j(x)$, varies from one groove to the next due to the roughness of the cutting tool. Abscissa of minima from the filtered linescan, $z'_j(x)$ are spaced equidistantly by an expected distance of 0.254mm. The abscissa of local minima of $z'_j(x)$ provide the groove seed position.

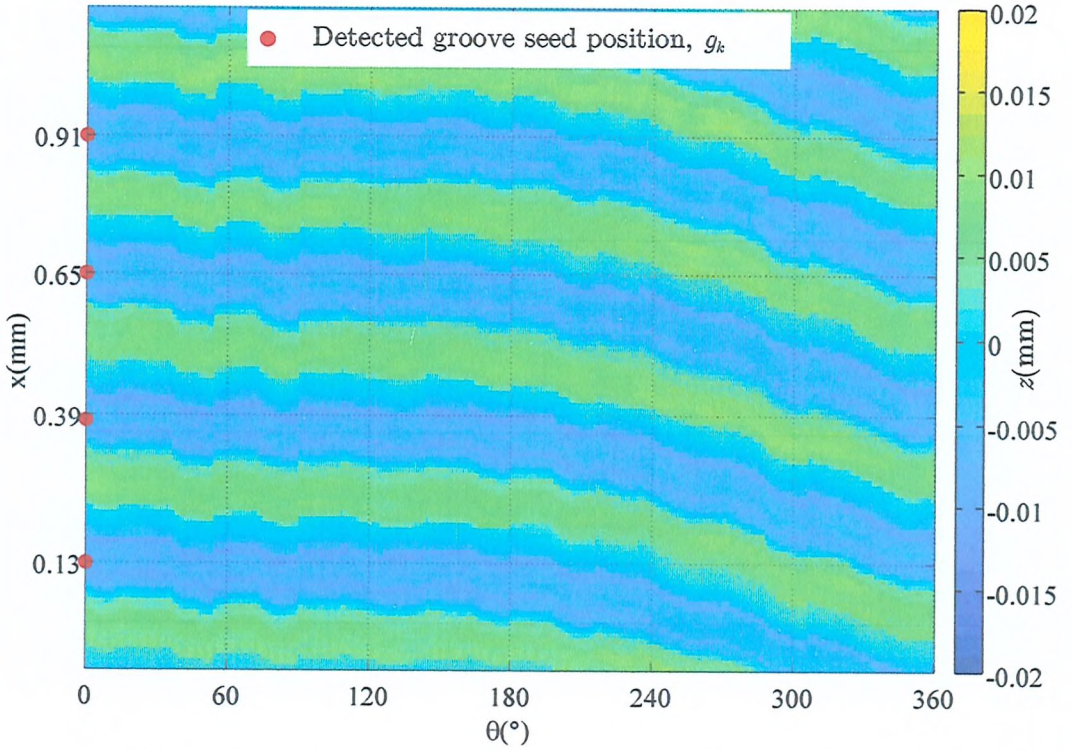

Figure 4.12: Groove seed positions overlaid onto 2D surface plot constructed from example of noisy linescan data (TC1) shown in Figure 4.11.

4.4.6 Stylus Trajectory Estimation

In previous work on NCFSM cylinder audio reproduction [5], the estimation of stylus trajectory and groove depth were not considered independently, but as a single process. In this work, signal recovery is considered in two independent stages – stylus trajectory estimation and audio signal estimation.

The purpose of stylus trajectory estimation is to form the vector, A which approximates the approximate path traced by the stylus in x and θ. This trajectory vector is used as the framework for a local estimate of groove depth, from which the final audio signal is derived. Reasons for forming a stylus trajectory in this way prior to estimating the audio signal directly are:

- **Coherency**: In ensures that the entire audio signal is extracted coherently and that no local tracking errors occur (as is common with stylus reproduction, the stylus 'skips' across the surface).
- **Visualisation**: The trajectory overlaid onto the surface can be use to find potential problems with the groove structure, for example groove misalignment (in the case of reconstructed cylinders).
- **Robustness**: It ensures that the 'virtual stylus' has some inherent knowledge of where to go, in the case that it arrives at an obstruction or damage site.
- **Speed**: A coarse estimate for the stylus trajectory can be found quickly, and can provide the user with a rough idea of the audio content.

For an ideal cylinder recording, the helical trajectory produced by the cutting tool varies at a constant rate along the x-axis. When the surface is unwrapped to a 2-D matrix of surface heights (see Fig. 4.3), this ideal helix appears as a series of straight line trajectories, separated by the groove pitch λ_x, spanning from $\theta = 0$ - $360°$. Due to the eccentricity of most cylinder surfaces, coupled with the geometric projection of the sensor's imaging plane, trajectories generally appear curved. The tracking of these curves forms the task of stylus trajectory estimation.

Methods of stylus trajectory estimation have been developed for cylinders in two categories:

1. Cylinders that have a consistent groove structure and are generally in good surface condition.
2. Cylinders with a complex groove structure and/or with poor surface condition.

For cylinders in good condition (e.g. commercial Edison cylinders), trajectory estimation is fully automated, and requires only the measured surface data $Z(x_i,\theta_j)$, and measurement grid parameters (Δx, $\Delta\theta$) as inputs. This method uses the phase information (ϕ) of our linescan model, introduced in Eq. (1).

For complex or poor condition cylinders, (e.g. broken, damaged grooves, multi-session recording), some manual interaction is required, because the groove trajectory is often ambiguous and thus difficult to track automatically. This method is based on local feature point tracking of groove minima features.

Method for Cylinders in Good condition

Figure 4.13 shows that if we observe a single groove valley at linescan at $\theta = 0°$ and track its shift along the x-axis around the cylinder circumference until linescan at $\theta = 360°$, the x-position of the valley propagates through a distance of λ_x. For a cylinder recorded with constant pitch, this shift is approximately equal for all grooves. An estimate of this shift vector denoted by $\tau = \tau(\theta)$, can therefore be used to form an approximate stylus trajectory.

The shift vector $\tau = \tau(\theta)$, can found by estimating the relative phase (shift) between linescans from 0 to 360°. With k_0 determined from Eq. (2), the phase estimate of the *j-th* linescan is given by:

$$\varphi_j = \arctan\left[\frac{\mathrm{Im}\{\tilde{z}_j(k_0)\}}{\mathrm{Re}\{\tilde{z}_j(k_0)\}}\right], \quad j \in \theta : [0°, 360°] \tag{5}$$

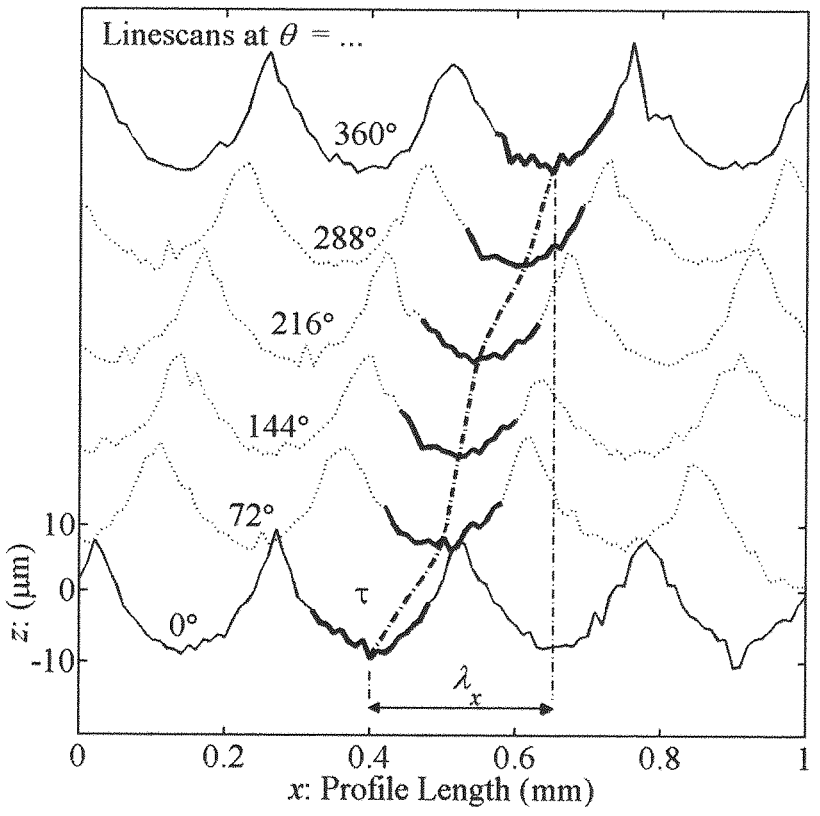

Figure 4.13: Linescan measurements taken at six angular intervals from $\theta = 0$ - 360° (linescans have been offset in z-axis to aid visualisation). The path traced by τ represents the shift in x of the highlighted groove valley. This shift vector is approximately equal for all groove valleys and can be estimated by calculating the relative phase shift between adjacent linescans.

The result from Eq. (5) is given in radians, thus in order to obtain τ, the phase angle vector φ_j, must be scaled such that a phase shift of 2π radians is equal to the distance λ_x. If the phase variation between successive linescans exceeds 2π (in the case of highly non-concentric cylinders for instance), a discontinuity occurs in φ_j. The procedure for constructing a continuous, natural phase variation is known as phase-unwrapping [83] and is accomplished by adding +/- 2π to φ_j, when absolute jumps between consecutive elements of φ_j are greater than or equal to π.

Discontinuous and continuous phase estimates from an Edison cylinder surface are shown in Figure 4.14.

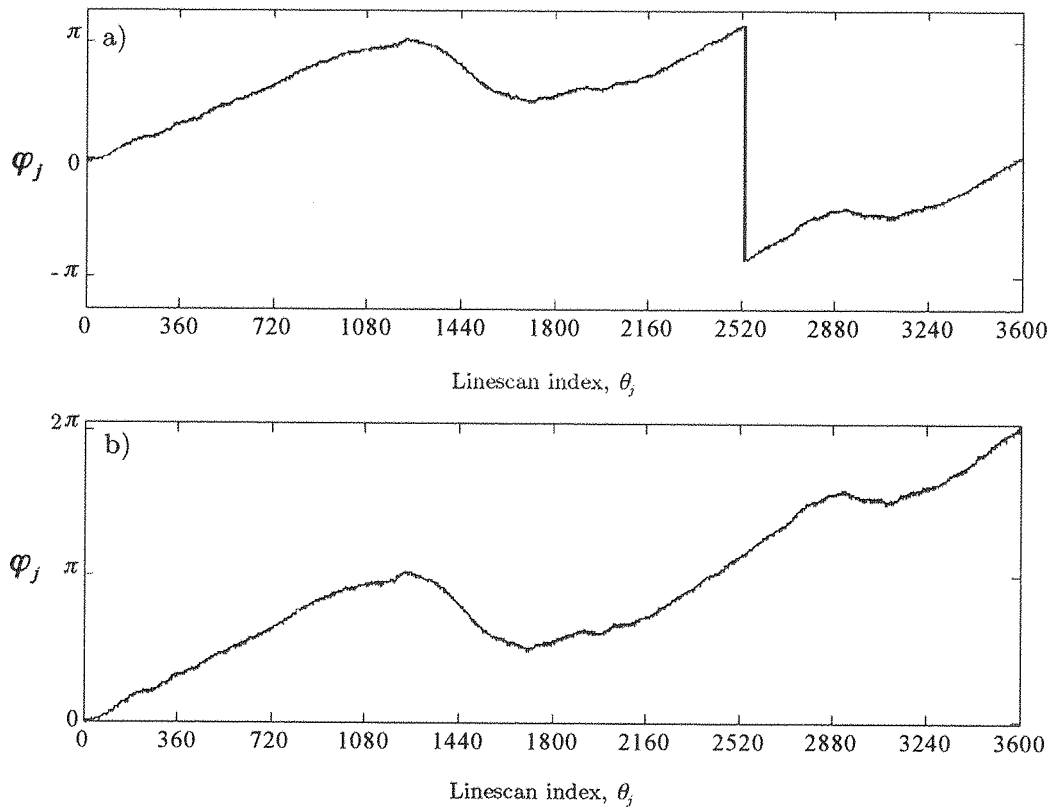

Figure 4.14: Trajectory estimation via examination of the phase angle at frequency k_0 , for all linescans around the cylinder circumference. Discontinuous phase estimate (a) and continuous, unwrapped phase estimate (b).

It should be noted that an equivalent trajectory estimate could be obtained using cross-correlation between linescans, but due to the linescan modelling step developed in Section 4.4.3, it makes sense to make use of the phase information already provided by the Fourier transform.

The resultant shift vector τ, is rounded to the nearest grid position along the x-axis, so that the discrete surface data can be referenced. Figure 4.15 shows examples of phase shift vectors τ, overlaid onto four different cylinder specimens. The vector τ is then used as a *template*, to form the approximate stylus trajectory for the full surface.

In general, τ characterises the average trajectory for all grooves of the recorded surface. However, in cases where the surface is heavily deformed/eccentric, the estimate is updated at intervals along the x-axis. The size of these update regions is chosen typically to be every 10 mm for standard 4" Edison cylinders. Figure 4.16 shows two

surface regions of an Edison cylinder (PBEAR) to demonstrate how the τ estimate varies for a highly non-concentric cylinder, from one end to the other.

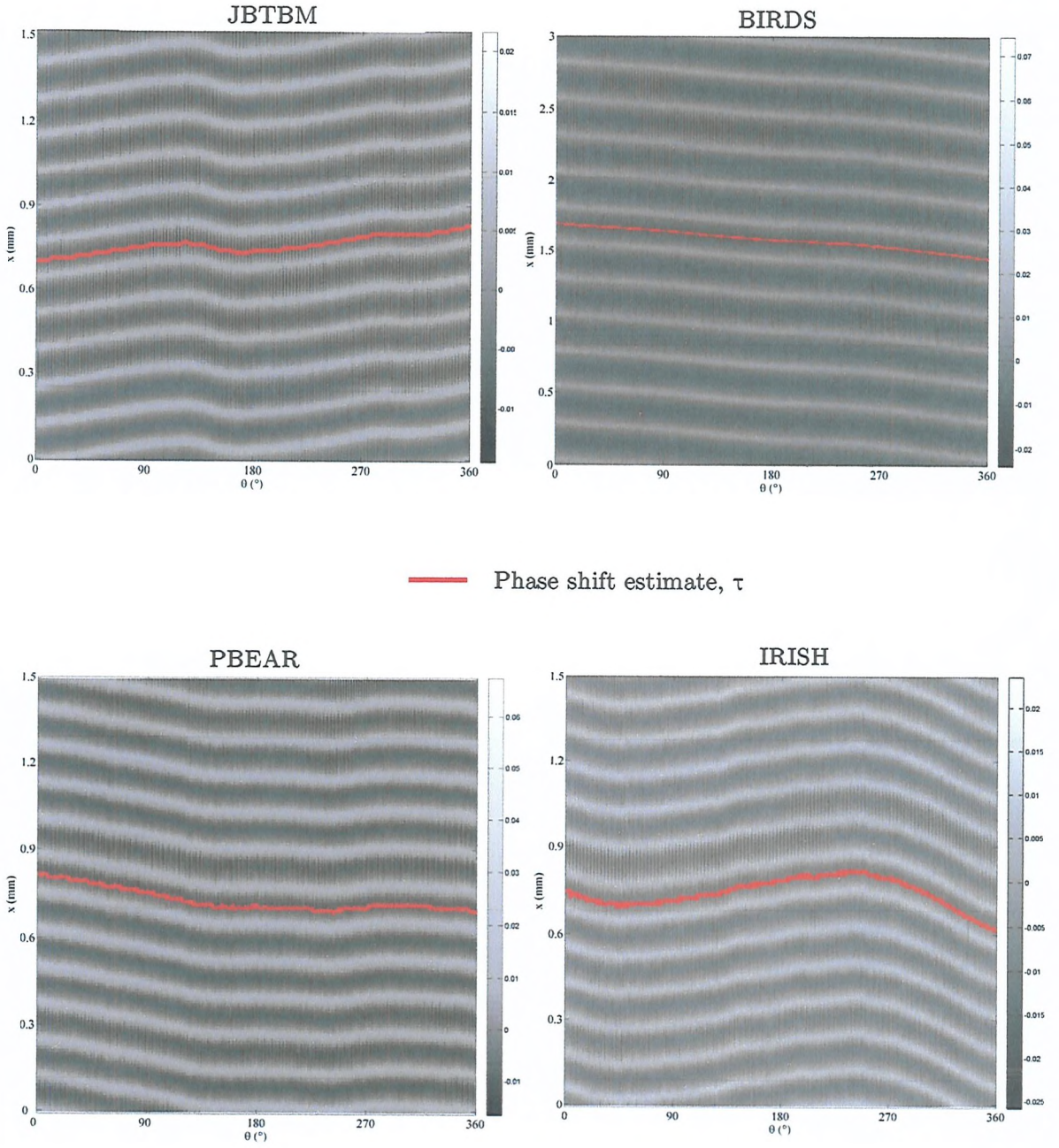

Figure 4.15: Four different cylinder surface plots with groove phase shift estimates τ overlaid. The cylinder BIRDS most closely resembles a straight line, suggesting that the surface is most concentric.
(Surfaces have been de-trended in the z-axis to aid visualisation)

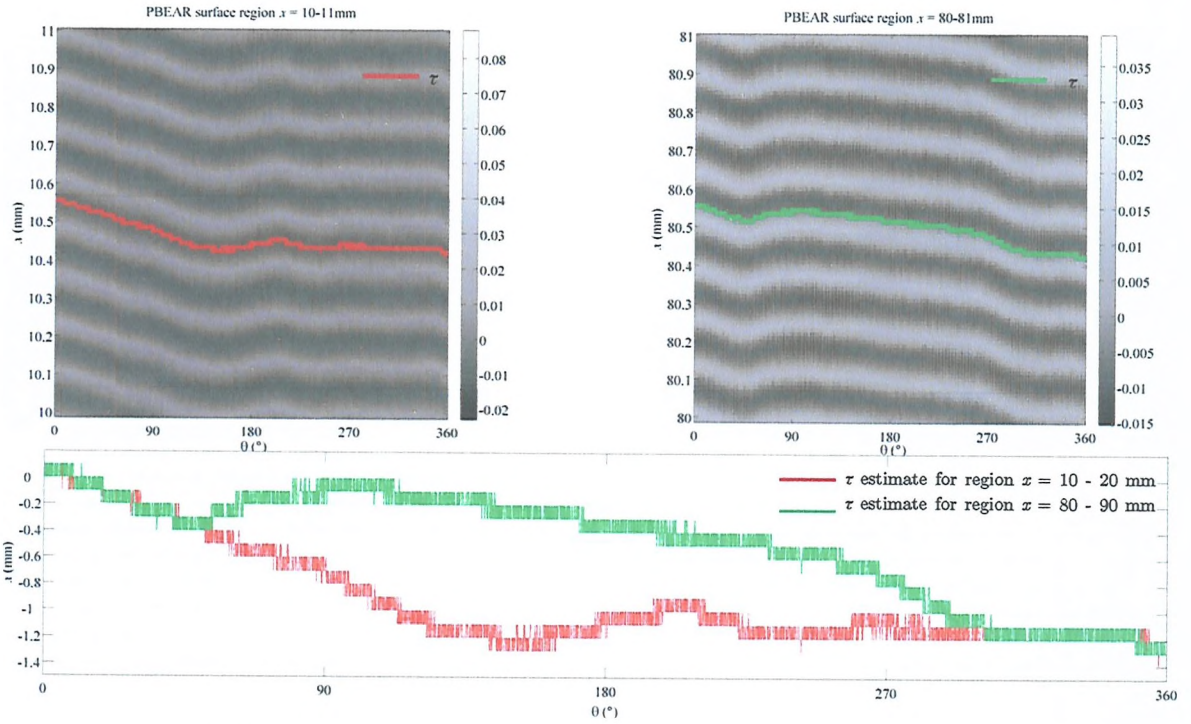

Figure 4.16: Two regions of the same cylinder, PBEAR (of recorded length = 90mm), demonstrating the variability in stylus trajectory curve τ from one length of the cylinder to the other. Reasons for this variability in τ are surface deformation, imperfect recording apparatus or a difference in geometric projection of the sensor's imaging plane from one length to the other.

Following groove seed detection, (as described in Section 4.4.5) an approximate stylus trajectory, containing x-θ point-pair coordinates, denoted by $\boldsymbol{A}_n$, is constructed (see Fig. 4.17b) by positioning τ at x-positions specified by the k-th groove seed, $g_1, g_2 \ldots g_k$:

$$\boldsymbol{A}_n \in \{x_n, \theta_n\} = \begin{bmatrix} [\tau + g_1]^T & [\theta]^T \\ \vdots & \vdots \\ [\tau + g_k]^T & [\theta]^T \end{bmatrix} \tag{6}$$

Where $\boldsymbol{A}_n$ is of size n rows by 2-columns, n is the number of groove cross-sections found along the trajectory. The column vectors $[\tau + g_1]^T$ and $[\theta]^T$ are both of length J, corresponding to the number of linescans measured over the cylinder circumference.

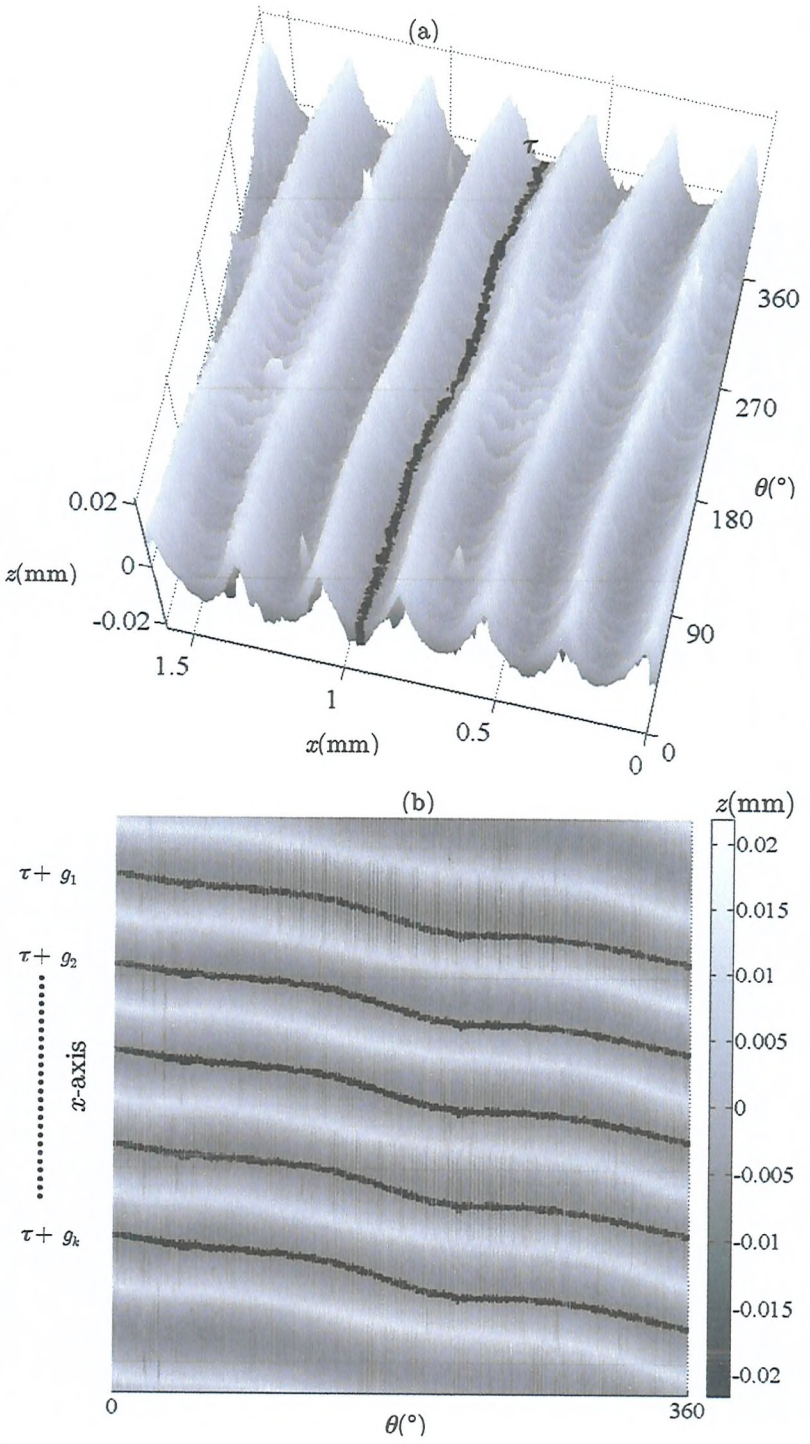

Figure 4.17: (a) Scaled, discretised shift estimation vector τ, overlaid onto a cylinder surface. (b) Formation of stylus trajectory vector A_n, using the phase shift estimate τ and groove seed positions, g_1, g_2 ... g_k .

The DFT method described provides a *global* (as opposed to *local*) method of trajectory estimation. By this we mean that the overall shift of the groove is assumed to be approximately equal over the full surface and a global. This global treatment of trajectory estimation has advantages over local tracking methods – namely its ability to

correct for linescan offset-error, introduced by the measurement process. Figure 4.18 shows the φ_j and τ estimates for a cylinder specimen scanned in 19 segments. This cylinder measurement exhibits linescan offset-error of up to 20 μm. Figure 4.18 also shows how the trajectory estimation stage has accounted for this linescan offset error, where local tracking methods may have failed to cope with discontinuities between segments.

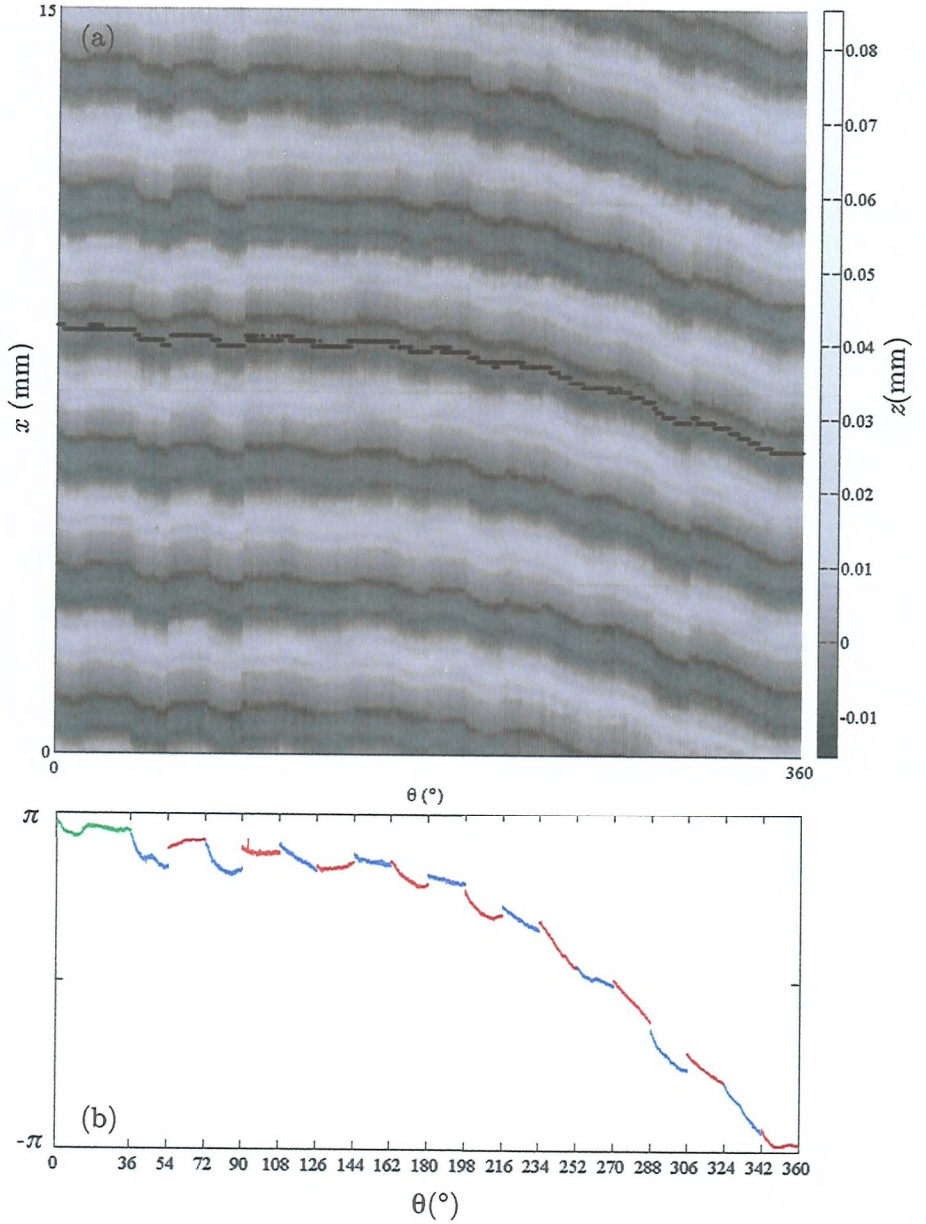

Figure 4.18: Example of (a) cylinder surface (TC1) exhibiting x-offset error, scanned in 19 segments (segment joins denoted by change in colour). The (b) groove phase shift estimate derived from the surface in (a), demonstrating the ability of this method to account for x-offset error.

Local tracking method for cylinders with complex groove structure

In the case of some specimens, stylus trajectory estimation via the DFT method described above will not warrant an accurate trajectory estimate, due to local surface complexity. Examples of cylinders with complex groove structures include:

- **Heavily damaged cylinders** – Grooves may be completely destroyed or follow a highly variable trajectory, which do not conform to the global helix of the recorded surface.
- **Broken recordings** – when broken fragments of cylinders have been re-assembled (see Figure 2.16b for example) the grooves do not generally match up exactly. This means that the helical trajectory is discontinuous, and cannot be well estimated using the phase shift estimation method.
- **Multi-session recordings** – some cylinders were not always recorded fully in one continuous helix, but as a series of independent recording events. Recordings of this type will generally have a blank region, between adjacent tracks, so that each track was distinguishable, and could be played back independently.

In the cases above, it is more appropriate to track each groove on a local basis, in order to account for extra complexity. To do this, a second trajectory estimation method was developed, based on the local tracking of groove features. This method of feature point tracking is similar to work carried out with images of flat disc recordings (See Fig 2.28 for instance) but in this case, the medial axis of the groove valley is tracked as opposed to grayscale image intensity. This procedure is as follows:

1) **SG filtering** - linescan data is smoothed using a cubic SG filter with the same parameters (P=3, L=λ_x) described in Section 4.4.4 to form the smoothed surface, $Z'(x,\theta)$. This stage ensures that the medial ridge-valley structure is enhanced, whilst reducing surface noise.

2) **Minima Detection** - using $Z'(x,\theta)$, a feature map of groove minima is produced by locating the local minimum point of the SG filtered surface, by the method described for groove seed detection in 4.4.5. Figure 4.11 showed how the SG filtered linescan has local minima positioned along the medial axis of the groove. Minima locations are stored in a logical array, with 'ones' at locations where minima have been detected, and 'zeros' elsewhere. These minima locations form the features to track. A minima map from a damaged and reconstructed cylinder, is shown in Figure 4.19.

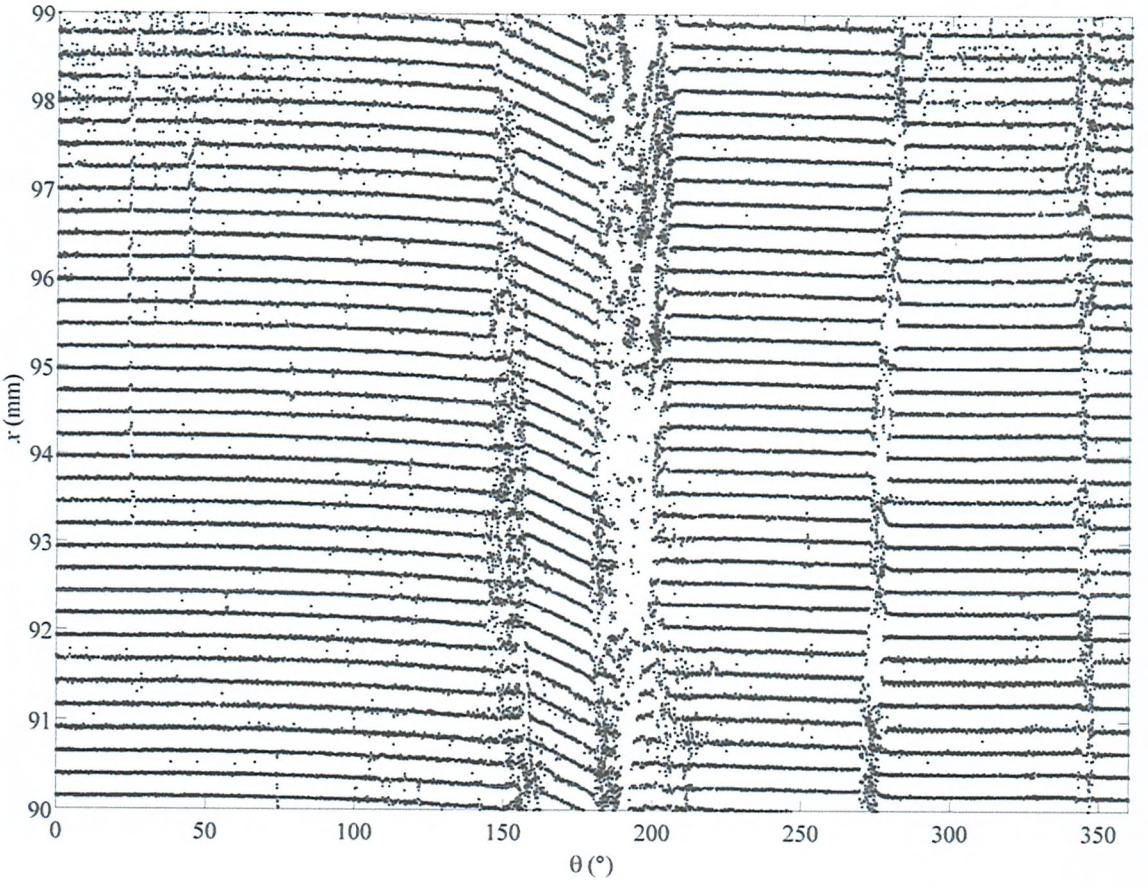

Figure 4.19: Minima map produced from damaged, reconstructed cylinder (EVAN). The main line structures indicate the medial axis of the grooves. Additional spurious feature points are due to damage sites.

3) **Feature Point Tracking** - Once all minima have been located, the next stage is to track feature points in the minima map, in order to estimate the trajectory produced by the cutting stylus. Initial algorithm development carried out by Fadeyev and Haber accomplished this tracking by examining each linescan individually. Groove minima positions in the next linescan were searched for in some minimal distance along the cylinder axis. There are two potential problems when applying this technique to a minima map:

i. The minimal distance constraint breaks down if minima exist in-between adjacent groove trajectories, due to damage sites etc. The effect of this can be that the track skips from one groove to the next.
ii. Computation time is high – for a cylinder with 700 grooves, scanned with 36,000 linescans, over 2 x 10^7 neighbourhood searches must be carried out.

To overcome these issues, a two-dimensional moving window denoted by W, is used to track the minima map, instead of treating each linescan individually. The dimensions of W are $[\lambda_x/2 \times N_\theta]$, corresponding to the linescan and circumferential dimensions respectively. The choice of $\lambda_x/2$ ensures that minima pixels inside W do not belong to adjacent grooves. The value of N_θ is dependent on the θ-grid sampling and groove pitch and is calculated such that in a range of N_θ linescans, the x-distance shifted by the groove trajectory ε_x , does not exceed half a linescan sample. Assuming a linearly varying trajectory, this is calculated by:

$$N_\theta = \frac{1}{2\lambda_x}\left(\frac{360}{\Delta\theta}\right) \tag{7}$$

Where λ_x is given in linescan samples, and $\Delta\theta$ is given in degrees. The calculation of N_θ ensures that within the limits of W , minima positions approximate a straight-line segment, which is useful for the tracking step. Following the design of W, tracking of minima features is carried out as follows (this is shown graphically in Fig 4.20):

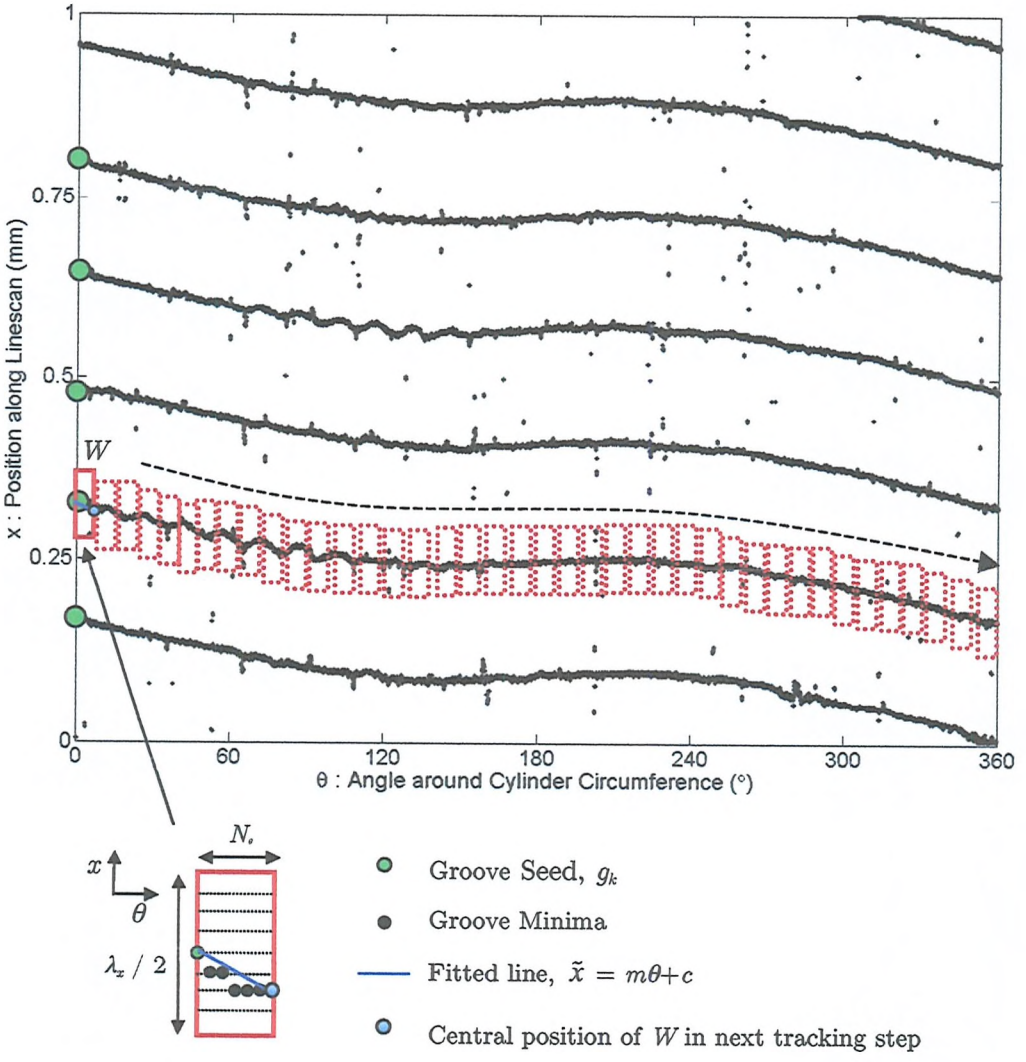

Figure 4.20: Figure to demonstrate the local tracking of groove minima using the 2-D moving window, W. (Minima map is derived from GRAPH surface)

I. Tracking starts at $\theta = 0°$, with the central x-index of W positioned at the first groove seed index, g_1.

II. A straight-line function: $\tilde{x} = m\theta + c$, is least-squares fitted to minima pixels inside the window W.

III. The window progresses by N_s samples along the θ-axis, and its central x-position is updated to equal the last ordinate position of $\tilde{x}$ found in (II).

IV. W advances in steps of N_s samples until reaching the linescan at $\theta = 360°$.

V. A spline interpolant is fitted to nodes given by the central positions of the tracking window W to form a single groove trajectory denoted by a_k, for the k-th groove.

VI. Steps I-V are repeated for all other groove seed positions, given in g_k.

VII. The full stylus trajectory A_n, is formed by concatenating the individual groove trajectories:

$$A_n \in \{x_n, \theta_n\} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{bmatrix} \tag{8}$$

Figure 4.21 gives an example of the trajectories formed by this local minima tracking method.

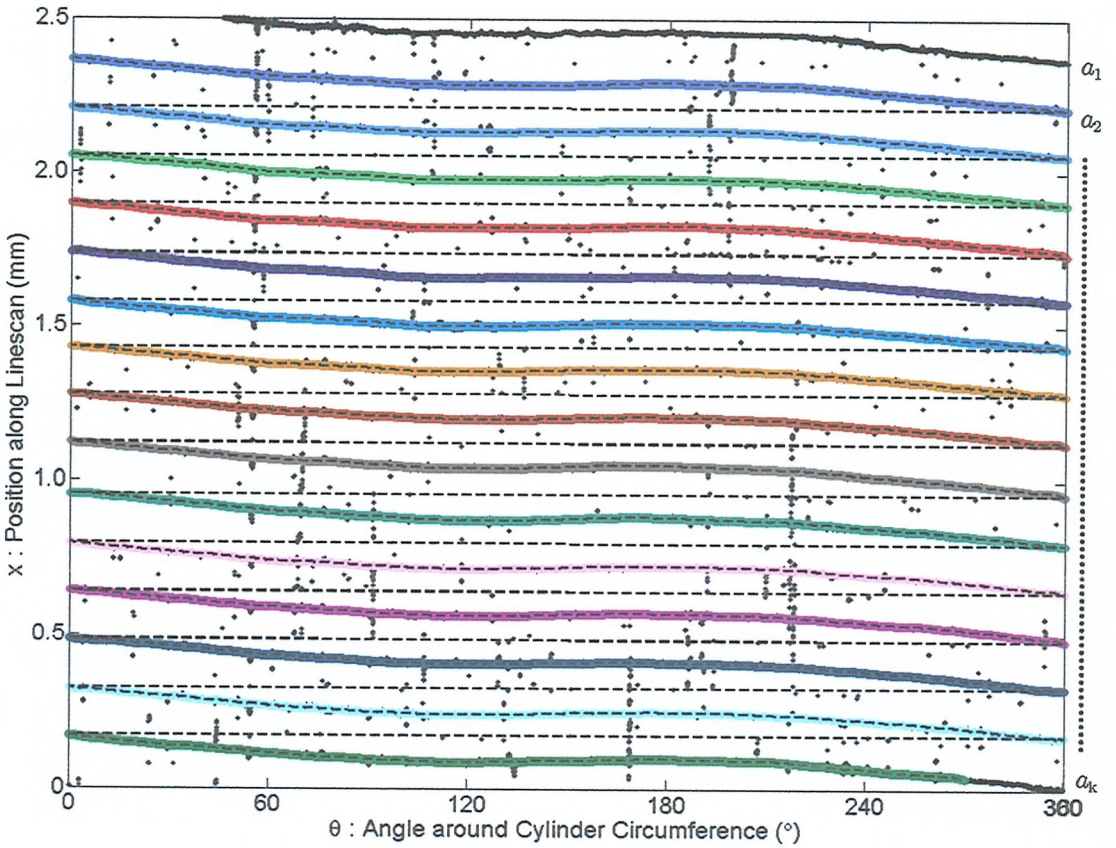

Figure 4.21: Example of minima map (from GRAPH) with stylus trajectory vector overlaid. Individual groove trajectories: a_1, a_2.., a_k , found in the minima tracking-stage are depicted by different colours.

Manual tracking

In the case of heavily damaged surfaces when the groove structure is ambiguous, (for example, the minima map shown in Fig. 4.19), some manual tracking is required, to ensure that the stylus trajectory follows a coherent path. In this case, it is necessary to manually modify the nodes produced by the minima-tracking step, to ensure that grooves line up correctly. This can be accomplished interactively in MATLAB. The red region in Figure 4.22 gives an example of manual tracking for the minima map shown in Figure 4.19. In the case of this cylinder, (which was broken and reconstructed), the correct groove structure was too ambiguous to be followed completely automatically, and needed to be tracked manually.

