

Department of Computer Science
University of Fribourg, Switzerland

PHONOGRAPHIC RECORD SOUND EXTRACTION BY IMAGE PROCESSING

Thesis

Submitted to the Faculty of Science, University of Fribourg (Switzerland)
to obtain the degree of Doctor Scientiarum Informaticarum

Sylvain STOTZER

from

Büren an der Aare (BE) and Geneva

Thesis N° 1534
Imprimerie St-Paul, Fribourg
2006

Accepted by the Faculty of Science of the University of Fribourg (Switzerland), on the proposal of:

Prof. Béat Hirsbrunner, University of Fribourg, Chairman
Prof. Rolf Ingold, University of Fribourg, Thesis Supervisor
Prof. Martin Vetterli, Swiss Federal Institute of Technology Lausanne (EPFL)
Prof. Ottar Johnsen, University of Applied Sciences of Fribourg
Prof. Frédéric Bapst, University of Applied Sciences of Fribourg

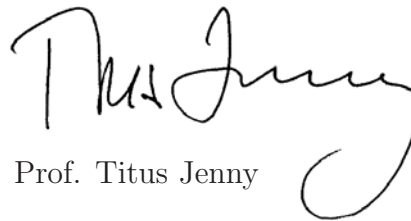
Fribourg, September 26, 2006

The Thesis Supervisor:



Prof. Rolf Ingold

The Dean:



Prof. Titus Jenny

Acknowledgements

I would personally like to acknowledge all those who helped me in the development and completion of this research.

First, I would like to thank my thesis advisors, Professors Ottar Johnsen, Rolf Ingold and Frédéric Bapst for their guidance in my studies and my work. They shared with me their expertise and offered fruitful discussions and valuable insights into the different topics required for this research. Their motivation and support was very helpful.

I want to thank Professor Martin Vetterli for accepting to be my fourth committee member. I was very pleased that he accepted to read my work.

Many famous inventions and discoveries started with a crazy idea. I would like to acknowledge Stefano Cavaglieri and Pio Pelizzari, who got the idea to take pictures of records to archive them and who shared this idea with us.

Many people took part in this research project. I am particularly grateful to Cédric Milan and Christoph Sudan for their involvement and their achievements in developing the hardware tools used in this work. Many thanks also to Thierry Fumey for sharing with us his expertise in photography. Finally I would like to thank all the students, assistants and professors who worked on the VisualAudio project for semester work, diploma or bachelor thesis.

My colleagues at both the Ecole d'ingénieurs et d'architectes and the University of Fribourg also offered interesting discussions and helped with practical matters, for which they all deserve thanks.

This research would not have been possible without the support of all my family. I particularly thank my wife, Ula, for her patience, understanding and encouragement. I would never have been able to achieve this work without her love and unconditional support. And finally, I would like to thank my daughter Sarah who let her father work during long hours and who learned very early to say "Dad's gone to work" ...

Abstract

The phonographic record was the only way to store sounds until the introduction of magnetic tape in the early 50's. Therefore there are huge collections of phonographic records, for example in radio stations and national sound archives. Such archives include pressed discs, which were produced in mass by record companies for commercial distribution, as well as direct cut discs obtained by the direct recording with often a great cultural value and available only as single copies. Many records are deteriorating with time. Worse, many records are in an advanced stage of decay and would be destroyed by the movement of the stylus from even the best turntables. Thus, we risk losing an important cultural heritage, if no alternative playback system is developed.

In record players, a needle follows the position of the groove and converts it into an electrical signal corresponding to the sound. This means that the radial displacement of the groove contains the sound information. Thus the sound information is visible and it is then possible to extract it by image processing techniques. These observations lead to the VisualAudio project, which proposes an optical system to extract the sound from phonographic records. The VisualAudio concept proposes first to take the record in picture, in order to have a photographic copy of the record and of the sound. This photographic sound copy can be stored for long term archiving. Then the film is digitized using a specially designed rotating scanner, and the image is processed in order to extract the recorded sound. Thus this system can play records without deteriorating them and it is also able to play severely damaged records.

This work focuses on the image processing parts of the VisualAudio project. The image acquisition system is thoroughly studied to understand the image formation process as well as all kinds of degradations, which may affect the final sound quality. Based on this analysis, a groove model is proposed, in order to develop dedicated image extraction and signal correction methods. The whole system is then evaluated to point out the strengths and weaknesses of the VisualAudio sound extraction process.

Résumé

Le disque phonographique était le principal support utilisé pour enregistrer et conserver du contenu sonore jusqu'à l'introduction de la bande magnétique au début des années 50. Ainsi il existe actuellement d'importantes collections de disques phonographiques dans les archives nationales, les stations radio et les phonothèques. Ces archives possèdent des disques pressés, produits de façon industrielle pour la distribution commerciale, ainsi que des disques gravés, obtenus par l'enregistrement en direct de conférences ou d'émissions radios. Les disques gravés sont ainsi des copies uniques, qui ont souvent une grande valeur culturelle. Mais beaucoup de disques se détériorent avec le temps, et certains sont tellement dégradés, qu'ils se décomposeraient au seul contact d'une aiguille de tourne-disque. Nous risquons ainsi de perdre une importante partie de notre patrimoine culturel, si nous ne trouvons pas un nouveau système de lecture pour ces disques.

Dans les tourne-disques, une aiguille suit le parcours du sillon le long du disque et convertit le déplacement latéral en un signal électrique correspondant au son. Cela signifie que le déplacement latéral du sillon contient l'information sonore. Ainsi le contenu sonore est visible et il est donc possible d'extraire le son par des techniques de traitement d'image. Ces observations sont à la base du projet VisualAudio, qui propose une méthode optique pour extraire le contenu des disques phonographiques. L'idée du projet VisualAudio consiste à prendre tout d'abord une photographie du disque, afin d'en conserver une copie pour un archivage à long terme. Ce film est ensuite numérisé à l'aide d'un scanner rotatif conçu spécialement dans le cadre de ce projet, et l'image numérique est analysée afin d'en extraire le contenu sonore. Ainsi VisualAudio permet d'extraire le son des disques sans contact avec leur surface. Ce système est également capable d'extraire le son de disques gravement endommagés.

Ce travail de thèse se concentre sur la partie traitement d'image du projet VisualAudio. Le système d'acquisition d'image est tout d'abord analysé en détail, afin de comprendre le processus de formation d'image ainsi que tous les types de dégradations qui peuvent affecter la qualité de l'image et du son extrait. Sur la base de cette analyse, nous proposons un modèle de sillon qui est utilisé pour développer les méthodes d'extraction du son et de correction du signal. Le système est ensuite évalué afin de comprendre les avantages et inconvénients du système de lecture et d'archivage VisualAudio.

Contents

1	Introduction	1
1.1	The VisualAudio project	1
1.2	Objectives of this thesis	4
1.3	Related works	5
1.3.1	Sound on film	5
1.3.2	Signal extraction by image processing	6
1.3.3	Phonographic optical playback systems	8
1.4	Overview of this dissertation	11
2	Phonographic recording	13
2.1	History	13
2.2	Mechanical recording	14
2.3	Recording process	16
2.4	Disc manufacturing	16
2.5	Record materials	17
2.5.1	Shellac	17
2.5.2	Acetate/lacquer	17
2.5.3	Vinyl	18
2.6	Recording speed	18
2.6.1	78 rpm	18
2.6.2	33 rpm	19
2.6.3	16 rpm	19
2.6.4	45 rpm	19
2.7	Record equalization	20
2.7.1	Recording characteristics	20
2.7.2	Playback characteristics	20
2.7.3	Equalization	20
2.7.4	Equalization transfer function: the RIAA case	22
2.8	Reading distortions and undesirable effects	23
2.8.1	At recording and manufacturing	23
2.8.2	At playback	24
2.8.3	Record storage and maintenance	25
2.9	Record and groove geometry	25
2.10	Dynamic range and signal to noise ratio	27

3	Acquisition system	29
3.1	Films	29
3.1.1	Film constraints for VisualAudio	29
3.1.2	Film structure	30
3.1.3	Sensitometry / response of a photographic film	30
3.1.4	Exposure time	32
3.1.5	Choice of the film	32
3.1.6	Storage and life expectancy	34
3.2	Camera	34
3.2.1	Camera construction	34
3.2.2	Illumination system	36
3.3	Scanning	37
3.3.1	Basic structure of the scanner	37
3.3.2	Short history	38
3.3.3	The 2004 scanner	39
3.3.4	Sampling process	40
4	Image analysis	43
4.1	Image formation and definitions	43
4.2	Blur and resolution	44
4.2.1	Shading blur	45
4.2.2	Camera resolution	45
4.2.3	Film resolution	46
4.2.4	Scanning optics	46
4.2.5	Motion blur	48
4.2.6	Sampling blur	49
4.2.7	Total system resolution	50
4.2.8	Effect of the blur on the tangential direction	51
4.2.9	Effect of the blur on the radial direction	51
4.2.10	Spatial variance of the system	51
4.3	Noise and local degradations analysis	52
4.3.1	Record	52
4.3.2	Records with shrinkage of the recording layer	53
4.3.3	Film and picture taking	54
4.3.4	CCD camera	55
4.3.5	Signal to noise considerations	56
4.4	Illumination variation	60
4.4.1	Radial variations	60
4.4.2	Indirect light	64
4.5	Nonlinear distortions	67
4.5.1	Nonlinearity of recording media	67
4.5.2	Geometrical distortions	67
4.5.3	Off-axis and pitch variation	73
4.6	Groove model	76
4.7	Conclusion	78

5	Groove extraction	81
5.1	Groove extraction algorithm	81
5.1.1	Basic properties of the groove image	81
5.1.2	Algorithm overview	85
5.2	Edge detection	86
5.2.1	Literature review	87
5.2.2	Edge detection algorithm	89
5.2.3	Coarse edge detection	90
5.2.4	Fine edge detection	91
5.2.5	Edge detection error	95
5.3	Groove reconstruction	97
5.3.1	Trace following	97
5.3.2	Reading way	102
5.3.3	Single and double trace grooves	102
5.3.4	Trace joint	103
5.3.5	Bypassing the ring structure	104
5.4	Extraction of an entire record	105
5.5	Conclusion	106
6	Signal restoration	107
6.1	Acquisition artifacts	107
6.1.1	Camera calibration	107
6.1.2	Pitch variation and wow correction	110
6.2	Deblurring	114
6.3	Local damage correction	115
6.3.1	Existing methods	116
6.3.2	Image smoothing	117
6.3.3	Corrupted pixel map	119
6.3.4	2-passes trace extraction	125
6.3.5	1D impulses detection	126
6.3.6	LMS signal fitting	127
6.3.7	Signal reconstruction	127
6.4	Shrinkage of the recording layer	129
6.4.1	Proposed correction	129
6.4.2	Lost points collection system	130
6.4.3	Detect the cuts	130
6.5	Conclusion	131
7	Sound processing	133
7.1	Stored sound reconstitution	133
7.1.1	Sound extraction in mono and stereo	134
7.1.2	Derivation	134
7.1.3	Low frequencies removal	135
7.2	Recorded sound recovering	136
7.2.1	Removal of high frequencies	137

7.2.2	Pre-amplification equalization	137
7.2.3	Equalization combined with derivation	139
7.3	Output file format	139
7.3.1	Sampling rate	140
7.3.2	Bit depth	141
8	Evaluation	145
8.1	Measures and tests protocols	145
8.1.1	Tests records	145
8.1.2	Acquisition naming conventions	146
8.1.3	SNR (Signal to Noise Ratio)	147
8.1.4	THD (Total Harmonic Distortion)	147
8.1.5	STD (Standard Deviation)	148
8.1.6	Measurement domain	149
8.1.7	Parameters	150
8.2	Evaluation of the acquisition process	151
8.2.1	Acquisition regularity	151
8.2.2	Optics	151
8.2.3	Opening time	153
8.2.4	Oversampling	156
8.2.5	Light intensity and grey levels dynamic	158
8.2.6	Quality variation over a record acquisition	159
8.3	Image processing evaluation	165
8.3.1	Camera calibration	166
8.3.2	Edge detection	167
8.3.3	Corrections	171
8.4	Mass testing	172
8.5	Processing time	175
8.5.1	Time optimization	176
8.6	Sound quality evaluation	177
8.6.1	Bit depth	177
8.6.2	Signal to noise ratio evaluation	178
8.7	Global evaluation	179
9	Conclusion, perspectives and discussion	181
9.1	Contributions	181
9.2	Perspectives	182
9.2.1	Frequency and time domain detection and correction	182
9.2.2	Multichannel correction	183
9.2.3	Discs with shrinkage of the recording layer	183
9.2.4	Film	183
9.2.5	Direct reading	183
9.3	Discussion on the VisualAudio project	184
9.3.1	Photographic film	184
9.3.2	Archiving	185

9.3.3 Optical playback 186

List of Figures

1.1	The VisualAudio concept	3
1.2	Lofargram	8
2.1	Phonautograph	14
2.2	Stereo recording	15
2.3	Records top views	15
2.4	Records profile views	18
2.5	RIAA pre and de-emphasis curves	23
2.6	Acetate record with shrinkage of the recording layer	26
3.1	Photographic D-log E curve	31
3.2	Camera design	35
3.3	Light reflection angle	36
3.4	Acquisition part of the VisualAudio scanner	38
3.5	Transmission system of the scanner	39
3.6	Scanner acquisition	41
4.1	Light reflection on the groove	44
4.2	Sample of a groove acquisition	45
4.3	Picture of record with scratch	53
4.4	Record with fungus	53
4.5	Samples of image degradations	54
4.6	Disc with shrinkage of the recording layer	54
4.7	Pepper fog	56
4.8	Noise standard deviation vs signal to noise ratio	59
4.9	The light reflected to the lens by an unmodulated groove	62
4.10	Groove cut views of the light reflection	63
4.11	Top view of the record and light source	64
4.12	Indirect lightening	65
4.13	Indirect light simulation	65
4.14	Indirect light on a photography	66
4.15	Light section illuminating an unmodulated groove	67
4.16	Planes definition for the scanning	68
4.17	Camera shift along the Y-axis	69
4.18	Camera rotation in the XZ-plane	70
4.19	Aligned and misaligned camera	71

4.20	Sound signal with a misaligned camera	71
4.21	Film off-axis	74
4.22	Sampling with film off-axis	75
4.23	Basic luminance model	77
4.24	The groove model	78
4.25	Linear and non-linear part of the transition	79
5.1	The trace and groove width show important variations	83
5.2	Asymmetric traces	84
5.3	Ring images gradients	84
5.4	Pseudo-code of the groove extraction algorithm	86
5.5	Samples of groove intensity profiles	87
5.6	Line of acquisition	89
5.7	Coarse scale edge detection	91
5.8	Edge detection by local thresholding	92
5.9	LMS edge detection: point selection	93
5.10	LMS edge detection: point selection	94
5.11	Candidate selection	98
5.12	Ranges around the traces	100
5.13	Orientation of the record spiral on the acquired image	102
6.1	Acquisition performed with non-homogeneous illumination	109
6.2	Acquisition performed with homogeneous illumination	109
6.3	spectra for wow correction	113
6.4	Corrupted pixel map	119
6.5	Small degradations map	122
6.6	Large degradations map	123
6.7	2-passes trace extraction	125
6.8	Shrunked record's acquisition zoom	130
6.9	Magnified view of a 40×40 pixels ring image acquisition	131
7.1	Frequency response of the digital differentiator filter	135
7.2	Analog and digital RIAA de-emphasis curves	138
7.3	Digital combined (derivative and RIAA) filter response	140
7.4	Multistage resampling	141
8.1	Measures of the <i>STD</i> between consecutive samples	149
8.2	Normalized difference between two consecutive acquisitions	152
8.3	Spectra of signals acquired with different lenses	154
8.4	<i>STD</i> ₁ measured with different opening times	155
8.5	<i>STD</i> of unmodulated grooves with different opening times	155
8.6	Spectrum of acquisitions with varying opening times	156
8.7	Spectrum of 78 rpm acquisitions with different scanning frequencies	157
8.8	Spectrum of 33 rpm acquisitions with different scanning frequencies	158
8.9	<i>STD</i> of unmodulated grooves with different scanning frequencies	159
8.10	<i>SNR</i> ₁ and <i>THD</i> ₁ measured with different light levels	160