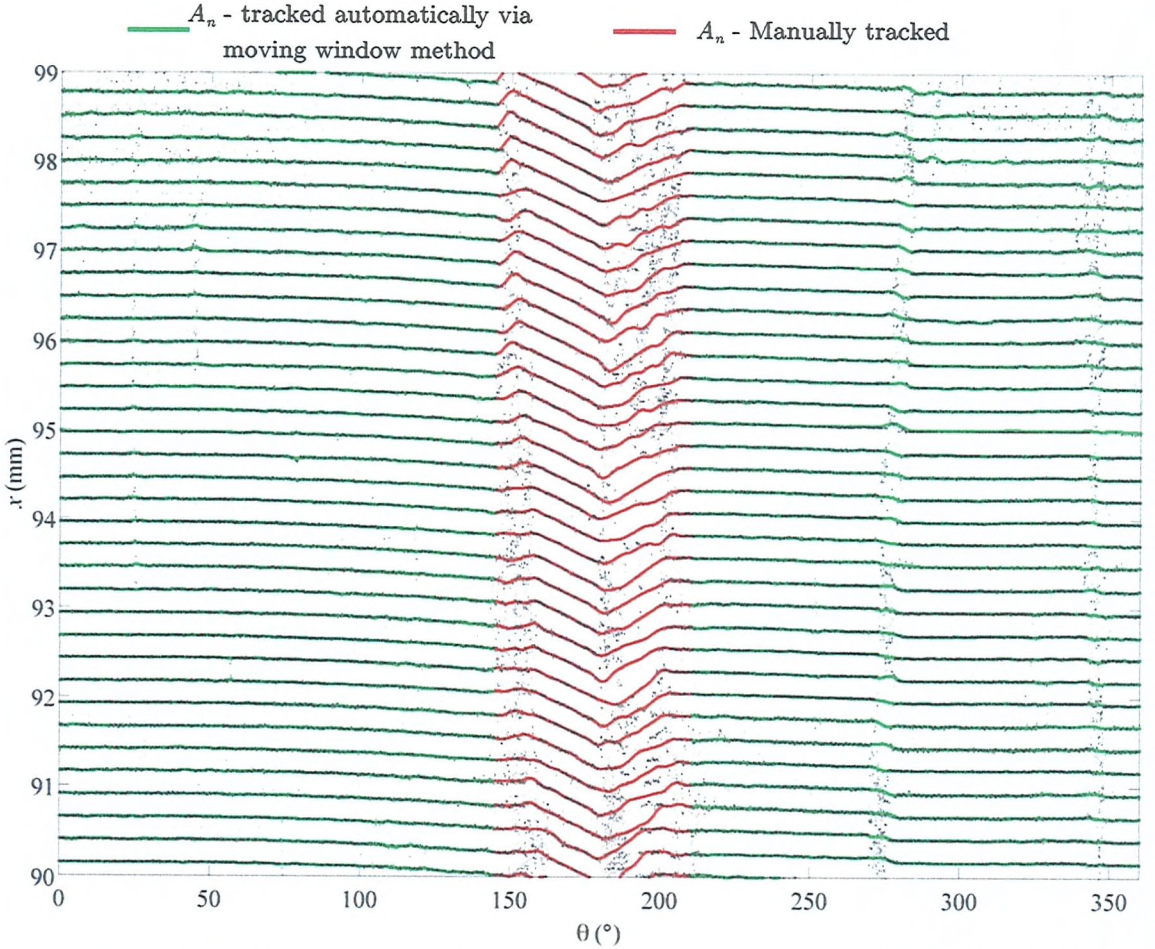

Figure 4.22: Stylus trajectory, A_n derived from the minima map shown in Fig. 4.19 via a combination of manual and moving window tracking methods. This particular surface (EVAN) has a complex damage structure towards $\theta = 180°$ and manual tracking was required here to ensure that the grooves lined up appropriately.

4.4.7 Groove Depth Estimation

The trajectory A_n, determined in 4.4.6, gives an approximation to the x-θ path followed by a reproduction stylus. In general, this trajectory indexes the medial axis of each groove cross-section, and is used as the initial location to estimate the groove depth.

The signal, $z_a(n)$ is defined as the local radial displacement of the groove cross-section, corresponding to the depth trace of the encoded sound. The n-th sample in this sequence is obtained by forming an estimate of the groove depth, in the vicinity of the n-th trajectory position: $A_n \in \{x_n, \theta_n\}$.

It should be noted that a number of different estimates for $z_a(n)$ can be derived from the same specimen, for example through model fitting methods (e.g. quadratic fitting in [5]), or by averaging samples across the groove cross-section. In order to determine the optimum method for estimating $z_a(n)$, we start by considering the composition of the optically measured groove. In the following section, six criteria for estimating $z_a(n)$ are stated and several estimation schemes are considered.

Composition of Optically Measured Groove

A consideration of the measured groove cross-section composition from ideal cutting facet to measured groove is shown in Figure 4.23. Ideally, we would like to obtain the original depth trace produced by the cutting facet. This is impossible in practice due to the following factors:

- Micro-roughness – inherently caused by surface imperfections.
- Groove wear – inherently caused by stylus playback or poor storage.
- Debris – caused be foreign objects that lie above the groove image plane.
- Sensor and quantisation noise – inherent to the measurement system.

Despite the outlier detection stage in Section 4.4.4, samples attributed to local damage/debris may still exist across the groove cross-section. In order to minimise

noise which can be introduced by surface roughness and debris measured by the optical method, the first two criteria for estimating $z_a(n)$ can be stated as:

CRITERION 1: The estimate for $z_a(n)$ should incorporate some form of smoothing across the groove cross-section, in order to suppress higher spatial frequencies, attributed to microscopic surface roughness and/or measurement noise.

CRITERION 2: Samples across the groove cross-section attributed to damage and/or debris, should be identified and not included when estimating $z_a(n)$.

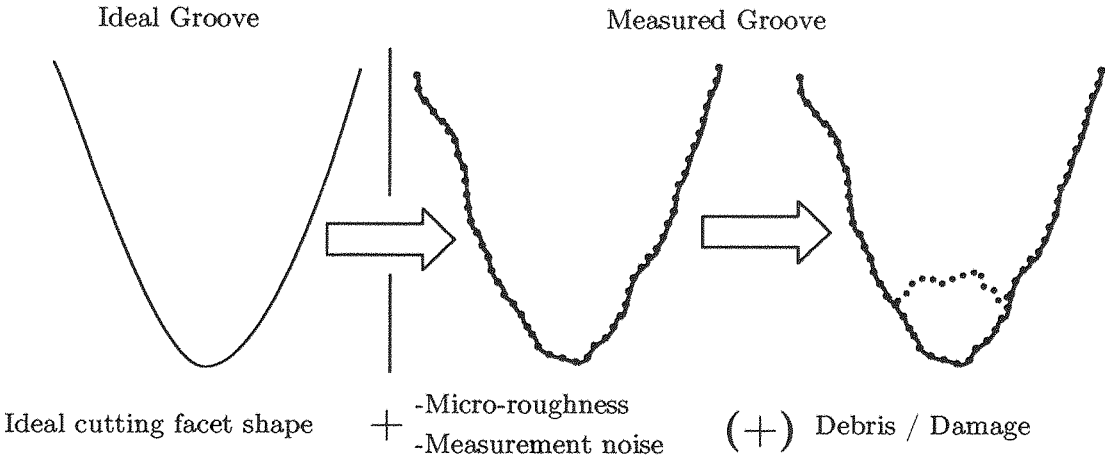

Figure 4.23: Consideration of cylinder groove cross-section from original cutting facet to measured groove.

For an ideal vertically cut groove, the cutting facet modulates in the z-plane as a solid body. This means that in theory, if one tracks a consistent x-position across the groove cross-section, for all grooves, the relative variation in depth (z-axis) should be the same, regardless of x-position. Figure 4.24 shows three depth variations, ε_{x1}, ε_{x2}, and ε_{x3} from consecutive linescans taken at grid positions, x_1, x_2, and x_3. In this example it is clear that $\varepsilon_{x1} \neq \varepsilon_{x2} \neq \varepsilon_{x3}$, hence the solid body assumption does not hold true. Reasons for this include:

- Local regions of the groove-cross section are likely to have been subject to varying amounts of deformation, as a result of stylus playback.
- Debris and mould growth are likely to be randomly distributed over the surface.
- Quantisation and discretisation error due to the measurement process.
- The recording process is imperfect.

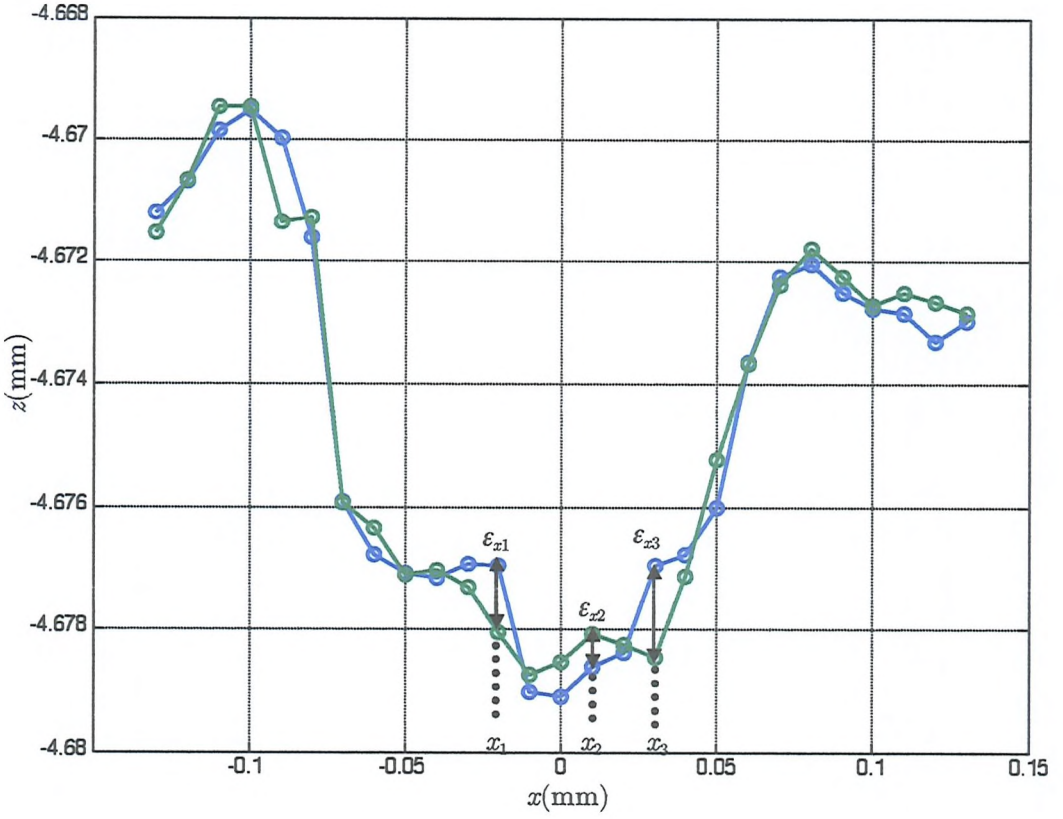

Figure 4.24: Two groove cross-sections taken from adjacent cylinder linescans (TC1). The three arrows indicate depth variations, ε_{x1}, ε_{x2}, ε_{x3} at consistent x-position across the groove cross-section. The variability in distance suggests that we cannot assume that the groove moves as a solid body.

Due to the lack of the solid body assumption and conditions stated above, the third criterion for estimating $z_a(n)$ is:

CRITERION 3: The estimate for $z_a(n)$ for a given cross-section should not be based on a single z-sample, but by a set of samples across the groove.

Groove Ridge Data

Care must be taken when including depth information from the groove ridge in order to avoid introducing temporal distortion (echo). In [78], it was observed that estimates for $z_a(n)$ which used samples towards the groove ridge exhibited temporal distortion, or 'echo'. This phenomenon was first noted with the laser-beam reflection method in [40], when the reproducing laser beam was wide compared with the groove width. Similarly in [5], Fadeyev et al. note that groove ridge data, provided audible, albeit noisy sound content, but with interference from different times in the recording. Echo is introduced due to the superposition of adjacent grooves at times t and $t+T$, where T is the time taken for one complete cylinder rotation (0.375 seconds in this case for 160 rpm). This leads us to the fourth criterion for estimating $z_a(t)$:

CRITERION 4: In order to avoid introducing temporal distortion, the estimate for $z_a(n)$ should not use data in the vicinity of the groove ridge (top).

Rationale for 1-D Estimate of Groove Depth

Due to the fact that the surface is built up by linear scans, measured independently of one another, and coupled with the issue of x-offset error (shown in Chapter 3, Figure 3.20), it is more appropriate to consider an estimate for $z_a(n)$ on a 'linescan-by-linescan' basis, along the x-axis only (as opposed to obtaining a smoothed estimate in both x and θ simultaneously). This rationale was also adopted in the work by Fadeyev et al. [5]. It is assumed that despite x-offset errors, the relative z-values (however shifted in x) must still be valid, for each groove cross-section.

With stylus reproduction, if a reproducer slides laterally (along the x-axis) across a vertically modulated groove, distortion will result, because the points of contact are no longer at a consistent depth in the groove for all time instances. A similar distortion may result with NCFSM signal reproduction if the $z_a(n)$ estimate is taken from inconsistent positions across the groove. These considerations lead to criterion five:

CRITERION 5: Each sample in $z_a(n)$ should be estimated on a linescan-by-linescan basis, at a consistent x-position across the groove.

Groove Shape

With stylus replay, reproducers of different geometry are often experimented with, in order to obtain desirable signal reproduction. This experimentation is due to the fact that the cutting facet used for many early recordings was not standardised, and so an ideal reproducer is initially unknown. Moreover, cylinders are likely to have undergone varying amounts of playback, and hence exhibit unique wear characteristics.

Figure 4.25 shows the clear variability in groove cross-section for three different cylinder specimens measured by the NCFSM method. Knowledge of the original groove shape is therefore important for signal reproduction and will differ for each specimen. This consideration forms the final criterion for estimating $z_a(n)$:

CRITERION 6: The average groove shape for a particular specimen should be known when estimating $z_a(n)$.

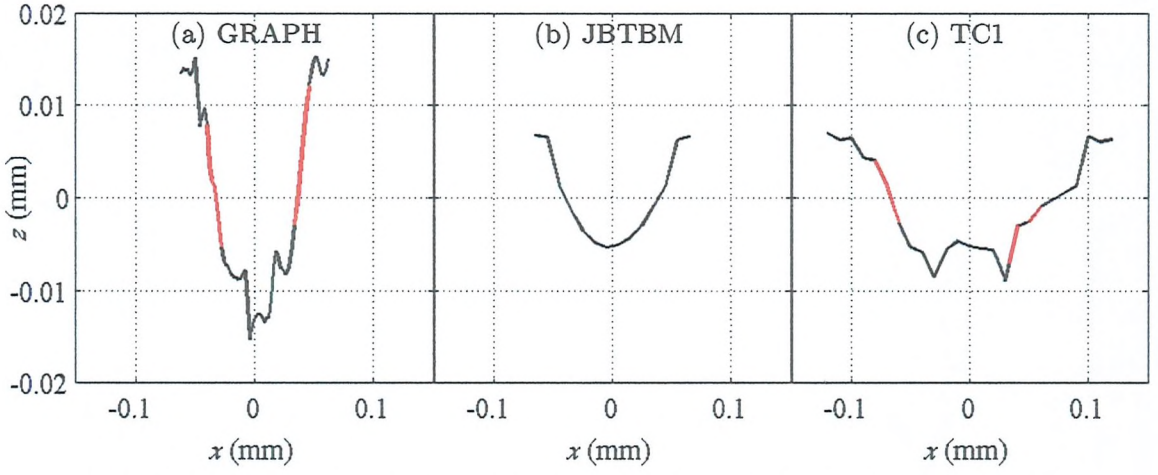

Figure 4.25: Average groove shape from three cylinder specimens. In (a), the groove (GRAPH) shows a wear feature produced by a reproduction stylus. In (b) the groove (JBTBM) has a parabolic cross-section. In (c) the groove (TC1) has an irregular shape, due to the manufacturing process. (Red denotes regions where data is generally missing due to the angular tolerance of the sensor)

Schemes for Estimating Groove Depth

In this thesis, the following schemes have been considered for obtaining an estimate for $z_a(n)$:

i. Single point z-data streams, (using raw z-data, no averaging of samples) [84]

ii. Averaging samples across the groove cross-section [85].

iii. SG Polynomial filtering method [85].

iv. Model fitting methods [5].

Scheme (i) : Single point z-data streams

This is the simplest estimation method, whereby $z_a(n)$ is derived from a single raw z-value found along the stylus trajectory vector A_n. For example, one might take the minimum point of each groove found along A_n as an estimate for $z_a(n)$, or a data-stream that was 30 μm to the left of the groove's medial axis for instance.

Examples of single point z-data streams are shown in Figure 4.26. Scheme (i) was investigated qualitatively in [84] by examining data streams taken from three different positions across the groove cross-section. The cylinder in question was a moulded Blue Amberol cylinder (JBTBM) with smooth, symmetric groove cross-section (as shown in Fig 4.26). A frequency analysis of the three signal estimates depicted in Figure 4.26 is shown in Figure 4.27. This figure shows how the signal-to-noise ratio (SNR) varies as a function of x-position across the groove cross-section. In general for this specimen, the SNR increases as the depth estimate approaches the groove bottom. This is to be expected for a parabolic groove shape, because the curvature of the groove is steeper at the side-wall, than at the bottom. The variation in z from data streams taken at the groove wall is therefore greater than at the groove bottom, and this increased variance leads to a greater noise level in this case.

The use of single-point data streams relies on the solid body assumption holding true, which, as Figure 4.24 showed, is not the case, (especially for noisy groove cross-sections). Although this scheme does not comply with Criterion 3 (and others), it is

worthy of consideration, because in theory, local regions of the groove may have undergone varying amounts of wear. In this situation, taking z-data streams from undamaged regions across the cross-section could result in a better signal.

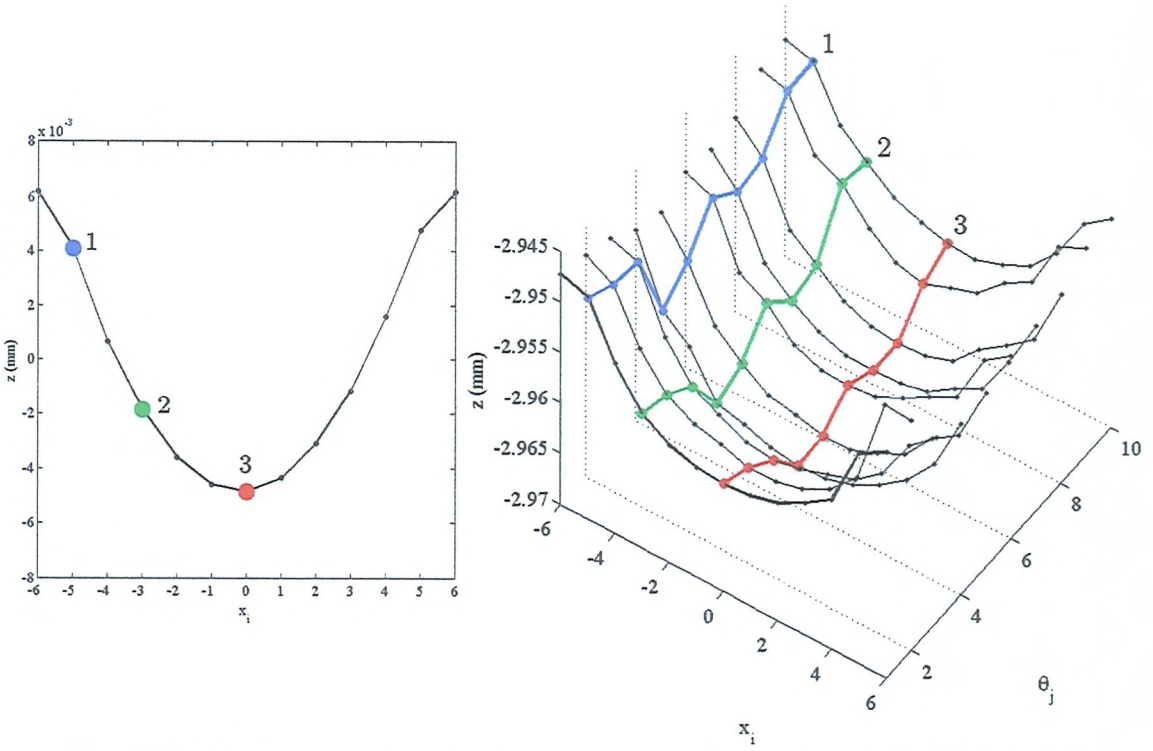

Figure 4.26: Single-point data streams estimates for $z_a(n)$ taken at three different positions across the groove cross-section (JBTBM).

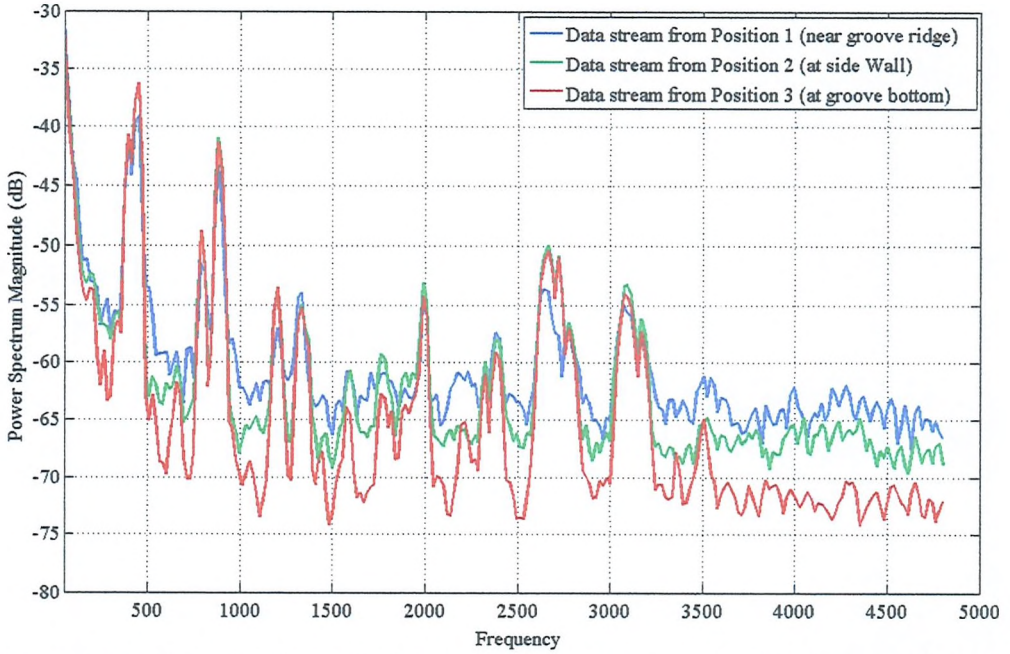

Figure 4.27: PSD estimates derived from the data-streams depicted in Figure 4.26. In this example, the signal-to-noise ratio increases as the data stream position tends towards the groove bottom.

Scheme (ii) : Averaging Samples across the groove cross-section

This scheme aims to address **Criterion 1** & **3**, whereby $z_a(n)$ is obtained by taking the mean depth of a distribution of samples across the groove cross-section, found along the stylus trajectory vector A_n. To explore averaging of samples, we introduce the notion of a 'groove matrix'.

The groove matrix, denoted by $\boldsymbol{G}$, is simply a matrix of groove cross-sections (z-values), found along the trajectory vector A_n. $\boldsymbol{G}$ is of size n-rows by m-columns where n is the number of groove-cross sections (equivalent to the number of samples in the time domain) and m is the number of samples across the groove-cross section ($m = \lambda_x$).

An estimate for $z_a(n)$ is derived from $\boldsymbol{G}$ by an un-weighted mean averaging process across the n-th row of $\boldsymbol{G}$.

$$z_a(n) = \frac{1}{N} \cdot \sum_{i=1}^{N} G(n,i) \tag{9}$$

The mean average is taken from the central column of the $\boldsymbol{G}$, which corresponds to the medial axis of the groove cross-section. The number of points included in the average varies from N=1 (central column of the $\boldsymbol{G}$ only) to N=25 points (±12 samples either side of the central column of $\boldsymbol{G}$).

The advantage of averaging an ensemble of points (as opposed to taking a single depth sample) is a suppression of high frequency noise, caused by surface roughness. Scheme (ii) therefore complies with **Criterion 1** & **3**.

Averaging is a low pass filtering operation since its effect is to allow lower spatial frequencies to be retained, while suppressing higher frequency components. The mean of a larger distribution will remove more surface noise, but will result in a loss of signal power. Care must be taken in the limit where $N \rightarrow \lambda_x$, because ridge data from adjacent grooves can introduce temporal distortion (**Criterion 4**).

There are problems associated with averaging samples to obtain $z_a(n)$. Applying sample averaging to cross-sections without first removing outliers (**Criterion 2**) is not

ideal, because outlying z-data points will skew the distribution, resulting in a loss of signal accuracy. Additionally for vertically cut grooves, the distance between adjacent groove ridges is dependent on the depth of cut (see Fig. 4.2), thus the number of samples attributed to a single groove cross-section can varies throughout the duration of the recording. This suggests that obtaining the mean of a fixed number of samples N for all groove cross-sections could introduce bias, if N is greater than the smallest groove width encountered in the specimen.

Scheme (iii): SG Polynomial filtering method

A third method has been developed [85] which utilises the SG polynomial smoothing filter to obtain an estimate for $z_a(n)$. The rationale behind this method is to incorporate a filter that suppresses surface roughness (**Criterion 1**), and also ensures that depth estimates are taken from a consistent x-position across the groove (**Criterion 5**). In addition, different combinations of SG filter parameters can be experimented with, in the same way that different reproducers are used with stylus reproduction. The method for obtaining an estimate for $z_a(n)$ is as follows:

- Raw surface data is filtered along each linescan via an SG filter of polynomial order P, and frame length L to form the SG filtered surface, $Z'(x_i,\theta_j)$.
- Along the trajectory estimate A_n the radial displacement estimate $z_a(n)$ is found by locating the local minimum point of the filtered surface: $z_a(n) = min\{ Z'(A_n) \}$
- SG filters with different combinations of order $P = \{1, 3... 9\}$ and frame length $L = \{11, 13,... 25\}$ can be used to obtain unique groove minima at different positions in the groove cross-section.

In this method, the SG filter is used to smooth the groove cross-section, in order to reduce the effect of surface roughness, caused by the cutting tool. The local minimum of the smoothed cross-section, $min\{ Z'(A_n) \}$ is then used as an estimate of $z_a(n)$. The

rationale behind this method is that by exploring different SG filter types of varying polynomial order and frame length, a number of unique estimates for $z_a(n)$ can be found (see Fig. 4.28 for example). This method is therefore analogous to experimenting with different styli tips, as is common with stylus reproduction.

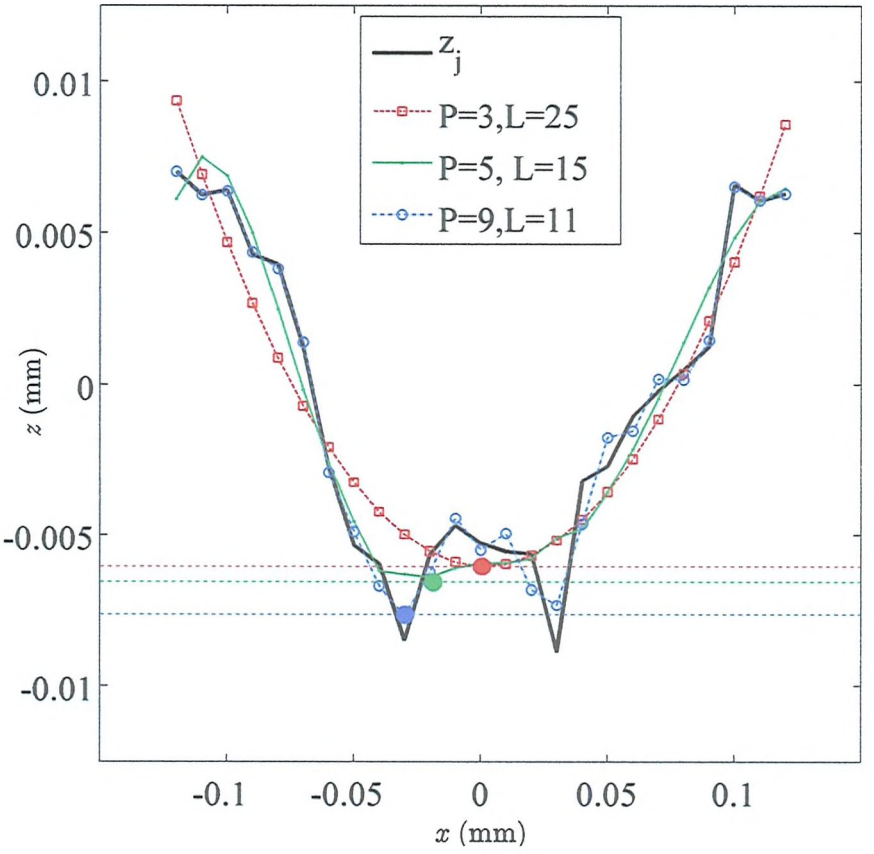

Figure 4.28: Figure to demonstrate estimating z_a(n), via different SG filters. Horizontal lines correspond to the depth estimate of the given SG filter with P (polynomial order) and filter length (L, in samples).

The SG filtering scheme described meets more of the six criteria than schemes (i) & (ii) and has the advantage of being adaptable to different groove types (by using different combinations of P and L). It does not however provide any inherent error checking, to account for damage and debris. The issue of impulsive noise rejection (**Criterion 2**) and groove shape (**Criterion 6**) implores us to explore scheme (iv).

Scheme (iv) : Model fitting Methods

In scheme (iv), a groove shape model $M(x)$, is fitted to the discrete groove cross-section. The minimum point of $M(x)$ is then used as the estimate for $z_a(n)$. One of the advantages of this scheme is that samples attributed to debris/damage, (which may not have been removed in stage 4.4.4) can be identified and removed by examining the similarity of raw z-data with $M(x)$. This means that impulsive noise can be reduced and a more accurate estimate for the true groove depth is obtained.

Model-fitting methods rely on the groove shape model $M(x)$ bearing strong similarity to the measured groove cross-section. For cylinders in good surface condition, the shape of the groove cross-section does not vary greatly over the recording duration. Exceptions to this rule will be regions of groove damage, debris sites and periods of extreme loudness or quietness during the recording.

A constrained quadratic fitting method was previously described in [5] for a Blue Amberol cylinder with smooth, symmetric groove cross-section, (shown in Fig. 4.25b). For this particular case, the parabola model agreed well with the groove shape. In [5], outliers we identified as those which deviated from the constrained quadratic fit by more than 3σ, where σ was the individual point error, based on the width of the distribution of fitted residuals. Outliers were removed, and then the fit was re-iterated. Fit iterations stopped if no points deviated by more than then 3σ (or only 5 samples across the groove remained).

For some cylinder specimens, (see Fig. 4.25a-c for example), the groove cross-section is irregular, asymmetric and is not necessarily a parabola. This means that constrained quadratic fitting, as described in [5], is not appropriate to all cylinder specimens and is therefore discounted as an optimum solution for *all* cylinder specimens.

In order to account for **Criterion 6** and to design a model-fitting method suitable for all cylinder types, a unique groove shape model $M(x)$ must be derived directly from the surface data for each specimen. The method for forming $M(x)$ is as follows:

- Using A_n, form the Groove matrix $\boldsymbol{G}$, from $Z(x_i,\theta_j)$. $\boldsymbol{G}$ is of size n by λ_x samples, where n is the length of A_n (the number of x-θ point pairs) and λ_x is the number of samples across the full groove cross-section.
- The mean of each groove cross-section is subtracted from each row in $\boldsymbol{G}$ to form the groove matrix $\boldsymbol{G}_*$ with 'zero-mean'.
- A groove shape template, $M(x)$ is formed by taking the mean across columns in $\boldsymbol{G}_*$.
- To avoid fitting to samples at the groove ridge (**Criterion 4**), an additional constraint of the groove model is to consider samples only in the lower half of $M(x)$. It is assumed that the shape of groove in the lower half is a consistent characteristic for all groove cross-sections, regardless of depth. Examples of groove shape models derived from two different cylinder specimens are shown in Figure 4.29.

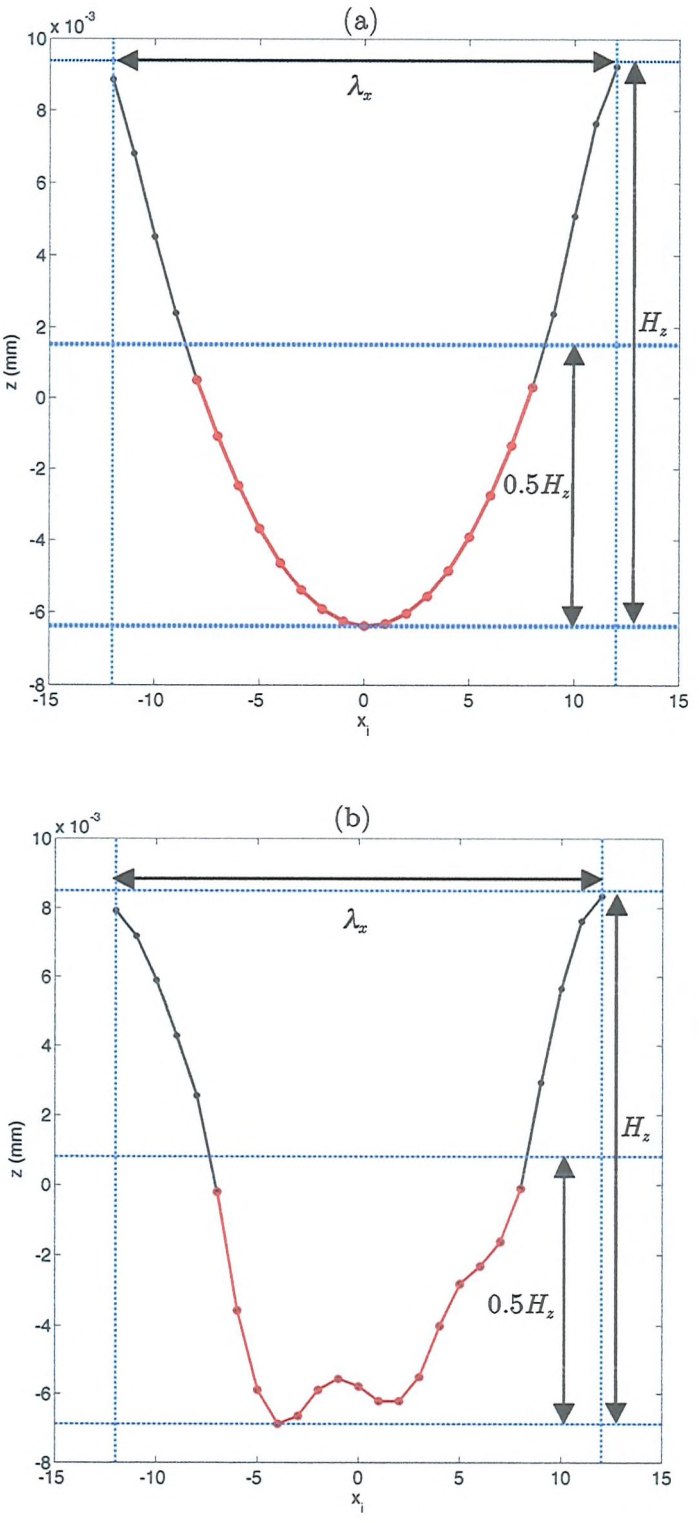

Figure 4.29: Groove shape model for two different cylinder specimens: (a) Black Amberol cylinder (BIRDS) and (b) custom-made wax cylinder (TC1). The model used for fitting, $M(x)$ is formed using only samples from the lower half of the groove cross-section to account for deep and shallow grooves, with different groove widths, and to avoid fitting to the groove ridge.

In [5], the constrained quadratic function is fitted via method of least squares. In the method described here, the groove model $M(x)$ is fitted to the n-th groove cross-section, G_n via a least-squares minimisation to form the fitted model, $F_n(x) = M(x) + c$. (see Fig. 4.30). We want to determine c, such that the distance between fitted model F_n and G_n is minimised. Using the method of least squares, the residual square error between F_i and G_i which we wish to minimise is given by:

$$R^2 = \sum_{i=1}^{n}\left[G_i - F_i\right]^2 = \sum_{i=1}^{n}\left[G_i - (M_i + c)\right]^2 \tag{10}$$

Since we are seeking to find the constant, c, which gives the minimum error, by differentiation we obtain:

$$\frac{\partial(R^2)}{\partial c} = -2\sum_{i=1}^{n}\left[G_i - (M_i + c)\right] = 0$$

$$\sum_{i=1}^{n}\left[G_i - (M_i + c)\right] = 0$$

$$\sum_{i=1}^{n} G_i - \sum_{i=1}^{n} M_i - nc = 0$$

$$c = \frac{1}{n}\sum_{i=1}^{n}\left[G_i - M_i\right]$$

So the required fitting function is simply:

$$F_i = M_i + \frac{1}{n}\sum_{i=1}^{n}\left[G_i - M_i\right] \tag{11}$$

The groove depth estimate for the n-th groove section is then given by the minimum ordinate of $F_n(x)$: $z_a(n) = min\{F_n(x)\}$

The method of least squares fitting assumes that groove data has a normal (Gaussian) distribution. This means that if outliers exist in G_n, the fitting becomes inaccurate. Despite the outlier removal stage described in Section 4.4.4, local outliers

(which do not resemble the groove shape template) may still exist and need to be removed from the fit. At each x-offset samples which deviate from the fitted template by more than $0.5H_z$ were considered outliers and removed the fitting process is reformed with outliers set to *NaN*. For the n-th groove, if 50% or more of the samples in the n-th groove cross-section G_n are equal to *NaN*, a valid depth estimate could not be obtained, and the value of z_a(n) is set to *NaN*.

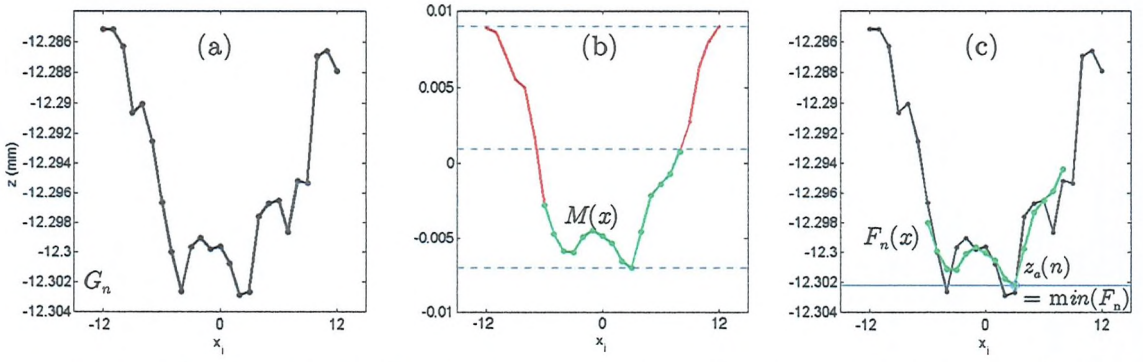

Figure 4.30: Procedure for fitting groove shape model. (a) raw groove cross-section G_n, (b) the groove shape model, (with lower half shown in green) and (c) the fitted groove shape model via method of least squares to obtain groove depth estimate $z_a(n)$.

The trajectory estimate found in 4.4.6 provides an approximation to the medial axis of the groove cross-section. In order to ensure that the model M(x) is positioned correctly in the groove, a local 2-D cross-correlation is carried out. The fitted model F_n is obtained at local offsets about the x-position, given by A_n. The fitted groove model with x-offset τ, is denoted by, $F_n\ (x + \tau)$. The local x-offset neighbourhood k is set to ±30 μm, which is the maximum expected drift assumed around the initial estimate in A_n. This procedure also ensures that any remaining outliers not identified in Section 4.4.4 are not included in the fitting process for F_n. This method is shown graphically in Figure 4.31 and is now described:

- For the n-th trajectory position given by A_n, perform a sliding-correlation with $M(x)$ and the $Z(A_n)$, to find $F_n\ (x + \tau)$ at offset $\tau\ = [-30\mu m : \Delta x : 30\mu m]$.
- Obtain the x-offset, τ_{min} with minimum, mean absolute z-point error, min $\{|C(\tau)|\}$

- At this offset check for outlying data points, based on a z-point error threshold.
- Remove outliers at this offset and re-perform previous steps until no outliers exist.
- Take $min\{F_n(x)\}$ as the depth estimate for $z_a(n)$.

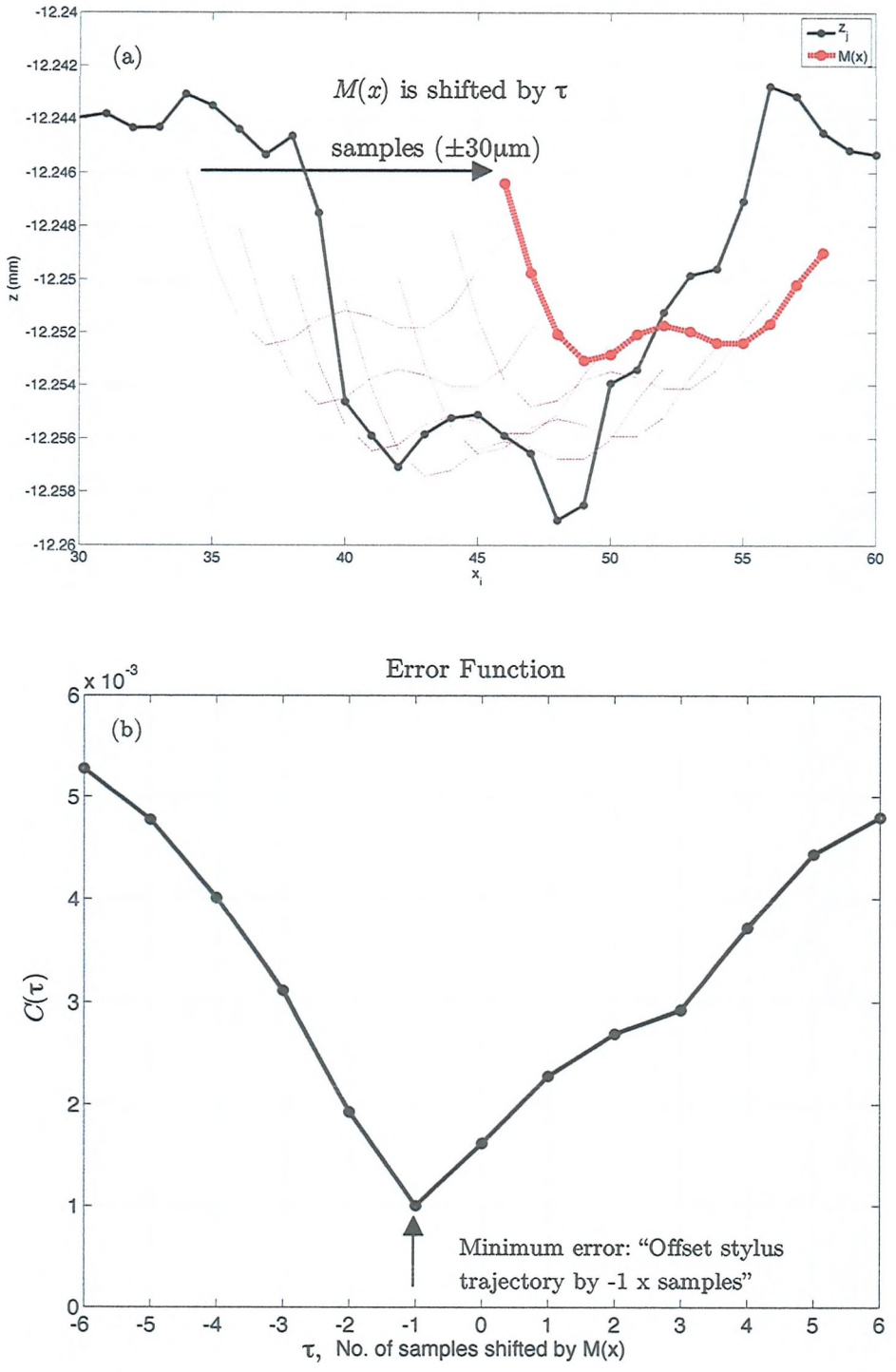

Figure 4.31: (a) The groove shape model, $M(x)$ is shifted across the linescan data z_j and the absolute mean error between the model and linescan data is computed. The error function shown in (b) is used to update the initial trajectory estimate, A_n.

Additional Outlier Removal from Depth Estimate

For each groove cross-section, the residual fitting error between the groove model $M(x)$ and raw data is stored in the vector $\varepsilon_M(n)$. This error vector ε_M gives a 'groove shape likelihood function', which can be used to discriminate between groove cross-sections that do not resemble the groove model, $M(x)$. Values from the groove shape likelihood function, ε_M which exceeded 3σ were found not to resemble the groove model, $M(x)$ sufficiently and are flagged as outliers in the displacement signal z_a(n).

Despite the groove shape similarity test, impulsive noise may still result in the case some cases. For example, debris or damage present in the groove can unwittingly resemble the model $M(x)$, despite being locally much higher or lower than surrounding groove cross sections. Again, it is possible to make use of the average groove height parameter, H_z as a means of identifying such outliers. Consecutive depth estimates $z_a(n+1)$, and $z_a(n)$ whose absolute difference exceeds $0.5H_z$ can be flagged as outliers. This stage reduces the requirement for impulsive noise removal in the audio signal domain, by making use of the groove geometry in the spatial domain. Figure 4.32 shows the effect of these additional outlier removal stages.

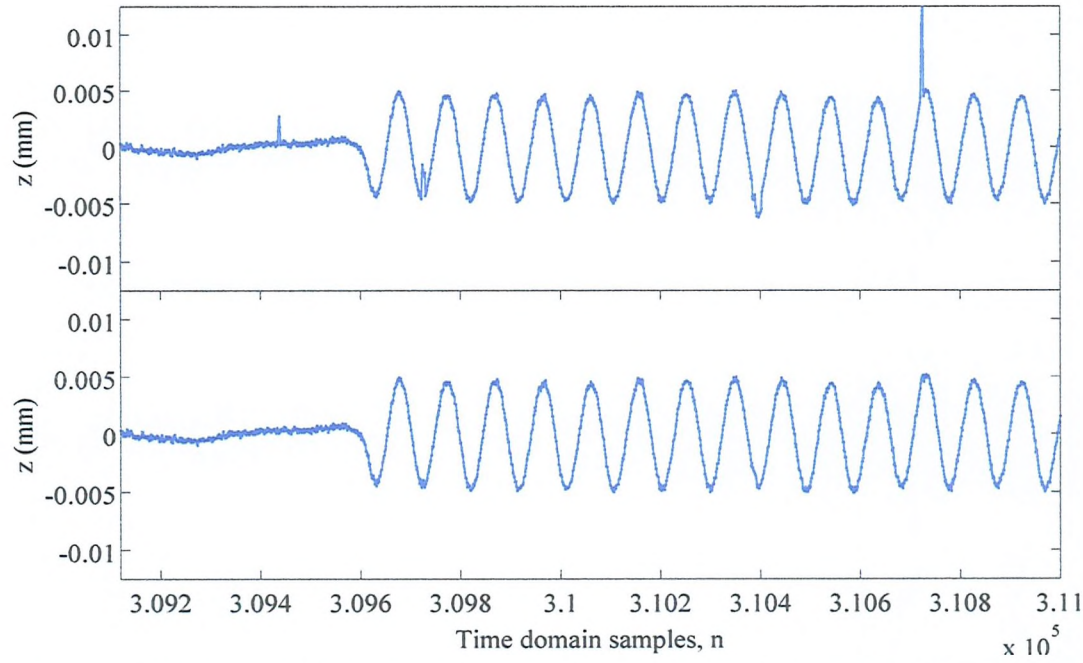

Figure 4.32: Impulsive noise removal by examining the groove shape likelihood function and observation of consecutive samples whose absolute difference exceeds $0.5H_z$

4.4.8 Equalisation and Signal Correction

In this section, the raw, unfiltered depth estimate $z_a(n)$ is adjusted to overcome issues of z-offset error (discussed in Chapter 3, Section 3.4), and is output to an audio signal file (.wav) that is suitable for playback.

Equalisation

For recordings that exhibit constant velocity characteristics (over a limited frequency range), such as cylinder recordings, the maximum amplitude of the recorded groove displacement decreases as the frequency of the recorded signal increases. Furthermore, the frequency response for acoustically recorded media is limited to within a bandwidth of ~200Hz–5 kHz, due to the characteristics of the horn and mechanical recording process. In order to produce a signal which approximates a flat frequency response over this given frequency range, a filter should be implemented which incorporates both the constant velocity constraint (differentiating filter) and the limited bandwidth in which the sound is audible (band pass filter).

Different numerical differentiation methods (e.g. forward/backward/central difference), will introduce different frequency responses to the resultant signal. Due to the complex transfer function of the horn, mechanical linkage, and specimen's surface condition, the required equalisation is likely to vary from cylinder to cylinder. It is therefore suggested that any final equalisation, (prior to public access) is left to the expertise of the transfer engineer or sound archivist.

Regardless of which filtering method is used, the most important stage in the signal reproduction chain following surface measurement, is to obtain the most accurate estimate of the displacement track, $z_a(n)$, with minimum distortion. The ideal equalisation required for purposes of providing public access to rare sound recordings remains a subjective task should be left in the hands of the sound archivist.

Correcting for z-sensor drift

Due to the measurement process and sensor drift, z-offsets generally occur between segment joins. In Chapter 3, Figure 3.17, an example of this offset was shown between linescans at the wraparound point between $\theta = 0°$ and $\theta = 360°$. The effect of this z-offset for the final audio signal will be a discontinuity, at regular intervals of nT samples, where n is the number of linescans between joined segments, or the number of linescans for a complete cylinder rotation.

This offset is accounted for by adjusting the displacement signal $z_a(n)$ every nT samples, by calculating the z-offsets and adjusting the displacement track accordingly, such that the error between samples $z_a(nT+1)$ and $z_a(nT)$ is zero. For surfaces measured in multiple segments, this z-correction is carried out between each segment join, using knowledge of the segmentation scheme described in Section 4.4.2. Figure 4.33 shows the uncorrected and corrected displacement track, from the surface of a Blue Amberol cylinder. The effect of this stage is a reduction in repetitive impulsive noise.

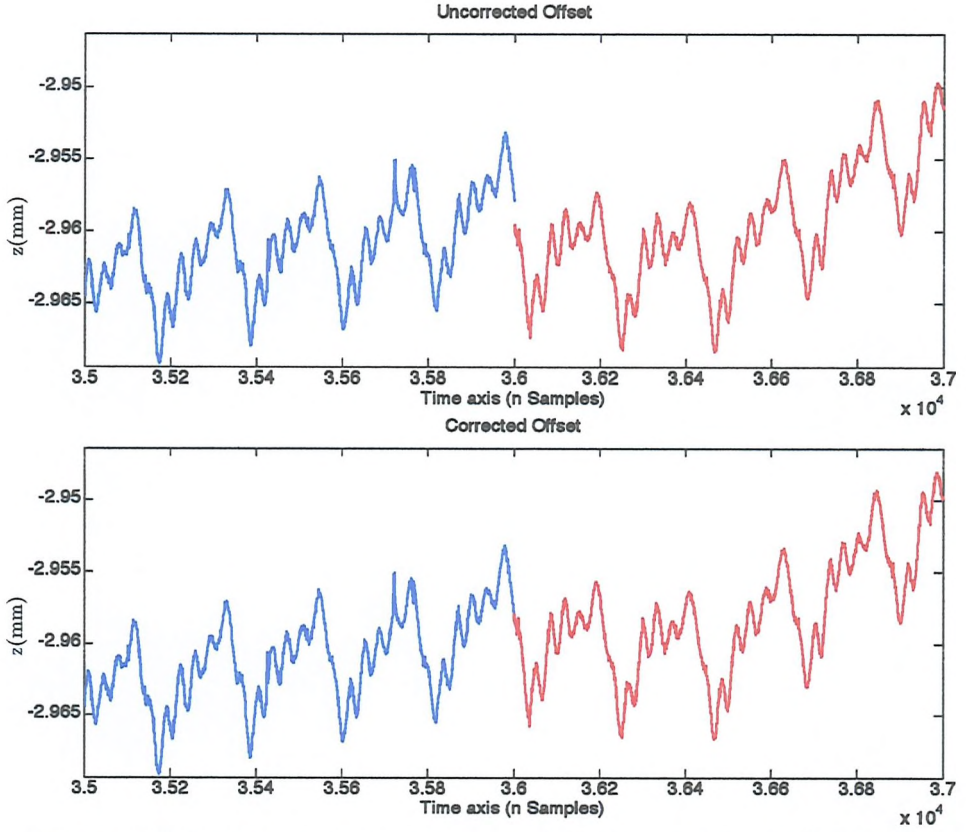

Figure 4.33: Uncorrected (top) and corrected displacement signals (bottom), from Blue Amberol cylinder (JBTBM).

Low Frequency Noise Removal

The measured surface data typically has a significant low frequency component, due to the eccentricity of the surface. This results in low frequency noise, which is perceived as a 'thump', or 'rumble'. This noise was not produced by the original incident acoustic pressure, and should thus be removed. For non-contact sound reproduction, low frequency noise is produced by the following phenomena:

1. **Eccentric surface form, due to radial deformation of the cylinder.** Typically, this noise is sub-audible and occurs at the frequency of rotation, which is 2.67 Hz for 160 rpm cylinder.
2. **Occurrence of segment joins.** For a surface scanned in segments there often exists differences in surface orientation between adjacent scans. This may be due to the adjustment in the sensor offset distance, causing small changes in sensor orthogonality. For a 160rpm cylinder scanned in 20 segments, a low frequency noise peak in the spectrum is observed at 53.3 Hz (2.67 Hz x 20).

Low frequency noise can be removed by applying a high pass filter, with a cut off frequency of around 150 - 200 Hz. The frequency response of such a filter, implemented as a 4^{th} order digital Butterworth filter is shown in Figure 4.34.

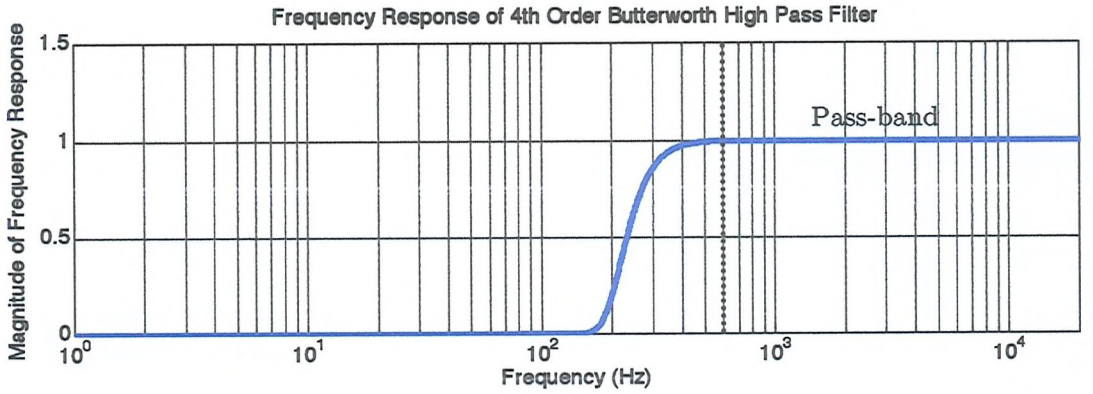

Figure 4.34: High pass filter frequency response (implemented via digital Butterworth filter, 4th order, with cut-off frequency of 150 Hz.

Un-filtered and filtered displacement signals (with the filter design given in figure 4.34) are shown in figure 4.35.

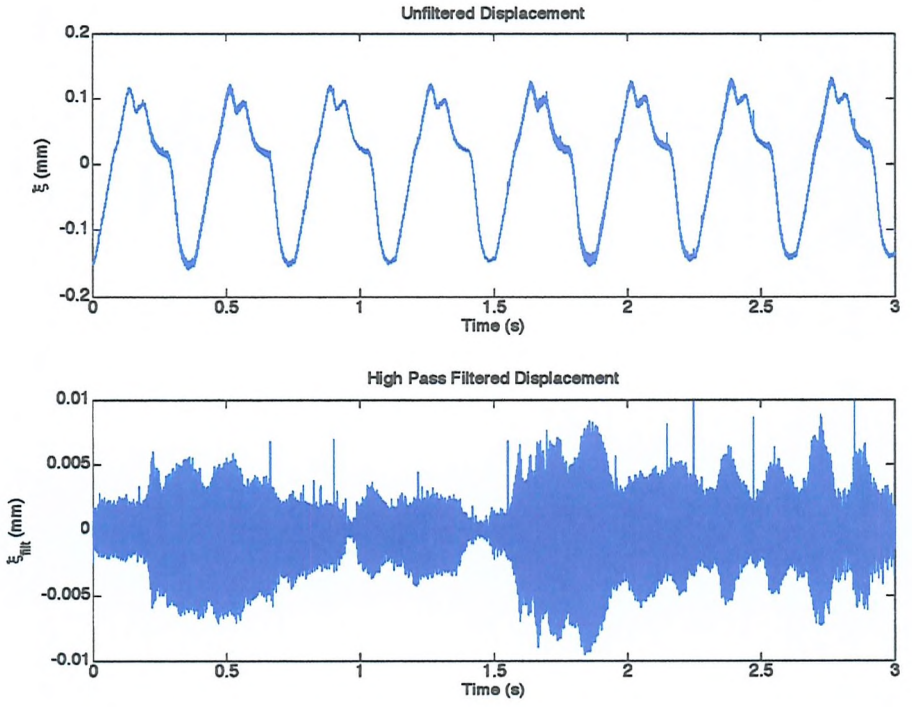

Figure 4.35: Un-filtered (top) and high-pass filtered (bottom) displacement signal, from Blue Amberol cylinder (JBTBM).

4.5 Methods developed for Disc Media

This section briefly outlines some preliminary investigations into signal reproduction of discs measured via the NCFSM method.

Reproduction of laterally encoded, monaural discs relies on the ability to resolve modulations along the x-axis (linescan axis), as opposed to modulations in the z-plane for cylinders. Despite this distinction, the same methods of groove seed detection and groove shift estimation for cylinders can be applied to disc surfaces.

As was shown in Figure 3.11-3.13 in Chapter 3, the air-bearing system measures the land (groove tops) and bottom of the groove only. From preliminary scans, it was also found that the groove tops (land), imaged better the groove bottom. Although not ideal

for surface preservation, enough information is available for signal recovery, albeit with poor audio fidelity.

The sensor's angular tolerance (and inability to resolve 45° surface inclines) is used to form an edge detection scheme. Due to a lack of information at the groove walls, depth information from the linescan is disregarded and instead, the surface is treated as a binary surface map, with 'ones' at the position of measured data, and 'zeros' at the position of unmeasured data. From this a simple edge detection scheme is used, based on the finding of edge pairs. The criteria for choosing these edge pairs is based a distance corresponding to the groove width. The groove width for monaural disc recordings is approximately constant throughout the recording. This method is now described:

1. The surface data is transformed into a binary map, with 'zeros' at the position of measured data and 'ones' at the position of unmeasured data (see Fig. 4.36)
2. The groove seed is determined by searching for minima along the interpolated linescan at 0°.
3. A groove phase shift estimate τ_{x_s} is determined for the surface (see Fig. 4.37), as per the method described for cylinders from the binary map. A skeleton stylus trajectory is formed as per the method described for cylinders in Eq. (6).
4. Edge points along the binary linescan are simply detected by observing when the value changes from '0' to '1', and '1' to '0'.
5. The left and right groove wall signals x_L and x_R are found by examining edge pairs separated by an optimum distance, specified by the groove width (shown in Figures 4.38-4.39).
6. Steps 4-5 are repeated for all grooves found along the trajectory formed in step (3).

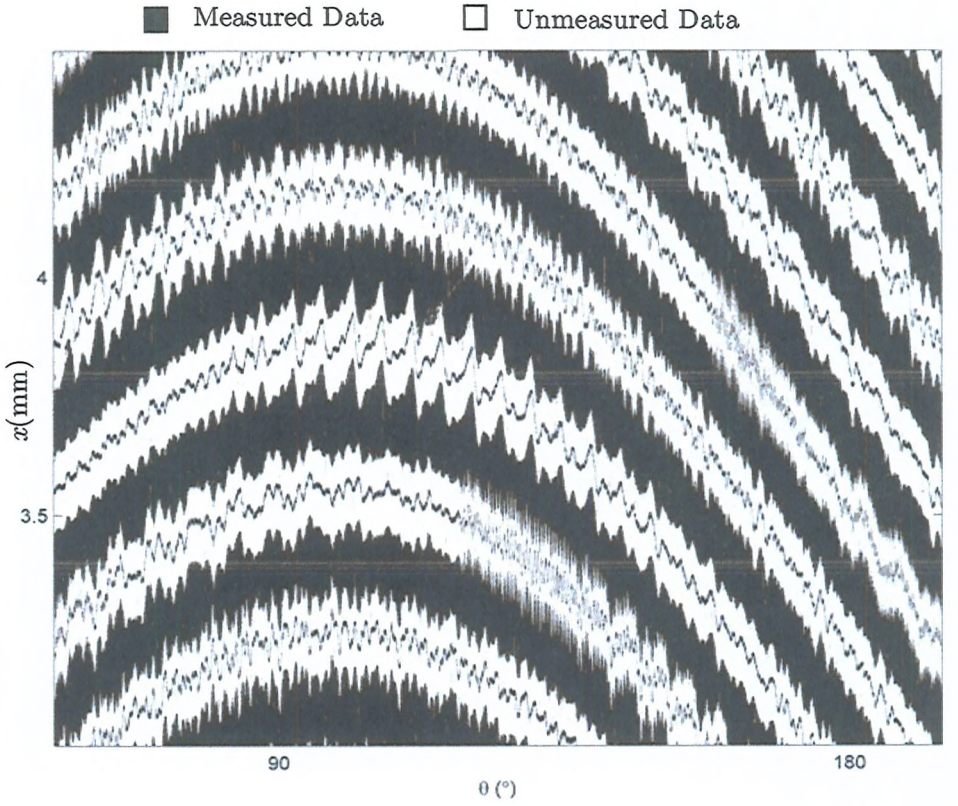

Figure 4.36: Binary map of 78 rpm disc surface.

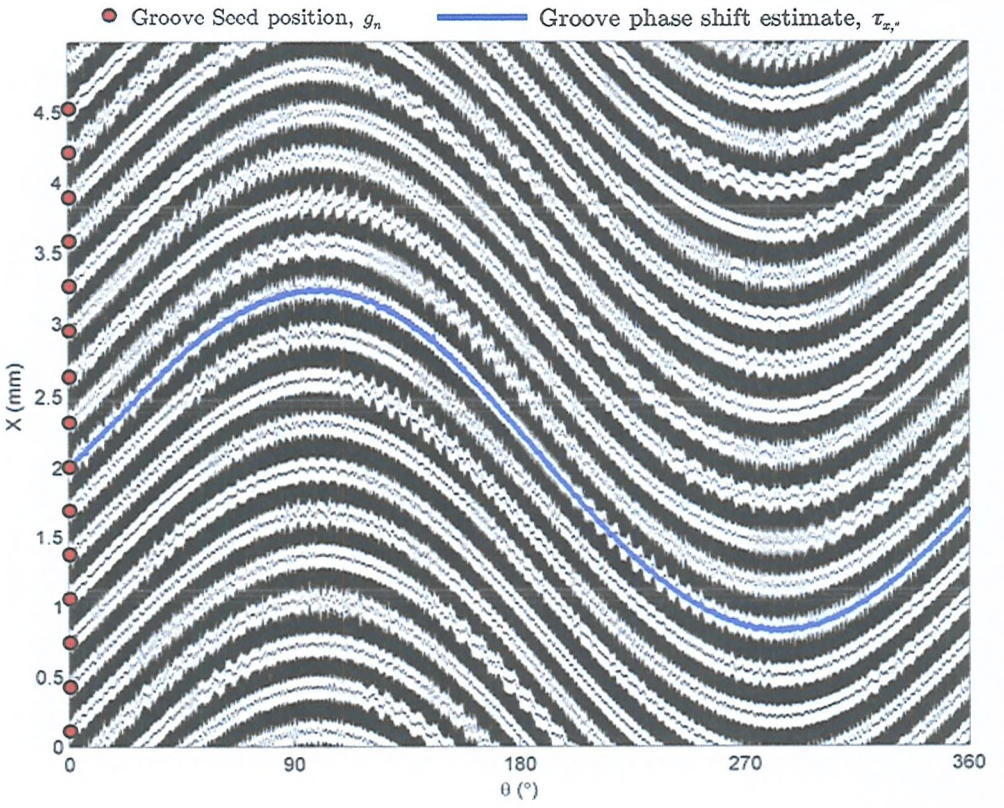

Figure 4.37: Binary map of 78 rpm disc record with groove seed positions and groove phase shift estimate overlaid.

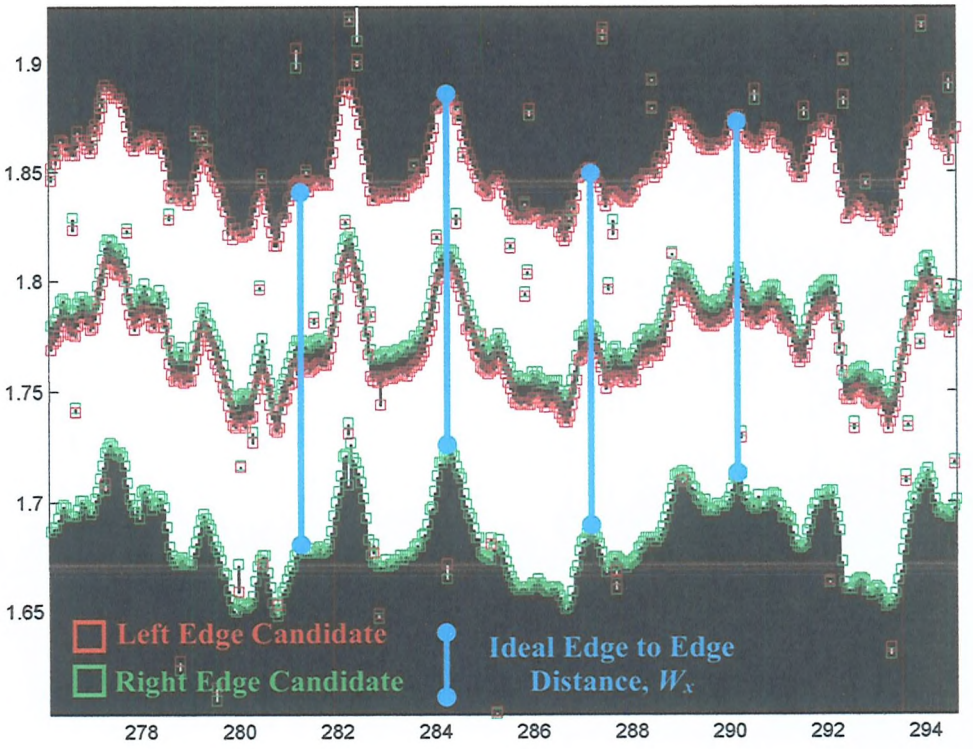

Figure 4.38: Close up of single 78 groove showing left and right edge candidate detection and the ideal groove width distance W_x used to select correct edge candidate pairs.

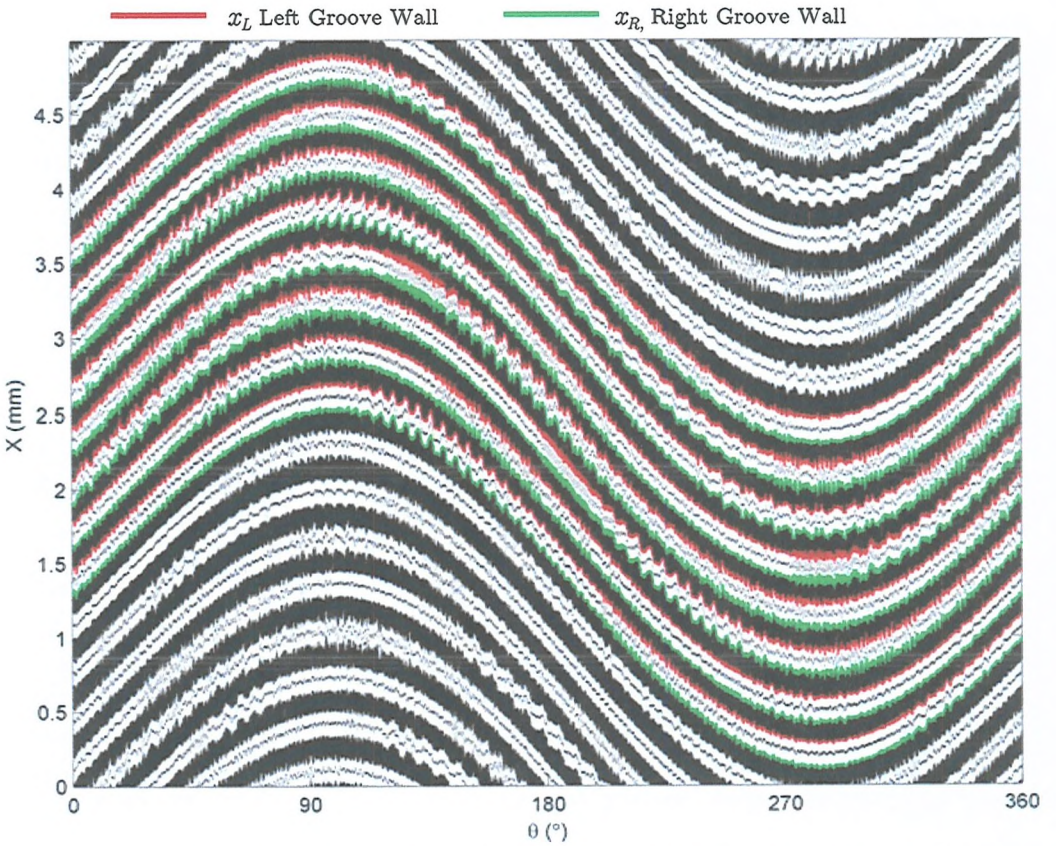

Figure 4.39: Binary map of 78 rpm disc record with left and right groove wall traces overlaid.

4.6 Chapter Summary

In this chapter, an 8-stage methodology has been described for audio signal reproduction from NCFSM measurements of cylinder recordings. Two stylus trajectory estimation methods have been designed for cylinders in good and poor surface condition. Both methods rely on the determination of groove pitch, and some *a priori* knowledge of the groove structure. The advantage of estimating the stylus trajectory via the phase shift estimation method is its ability to account for linescan x-offsets encountered with troublesome datasets, for example, the TC1 data shown in Figure 4.16. The local tracking method is slower than the groove shift estimation method, but is necessary for tracking damaged groove structures, or multi-track recordings for instance. In certain circumstances, manual tracking of the grooves is required, however this stage is simplified by transforming the 3D surface map into a binary map of feature points (minima map). The choice of tracking method depends on the condition of the surface, whether or not x-offset errors exist, and the type of recording, hence it is encouraged that the surface data is visually inspected, prior to selecting the groove tracking method.

In Section 4.4.7, four different groove depth estimation schemes have been considered, with a view to obtaining the optimum groove depth estimate. To verify which scheme produces the best signal reproduction, a quantitative signal analysis shall be carried out in the next chapter. The various depth estimation methods schemes can be rated in terms of signal-to-noise ratio (SNR), and total harmonic distortion (THD). This shall be accomplished using a cylinder recording (TC1), encoded with test signals.

The methods developed from preliminary measurements of a 78 rpm disc should be seen as a proof of concept and not as a final method for signal reproduction from the air-bearing surface data. Improvements to the scanning methods should provide a signal with higher fidelity. It has been shown that similar groove seed and groove shift estimation, and local tracking methods developed for cylinder signal reproduction, can also be applied to disc surfaces.

Chapter 5

Results

5.1 Chapter Overview

This chapter details the results and analysis of signal reproduction methods developed in Chapter 4 from the NCFSM method. The majority of this chapter focuses on quantitatively assessing the depth estimation schemes described in Chapter 4, Section 4.4.7. Results in this chapter are described for four cases:

i) Test-tone cylinder (custom made cylinder for quantitative analysis)

ii) Cylinders in good condition - commercial cylinder recording (typically transferred by stylus methods)

iii) Cylinders in poor condition (a broken cylinder which was reconstructed)

iv) Flat media - tinfoil recording, measured via the air-bearing system.

The test-tone cylinder in case (i) provides a means of quantitatively assessing signal reproduction methods, using signals of known type. From these results, conclusions can be drawn as to the optimal method of signal reproduction and comparisons can be made with stylus replay.

Cases (ii) and (iii) represent a qualititative analysis of signal reproduction. These testing procedures allow for a direct comparison of the NCFSM method with stylus reproduction. Where possible, methods are sought to quantify signal quality numerically. Sound reproduction quality of musical signals is however a subjective matter dependent on the listener, and the reader is encouraged to explore signals described in this thesis here: [86].

5.2 Test Cylinder

Unlike calibrated 78 rpm disc recordings manufactured by the Audio Engineering Society [87], standardised cylinder test specimens are not commercially available for testing purposes. Meulengracht-Madsen produced acoustically recorded cylinders with original phonograph apparatus, encoded with pure-tones, for the purposes of evaluating his own playback apparatus [32], however the whereabouts and condition of these cylinders is unknown.

In order to quantitatively assess the audio signals reproduced via the NCFSM method, a custom-made test cylinder was produced at Poppy Records [88] (see Fig. 5.1). The production of a test cylinder has the following desirable properties:

- It provides a benchmarking tool for both the scanning and signal reproduction stages.
- It is possible to have control over the playback conditions, prior to scanning, (i.e. the grooves can be left untouched from stylus replay).
- A measure of signal to noise ratio (SNR) of the non-contact sound extraction method can be calculated numerically (as Stotzer et al. investigated for discs [54,57]), and can be used to conclude which the optimal depth estimation scheme, described in Chapter 4, Section 4.4.7.

In contrast to Meulengracht-Madsen's experiments, the recording stylus was not driven by acoustic pressure, but was directly cut by an electrical transducer. This bypasses the complex frequency response introduced by recording horns and allows for recording

above the 5 kHz limit if necessary. A wax cylinder blank was firstly 'shaved' using a skimming device, to produce an unmodulated surface on which to cut. WAVE files were generated, with each tone being of equal amplitude, and recorded onto CD (16-bit / 44.1 kHz). The CD output was fed into a recording amplifier with input at standard BBC line level (0 dBm into 600 Ω). The cutting tool head is then driven directly by the electrical current fed from the amplifier. Pure tones (sinusoids) in the range of 200 Hz – 5 kHz were recorded in three second tone bursts onto a cylinder rotating at 160rpm, with a pitch of 100 tpi. An image of this cylinder is shown in Figure 5.2 and a description of the tones encoded and cylinder description is given in Figures 5.3-5.4.

A second test cylinder, (TC2), was also recorded at Poppy Records, with a series of different waveforms (Sawtooth, Square, Triangular), and a tone sweep to test for signal reproducibility. Due to problems with the skimming apparatus and time constraints regarding measurement, the second cylinder was not measured satisfactorily to provide significant results. A comparison of the two measured test cylinder surfaces are shown in Figure 5.5-5.6.

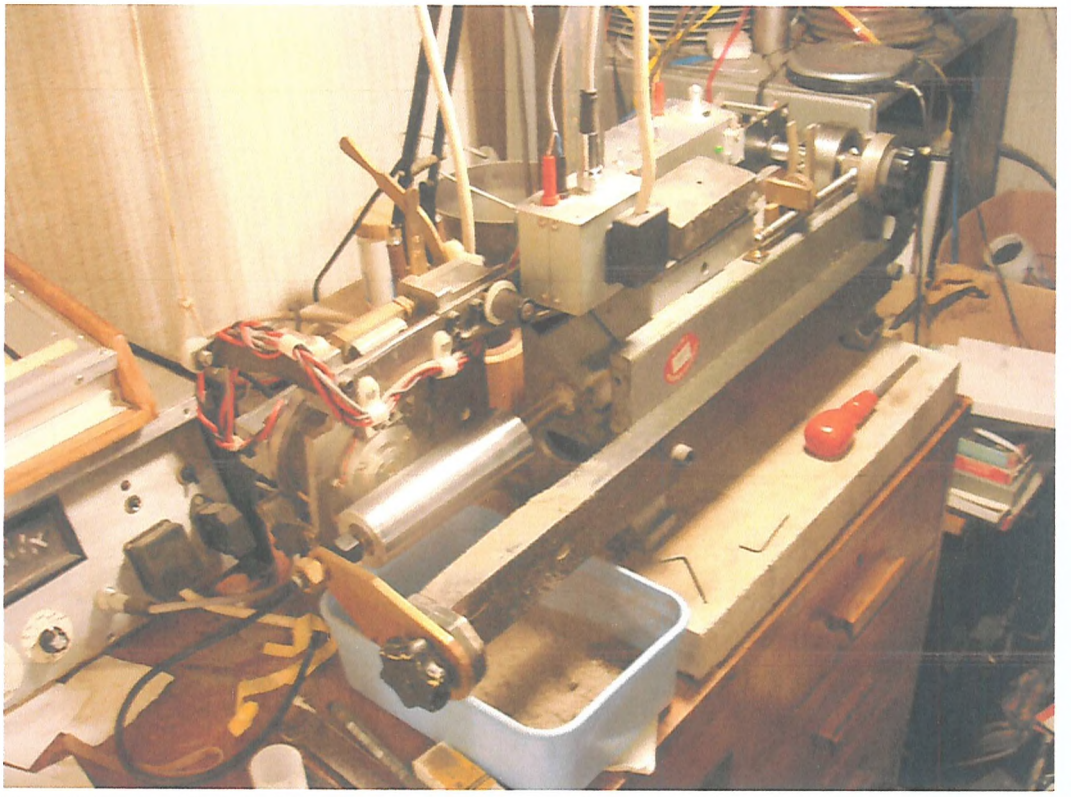

Figure 5.1: Electrical apparatus used at Poppy Records [88] for production of test cylinder recording. The cylinder is not shown mounted onto the mandrel here.

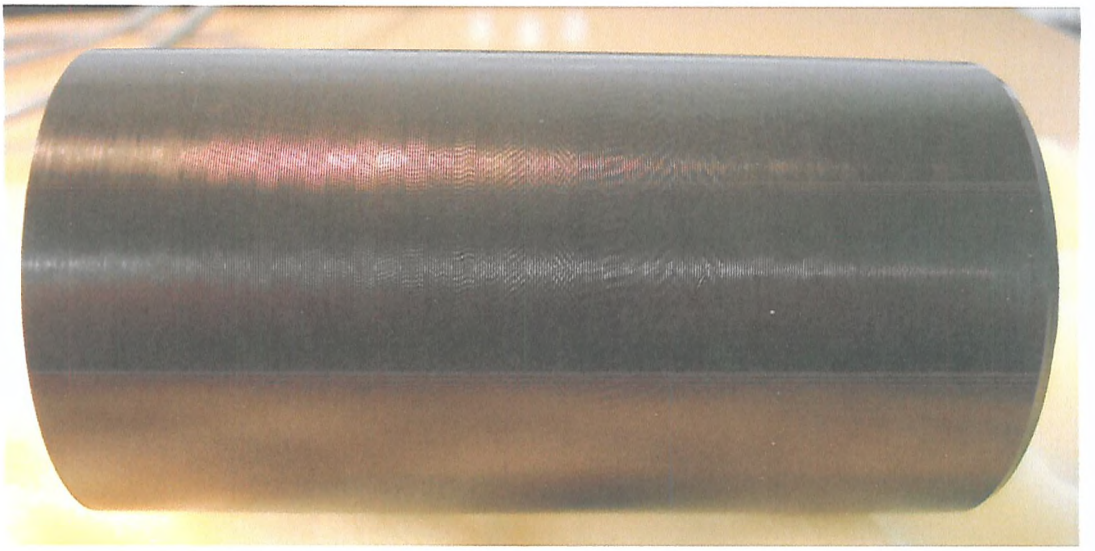

Figure 5.2: Image of test cylinder (TC1) encoded with sinusoids, produced at Poppy Records [88]. Bands of tones are visible. The cylinder was produced by an electrical cutting process, as opposed to acoustic coupling of a horn-diaphragm arrangement, as with the original Phonograph.

Frequency (Hz)	
Octave bands	1/3-Octave bands
250	**200** **250** **315**
500	**400** **500** **630**
1000	**800** **1000** **1250**
2000	**1600** **2000** **2500**
4000	**3150** **4000** **5000**

Figure 5.3: Details of recorded tones for TC1 cylinder. Each tone has a recording duration of 3 seconds.

Parameter	Value
Production Year	2006
Length	100 mm
Diameter	55.56 mm
Recording Speed	160 rpm
Recording Pitch, tpi (λ_x)	100 tpi (254 µm)
Recording Method	Electrical, direct cut
Recording Description	Test tones
Recording Material	Black wax cylinder
Measurement Sampling	$\Delta x = 10$ µm $\Delta\theta = 0.01°$ $\Delta z = 10$ nm
Scan details (measurement points)	**1 segment**: 98 mm x 36° (9801 x 3600) **18 segments:** 98 mm x 18° (18 x 9801 x 1800) **Total Measurement**: 98 mm x 360° (9801 x 36000)
Total Scan Time	~140 hours
Playback Sample rate	96 kHz

Figure 5.4: Description of test cylinder (TC1) recording and measurement details.

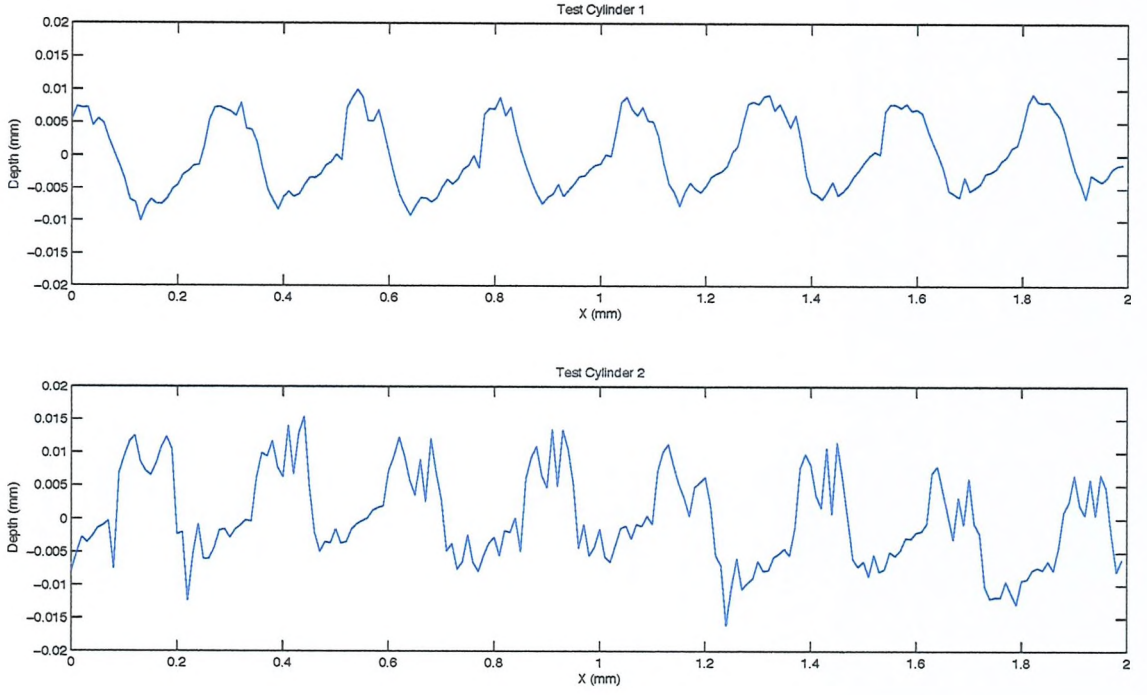

Figure 5.5: Comparison of linescan measurements taken from test cylinders produced at Poppy Records. The groove structure in test cylinder 2 is noisier and less defined due to issues with the cylinder shaving process.

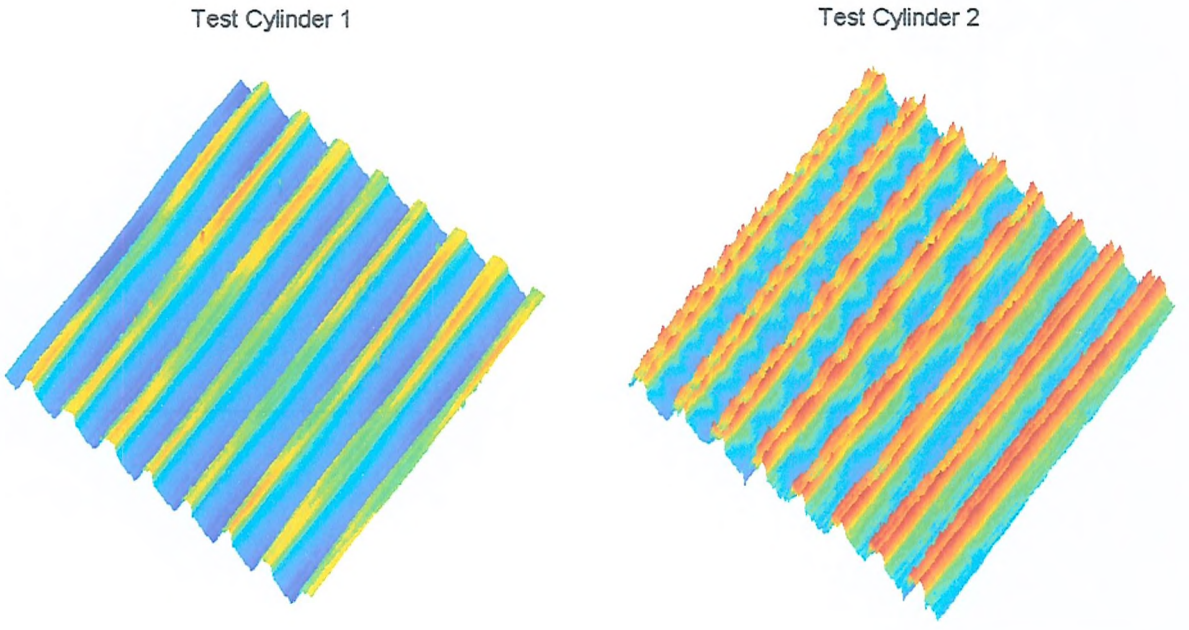

Figure 5.6: Comparison of surface measurements taken from test cylinders produced at Poppy Records. With test cylinder 2, a residual groove structure is visible towards the groove ridge (top), due to 're-shaving' of the cylinder during the recording process.

5.2.1 Testing Details

The sampling defined in Figure 5.4 gives approximately 25 data points per groove cross-section, and a playback sample rate of f_s = 96 kHz. The cylinder was measured in 19 surface patches. The recovered signals exhibited some frequency modulation and harmonic distortion, which is in part due to the condition of the recording apparatus shown in Figure 5.1.

Following data import, and outlier removal, the TC1 surface was segmented into 15 regions, with each region containing one of the fifteen, three-second tone bursts, as defined in Figure 5.4. A stylus trajectory A_n for each of the 15 TC1 tone surfaces was determined using the phase shift estimation method, described in Section 4.4.6. The phase shift estimation method was required in these circumstances because the data sets exhibited x-offset of error, especially between the segment joins. The four depth estimation schemes described in Chapter 4, Section 4.4.7 were derived from the surface:

Scheme 1: Single-point data streams, $z_{a\text{-}Xoff}$, found at consistent x-offset position across the groove cross-section, using trajectory A_n. The x-offset vector X_{off} is the set of integers: X_{off} = [-6, 6]. An x-offset of X_{off} = 0 is equivalent to the depth estimate found along the trajectory A_n.

Scheme 2: Averaging Samples across the groove cross-section $z_{a\text{-}N}$ - the mean of $\pm N$ samples around the central x-position, given by A_n. N is the set of integers: N = [0,12]. An average of $\pm N$ = 12, gives a 25 point average across the full groove cross-section, where λ_x = 25 samples (~250μm). The case where N=0 is equivalent to the depth estimate found along the trajectory A_n.

Scheme 3: SG Polynomial Filtering method, $z_{a\text{-}P,L}$ - denotes the depth signal estimate given by Savitzky-Golay (SG) filtering method described in Section 4.4.7. The depth estimate $z_{a\text{-}P,L}$ is derived from the SG filtered surface of order P={1,3,5,7,9} and filter length L={11,13,...25}.