8.11 SNR_c measured on twelve groove circumvolutions	161
8.12 SNR_q measured over six consecutive groove circumvolutions	162
8.13 Quality of the top and bottom edges	163
8.14 Pepper fog on the inner and outer tracks	165
8.15 Spectrum of the inner and outer tracks of a 33 rpm	166
8.16 STD measured on the inner and outer tracks	167
8.17 Spectrum of the inner and outer tracks of a 78 rpm	168
8.18 Acquisition of a sweep with different edge detection methods	171
8.19 Spectra for different methods of signal correction	173
8.20 Shrinked record, which content was extracted with VisualAudio	173
8.21 Mould record, which content was extracted with VisualAudio	174
8.22 Local view of the acquisition of a record affect by mould	175

List of Tables

1.1	UNESCO Survey of Endangered Audiovisual Carriers	2
2.1	Audio recording history	14
2.2	Equalization chart for 78 rpm records	21
2.3	Equalization chart for 33 rpm records	22
2.4	Record characteristics	26
2.5	Groove geometry	27
2.6	Reading stylus properties	27
2.7	Wavelength on the record	28
2.8	Dynamic and frequency ranges	28
2.9	Signal to noise ratio performance	28
3.1	Groove width on the film	32
3.2	Films gamma and edge standard variation	33
3.3	Comments about the films	34
3.4	Light reflection angle	37
3.5	Image and audio sampling frequencies	41
3.6	Effective sound sampling frequencies	42
3.7	Integrated areas according to the radial position	42
4.1	Properties and resolution of the 420 mm lens	46
4.2	Scanning blur with $DOF = 10 \mu m$	47
4.3	Scanning blur with $DOF = 30 \mu m$	47
4.4	Sampling blur with a $10\times$ magnification optics	49
4.5	Sampling blur with a $4\times$ magnification optics	50
4.6	Maximum noise standard deviation allowed	60
4.7	Variation of the SNR and THD for camera rotations in the XY-plane	73
7.1	Effective audio sampling frequencies	140
7.2	Resampling to 44.1 kHz	142
7.3	Resampling to 48 kHz or 96 kHz	142
8.1	Description and references of the records used for the evaluation . . .	146
8.2	Description of the various tracks used for the evaluation	146
8.3	Measurement domains for the SNR , THD and STD	150
8.4	Real magnification of the lenses	153

8.5	SNR and THD measurements for signals acquired with different lenses	153
8.6	Peak and harmonics of signals acquired with different lenses	153
8.7	SNR_1 measured on acquisitions performed using different scanning frequencies	157
8.8	STD_1 measured on the four edges of a groove	163
8.9	SNR_{total} for the outer and inner edges	164
8.10	STD_1 for the outer and inner edges	164
8.11	Harmonics for the outer and inner edges	164
8.12	SNR_1 measured on the inner and outer tracks	165
8.13	THD_1 measured on the inner and outer tracks	166
8.14	STD_1 measured on the inner and outer tracks	166
8.15	SNR with and without camera calibration	167
8.16	Edge detection methods used in the current evaluation test	169
8.17	SNR_{total} with different edge detection methods	169
8.18	SNR_1 with different edge detection methods	169
8.19	THD_{total} with different edge detection methods	169
8.20	STD_1 with different edge detection methods	170
8.21	Acquisition statistics using different edge detection methods	170
8.22	SNR_{total} with different methods of signal correction	172
8.23	SNR_1 with different methods of signal correction	172
8.24	THD_{total} with different methods of signal correction	172
8.25	Ratio of the recorded sound duration to the processing and scanning time	176
8.26	SNR_1 measured on acquisitions with different bit depth	177
8.27	Signal to noise ratio measured according to the NAB standard	178
8.28	Comparison of the signal to noise level	179

Chapter 1

Introduction

1.1 The VisualAudio project

The sound is propagated in the space by air pressure variations. Thus in the first phonographic sound recorders, a diaphragm vibrated to these changes in air pressure. This diaphragm was connected to a cutting stylus and the diaphragm's movements were then written as a modulated groove on a record. At playback, a needle followed the groove and transmitted its movements back to the diaphragm, which produced air pressure variations again.

The phonographic recording technology was introduced in the late nineteenth century and was then widely used during the twentieth century. Thus sound archives now own large collections of phonographic records, including direct cut discs, which were used to record speeches and conferences as well as to edit and mix sound recordings for radio stations, and pressed records, which were produced in mass for commercial distribution. Sound archives face numerous problems with phonographic records. To get an overview of these problems, the IASA (International Association of Sound and Audiovisual Archives) carried out a "Survey of Endangered Audiovisual Carriers" in 2003 at the request of the UNESCO (United Nations Educational, Scientific and Cultural Organization). They asked 2093 institutions in 184 countries to classify their audio carriers in three categories. This survey, which results are summarized in Table 1.1, is not intended to be an accurate, scientific piece of research; but it gives an interesting picture of the various audio carriers rate of decay [1]. This survey confirms the conclusions from practical experience and from past surveys: acetate direct cut discs are the audio recordings most at risk. This risk is compound by the fact that the vast majority of acetates are unique recordings, and that substantial numbers of acetate discs are being lost each year because the final stage of the decay is unpredictable and catastrophic. Thus there is an urgent need to transfer the content of all these records on a new audio carrier. Moreover, decaying records are no more playable with usual turntables and require a contactless reader and an automated system to follow the groove in case of important surface damages.

Digitizing records with high quality turntables is a satisfying solution to preserve the content of many records that are still in good condition; but the transfer time

Audio carriers	Nb of institutions / Nb of items	In good con- dition	Giving some concerns	Obviously decaying
Cylinder recordings	20 / 43965	14.65 %	58.73 %	26.62 %
Shellac discs	41 / 614935	95.06 %	4.93 %	0.02 %
Direct cut discs	23 / 60332	2.84 %	35.03 %	62.13 %
Vinyl discs	55 / 1855120	88.43 %	11.56 %	0.01 %
Magnetic tapes	49 / 2161941	76.94 %	21.28 %	1.78 %
Recordable CDs	48 / 193062	86.94 %	9.95 %	3.11 %
Audio CDs	52 / 1128400	95.36 %	4.58 %	0.06 %
R-DAT digital tapes	29 / 198477	45.40 %	27.85 %	26.75 %

Table 1.1: Rate of decay classification for some audio carriers, based on the UNESCO "Survey of Endangered Audiovisual Carriers". These results must be considered with caution, given the low rate of response, the loosely defined categories, the variable number of items and some results variations with former surveys [1].

vary considerably according to the record's state of conservation. According to the IASA "Guidelines on the Production and Preservation of Digital Audio Objects", the average transfer time for a 3 minutes 78 rpm record is of 45 minutes, including the time to find the correct settings for the equipment and the choice of the reading stylus based on the recording analysis [2]. But while a complex transfer may easily take 20 hours, the mass saving of records in good conditions with experienced people take only around 10 minutes per face, according to the Swiss National Sound Archives. However, digitizing analog audio recording (such as phonographic discs) is a controversial topic: the analog to digital transfer is lossy and people always debate to know what are the minimal requirements (in terms of bit depth and sampling frequency) for a digitizing to restore a perfect sound [3]. This debate incites some sound archivists to wait for newer, more powerful technologies for digitizing, at the risk of getting some of their records decaying with time. Another interesting conclusion of the UNESCO survey is that the most endangered carriers are not necessarily the oldest: as presented in Table 1.1, some of the digital media used in the last twenty years give more concerns than the shellac records produced between 1890 and 1950 [1].

To bypass the analog/digital debate, Stefano Cavaglieri, who works at the Swiss National Sound Archives, wanted to keep an analog copy of a record. Since the groove modulations are visible on the record surface, he proposed to take a photograph of a disc, in order to keep an analog copy of the recorded sound. The problem was then to be able to playback the photograph to reproduce the sound. This idea gave birth to the VisualAudio concept, which is presented in Figure 1.1 and consists in three steps [4, 5, 6]:

1. An analog picture of each side of a disc is shot. The photographic film must have a high spatial resolution and be as large as possible (about 1:1), in order to catch the finest details of the groove. This process can be done quickly. The film is cheap, and can be stored for a long time (more than 100 years). That way, the sound information is preserved in case the original discs deteriorate.
2. When one wants to recover the sound, the film can be digitized using a specially designed rotating scanner.

3. The sound must then be extracted from the digital image. This requires image processing techniques to extract the radial displacement of the groove, which contains the sound. Additional processing are applied to detect cuts and to correct other defects. Digital signal processing must be applied to the groove signal to extract the sound.

The VisualAudio process is contactless, and can then be used on any records, even broken or decaying. But the photographic sound storage system presents also some other advantages. On an optical record reader, the groove position must be evaluated very accurately, requiring high image resolution that can be reached only with a microscope lens. Unfortunately the groove depth exceeds the microscope depth of field, and many discs have some warping which is even more important than the groove depth. Using the intermediate photography step allows us to work with a larger depth of field while imaging the disc, but ensures that the image to be digitized (the film) fits in the reduced depth of field required by the microscope's optics. Time is also critical issue in such an archiving system: the challenge is to save a large amount of records quickly, before their complete physical destruction. Taking a picture of the discs is a quick way to store a copy of the sound content in its current stage of conservation. The sound extraction could then be done on demand, without time pressure.

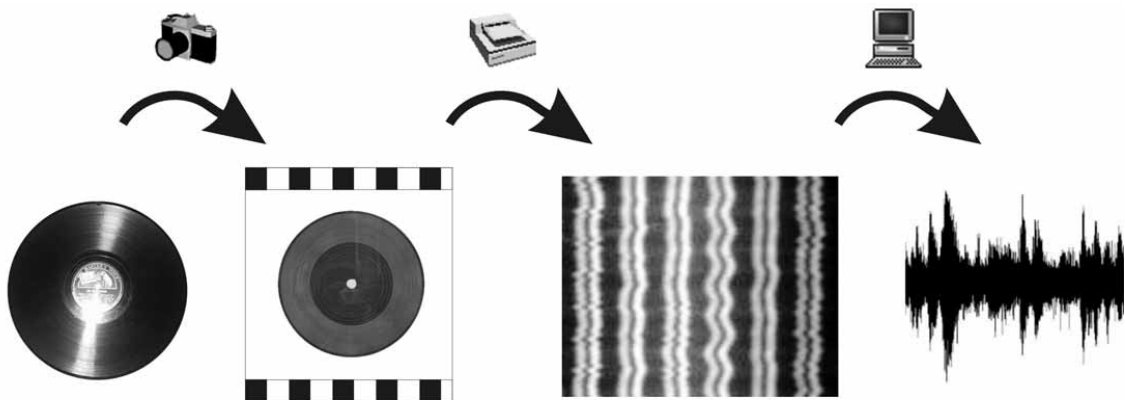


Figure 1.1: The VisualAudio concept.

Thus the main advantages of the VisualAudio system compared to conventional turntables can be summarized as follows:

- The record is preserved with an analog copy (the film).
- The analog copy can be stored for long term archiving.
- No need to use an autofocus system as the photography depth of field is much larger than the microscope lens depth of field.
- The archiving time (time to take the photograph) is low and thus well adapted for mass archiving.

- The archiving time is considerably lowered for damaged records.
- The optical playback system is not intrusive and does not damage records.
- The optical playback system is able to read records that are currently considered as unreadable with conventional turntables.

Based on this concept, three diploma works have been performed at the University of Applied Sciences of Fribourg: [7, 8, 9]. During these works, the students took a few high resolution pictures of records and built a first scanner prototype to digitize these photographic films. They were able to extract some sound samples out of these images, and thus they showed the feasibility of the VisualAudio concept.

Based on these preliminary works, four objectives were defined for the VisualAudio project in order to build a fully operational system:

1. Efficiently transform a collection of discs into a film archive.
2. Scan the films with high resolution.
3. Accurately extract the groove position out of the image.
4. Extract good quality sound from record pictures.

To reach the first two objectives, new hardware have been designed. A dedicated camera has been built to match the specific requirements for disc pictures and a new scanner has been developed, based on the flaws and knowledge acquired with the previous prototype. These hardware devices are fully described in [10] and will be further presented in this work.

The last two objectives are the main topics of this thesis and are detailed in the following section.

1.2 Objectives of this thesis

The current work focuses on the image processing and sound extraction parts of the VisualAudio project, in order to get the best and most faithful sound quality. Extracting the sound from photographs of disc records is an original process which has not been studied up to now. Moreover, and to the best of our knowledge, the VisualAudio was the first project which uses 2D image processing to extract sound from phonographic records. Thus this thesis presents a new approach, which first requires a theoretical analysis. Therefore the first objective of this work is to fully analyze the image acquisition system, which includes the photography and scanning processes. This analysis encompasses the following steps:

- Investigate the phonographic record technology and point out the groove characteristics.
- Analyze the resolution and the distortions produced by the picture taking stage, including the effect of the light source and the optics.

- Study the resolution and the properties of photographic films.
- Analyze the distortions produced at scanning. This includes the effect of the light source, the optics as well as geometric distortions and mechanical perturbations.

This theoretical analysis points out the VisualAudio system's limitations and establishes the necessary foundations to design the operational camera and scanner prototypes. It is also used to develop a groove model, which is the base for the image processing algorithms. The resulting model must represent the evolution of the groove information from the physical disc to the extracted sound.

Image processing algorithms are then developed to extract the groove accurately and to correct the degradations, based on the groove model and on the theoretical analysis. Denoising and sound restoration are well known research domains, as presented in [11, 12, 13, 14], and several high quality denoising software and hardware are commercially available [15, 16]. Therefore, this work doesn't focus on classical sound restoration methods, which consider the sound as a one dimensional signal; but it tries to concentrate on the specific part of this sound extraction application, which is the two dimensional information provided by the image. We try to reduce the effect of the damages, noise and uncertainty using the contextual information available on the image.

The final objective is to perform practical experiments with the available camera and scanner, in order to validate the model. Mass testing must then be performed to validate the process for many different records showing different types of degradations. The final sound quality depends on the record stage of conservation. Therefore the aim is to reach good sound quality for records in good conditions and to get understandable speeches for transcription. These experiments will then determine the reachable quality and usability of the VisualAudio system.

1.3 Related works

The VisualAudio project proposes to make analog copies of old records by the mean of photography, in order to read these records contactless. Thus this process basically consists in two things: to store sound on a film and to extract a signal out of an image. Several works have already performed in these two fields: Subsection 1.3.1 describes some existing applications which store sounds on film and Subsection 1.3.2 provides an overview of applications, which aim is to accurately extract a signal out of an image. Subsection 1.3.3 describes the existing contactless optical techniques that have been proposed up to now to read mechanically recorded sounds.

1.3.1 Sound on film

Efforts to store sounds by the mean of photography have already been performed a long time ago: in 1878, E. W. Blake published a paper on "A Method of Recording Articulate Sounds by Means of Photography". He photographed the vibrations of a

microphone diaphragm by means of a mirror which reflected a beam of light. In 1902, W. Duddell filed a patent application covering a method of variable area recording and reproducing, under the title of "An Improved Phonograph". These were the beginning of the sound-on-film technology which will later be used for the movie soundtracks. Two techniques were basically used for optical soundtracks: variable density and variable area. For the variable density, a shutter was used to vary accurately the amount of light that reached and exposed a moving film. Variable area tracks were generated by exposing the film with a constant light source and a gate whose aperture was modulated by the audio signal. Variable area became the standard in the movie format since the 1930's [17, 18, 19, 20].

The width of the soundtracks recorded area is 1.9 mm and the film moves at 0.8 meter per second. For a quick comparison with a phonographic record in terms of resolution, the moving speed of the film soundtrack is about twice the speed of the outermost groove of a 78 rpm record and the amplitude of the audio signal is twelve times larger than on a record. This means that the soundtracks reading system requires less accuracy than needed for the VisualAudio project, and that soundtracks are also less affected by the noise produced by the graininess of the film.