Scheme 4: Model fitting method $z_{a\text{-}M}$: depth estimate produced via the model fitting method which incorporates outlier rejection described in Section 4.4.7. The groove model $M(x)$ is derived for each of the 15 tone surfaces.

5.2.2 Signal Quality Metrics

In previous work with test disc recordings Stotzer [57] used two signal quality metrics to quantify signal reproduction from the VisualAudio method. With the test cylinder (TC1), a similar analysis can be carried out to assess the quality of sinusoidal signal reproduction and critically assess the four depth estimation schemes described above. The metrics used here to assess for each depth signal estimate are the signal to noise ratio (SNR) and total harmonic distortion (THD).

Signal to Noise Ratio

The signal to noise ratio (SNR) is computed by locating the peak frequency and then calculating the total noise power in the remaining frequency bins of the DFT, (considered to be noise). This is expressed in decibels and is given by:

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

Definitions of the signal and noise powers, P_{sig} and P_n , are not immediately obvious for pseudo-sinusoidal signals which are frequency modulated and contain harmonic distortion. Care must therefore be taken when deciding what constitutes the signal and noise. For example, it is questionable whether energy from harmonics in the spectrum should be considered as part of the signal or noise band. The choice of frequency band in which to search for P_{sig} is complicated by the fact that the recovered signals exhibit frequency modulation due to the recording process. Other factors affecting the calculation include window length and frequency resolution of the DFT.

For each groove depth estimate defined in Section 5.2.1, the following procedure was carried for each tone:

- A region of 2.5 seconds of clear tone burst was identified from the depth estimate.
- The short-time Fourier transform (STFT) is used to carry out spectral analysis on windowed segments of the each tone. Each frequency window has of duration 0.125 seconds, with a frequency bin resolution of 2 Hz.
- The signal power P_{sig} is found by locating the peak in the power spectrum (fundamental frequency) around the n-th tone of interest, f_n.
- Frequencies which contribute to the noise power P_n band were considered in two ranges:
    i. Below f_n : 150 Hz $\rightarrow$ $0.95f_n$
    ii. Above f_n : $(1.05f_n)$ $\rightarrow$ 20 kHz.
- Harmonic frequencies (integer multiples of f_n) above the fundamental were considered as undesirable and were included in the noise power band. Frequencies below 150Hz were not included in the noise band calculation, as these were attributed to the surface form and not the audio signal.
- SNR and THD are calculated for each of the windowed segment, and the mean of these metrics is taken as the SNR and THD scores.

Total Harmonic Distortion analysis

Total Harmonic Distortion (THD) gives a measure of the geometric distortion of the sinusoid. The lower THD, the more the signal resembles a true sinusoid. It is defined as the ratio of the sum of the powers of all harmonics components above the fundamental frequency, to the power of the fundamental. Expressed as a percentage, the THD is given by:

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

Where F_n is the Fourier component of the n-th harmonic, $n = 1$ denotes the fundamental frequency.

In the case of a 1 kHz tone, the first four harmonics - 2 kHz, 3 kHz, 4 kHz and 5 kHz - are included in this calculation, as harmonics above 5 kHz become lost in background noise.

5.2.3 SNR and THD Results

SNR and THD results for the four depth estimation schemes (as described in 5.2.1), are shown in Figures 5.7 - 5.8 (tabulated results are given in Appendix D). It can be seen in Figure 5.7 that for all tones, the model-fitting method with outlier rejection (Scheme 4, $z_{a\text{-}M}$) provides the highest SNR ratio score for the TC1 cylinder. The SG filtering method (Scheme 3, $z_{a\text{-}P,L}$) provided better SNR scores than mean averaging (Scheme 2, $z_{a\text{-}N\mu}$) and Scheme 1 (averaging across the groove cross-section, $z_{a\text{-}Xoff}$). Of all the schemes, Scheme 1 produced the lowest SNR score across the full frequency range.

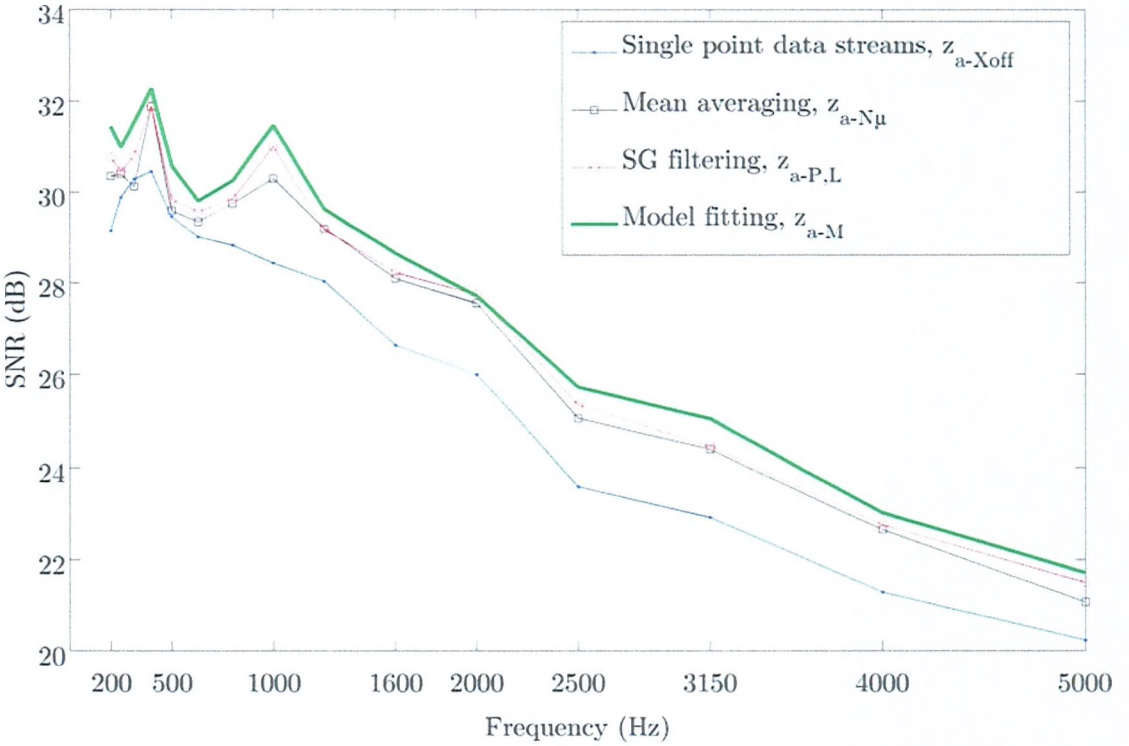

Figure 5.7: Results for best average SNR scores for the four different depth signal estimation schemes, (higher SNR is better).

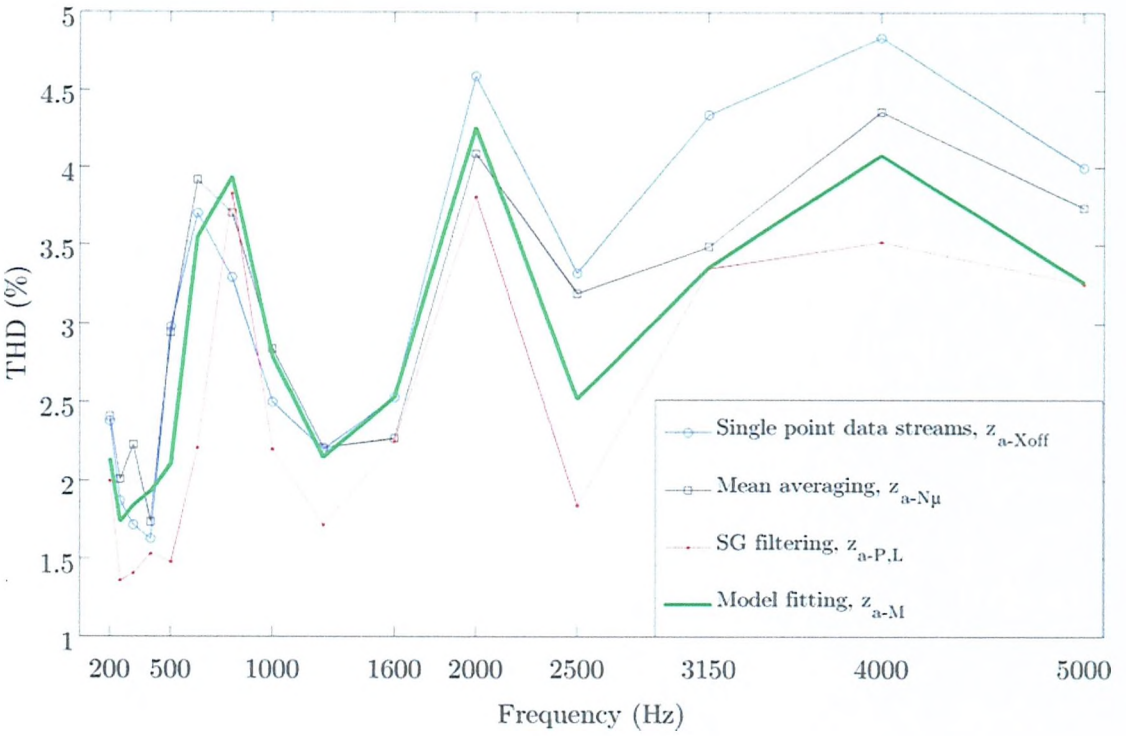

Figure 5.8: Results for best average THD scores for the four different depth signal estimation schemes, (lower THD is better).

The SG-filtered depth estimate (Scheme 3) gave consistently low THD scores (bar 800 Hz), however the combination of best filter parameters (P,L) varied from tone to tone. This demonstrates the ability of this method to obtain different variable signal quality by experimenting with different combinations of P and L which may be desirable to sound archivists.

Figure 5.9 shows the variation in SNR of the single point data streams (Scheme 1, $z_{a\text{-}Xoff}$) as a function of position across the groove, for the 1 kHz tone surface. It can be seen that for this sample, the SNR scores was greater towards the groove bottom, and lower at positions where the gradient of the groove shape is steeper. This is consistent with the results for a Blue Amberol cylinder, previously described in Section 4.4.7 and [84]. The variation in z from data streams taken at the groove wall is greater than at the groove bottom, and this increased variance leads to a greater noise level (and lower SNR).

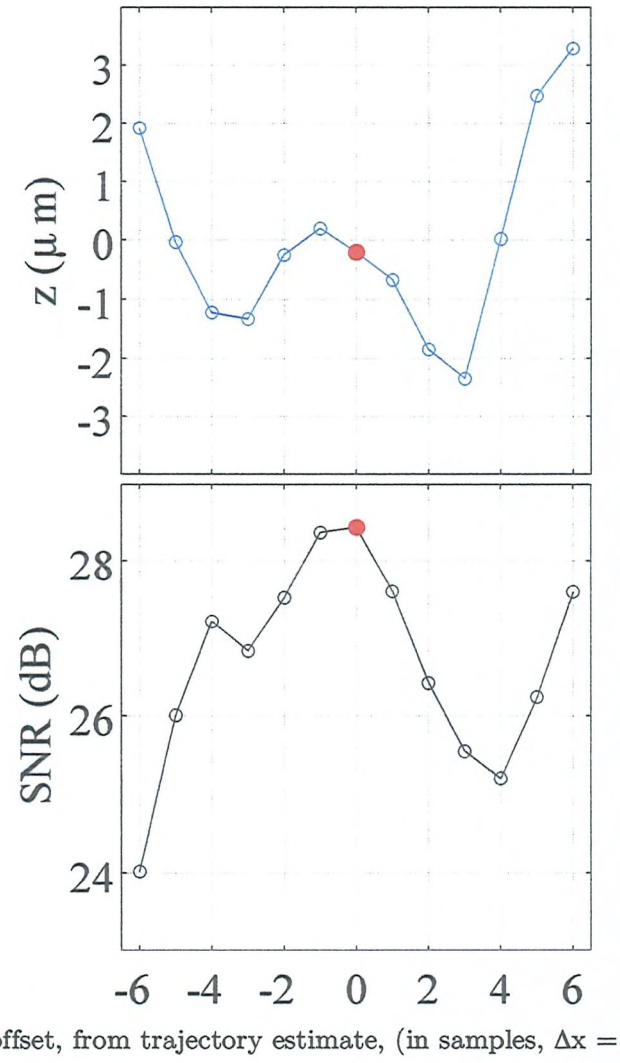

Figure 5.9: SNR vs. x-position across the groove cross-section for Scheme 1, (single-point data streams) for 1 kHz tone surface of TC1 cylinder. The red dot, represents the x-offset at which the SNR was maximum. For this example the maximum is obtained from the centre of the groove at x-offset = 0.

Figures 5.10 and 5.11 show SNR and THD scores as a function of the parameters N (for Scheme 2) and P-L (for Scheme 3), for the 1 kHz tone surface. Figure 5.10 shows how the SNR increases as the number of points in the average window, N increases, but care must be taken when interpreting this trend. The best SNR for Scheme 2 comes from a 21pt mean average, however this signal did not have the lowest THD, and upon listening to the signal, an echo effect is noticeable. This kind of temporal distortion is not accounted for in the spectrum averaging technique and is discussed later. The signal which had lowest THD distortion (and hence the most sinusoidal) was from a 15pt mean average. There is a rapid increase in THD above $N = 19$, and this is due to

the fact that as $N \rightarrow \lambda_x = 25$, more samples towards adjacent grooves are included in the average, which resulted in increased temporal (and hence, harmonic) distortion.

Mean averaging is a low pass filtering operation, since its effect is to allow lower spatial frequencies to be retained, whilst suppressing higher frequency components. A larger mean average will remove more noise (high frequency), but can reduce the level of detail in the groove cross-section. The result of this for the recovered sinusoid signals is a reduction in peak signal power and increased harmonic distortion when N is large (greater than 19 samples). In addition, by listening to signals where N is large, temporal distortion is also present in the form of an echo. At 1 kHz, a 15pt mean produced a signal with lowest THD, reasonably high SNR score, no noticeable temporal distortion, and was therefore considered to be the best signal produced by Scheme 2.

Figure 5.11 shows how the SNR score for Scheme 3 decreases as the filter order P is increased. This is due to the fact that higher order polynomial filters retain the higher spatial frequency content (roughness) of the groove cross-section, which results in increased high frequency noise for the audio signal. The highest SNR came from filter [P=1, L=15]. An SG filter of order $P = 1$ is essentially an un-weighted, moving average filter, similar to the mean averaging in Scheme 2. This similarity can be observed in Figure 5.10 and 5.11 by noticing that THD curves for Scheme 2 and Scheme 3, ($P = 1$) have similar trends. An echo effect was also noticeable for [$P = 1$, L= 19-25], as was found with Scheme 2 for large N.

Figure 5.12 shows a plot of measured frequency versus measured sinusoid amplitudes. It can be seen that at 4 kHz, the displacement amplitude of the recovered tone is 1μm (or 2 μm, from peak to peak). Comparisons of recovered signals at 400 Hz and 4 kHz are shown in Figure 5.13. At these two groove amplitudes, the sensor can resolve the sinusoid over approximately 1200 discrete depth values for 400 Hz, and 200 discrete values for 4 kHz. This matter of quantisation, coupled with the fact that smaller amplitude signals become masked by surface noise is the reason why the SNR generally decreased with frequency for all depth estimation methods.

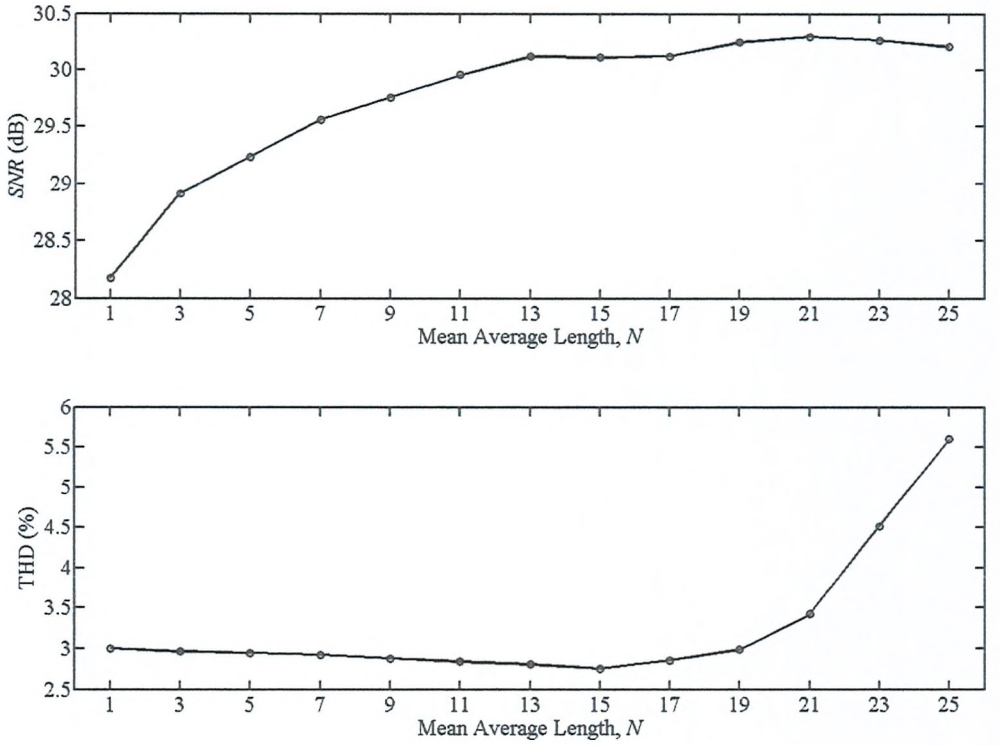

Figure 5.10: SNR & THD scores for Scheme 2 (averaging across the groove matrix), at 1kHz.

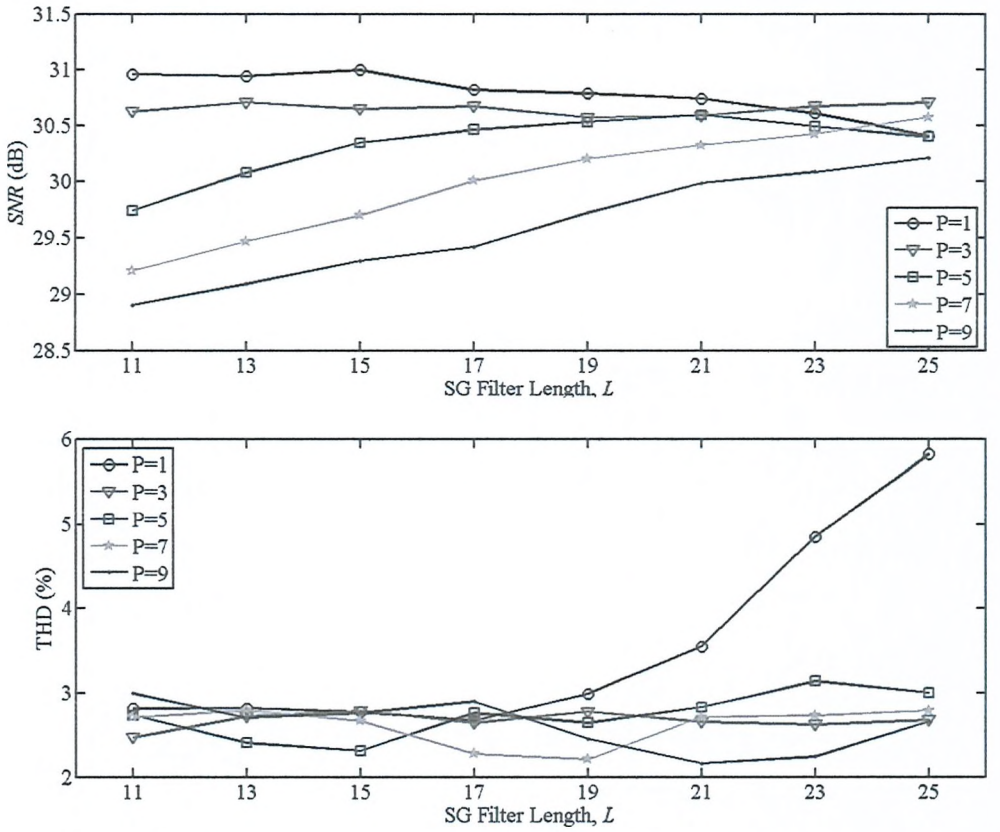

Figure 5.11: SNR & THD scores for Scheme 3 (averaging across the groove matrix), at 1kHz.

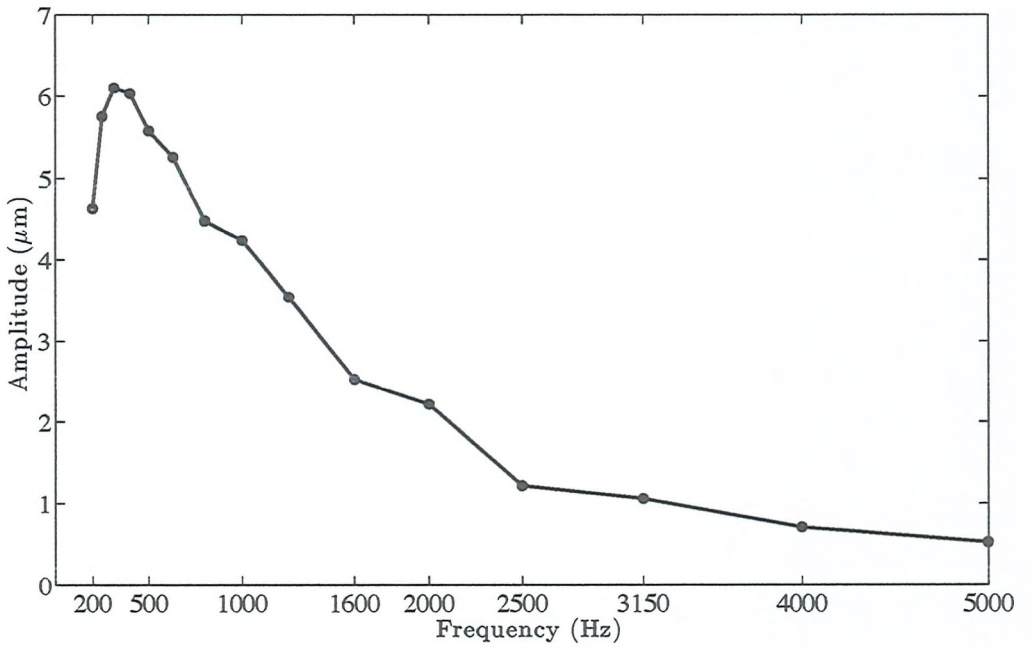

Figure 5.12: Frequency vs. Amplitude for recovered sinusoidal signals on test cylinder (TC1) via the NCFSM method.

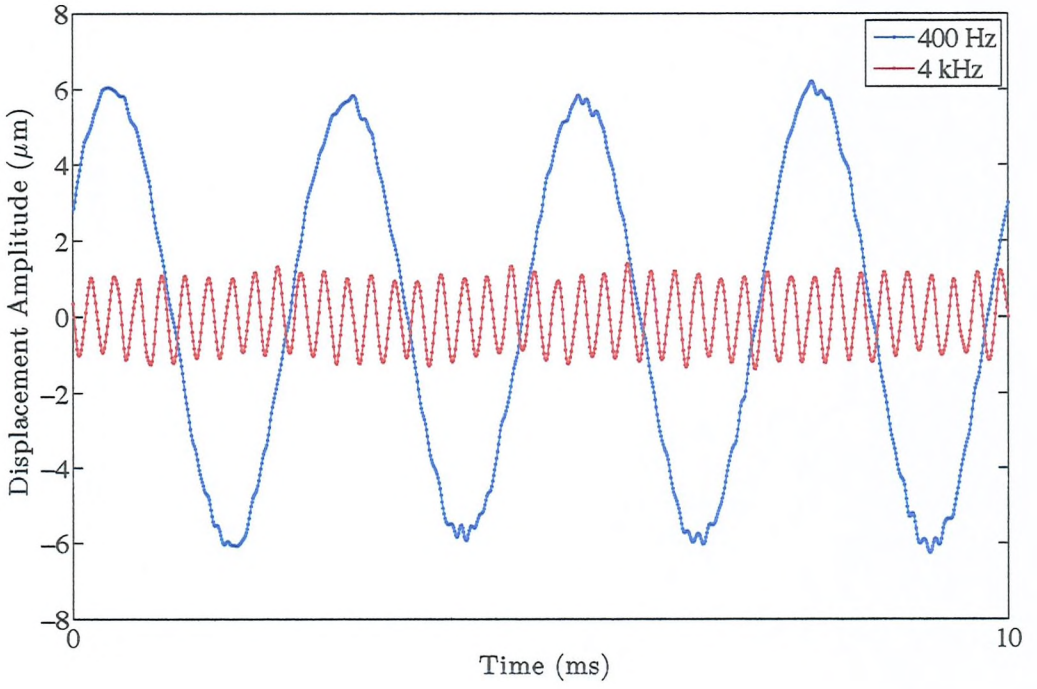

Figure 5.13: Time histories of tones recovered from the test cylinder (TC1) showing the difference in amplitudes of the 400 Hz and 4 kHz tones. The range of displacement for the 4 kHz tone (peak-to-peak) is approximately 2 μm, compared to 12 μm for the 400 Hz tone.

5.2.4 Comparison of Optical and Stylus methods

Once the test cylinder (TC1) was fully scanned, a number of stylus transfers were produced by the British Library Sound Archive [4], using a stylus, fitted with a stereo cartridge. In total, eight stylus transfers were made. The cylinder had not been played back by a stylus prior to this reproduction. All transfers were captured at 96kHz, 24-bit resolution WAVE files. The cylinder player outputted a stereo channel signal, which was summed to mono to cancel any reverse phase noise (as described by [4] in the accompanying transfer log).

No filtering or equalisation was applied to the transfer. The transfer engineer suggested that the audible 'wow' is in part due to the movement of the cylinder and off-centre mounting on the mandrel. It was noted that on initially mounting the cylinder on the mandrel, one end of the cylinder appeared elliptical, thus this end did not fit well around the mandrel, and was free to move to a small degree.

Due to differences in equalisation and considerable 'wow' of the stylus transfer, *direct* SNR/THD comparisons between stylus and optical signals are not presented in this thesis. Subjective listening suggests that stylus reproduction has a higher SNR, in terms of a lower noise floor power and less impulsive noise, (the reader is encouraged to listen and compare tones here [86]). Despite the general increase of SNR with the stylus method, it appears that the NCFSM has two advantages over conventional replay: (i) increased accuracy of waveform reproduction (lower harmonic distortion) and (ii) less susceptibility to 'wow', or frequency modulation.

Harmonic Distortion

The *accuracy* of sine-wave reproduction in terms of harmonic distortion and frequency modulation appears to be greater with the optical method. Figure 5.14 gives a comparison of frequency spectra for recovered 2 kHz tones from stylus and optical methods. Stylus reproduction appears to show increased harmonic distortion compared with the optical signal. This is likely due to the stylus-groove interaction, known as 'tracing distortion' (discussed in Section 2.3.1), where the curvature of the reproduction

stylus is large compared with the curve of the modulated groove. As a purely non-contact method, the NCFSM method is not affected by such effects due to contacting the groove.

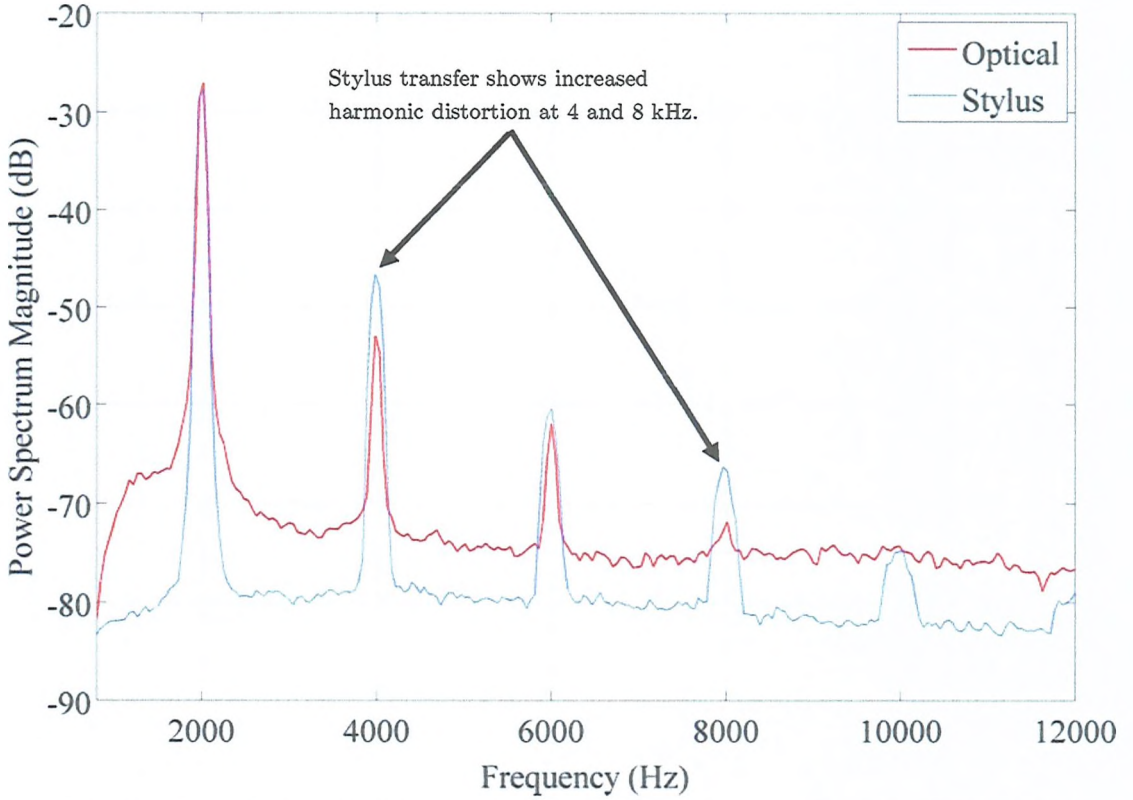

Figure 5.14: Comparison of frequency spectra (PSD of 2.5 seconds of sustained tone, NFFT=2048) for 2kHz tones recovered from test cylinder. The optical signal was high-pass filtered (f_c=500Hz) and normalised to give the same RMS as the stylus transfer. Although the stylus transfer has a greater SNR than the optical method, it is seen here to exhibit more harmonic distortion.

Frequency Modulation

The degree of frequency modulation was also compared by locating the position of the peak frequency in the DFT for windowed segments of recovered tones. The standard deviation in measured frequency from peaks in the DFT was used as an indication of frequency modulation. Figure 5.15 compares the measured frequency modulation for all tones reproduced from the test cylinder by optical and stylus methods. At 5 kHz, the frequency modulation with stylus reproduction is approximately ± 40 Hz, and for optical reproduction is less than ±10 Hz. The increased frequency modulation with stylus reproduction is largely due to the mounting of the cylinder on the mandrel. If

mounted incorrectly, the surface speed of the cylinder will vary, meaning that for a constant playback sample rate a modulation in frequency occurs

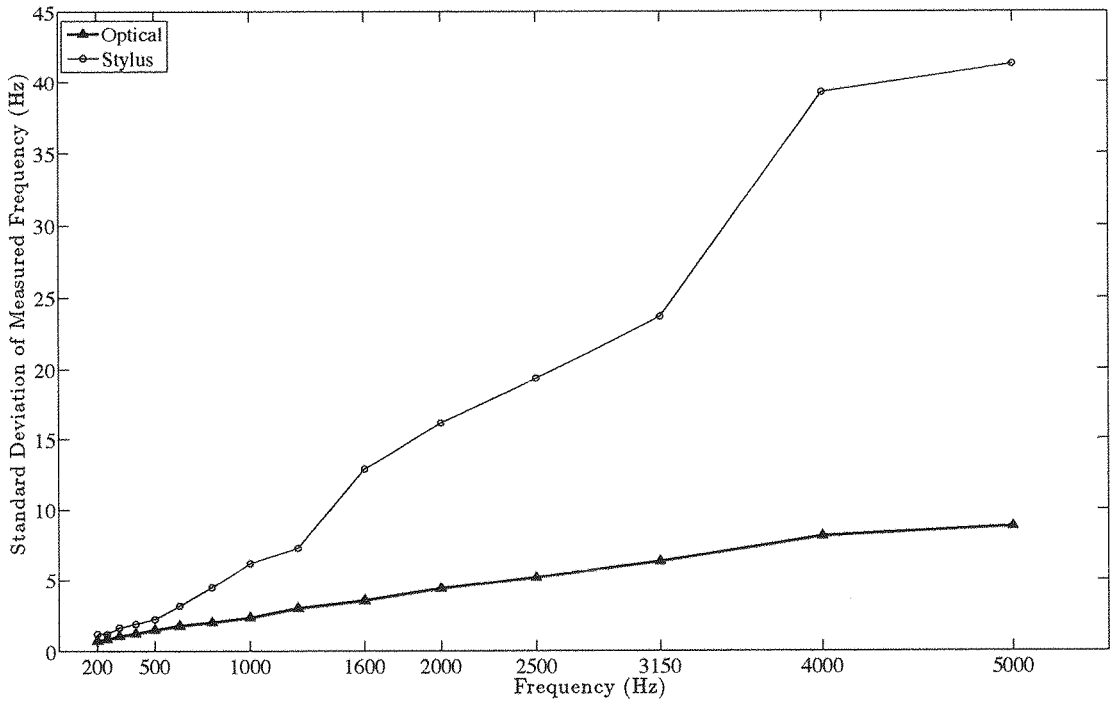

Figure 5.15: Comparison of measured frequency modulation between optical and stylus transfer of tones from the test cylinder (TC1). At 5kHz, the frequency modulation of the stylus signal is approaching ±42 Hz, compared with ±8 Hz for the optical transfer.

5.3 Results for Cylinders in Good Condition

This section describes results obtained from a commercial Blue Amberol cylinder (recorded on celluloid), recorded in 1913, with a musical recording called "The Preacher and the Bear" (PBEAR). This represents a cylinder in good surface condition, which would typically be played back via stylus, without risk of causing damage. The purpose of this analysis is to give a qualititative comparison of the NCFSM method with stylus replay, for cylinders in good condition. Details of the cylinder and scanning parameters are given in Figure 5.16.

Parameter	Value
Production Year	1913
Recording Method	Acoustic
Material	Celluloid
RPM	160
Recording Pitch, tpi (λ_x)	200 tpi (127 µm)
Duration	4 mins
Description	(1560: Edison Blue Amberol) "The Preacher and the Bear", Folk song.
Measurement Sampling	$\Delta x = 10$ µm $\Delta\theta = 0.1°$ $\Delta z = 10$ nm
Scan details (measurement points)	**4 segments:** 96 mm (x) by 90° (θ) (4 x 9601 x 900) **Total Measurement**: 98 mm x 360° (9801 x 3600)
Playback sampling rate	9.6 kHz

Figure 5.16: Details of cylinder used for testing well-conditioned cylinder (PBEAR)

The groove pitch λ_x was correctly identified using Eq. (2) from Chapter 4 as 127 µm. With a linescan grid spacing of 10 µm, the groove cross-section was made up of approximately 13 samples.

For this cylinder, the stylus trajectory was estimated via the groove shift estimate described in Chapter 4. This method was chosen because the cylinder was considered to be in good surface condition, and also x-offsets due to measurement error (especially at segment joins) were observed.

The groove shape model fitting method was used to form an estimate for $z_a(n)$ as it produced the best SNR for the TC1 dataset and the groove shape was observed to be relatively unchanged throughout the duration of the recording. A groove shape model $M(x)$ was formed and used to correct for linescan x-offsets. The groove shape model is shown in Figure 5.17 and the effect of reducing impulsive noise by identifying groove cross-sections that did not resemble the groove shape model (as discussed in Section 4.4.7) is shown in Figure 5.18.

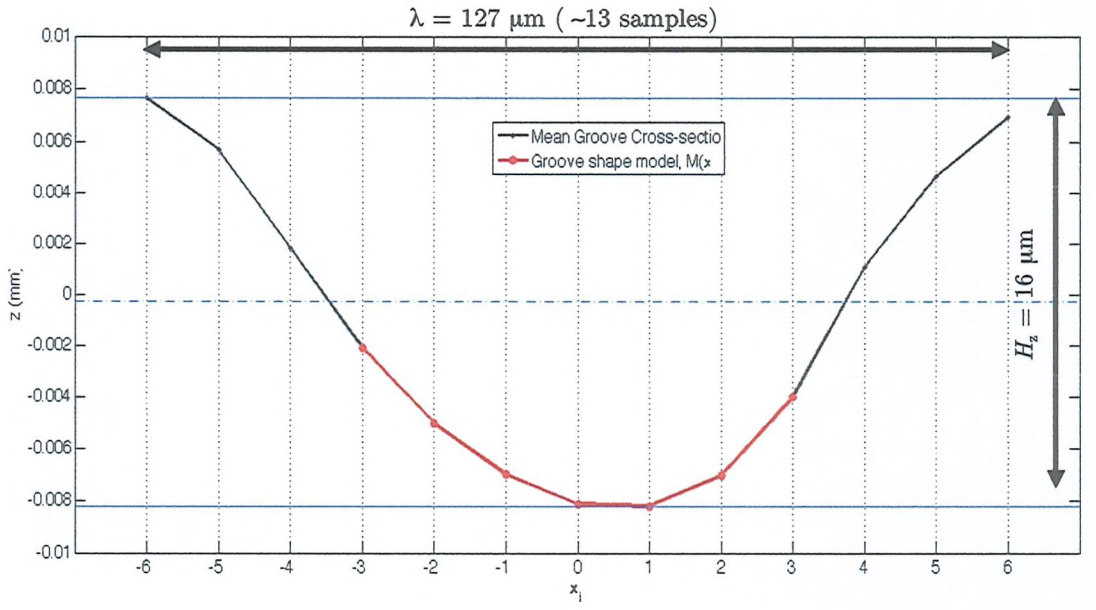

Figure 5.17: Groove shape parameters for PBEAR cylinder.

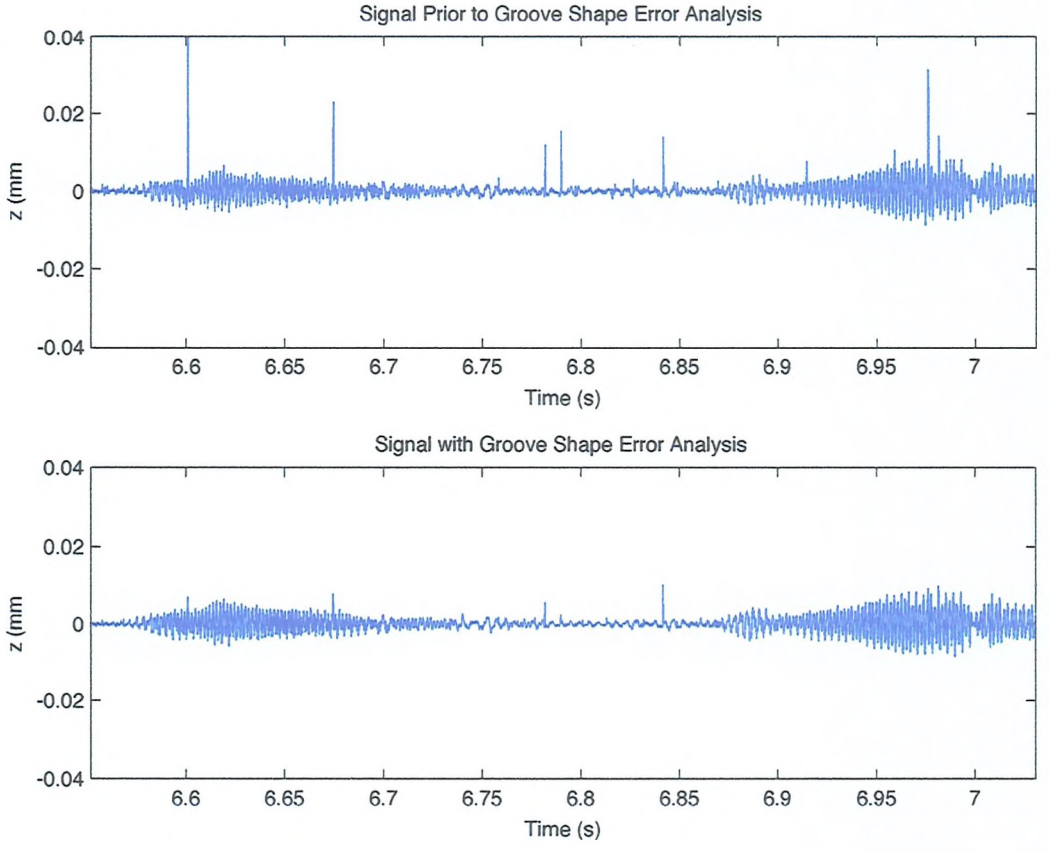

Figure 5.18: Segment of audio signal from PBEAR cylinder to show the effect of impulsive noise removal via groove shape error analysis. The majority of impulsive noise attributed with debris and outlying data points has been removed in the bottom plot.

In order to make comparisons with stylus replay, the PBEAR cylinder was transferred via stylus at the British Library Sound Archive, using the same apparatus as for the TC1 transfer. No noise reduction was applied, in order to make a fair comparisons with the non-contact method. The same high-pass filter was applied to both the stylus and optical replay in order to make a like-for-like comparison. This filter was a 4-th order Butterworth filter (high-pass), with cut-off frequency of $f_c = 150$Hz, (as was shown in Figure 4.34). The raw displacement signal $z_a(n)$ was high pass filtered. (It should be noted that response from the magnetic stylus cartridge is a velocity signal, not the displacement as with $z_a(n)$, hence equalisation is not identical)

Figure 5.19 shows a comparison of the optical and stylus time histories, and Figure 5.20 shows time-frequency plots of identical regions of the cylinder. The similarity of the time history between optical and stylus methods shows that optical reproduction is comparable with stylus replay. The optical playback sample rate of $f_s = 9.6$ kHz means that the highest frequency which can be resolved by Nyquist criterion is 4.8 kHz. With stylus reproduction, a 96 kHz playback sample rate was used. From the spectrograms shown in Figure 5.20, most of the signal content attributed to the singing, voice or instruments lies below 4 kHz. The highlighted region in Figure 5.20 shows that the stylus transfer has signal content which lies above 4.8 kHz, which suggests that the cylinder should be rescanned with revised θ-grid spacing, in order to make a direct comparison.

In general, the stylus transfer is of higher fidelity (due in part to its greater frequency sampling), with less surface noise and a better signal to noise ratio. The optical transfer suffers from more low frequency noise, caused by the surface eccentricity and a general lack of fidelity towards the higher end of the frequency spectrum. Again, the reader is encouraged to visit [86] for audible comparisons.

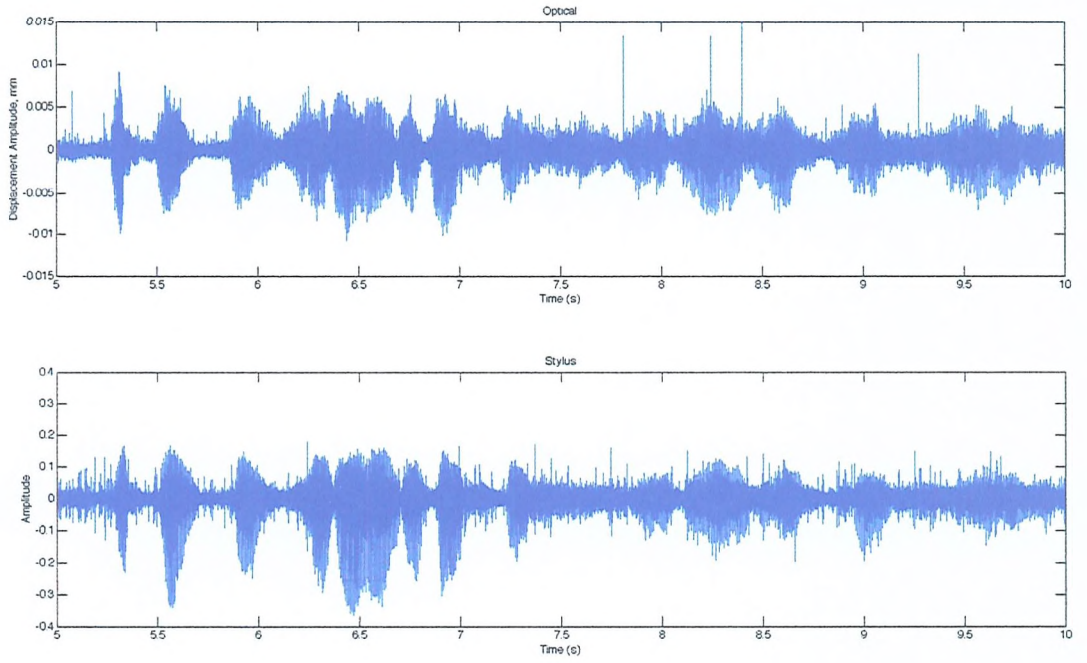

Figure 5.19: Time history of optical (top) and stylus reproduction (bottom) from PBEAR cylinder.