Digitizing of soundtracks leads to image processing techniques, which are similar to the ones presented in the present work: a sound signal is extracted out of an image containing modulated edges. For example, Technicolor proposes an image restoration technique for movie soundtracks, resulting in better audio reproduction quality than usual reading and sound restoration techniques [21].

It should be noticed that the films used for recording variable area optical soundtracks present the same properties than the films chosen for the VisualAudio project: black and white orthochromatic films with a high contrast, for example: Agfa Sound ST9 or Kodak EASTMAN EXR Sound Recording Film 2378.

1.3.2 Signal extraction by image processing

A 1D signal can be basically stored either in the grey level variation of a 2D image or in the modulation of a line over the image. Several applications need to extract such a signal out of an image. The next two paragraphs provide two examples: bar code reading, where the signal is extracted from the grey level variations of a 1D image, and Lofar lines detection, where the signal is stored in the modulation of a line on a 2D image. These examples also show the importance of the a priori knowledge of the acquired image and the contained signal properties.

1.3.2.1 Bar codes

The bar code reading is a typical application, which needs to extract a signal stored in the varying grey levels of a 1D image. Bar codes are a finite sequence of parallel light and dark lines of variable width, where the information is encoded in the width of the dark stripes. The basic structure of bar codes contains start and stop patterns, as well as one or two check characters. Thus the bar code reader must retrieve the

width of the black lines in order to reconstruct the original code. Knowledge of the bar code basic structure also provides a great help to read and validate the extracted signal.

Standard decoding techniques are based on classical edge detectors, such as zero crossings of the second derivative. Using the detected edges, the width of each stripe is estimated separately to find the narrowest width, and then to map the ratio of all stripe widths over the narrowest into an integer sequence [22]. While the standard supermarket scanners are equipped with mirrors and retrieve signals from multiple angles, handheld scanners are much more subject to image degradation: images may be affected by the ambient light or blurred, depending on the distance between the reader and the surface where the bar code appears. Thus there was a need for more powerful recognition algorithms, using some more a priori knowledge of the signal to extract. Fortunately, bar codes are functions that have a very specific form: square signals with constant amplitude. Knowledge of this form is highly useful to elaborate more sophisticated techniques.

Esedoglu estimated the unknown parameters of its deblurring kernel using global information contained in the observed signal: the square pulse form as well as the constant amplitude property of the bar code. His model is well suited for recovering bar codes from very blurred and noisy images [23]. Joseph et al. based their recognition algorithm on peak detection instead of edge detection. They stated that for the specific case of blurred bar codes images, the waveform peaks are a more reliable feature than edge information, due to their tolerance to convolution. In case of blur, while the edges from close stripes will interact, the peaks will not be affected [24].

1.3.2.2 Lofar lines detection

Lofar is an acronym for Low Frequency Analysis and Recording. The objective of passive sonar system is to detect the presence of signals emitters in underwater acoustic fields. Such systems record acoustic signals using an array of hydrophones. These signals are beamformed, spectrum analyzed and result in an image of frequency power versus time, commonly referred as lofargram. Signal emitters are characterized by fluctuating curves in the lofargram, which are called spectral lines. By extracting the spectral lines on a lofargram, it is possible to determine the acoustic source of the sound. But there is a tradeoff involved in this processing: lofargram images are very noisy, but the lines must be detected with very fine structure to get a good localization. Di Martino et al. presented a three steps algorithm to detect Lofar lines [25]:

- Edge detection
- Region location
- Line tracing process

As the images are very noisy, the edge detection step is performed on highly filtered images to ensure a good detection, at the expense of a poor localization. This allows detecting regions of interest, where an accurate line tracing process is further applied.

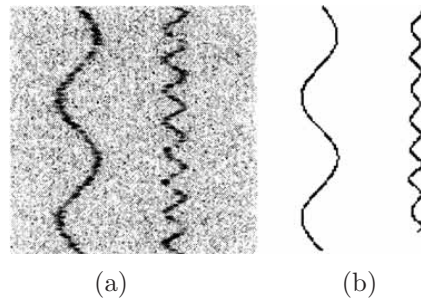


Figure 1.2: (a) Lofargram image containing two spectral lines and (b) extracted spectral lines [25].

The line tracing and gap bridging process is based on the low frequency particularity of the signals: the curvature presents only slow variations and continuous segments are close to each other. A lofargram and the extracted patterns are displayed on Figure 1.2.

1.3.3 Phonographic optical playback systems

The first sound recording and reproducing methods consist in mechanically engraving a modulated groove in a storage medium, which can be either a cylinder or a disc. However, the mechanical playback may damage some old fragile carriers. Moreover, many mechanically recorded media are subject to physical degradations and are therefore no more playable. Thus there is a real need for optical playback techniques. Such techniques can be classified in three categories: the methods involving an optical device to follow the groove, the 2D image processing methods and the 3D techniques which acquire a surface map of the recorded medium, where the depth information can also be used to extract the sound.

It should be noticed that all of the below mentioned projects directly scan the recorded medium, extract the sound from the image and output a digital sound file. None of them provides an intermediary analog storage medium such as proposed in VisualAudio.

1.3.3.1 Optical stylus

In a first category of systems, a mechanical device is used to follow the groove, but with a much lower weight than a pickup stylus. This low mass stylus reflects light out of the groove to a measurement tool, and the sound is then extracted from the measured position of the light spot. Poliak used an optical fiber to follow the groove from old records [26]. He reached good quality results: for modern records, the background noise is of the same order of magnitude than with a classical system. But operator's interventions are still needed to position the optical fiber on crackled records and to clean the dust that pile on the optical stylus. Poliak's optical fiber system is currently in use at the Swiss National Sound Archives and the Radio Suisse Romande. Petrov also worked with an optomechanical method, using

a laser interferometer system to read Edison cylinders [27]. Working at low rotation speed required consequent mechanical and anti-vibrations devices, but resulted in important noise reduction.

In a second category of project, the optical stylus is contactless. The Syracuse University Library Radius Project developed an optical playback system for cylinder recordings using a light beam instead of a mechanical stylus: an interferometer gets a precise measurement of the Doppler frequency shift caused by the vertical modulation of the groove in the cylinder [28].

Optical turntables using a laser beam to follow the groove of phonographic records are also commercially available [29].

All of the above mentioned techniques still follow one circumvolution of the groove at a time, which is not a suitable technique for broken or severely damaged discs, where the groove has large discontinuities.

1.3.3.2 Image processing

The second optical method to read mechanically recorded sounds is to acquire a 2D high-resolution image of a medium, to process the image in order to measure the groove displacement and to extract the sound. One of the advantages of the image processing method is that an acquired image contains several circumvolutions of the groove, and thus it is possible to automate the correction of larger discontinuities of the groove. It should be noticed that such image processing methods are not applicable to the vertical cut grooves, as the depth information is (almost) absent on a 2D image.

Several personal work and students projects using desktop scanners are described on the internet [30, 31, 32, 33]. They all used a desktop scanner to acquire the record images, which generates several problems:

- Due to the size of the scanner, the record must be digitized in several slices to get the whole record. Therefore the extracted sound from the different slices must be later realigned and processed to remove the overlapping parts between the slices.
- The image must be warped to transform the circular slices into rectangular images. Another way to perform this step is to transform the X-Y coordinates of the extracted sound signal into the polar coordinates.
- The used desktop scanners have a too low resolution: 1600 to 2400 dpi. To get a good audio quality, a resolution of about $1\ \mu\text{m}$ is needed, which would mean 25000 dpi.

The groove detection algorithms proposed in these works are pretty similar and can be summarized in the following steps:

- Acquisition of the record image slices using the desktop scanner.
- Image warping to transform the X-Y into polar coordinates.

- The groove positions are located by the maximal intensity pixels at each sampling time.
- Track following: a groove is built of consecutive samples in the time direction, which lay in a close neighborhood in the radial direction.
- Alignment of the tracks resulting from the different slices. The final extracted sound quality of these projects is pretty low, which is mainly due to the non-adapted hardware

The final extracted sound quality of these projects is pretty low, which is mainly due to the non-adapted hardware.

PrestoSpace is a project which objective is to provide technical solutions and integrated systems for digital preservation of all types of audiovisual collections [34]. An optical phonographic record player has been developed as part of the PrestoSpace project. The principle of this system is to illuminate the walls of the groove through a condenser that provides an angle-dependent lighting color. A CCD camera captures the resulting images, where the orientation of the groove is represented by color coding. The advantages of this system is that it requires low resolution (about 10 μm), and that it captures the groove orientation and therefore does not need any numerical differentiation to get the recorded sound. However, this system seems to be very sensitive to dusts and groove deformation.

Fadeyev et al. used methods derived from their work on instrumentation for particle physics to recover audio data from discs and musical cylinders. They developed a 2D imaging system to extract sound out of the bottom of the groove from monaural phonographic records. They work with a rotating scanner and an auto focus system to stay in the limited depth of field required by the microscope video zoom. This system results in good quality sound extraction, and important efforts are now underway to decrease the processing time to around 10 minutes and to develop a final tool for mass digitizing [35].

1.3.3.3 3D method

Fadeyev et al. also introduced a 3D surface measuring technique, using confocal scanning microscopy [36]. This results in a surface map, which is used to detect the groove position as follows:

- Minima candidates are identified as data points which were not higher than the four nearest points.
- These candidates are removed, using the inherent data periodicity of the known groove structure.
- A parabolic fit is applied on the deepest point of the "valleys" of the map.
- The list of fit minima is filtered using the inherent data periodicity of the known groove structure.

The downside of this 3D technique is the long acquisition time, which may last up to one day for one side of a record. But it also presents several advantages as the resulting sound extraction quality is better than the 2D systems, and as it allows reading of vertical cut discs and cylinders, which is not possible with a 2D technique. Further investigations in this domain are underway to reach better accuracy in the mechanical displacements and lower processing time.

Lutz et al. enhanced this 3D technique and adapted it for the special case of dictation belts, which grooves present relatively low excavations [37].

1.4 Overview of this dissertation

This thesis is composed of nine chapters. Chapter 2 gives an overview of the phonographic recording technology and describes the disc and groove characteristics that are useful for the current work. A detailed analysis of the VisualAudio acquisition system is given in Chapter 3. This description focuses on the physical devices and on the sampling process. It encompasses the camera, the photographic film and the scanning stage. Chapter 4 provides then an in-depth analysis of the imaging chain in terms of resolution and image degradations. This analysis results in the definition of a groove model. Based on this model, Chapters 5 and 6 present the image processing algorithms, which perform the groove extraction, respectively the signal corrections. Once the groove has been correctly extracted from the image, the detected positions are transformed in a sound signal, as explained in Chapter 7. Chapter 8 is an evaluation of the extraction process and of the VisualAudio system. Finally, the conclusions drawn from our experiments are discussed in Chapter 9.

Chapter 2

Phonographic recording

This chapter describes the phonographic technology. Subsection 2.1 first presents a brief history. The different technological aspects of the phonographic recording are then presented in the following subsections. This description is not intended to be exhaustive; however it points out the main features that are of interest for an optical sound extraction process. Interested readers are referred to [38] and [39] for more details about the phonographic technology.

2.1 History

The phonographic recording technology was progressively introduced in the nineteenth century. In 1857, Léon Scott created the first machine to record sound: the phonautograph (Figure 2.1). This device only made a visual image of the sound waves on a cylinder, but was not able to play or reproduce sounds. Thus the sound reproduction technology really started in 1877, when Edison recorded human voice for the first time on a tinfoil cylinder phonograph, and played it back using the same device. Berliner introduced the gramophone in 1887, and used discs instead of cylinders to store sounds. Discs were more resistant, more economical to produce and easier to store; thus they gradually overtook the cylinder as the dominant medium for sound recording.

Although the phonograph denotes Edison's cylinder player, it is commonly used as a generic term for any early sound reproducing machine and Emile Berliner's gramophone is then considered a type of phonograph. Therefore, we decide to use the term "phonographic" to describe the disc recording technology in this work.

Phonographic records played a major role in the sound recording history, as it was the only means to store sound until the popularization of the magnetic tape in the late 40's. Thus direct cut disc recording was the only means to archive radio broadcasting, conferences and speeches until the 50's. The LP (Long Play) 33 rpm (rotations per minute) still remained the most popular audio media until the commercial introduction of the compact disc in 1982. Table 2.1 briefly presents the more important dates and facts in the audio recording and phonographic history.

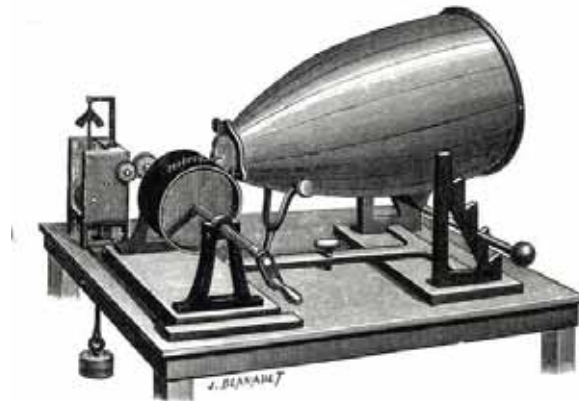


Figure 2.1: Leon Scott's phonautograph: a large horn captures the sound and transmits these vibrations to a needle, which engraves a modulated groove on a rotating cylinder.

1857	Léon Scott creates the phonautograph
1877	Edison makes the first recording on a tinfoil cylinder phonograph
1887	Berliner develops the gramophone, and introduces the disc as a recording media
1925	Electrical amplification is introduced for recording and playback
1925	First 33 rpm (rotations per minute) records
1929	Transcription on 33 rpm records for the use of radio stations
1929	Use of vinyl for record manufacturing
1931	Blumlein patents the stereo recording method
1935	Introduction of the magnetic tape recording
1947	Commercialization of the magnetic tape recording on a large scale
1948	Introduction of the microgroove and LP (Long Play) 33 rpm
1949	Introduction of the 45 rpm microgroove
1954	Standardization of the equalization process (RIAA)
1972	First digital recording
1982	Commercial introduction of the compact disc

Table 2.1: Important dates and facts in the phonographic and audio recording history [40, 41].

2.2 Mechanical recording

Mechanical sound recording techniques store the signal as the groove modulation over the surface of a carrier, which can be a cylinder, a disc or a strip.

The groove modulation of monophonic records is either vertical or lateral. Vertical recording, which is also called hill-and-dale, was used mainly for cylinders, but only by a few record manufacturers (Edison, Diamond Disc, Pathé...) [2]. It rapidly appeared that lateral recording presents less distortion and thus a better sound quality. Therefore the vertical recording has been abandoned and it will not be further described in the current document.

For stereophonic records, the two channels must be stored separately on the same groove using a combination of vertical and lateral recording. The quality difference between the lateral and vertical recording system led to a mixed system called 45/45 system: the groove has a triangular shape and each groove's wall has a 45 angle with the record surface. Left channel signal is stored in the displacement of the groove inner wall perpendicularly to its plane, while the right channel signal is stored in the outer wall displacement (Figure 2.2). Signals that are in phase in both

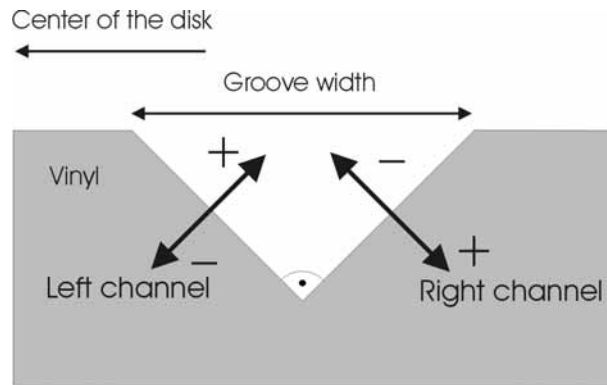


Figure 2.2: Stereo groove cut view: the left channel signal is stored in the displacement of the inner wall of the groove perpendicularly to its plane, while the right channel signal is stored in the outer wall displacement.