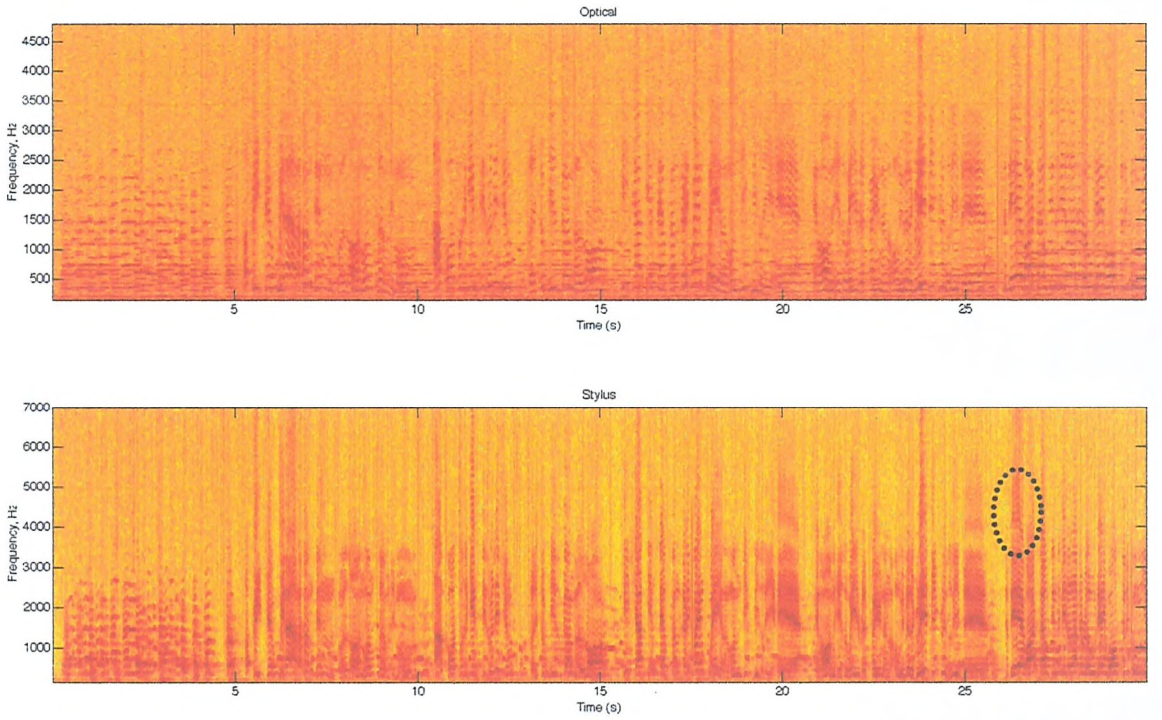

Figure 5.20: Time-frequency plots from identical regions of optical (top) and stylus reproduction (bottom)of PBEAR cylinder. The highlighted region in the stylus transfer suggests that harmonic content attributed to vocals exists up to 4.8 kHz, which is approaching the upper reproducible limit of the optical transfer with $\Delta\theta = 0.1°$, $f_s = 9.6$ kHz.

5.4 Results for Cylinders in Poor Condition

This section gives an example of signal reproduction for a cylinder which is in very poor surface condition. The cylinder in question was originally broken and reconstructed and contains the voice of a Welsh Preacher, (Evan Roberts) with chorus. Figure 5.21 shows the reconstructed cylinder. Red regions introduced during the reconstruction denote missing portions of the cylinder, filled-in with wax. In these regions, signal dropout is certain. Details and measurement parameters of this cylinder are given in Figure 5.22.

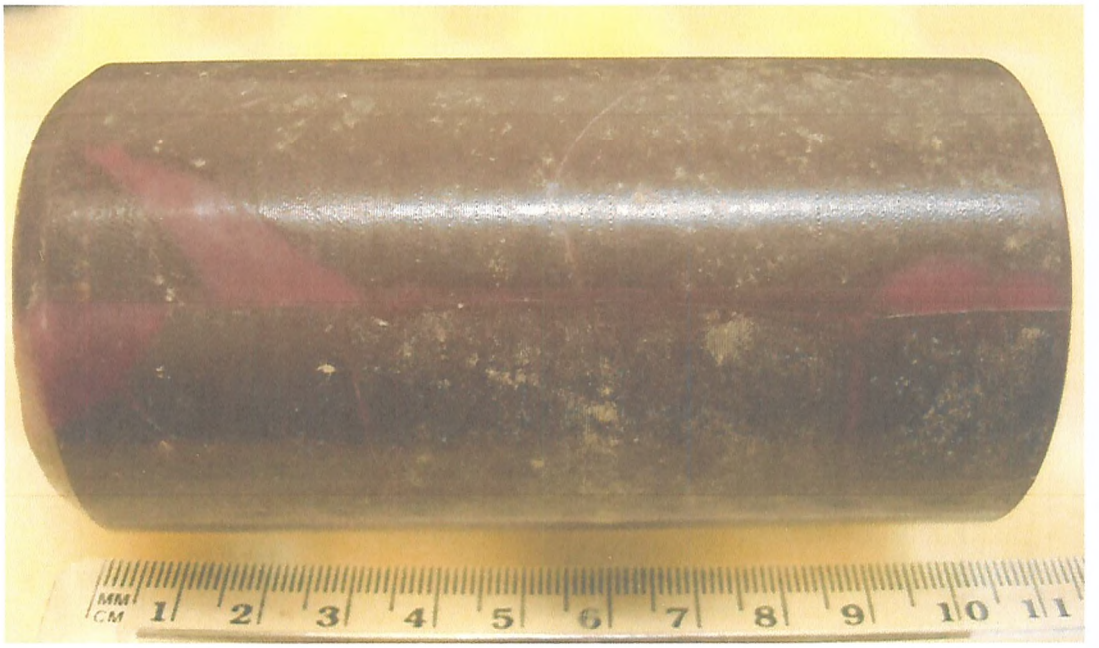

Figure 5.21: Reconstructed cylinder (EVAN), believed to contain the voice of Welsh preacher, Evan Roberts.

Figure 5.23 shows a surface plot of the cylinder, highlighting the misalignment of cylinder segments following the reconstruction process. The implication for signal reproduction is that the stylus trajectory no longer follows a perfect helix. For this reason, the cylinder was tracked using a combination of the minima-map tracking method described in section 4.4.6, and manually adjusting the trajectories node positions, by visually assessing where the track should go. This manual tracking was required because in some cases, the stylus trajectory became ambiguous.

Parameter	Value
Production Year	1905
Recording Method	Acoustic
Material	Wax
RPM	~160
Groove pitch (tpi)	100
Duration	~2 minutes
Description	Dialogue from Welsh Preacher Evan Roberts and Chorus
Measurement details	Measured in 35 unequal segments
Δx	10 μm
$\Delta\theta$	0.02°
Δz	10 nm
Playback sampling frequency, f_s	48 kHz

Figure 5.22: Parameters and measurement details of Evan Roberts Cylinder (EVANS)

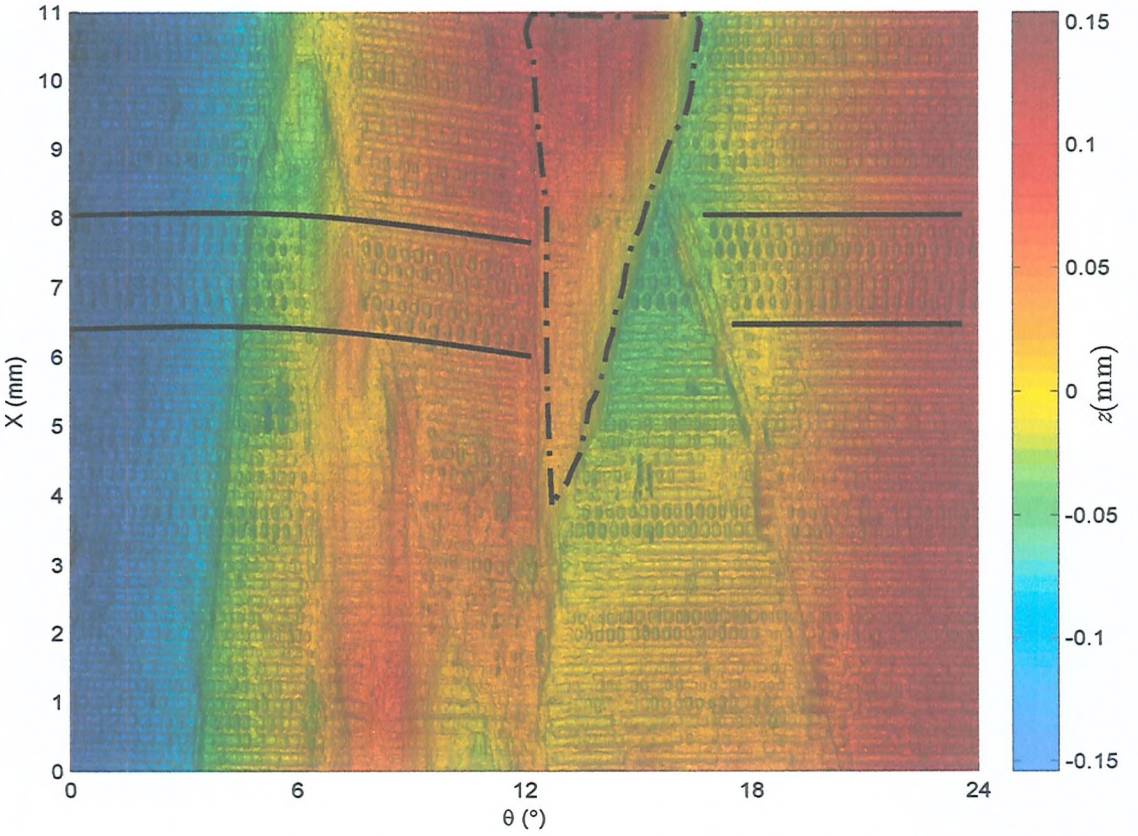

Figure 5.23: 2-D surface plot of damage region from Evan Roberts Cylinder (EVANS). The dashed region indicates a region of the surface where no apparent groove structure is present. Within the bands shown by solid lines groove modulations are clearly visible but the stylus trajectory follows an ambiguous path through the damage region. Manual tracking of the groove is required in this case.

The groove shape model fitting method was not used in this case to obtain an estimate for $z_a(n)$, because it was noted that the groove shape was highly variable over the duration of the recording, due to heavy groove damage. Instead, the SG-filtering method was used to form the track, with parameters $P = 1$. $L = 15$. This was judged subjectively to give the best SNR of all groove depth estimates.

The Evan Roberts cylinder had previously been transferred at the British Library sound following its reconstruction (but prior to surface measurement) via the same playback apparatus as previous transfers. The number of playbacks and equalization is unknown, however it is observed that a certain amount of 'de-noising', and 'de-clicking' was carried out, using CEDAR [89].

Subjectively, the optical transfer appears to have better signal fidelity and less modulation in frequency. Additionally, the amount of 'skipping' of the track is greater with the stylus transfer. The condition of the surface shown in Figure 5.24 perhaps explains why this is so, because the stylus is unable to accurately track the groove where reconstructed segments are misaligned.

With the Evan Roberts cylinder, the advantage of tracking the groove in post-processing is very clear. A stylus is not able to navigate the complex groove structure correctly, which means that 'skipping' and signal loss occurs. This is shown graphically in the spectrogram of Figure 5.24, whereby the tonal content can be seen to be more consistent with the optical transfer, than with the stylus, where there is a greater amount of signal drop out, due to incorrect tracking. Comparisons of audio from the stylus and optical transfers can again be found at [86].

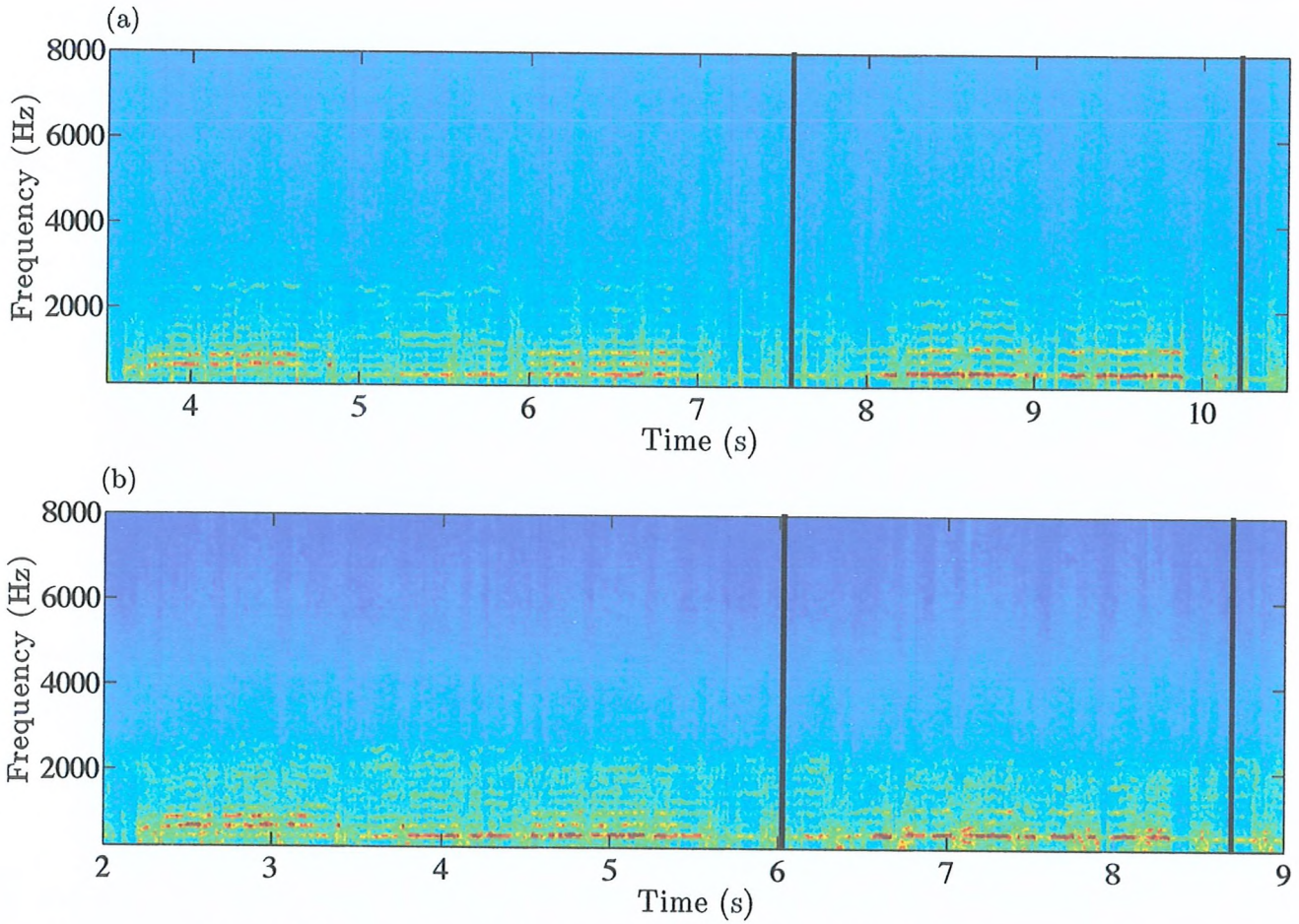

Figure 5.24: Spectrogram comparison of (a) optical and (b) stylus reproduction of Evan Roberts cylinder (EVANS). The time axes differ, but audio signal is taken from the same opening passage of the cylinder. The highlighted section denotes the region of cylinder which has been manually tracked. It can be seen that the optical transfer shows a more consistent tonal section, whereas the stylus transfer breaks up due to incorrect tracking of the groove.

5.5 Tinfoil Sound Recovery

In October 2008, the British Library Sound Archive provided the SAP with a tinfoil for replay using the NCFSM method. Details of the recording are unknown, with little documentation regarding the artefact, though it was believed by the curator to contain speech. There is no known stylus transfer of this recording, hence this represents the first attempt to playback the artefact. It is likely however, that the tinfoil was originally played back a number of times, given its poor surface condition.

A summary of the known tinfoil parameters and measurement details are given in Figure 5.25.

Parameter	**Value**
Production Year	Unknown, pre-1900
Recording Method	Acoustic
Recorded Length	75mm
Groove pitch, tpi (λ_x)	~8.5 (3000 μm)
Diameter (unravelled length)	325 mm
Material	Tinfoil
RPM	~60
Duration	~20 seconds
Description	Tinfoil recording of unknown origin.
Measurement details	24 Grooves over three sheets, measured in a total of 330 surface patches (variable size).
Δx	5 μm
Δy	25 μm
Δz	10 nm
Assumed Playback sampling frequency	63.8 kHz

Figure 5.25: Details of Tinfoil recording and measurement parameters.

The tinfoil was broken into two separate pieces, with the larger piece (referred to as 'Sheet 2' and 'Sheet 3'), folded in half and a smaller section ('Sheet 1'). An image acquired using a flat bed scanner is shown in Figure 5.26. This made scanning of the artefact more complex than for complete specimens.

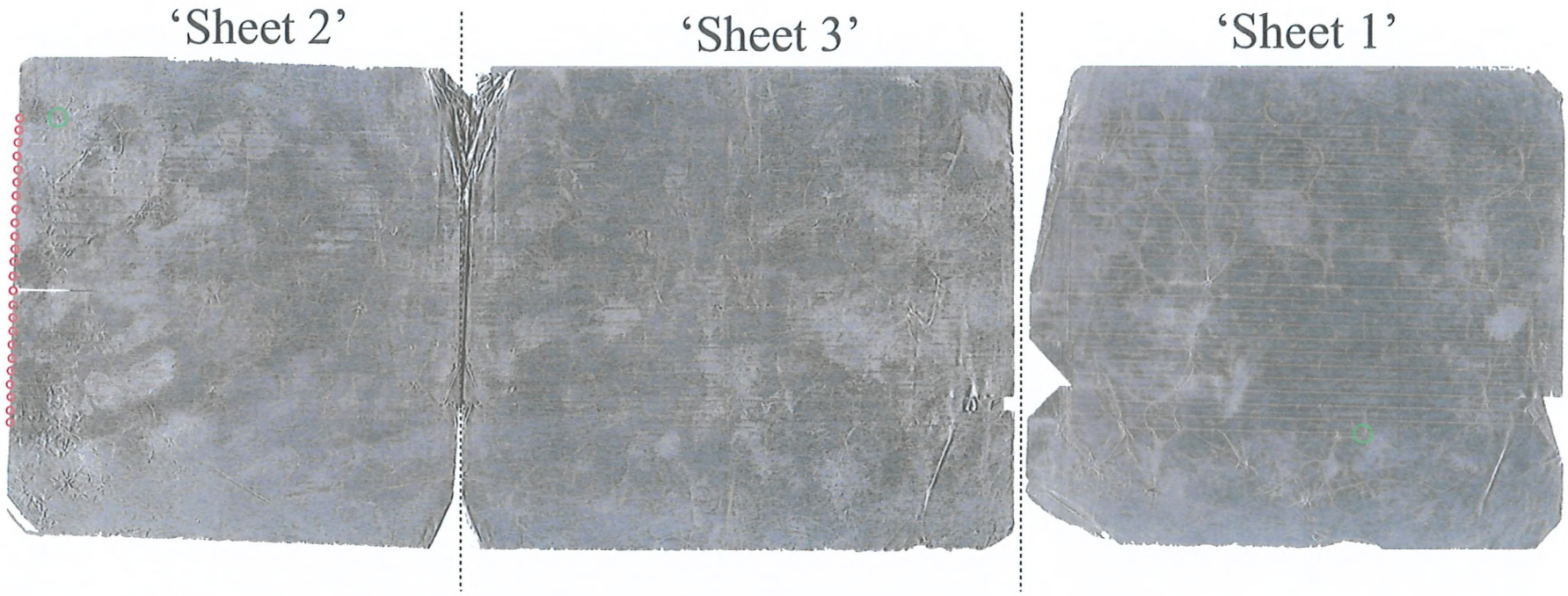

Figure 5.26: Images of the British Library Tinfoil measured via flatbed scanner. The green circles denote the assumed start and end points of the recording. Red circles denote the groove seed.

Due to its flattened surface, the air-bearing system was used to measure the tinfoil sheets. Unlike the disk measurement system which used x (translation), θ (rotation) and z (depth) axes, the tinfoil was measured over an x-y-z grid, with sampling intervals Δx and Δy and Δz. The tinfoil was oriented in such a way that measurements in x corresponded to the linescan axis, (as with cylinder measurement) and the y-axis corresponded to the playback direction (circumferential).

Each sheet was measured individually, with 24 grooves per sheet. Each groove was sub-divided into a variable number of segments, which was dependent on the degree of deformation, and calculated using the optimisation routine described in Section 3.2.4. In total, 330 surface patches were acquired using the air-bearing system and stitched together.

One problem that was noted during scanning was that the tinfoil deformed noticeably over the course of measurement, due to airflow/temperature changes in the lab. Each groove was measured individually across Sheet 1, 2 and 3. The order in which the sheets 'fitted together' was determined by noting two termination points of the groove track, where the recording tool had 'entered' and 'exited' the sheet.

Due to the complex construction of the discrete surface, each groove was tracked individually using the same methods used for the EVAN cylinder, using a minima map and tracking method, followed by some manual correction for regions where the grooves were torn, or did not exist. The depth estimation method that subjectively provided the best signal to noise ratio was a 7pt mean average (Scheme 2), found along the trajectory A_n. The exact playback speed of the tinfoil was unknown, as tinfoil phonographs were generally controlled via a hand crank. The sampling in the y-axis was chosen assuming a rotational speed of 60 rpm, which seemed reasonable for a hand-cranked cylinder (one rotation per second). On playback, it was estimated that a resultant sampling frequency of 63.8 kHz gave the most likely playback speed, based on the observed frequencies attributed to speech. Figure 5.27 shows a short section of the recording, with vocal frequencies (500-1000 Hz) present. Speech intelligibility is very low and there is very little signal content above 1.2 kHz, but the signal content is clearly attributed to a human voice.

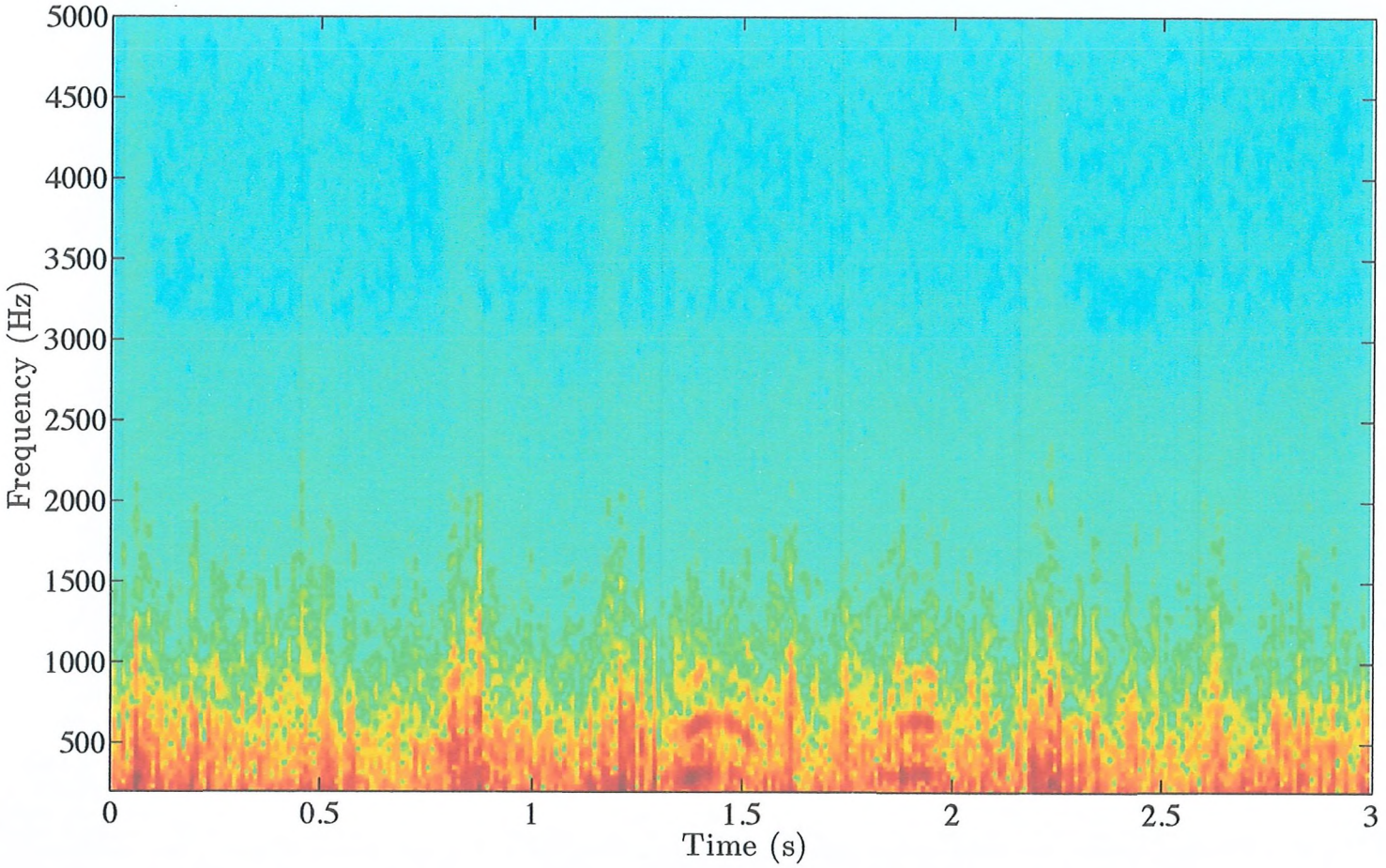

Figure 5.27: Three second spectrogram of tinfoil signal reproduction via the NCFSM method. There is no considerable audible content attributed to speech above 1.2 kHz. Harmonics attributed to the human voice are visible between t = 1-2 seconds.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

In this thesis, the development of methods for sound reproduction from discrete surface maps of early mechanical recordings has been described. The non-contact measurement system that has been developed allows for accurate sound reproduction from cylinder recordings. The advantage of mapping the full surface topology (as opposed to real-time optical playback) is that numerous estimates for the audio signal can be derived from a single data set in post-processing. In addition, the NCFSM method allows for playback of fragile media such as tinfoils, which is not possible with conventional contacting methods.

In order to extend the methods initially developed by Haber et. al, a thorough investigation into the noise generation processes caused by scanning and sound recovery methods was first carried out.

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

tracking of the groove in post-processing. It was shown that the same recording transferred via stylus was incorrectly tracked, which led to a loss of signal information.

In order to avoid the introduction of temporal distortion (echo), the use of data from the groove ridge is not advisable. In addition, it was shown that the selection of data points from the groove side-wall should be avoided, as noise variance was greater here than at the groove bottom.

Savitzky-Golay polynomial smoothing filters are well suited for feature extraction from linescans with a known periodic structure. The ridge-valley features are easily identifiable by matching the filter parameters with the features of interest.

The condition and geometry of groove cross-sections is highly variable across the different specimens encountered in this work. Therefore, (in the same way that archivists experiment with different stylus types), a number of different signal estimation routines must be explored, in order to obtain an 'optimal signal'. What one considers 'optimal' for specimens containing musical signals is highly subjective, and is beyond the scope of this Thesis. It is highly likely that if the NCFSM method is to be used commercially by sound archivists, that a number of different signals would be extracted, and then rated subjectively via a listening tests.

In an attempt to quantify the sound quality, a test cylinder was specially produced, in order to obtain signal to noise ratio measurements. This process proved difficult due to the lack of professionally available manufacturing equipment. It does however provide a unique reference for conducting signal to noise tests, and gives a way of directly comparing the NCFSM method with conventional stylus methods. For test signals such as sinusoids, it is possible to deduce and rate signal reproduction quantitatively, and to judge which estimate provided the optimal signal. A number of different groove depth estimation methods were investigated and rated in terms of signal to noise ratio and total harmonic distortion. Criteria suggested in Section 4 lead to the development of a groove model-fitting method, which accounted for linescan offset and groove damage. It was found that this model fitting method, gave

the best signal to noise ratio, and the Savitzky-Golay method produced the lowest total harmonic distortion for sinusoidal signals.

Due to the accuracy and evenness of angular rotation at low speed during measurement, the NCFSM method does not appear to be affected by speed fluctuations caused by the non-compliance with cylinder and mandrel. The resultant signal has been shown to suffer less from speed fluctuations known as 'wow', which are commonly associated with stylus reproduction.

The interaction between stylus and groove, which causes geometric distortion ('tracing distortion'), does not affect the NCFSM method. As a result, signal reproduction, (especially at high frequencies), may be considered more accurate with the NCFSM method than with stylus reproduction. This can be seen by reduced harmonic distortion, present in the test cylinder signal reproduction for NCFSM, when compared with the stylus reproduction.

The sound reproduction methods developed rely on an accurate measurement at the data acquisition stage. With all discrete measurement techniques, there is a compromise between data density, measurement time and memory constraints.

It can be concluded that in order for the NCFSM method to meet the recommended archival standards for digitisation of analogue media, a minimum of 48 kHz playback sample rate is needed at 16-bit word length is required. The current system permits a playback sample rate of up to 96 kHz if required, but fails to meet the bit-resolution requirement of 16-bit by some margin. The ability to resolve depth to sufficient resolution is the one of the limiting factors of an optical system.

Another limiting factor of the data acquisition system is the sensor gauge range. In order to measure the full surface of a cylinder recording using a single displacement sensor, it is required that the surface be scanned in piece-wise segments. This is not ideal for mapping continuous surfaces. In fact, the increased number of scans increases the complexity of surface reconstruction.

Difficulties encountered from very early wax cylinders include poor surface condition, such as cracking, mould growth and wear caused to the groove structure through repeated stylus playback or poor storage. In cases where a cylinder is

cracked, it is possible to scan the broken segments and then reconstruct the surface. In cases where the groove structure is damaged, it is possible to track more closely in post-processing, than with playback in situ.

6.2 Future Work

The measurement systems described in this thesis remain in development at time of writing. Future improvements to the data acquisition system are required – especially for measuring disc recordings. The measurement requirements for disc reproduction are more demanding than for cylinders, due to their increased surface area, slower playback speeds, and increased angle of groove wall. To obtain the equivalent of 16-bit audio resolution for discs, a grid sampling of approximately 4 nm is required. Investigation needs to be carried out as to whether this is realistically achievable with a WL sensor spot size currently of 7μm.

It future, multiple sensors may be required, in order to reduce scan times, and sensors. The sensor(s) may also need to be oriented differently, in order to resolve the 45° groove walls of disc recordings for instance. This brings about another complexity to the signal reproduction stage, in that the surface must first be reconstructed from multiple surface patches, and possibly in a new coordinate system.

Throughout the project the surface segments have been acquired without any overlap. In many data stitching applications, it is advantageous to include an overlap between adjacent segments, in order to aid the registration process during image reconstruction. It is therefore suggested that a future implementation of the scanning software should incorporate an overlap between segments, to aid surface reconstruction.

The use of additional test cylinders to rate the NCFSM method quantitatively is of vital importance to the SAP. Such recordings are unfortunately not available as stock artefacts from sound archives, and must be specially made using custom recording equipment. The SAP is currently in possession of calibrated 78 rpm disc

recordings, which will be of great use for quantifying the signal reproduction from discs measured by the NCFSM method.

Future algorithm development is required for dealing with fragment reconstruction, in the cases of broken recordings for instance, which have not been reconstructed. In this case, the task for post-processing is similar to that of a 'jigsaw' puzzle. Boundaries and ridge orientations will likely form the features, which can be matched when piecing together the jigsaw. Coupled with this idea is the idea of 'blind audio fragment reconstruction'. Is it possible for instance, to simply extract all audio information in the grooves and then reconstruct the audio signal automatically in the time domain, using statistical measures at the segment boundaries, or will this procedure always require manual interaction?

It may be possible to merge macroscopic, colour imaging with surface data from the NCFSM method, in order to identify mould or damage sites. Such a source of data fusion would be an advantage for automatically identifying troublesome areas.

From an archival perspective, there also needs to be a formal definition of a 'preservation surface' for archival purposes. This would need to be derived by communicating with archivists as to their specific resolution requirements.

References

[1] US Patent No. 200,521, Phonograph or Speaking Machine. Patented awarded to Thomas Alva Edison, February 19th 1878.

[2] Thomas Edison's Notebooks, Edison National Historic Site Archive, <http://www.nps.gov/edis/sounds.htm>, last accessed 25/07/09.

[3] Brock-Nannestad, G. (1998). Up-To-Date Replay Facilities for Obsolete Mechanical Sound Recording Formats, 104th AES Convention, Amsterdam.

[4] British Library Archival Sound Recordings, <http://sounds.bl.uk>, last accessed 25/07/09.

[5] Fadeyev, V. and C. Haber (2005). "Reconstruction of Recorded Sound from an Edison Cylinder using Three-Dimensional Non-Contact Optical Surface Metrology." JAES 53(6): 485-508.

[6] Fadeyev, V. and C. Haber (2003). "Reconstruction of Mechanically Recorded Sound by Image Processing." JAES 51(12): 1172-1185.

[7] Millard, A, *America on Record: A History of Recorded Sound,* Cambridge University Press, Cambridge, 1995.

[8] Read, O. and Welch, W. L., From Tin Foil to Stereo, (1976), Howard W. Sams & Co., Inc., 1976

[9] French patent No. 31470, Phonautograph, awarded to Édouard-Léon Scott, 18th May 1857.

[10] Early Recorded Sounds and Wax Cylinders Website, <http://www.tinfoil.com>, last accessed 25/07/09.

[11] The Edison Papers Website, <http://edison.rutgers.edu/tinfoil.htm>, last accessed 25/07/09.

[12] Recording Devices in Smithsonian National Museum of American History, <http://history.sandiego.edu/gen/recording/nonelectrical>, last accessed 25/07/09.

[13] US Patent No. 341,214, Graphophone Patent - Recording and reproducing speech and other sounds, awarded to C. A. Bell and S. Tainter 4th May 1886.

[14] Edison, T. A., (1888), The Perfected Phonograph, North American Review, Volume 146, Issue 379.

[15] Klinger, B. (2007). " CYLINDER RECORDS: Significance, Production, and Survival ." ARSC.

[16] Guy, P. J. (1964). Disc recording and reproduction. New York, Focal Press.

[17] Klinger, B. (2002), "Stylus Shapes and Sizes: Preliminary Comments on Historical Edison Cylinder Stylii", ARSC

[18] Webster, A. G., (1919) Acoustical Impedance, and the Theory of Horns and of the Phonograph, Proc. Nat. Academy of Sciences of the USA, 275-282

[19] Wilson, P, Wilson, G. L, (1974) Horn Theory and the Phonograph, AES Preprint 1002-b, AES Convention 49.

[20] Maxfield, J.P. and Harrison, H.C., The Bell System Technical Journal, Volume V, No.3, July 1926, pp. 493-524.

[21] Olson, H. F., (1957) Acoustical Engineering, D. Van Nostrand Company, Inc.,

[22] Galo, G, Disc Recording Equalization Demystified, (1996), ARSC Journal, 27:2, 188-211.

[23] Vaseghi, S. V. (1996), Advanced Digital Signal Processing and Noise Reduction, p29.

[24] Hunt, F. V. (1955). "On Stylus Wear and Surface Noise in Phonograph Playback Systems." AES 3(1): 2-18.

[25] Godsill, S. and P. Rayner (1998). "Digital Audio Restoration.", Springer.

[26] Godsill, S. and C. H. Tan (1997). "Removal of low frequency transient noise from old recordings using model-based signal separation techniques." IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics.

[27] Godsill, S. J. and P. J. W. Rayner (1995). "A Bayesian approach to the restoration of degraded audio signals." Speech and Audio Processing, IEEE Transactions on 3(4): 267-278.

[28] Photomicrograph image, Ediphone School Record, No. 9. by Jerry Fabris, March 2002.

[29] H. E. Roys, (Ed.), *Disc Recording and Reproduction, Benchmark Papers in Acoustics*, vol. 12, Dowden, Hutchinson & Ross, 1978.

[30] Bachman, W. S., B. B. Bauer, et al. (1962). "Disk Recording and Reproduction." Proceedings of the IRE 50(5): 738-744.

[31] U.S Patent No. 372,786, *Gramophone*, award to Emile Berliner, Nov 8, 1887.

[32] Meulengracht-Madsen, Hans, (1976) On the Transcription of Old Phonograph Wax Records, JAES 24(1): 27-32.

[33] Publication no. IASA-TC 04 (2004) IASA Technical Committee, Standards, Recommended Practices and Strategies, Guidelines on the Production and Preservation of Digital Objects.

[34] Publication no. IASA-TC 03, (2005) IASA Technical Committee, Standards, Recommended Practices, The Safeguarding of the Audio Heritage: Ethics, Principles and Preservation Strategy.

[35] W. D. Lewis and F. V. Hunt (1941). "A Theory of Tracing Distortion in Sound Reproduction from Phonograph Records", J. Acoust. Soc. Am. 12 348.

[36] Pierce, J. A. and F. V. Hunt (1938). "On Distortion in Sound Reproduction from Phonograph Records." JASA 10(1): 14-28.

[37] Bauer, B. B. (1945). "Notes on Distortion in Phonograph Reproduction Caused by Needle Wear." JASA 16(4): 248-253.

[38] Rabinow, J. and E. Codier (1952). "Phonograph Needle Drag Distortion." JASA 24(2): 216-225.

[39] Roys, H. E. (1953). "Distortion in Phonograph Reproduction." JASA 25(6): 1140-1144.

[40] Bastiaans, C. R. (1967). "Factors Affecting the Needle/Groove Relationship in Phonograph Playback Systems." AES 15(4): 389-399.

[41] Petrov, Kryuchin, et al. (1997). Optomechanical Method of Sound From Edison Cylinders. International Conference on Optical Storage, Imaging and Transmission of Information.

[42] Shanoylo, Kosyak, et al. (2001). Reading and processing of audio information reproduced from Edison phonograph cylinders by method of laser interferometry. Laser Techniques and Systems in Art Conservation.

[43] Penn, W. A. and M. J. Hanson (2003). "The Syracuse University Library Radius Project: Development of a non-destructive playback system for cylinder recordings." First Monday 8(5).

[44] Iwai, T., Asakura. T., et al., Reproduction of sound from old wax phonograph cylinders using the laser-beam reflection method, Appl. Opt. 25 (5) (1986) 597–604.

[45] Uozumi, J., T. Ushizaka, et al. (1998). Optical Reproduction of Sounds from Negative Phonograph Cylinders. Contributions to the Fifth International Conference on Optics Within Life Sciences OWLS V, Crete, Springer.

[46] Nakamura, T. and T. Asakura (1999). "Reproduction of Sounds from an Old Russian Wax Phonograph Cylinder by Various Optical Methods." OWLS: 54-57.

[47] Asakura, T. and J. Uozumi (1999). "Optical methods for Reproducing Sounds from Old Phonograph Records." International Trends in Optics and Photonics ICO IV: 65-81.

[48] Heine, W. K. (1977). A Laser Scanning Phonograph Record Player. 57th AES Convention.

[49] ELP (1997). ELP laser turntable. <http://www.elpj.com>, last accessed 25/07/09.

[50] Stoddard, R. E. and R. N. Stark (1988). The Optical Turntable, Finally a Reality. 85th Audio Engineering Society Convention. Los Angeles, California.

[51] Uozumi, J. and T. Asakura (1988). "Reproduction of Sound from old disks by the laser diffraction method." Applied Optics 27(13): 2671-2676.

[52] Digital Needle Project - Team Light Blue, KTH, Stockholm, Sweden, <http://www.s3.kth.se/signal/edu/projekt/students/03/lightblue>, last accessed 25/07/09.

[53] Digital Needle Project - Team Light Green, KTH, Stockholm, Sweden, <http://www.s3.kth.se/signal/edu/projekt/students/03/lightgreen/>, last accessed 25/07/09.

[54] Stotzer, S., O. Johnsen, et al. (2007). "Phonographic sound extraction using photography and signal processing." Digital Signal Processing 17(2): 433-450.

[55] Johnsen, O., S. Stotzer, et al. (2007). Detection of the Groove Position in Phonographic Images.. ICIP 2007. IEEE International Conference on Image Processing.

[56] Stotzer, S., O. Johnsen, et al. (2006). Groove Extraction of Phonographic Records. DAS 2006, 7th IAPR Workshop on Document Analysis Systems, Nelson New Zealand,.

[57] Stotzer, S. (2006). Phonographic Record Sound Extraction by Image Processing. Faculty of Science, University of Fribourg. Doctor Scientiarum Informaticarum: 217.

[58] Stotzer, S., O. Johnsen, et al. (2004). "Phonographic Sound Extraction Using Image and Signal Processing." Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). 4(5): 289-292.

[59] Stotzer, S., O. Johnsen, et al. (2004). VisualAudio: an Optical Technique to Save the Sound of Phonographic Records. Joint Technical Symposium 2004, Toronto, Canada.

[60] Perceptual Evaluation of Audio Quality (PEAQ), <http://www.peaq.org>, last accessed 25/07/09.

[61] Huber, R. and B. Kollmeier (2006). "PEMO-Q : A New Method for Objective Audio Quality Assessment Using a Model of Auditory Perception." Audio, Speech and Language Processing, IEEE Transactions on Speech and Audio Processing, 14(6): 1902-1911.

[62] Green, D. M., & Swets, J. A. (1966). Signal detection theory and psychophysics. New York: John Wiley and Sons.

[63] A Whitaker (1928), "Progress in the recording and reproduction of sound", J.Sci. Instrum. 5 35-41.

[64] Fermilab Symposium on the Nature of Science, October 8, 2008, <http://irene.lbl.gov/Fermilab-Symposim-2008.pdf>, last accessed 25/07/09.

[65] Haralick, R. M. (1983). "Ridges and valleys on digital images." Computer Vision, Graphics, and Image Processing 22(1): 28-38.

[66] Eberly, D., R. Gardner, et al. (1994). "Ridges for image analysis." Journal of Mathematical Imaging and Vision 4(4): 353-373.

[67] Lopez, A. M., F. Lumbreras, et al. (1999). "Evaluation of methods for ridge and valley detection." Pattern Analysis and Machine Intelligence, IEEE Transactions on 21(4): 327-335.

[68] Gauch, J. M. and S. M. Pizer (1993). "Multiresolution analysis of ridges and valleys in grey-scale images." Pattern Analysis and Machine Intelligence, IEEE Transactions on 15(6): 635-646.

[69] Yin, J., E. Zhu, et al. (2007). "Two steps for fingerprint segmentation." Image and Vision Computing 25(9): 1391-1403.

[70] Press, W. H., Teukolsky, et al. (1996). Numerical Recipes in C., Cambridge University Press.

[71] Savitzky, A. and M. J. E. Golay (1964). "Smoothing and Differentiation of Data by Simplified Least Squares Procedures." Analytical Chemistry 36(8): 1629-1639.

[72] Gilmore, I. S. and M. P. Seah (1996). "Savitzky and Golay differentiation in AES." Applied Surface Science 93(3): 273-280.

[73] Turton, B. C. H. (1992). "A Novel variant of the Savitzky-Golay filter for spectroscopic applications." Measurement Science and Technology 3: 858-868.

[74] Orfanidis, S.J. Introduction to Signal Processing (1996), Prentice-Hall, Englewood Cliffs, NJ.

[75] Bromba, M. and H. Ziegler (1981). "Application Hints for Savitzky-Golay Digital Smoothing Filters." Analytical Chemistry 53(11): 1583-1586.

[76] Boltryk, P. J., M. Hill, et al. (2008). "A comparison of precision optical displacement sensors for the 3D measurement of complex surface profiles." Sensors and Actuators A: Physical 142(1): 2-11.

[77] Confocal laser principle from Keyence website: <http://www.keyence.com>, last accessed 25/07/09.

[78] White Light confocal measurement principle image, http://www.stilsa.com

[79] Boltryk, P.J., McBride, J.W., Hill, M., Nascè, A.J., Zhao, Z. and Maul, C., Non-contact surface metrology for preservation and sound recovery from mechanical sound recordings, (2008), J. Audio Eng. Soc, 56, (7/8) 545-559.