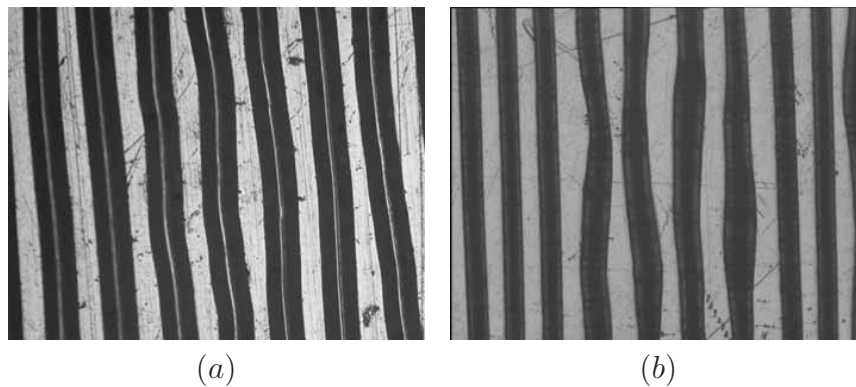


Figure 2.3: Top views of (a) a 78 rpm monophonic record and (b) a 33 rpm stereophonic microgroove record. While the monophonic groove has a constant width and is only radially modulated, both stereophonic groove's walls are modulated resulting in a groove width variation.

channels produce lateral motion of the groove, and out of phase signals produce vertical motion. With the 45/45 system, mono and stereo cartridges and records are compatible: while mono records can be played with stereo playback cartridge, it is not recommended to use stereophonic records with monophonic cartridges, as their resistance to the groove vertical motion may damage the stereo groove.

The inability of mono turntables to track properly the vertical component of the groove, results in distortion of the lateral signal. Thus the vertical component of the low-frequencies is attenuated to limit the vertical movement. Figure 2.3 shows top views of monophonic and stereophonic records.

A few other engraving formats have been developed, but were not widely used, as the quadrasonic sound format for example, which encoded two channels in the displacement of each groove's wall [42].

2.3 Recording process

First records were acoustical recording, which means that the recording process converted the acoustic energy of the sound directly into the groove modulation. The first acoustical recording devices were quite simple, such as the phonograph presented on Figure 2.1: it consisted of a light membrane carrying a stylus, which was cutting a record. A horn was used to concentrate the sound on the membrane. Thus the membrane vibrated under the influence of a sound and the vibration was transmitted to the stylus that moved the groove perpendicularly to the stylus direction. The playback device worked the opposite way: the stylus movements were transmitted to the membrane, which vibrations rendered the recorded sound. There is no amplification in such process, therefore all the energy of the sound is contained in the groove modulation. This imposes large groove amplitude which limits the track duration on each face of the record.

Acoustic records had a limited frequency bandwidth due to the mechanical limitations of the recording process and to the shape of the horn. The shape of the horn also produced resonances that led to great variations in the frequency response of the system. Different horn shapes have been developed in order to control the effect on the recorded sound [43].

Electrical recording was introduced in 1925 after the invention of the microphone and the amplifier. Electrical recording devices use a transducer to convert the mechanical vibrations of the air into corresponding electrical signals. This signal is transmitted to a cutting stylus that transforms the electrical pulses into lateral stylus vibrations to modulate the groove. At playback, the pickup converts the lateral stylus displacements into electrical pulses. These pulses are finally transformed into air sound waves by the loudspeakers. Lots of disturbances are avoided by the electrical recording in comparison to the acoustical one, as most of the process is done on the electrical form of the vibrations and no more the mechanical form [44].

2.4 Disc manufacturing

There are mainly two kinds of record manufacturing: direct cut and pressed records. Direct cut discs were used for studio recording, to record events like speeches and conferences and to edit and mix sound recordings for radio stations. Thus each direct cut record was available as a single copy. The advantage of the direct cut disc is that they can be directly played back after recording, without having to wait for an industrial reproduction of their content. Thus they are also called instantaneous records.

The records manufactured in mass production are also called pressed records. They are basically manufactured in five steps:

1. A direct recording is cut on a lacquer record. Record is metalized.
2. A metal negative master is made by electroforming.
3. A positive master is made from the negative one to obtain a mold.

4. This mold serves to generate pressing masters or "stampers".
5. Stampers are used to press the shellac or vinyl records.

The metal negative master could be directly used to stamp out positive records in plastic; but since it is the only existing copy of the original, it is used instead to produce the stampers [44, 45].

2.5 Record materials

Several kinds of materials and compounds have been used to manufacture records. While pressed records are manufactured with resistant materials to avoid surface wear and allow multiple playbacks, the instantaneous records are made of softer material, which is a compromise between ease of engraving and long playback life. The next three subsections briefly describe the three most used materials, which are the shellac and the vinyl for pressed records and the acetate for direct cut discs.

2.5.1 Shellac

Shellac records were produced during the first half of the 20th century. Shellac is an organic material collected from the secretions of an Asian insect called *Coccus Lacca*. In spite of their name, shellac records contain only approximately 15% of shellac. They are in fact made of a compound of shellac, fillers, binder, lubricant, colorants... These ingredients were added in order to reach the best audio quality requirements and to optimize the automated record manufacturing process. Although shellac itself is resistant to mould and fungus growth, the organic materials used in the compound are susceptible to fungus attack. Thus there is a very wide range of shellac records quality, and stored shellac records may behave in several manners as their chemical compositions differ [46, 47].

2.5.2 Acetate/lacquer

These records consist of a strong base, of either glass or aluminum, covered with a layer of cellulose nitrate lacquer. The softening agents, such as castor oil or camphor, are added in lacquer to insure better cutting properties, as the acetate records were mainly used for direct cut recording. The lacquer color was usually black, but one can find green, yellow or transparent records. Acetate records were mainly used between 1930 and 1955, but the BBC (British Broadcasting Corporation) used instantaneous discs as late as the 1970s [1].

Acetate records have a much finer graininess than the shellac records, which considerably increases the sound quality at playback. Unfortunately, acetates are the most fragile records in comparison to the shellac and vinyl: the different rates of expansion between the layers and the ongoing contraction of the lacquer create stress on the layers and shrinkage of the lacquer coating [2].

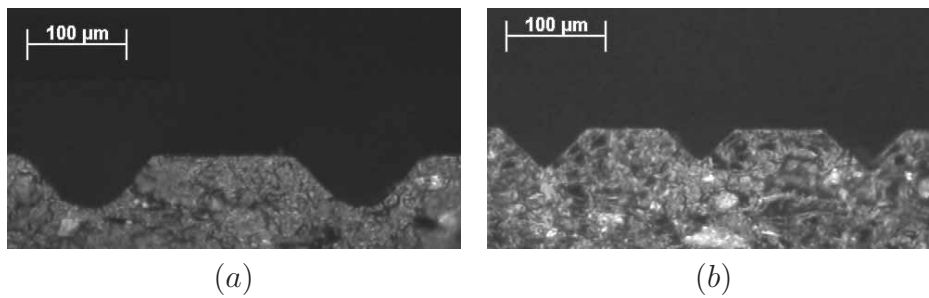


Figure 2.4: Profile views of (a) a 78 rpm monophonic shellac record coarse groove and (b) a 33 rpm stereo microgroove vinyl record.

2.5.3 Vinyl

Since the '40s, the raw material used for the manufacture of pressed records is the polyvinyl chloride (PVC). Vinyl records are also a compound of PVC, stabilizers, colorant, fillers and additives. The PVC has a finer graininess than the shellac. Thus they produce less surface noise and the rotation speed of the records could be reduced. Vinyl is the most stable material used for records manufacturing: it is resistant to fungus and unaffected by high humidity levels.

2.6 Recording speed

Many different recording speeds have been used at the beginning of the phonographic technology. This was not only due to the lack of standardization, but also to the mechanical drive motors that could not always be locked to a specific speed. Thus each manufacturer established its own standard, and recording speed ranged from 60 rpm (rotations per minute) to at least 100 rpm.

Record types may be roughly classified in two categories: coarse groove (78 rpm and a few 33 rpm) and fine groove or microgroove (16, 33 and 45 rpm), which are also displayed on Figure 2.4. The most widely used recording speeds are described in the next subsections [48, 49].

2.6.1 78 rpm

Early speeds of rotation varied widely, but by 1910 these speeds gradually settled around the standard of 78 rpm. Several standards were used in this 78s family, depending on the motor gear ratio and power supply (50/60 Hz) for example: 71.29 rpm, 76.59 rpm, 78.26 rpm, 78.8 rpm and 80 rpm. All these records were later called 78s or coarse groove, to distinguish them from the newer rotating speeds and finer grooves. A 30 cm 78 rpm record could hold up to 5 minutes audio content per side.

2.6.2 33 rpm

The first 33 rpm were manufactured in 1925. The need for longer audio content came about with the advent of early sound films. In the early days of film, the audio was recorded on a separate record, and a reel of film in those days, could run for about 11 minutes. Simply dropping the record recording speed down to 32 rpm would enable the full length film audio to be recorded non-stop on one side. Ultimately, $33\frac{1}{3}$ rpm was agreed on as the final modern standard as it could be yield with gear reduction with either 50 Hz or 60 Hz power supply. Even if the effective speed is $33\frac{1}{3}$ rpm, these records are usually referred to as 33 rpm.

The use of vinyl for record manufacturing came in 1929. Working with a finer grain material allowed cutting finer grooves, which were called microgrooves. With the use of the microgroove and of the variable groove spacing technology (which vary according to the signal amplitude), it was possible to record up to 25 minutes per side. These records are often called LP (Long Play), microgroove or fine groove records in opposition to the earlier coarse groove records.

Thus the stereo, microgroove and 33 rpm technologies were all developed in the early 30's, but the first stereo LPs appeared only in 1948, under the commercial pressure of the tape, which was introduced in the late 40's.

2.6.3 16 rpm

The 16 rpm disc appears towards the 50's. These records rotate in fact at $16\frac{2}{3}$ rpm, which is half the rotational speed of 33 rpm records. This low speed was used to maximize continuous recording times: the low speed combined with large sized record (40 cm diameter) could yield up to about 30 minutes per side for coarse groove or almost 60 minutes for microgroove records. Unfortunately, low speed means more surface noise and therefore less bandwidth (up to 3 kHz), thus these discs were mainly used for spoken words content: transcriptions, language courses, books reading for visually impaired people [50].

2.6.4 45 rpm

The 45 rpm was introduced in 1949 due to a commercial policy: the decision was to have a high-quality fine groove $5\frac{1}{2}$ minutes single record. Since all of the variables have been defined (with certain assumptions about the bandwidth and tolerable distortion), the rotation speed was easy to determine: 45 rpm. However some people still pretend that the choice of the 45 rpm speed is only due to the fact that 78 minus 33 equals 45!

Since the 45 rpm record surface is large enough and to get a better sound quality, the 45 rpm discs are usually recorded with a 3 dB gain in comparison to the 33 rpm.

2.7 Record equalization

2.7.1 Recording characteristics

There are basically two different recording modes to store a signal on a modulated groove: constant amplitude and constant velocity. For a signal being recorded at the same level for all frequencies, constant amplitude recording engraves the same groove amplitude over the whole frequency band. Thus the lateral stylus velocity increases as the frequency rises. Constant velocity records keep a constant lateral stylus velocity. Therefore the amplitude is cut in half each time the frequency doubles, in order to keep a constant velocity. In other words, the amplitude of a sound signal may be stored either in the groove position for constant amplitude recording, or in the derivative of the groove position for constant velocity recording.

Each of these recording characteristics presents some limitations. As frequency increases, the radius of curvature of the groove becomes shorter in constant amplitude mode. This limits the frequency band at recording; otherwise the radius of the groove may be smaller than the size of the cutting or reading stylus, resulting in tracing distortion. The physical mass of the stylus also limits its velocity: recording high frequencies with constant amplitude will require a very high energy to move the stylus, which renders this technique unusable in practice. With constant velocity, the low frequencies have wide excursions, which limit the duration of sound that can be stored on a given surface. High frequencies on the opposite are recorded at a very low level and get lost in the surface noise. Thus none of these recording modes is entirely satisfactory.

2.7.2 Playback characteristics

Pickups use several means to translate the stylus motion into electrical energy. The four basic transducers are called magnetic, dynamic, piezo-electric and capacitance. The first two have a constant velocity response, which is the exact reverse of the process by which the cutting head cuts the record. Piezo-electric and capacitance pickups have a constant amplitude response. Most of the high-fidelity pickups are constant velocity (magnetic or dynamic).

2.7.3 Equalization

To overcome the limitations of the constant amplitude and constant velocity modes, Maxfield and Harrison proposed a hybrid recording characteristic using both constant amplitude and constant velocity characteristics [51]. They defined a bass turnover frequency: all the frequencies below are recorded with constant amplitude and the frequency band above at constant velocity. As the electrical recording progressed, it became possible to extend the frequency range. Thus equalization curves have been enhanced with a treble turnover in order to boost the high frequencies to overcome the surface noise problem. The recording characteristic of an equalized record is then split in three sections:

1. Constant amplitude section from the lower frequency to the low bass turnover.
2. Constant velocity from the bass to the treble turnover.
3. Constant amplitude for frequencies higher than the treble turnover.

In other words, the recording equalization process attenuates the low frequencies and boosts the high frequencies. At playback, the preamplifier does the opposite to restore the original signal. These equalizations are called pre-emphasis respectively de-emphasis. An example of equalization curves is shown on Figure 2.5.

Equalization curves for recording and playback are usually defined by turnover frequencies and roll-off, which is the rate of treble attenuation in dB at 10 kHz during record playback. There is no optimum turnover frequency, as it is a compromise that depends on the audio content, as well as on the record speed and material. However several physical characteristics give some limitations for the definition of the turnovers, and the three above mentioned sections could be redefined as [52, 39]:

1. Amplitude limited region (<1 kHz): limited by the displacement overload, when a groove overloads the adjacent grooves.
2. Velocity limited region (from 1 kHz to 4 kHz): limited by the slope overload, which occurs when the slope of the modulated groove's wall becomes greater than the slope of the trailing faces of the recording stylus.
3. Acceleration limited region (>4 kHz): limited by the curvature overload, which occurs when the radius of curvature of the groove modulation is less than the tip radius of the pickup stylus.

Many equalization curves have been used during the 78 rpm era and the first years of the LP. A few of them are displayed on Tables 2.2 and 2.3. The RIAA equalization was standardized and broadly used since 1955. The RIAA equalization will be explained in details in the following subsection. As the turntable speed decreased and the frequency range increased, there was a need to reduce also the low frequency noise due to the record warping and turntable rumble. Consequently a low bass turnover has been set to use a constant velocity characteristic below this frequency. Thus there is no bass boost at playback below this turnover.

System	Treble turnover	Bass turnover	Low bass turnover	Cut at 10 kHz	Boost at 50 Hz
US MID 30	-	400 Hz	70 Hz	-	16 dB
WESTREX	-	200 Hz	-	-	15 dB
HMV/Blumlein	-	250 Hz	50 Hz	-	12 dB
FFRR 78 (1949)	6.36 kHz	250 Hz	40 Hz	5 dB	12 dB
Early DECCA	5.8 kHz	150 Hz	-	6 dB	11 dB
Columbia	1.6 kHz	300 Hz	-	16 dB	14 dB
BSI	3.18 kHz	353 Hz	50 Hz	10.5 dB	14 dB

Table 2.2: Equalization chart for 78 rpm records [53, 54].

System	Treble turnover	Bass turnover	Low bass turnover	Cut at 10 kHz	Boost at 50 Hz
FFRR LP	3 kHz	450 Hz	100 Hz	11 dB	12,5 dB
EMI LP	2,5 kHz	500 Hz	70 Hz	12 dB	14,5 dB
NAB	1,6 kHz	500 Hz	-	16 dB	16 dB
Columbia	1590 Hz	500 Hz	100 Hz	16 dB	12,5 dB
RIAA	2.1215 kHz	500,5 Hz	50,5 Hz	13,6 dB	17 dB

Table 2.3: Equalization chart for 33 rpm records [53, 54].

2.7.4 Equalization transfer function: the RIAA case

A standardization of the recording and playback curves has been proposed by the RCA company, adopted by the RIAA (Recording Industry Association of America.) in 1954, and later adopted by the IEC (International Electrotechnical Commission) [55]. This equalization curve is now called "RIAA" and has been used for almost all the records since 1955. This norm defines turnovers at three time constants: a treble turnover at $75 \mu s$, a bass turnover at $318 \mu s$ and a low bass turnover at $3180 \mu s$, which correspond to 2122.06 Hz , 500.5 Hz and 50.05 Hz . Equations 2.1 display the turnover definitions, where the symbol ω denotes an angular frequency.

$$\frac{1}{\omega_1} = \tau_1 = 3180 \mu s \quad f_1 = \frac{\omega_1}{2\pi} = 50.05 \text{ Hz} \quad (2.1)$$

$$\frac{1}{\omega_2} = \tau_2 = 318 \mu s \quad f_2 = \frac{\omega_2}{2\pi} = 500.5 \text{ Hz} \quad (2.2)$$

$$\frac{1}{\omega_3} = \tau_3 = 75 \mu s \quad f_3 = \frac{\omega_3}{2\pi} = 2122.06 \text{ Hz} \quad (2.3)$$

The recording pre-emphasis equalization is composed of two high-pass filters (at $3180 \mu s$ and $75 \mu s$) and one low-pass filter (at $318 \mu s$), which transfer functions are:

$$H_1(j\omega) = 1 + j\omega\tau_1 = 1 + \frac{j\omega}{2\pi f_1} \quad (2.4)$$

$$H_2(j\omega) = \frac{1}{1 + j\omega\tau_2} \quad (2.5)$$

$$H_3(j\omega) = 1 + j\omega\tau_3 = 1 + \frac{j\omega}{2\pi f_3} \quad (2.6)$$

The recording complex transfer function is then defined by the combination of these functions

$$H_r(j\omega) = H_1(j\omega)H_2(j\omega)H_3(j\omega) = \frac{(1 + j\omega\tau_1)(1 + j\omega\tau_3)}{1 + j\omega\tau_2} \quad (2.7)$$

The playback de-emphasis complex transfer function must then be the inverse of the recording transfer function:

$$H_r(j\omega) = \frac{1 + j\omega\tau_2}{(1 + j\omega\tau_1)(1 + j\omega\tau_3)} \quad (2.8)$$

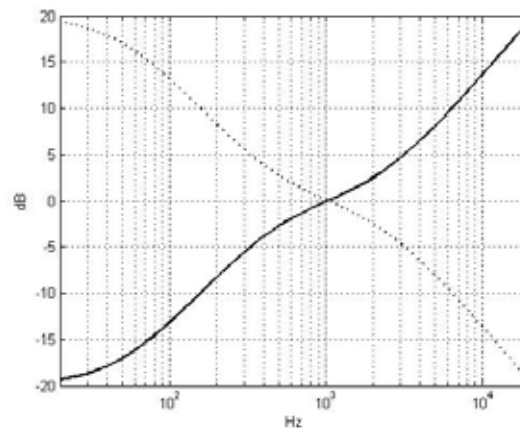


Figure 2.5: RIAA pre-emphasis (plain line) and de-emphasis (dashed line) equalization curves.

The RIAA pre-emphasis (Equation 2.7) and de-emphasis (Equation 2.8) curves are displayed on Figure 2.5.

This was the standardized documented RIAA filter; but according to Wright, there is another turnover used in practice. He cites the "Neumann cutting amp manual", which mentions that the pre-emphasis high frequency boost is being rolled off at about 50 kHz (or sometimes 30 kHz), which corresponds to a high-frequency turnover set at $3.18 \mu\text{s}$. This high-frequency corner must be taken into account to get a more accurate de-emphasis corresponding to the original [56]. This RIAA pre-emphasis version is also called enhanced RIAA. It should be noticed that the reason to use this high turnover is only technical and it has no influence on the sound recording or playback, as it is much higher than the audible frequencies (which range from 20 Hz up to 20 kHz).

2.8 Reading distortions and undesirable effects

Discs can be damaged and recorded sounds may be distorted by several means which can be classified into three categories described in the next three subsections: damages produced during manufacturing, at playback, or caused by bad storage or maintenance.

2.8.1 At recording and manufacturing

Groove overload : If the signal level exceeds a certain limit it produces distortions of the recorded signal. These limits are determined by geometrical factors: the dimensions of the groove and playback styli. Three kinds of overload may appear [52]:

- Displacement overload: the groove overloads the adjacent grooves.

- Slope overload: the slope of the modulated groove's wall becomes greater than the slope of the trailing faces of the recording stylus.
- Curvature overload: the radius of curvature of the groove modulation is less than the tip radius of the pickup stylus.

Time delay : The instantaneous recording cutting stylus was sometimes deliberately rotated over its longitudinal axis, to facilitate the swarf collection by a co-rotating central brush. This produced a time delay between the two groove's walls, resulting in a time delay between the right and left channels of a stereo signal. This time delay is variable across the record surface, but can be digitally corrected. [2].

2.8.2 At playback

Tracing distortion : Non-linearity introduced at playback because the curve traced by the centre of the tip of the reproducing stylus is not exactly the same as the modulated groove. This produces a high frequencies loss, known as scanning loss, and introduces a complex distortion called tracing distortion, which is function of the recorded audio frequencies and level.

Pinch effect : The pinch effect results in the shape difference between the cutting stylus (flat face) and replay stylus (spherical or elliptical). Thus the width of the groove is narrowed twice a cycle resulting in a vertical movement of the needle.

Mechanical resonance : Transient noise is produced by the mechanical resonance of the needle and arm resonance on a damaged groove that is propagated over time. This resonance also causes surface and stylus wear.

Tracking : With a pivoted pickup arm, the angle between the axis of the pickup and the tangent of the groove changes between the outside and the inside of the record. This results in harmonics caused by frequency modulations and surface wear.

Rumble : Low frequency noise resulting from vibrations in platter and motor of a turntable, and from record warping.

Pitch variation or wow : Slow pitch variation (once-per-revolution pitch variation) caused by speed fluctuation in the record rotation (either at recording or at playback), by record warping or by an off-axis spindle hole, which means the spiral of the groove and the spindle hole do not exactly have the same center.

Flutter : Quick pitch variation caused by speed fluctuations in the record rotation.

2.8.3 Record storage and maintenance

Playback loss : It is the difference between the recorded and the reproduced level, which is due to physical properties of the record material. Vinyl and lacquer are much more flexible than shellac, and thus present more playback loss. It is also function of the used stylus and pickup.

Surface wear : 78 rpm are made of a wide array of materials which are very fragile: they will peel, crack, mould, warp, shatter, and suffer about any type of damage due to improper handling or storage. Surface wear is increased by the use of a non-adapted reading stylus which could be too small, too large or too heavy for the groove.

Unroundness : Records exposed to heat may be warped. This kind of deformation appears as non-audible low frequencies on the sound. Bad manufacturing process may be the cause of record warping, as the pressed records are heated and chilled during their production.

Fungus : The organic materials used to press or cut the records are attacked by fungus, which eat the surface of the record. The fungus and the mechanical damages they made will produce noise at playback.

Record contamination : Record can be contaminated by all kind of objects (dust, saliva deposits, tobacco smoke...). This contamination shows up mainly in spurious noise becoming more obtrusive, increased transient noise response and high frequencies loss. The electrostatic nature of vinyl increases the attraction of dust on the record surface in comparison to previously used materials.

Lacquer shrinkage : The lacquer contraction and the differences in rate of expansion between the layers of the lacquer records create stress on the recorded layer which may overcome the force of adhesion. Therefore cracks develop and the record will be unplayable, even untouchable. A picture of a record with shrinkage of the recording layer is shown on Figure 2.6.

2.9 Record and groove geometry

Discs present a flat smooth surface, where a groove is engraved. The record surface between the grooves is called the land. The groove is made of two walls and a rounded bottom, which curvature depends on the recording stylus. The disc is recorded clockwise with a concentric spiral groove running from outside to inside. However, some direct cut radio transcription and some early commercial discs are recorded from inside to outside. Table 2.4 and Table 2.5 gather some measures of grooves and records [57, 55, 38]. It should be noticed that early commercial 78 rpm records as well as direct cut discs were not standardized. Thus the measures of some such records may slightly differ from the ones presented in these tables. Table 2.4 mentions a maximum spindle hole eccentricity, which corresponds to the distance



Figure 2.6: Acetate record with shrinkage of the recording layer: the aluminum base layer is visible at many places (white cracks).

between the groove spiral center and the center of the turntable rotation. This maximum allowed eccentricity is $200\ \mu\text{m}$ for 33 rpm (NAB 1963 and DIN 45507 [55]), which is relaxed in comparison to the $50\ \mu\text{m}$ allowed for the 78 rpm norm (NAB 1953 [58]). This could appear as a step backwards, but in fact the $50\ \mu\text{m}$ were barely attained in practice and the norm has then been adapted to be more realistic. An other interesting value for the eccentricity is the sum of the tolerances for all the pieces involved for playback: the center pin of the turntable and the center hole of the disc. If all these tolerances are observed, the maximum off-axis is smaller than $241\ \mu\text{m}$ [58]. Finally, Table 2.7 gives some wavelength measures for 78 and 33 rpm on the innermost groove of a record, to get an order of magnitude of the groove modulations.

Reading stylus are made of needle shapes and materials. Table 2.6 points out a few stylus properties in order to show the influence of the record material and groove geometry on the turntable needle.

	78 rpm \varnothing 25 cm	78 rpm \varnothing 30 cm	33 rpm \varnothing 17 cm	33 rpm \varnothing 30 cm
Outer disc diameter	250 mm	300 mm	170 mm	300 mm
Outermost groove diameter	241.3 mm	292.1 mm	168.3 mm	292.6 mm
Innermost groove diameter	95 mm	95 mm	97.4 mm	106.4 mm
Max. spindle hole eccentricity	$50\ \mu\text{m}$	$50\ \mu\text{m}$	$200\ \mu\text{m}$	$200\ \mu\text{m}$

Table 2.4: Record characteristics.

	78 rpm	33 rpm
Grooves per mm	2.8 - 4.72	6.85 - 11.81
Groove spacing	211 - 357 μm	84 - 140 μm
Width of groove at top	150 - 200 μm	25-100 μm
Depth of groove	40 - 80 μm	25 - 100 μm
Radius at bottom of groove	20.32 - 63.5 μm	6.35 μm
Angle of groove	$90^\circ \pm 8^\circ$	$90^\circ \pm 5^\circ$
Maximum peak velocity	28 cm/sec	14 cm/sec
Maximum amplitude	75 μm	25 μm

Table 2.5: Groove geometry.

	78 rpm	33 rpm
Radius of the reading stylus	63.5 -100 μm	12.5 - 17.5 μm
Stylus angle	40° - 55°	40° - 50°
Weight of the reading stylus	3 - 6 grams	0.5 - 2 grams

Table 2.6: Reading stylus properties.

2.10 Dynamic range and signal to noise ratio

Records performance and sound quality can be measured by their dynamic and frequency ranges. The dynamic, which is also called the maximum signal to noise ratio, is the ratio between the maximum recorded level (at the maximum peak velocity) and the noise level, and is usually given in dB. The frequency range is the frequency band between the lowest and highest recordable (and reproducible) frequencies. Since the highest frequencies are limited by the surface noise (produced by the record material), the dynamic and the frequency ranges depend mainly on the material used to build the records. Sample values are given on Table 2.8 for several kinds of record materials and technologies.

For compatibility and performance comparison, the sound quality of a record may also be expressed by the S/N (signal to noise ratio), which is given in dB. The S/N is the measure of the noise level below a reference signal. This reference signal is defined by its peak velocity at a given frequency. On monophonic laterally modulated groove, the reference speed is equal to the lateral speed of the stylus. On the stereo 45/45 groove, the stylus is driven at an angle of 45 to the surface. Thus the surface velocity is equal to the stylus velocity in its driven direction, multiplied by the square root of two. For a full compatibility between the two systems, the reference velocity for the stereo channel is divided by the square root of two. Unless specified, the S/N was usually measured on a flat frequency response (constant velocity). The S/N varied according to the material and the recording technologies. Some common values are presented in Table 2.9 [59, 60, 38, 58].

Recorded frequency	Wavelength on a 33 rpm	Wavelength on a 78 rpm
100 Hz	1849.64 μm	4343.54 μm
6000 Hz	30.83 μm	72.39 μm
12000 Hz	15.41 μm	36.19 μm
15000 Hz	12.33 μm	28.95 μm

Table 2.7: Wavelength for several frequencies recorded on the innermost groove (\varnothing 10.6 cm).

	Dynamic range	Frequency range
Shellac 78 rpm	30-50 dB	150-6000 Hz
Acetate / cellulose 78 rpm	50-60 dB	30-10000 Hz
Vinyl 33 rpm Microgroove	66 dB	30-15000 Hz

Table 2.8: Dynamic and frequency ranges for different recording materials and techniques.

	Reference signal	Frequency band	S/N
Shellac 78 rpm	7 cm/sec @ 1kHz	500-6000 Hz	17-37 dB
Acetate / cellulose 78 rpm	7 cm/sec @ 1kHz	500-10000 Hz	37-47 dB
NAB standard (1949) mono	7 cm/sec @ 1kHz	500-10000 Hz	40 dB
NAB standard (1963) mono	7 cm/sec @ 1kHz	500-15000 Hz	55 dB
NAB standard (1963) stereo	5 cm/sec @ 1kHz	500-15000 Hz	50 dB

Table 2.9: Signal to noise ratio performance and standards for different kinds of record.

Chapter 3

Acquisition system

The VisualAudio image acquisition consists in two steps: the record is first pictured by an analog photographic camera and the photographic film is then digitized using a dedicated scanner. The camera and scanner used for the image acquisition are described in Subsections 3.2 and 3.3 respectively. Apart from these two devices, the photographic film also takes a high importance in the imaging process, as is it used as an intermediate storage medium. Subsection 3.1 describes the structure and properties of photographic films as well as the specific needs for the VisualAudio project and the kind of films which have been chosen for the record pictures.

3.1 Films

Photographic film captures the image formed by light reflecting from the surface being taken in picture. Most phonographic discs are black, and thus their surface has a very low reflectivity factor, as most of the light is absorbed. But since the discs are smooth and bright, their reflectivity is mainly specular, meaning that most of the reflected light has a reflective angle equal to the incidence angle.

We will first describe the needs for films to be used to picture records. Then we will describe the film structure and processing, in order to explain the choice we have made for the VisualAudio system.

3.1.1 Film constraints for VisualAudio

The film to be used for the VisualAudio process must satisfy the following constraints regarding the image quality and the mass saving and archiving process:

- Fine grain.
- High contrast.
- Image stability: a local change in luminosity on a record shouldn't lead to a loss of information in case of non-homogeneous record material. The difference of reflectivity between two records (color and recording material) should also not lead to a loss of information on the film.

Some additional constraints must be satisfied for a commercial use of this system but they will not be discussed more in detail in this chapter, as they are not of high interest in the scope of this work:

- Speed of the film and of the development: to have a fast archiving process.
- Easy to manipulate / resistant to manipulation.
- Mid-term availability of the film and development chemicals.
- Available in the desired format (30x30 cm or 40x40 cm).
- Cost.

3.1.2 Film structure

A photographic black and white film is basically made of a polyester base layer and the emulsion, which is the sensitive part of the film. Additional protection layers are added as for example anti-UV, anti-halo and anti-curl layer.

The emulsion is made of a 5 to 10 μm layer of gelatin in which silver halide crystals are randomly distributed. The size of these developed grains is 0.2 to 3 μm , depending on the film, the exposure time and on the development process. The apparent graininess of a film is due to grain clumps created by the overlap of individual grains at different depths of the emulsion.