[80] Kay, S. M, Modern Spectral Estimation: Theory and Application, (1988). Prentice Hall.

[81] Hawkins, D. M. , "Identification of Outliers", Chapman Hall (1980).

[82] Nixon, M. and Aguado, A., Feature Extraction & Image Processing, Newnes, 2002

[83] Yu, C. and Q. Peng (2007). "A correlation-based phase unwrapping method for Fourier-transform profilometry." Optics and Lasers in Engineering 45(6): 730-736.

[84] Nascè, A.J., McBride, J.W., Hill, M. and Boltryk, P.J. (2007) Processing Techniques for the Recovery of Audio from Edison Cylinder Recordings, via non-contact Surface Measurement. 31st AES International Conference, London, June 2007.

[85] Nascè, A.J., Hill, M., McBride, J.W., and Boltryk, P.J. (2008). "A quantitative analysis of signal reproduction from cylinder recordings measured via noncontact full surface mapping." The Journal of the Acoustical Society of America 124(4): 2042-2052.

[86] Sound Archive Project Website: <http://archivesound.co.uk>, last accessed 25/07/09.

[87] Calibration Disc Set. Catalogue number AES-S001-064: <http://www.aes.org/publications/standards/calibration>, last accessed 25/07/09.

[88] Poppy Records: <http://www.poppyrecords.co.uk>, last accessed 25/07/09.

[89] Cedar Audio Restoration Systems: <http://www.cedar-audio.com>, last accessed 25/07/09.

Appendix A

Cylinder Reference Details

Below is a list of the naming conventions used throughout this thesis in reference to cylinder surfaces measured by the NCFSM method during the Sound Archive Project:

- GRAPH – Graphophone wax cylinder, scanned at British museum.
- TC1 – Test cylinder, encoded with pure tones, produced at Poppy Records.
- TC2 – Test cylinder, encoded with various waveforms at Poppy Records.
- JBTBM – 'Just Before the Battle Mother', Blue Amberol Cylinder.
- PBEAR – 'The Preacher and the Bear', Black Amberol Cylinder.
- BBIRDS – 'Beautiful Birds Sing On', Black Amberol Cylinder
- EVAN – Brown wax cylinder - voice of Evan Roberts and chorus.

Appendix B

Cylinder measurement system

This section gives specifications of the non-contact full-surface mapping system developed by the Sound Archive Project for cylinder recordings. The cylinder system consists of:

- A STIL confocal white light probe,
- A Newport VP-25XL translational x-y stage,
- A Newport RV12- CCHL Rotary Stage.

A list of the measurement specifications are given below:

Parameter	**Value**
Resolution of X-Y linear stages	100 nm
Resolution of Z-stage	10 nm
Sensor Spot Size (diameter)	7 μm
Z-axis sensor resolution (WL)	10 nm
Resolution of rotary stage	0.001°
Gauge range	350 μm
Range of sensor movement in Z	25 mm

Appendix C

Air-bearing measurement system

This section gives specifications of the air-bearing system developed by the Sound Archive Project for measurement of disc media and tinfoil recordings. The air-bearing system designed for measurement of tinfoil and disc media consists of:

- A STIL confocal white light probe
- A Newport dynamYX with 2 air-bearing axes
- A Newport VP-25XL translational Z-stage
- A Newport URS150 Rotary Stage.

A list of the measurement specifications are given below:

Parameter	Value
Resolution of X-Y linear stages	2 nm
Resolution of Z-stage	10 nm
Resolution of rotary stage	0.0005°
Z-axis sensor resolution (WL)	10 nm
Sensor spot size (diameter)	7 μm
Range of sensor movement in Z	25 mm
Range of measurement in X and Y	320mm

Appendix D

SNR and THD Results

This section gives tabulated results of SNR and THD scores for the four depth estimation schemes described in Chapter 4.4.7 and 5.2.1, for the TC1 test cylinder.

Tone (Hz)	Signal-to-Noise Ratio (dB)			
	$z_{a\text{-}X_{off}}$ $[X_{off}]$	$z_{a\text{-}M}$ $[M]$	$z_{a\text{-}P,L}$ $[P,L]$	$z_{a\text{-}M}$
200	29.14 [+1]	30.35 [3]	30.76 [1,15]	**31.44**
250	29.88 [+1]	30.40 [1]	30.46 [1,15]	**30.98**
315	30.29 [+2]	30.12 [2]	30.79 [1,19]	**31.53**
400	30.46 [+1]	31.88 [4]	31.89 [1,15]	**32.26**
500	29.45 [+2]	29.58 [1]	29.82 [1,11]	**30.58**
630	29.00 [-1]	29.34[1]	29.56 [1,17]	**29.80**
800	28.81 [0]	29.74 [1]	29.85 [1,19]	**30.25**
1000	28.43 [0]	30.30[5]	31.0 [1,15]	**31.47**
1250	28.03 [+1]	29.19 [4]	29.16 [1,11]	**29.62**
1600	26.63 [0]	28.09 [5]	28.22 [1,11]	**28.64**
2000	25.99 [+1]	27.54 [5]	27.74 [1,15]	**27.71**
2500	23.59 [+1]	25.03 [5]	25.32 [1,15]	**25.71**
3150	22.93 [+1]	24.38 [5]	24.44 [1,15]	**25.02**
4000	21.28 [+1]	22.66 [5]	22.76 [1,15]	**23.03**
5000	20.23 [+2]	21.06 [4]	21.49 [1,15]	**21.71**

Values in square brackets correspond to the parameters that gave the best SNR. Bold values denote the best result across the four different estimation schemes.

Tone (Hz)	Total Harmonic Distortion (%)			
	$z_{a\text{-}AXoff}$ [X_{off}]	z_{aN} [N]	$z_{a\text{-}P,L}$ [P,L]	$z_{a\text{-}M}$
200	2.38 [+2]	2.41 [3]	**2.00** [3,19]	2.14
250	1.87 [+2]	2.01 [1]	**1.36** [5,21]	1.74
315	1.72 [+1]	2.23 [1]	**1.41**[3,19]	1.84
400	1.63 [0]	1.74 [1]	**1.53** [3,17]	1.93
500	2.98 [+2]	2.94 [1]	**1.48** [3,17]	2.11
630	3.70 [-1]	3.92 [1]	**2.21**[3,21]	**3.55**
800	**3.29** [-1]	3.70 [4]	3.83 [3,21]	3.94
1000	2.50 [-2]	2.84 [4]	**2.20** [9,21]	2.79
1250	2.20 [-1]	2.21 [2]	**1.72** [5,15]	2.15
1600	2.53 [2]	2.27 [5]	**2.25** [7,17]	2.53
2000	4.59 [+1]	4.09 [5]	**3.81**[7,17]	4.26
2500	3.32 [+1]	3.19 [4]	**1.84** [3,15]	2.52
3150	4.34 [+1]	3.49 [4]	**3.35** [3,17]	3.36
4000	4.84 [+1]	4.36 [5]	**3.52** [1,15]	4.08
5000	4.00 [+2]	3.74 [5]	**3.25** [1,15]	3.26

Values in square brackets correspond to the parameters that gave the best SNR. Bold values denote the best result across the four different estimation schemes.

A quantitative analysis of signal reproduction from cylinder recordings measured via noncontact full surface mapping

Antony Nascè, Martyn Hill,[a] John W. McBride, and Peter J. Boltryk
Electromechanical Research Group, University of Southampton, Hampshire SO17 1BJ, United Kingdom

(Received 28 March 2008; revised 24 July 2008; accepted 26 July 2008)

Sound reproduction via a noncontact surface mapping technique has great potential for sound archives, aiming to digitize content from early sound recordings such as wax cylinders, which may otherwise be "unplayable" with a stylus. If the noncontact techniques are to be considered a viable solution for sound archivists, a method for quantifying the quality of the reproduced signal needs to be developed. In this study, a specially produced test cylinder recording, encoded with sinusoids, provides the basis for the first quantitative analysis of signal reproduction from the noncontact full surface mapping method. The sampling and resolution of the measurement system are considered with respect to the requirements for digital archiving of cylinder recordings. Two different methods of audio signal estimation from a discrete groove cross section are described and rated in terms of signal-to-noise ratio and total harmonic distortion. Noncontact and stylus methods of sound reproduction are then compared using the same test cylinder. It is shown that noncontact methods appear to have distinct advantages over stylus reproduction, in terms of reduced harmonic distortion and lower frequency modulation.
[DOI: 10.1121/1.2973238]

PACS number(s): 43.38.Ne, 43.38.Md, 43.58.Ry, 43.60.Dh [AJZ] Pages: 2042–2052

I. INTRODUCTION

Sound reproduction from early mechanical recordings via noncontact optical methods has been the topic of increasing interest in recent years.[1–13] A noncontact method ensures that no further damage or pressure is applied to the often fragile grooves, wherein the sound was originally encoded. In general, two noncontact reproduction strategies have been proposed in literature: (i) signal reproduction, *in situ*;[1–4] and (ii) surface mapping via optical sensors, followed by postprocessing.[5–13]

In (i), accurate sound reproduction requires robust tracking of the grooves with a laser[1–3] or an interferometer system,[4] *in situ*. This strategy can be problematic in cases where the recording is damaged or deformed. Furthermore, only a single audio signal estimate can be obtained from each transfer attempt.

An alternative transfer strategy (ii) uses optical sensing to map the full surface of cylinder recordings. The philosophy behind this technique is aimed at preservation of the full recorded surface, not real-time playback. Most of these approaches have stored a two-dimensional (2D) image of the surface, but recent work[9–12] has measured the full three-dimensional (3D) surface topology. The advantage of mapping the full surface topology is that sound can be extracted from anywhere within the groove, any number of times, at any point in the future. Tracking of the grooves in postprocessing can therefore be more accurate than *in situ*, especially in cases where the surface is highly deformed or damaged. Furthermore, this technique has potential for recovering sound from broken artefacts, by scanning fragments of the recording, and then stitching data sets together.

In previous work[10] based on (ii), signal reproduction from cylinder recordings has been judged qualitatively to provide "faithful playback of the recorded information." In this study, we seek a method of quantifying the signal quality, in terms of metrics such as signal-to-noise ratio (SNR) and total harmonic distortion (THD). This is achieved through the use of a specially produced cylinder, encoded with sinusoids. In addition, the concept that multiple audio signal estimates can be derived from the same dataset is investigated, by comparing signal reproduction from two groove depth estimation methods. The noncontact signal reproduction described in this paper considers cylinder recordings only, but similar methods can also be applied to other early recordings such as phonograph tinfoil recordings and vertically cut disks.

This study is organized in the following way. Following a brief summary of cylinder recordings, the noncontact full surface mapping technique is described, with emphasis on sampling requirements for audio signal reproduction. In Sec. II, a review of previous work relating to signal reproduction and quality assessment from noncontact mapping methods is given. Section III gives details of the sound reproduction process and signal quality testing methods. The results are discussed in Sec. IV, and conclusions are given in Sec. V.

A. Cylinder recordings

Cylinder recordings represent the earliest mechanical means for archiving and playing back sound. Recording systems and cylinder technology developed over many years, but the basic recording principles remained the same (see Fig. 1). In recording mode, a cylinder "blank" is set to rotate with constant angular velocity, $\dot{\theta}$. A cutting tool (stylus),

[a]Author to whom correspondence should be addressed. Electronic mail: m.hill@soton.ac.uk

 0001-4966/2008/124(4)/2042/11/$23.00

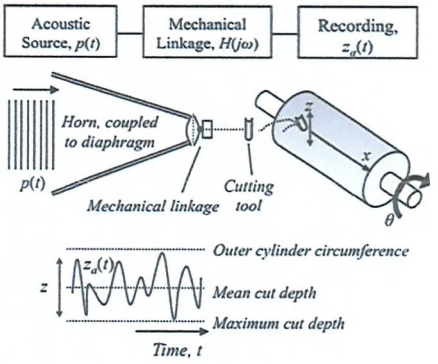

FIG. 1. Overview of the recording process for an acoustically recorded cylinder.

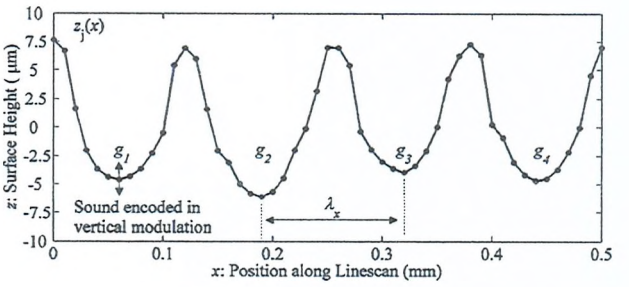

FIG. 3. A typical linescan of length L=0.5 mm, measured from a Blue Amberol cylinder, ca. 1912 (Δx=10 μm). This cylinder was recorded with 200 turns/in. and has a groove spacing λ_x of 127 μm. The *groove seed*, denoted by g_n is the nth groove cross section along the linescan.

which is coupled to a diaphragm-horn arrangement, advances along the cylinder axis (x-axis), via a lead screw with constant pitch, λ_x. Prior to recording, the cutting stylus is forced into the outer cylinder surface, producing an unmodulated cut of fixed radial depth. During recording, incident acoustic pressure variations are channeled via a horn from the surrounding air, toward a thin diaphragm, forcing the stylus to cut undulations that are normal to the surface (in the radial plane) about its mean cut depth. This radial modulation, $z_a(t)$, is often referred to as "hill-and-dale" or "vertical," as opposed to the majority of monophonic flat disk recordings, which are encoded with a "side-to-side" or "lateral" modulation. The encoded groove follows a helix of pitch λ_x, with cross section dependent on the stylus tip geometry. Cylinder recordings were typically recorded with a pitch of 100–200 turns/in. at a rotational speed of 100–160 rpm. For the interested reader, a history of the acoustic era of recorded sound is given by Millard[14] and Read and Welch.[15]

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

B. Cylinder surface measurement and definitions

Figure 2 shows a cylinder recording mounted onto a noncontact measurement system. The methodology, which is typical of the approaches described in literature[9–12] uses a single-point optical sensor to measure the full surface topology of the cylinder recording. A motion system is used to traverse the sensor linearly along the cylinder's axis, forming a "linescan." The cylinder is mounted on a rotatable mandrel, allowing the sensor to map the continuous cylinder surface $z(x,\theta)$ via a succession of linescan measurements. An additional linear stage provides a third axis of motion orthogonal to the linescan axis, which controls the sensor's standoff distance from the artefact to keep the surface within the sensor's gauge range.

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

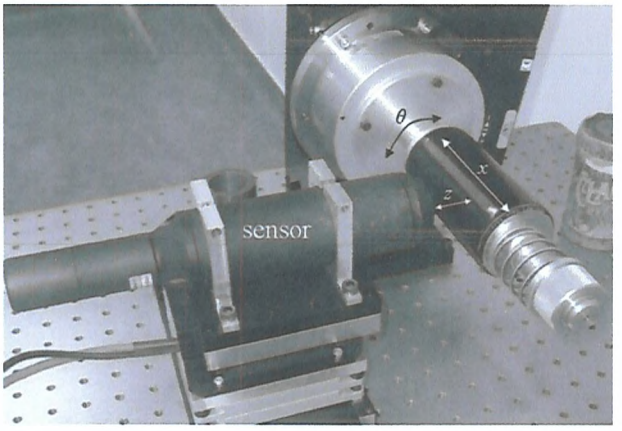

FIG. 2. Cylinder recording mounted on the measurement system, showing the three measurement axes.

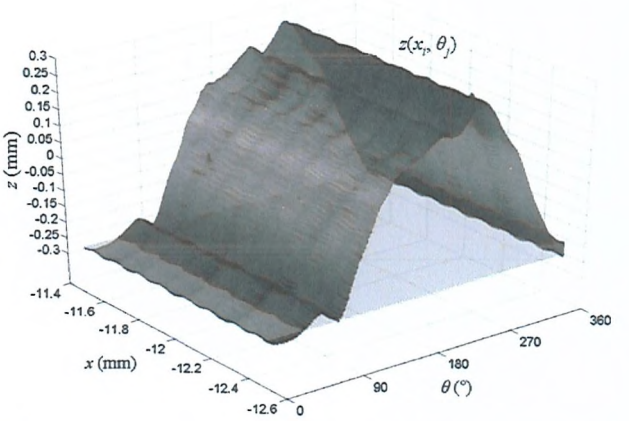

FIG. 4. Surface measurement of a Blue Amberol cylinder surface. The surface is built up through a succession of linescans. The large deviation in z is due to the lack of cylinder concentricity.

C. Resolution requirements for sound reproduction

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

The choice of Δx determines the number of samples over which the groove cross section is resolved along the linescan. This is dependent on the cylinder pitch and complexity of the groove cross section. For a cylinder recorded with 200 turns/in., a linescan grid spacing of $\Delta x = 10\ \mu$m gives approximately 13 samples per groove cross section (see Fig. 3, for example). This sampling has been previously judged to provide sufficient resolution for sound reproduction.[10]

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

$$f_s = \frac{6\omega}{\Delta\theta}. \tag{1}$$

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

$$f_s > 2f_{\max}, \tag{2}$$

where $f_{\max}$ is the maximum frequency limit of the cylinder record. For cylinder recordings, this upper limit is around 5 kHz.[10,16] For 160 rpm cylinders, an angular sampling increment of $\Delta\theta = 0.1°$ gives a playback sample rate of 9.6 kHz, which should be considered as an absolute minimum for accurate sound reproduction. The technical committee of the International Association of Sound and Audiovisual Archives (IASA) recommends a minimum sample rate of 48 kHz at 24 bit word length for digitizing analog recordings.[17] In some cases, a sampling of 96 kHz at 24 bits has been adopted.[17] With the noncontact system, a playback sample rate of 96 kHz is achievable with an angular sampling of $\Delta\theta = 0.01°$ (for cylinders recorded at 160 rpm). The datasets discussed in this paper use a sampling increment of $\Delta\theta = 0.01°$, giving $f_s = 96$ kHz.

The third sampling dimension is the axial resolution of the sensor, Δz, which is the ability of the sensor to resolve displacements in the z-plane. It was shown by Boltryk *et al.*[2] that in order to resolve the smallest displacement amplitudes (high frequency and low amplitude) for typical cylinder recordings, an axial resolution in the order of 10 nm is required. For vertically modulated grooves, Δz determines the dynamic range of the digital audio signal. Current hardware permits a *measurement dynamic range* of ~15 bits (3.5×10^4 discrete z-values) over the sensor's full gauge range. However, in order to relate Δz to the dynamic range of the digital audio signal, the range over which the groove depth modulates in the z-axis must be known. A typical range value for the test cylinder described in this study is 10 μm (for a 1 kHz tone, measured from peak to trough). The current sensor has an axial resolution of $\Delta z = 10$ nm, which in this case gives an equivalent digital audio bit depth of just under 10 bits (1000 discrete z-values).

II. REVIEW OF RELATED TECHNIQUES

A. Sound recovery from full surface mapping of cylinder recordings

Sound reproduction from full surface mapping of cylinder recordings presents a relatively recent area of research. There are currently no commonly used or universally approved methods for audio signal estimation from discrete surface maps of cylinders in literature.

In a previous collaborative study,[10] between Lawrence Berkeley National Laboratory and the University of Southampton, a method was described for extracting sound from a discrete surface map of a Blue Amberol cylinder (ca. 1912), measured via single-point optical displacement sensor, as described in Sec. I B.

In Ref. 10, the radial displacement track encoded by the cutting tool $z_a(t)$ is found by initially estimating the local minimum height (minima) of each groove cross section for all linescans in the discrete surface, $z(x_i, \theta_j)$. Groove minima data are then reorganized into a time series, which follows a helical trajectory. A list of groove valleys (minima) is initially located via a neighborhood search along each linescan, and outlying candidates are removed. A set of points in the kth groove valley is then fitted to a quadratic function, $H_k(x) = A_k x^2 + B_k x + C_k$. The quadratic term ($A_k$) remains fixed and is chosen such that the curvature of the parabola is an approximation to that of the original cutting tool, which produced the recording. The depth (minimum ordinate) of the fitted quadratic curve is then used to form a radial displacement track estimate, $z_a(t)$, corresponding to the encoded sound. The concept of three stand-alone data streams were considered: (1) the groove bottom (valley), (2) the groove top (ridge), and (3) the groove bottom position with respect to the groove top position (top-bottom).

It was noted that the use of the ridge data stream provided audible albeit noisy sound content, but with interference from different times in the recording. This kind of temporal distortion was also noted for playback systems *in situ*[1,2] when the width of the laser beam spot was large compared with the groove width. The use of "top-bottom" subtraction[10] was shown to remove low frequency structures (below 150 Hz) due to surface form.

The audio signal reproduced by Fadeyev *et al.*[10] was compared with a stylus transfer of the same recording. Visual comparisons of the time histories and spectra showed that the optical method provided an "accurate audio transcription," compared with the stylus playback, but no quantitative analysis of the audio signal quality was presented. In this paper, we seek to evaluate signal reproduction of the noncontact system for cylinders using quantitative measures. This allows for different audio signal estimates to be rated and compared numerically without subjectivity.

B. Signal quality testing for noncontact sound reproduction

An audio signal quality analysis was developed by Stotzer[13] for their noncontact reproduction method. This analysis was carried out to determine the feasibility of photographic methods of sound reproduction for flat disk record-

ings and to make comparisons with stylus playback. The SNR and THD are determined through the use of test records encoded with single frequency tones. In this way, the recovered signal is of known type; thus, the SNR and THD are calculated by estimating the ratio of powers in the signal and noise bands.

A summary of discrete-time methods for estimating the signal-to-noise power ratio estimates is given by Jenq.[18] Two algorithms are described for estimating the SNR of a noisy sinusoid from discrete-time data. The first is based on estimating directly the four parameters of a sine wave (amplitude, frequency, phase, and dc offset), and the second is based on a spectrum averaging method via the discrete Fourier transform (DFT).

Estimates for SNR using spectrum averaging techniques are given by Stotzer[13] for various test disks recorded with a 300 Hz tone. The SNR is defined as the ratio between the signal power P_{sig} and noise power P_n for tracks containing a single frequency, f_0. The SNR is expressed in decibels and is given by

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

The signal and noise powers are derived from discrete power spectral estimates of the recovered tonal waveform, where P_{sig} is calculated over a frequency range of $\pm f_{\text{margin}}$ around the center frequency, f_0. (In Ref. 13, $f_{\text{margin}}=5$ Hz.) The noise power P_n is the power in the remaining frequency bins, outside of the range $f_0 \pm f_{\text{margin}}$, from 100 Hz up to 10 kHz (these limits were chosen arbitrarily).

Another method of measuring the quality of sinusoidal signals is to examine the THD. The spectrum of an ideally recovered sinusoid should contain a single peak at the fundamental frequency. In practice, geometric distortion of the sine wave can be introduced at time of recording or at time of measurement. Harmonic distortion can also be introduced by errors in the signal recovery method itself and this effect must be minimized. Stotzer[13] gave the following definition for THD:

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

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

III. METHODOLOGY

A. Proposed methods of audio signal recovery

In previous work,[10] the estimation of stylus trajectory and groove depth was not considered independently but as a single process. For damaged cylinder surfaces, it is useful to visually assess the estimated trajectory before extracting the sound to ensure that the trajectory is coherent (no skipping of grooves occurs) and complete (no regions containing sound are missed). For this reason, we consider sound recovery in two independent stages—trajectory estimation and audio signal estimation—as was previously discussed.[11]

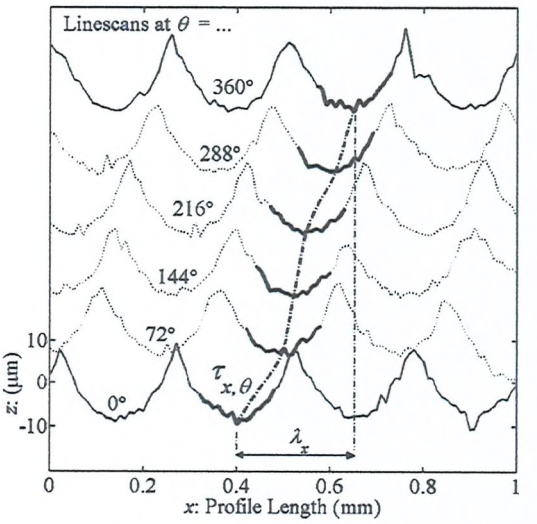

FIG. 5. Linescan measurements taken at six angular intervals from 0° to 360° around the cylinder circumference (linescans have been offset in the z-axis for visualization purposes). The path traced by $\tau_{x,\theta}$ shows the shift along the x-axis of the highlighted groove valley. This shift vector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

1. Trajectory estimation

A method of rapidly approximating the trajectory produced by the cutting tool was previously introduced,[11] by modeling the linescan as a periodic signal and by measuring the relative shift between linescans around the circumference. Reasons for adopting this strategy include the reduction in computation time and the ability to observe and account for any potential phase offsets between adjacent linescans caused by measurement error.

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

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

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

where k_p is the range of possible frequency bins in which k_0 is valid, which for cylinder recordings of the era is between 3.9 and 7.9 mm^{-1} (for 200 and 100 turns/in., respectively).

Figure 5 shows that if we observe a single groove valley at linescan at $\theta=0°$ and track its shift along the x-axis around the cylinder circumference until linescan at $\theta=360°$, the x-position of the valley propagates through a distance of λ_x. For a cylinder recorded with constant pitch, this shift is ap-

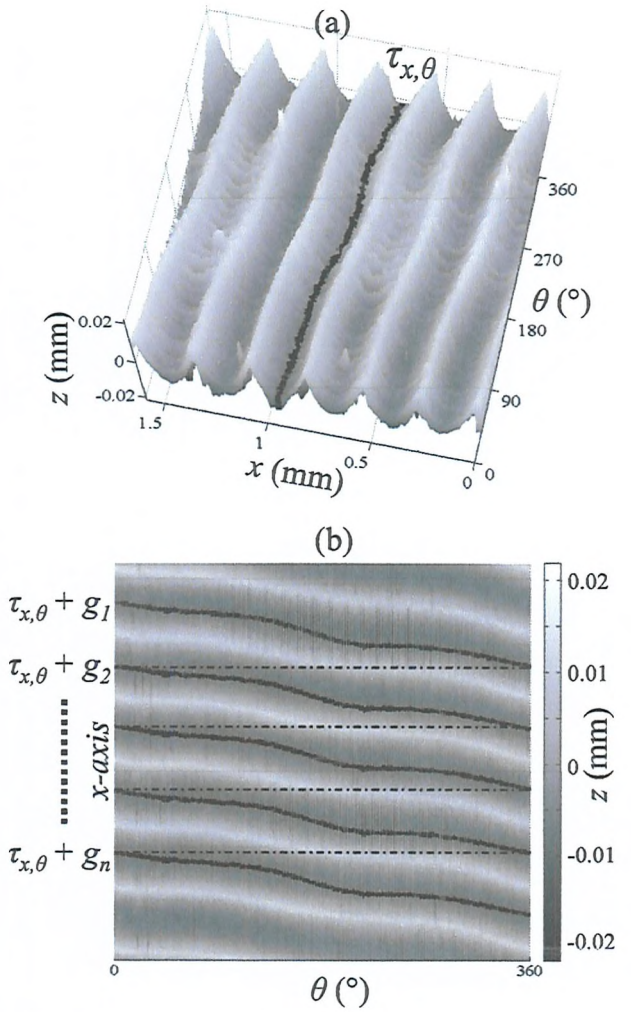

FIG. 6. (a) Scaled discretized shift estimation vector $\tau_{x,\theta}$ overlaid onto a 3D surface plot of the test cylinder data. (b) Formation of stylus trajectory vector $A(x,\theta)$ using the phase shift estimate $\tau_{x,\theta}$ and groove seed positions, $g_1, g_2, \ldots, g_n$.

proximately equal for all grooves. An estimate of this shift vector, $\tau_{x,\theta}$, can therefore be used to form an approximate stylus trajectory.

The shift vector $\tau_{x,\theta}$ can found by first estimating the relative phase (shift) between linescans from 0° to 360°. With k_0 determined from Eq. (6), the phase estimate of the jth linescan is given by

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

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

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

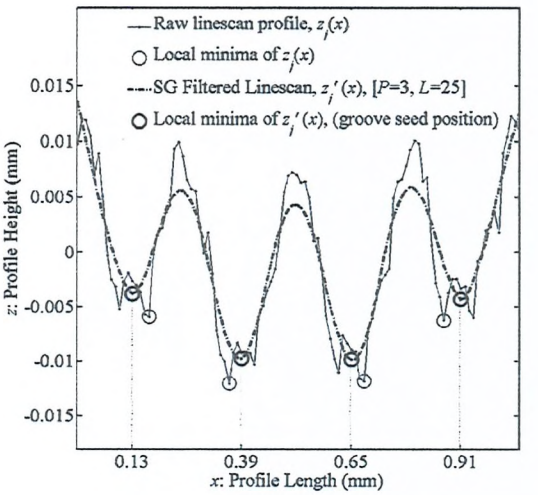

FIG. 7. The abscissa of local groove minima from the raw linescan profile, $z_j(x)$, varies from one groove to the next due to the roughness of the cutting tool. The abscissa of minima from the filtered linescan $z_j'(x)$ can be seen to be spaced equidistantly, by a distance of λ_x=0.254 mm, which is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

This shift vector $\tau_{x,\theta}$ can then be used as a "template" for segmenting the surface into groove regions (groove seed). The nth groove seed, g_n (see Fig. 3), is found by locating the medial axis of each groove valley at 0°. For smooth symmetric groove cross sections, the medial axis is indexed by the local minimum point of the groove and can be found via a simple neighborhood search. In cases where the groove is rough and/or asymmetric (see Fig. 7, for example), the medial axis does not always lie at the local minimum point.

To ensure that each groove is indexed by its medial axis, an appropriately designed Savitzky–Golay (SG) polynomial smoothing filter is applied to the linescan $z_j(x)$ to form the filtered linescan $z_j'(x)$.

SG filters were initially developed to identify the relative widths and heights of spectral peaks in noisy spectrometric data.[19] In this work, the SG filter can be used in a similar way to obtain the position of groove valleys (to the authors' knowledge, this is the first time that SG filters have been applied to aid groove extraction). The SG filter works by a local polynomial regression (of degree P) on a distribution of length L (of at least P+1 equidistant samples) to determine the smoothed value for each point.

The frame length, L of the SG filter is chosen to match the groove pitch, λ_x. In this way, the SG filter operates locally on each groove valley, replacing the rough groove cross section with a smoothed version. Figure 7 shows a linescan filtered with a cubic SG filter (P=3) with frame length L =25 (the approximate number of samples across one groove cross section for 100 turns/in. cylinder/Δx=10 μm). The position of minima in the filtered linescan, $z_j'(x)$, are located closer to the middle (medial axis) of the groove valley than the minima found using the raw linescan, $z_j(x)$. The local minima of this SG filtered linescan are used to locate the groove seed positions.

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

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

where A is of size nJ rows by two columns, n is the number of complete grooves, and J is the number of linescans measured around the full cylinder circumference. The column vectors $[\tau_{x,\theta}+g_1]^T$ and $[\theta]^T$ are both of length J.

2. Audio signal estimation

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

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

(i) single point data streams (using raw z-data, no averaging);
(ii) averaging samples across the groove cross section;[11]
(iii) stylus model, fitting based methods;[10] and
(iv) rule-based methods.

In previous work,[11] the notion of a "groove matrix" was introduced. The groove matrix, denoted by G_t, is a matrix of z-values of size n by m, where n is the number of groove cross sections (equivalent to the number of samples in the time domain) and m is the number of samples across the groove cross section found along the trajectory vector A.

Representing the groove depth by a single discrete data point (i) is not ideal for rough groove cross sections, as the selection of the sample is very sensitive to noise. Scheme (ii) was investigated previously by averaging data streams at the groove bottom and sidewall.[11] The cylinder in question was a molded Blue Amberol cylinder. In this case, the groove shape was symmetric and the abscissa of groove minima remained relatively unchanged from one groove to the next. A comparative examination of three different signal estimates in the time and frequency domains showed that the SNR increased as the estimate approached the bottom of the groove cross section.

The parabola fitting method[10] described in Sec. II A, based on scheme (iii), is well suited for cylinder recordings with a smooth symmetric groove cross section, as shown in Fig. 8(b), where the quadratic fit agrees well with the groove shape. In some cases, however [see Figs. 8(a) and 8(c), for example], the cylinder groove cross section is irregular, asymmetric, and does not resemble the quadratic fit. This irregularity is likely to have been caused due to the roughness of the cutting tool [as with the case of Fig. 8(c)] or as a result of wear due to repeated stylus playback [in the case of Fig. 8(a)].

The cylinder used for test purposes in this study has an irregular groove cross section, which is in contrast to the smooth groove cross section investigated previously.[10,11] The test cylinder, which contains signals of known type, allows for a quantitative assessment of signal quality and a comparison of the two different estimation methods for $z_a(t)$.

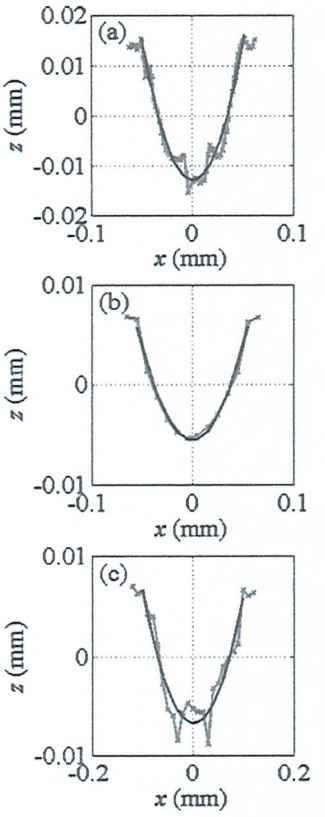

FIG. 8. Three groove valley examples. The smooth dark line in each plot shows a polynomial fit of order 2, which best fits the groove data, in a least squares sense. The Blue Amberol Cylinder (b) has strong agreement with the quadratic curve, whereas the graphophone (a) and test cylinder (c) example show a more complex groove shape, due to the cutting tool that produced the recording, and increased surface roughness or damage.

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

a. Method A: Unweighted mean sample averaging across the groove matrix. A groove matrix G_t of size n by m is formed from the trajectory vector $A=A(x,\theta)$, where n is the number of samples in the time domain and m is the number of samples across the groove cross section.

Radial displacement estimates $z_a(t)$ are then derived from G_t by an unweighted mean averaging process across the nth row of G_t.

The mean average is taken from the central column of the G_t, which corresponds to the medial axis of the groove cross section. The number of points included in the average varies from $N=1$ (central column of the G_t only) to $N=25$ points (± 12 samples either side of the central column of G_t).

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

TABLE I. Description of test cylinder recording and measurement details.

Parameter	Value
Length	100 mm
Diameter	55.56 mm
Recording speed	160 rpm
Recording pitch	100 turng/in.
Recording method	Electrical, direct cut
Recording medium	Wax cylinder
Sampling details	$\Delta x = 10\ \mu$m
	$\Delta\theta = 0.01°$
	$\Delta z = 10$ nm
Playback sample rate	96 kHz

b. Method B: Use of Savitzky–Golay filters of to locate unique groove minima. Raw surface data are filtered along each linescan via SG filter of polynomial order P and frame length L.

Along the trajectory vector $A = A(x, \theta)$, the radial displacement estimate $z_a(t)$ is found by locating the local minimum point of the filtered surface.

Different SG filters of orders $P = \{1, 3, \ldots, 9\}$ and frame length $L = \{11, 13, \ldots, 25\}$ are used to obtain unique groove minima at different positions in the groove cross section.

In Method B, a SG filter is used to firstly smooth the groove cross section in order to account for the increased surface roughness caused by the cutting tool. The local minima of the smoothed cross section are then used as an estimate of the groove depth. Through the use of different combinations of filter polynomial order (P) and filter length (L), an optimal filter can be found by evaluating the SNR and THD of the recovered signal.

The best signal estimates derived by Methods A and B can then be compared in terms of SNR and THD.

B. Method for comparing audio signal quality

1. Details of test cylinder

A wax cylinder recording was produced for purposes of assessing the quality of the recovered signals. The cylinder was recorded at Poppy Records.[20] The recording stylus was not driven by acoustic pressure but was directly cut by an electrical transducer. Pure tones (sinusoids) in the range 200 Hz–5 kHz were recorded onto a cylinder at 160 rpm with a pitch of 100 turns/in. A geometrical description of the test cylinder is given in Table I.

Following the recording process, the groove structure remained untouched by stylus or cleaning. The surface map produced by the optical scanning system used a measurement grid of [$\Delta x = 10\ \mu$m, $\Delta\theta = 0.01°$]. This sampling scheme gave approximately 25 data points per groove cross section and a playback sample rate $f_s = 96$ kHz. The recovered signals exhibited some frequency modulation and harmonic distortion, which is in part due to the nature of the recording process. The degree of noise and distortion introduced at the time of recording is unknown. To this end, the experiment does not give an "absolute" measure of SNR/THD, but is useful for a comparative study of different audio signal estimation methods and comparisons with stylus replay. Recovered signals from the 1 kHz tone region provide the basis for comparison of audio signal estimates.

2. SNR analysis

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

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

Frequencies that contribute to the noise power P_n band were considered in two ranges: (1) below 1 kHz: 100–950 Hz and (2) above 1 kHz: 1050 Hz–10 kHz.

Harmonic frequencies above the fundamental were considered as undesirable and were also included in the noise band calculation. Frequencies below 100 Hz were not included in the noise band calculation, as these were attributed to the surface form and not to the audio signal.

3. Total harmonic distortion analysis

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

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

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

IV. RESULTS

A. Signal quality comparisons for Methods A and B

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

Figure 9(a) shows that the SNR increases as the number of points in the average window, N, increases, but care must be taken when interpreting this trend. The best SNR for Method A comes from a 21 point mean average ($N = 21$); however, this signal did not have the lowest THD [see Fig.

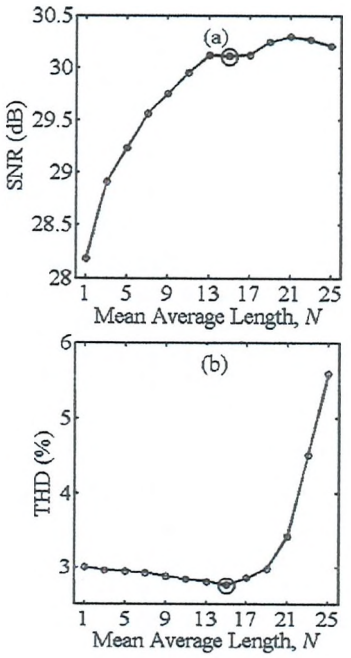

FIG. 9. SNR (a) and THD (b) results for Method A (mean averaging across the groove matrix). Larger circles indicate the optimal points on the curve.

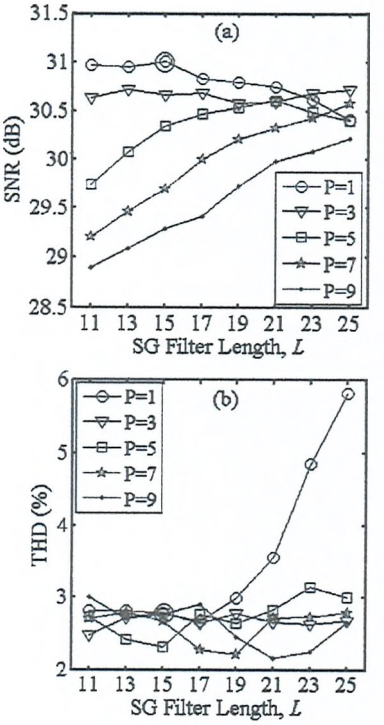

FIG. 10. SNR (a) and THD (b) results for Method B (SG filtering). Larger circles indicate the optimal points on the curve.

9(b)], and upon listening to the signal, a small echo effect was noticeable. This kind of temporal distortion is not accounted for in the spectrum averaging technique and is discussed in Sec. IV B. The signal that had the lowest THD (and hence the "purest" sinusoid) was from a mean average of $N=15$.

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

SNR and THD results for Method B for the different SG filters are shown in Fig. 10.

In general, the SNR decreased as the filter order P increased. This is due to the fact that higher order polynomial filters retain the higher spatial frequency content (roughness) of the groove cross section (see Fig. 11), which results in increased high frequency noise for the audio signal. The highest SNR came from the filter $[P=1, L=15]$. A SG filter of order $P=1$ is essentially an unweighted moving average filter, similar to the mean averaging in Method A. This similarity can be observed by noticing that the THD curves for Methods A and B ($P=1$) have similar trends. An echo effect was also noticeable for $[P=1, L=19-25]$, as was found with Method A for large N.

With the exception of the first order filter ($P=1$), the THD scores remained relatively constant as L increased. The lowest THD came from the higher order filters ($[P=9, L=21]$, for example). This may be due to the fact that the unique groove minima produced by higher order SG filters tended toward the positions of raw groove minima, as opposed to being positioned higher up in the groove cross section, where raw displacement data do not exist (as can be seen in Fig. 11).

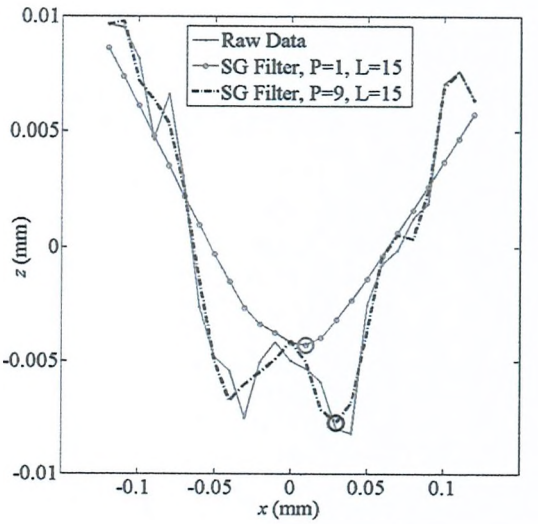

FIG. 11. First and ninth order SG filters ($L=15$) applied to the groove cross section of test cylinder data. The ninth order filter retains more of the higher spatial frequencies, whereas the first order filter produces a smoother cross section. Large circles indicate the position of unique groove minima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

TABLE II. Summary of best SNR and THD results for Methods A and B.

	Method A (Mean averaging)	Method B (SG filtering)
Highest SNR (dB)	30.3	31.0
Lowest THD (%)	2.78	2.16

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

B. Sources of noise and distortion

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

1. Low frequency noise

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

2. Impulsive noise

Impulsive noise can be caused by surface debris, which was not cleaned from the cylinder prior to scanning. Any debris that lies above the imaging plane will therefore be measured by the sensor. This can result in transients for the recovered audio signal, as can be seen by the vertical bands in the spectrograms in Figs. 12 and 13. In order to minimize this effect, any loose debris should be removed from the surface prior to scanning wherever possible. Additionally, if the cylinder surface is measured in piecewise segments (as is required for heavily warped surfaces,[12] and the test cylinder in this case), impulsive noise can be introduced at segment joints if the segments are not appropriately aligned in software prior to sound extraction.

3. Temporal distortion

It is observed that displacement estimates that used groove cross section data toward the tops of the groove exhibited an echo. This phenomenon was first noted by Iwai *et al.*[1] for real-time optical playback of phonograph cylinders, when the width of their laser beam was wide compared with the groove width. Similarly, Fadeyev *et al.*[10] noted that groove ridge data provided audible albeit noisy sound content, but with interference from different times in the recording. The test cylinder data allow us to quantify the interference observed from the ridge data stream. Figure 12 shows a time-frequency analysis (spectrogram) of data streams derived from the top and bottom of the groove cross section. It

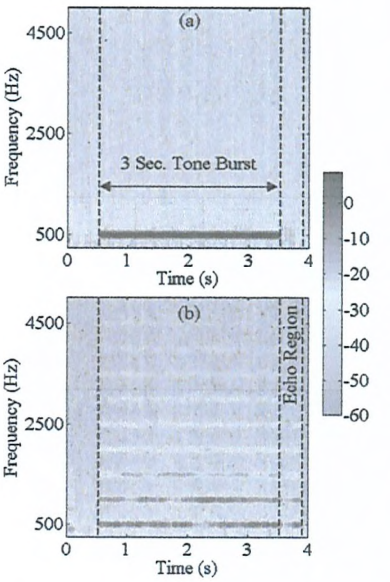

FIG. 12. Spectrogram analysis of signals recovered from the bottom (a) and top (b) of the groove cross section for the 500 Hz tone. An echo region is clearly visible in the ridge spectrogram (b) due to the superposition of adjacent groove displacement data at times t and $t+T$. In addition, the ridge data stream has more harmonic distortion.

appears that an echo is introduced due to the superposition of adjacent grooves at times t and $t+T$, where T is the time taken for one complete rotation (0.375 s in this case for 160 rpm). To avoid introducing temporal distortion, it is therefore suggested that the use of displacement data from the groove ridge should be avoided.

C. Comparison of stylus and noncontact reproduction

In addition to optical reproduction, the test cylinder audio was also transferred using conventional stylus techniques

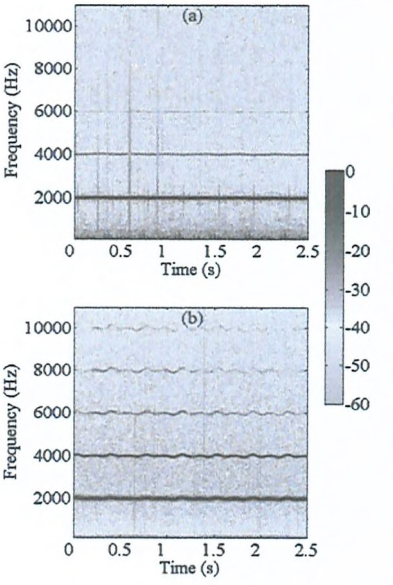

FIG. 13. Spectrogram comparisons for noncontact (a) and stylus reproduction (b) of 2 kHz tone. Harmonic distortion is greater with stylus reproduction. The noncontact method (a) shows less harmonic distortion than in (b) but more low frequency (below 500 Hz) and impulsive noise. In addition, frequency modulation is less apparent for the noncontact method.

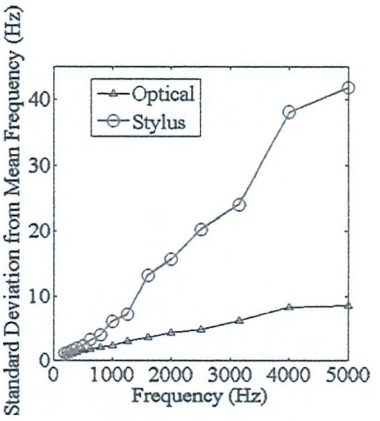

FIG. 14. Comparison of measured frequency modulation of tones recovered from the test cylinder via noncontact and stylus methods.

at the British Library Sound Archive. Due to difficulties in matching the equalization of the stylus and optical reproduction, direct signal quality comparisons are not presented in this paper. Subjective listening suggests that stylus reproduction has a higher SNR, in terms of a lower noise floor power and less impulsive noise (the increased impulsive noise found with optical reproduction is likely due to measurement of remaining debris above the imaging plane, as described in Sec. IV B 2). However, the accuracy of sine wave reproduction in terms of harmonic distortion and frequency modulation appears to be greater with the noncontact method. Examples of the recovered signals from optical and stylus methods can be found here.[21]

1. Harmonic distortion

The spectrogram in Fig. 13 compares the reproduction of a 2 kHz tone from both stylus and optical methods. Stylus reproduction appears to show increased harmonic distortion compared with the optical signal. This is likely due to the stylus-groove relationship, known in literature,[22,23] as "tracing distortion," which arises from the fact that the reproducing stylus tip is of finite size. When the stylus tip traces over a sinusoidal groove, a mismatch in curvature between the tip radius and the groove contour means that the reproduced signal is not sinusoidal.[22] This geometric effect results in harmonic distortion. The noncontact method of sound reproduction is not affected by tracing distortion.

2. Frequency modulation

The degree of frequency modulation was also compared by locating the position of the peak frequency in the DFT for windowed segments of recovered tones. The standard deviation in measured frequency from peaks in the DFT was used as an indication of frequency modulation. Figure 14 shows frequency modulation for all tones reproduced from the test cylinder by optical and stylus methods. At 5 kHz, the frequency modulation with stylus reproduction is approximately ± 40 Hz, and for optical reproduction is less than ± 10 Hz. The increased frequency modulation with stylus reproduction is likely due to the poor compliance of the cylinder bore with the rotating mandrel (due to lack of cylinder concentricity). If mounted incorrectly, the surface speed of the cylinder will vary, meaning that for a constant playback sample rate a modulation in frequency occurs. In this particular case, it appears that the stylus reproduction is more heavily affected by speed fluctuations, than with the noncontact method.

V. CONCLUSIONS

The noncontact full surface mapping system, which has been developed, allows for accurate sound reproduction from cylinder recordings. The advantage of mapping the full surface topology (as opposed to optical playback *in situ*) is that numerous estimates for the audio signal can be derived from a single data set in postprocessing. To the authors' knowledge, this study represents the first quantitative analysis of test signal reproduction using the noncontact surface mapping method.

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

The test cylinder encoded with sinusoids provides a basis for quantitative assessment of sound reproduction quality. Two different methods of estimating the radial displacement track from a discrete surface have been shown to produce variable audio signal quality, in terms of SNR and THD. The use of appropriately designed SG filters (Method B) to produce an estimate for the groove depth provided a better audio signal estimate signal in terms of SNR and THD compared with unweighted mean averaging across the groove matrix (Method A).

Audio signal estimation from discrete surface maps is not restricted to the two methods described in this paper. With stylus reproduction, the use of different styli are chosen, based on the groove geometry and surface condition of the recording. The same principle applies to the noncontact method, in that the "optimal" method of signal reproduction is likely to be dependent on the cylinder under test. To improve further the quality of the recovered signal, other rule- or model-based navigation methods may be used to avoid the introduction of outliers (caused by debris, etc.).

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

The audio signal recovered via stylus reproduction exhibited more frequency modulation and harmonic distortion compared with the optical method. Although this is a feature of this particular transfer (and not necessarily inherent for all stylus playback), it highlights some advantages of noncontact reproduction. For example, geometric effects of the

stylus-groove interaction, such as tracing distortion, do not affect optical reproduction. Additionally, in this particular case, the noncontact method appears to be less affected by speed irregularities, found with the stylus reproduction.

Difficulties encountered from very early wax cylinders include poor surface condition, such as cracking, mold growth, and wear caused to the groove structure through repeated stylus playback or poor storage. In cases where a cylinder is cracked, it is possible to scan the broken segments and then reconstruct the surface. In cases where the groove structure is damaged, it is possible to track more closely in postprocessing than with playback *in situ*.

Topics for further investigation in this field include the development of signal reproduction methods for cylinders in poor surface condition, including broken or shallow grooved recordings. In addition, a reduction in measurement time remains a priority for the data acquisition process. The use of other specially produced test cylinders encoded with known waveforms may also provide further scientific analysis of signal reproduction from the noncontact full surface mapping method.

ACKNOWLEDGMENTS

The authors would like to thank The British Library Sound Archive for assisting with the stylus transfer and The Engineering and Physical Sciences Research Council for funding this research.

[1]T. Iwai, T. Asakura, T. Ifukube, and T. Kawashima, "Reproduction of sound from old wax phonograph cylinders using the laser-beam reflection method," Appl. Opt. **25**, 597–604 (1986).

[2]T. Ifukube, T. Kawashima, and T. Asakura., "New methods of sound reproduction from old wax phonograph cylinders," J. Acoust. Soc. Am. **85**, 1759–1766 (1989).

[3]T. Asakura and J. Uozumi, "Optical methods for reproducing sounds from old phonograph records," *International Trends in Optics and Photonics*, ICO Vol. IV (Springer-Verlag, Berlin, Germany, 1999), pp. 65–81.

[4]S. M. Shanoylo, I. V. Kosyak, V. Petrov, and A. A. Kryuchin, "Reading and processing of audio information reproduced from Edison phonograph cylinders by method of laser interferometry," Proc. SPIE **4402**, 194–201 (2001).

[5]S. S. Cavaglieri, O. Johnsen, and F. Bapst, "Optical retrieval and storage of analog sound recordings," in AES 20th International Conference: "Archiving, Restoration, and New Methods of Recording," 2001.

[6]S. Stotzer, O. Johnsen, F. Bapst, C. Milan, and R. Ingold, "Phonographic sound extraction using photography and signal processing," Digit. Signal Process. **17**, 433–450 (2007).

[7]S. Stotzer, O. Johnsen, F. Bapst, and C. Sudan, "Phonographic sound extraction using image and signal processing," *Acoustics, Speech, and Signal Processing*, in Proceedings ICASSP'04, 2004, Vol. **4**, pp. 289–292.

[8]V. Fadeyev and C. Haber, "Reconstruction of mechanically recorded sound by image processing," J. Audio Eng. Soc. **51**, 1172–1185 (2003).

[9]V. Fadeyev, C. Haber, Z. Radding, C. Maul, J. W. McBride, and M. Golden, "Reconstruction of mechanically recorded sound by image processing," J. Acoust. Soc. Am. **115**, 2494 (2004).

[10]V. Fadeyev, C. Haber, C. Maul, J. W. McBride, and M. Golden, "Reconstruction of recorded sound from an Edison cylinder using three-dimensional non-contact optical surface metrology," J. Audio Eng. Soc. **53**, 485–508 (2005).

[11]A. J. Nascè, J. W. McBride, M. Hill, and P. J. Boltryk, "Signal processing methods for the recovery of audio from early acoustic cylinder recordings, measured via non-contact optical sensor," in AES 31st International Conference: New Directions in High Resolution Audio, Gainesville, Florida, 2007.

[12]P. J. Boltryk, M. Hill, J. W. McBride, and A. J. Nascè, "A comparison of precision optical displacement sensors for the 3D measurement of complex surface profiles," Sens. Actuators, A **142**, 2–11 (2008).

[13]S. Stotzer, "Phonographic record sound extraction by image processing," Ph.D. thesis, University of Fribourg, Switzerland.

[14]A. Millard, *America on Record: A History of Recorded Sound* (Cambridge University Press, Cambridge, 1995).

[15]O. Read and W. Welch, *From Tin Foil to Stereo: Evolution of the Phonograph* (Howard W. Sams & Co., Inc., Indianapolis, IN, 1977).

[16]H. Meulengracht-Madsen, "On the transcription of old phonograph wax records," J. Audio Eng. Soc. **24**, 27–32 (1976).

[17]IASA TC-03, "The safeguarding of the audio heritage: Ethics, principles and preservation strategy," 2005.

[18]Y.-C. Jenq, "Discrete-time method for signal-to-noise power ratio measurement," IEEE Trans. Instrum. Meas. **45**, 431–434 (1996).

[19]A. Savitzky and M. J. E. Golay, "Smoothing and differentiation of data by simplified least squares procedures," Anal. Chem. **36**, 1629–1639 (1964).

[20]Poppy Records, http://www.poppyrecords.co.uk/ (Last viewed 8/23/2008).

[21]The Sound Archive Project Website, "Examples of signal reproduction from test cylinder recordings," http://www.sesnet.soton.ac.uk/archivesound/audio/testcylinder/ Last viewed 8/23/2008.

[22]J. A. Pierce and F. V. Hunt, "On distortion in sound reproduction from phonograph records," J. Acoust. Soc. Am. **10**, 14–28 (1938).

[23]W. D. Lewis and F. V. Hunt, "A theory of tracing distortion in sound reproduction from phonograph records," J. Acoust. Soc. Am. **12**, 348–365 (1941).

SIGNAL PROCESSING METHODS FOR THE RECOVERY OF AUDIO FROM EARLY ACOUSTIC CYLINDER RECORDINGS, MEASURED VIA NON-CONTACT OPTICAL SENSOR

A. J NASCÈ[1], J. W. MCBRIDE[1], M. HILL[1], P. J BOLTRYK[1]

[1] *Electromechanical Research Group, University of Southampton, UK*
antony.nasce@soton.ac.uk

A non-contact method for the recovery of audio signals from early acoustic cylinder recordings is presented. The cylinder surface is scanned via optical displacement sensor, capable of submicron axial resolution. Sound recovery is achieved by estimating the trajectory of a playback stylus over the measured surface. The processing methods required to extract audio from a discrete height map are described. The signal to noise ratio of the extracted signal as a function of position across the groove cross-section is examined, for a Blue Amberol cylinder, (circa 1912).

INTRODUCTION

The development of non-contact, optical methods for the reproduction of audio from early mechanical sound carriers has been subject to some research in recent years, (see for example [1-4]). A non-contact method ensures that no further damage or pressure is applied to the fragile groove structure, wherein the audio signal was originally encoded. High resolution mapping of the surface also has the potential for sound recovery from recordings which are broken, warped or otherwise unplayable via stylus methods. Such techniques have significant potential for sound archivists dealing with early acoustic recordings, who are aiming to maintain quality throughout the playback chain.

The philosophy behind mapping the full recorded surface (as opposed to real time playback) is aimed at providing sound archives with digital preservation copies of early mechanical sound carriers, and also to provide public access to early and rare sound collections.

In this paper the methodology and resolution requirements for the full surface mapping of an acoustic cylinder recording are described. The post-processing stages required for recovering the audio signal are then presented. These methods are based on parameter estimation and spectral examination of the scanned surface, via the Discrete Fourier Transform. An example of different signal estimates derived from the scanned surface, are compared and strategies for attaining the best signal to noise ratio, prior to filtering and equalisation are discussed.

1.1 Acoustic Cylinder Recordings

Cylinder recordings represent the earliest mechanical means for archiving and playing back sound. In recording mode, the acoustic source is usually channelled via a horn towards a diaphragm that responds to the fluctuating acoustic pressure, $p(t)$. The diaphragm in turn moves the cutting stylus, producing radial (hill-and-dale) indentations on the cylinder surface. Upon playback, the same mechanism takes place, but in reverse, so the movement of the playback stylus produces a movement of the diaphragm, creating a fluctuating acoustic pressure that is representative of the recorded signal.

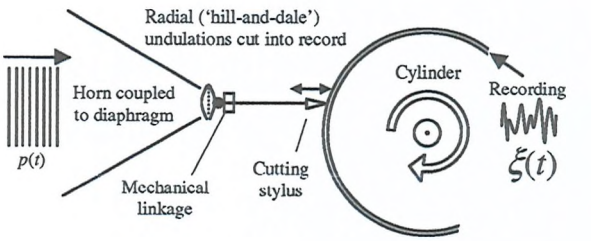

Figure 1: Simplified model of the recording process for an acoustic cylinder recording

The aim of non-contact audio extraction is to obtain a signal which is representative of the acoustic signal that would be reproduced on playback from the physical indentations on the cylinder. Ideally, this reproduced signal should have a flat frequency response. In practice however, horn resonances and the mechanical impedance of the recording apparatus mean that the recovered signal will exhibit irregular spectral characteristics.

Over a limited frequency range (~150Hz - 4kHz) acoustic recordings approximate to a constant-velocity characteristic [5].

Through the scanning process, the radial displacement $\xi(t)$ produced by the cutting stylus can be estimated. Hence, to obtain a signal which is proportional to the acoustic pressure recorded, the first derivative of this vector should be taken to obtain a signal which is equivalent to the radial velocity of the stylus. For a continuous harmonic signal of the form: $\xi = \xi_p e^{j(\omega t - kx)}$, the first time derivative can be obtained, simply by multiplying by $j\omega$. For discrete signals, the limit over which the numerical gradient at any one point by definition, cannot be infinitesimal. Different numerical differentiation methods (e.g. forward/backward/central difference), will introduce different frequency responses to the resultant signal. It is therefore advised that some discrete filtering operation, incorporating both the differentiation stage and desired equalisation is applied to ξ. This approach has been implemented in previous work [2], via FFT filtering.

Due to the complex transfer function of the horn, mechanical linkage, and cylinder surface roughness, the required equalisation is likely to vary from cylinder to cylinder. It is therefore suggested here that any final equalisation is left to the expertise of the transfer engineer or sound archivist. Regardless of which filtering method is used (prior to public access), the most important stage in the playback chain following surface measurement, is to obtain the best estimate for the displacement track, $\xi(t)$.

2 NON-CONTACT SURFACE MEASUREMENT

The cylinder surface is mapped via a single-point, displacement measuring sensor, coupled to a precision linear stage system. The sensor traverses linearly, parallel to the cylinder axis, to measure a *linescan*. The linescan, denoted by $S_n(x)$, is a linear scan of points, measured along the cylinder axis, at some fixed azimuth, θ (see Fig. 2). The cylinder, mounted onto a mandrel, (which is coupled to a rotary stage system) is rotated, and linescans are recorded in this way until the full the surface is mapped over the full circumference. A typical surface height map from a Blue Amberol cylinder is shown in Fig 3. More details regarding the sensor technology and acquisition process can be found in: [6]

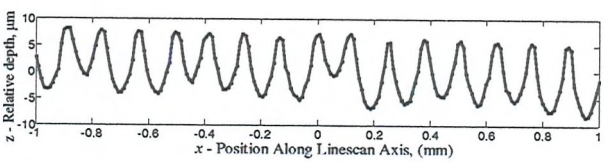

Figure 2: 2mm segment of linear scan from a Blue Amberol cylinder, circa. 1912 (Δx = 10µm).

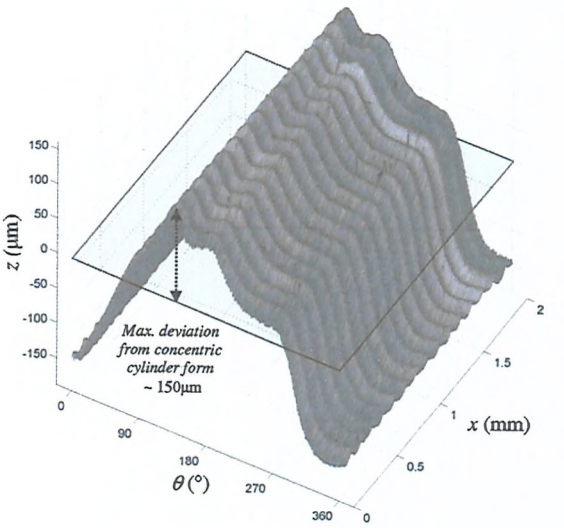

Figure 3: 3-D surface plot of 2mm segment from a Blue Amberol cylinder. This particular cylinder is non-concentric (a perfect cylinder would assume a flat form at a fixed z position). The implication for the extracted audio signal is a low frequency modulation at 2.67Hz, (this is the rotational frequency for 160rpm recordings).

3 SAMPLING AND RESOLUTION

The sensor acquires displacement measurements over a discrete sampling grid, with Δx, corresponding to the increment of the linear stages parallel to the cylinder axis and $\Delta\theta$, the angular sampling increment of the rotary stage. These parameters are now discussed.

3.1 Linescan Sampling, Δx

The choice of Δx is made so that the number of points across the linescan profile sufficiently characterises the groove shape. Strictly, to avoid aliasing of the groove structure, the following equality must be satisfied:

$$2\Delta x < \lambda_x \qquad (1)$$

where λ_x is the spacing between adjacent grooves, (For 200tpi cylinder, λ_x = 127 µm). In practice, the choice of Δx is much lower than (1) suggests, (typically Δx = 10µm), in order to resolve any complex damage structures, which may exist. The linear stages allow for higher resolution linescan sampling, if required.

3.2 Circumferential Sampling, $\Delta\theta$

The sampling grid dimension corresponding to the playback direction is the circumferential axis, θ, thus the choice of $\Delta\theta$ determines the playback sampling

rate, f_s, of the recovered sound. To avoid aliasing effects:

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

Where f_{max} is the highest audible frequency contained within the audio signal. For early acoustic recordings, f_{max} is ~ 5kHz [7], therefore f_s should be at least 10kHz. The rotary stage allows for angular incremental sampling of 0.01° (and higher), which means that 36,000 linescan measurements are recorded over the full circumference. For a record with playback speed of 160rpm, this equates to f_s = 96 kHz.

3.3 Axial Resolution, Δz

The third sampling dimension which is of interest is the axial resolution Δz, of the sensor. This is the ability of the sensor to be able to resolve the surface undulations of the recorded grooves. Previous studies [6] have shown that an axial resolution of 10nm (1x 10^{-8} m) is required for resolving the smallest of recorded groove modulations.

In relation to the extracted audio signal, the choice of Δz is analogous to the dynamic range (or 'Bit resolution') of a digital audio signal. The dynamic range of the measurement system can be estimated by dividing the gauge range (the range of surface heights over which the sensor will operate) by Δz. Current hardware permits a measurement dynamic range of ~15-Bit (3.5 x 10^4 discrete displacement values), although this value may be improved upon in future, by incorporating an adaptive z-axis control system, to track the surface more closely.

A summary of the discussed sampling requirements are given in Table 1.

Sampling Parameters	Recommended Values
Δx : Linescan Sampling increment	10μm (1 x 10^{-5} m)
Δθ: Circumferential Sampling increment	0.01 ° (f_s = 96 kHz)
Δz : Axial Resolution,	10nm (1 x 10^{-8} m)

Table 1: Summary of recommended resolution requirements for 160rpm, 200tpi acoustic cylinder recordings.

4 SURFACE MAPPING

One issue that arises from the measurement process is a discontinuity between linescans s_N and s_1, (see Fig. 4). The helical audio path of the cylinder recording is no longer a continuous trajectory in the matrix representation of surface heights. Instead, the audio path must now be considered as a set of trajectories, which propagate in some way from 0 - 360°. These trajectories, when appropriately ordered and concatenated, form the complete path produced by the cutting stylus. Figure 5 demonstrates the continuous and discrete audio paths.

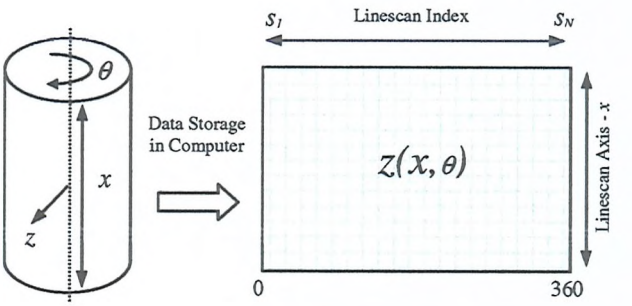

Figure 4: Mapping of the cylinder surface to a 2-D matrix representation of surface heights.

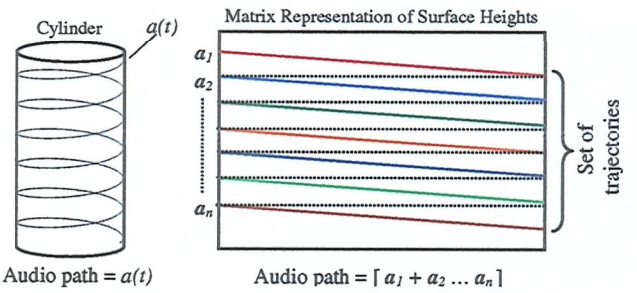

Figure 5: Continuous (left) and discrete (right) audio paths for cylinder and matrix representation of cylinder surface.

The sensor captures the 3-D surface with respect to a 2-D image plane. Any objects, (e.g. cotton fibres, grit, dust), which lie above the image plane will obstruct the underlying groove structure beneath (see fig. 6 for example). The effect of this for the extracted audio signal will be heard as an impulsive click.

It is therefore advisable that prior to measurement, any loose debris lying on top of the grooves is removed via compressed air gun (or otherwise), to ensure that only

the groove structure and originally recorded sound is measured by the sensor.

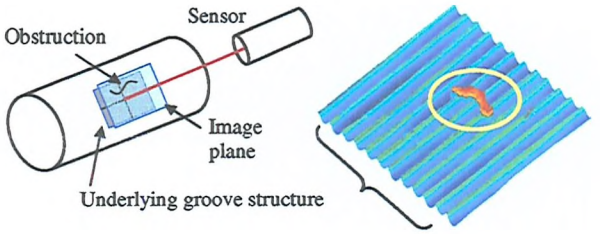

Figure 6: Debris lying upon the cylinder surface is obstructing four of the underlying grooves.

5 AUDIO EXTRACTION

After the surface has been scanned, the audio extraction stage is broadly defined in two stages:

i) **Trajectory Estimation** – locating the audio path produced by the recording stylus, in order to form a coherent time series.
ii) **Audio Signal Estimation** – using surface height data located at the trajectory found in (i) to estimate the raw displacement signal, encoded by the recording stylus, $\xi(t)$.

In previous work, [2], these stages have been accomplished by detecting the locations of ridges and valleys (maxima and minima) for each linescan. The groove bottoms are then fitted to a constrained quadratic function, with curvature that agrees well with the recording stylus geometry. Data is then re-organized into a time series which follows the spiral trajectory.

In this paper, a method is described which relies on sinusoidal modelling of the linescan, and examination of the relative phase shift between adjacent linescans, in order to produce a trajectory estimate. Reasons for adopting this method for trajectory estimation include reduction in processing time and the ability to align angular scans (measured at different time instances), which may contain phase errors between consecutive scans.

5.1 Linescan Parameter Estimation

In order to estimate the path produced by the cutting stylus and hence extract sound from the surface, some parameters of the groove profile must first be identified. For cylinder recordings, the number of turns per inch (tpi) dictates the spacing between adjacent groove valleys. In most cases, the groove spacing, λ_x can be derived from historical records, but for completeness and records of unknown type, it is calculated numerically via spectral analysis of the linescan.

Due to the nature of the recording process and constant pitch of the grooves, the periodic structure of the linescan can be modelled as a single sinusoid:

$$x[n] = A\cos(2\pi k_0 n + \phi) \qquad (3)$$

The amplitude A of the model gives an approximation of the peak to trough distance of the linescan and can be used as a threshold for outlier detection.

The fundamental spatial frequency k_0 is related to λ_x via the reciprocal relationship of period and frequency. For cylinder recordings of the era, with 100-200tpi, k_0 is defined over the range:

$$3.9\text{mm}^{-1} < k_0 < 7.9\text{mm}^{-1}. \quad (4)$$

The maximum likelihood estimate for the frequency of a single sinusoid in Gaussian noise can be found by locating the frequency at which the Fourier transform of the signal attains its maximum [8]. For a linescan sequence $s(n)$ of N samples, the Discrete Fourier Transform (DFT) is defined as the complex sequence $S(k)$ with N uniformly spaced frequency samples:

$$S(k) = \sum_{n=0}^{N-1} s(n) e^{-\frac{j2\pi kn}{N}}, \quad k = 0, \ldots, N\text{-}1 \qquad (5)$$

The amplitude spectral density $|S(k)|$, is given by the magnitude of the DFT. Figure 7 shows the linescan and sinusoid model profiles, and their associated amplitude spectral density estimates.

The phase angle of the DFT, $\phi(k)$, is given by:

$$\phi(k) = \tan^{-1}[\text{Im}\{S(k)\} / \text{Re}\{S(k)\}] \qquad (6)$$

The phase $\phi(k)$ of the sinusoid model tells us by how much the groove profile is 'shifted' along the cylinder axis, and is valid from $-\pi$ to π. For a perfectly concentric cylinder, recorded at constant angular velocity, the expected audio path is a helix. The phase angle of our groove model would therefore be expected to vary linearly around the cylinder circumference from 0 to 2π, through a distance λ_x. In reality, the shift of the measured grooves is non-linear. Examining the phase angle of $S(k)$ at k_0 gives a useful method for estimating the average trajectory made by the cutting stylus. This procedure also aids the process of stitching data sets

together, where the correct alignment of adjacent linescans is important.

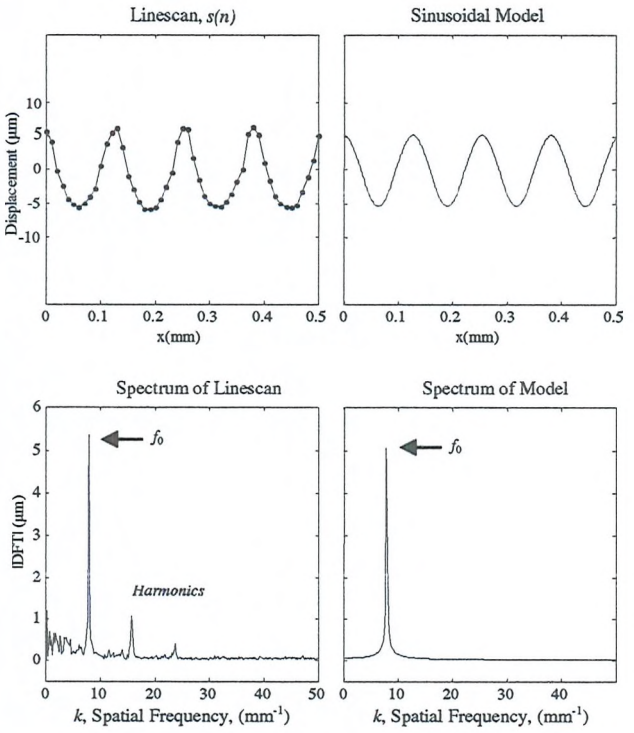

Figure 7: Comparison of profiles and spectra for linescan and signal model. The DFT of the linescan shows a spectral peak at 7.9mm^{-1} which is consistent with the expected value for a 200tpi recording. The sinusoid model in this has: $A = 5 \times 10^{-3}$mm, $k_0 =$ 7.9mm^{-1}, $\phi = 0$.

5.2 Trajectory Estimation

A phase shift vector Φ is formed by extracting the phase angle $\phi(k_0)$, for all linescans around the cylinder circumference. If the phase variation between successive linescans exceeds 2π, a discontinuity occurs. The procedure for constructing a continuous, natural phase variation is known as phase-unwrapping and is accomplished by adding $\pm 2\pi$ to Φ, when absolute jumps between consecutive elements of Φ are greater than or equal to π. Discontinuous and continuous phase estimates from an Edison cylinder scan are shown in fig. 8.

Following phase unwrapping, Φ is scaled in such a way that a phase shift of 2π is equal to the distance λ_x. The vector is then rounded to the nearest integer value, so that the discrete surface data can be referenced. Figure 9 shows the scaled, discretised prototype trajectory estimate $\mathbf{P}_\Phi$, overlaid onto the surface.

If the trajectory of the grooves is seen to vary substantially from one length of the cylinder axis to the other, (e.g. one end of the cylinder surface is more heavily deformed than the other), then the prototype estimate can be updated along the cylinder axis, by examining the phase shift as before, but with windowed linescan segments. In practice, from surfaces scanned thus far, a single prototype is sufficient for segmenting individual grooves over the full cylinder length.

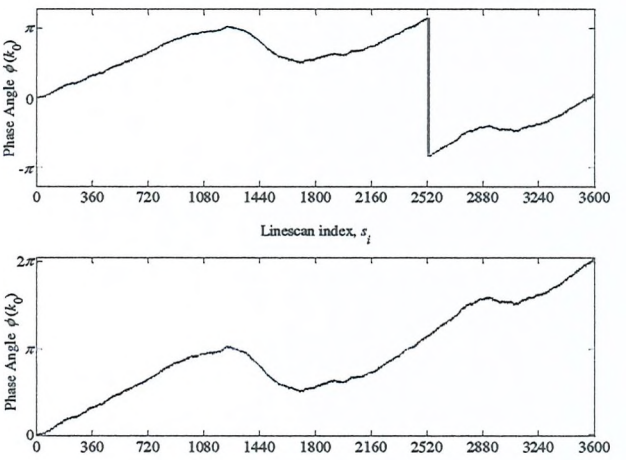

Figure 8: Trajectory estimation via examination of the phase angle at frequency k_0, for all linescans around the cylinder circumference. Discontinuous phase estimate (top) and continuous, unwrapped phase estimate (bottom).

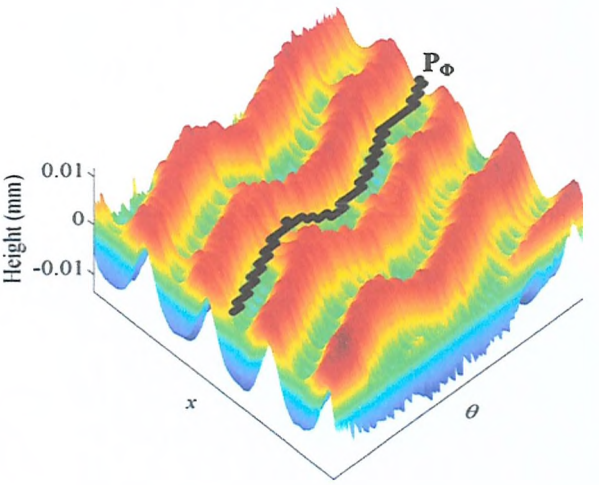

Figure 9: Scaled trajectory estimation vector, overlaid onto 3-D surface plot.

It should be noted that the trajectory estimate $\mathbf{P}_\Phi$, does not produce the final audio signal directly, but is used as a tool for segmenting the individual grooves. In order to create a coherent time series, the groove "seed", (i.e. the position along the linescan at which a groove exists at 0°) must first be established. This is achieved by detecting local minimum regions of the grooved surface, at 0°.

$\mathbf{P}_\Phi$ is then used as a template for segmenting the surface into groove regions and provides a coarse approximation for the position of our displacement track, ξ. The prototype is positioned at each seed location and a new groove matrix G_t, is formed from neighbouring points which surround $\mathbf{P}_\Phi$. The neighbourhood is set to $\lambda_x / 2$, in order to avoid locating points belonging to adjacent grooves (see fig. 10).

The resultant G_t matrix is composed of rows that correspond to the groove cross-section, produced by the cutting stylus at time t.

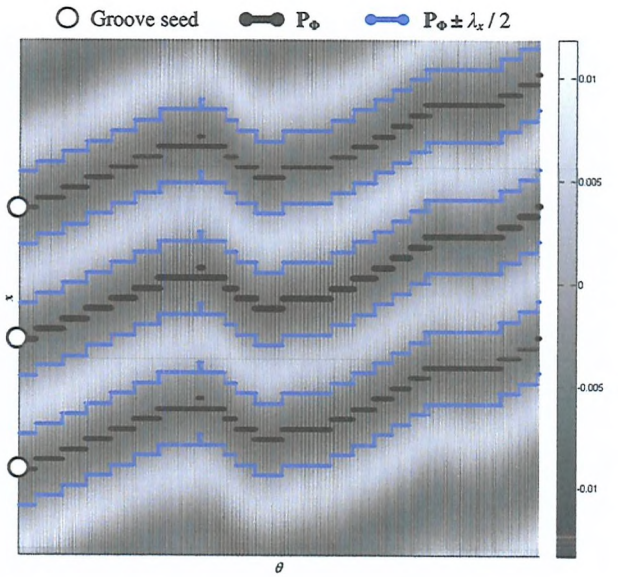

Figure 10: Obtaining the groove matrix, G_t

5.3 Audio Signal Estimation

Once G_t has been determined, one question arises : "Which region of the groove will return a displacement track with best signal-to-noise ratio?". With conventional stylus transfer, this question is answered by trying styli of different sizes/shapes and selecting the one which offers the best SNR. In the case of our discrete groove matrix G_t, we must achieve this through the selection of samples, and in doing so, seek to maximize the SNR of our displacement track $\xi(t)$. It is possible to derive many estimates for $\xi(t)$, for example:

i. data streams across the groove cross-section (columns of G_t)
ii. un-weighted/weighted sample averaging across the groove cross-section.
iii. stylus model, fitting based methods [2].
iv. rule-based navigation through G_t.

An example of how different estimates for $\xi(t)$ can produce different SNRs, based on estimate (i) was investigated for a Blue Amberol cylinder. The cylinder in question had a symmetric, 'U-Shape profile', thus the minimum point (or turning point), indexes the centre (medial axis) of the groove. The groove matrix G_t was set so that its central column contained the minimum point for each groove cross-section. Three data streams were derived (shown diagrammatically in Figure 11):

- $\xi_1(t)$: Mean of columns [-4, +4], from central column of G_t
- $\xi_2(t)$: Mean of columns [-2, +2], from central column of G_t
- $\xi_3(t)$: The central column of G_t (corresponding to the minimum sample for all groove cross-sections)

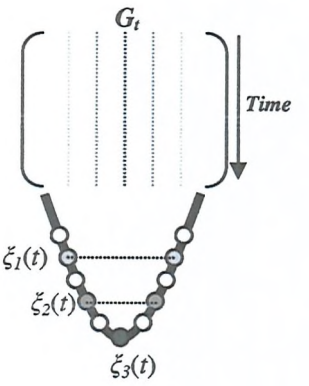

Figure 11 : Comparing displacement signals as a function of the groove cross-section.

The 10ms time histories in Figure 12 show how the apparent signal noise decreases as the groove bottom is approached. This trend can also be seen by the level of the noise floor, (assuming the noise lies mainly above 5 kHz), shown in the Power Spectral Density plots in fig. 13.

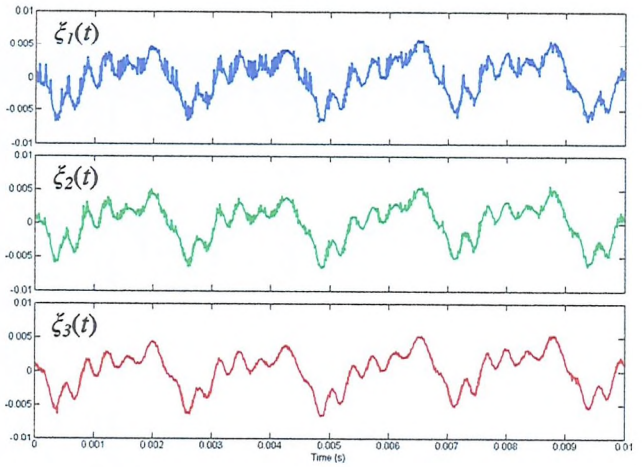

Figure 12: 10ms time histories of displacement signals, as a function of the groove cross-section. (Y-axis amplitudes given in mm).

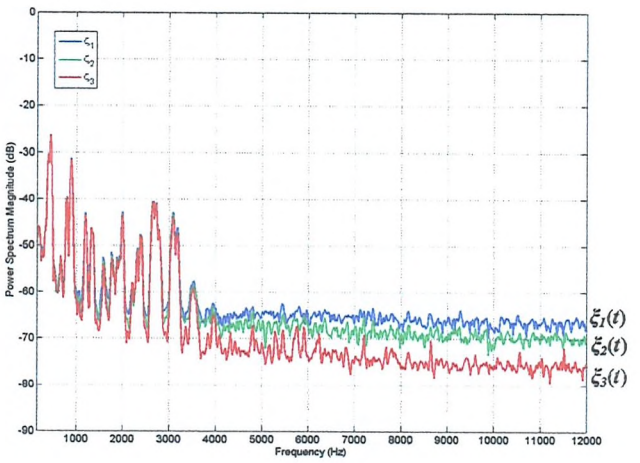

Figure 13 : PSD estimates for displacement signals $\xi_1(t)$, $\xi_2(t)$ and $\xi_3(t)$ (FFT of 1 second duration).

The method of choosing a single column of the G_t matrix as the estimate for the recorded displacement signal is perhaps an oversimplified one. It does however demonstrate that one source of noise (which must be avoided) when estimating $\xi(t)$, can be generated by the transverse (side-to-side) selection of points across the groove cross-section, between adjacent time steps, $\xi(t)$ and $\xi(t+\Delta t)$.

An optimal estimate for obtaining the best signal to noise ratio has yet been concluded, and is likely to vary from cylinder to cylinder, depending the condition of the grooved surface. The answer may lie in some combination of estimates (ii-iv) as described above and remains the main focus for the signal processing required for audio extraction.

6 CONCLUSIONS

The resolution and sampling requirements for non-contact surface mapping of acoustic cylinder recordings have been discussed. A new method for estimating the groove trajectory based on sinusoidal modelling and examination of the relative phase shift between linescans has been successfully implemented and described.

An example has been shown which demonstrates how the signal to noise ratio of our output signal may vary, depending on the choice of samples across the groove cross-section. The development of an optimal method for maximizing the SNR remains one of the key signal processing issues.

In order to help quantify the signal to noise ratio of the non- contact method, a cylinder recording has been is currently under test, which contains test tones and known reference signals.

Although the sound encoded on the earliest acoustic recordings may not be considered 'high resolution' by today's standards, the sensing technology required for non-contact audio recovery must be capable of resolving submicron-scale groove modulations. This raises an interesting question of 'what constitutes high resolution audio?'

7 ACKNOWLEDGEMENTS

The authors gratefully acknowledge support from the UK's Engineering and Physical Sciences Research Council, the British Library Sound Archive and TaiCaan Technologies Ltd.

REFERENCES

[1] Fadeyev, V. et. al, "Reconstruction of Recorded Sound from an Edison Cylinder using Three-Dimensional Non-Contact Optical Surface Metrology." *JAES* 53(6): 485-508 (2005).

[2] Fadeyev, V, et. al, "Reconstruction of Mechanically Recorded Sound by Image Processing." *JAES* 51(12): 1172-1185 (2003).

[3] Stotzer, et al., "Phonographic sound extraction using photography and signal processing." *Digital Signal Processing* 17(2): 433-450 (2007).

[4] Stotzer, et al., "VisualAudio: an Optical Technique to Save the Sound of Phonographic Records", *Joint Technical Symposium*, Toronto, Canada, June 24-26, (2004).

[5] Galo, Gary A. "Disc Recording Equalization Demystified" *ARSC Journal*, 17(2): 188-211 (1996).

[6] Boltryk, et al. "A comparison of precision optical displacement sensors for the 3-D measurement of complex surface profiles." *Sensors and Actuators A: Physical* doi:10.1016/j.sna.2007.03.006., (2007)

[7] Olson, F. H., *Acoustical Engineering*, p359, (1957)

[8] Kay, S. M, *Modern Spectral Estimation: Theory and Application*, Prentice Hall (1988)

Available online at www.sciencedirect.com
ScienceDirect
Sensors and Actuators A 142 (2008) 2–11
SENSORS AND ACTUATORS A PHYSICAL
www.elsevier.com/locate/sna

A comparison of precision optical displacement sensors for the 3D measurement of complex surface profiles

Peter J. Boltryk *, Martyn Hill, John W. McBride, Antony Nascè

School of Engineering Sciences, University of Southampton, Southampton, UK

Received 1 October 2006; received in revised form 1 March 2007; accepted 1 March 2007
Available online 14 March 2007

Abstract

Non-contact optical sensors, combined with a raster scanning system, can be used in surface metrology applications, such as MEMS processing to build high-precision topological surface maps representing the surface profile. We compare three optical displacement sensors for the measurement of complex surfaces, focussing on early sound recordings and recommend sensor selection based on parameters, including measurement area, time and resolution.

Keywords: Non-contact metrology; Optical displacement sensor; Profilometer; Sound recovery

1. Introduction

The 3D measurement of complex surface topologies is important in many fields, in diverse applications, such as tribology, profiling of optical components and MEMS processing. In many circumstances, it is desirable to use non-contact measurement processes to avoid potential surface damage or contamination of the surface resulting from the use of metrology instruments, such as a mechanical stylus profilometer. Atomic force microscopy provides extremely high spatial resolution [1], but is usually only practical for relatively small surface areas. For larger artefacts, optical displacement sensors are preferable. The single point measurement sensors can be combined with a raster scanning system to create a 3D topological map of the surface, composed of (x, y, z) data, whose resolution is dependent on the sensor's axial resolution (z-axis) and the chosen resolution of the lateral sampling (x- and y-axis). The meaningful lateral resolution in the x- and y-axis directions is also limited by the lateral resolution of the sensor.

Different types of optical sensors are affected to varying degrees by surface reflectivity, surface inclination angle, spot size and axial resolution [2]. This paper describes a study of competing optical displacement sensors for applications where the sensor's axial resolution is a fundamental parameter. The sensors that have been tested include a triangulation laser (TL), a confocal laser (CL) and confocal white light probe (WL). The operating principles of each sensor are briefly summarised, and then each sensor tested on representative media. Direct comparisons of measured surface profiles are performed using a reference cylinder machined with angular features. The sensors are further evaluated for application in the measurement of early cylinder recordings as part of a process for the non-contact recovery of sound, where the key parameters of interest are sensor response, scan times and ability to resolve surface details for effective sound recovery.

2. Overview of sensing technologies

The triangulation principle is arguably the simplest sensing technology considered in this paper, but suffers from a number of crucial drawbacks which are discussed later. A laser source is directed onto the surface of the artefact, at a nominally normal angle to the surface profile. An off-axis lens collects the diffuse surface reflections and focuses the resultant beam onto a CCD array contained within the sensor head. The deflection in the position of the maximum intensity of this beam is representative of the displacement of the surface height. Unlike the CL and WL confocal sensors, where the reflected light used by the sensor uses the same transmission path as the incident beam,

* Corresponding author. Tel.: +44 23 80592339.
E-mail address: boltryk@soton.ac.uk (P.J. Boltryk).