High-resolution films have a resolution of about 600 lp/mm (line pairs per millimeter), corresponding to 1200 dots/mm. This resolution is limited by the film grain size.

3.1.3 Sensitometry / response of a photographic film

The common method to describe the response of a photographic film is to plot the density against the logarithm of the exposure, which is called D-log E curve. This characteristic D-log E curve can be divided into five sections, named A to E on Figure 3.1:

1. A: D_{min} area, where the density is constant.
2. B: non-linear section called "toe".
3. C: linear exposure portion.
4. D: non-linear section called shoulder.
5. E: D_{max} area, where the density is constant.

The density of the D_{min} area is also known as fog, which can be defined as the developed density that is not due to the image forming exposure. Fog may result from the action of stray radiation (cosmic radiation, not adapted safelight in the

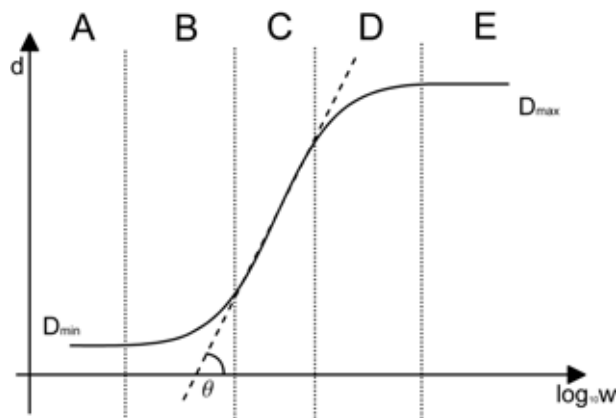


Figure 3.1: D-log E curve of a photographic film: the gamma is the slope of the dashed tangent line: $\gamma = \tan(\theta)$. The dotted lines separate the five sections of the D-log E curve.

darkroom) or from the unselective action of the developer [61]. The linear exposure portion of the curve is modeled as:

$$d = \gamma \log_{10} w - d_0 \quad (3.1)$$

where w represents the incident light intensity and d is called the optical density. γ is called the gamma of the film; it represents the ratio between the film density and the image luminance (Figure 3.1) and is defined as the slope of the linear exposure portion. A negative film has a positive γ . The value of γ depends mainly on the film emulsion, but it can also be altered by changing the exposure time or the development mode. When $\gamma > 1$, the density difference on the film is higher than the luminance difference on the image. Thus a high γ film enhances the contrast, but leads to saturation with a loss of contrast in the high and low luminance parts of the image.

Some other non-linear phenomena alter the exposure-density relationship, as for example the diffusion halo and the chemical adjacency effects. Diffusion halo is caused by light scattering inside the photographic layer of the film during the exposure time. Adjacency effect refers to the flow of active ingredients during development, which could locally enhance the contrast of straight edges. These two phenomena are more pronounced on the highly contrasted areas of the image. They do not affect the linear part of the photographic film response, but alter the toe and shoulder of the curve. Therefore the grain density is not only determined by the exposure time alone, but depends also on the local image configuration and on the development conditions [62].

Contrast, graininess and speed (how quickly the film reacts to light) are closely correlated: slow films contain small grains which require longer time to be sufficiently exposed. The homogeneity of the grain size and sensitivity results in high contrast images with low graininess. Fast films contain grains of different sizes, with larger grains that capture light more efficiently. Thus the different grain size and sensitivity result in grainy images with smooth contrast.

3.1.4 Exposure time

The exposure is the duration of the shutter opening, while the film is exposed to light. Given an illumination system and an aperture (which will be discussed in Chapter 4), and knowing the film sensitivity, the exposure time must be set according to the luminosity of the object, which can be measured with a luminance meter and is called the exposure value.

A bad exposure time may lead to underexposed or overexposed films. Working with a specular reflectivity and a correct illumination system, it is almost impossible to get an overexposed film. But a too long exposure time produces a diffusion halo, leading to an enlargement of the black areas on the negative film. Diffusion halo is a problem for sound extraction on record pictures, as it will produce low-pass filter and harmonics on the extracted sound.

We have measured the diffusion halo on record pictures. The test record HFN001 (see Section 8.1.1) contains an unmodulated track, which has a constant width. Taking pictures with varying exposure time and development time, we have evaluated the width variation and thus the edge displacement due to the diffusion halo. Edge displacements of 20 μm have been measured for 400% enhancement of the exposure time (Table 3.1).

According to the stability constraint (see Subsection 3.1.1), we must guarantee that the film will capture the record image without degradation, even with slight variations of the reflected light. A correspondence table returns the exposure time (ET), according to the measured record luminosity [10]. We have measured the groove width variation produced by the variation of the exposure time, using several records. Our conclusion is that that a variation of $\pm 30\%$ around the calculated ET produces almost no groove width variation. Thus the stability of the image is guaranteed. Working with less than 50% or more than 200% of the ET will lead to underexposed, respectively overexposed pictures.

		Development time (seconds)		
		15	30	60
Exposure time (seconds)	15	121	120	120
	30	115	110	107
	60	83	80	82

Table 3.1: Groove width in μm , with varying exposure and development times. The original width of the groove is 120 μm . The diffusion halo is clearly correlated to the exposure time.

Working with specular reflection, it should be noted that a luminosity variation due to the non-planarity of the record cannot be corrected by a longer exposure time, but with a well-adapted illumination system (see Subsection 3.2.2).

3.1.5 Choice of the film

The sharpness of black and white films is much better than color films due to their emulsion. Thus only black and white films have been considered for VisualAudio, as the main information in disc pictures is perfectly rendered by gray levels. Moreover, picturing with color films would force to work with a larger depth of field during the

scanning process, as these films have three superimposed sensitive layers instead of one for the black and white.

We chose to work with orthochromatic films, which are particularly sensitive to the blue light and insensitive to the red light. Thus the record can be illuminated by a monochromatic blue light during exposure, which short wavelength ensures more sharpness. Another practical advantage of the orthochromatic film is that it is possible to have a red light in the photographic room during the picture taking, which is much easier than to work in the full darkness.

Several films have been tested for record pictures. Four films are presented in this section:

- Ilford Ortho Plus
- Kodak Aerographic Direct Duplicating Film 2422
- Typon TO-G
- Kodak Accumax ABX7

It is not trivial to compare films with the information given by the manufacturers in the fact sheets, as these are given under certain conditions (chemicals, lightening, lens aperture) which do not correspond to our record picture conditions.

Thus film graininess has been measured on record pictures using the VisualAudio application. The standard deviation of the groove edges (*STD*) has been measured on the same track from the same record for all the films (for a full specification of the *STD* measure, please refer to Section 8.1). The resulting standard variations and gamma are displayed on Table 3.2.

Several other parameters may affect these measures, such as the depth of field, but the main *STD* variation between films is due to the film graininess.

Finally, the films have been evaluated according to the above mentioned constraints. The low-gamma films present too much graininess. High gamma films clearly appear as the best compromise between fine grain, high contrast and image stability. Table 3.3 summarizes the advantages and disadvantages relatively to the constraints defined in Subsection 3.1.1. Typon and Accumax were both the best solutions, and we finally chose Typon TO-G for non-technical reasons (delivery, mid-term availability and cost).

Film	Gamma	Edge <i>STD</i>
Ilford Ortho	1	2.12
Kodak Aerographic	1.3	1.55
Typon TO-G	8	0.87
Kodak Accumax	15	0.72

Table 3.2: Films gamma and edge standard variation in pixels. The *STD* measure is detailed in Section 8.1.

Film	Comments
Iford Ortho:	Too much graininess, hard to get a good contrast
Kodak Aerographic:	Too much graininess, hard to get a good contrast, fragile film, lot of physical damages on the film
Typon TO-G:	Low graininess, good contrast, may get a loss of information with a non-perfect illumination
Kodak Accumax:	Low graininess, good contrast, may get a loss of information with a non-perfect illumination, stiff resistant base layer

Table 3.3: Comments about the films.

3.1.6 Storage and life expectancy

The polyester-based film is a very stable media. The ANSI/ISO norms assign a LE-500 (life expectancy of 500 years) rating for polyester-based black and white photographic films. This LE is defined for developed films stored under extended term storage conditions: under 21°C with a relative humidity from 30% to 50% [63]. Although the base layer of polyester film is very stable, the silver-gelatin emulsion layer is sensitive to mould attack and environmental conditions, especially moisture, pollutants and acidic enclosures. These conditions cause oxidation and fading of the images. Therefore the recommended temperature to avoid image decay is of 12°C [64].

3.2 Camera

The purpose of the picture taking step is to get an analog copy of the record, which can be stored for long-term archiving.

The camera that is used for the VisualAudio system must be able to take a 1:1 picture of records having a diameter up to 40 cm. The illumination must be homogeneous on the whole record and ensure enough contrast to clearly distinguish the grooves modulation on the picture.

Several camera models have been tried out, but no commercial model matched the needs required for record picture. Thus a dedicated camera has been designed and built. A complete description of this camera is available in [10].

3.2.1 Camera construction

The camera is built of three planes, as shown on Figure 3.2. The ground level is used to lay the record horizontally. The lens is embedded in the middle layer, and the top layer is made of glass, to put the film on it. Distances between the layers are called the object distance p (between the object and the lens) and the image distance p' (between the lens and the film). These distances are determined by the focal length f of the lens and the magnification ratio m using formula 3.2 and 3.3:

$$m = \frac{p'}{p} \quad (3.2)$$

$$\frac{1}{f} = \frac{1}{p} + \frac{1}{p'} \quad (3.3)$$

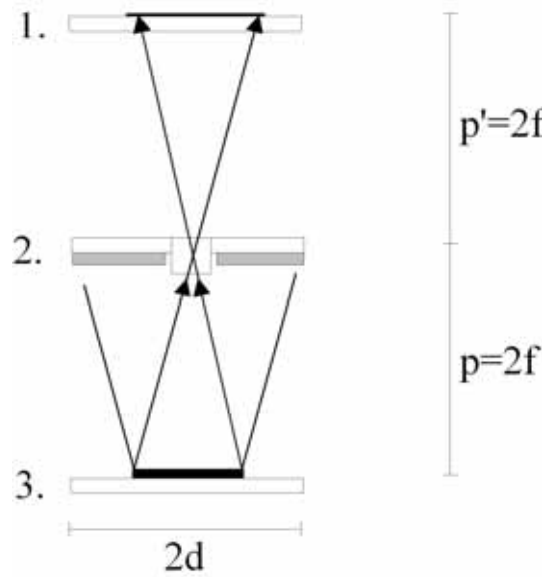


Figure 3.2: The camera designed to take record pictures is built of three planes: 1. the film plane, 2. the lens and illumination plane, 3. the object plane. The width of the illumination system is twice the maximum record diameter d .

In order to limit the geometric aberrations on the outer part of the record image, a large format lens has been chosen with a focal length $f = 420$ mm. Thus for a 1:1 magnification, p and p' equal twice the focal length:

$$p = p' = 2f = 84\text{cm} \quad (3.4)$$

While the working distances p and p' define the minimum camera height, the width of the camera is determined by the width of the illumination system, which must be at least twice the maximum record diameter d (see Subsection 3.2.2). Thus the camera chassis measures 2.2 meters in height and 1 meter in width.

Working with a larger focal length would enhance the resolution, but it would also pose problems as it would reduce the inner part of the illumination system (as the size of the lens increases), thus the light homogeneity would not be guaranteed for the inner part of the record. The height of the camera would also increase proportionally to the focal length f .

While the top layer is fixed to the chassis, the height of the two lower layers of the camera is adjustable by a coarse and a fine tuning system. The coarse adjustment allows to change the enlargement ratio from 1:0.7 to 1:1.2, or to adapt the working distances p and p' to different lenses with focal length from $f = 420$ mm to $f = 360$ mm. A fine tuning system, with an accuracy of ± 1 mm, is used to adjust the sharpness of the image, according to the record thickness or groove depth.

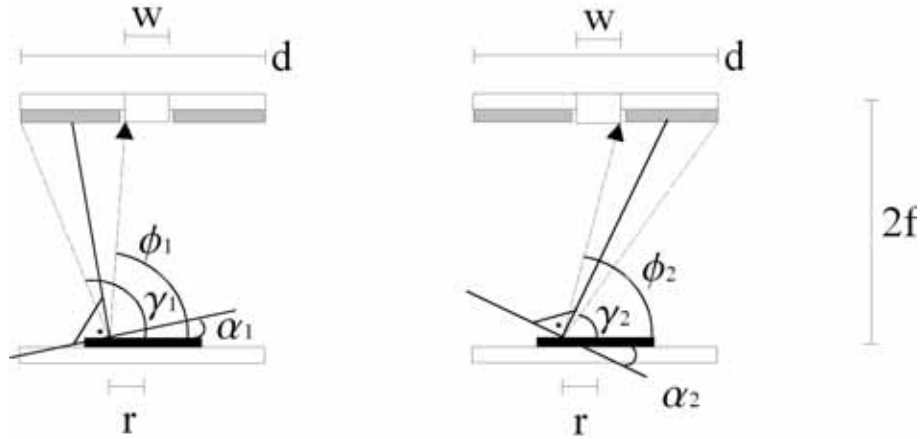


Figure 3.3: Surfaces having an inclination angle between α_1 and α_2 will reflect light to the lens.

3.2.2 Illumination system

The record picture is mainly based on specular reflection, where the light reflection angle on a surface is equal to the incident angle with respect to the surface normal. Therefore the illumination system is of high importance for the photography step in order to distinguish the walls of the grooves from the flat areas of the record (land and bottom of the groove). To have the same light reflection by the flat surfaces on the inner and outer part of the record, the illumination system must be at least twice the size of the object.

Thus the VisualAudio camera illumination system is made of a spiral-shaped blue (with a wavelength of 463 nm) fluorescent tube attached below the middle tray and located behind a diffusing glass, which ensures a homogeneous illumination. Considering records diameter up to 40 cm, the outer diameter of the neon tube is of 83 cm. A wider illumination system would produce a more homogeneous illumination for slightly warped records, but it could also produce undesirable reflections and parasite illuminations.

By working with a large diffuse (non-directive) illumination, the record areas reflecting the light to the lens will not only be the flat surfaces, but all the surfaces which orientation to the horizontal plane is between the angles α_1 and α_2 , as displayed on Figure 3.3. These orientation angles α_1 and α_2 are the normal to the bisectors between the light incident angle coming from the outer part of the light source (γ_1 and γ_2) and the reflection angles (ϕ_1 and ϕ_2) that are needed to reflect the light to the lens at a given radial position r on the record.

Using a lens with a focal length f , these orientation angles α_1 and α_2 can be formulated as follows for a lens width w and a point located at a radius r from the record center:

$$\phi_1 = \arctan\left(\frac{4f}{2r - w}\right), \quad \phi_2 = \arctan\left(\frac{4f}{2r + w}\right) \quad (3.5)$$

$$\gamma_1 = 180^\circ - \arctan\left(\frac{4f}{d-2r}\right), \quad \gamma_2 = \arctan\left(\frac{4f}{d+2r}\right) \quad (3.6)$$

$$\alpha_1 = \frac{\gamma_1 + \phi_1}{2} - 90^\circ, \quad \alpha_2 = \frac{\gamma_2 + \phi_2}{2} - 90^\circ \quad (3.7)$$

Several values for these angles α_1 and α_2 for different radial positions r on the record are given on Table 3.4, considering a light diameter $d = 83$ cm, a focal length $f = 42$ cm, a lens width $w = 2.7$ cm.

r	α_1	α_2
5 cm	10.5°	-16.6°
10 cm	7.3°	-19.6°
15 cm	4.1°	-22.5°
20 cm	0.9°	-25.2°

Table 3.4: At a given radius r on the record, surface having an orientation to the record surface between α_1 and α_2 will reflect the light to the lens.

These maximal inclination angles are lower than the groove's walls inclination angles, which are usually between 40° and 50° (see Table 2.5). This means that the illumination system is well dimensioned and guarantees that the groove's walls won't reflect the light and thus will be visible on the record picture.