0924-4247/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.sna.2007.03.006

the TL sensor requires unobstructed light transmission towards the off-axis lens and detector. Under certain conditions, such as scanning adjacent to deep troughs with steep sides, the surface form can obstruct the light path to the detector, causing a shadow.

CL sensors use a continuous wave, red light (670 nm for the sensor used in this paper), laser source, which is first passed through a collimator lens to limit divergence of the beam. A lens focuses the light into a small spot on the surface, whose reflection passes back through the focussing lens and is then diverted towards a pinhole using a beam splitter. The optics are arranged to ensure that the intensity of light received through the pinhole on a CCD detector is attenuated, if the surface is not within the focal point of the incident light. The focussing lens is oscillated at high frequency in the axis of the light direction using a tuning fork arrangement, such that the focal point for the light spot varies as a function of the instantaneous tuning fork position. Surface height is thus inferred by detecting the lens's position at the point of maximum received light intensity, which is then cross-referenced to the calibrated height measurement.

The WL sensor operates on the confocal principle, where polychromatic light is focussed onto the specimen's surface. Lens-induced chromatic aberration causes the light to be split into a continuous distribution by wavelength, whose individual focal points vary with wavelength. The resultant reflections are transmitted through the confocal system to a spectrometer system via an optic fibre. The wavelength corresponding to the highest detected intensity is representative of the displacement between sensor lens and the surface, which is calibrated to relate the wavelength to a given surface displacement (i.e. height) value. Unlike the CL which operates using the oscillating objective lens, the WL sensor lacks moving optical components.

The three sensors presented in this study are commercial devices which have been specifically selected for this application.

3. Sensor parameters

Table 1 lists a series of parameters that characterise the performance of the three sensors selected here for use on complex surface profiles. In this study, we specifically consider the measurement of mechanical cylinder recordings, whose surface features the sound-carrying groove. A cross-section through such a groove reveals a trough-like feature, with angular sides, typically a depth of up to 30 μm and a helical pitch of approximately 130 μm. Included in this table are parameters for a stylus profilometer equipped with a sapphire stylus, used in this study as a reference measurement system.

Since each sensor has a finite spot size, a given measurement results from an averaging process of contributions from throughout the light spot. In addition, due to the number of samples required for the surface scans, it is impractical for the stages to stop at each raster point so the displacement measurement is triggered for a finite period of time as the sensor moves past the required grid position. The surface area addressed by the sensor during a single measurement is therefore, elongated through the stage's movement. To accurately resolve troughs, it is necessary for the spot size to be sufficiently small. As shown later, the large spot size of the TL sensor limits the ability of the TL sensor to successfully measure the sound-carrying grooves of a typical cylinder.

For this application, the sample frequency has the strongest influence on how quickly a surface may be measured. Whilst the CL sensor has a higher sample rate than the WL sensor, to achieve reasonable surface profiles on these artefacts requires implementation of the in-built averaging process. The CL sensor's output measurement is obtained by selecting the average of 2, 16 or 128 consecutive measurements, which, operating with the 1.4 kHz internal sample rate requires an effective data acquisition rate of approximately 454, 82 and 11 Hz, respectively. Tests using the CL sensor on representative test surfaces show that the most appropriate setting is 16 averaged measurements, as a compromise between maximising the scan rate and maintaining measurement quality. Whilst measurements using the 2-measurement average permit significantly faster scan rates, the surface profiles obtained tend to be unexpectedly irregular, based on *a priori* knowledge of the surface form. In contrast, whilst the WL sensor permits use of averaging consecutive measurements, the point-to-point measurement quality for single value operation is found to be robust for the surface profiles found on typical mechanical sound recordings and the requirement to reduce scan time makes averaging less attractive. As shown later, the consequence of this choice can result in data-loss on steeper inclines when compared with the averaged CL sensor profiles. However, for the sound recovery application described later in this paper, the sides of the groove are arguably less important than robustly measuring the form of the groove bottoms, meaning the data-loss is not necessarily problematic.

Gauge range defines the range of surface heights over which the sensor will operate. As shown in the next section, the reduced gauge range of the WL compared with the two laser sensors results in data-loss when measuring the deepest features in a

Table 1
Comparison of optical sensors plus reference stylus profilometer

Sensor	Spot size (μm)	Sample frequency (Hz)	Gauge range (mm)	Axial resolution (nm)	Angular tolerance (°)
WL	7	1000	0.35	10	27
CL	2	1400	0.6	100	17
TL	30	2000	10	1000	17

Sensor	Stylus size (μm)	Lateral measurement grid spacing (μm)	Gauge range (mm)	Axial resolution (nm)	N/A
Stylus	2	1	6	10	N/A

reference cylinder. Furthermore, this also represents a problem for typical cylinder mechanical recordings, where experience shows that the macroscopic surface form can vary more than the WL sensor's maximum range. Whilst surface tracking can keep the sensor within range of the surface, it is desirable to minimise adjustments, both to reduce scan times and also to limit uncertainties in motion stage positioning affecting surface profile measurements.

The ability of the sensors to measure on inclined surfaces is defined by the limit of angular tolerance. The WL sensor used for this work had the highest available angular tolerance, although it will be observed that out of the three sensors used in this study, the WL sensor was affected most by the angular nature of the surface features encountered.

4. Sensor comparison using machined reference cylinder

Sensor measurement quality was initially assessed with a steel reference cylinder (Fig. 1), machined with a series of features designed to test the sensors' performance on both generic angular surfaces and grooves reminiscent of the sound-carrying grooves found on mechanical sound recordings.

The cylinder's external diameter was ground to achieve a high level of concentricity, into which a set of grooved rings were machined using HSS tool steel blanks, ground such that the cut grooves have sides angled at approximately 15°, 20°, 25°, 30° and 45° to a plane tangential to the cylinder surface. The size of these grooves is an order of magnitude larger than those found on mechanical recordings and are designed to test the sensor's ability to measure inclined surfaces. The cylinder's orientation is identified using a pair of machined cross hairs at each end of the cylinder, which enables measurements to be taken along the same line for direct comparison. The cylinder also features a set of five 30 μm-deep rings and a single 150 μm-deep ring, all of which were machined using a custom ground diamond tipped tool. The designed profile of the diamond tip matches the specification for Edison-manufactured recording styli, as a formal test for the sensors operating on grooves similar to those found on mechanical sound recordings.

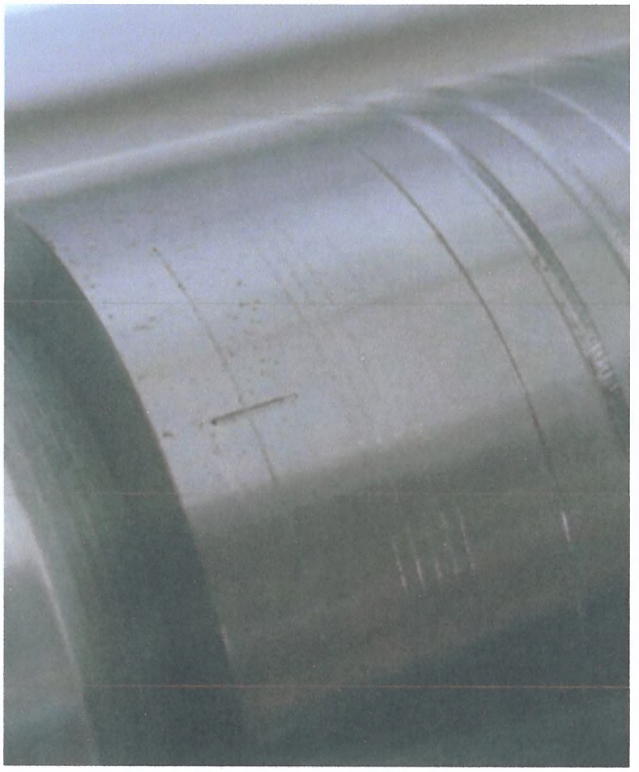

Fig. 1. One end of reference cylinder, mounted between conical centres on scanning mandrel. From left to right: location cross hair can be seen on the left-hand end, five diamond machined rings plunged to 30 μm depth, one diamond machined ring plunged to 150 μm depth and four angled grooves machined using custom-ground tool steel blanks.

4.1. Results on the deeper, HSS steel cut grooves

Fig. 2 compares the profiles of the larger features machined into the cylinder, as measured by the three optical non-contact sensors, with the measured profile obtained using a mechanical stylus profilometer (bottom axis), equipped with a 2 μm sapphire stylus. As discussed in the previous section, when using the TL sensor, the scanning system operates at an effective data acquisition rate of 2 kHz, whilst data are acquired at 82 Hz and 1000 Hz when using the CL and WL sensors, respectively. The WL sensor is shown to lose data on the steeper features (A) and the loss of measured data at the bottom of the two deepest grooves (B) demonstrate the lower gauge range of this sensor. In contrast with their larger gauge ranges, both laser displacement sensors are successful in measuring the height ranges of all features, and are generally better able to operate on the inclined surfaces, although for the case of the CL sensor, this may be partly explained by the averaging of 16 consecutive measurements.

The measured profiles for the fifth groove from the left (groove with approximately 45° angled sides) are plotted in Fig. 3. In this plot, all three optical displacement sensor profiles are compared with the measurement obtained using a stylus profilometer. Each dataset plotted here has a linear trend removed using a least squares fit, which compensates for the offsets introduced by misalignment between the scanning axis and the surface. The bottom of this groove does not form a perfect 'V' groove, caused by tool chatter during the machining process of this deep cut, and therefore, features an undulating surface form.

The TL sensor's profile is subject to the highest apparent surface roughness, which is particularly evident at the top of the groove. Considering the level of surface roughness indicated by the stylus profilometer and CL and WL sensors at the outer diameter's ground surface, it is reasonable to assume that the TL sensor's higher profile irregularity is the result of higher point-to-point uncertainty affecting the measurements, rather than it being a genuine surface roughness feature. Where the surface genuinely exhibits high levels of surface irregularity, such as the plateau at the bottom of the inclined surfaces, the TL sensor profile closely resembles the profiles obtained using the CL sensor and stylus profilometer.

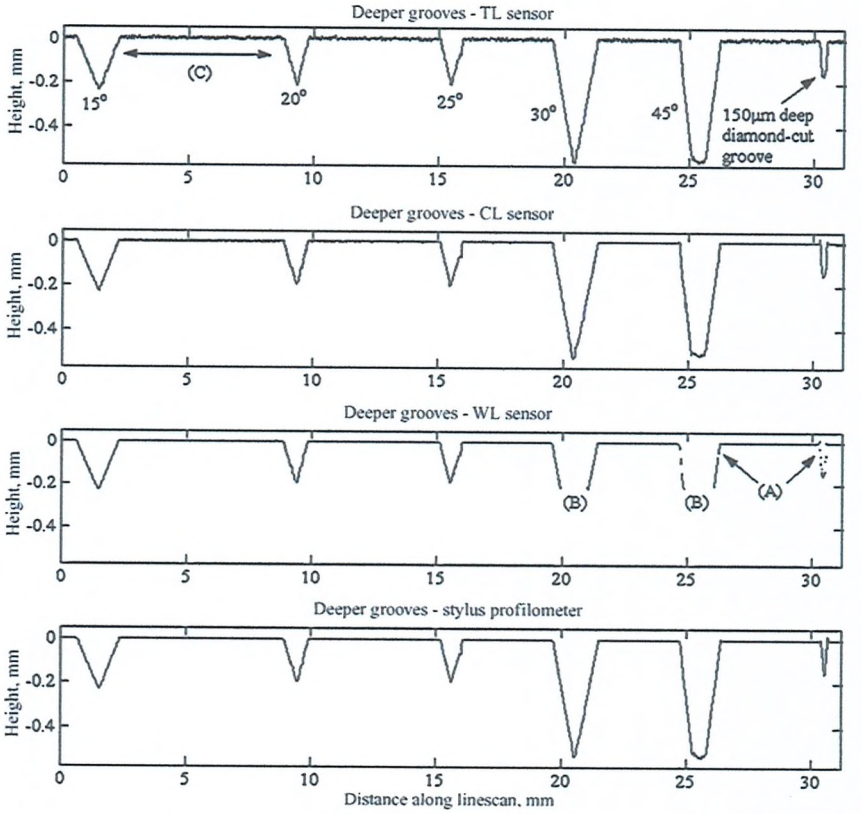

Fig. 2. Cross-section of deeper grooves, measured using TL, CL and WL sensors, respectively, compared with bottom graph which plots the profile obtained using mechanical stylus profilometer instrument. Features from left to right: 15°, 20°, 25°, 30° and 45°angled grooves to varying depths and 150 μm deep diamond-cut groove (*y*-axis exaggerated and least squares component removed). For the optical displacement sensors, the effective data acquisition rate achieved by the scanning system (and therefore, the speed of the scanning stages) is dependent on each sensor's selected averaging function and sampling frequency. When using the WL, CL and TL sensors, the scanning system can successfully acquire data at 1 kHz, 82 Hz and 2 kHz, respectively. However, whilst the TL can achieve the highest data rates, it is apparent from this figure that the measured profiles are unexpectedly irregular when compared with the alternative sensor profiles, suggesting that there is a compromise between measurement quality and speed of measurement.

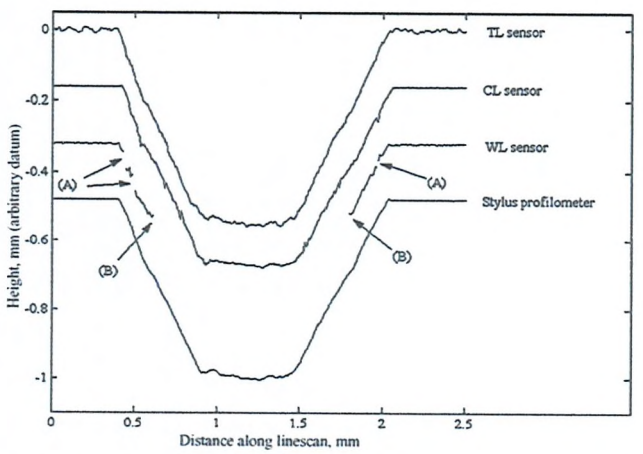

Fig. 3. Plots of the deeper 45° groove (fifth angular feature) as measured by the three optical sensors and compared with the stylus profilometer (bottom trace). For clarity, the *y*-axis offsets have been adjusted, the linear trend removed using least squares and the *y*-axis scaling is exaggerated. In common with Fig. 2, the scanning system was acquiring data at 1 kHz, 82 Hz and 2 kHz when using the WL, CL and TL sensors, respectively.

Of particular interest in this plot is the loss of data for the WL profile. One cause of data-loss in this case is that the depth of the groove exceeds the WL sensor's gauge range, resulting in loss of measurement in the regions labelled (B) for parts of the groove whose depth exceeds 350 μm below the cylinder's nominal outer diameter surface. Loss of data is also observed within the sensor's gauge range (for example, in the region labelled (A) of the linescan, where the depth of the groove surface is between 150 and 250 μm), which is attributed to angular tolerance of the WL sensor. Operating with a single measurement value (rather than averaged measurements as chosen for the CL sensor), it is likely that the intensity of the reflected light from the steeply inclined groove walls was too weak for a measurement to be taken.

The measured output from the displacement sensors can be viewed as the combination of the true surface profile plus an error component that is related to the point-to-point uncertainty introduced by each sensor. If each sensor measures the same region on the reference cylinder, then the indicated surface irregularity can be used as a metric to characterise the point-to-point uncertainty affecting each sensor. We define the indicated surface irregularity as the standard deviation of the measurements of a given region for each of the profiles, using N discrete height

measurements z, as shown in Eq. (1).

$$\delta = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(z_i - \bar{z})^2} \quad (1)$$

In the cylinder measurement system, the sensor is located at a known stand-off distance from the surface so the height measurements z are provided by the displacement measurements to the surface. Whilst the calculated standard deviation of the measurement values can only be used as a guide, as the differing spot size and scan speeds between the different sensors results in a different averaged area that is represented by each measurement point (lateral resolution), which has bearing on the calculated roughness [3], it is nonetheless a useful numerical comparison of the sensor's measured profile, as it can provide an indication the sensor point-to-point uncertainty on an identical test surface. An appropriate region for this study is the ground external surface of the cylinder, between the first and second angular feature (referring to Fig. 2, in the region labelled (C) along the linescan axis). Using data, where the least squares linear coefficient has been removed makes this calculation robust to significant macroscopic surface details influencing the δ figure. Note, however, that whilst the sampling resolution between measurement points on the optical sensors was a constant 5 μm, the stylus profilometer resolution was 1 μm, making it difficult to directly compare the δ figures for the mechanical system with any of the optical sensors. The resulting δ values are 3.67, 0.631, 0.428 and 0.365 μm for the TL, CL, WL and stylus profilometer, respectively. Considering the optical sensors, the TL sensor is characterised by a δ value that is roughly five and eight times larger than the CL and WL sensors, respectively. This large δ value is directly related to the TL sensor's higher point-to-point measurement uncertainty and suggests that caution should be used in using this sensing technology for surface roughness measurements as it could provide misleading results.

4.2. Measurements of 30 μm deep diamond-cut rings

Measured profiles of the five shallow-cut grooves replicating the sound-carrying grooves found on cylinder recordings are plotted in Fig. 4. Of particular note is the missing data on two of the groove profiles for the WL sensor near the regions labelled (A), where it is likely that the angle of the groove edge is too large for the sensor to return a reading. The CL sensor appears to suggest a slightly higher surface roughness on the ground outside diameter of the cylinder, whilst the TL sensor is shown to exhibit large point-to-point measurement variability.

In summary, results from measuring the reference cylinder suggest that there is a compromise between measurement noise and gauge range. The following section describes application of the sensors to the measurement of a mass-produced Edison

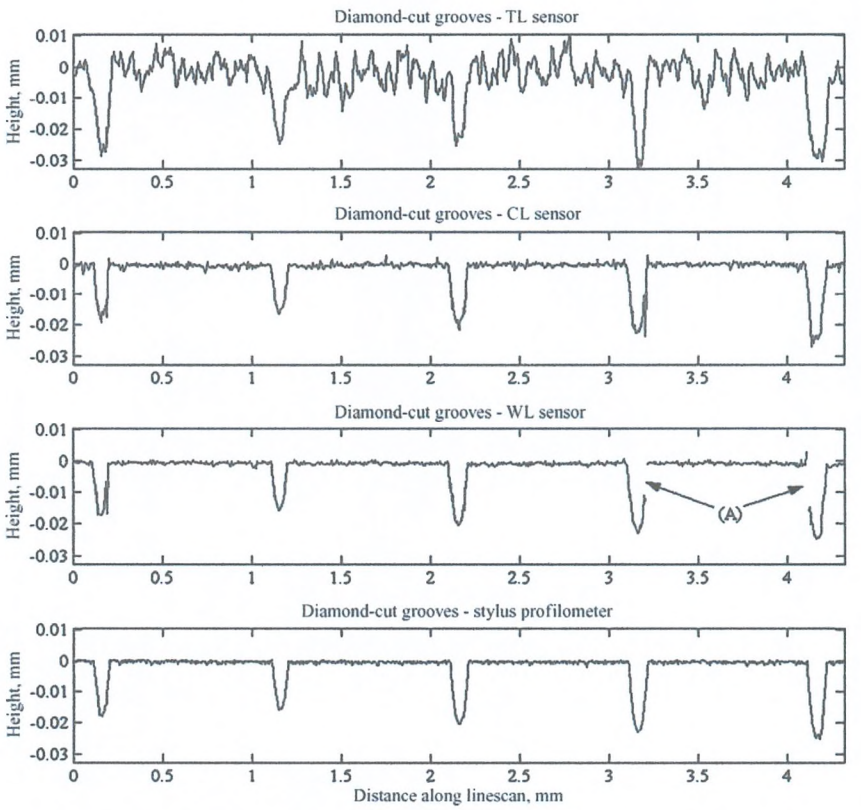

Fig. 4. Top three axis show cross-sections of five 30 μm-deep diamond-cut grooves measured using TL, CL, WL sensors, respectively, again compared with the stylus profilometer (bottom axis). In all cases, y-axis scaling is exaggerated and least squares fit removed for clarity. Referring to Section 3 and Fig. 2, the scanning system was able to acquire data at 1 kHz, 82 Hz and 2 kHz when using the WL, CL and TL sensors, respectively.

cylinder recording, where sensor resolution is highlighted as a fundamental criterion for sensor selection.

5. Application to mechanical cylinder recordings

5.1. Cylinder measurement systems

The development of optical measurement systems for cylinder recordings has been the subject of much research in recent years, primarily for the purposes of art preservation and providing public access to rare music collections. Systems developed by the Institute for Information Recording Problems in the Ukraine [4,5], Syracuse University [6] and Hokkaido University group in Japan [7–10] have been developed to measure the sound track profile only and not the recording surface in its entirety. The philosophy behind these measurement systems is for sound reproduction only and not for purposes of surface preservation. Furthermore, the non-contacting methods require a tracking system which must be accurate at time of measurement. This means that tracking groove features in situ, often results in unsatisfactory sound recovery, due to the nature of damage and deformation of the cylinder surface.

5.2. 2D measurement systems

An alternative measurement approach (and one which is most suitable for sound reproduction from damaged surfaces), is to measure the entire recorded surface, and apply image and signal processing, post-measurement. Other research groups, such as Refs. [11–15] have developed 2D imaging systems for flat disc recordings. The sound encoded in these disc recordings is held in the lateral (side-to-side) undulations of the grooves. The entire disc surface may be imaged via video zoom microscope [12], high-resolution flat-bed scanner [13] or by high-resolution photography [11,14,15]. Once the surface has been measured, edge detection and feature extraction may be applied to locate the grooves and acquire the audio data. The advantage of post-processing the measured data is that any scratches or deformation of the grooves may be treated, post-measurement in the computer. These 2D methods are intended for flat disc recordings only. For cylinders, the measurement system must operate in 3D, because the sound is encoded as vertical modulations, which therefore, requires measurement of the depth of the sound-carrying groove.

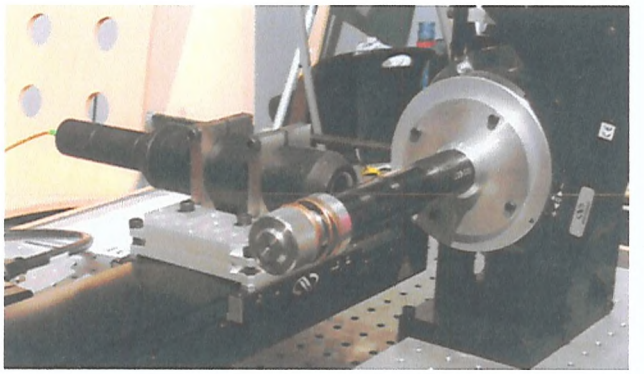

Fig. 5. Mandrel-mounted cylinder and WL sensor mounted on x–y translation stages. Linear stages are used to move the sensor platform in a linescan direction and to advance the sensor platform to the required stand-off distance, in the radial direction. The rotary stage increments the cylinder in the angular direction between linescans.

5.3. 3D measurement system

Research carried out in collaboration with Berkeley Laboratories and Taicaan Technologies has resulted in the development of a 3D measurement system (Fig. 5) for cylinder recordings [16]. A linear stage traverses the sensor platform in a linescan along the axis of the cylinder, which is mounted on a rotary stage, and allows rotation of the cylinder after every linescan to a resolution of 0.01°. For the mass-produced Edison cylinder shown in Fig. 5, this angular resolution equates to sampling the sound track at 96 kHz, which is chosen to provide redundancy in the data, and to allow for spatial filtering without reducing the audio bandwidth. An additional linear stage is used to advance the sensor platform to the sensor's required stand-off distance from the cylinder surface, in the radial direction.

The resulting surface map (Fig. 6) represents the radial height profile of the unwrapped cylinder surface, which is first analyzed to locate the groove trajectories. A continuous stylus displace-

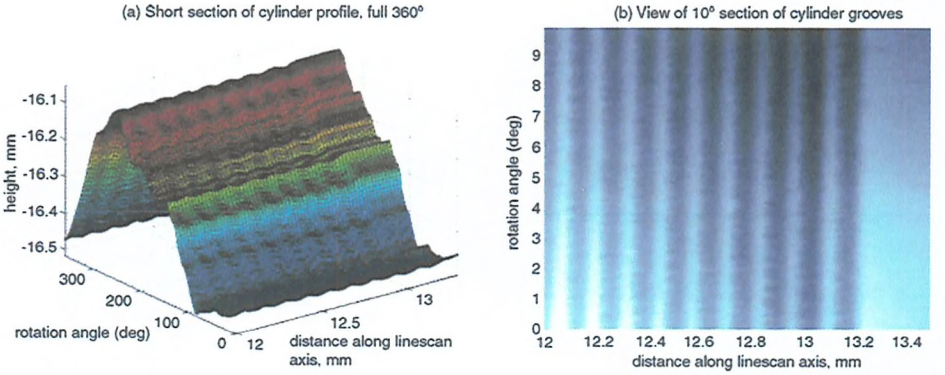

Fig. 6. Graph (a) plots radial height of cylinder surface and plot (b) zooms on 10° angular segment of surface: the ungrooved end of the cylinder can be clearly observed, together with the first 10 sound-carrying grooves. Surface measurements are provided by the WL sensor.

ment signal can thus be inferred by joining the grooves into a coherent time series. The audio signal is calculated by differentiating the displacement signal.

5.4. Sound-carrying groove measurement

The three optical sensors were further compared by measuring the same 10 mm length of a mass-produced Blue Amberol Edison cylinder recording, comparing the resulting surface profiles and recovering the audio signal. Whilst Blue Amberol cylinders are not generally considered to be precious, the use of such a cylinder as a test artefact may be viewed as a conservative choice, as shrinkage and deformation over time is worse than for alternative types of cylinder recordings. The resulting non-concentricity in Blue Amberol cylinders, thus, places higher demands on the chosen sensor and the raster scanning system, to ensure the sensor remains within its range.

Apart from subjective assessment of the resulting sound quality, analysis of the amplitude spectra for the derived stylus trajectory demonstrates that a surprisingly high resolution is required for this application.

5.5. Measured surface form comparison

In common with the reference cylinder results, the application of the optical sensors to a genuine cylinder recording demonstrates that the TL sensor suffers high levels of surface irregularity. Referring to Fig. 7, the TL measured surface is extremely irregular (compared with the results for the CL and WL sensor), and inspection of the groove cross-sections demonstrates that the TL fails to resolve the groove profiles in many instances making it unsuitable for use in the sound recovery system. Given that the pitch of the Blue Amberol cylinder's grooves is 130 μm, and the spot size of the TL sensor is 30 μm, it is perhaps unsurprising that the TL sensor is generally unsuccessful at properly resolving groove details.

Comparing the CL and WL data, it is apparent that the measured grooves are slightly smoother for the WL data, and the tops of the grooves more rounded. An interesting observation is that, unlike for the reference cylinder, the WL does not generally suffer data-loss for the steepest parts of the groove cross-sections. This might be explained by the fact that reflections from the steel reference cylinder are predominately specular, whilst the cellu-

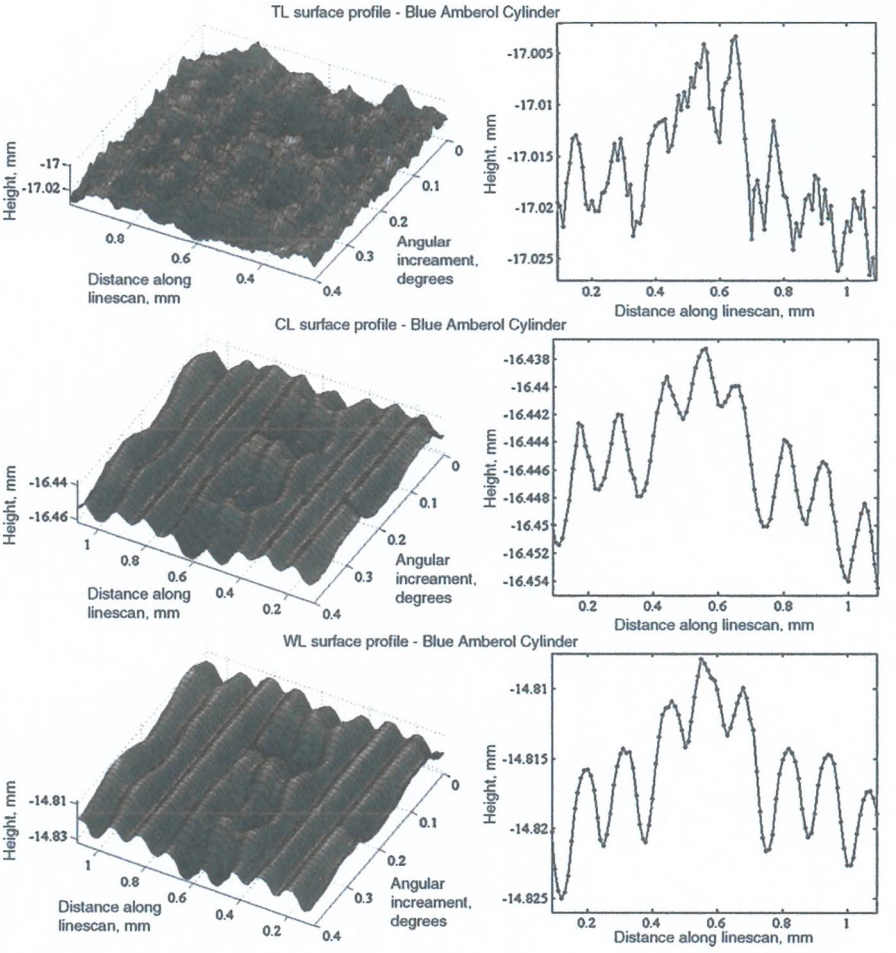

Fig. 7. Comparison of measured surface profiles, for identical region on Blue Amberol cylinder surface, measured using three optical displacement sensors. Right-hand set of axes plot linescan profiles across groove structure at 0.2° angular increment for each sensor. Again, data acquisition rate was dependent on the sensor, with the scanning system acquiring data at 1 kHz, 82 Hz and 2 kHz when using the WL, CL and TL sensors, respectively.

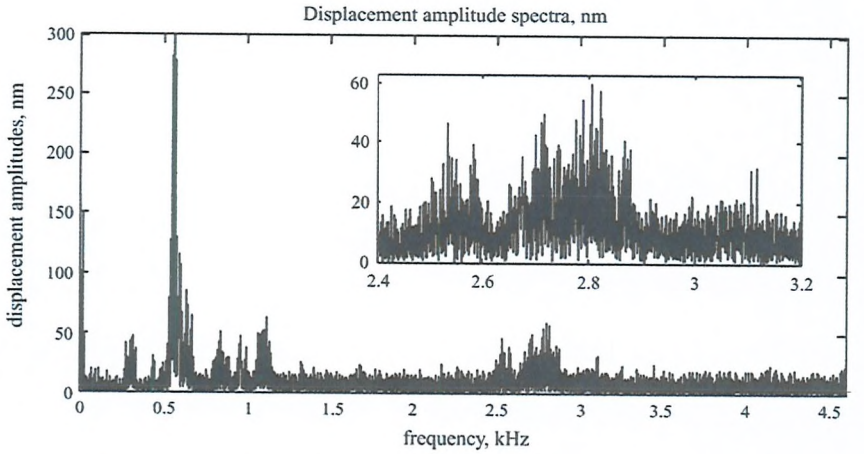

Fig. 8. FFT analysis of 2.5 s excerpt recovered from a Blue Amberol cylinder by the WL sensor. The inset plot provides a zoomed view of important vocal content in the 2.5–3.2 kHz range. The observed displacement amplitudes in this frequency range demonstrate that there are important spectral components whose amplitudes are sub 50 nm.

lose material used in the Blue Amberol cylinder may produce a combination of diffuse and specular reflections due to its dielectric surface properties, resulting in higher intensity confocal reflections at these steeper angles.

5.6. Scan time

For archival feasibility, a short scan time is crucial. This represents a non-trivial problem since a 4 min-long Edison cylinder (approximately 92 mm length of recorded surface), with spatial sampling of 10 μm and 0.01° in the linescan and radial directions, respectively, requires in excess of 300 million data points. Scan times are directly related to the sensor sample frequencies, so it is not surprising that the CL sensor, when operating with 16 averages (82 Hz) takes 160 h to complete the 10 mm section, the WL sensor takes 81 h, whilst the TL sensor operating at 2 kHz takes 19.5 h.

5.7. Cylinder non-concentricity

Typical Edison cylinders are non-concentric, due to manufacturing processes and factors, such as shrinkage and creep caused by long-term storage. Tests on the Blue Amberol artefact demonstrate that the height variation along any given linescan can exceed the stated gauge range of the WL sensor, so the WL sensor therefore, requires surface tracking to remain within the range. Larger gauge ranges can be achieved using combinations of lens components in the optical head, but at the expense of axial resolution. In the following section, it is observed that this system requires extremely high axial resolution, so it is unlikely that a reduction in axial resolution to improve gauge range is tolerable.

5.8. Recovered modulation signal

Fig. 8 plots FFT analysis for a quiet (i.e. smaller displacement amplitude) 2.5 s excerpt from an early twentieth century Blue Amberol cylinder, recovered using the WL sensor. The audio signal contains male vocals, accompanied by a series of instrumentalists. The spectral content in the 2.4–3.2 kHz range correspond to some spectral components of the vocals (high frequency speech content of the singing, such as syllables). Referring to the inset plot on this figure, which zooms in on the amplitude spectra in this range, there are significant spectral components whose amplitude spectra are sub 50 nm, suggesting that at least 20–30 nm resolution is required to measure the displacement modulations for this particular recording. This displacement amplitude figure is comparable with electron microscopy studies of flat disc recordings [17], where typical amplitude modulations can be a few millionths of an inch (25–75 nm).

Fig. 9 compares the spectrograms of the recovered sound signal, for both the WL and CL sensors for identical excerpts from the same cylinder. It is clear from this analysis that the audio signal recovered by the WL sensor exhibits a less pronounced broadband noise component than the CL sensor, suggesting that the WL sensor provides the sound recovery system with a higher signal to noise ratio. This can be observed in Fig. 9 by comparing the ellipse regions (D) and (E), which correspond to a moment of silence in the original sound track. The audio signal recovered using the CL sensor (D) has a much higher broadband spectral power content, across the whole bandwidth, than the equivalent WL signal (E).

5.9. Audio assessment

Qualitative comparison of the audio signals recovered by the WL and CL sensors suggest that the WL sensor reproduces the highest quality sound field. Whilst our results suggest that the CL is capable of resolving the groove profile, the WL's superior resolution appears to make the WL sensing technology more attractive, as also highlighted by the higher signal to noise ratio in Fig. 9 (compare, for example, the regions in (F) and (G), where the tonal content of the music is more pronounced against the background broadband noise for the WL when compared with the CL sensor signal).

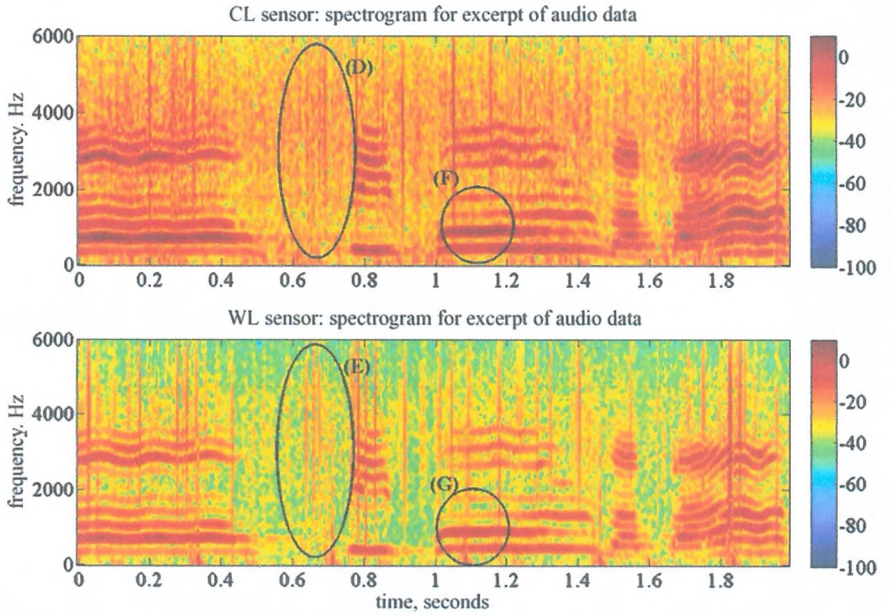

Fig. 9. Comparison of CL and WL spectrograms for 2 s audio excerpt. Note the stronger broadband noise component that affects the CL spectrogram. Despite the high effective sampling frequency (96 kHz), for clarity, these spectrogram axes are limited to the bandwidth of interest for early mechanical recordings, where an estimate of the bandwidth is up to 6 kHz. The prominent modulating frequency signals result from vibrato in the vocalist's style.

6. Discussion

This study compares three different types of optical displacement sensors for measuring complex surface forms, with particular application to artefacts with micron-scale grooved structures. Referring to Table 1, the WL sensor is more limited in its gauge range than the other two sensors, which is observed in practice by the loss of data at the bottom of the deeper HSS-machined grooves on the reference cylinder. In the sound recovery application described in this paper, extremely high spatial resolution scanning is required, over a relatively large area. Maximising the scanning system's data acquisition rates is therefore, vitally important, the limit of which is dependent on the sensor used. Referring to Table 2, data acquisition rates using the TL sensor are significantly higher than the two confocal sensors, but this study has demonstrated that the data from the TL sensor has neither the resolution nor the accuracy required for the recovery of sound from the cylinders. Empirically, to resolve the groove structure requires the CL to operate at 82 Hz, whilst the WL sensor is able to operate at its highest sampling rate of 1 kHz. In terms of scan times, the WL sensor is therefore preferable, for the surface contours measured as part of this study.

Table 2
Summary of sensor sampling frequency, the number of measurements averaged for an output reading to successfully measure the surface of the cylinder artefacts and the scanning system's effective data acquisition rate

Sensor	Sampling frequency (Hz)	Averages	Effective data acquisition rate (Hz)
WL	1000	1	1000
CL	1400	16	82
TL	2000	1	2000

Given the higher angular tolerance of the WL sensor, it is perhaps surprising at first sight that it was affected most by the steepest angular features found on the steel reference cylinder. Considering first, the CL and WL sensor performances, the proposed explanation for this discrepancy lies in the use of the CL's in-built averaging option used to ensure that the CL profiles accurately represent the cylinder recording groove profiles. For the case of the WL, no such averaging was necessary as the measured sound-carrying grooves were already highly representative of the actual groove profiles. It is envisaged that the use of averaging for the WL sensor would overcome the infrequent loss of data observed for steep sided features on metallic surfaces. The robustness of the WL sensor may also be improved on inclined surfaces by use of a lower sampling frequency, which permits its in-built spectrometer to receive a higher quantity of reflected light. However, this solution is at the expense of longer scan times, and since typical surfaces on the non-metallic Blue Amberol cylinders do not appear to present a problem for the WL sensor, this is an unnecessary process to implement.

A possible reason why the TL sensor was able to return measurements more consistently than the WL sensor on the steeply inclined surfaces of the reference cylinder is that the TL has a significantly larger spot size, increasing the likelihood of surface roughness features causing strong specular reflections.

7. Conclusions

Selection of non-contact displacement sensors for measurement of complex surface forms application is shown to be a

compromise between factors that include scan times, gauge range, sensor noise and required resolution.

Three competing sensing technologies have been compared both on inclined surfaces, on shallow diamond-cut grooves and finally applied to the measurement of the surface of mechanical recording media. It has been observed that the WL sensor is more strongly affected by surface inclination angles, but offers the highest axial resolution and a realistic data acquisition rate for applications, where large surface areas are to be measured. However, for applications, where the height of the surface's macroscopic form varies considerably, the limited gauge range of the WL sensor may require the sensor stand-off distance to be adjusted to keep the sensor focus point within range.

Acknowledgements

The authors gratefully acknowledge support from the UK's Engineering and Physical Sciences Research Council, the British Library Sound Archive and TaiCaan Technologies Ltd.

References

[1] K.J. Stout (Ed.), Developments of Methods for the Characterisation of Roughness in Three Dimensions., Penton Press, London, 2000, pp. 17–29.
[2] J.W. McBride, C. Maul, IEICE Trans. Electron. E87 (8) (2004) 1261–1267.
[3] J.M. Bennet, L. Mattsson, Introduction to Surface Roughness and Scattering, Optical Society of America, Washington, DC, 1989, p. 38.
[4] V. Petrov, A. Kryuchin, et al., Optomechanical method of sound from Edison cylinders, in: International Conference on Optical Storage, Imaging and Transmission of Information, 1997.
[5] S. Shanoylo, I. Kosyak, et al., Reading and processing of audio information reproduced from Edison phonograph cylinders by method of laser interferometry, Laser Tech. Syst. Art Conserv. (2001).
[6] W.A. Penn, M.J. Hanson, The Syracuse University Library Radius Project: development of a non-destructive playback system for cylinder recordings, First Monday 8 (5) (2003).
[7] T. Asakura, J. Uozumi, Optical methods for reproducing sounds from old phonograph records, Int. Trends Opt. Photonics ICO IV (1999) 65–81.
[8] T. Iwai, T. Asakura, et al., Reproduction of sound from old wax phonograph cylinders using the laser-beam reflection method, Appl. Opt. 25 (5) (1986) 597–604.
[9] J. Uozumi, T. Asakura, Reproduction of sound from old disks by the laser diffraction method, Appl. Opt. 27 (13) (1988) 2671–2676.
[10] J. Uozumi, T. Ushizaka, et al., Optical reproduction of sounds from negative phonograph cylinders, in: Contributions to the Fifth International Conference on Optics Within Life Sciences OWLS V, Crete, Springer, 1998.
[11] O. Johnsen, F. Bapst, et al., Visual audio: an optical technique to save the sound of phonographic records, in: Joint Technical Symposium, 2004.
[12] V. Fadeyev, C. Haber, Reconstruction of mechanically recorded sound by image processing, JAES 51 (12) (2003) 1172–1185.
[13] Light Green & Light Blue Projects, 2E1366 Project Course in Signal Processing and Digital Communication, KTH, Stockholm, Sweden, http://www.s3.kth.se/signal/edu/projekt/students/03/, 2003.
[14] S. Stotzer, O. Johnsen, et al., Groove Extraction of Phonographic Records. DAS 2006, Seventh IAPR Workshop on Document Analysis Systems, Nelson, New Zealand, 2006.
[15] S. Stotzer, O. Johnsen, et al., Phonographic sound extraction using image and signal processing. Acoustics, speech and signal processing, 2004. Proceedings (ICASSP '04), vol. 4, no. 5, 2004, pp. 289–292.
[16] V. Fadeyev, C. Haber, C. Maul, J.W. McBride, M. Golden, J. Audio Eng. Soc. 53 (6) (2005) 485–508.
[17] J. Walton, Pickups: The Key to Hi-Fi, Sir Isaac Pitman and Sons, Ltd., London, 1965, p. 6.

Biographies

Peter J. Boltryk received his MEng and PhD degrees in mechanical engineering from the University of Southampton in 2000 and 2004, respectively. After a postdoctorate position at the School of Electronics, University of Southampton, he is currently a research fellow for the Electro-Mechanical Research Group in the School of Engineering Sciences, University of Southampton. He is an associate member of the IMechE.

Martyn Hill is reader in Electromechanical Systems and deputy head of the School of Engineering Sciences, Southampton University. He graduated in 1985 from the Institute of Sound and Vibration Research, was appointed lecturer in 1990 and has research interests in acoustics (particularly ultrasonics) and in the measurement and characterisation of surfaces. His work has led to more than 100 technical publications, he is a chartered engineer, a member of the Institute of Measurement and Control and a member of the Institute of Acoustics.

John W. McBride received a degree in aeronautical engineering from the University of Southampton in 1978. In 1986, he received a PhD for his work at Plymouth University with a thesis on Electrical Contact Phenomena. From 1985 to 1987, he lectured in the Mechanical Engineering Department at Plymouth University and from 1987, he has been a lecturer, senior lecturer and now professor of electro-mechanical engineering in the School of Engineering Sciences at the University of Southampton. He is Chair of the Electro-Mechanical Research Group, and head of research in the school (2001–2005). His main research interests include electrical contacts, metrology and instrumentation.

Antony Nascè received his BEng degree from the Institute of Sound and Vibration Research at Southampton University, in acoustical engineering. His current PhD research involves the development of image and audio signal processing techniques for the recovery of sound from non-contact surface scans of early mechanical recordings.