3.3 Scanning

In VisualAudio, the scanner acquires a digital image of the record picture, in order to transmit it to a computer for image and sound processing. The main requirements for a record picture scanner are listed as follows:

- Able to scan pictures of records with an outer diameter of up to 40 cm
- Able to scan pictures of records with an inner recorded diameter of 9 cm
- Rotating scanner, to unwind the spiral of the groove
- High resolution, with submicron accuracy in the radial direction
- Low distortion and noise levels.

The current Section focuses on the mechanical and electronic requirements and design of the scanner. The complete scanner design is detailed in [10]. The analysis of the resolution, distortion and noise levels will be developed later in Chapter 4.

3.3.1 Basic structure of the scanner

The film scanner is built of a mechanical part, a digitizing and optical part and an electronic control board.

The film is put on a rotating transparent glass tray, with the photosensitive area on the top. A 2048 sensors linear CCD camera is located on the top to digitize the

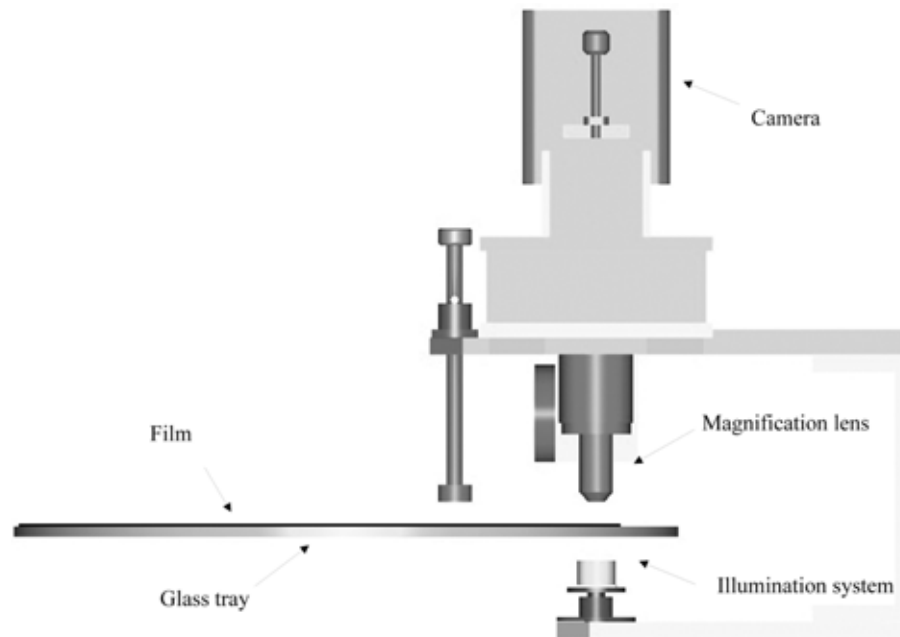


Figure 3.4: Acquisition part of the VisualAudio scanner.

image. A lightening system is fixed below the glass tray on the same axis as the camera (Figure 3.4). A first motor rotates the glass tray counterclockwise (as seen from above) with a constant speed. A second motor, mounted on a worm screw performs a radial displacement, in order to be able to digitize a complete record picture.

3.3.2 Short history

Four scanners prototypes have been built up to now. In 2000, a video camera was fixed on a vertical arm and scanned directly the record picture. There was no radial displacement system, and thus only one ring of the record picture could be acquired, resulting in a few seconds sound extraction. The camera was triggered by its internal clock, leading to an asynchronous sampling [7].

To ensure sufficient resolution, the 2001 prototype was based on a microscope structure. A linear CCD camera fixed on the top of the scanner digitized the record picture. A radial displacement system of the glass tray enabled to scan a complete record (up to a limited diameter of 25 cm). The camera was triggered by the encoder fixed on the glass tray axis, to ensure synchronization of the rotation and sampling [9]. The transition to the 2002 prototype was motivated by the need to get rid of the microscope structure. Thus the 2002 prototype was built from scratch, as a compact portable device. The magnification lens was mounted on an optical tube fixed under the CCD camera. The synchronous rotation motor was located on the glass tray axis, and the camera sampling was driven by the encoder fixed on the same glass tray axis. Unfortunately, the vibrations of the motor were transmitted to

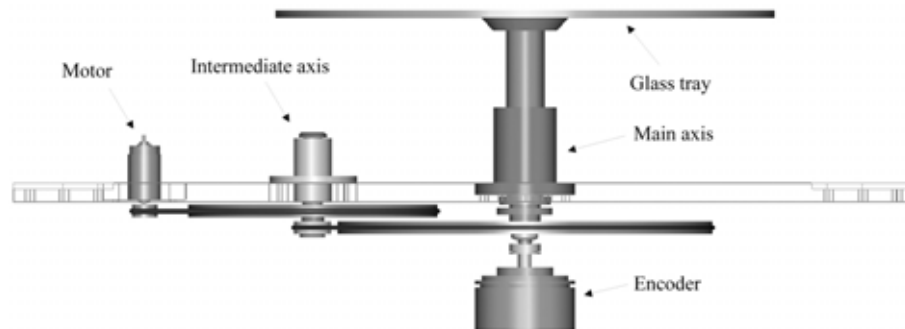


Figure 3.5: Transmission system of the scanner.

the whole scanner including the glass tray and camera. This resulted in a constant hiss and intermodulations in the extracted sound.

3.3.3 The 2004 scanner

Based on the flaws and knowledge acquired with the previous prototypes, the main objectives of the 2004 scanner have been defined as:

- to drastically lower the motor vibrations and their transmission into the scanner
- to be able to read most of the records (outer diameter up to 40 cm, inner recorded diameter 9 cm).
- to have a homogeneous lightening
- to enhance the resolution and to have an adjustable system

A new prototype has been designed to reach these objectives. The motor and the glass tray are located on separate axes, which are called the motor axis and the main axis. The rotation is ensured by a 2 belts driven transmission, with a third intermediate rotation axis. The speed ratio (100:1) between the motor and the glass tray is defined by the size of the pulleys on the three axes (Figure 3.5).

The high inertia ratio (1:100) between the motor and the glass tray prevents the motor vibrations and instantaneous speed variations to be transmitted to the rotating glass tray platter. The drawbacks of the belt driven system are that the overall system takes more space, and that it is harder to precisely control the speed. The accurate knowledge of the speed is not important, as the camera is triggered by an encoder located on the main axis and is thus synchronized with the glass tray rotation. The main constraint over speed is then to keep it constant, which is done by a speed regulation unit, located in the electronic board. Thus both instantaneous speed and average speed (over one circumvolution) variations are lower than 0.1%.

Due to its stability requirements and its heavy mass, the transmission system is fixed to the chassis and the radial displacement is performed by the camera and lightening system, which are mounted on a mobile carriage.

The scanner is equipped with a 2048 sensors CCD linear camera: Piranha CL-P1 from Dalsa [65]. This camera has $10 \times 10 \mu\text{m}$ square sensors and provides 8 bits of data in RS422 differential format at a maximum data rate of 25 MHz. The camera is positioned to have its sensors in line with the glass tray axis center, in order to be perpendicular to the record groove. The shutter of the camera is triggered by the signal of the encoder located on the main axis, which is resampled with a PLL (Phase-Locked Loop) clock synthesizer to guarantee a regular sampling frequency, even in case of rotation speed variations.

A red LED (light emitting diode) is used as a lightening source. This monochromatic light has a 617 nm typical wavelength, which ensures a sharp acquired image by minimizing chromatic aberrations. A monochromatic blue light would provide even sharper images, but the camera is more sensitive to the red light and the power of blue or even green LED is currently too low for our needs.

The LED does not guarantee a homogenous lightening over the whole scanned area. Thus a condenser is placed between the LED and the glass tray to concentrate the light on the scanned area of the film plane. This ensures a more homogeneous lightening. The center and sides of the scanned area are still not equally illuminated, depending on lens magnification and the LED position adjustment.

The 2004 prototype is equipped with a $4 \times$ magnification achromatic lens, with a numerical aperture $NA = 0.1$ that maximizes the resolution (cf. Subsection 4.2.4). This achromatic lens is corrected for axial chromatic aberration in two wavelengths (blue and red; about 486 nm and 656 nm), which is about the wavelength chosen for the LED lightening system.

The glass tray has an external diameter of 40 cm. Thus the camera is able to move and acquire areas of record pictures having a diameter between 8 and 40 cm, which is sufficient for all the standard sized records. The glass tray is mounted on the rotating axis with three adjustable precision screws, thus ensuring a parallelism error between the film plane and the camera plane of less than $10 \mu\text{m}$.

3.3.4 Sampling process

On one rotation of the film, we scan a ring of the disc image (Figure 3.6). Working with a 2048 sensors linear camera and a $4 \times$ magnification lens, the width of the ring is 5 mm. By radial displacements of the camera, adjacent rings are scanned in order to digitize the whole record picture. Two acquisition modes are currently defined to acquire the ring images for a whole disc:

- Online mode: the acquisition and processing of each ring are performed prior to move the camera radially to the next ring position. Thus, the radial displacement is computed for each ring as the maximum possible displacement, without any information loss (see Subsection 5.3.4).
- Offline mode: the radial displacement is defined once at the beginning of

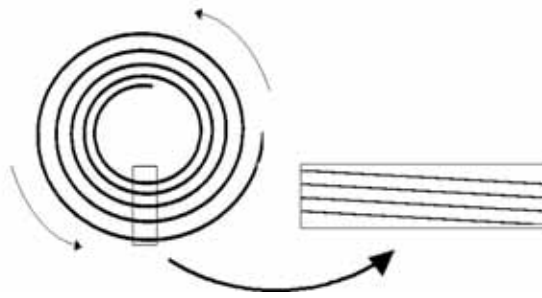


Figure 3.6: The rotating scanner acquires one ring of the disc picture in a rectangular image file, corresponding to one circumvolution of the film.

the scanning. Thus the acquisition can be performed without processing the image, and the image processing can be performed later offline. However, the predefined radial displacement must be large enough to get a sufficient overlapping between the rings in order not to lose any information.

The image sampling frequency is defined by the camera line rate (number of acquired lines per second) and the glass tray rotation speed. This frequency is limited by the hardware capacity and by the light level. Since the LED power is limited, a too fast acquisition will lead to underexposed images. Considering these limits and for electronic implementation purposes, it has been chosen to keep the camera line rate constant at 13 kHz. The sensors opening time must then be smaller than $76 \mu\text{s}$ ($= 1/13000$ seconds), and greater than $20 \mu\text{s}$, which is the maximum shutter speed for this CCD camera. The different image sampling frequencies are therefore defined by varying the rotation speed. The encoder, which triggers the camera, has 4096 points. A PLL controls the output signal of this encoder and multiplies it accordingly to reach the desired camera line rate (Table 3.5).

	12 rpm	6 rpm
Encoder: number of points / rotation	4096	4096
Encoder: output frequency	819.2 Hz	409.6 Hz
PLL multiplier	16	32
Camera line rate	13107.2 Hz	13107.2 Hz
Acquired lines / rotation	65536	131072
Output audio sampling frequency for 78 rpm	85'196 Hz	170'393 Hz
Output audio sampling frequency for 33 rpm	36'408 Hz	72'817 Hz

Table 3.5: Image and audio sampling frequencies.

Two rotation speeds have been defined for the turntable: 6 and 12 rpm. These speeds, combined with the line rate of the camera define the image sampling frequency, ranging from 65.5 to 131 k-samples per ring. These image sampling frequencies combined with the record speeds (33 rpm, 78 rpm ...) define the effective audio sampling frequency, as displayed on Table 3.6.

Since the camera is mounted on a $4\times$ magnification optic and as the sensor size is $10\times 10 \mu\text{m}$, each captured pixel corresponds to a $2.5\times 2.5 \mu\text{m}$ area. But as the glass tray rotates, each CCD-sensor integrates the light on a rotating length of the

	65 k-lines/rotation	131072 k-lines/rotation
33 rpm	36 kHz	85 kHz
78 rpm	72 kHz	170 kHz

Table 3.6: The effective sound sampling frequency is defined by the original record speed and the number of acquired lines per scanner rotation.

film. This integration time is constant but corresponds to different surface sizes, depending on the radial position on the disc. Thus, although the integrated area corresponds to the same sound content duration, it is non-isometric and the disc surface integrated by one pixel may vary from $2.5 \times 4.9 \mu\text{m}$ up to $2.5 \times 16 \mu\text{m}$ depending on the sampling rate, the camera opening time and the radial position on the disc (Table 3.7).

Rotation speed	Recorded groove diameter	Relative sensor move between two samples	Length of the integrated area (per sample)
6 rpm	10 cm	$2.40 \mu\text{m}$	$4.90 \mu\text{m}$
6 rpm	29 cm	$6.95 \mu\text{m}$	$9.45 \mu\text{m}$
12 rpm	10 cm	$4.79 \mu\text{m}$	$7.29 \mu\text{m}$
12 rpm	29 cm	$13.90 \mu\text{m}$	$16.40 \mu\text{m}$

Table 3.7: The relative sensor move between two samples increases proportionally to the radial position and to the scanner rotation speed. Current values are given for a $4\times$ magnification lens and a camera opening time of $76 \mu\text{s}$. The effective integrated length is equal to the move plus the length of the area captured by one sensor.

Chapter 4

Image analysis

During the VisualAudio process, the sound signal is stored in the image of the groove. The image quality should be as good as possible in order to reach the best sound quality at the extraction. Then we need to understand the image formation process as well as the degradations, which could occur in the imaging chain.

This chapter describes first the image of the records that are acquired with the VisualAudio scanner, in order to have a general overview of the image formation and to define some specific terms. Image degradations are split into four categories, which are described in Subsection 4.2 to 4.5: blur, noise and local degradations, illumination variations and nonlinear distortions due to acquisition artifacts. This image analysis leads to the groove model presented in Section 4.6, which will then be used for the groove extraction step. Section 4.7 finally summarizes the image analysis.

4.1 Image formation and definitions

The photography of an object is the projection of the 3D object on a 2D film. Our image is formed by the light reflection on the different surface inclinations of the record.

The light reflection on record material (shellac, vinyl or cellulose) is mainly specular. Thus we first assume a purely specular reflection model.

The camera light source is located on the lens plane and illuminates the disc from above. The groove's walls do not reflect the light to the lens and appear in light grey levels on the acquired image. Flat sections of the disc, i.e. land (flat surface of the disc between the grooves) and bottom of the groove, reflect the light to the camera lens and thus appear black on the negative film, as displayed on Figure 4.1. The groove bottom radius has an important impact on the disc picture: while the bottom of the coarse grooves is clearly visible on the record photography, the microgrooves have a very small bottom radius, and the microgroove bottom is therefore not visible on photography.

The VisualAudio scanner uses a linear camera, which acquires one line of data at each *sampling instant* or *time instant*. These *lines* are also called *acquired lines*. The acquired lines over exactly one circumvolution of the film build a rectangular

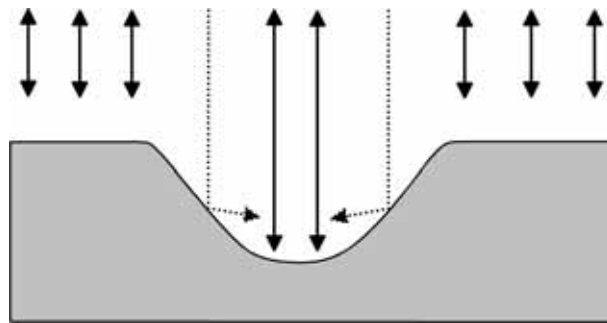


Figure 4.1: While the groove's walls do not reflect the light to the lens and appear as light grey intensities on the photographic film, flat sections of the disc, i.e. land and bottom of the groove, reflect the light to the camera lens and thus appear black on the negative film.

image, called *acquired images* or *rings*. Each ring contains the image of several *groove circumvolutions*. The dimensions of these 2D rings refer to the position on the circular record: *radial* and *tangential*. Thus the radial size of a ring is defined by the number of sensors on the CCD camera, and the tangential length depends on the used scanning frequency: 65K or 131K lines per ring. Consecutive rings are acquired to get a complete record image. Disregarding noise and damages, the record images are built of three kinds of regions:

1. Light traces representing the walls of the groove and which will be further called *traces*.
2. *Dark areas* representing the land (the flat surface of the record between the grooves) and the bottom of the groove.
3. Grey level *transitions*, which lay between the traces and the dark areas. These transitions are mainly produced by the blur during the image acquisition process.

A groove image is built of one or two traces, depending on the original groove bottom radius and shape. These are called *single trace groove*, respectively *double trace groove*. Each trace has two *edges*: a *raising edge* (black to white) and a *falling edge* (white to black). On the double trace groove images, we must distinguish between the *top edges* and the *bottom edges* (Figure 4.2). *Top edges* represent the transitions between the wall and the land and *bottom edges* correspond to the bottom of the groove. Thus bottom edges represent the shifted position of the groove bottom. Relative positions of the grooves, walls and traces can be expressed by their radial position on the record: *inner* (close the record center) or *outer*.

4.2 Blur and resolution

Several kind of blur may affect the groove image in the imaging chain. These different kinds of blur are described in the next subsections. Based on this analysis,

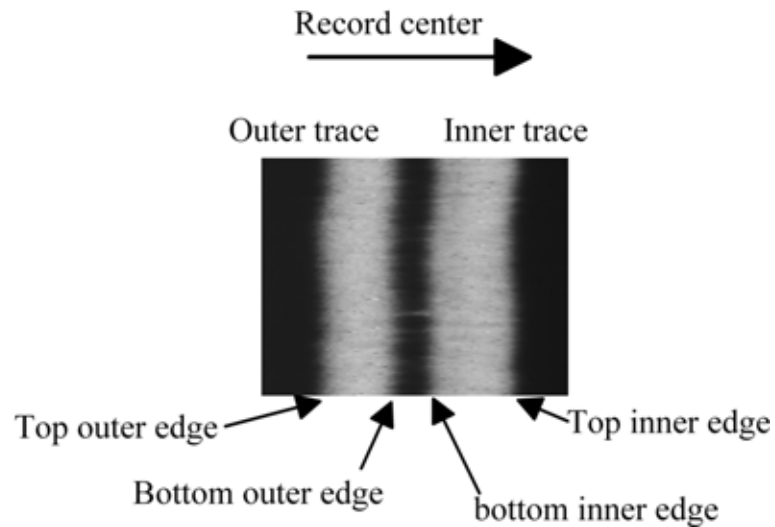


Figure 4.2: Sample of a groove acquisition: the four edges and the two traces corresponding to the two groove's walls are clearly visible.

Subsection 4.2.7 evaluates the order of magnitude of the resolution of the acquisition system.

4.2.1 Shading blur

On a rough surface, the reflection of light is diffuse, which means the light bounces off in all directions. Reflection of smooth surface leads to specular reflection: the incidence angle of the light beam is equal to the angle of reflection.

Most real surfaces reflect with a mix between specular and diffuse reflection. Therefore the luminance transition of a rounded edge will present a reflection that is proportional to the light reflection angle and will produce shading blur on the image [66].

Vinyl, acetate and shellac presents very smooth surface, with mainly specular reflection. Thus the shading blur is negligible on record pictures.

4.2.2 Camera resolution

A camera lens is limited by its sharpness and aberrations. Chromatic and geometric aberrations could be considerably lowered by working with a narrowband light and using a diffraction limited lens with a large focal length. These aberrations can then be considered as negligible, as we currently work with a large focal lens and with a narrowband blue light at picture taking (cf. Chapter 3). The spatial resolution of an optical system is then limited by its depth of field (DOF) and the Airy patterns (see [67]) produced by diffraction. The diameter of the Airy disk d_{airy} and the circle of confusion C produced by the depth of field can be computed using the focal length f and the opening diameter D of the lens. The ratio between f and D is also called

the f-number N . For a 1:1 enlargement, we get the following formulas using the light wavelength λ [67]:

$$C = \frac{DOF \cdot D}{4 \cdot f}, \quad d_{airy} = 2.44 \cdot \lambda \cdot \frac{f}{D} \quad (4.1)$$

With the groove depth and the disc warping, we estimate that the maximal depth of field needed to take picture of a disc is $DOF = 1$ mm. Thus we can estimate the total blur B_{camera} (in μm) of the camera:

$$B_{camera} = \sqrt{C^2 + d_{airy}^2} \quad (4.2)$$

The resolution R_{camera} (in lines pairs per millimeter) is equal to $1/B_{camera}$ and is optimal when $C = d_{airy}$, which defines the optimal f-number:

$$N = \frac{f}{D} = 15 \quad (4.3)$$

In order to limit the geometric aberrations, we have chosen to work with a large aperture lens having a focal length $f = 420$ mm and an opening diameter $D = 27$ mm. Table 4.1 displays the properties and resolution of such a lens, working with a monochromatic blue light with $\lambda = 463$ nm. This table compares different DOF values, corresponding to no depth difference ($DOF = 0$ μm), to the mean depth of a 78 rpm groove ($DOF = 70$ μm), and to the DOF estimated to encompass record warping ($DOF = 1000$ μm).

DOF	D	N	C	d_{airy}	B_{camera}	R_{camera}
0 μm	27 mm	15	0 μm	16.9 μm	16.9 μm	59.2 lp/mm
70 μm	27 mm	15	1.2 μm	16.9 μm	17.0 μm	58.8 lp/mm
1000 μm	27 mm	15	16.6 μm	16.9 μm	23.7 μm	42.1 lp/mm

Table 4.1: Properties and resolution of the 420 mm lens working with an opening diameter $D = 27$ mm and monochromatic blue light ($\lambda = 463$ nm).

4.2.3 Film resolution

The apparent graininess of a film is due to grain clumps created by the overlap of individual developed grains at different depths of the emulsion. This graininess, as well as the different effects produced during the film processing (diffusion halo, chemical adjacency effect) limit the resolution of the photographic film. High-resolution films have a resolution of about 600 lp/mm, corresponding to 1200 dots/mm. The blur of the film is then:

$$B_{film} = 1.7 \mu\text{m} \quad (4.4)$$

4.2.4 Scanning optics

The scanner resolution can also be estimated using the DOF and Airy disk. The circle of confusion is defined with the DOF , the numerical aperture NA of the lens

and the refraction index n . The needed DOF depends mainly on the glass tray warping and on the film emulsion layer depth. Typical film emulsion depths are between 10 to 20 μm , and the measured warping of standard quality glass tray were lower than 20 μm . The equations for the resolution of microscope optics are similar to those presented about the camera optics in section 4.2.2, except that they are based on the numerical aperture (NA) that defines the light gathering ability and resolving power of a lens:

$$C = DOF \cdot \frac{NA}{\sqrt{1 - NA^2}}, \quad d_{\text{airy}} = \frac{0.61 \cdot \lambda}{NA} \quad (4.5)$$

If we consider that the DOF is limited by the depth of the film's sensitive layer, than $DOF = 10 \mu\text{m}$. This resolution has been computed for normal quality glass. Better resolution would be reachable with a flatter high quality glass tray or the use of an autofocus, which would lower the needed DOF to the value defined by the film emulsion depth only. The ideal scanning blur would then be obtained with $NA = 0.19$:

$$B_{\text{scanning}} = 2.8 \mu\text{m} \quad (4.6)$$

Table 4.2 shows the resolution for $DOF = 10 \mu\text{m}$ and different magnification lenses: the best resolution is obtained by the 10 \times magnification lens with $NA=0.25$.

It is more realistic to consider $DOF = 30 \mu\text{m}$, as it encompasses the photosensitive emulsion depth and the warping of the rotating glass tray. Thus when working with $\lambda = 0.615 \mu\text{m}$, the total scanning blur is minimized by choosing $NA = 0.11$:

$$B_{\text{scanning}} = \sqrt{C^2 + d_{\text{airy}}^2} = 4.7 \mu\text{m} \quad (4.7)$$

Table 4.3 shows the resolution for $DOF = 30 \mu\text{m}$ and different magnification lenses: the best resolution is obtained with the 4 \times magnification lens with $NA=0.1$. Thus, as the measured glass tray warping is of around 20 μm , we decided to work with the 4 \times magnification lens.

NA	Magnification	C	d_{airy}	B_{scanning}
0.06	2 \times	0.6 μm	6.2 μm	6.3 μm
0.1	4 \times	1 μm	3.7 μm	3.9 μm
0.25	10 \times	2.6 μm	1.5 μm	3 μm
0.4	20 \times	4.4 μm	0.9 μm	4.5 μm

Table 4.2: Scanning blur with $DOF = 10 \mu\text{m}$, working with monochromatic red light with $\lambda = 0.615 \mu\text{m}$: the best resolution is obtained by the 10 \times magnification lens with $NA=0.25$.

NA	Magnification	C	d_{airy}	B_{scanning}
0.06	2 \times	1.8 μm	6.2 μm	6.5 μm
0.1	4 \times	3.0 μm	3.7 μm	4.8 μm
0.25	10 \times	7.7 μm	1.5 μm	7.9 μm
0.4	20 \times	13.0 μm	0.9 μm	13.1 μm

Table 4.3: Scanning blur with $DOF = 30 \mu\text{m}$, working with monochromatic red light with $\lambda = 0.615 \mu\text{m}$: the best resolution is obtained with the 4 \times magnification lens with $NA=0.1$.

4.2.5 Motion blur

The motion blur is produced by the displacement of the groove relatively to the camera during the acquisition of a single sample at scanning. It should be noticed that as the CCD linear camera is aligned with the center of the glass tray rotation and perpendicular to the groove, this motion blur appears only in the radial direction.

This groove displacement has three causes: the spiral of the groove, the film centering and the recorded audio signal which modulates the groove. The motion due to the spiral over one pixel is defined as:

$$m_{spiral} = \frac{s \cdot c \cdot f_s}{n} \quad (4.8)$$

where s is the spacing between two consecutive grooves, and n the number of samples per rotation, f_s is the sampling frequency of the camera and c the sensor's opening time.

The bad centering of the disc must be limited in order to be able to read a sufficient number of grooves at each rotation of the scanner. An acceptable centering will allow a $200 \mu\text{m}$ peak-to-peak deviation at most. The off-axis can be approximated by a sinus function $f(t)$ of the form:

$$f(t) = \frac{a}{2} \cdot \sin\left(\frac{2\pi}{n}t\right) \quad (4.9)$$

where a is the maximal peak-to-peak deviation due to the bad centering in one circumvolution. This means that the maximum groove displacement due to the bad centering is equal to the maximum of the derivative of $f(t)$ defined in Equation 4.9:

$$m_{centering} = \frac{a\pi}{n} \quad (4.10)$$

The maximum motion of the groove produced by the recorded signal during one sampling period may be defined as:

$$m_{audio} = \frac{V_{max} \cdot c \cdot f_s}{n \cdot rps} \quad (4.11)$$

where V_{max} is the maximum recorded velocity and rps is the recorded speed in rotation per second. With $n = 65$ k-samples, and considering $V_{max} = 28$ cm/sec for a 78 rpm disc, then $m_{audio} = 3.3 \mu\text{m}$. This value does not consider the CCD transfer time and assumes that the opening time of the CCD sensor is equal to a complete sampling period. This is not the case in practice, but this calculation gives an upper bound for the motion blur in the VisualAudio image acquisition. m_{spiral} and $m_{centering}$ are both smaller than $0.01 \mu\text{m}$ and can be neglected. Thus the maximum motion blur could be approximate as the audio motion blur acquiring with the maximum camera opening time:

$$B_{motion} = m_{audio} = 3.3\mu\text{m} \quad (4.12)$$

Working with smaller sensor opening time (see Section 3.3.4), this motion blur is linearly decreased. Motion blur could also be lowered by increasing the sampling frequency.

4.2.6 Sampling blur

The acquisition process integrates a ring angular section of the record surface over a rectangular pixel. The square camera pixels have a physical surface of $p \times p$, which correspond to a $d \times d$ area on the focal plane, where $d = p/m$, and m is the lens magnification ratio. Since the film is rotating during the acquisition, a digitized image pixel then represents a area on the image on the focal plane, considering the sensor move t during the camera opening time:

$$t = \frac{2 \cdot \pi \cdot r \cdot c \cdot f_s}{n} \quad (4.13)$$

where r is the radial position on the disc and f_s is the sampling frequency of the camera and c the sensor's opening time for one sample acquisition. The n samples of a ring acquisition are separated by a distance Δ :

$$\Delta = \frac{2 \cdot \pi \cdot r}{n} \quad (4.14)$$

If the distance Δ is smaller than $d + t$, then some film areas will be sampled by two consecutive pixels, resulting in sampling blur $B_{sampling}$:

$$B_{sampling} = \begin{cases} 0 & \text{when } \Delta > d + t \\ d + t - \Delta & \text{otherwise} \end{cases} \quad (4.15)$$

Table 4.4 and 4.5 show some sensor move t , according to various opening time c , to various radial positions over the record and for magnification optics $4\times$ and $10\times$. These tables put also in evidence the configurations which lead to blur.

n	r	Δ	$\Delta - d$	t for $c = 20 \mu s$	t for $c = 40 \mu s$	t for $c = 60 \mu s$
65k	5	4.79	3.79	1.26	2.51	3.77
	10	9.59	8.59	2.51	5.03	7.54
	15	14.38	13.38	3.77	7.54	11.31
131k	5	2.40	1.40	0.63	1.26	1.88
	10	4.79	3.79	1.26	2.51	3.77
	15	7.19	6.19	1.88	3.77	5.65

Table 4.4: This table shows several Δ and t values using a $10\times$ magnification optics and various opening times, sampling lines/circumvolution and radial positions on the record. For an $10\times$ optics, the surface on the focal plane corresponding to one pixel is $d \times d = 1 \times 1 \mu m$. The t values in bold mark the configurations of n , r and c that result in $\Delta > d + t$ and to blur for the $10\times$ magnification.

The definition of $B_{sampling}$ in Equation 4.15 considers that the surface areas sampled by consecutive pixels are independent. In practice, as the image at the image plane is blurred by the camera and scanning optics, the sampling blur $B_{sampling}$ must be taken into account to evaluate the resolution, even when $\Delta < d + t$.

n	r	Δ	$\Delta - d$	t for $c = 20 \mu\text{s}$	t for $c = 40 \mu\text{s}$	t for $c = 60 \mu\text{s}$
65k	5	4.79	2.29	1.26	2.51	3.77
	10	9.59	7.09	2.51	5.03	7.54
	15	14.38	11.88	3.77	7.54	11.31
131k	5	2.40	-0.10	0.63	1.26	1.88
	10	4.79	2.29	1.26	2.51	3.77
	15	7.19	4.69	1.88	3.77	5.65

Table 4.5: Same as Table 4.4, but for a $4\times$ magnification optics. For a $4\times$ optics, the surface on the focal plane corresponding to one pixel is $d \times d = 2,5 \times 2,5 \mu\text{m}$.

Since the combination of B_{camera} and $B_{scanning}$ is larger than Δ , we can consider the following $B_{sampling}$ as the maximum bound for further resolution calculation using $1 \times 1 \mu\text{m}$ sensor size and a $4\times$ magnification lens:

$$B_{sampling} = d = 2.5\mu\text{m} \quad (4.16)$$

Equation 4.16 is the maximum blur limit for small radius r when n increases and when Δ and t diminish due to oversampling.

The integration is in fact a much more complex process, as it is not uniform on the whole surface acquired by one pixel: surface integration is a weighted integration, where the weight is a function of Δ , d and t . Since the sound signal is largely oversampled in the VisualAudio system, as we are looking for a total system resolution and as $B_{sampling}$ is only a minor component of the whole resolution calculation, we do consider that $B_{sampling}$ is uniform and we do not go more into details to analyze the integration process in the scope of this work.

4.2.7 Total system resolution

The calculation of the total system resolution is useful in order to dimension the mechanics and the optics of the whole process. It also helps to determine whether the resolution is sufficient to extract the groove accurately.

Disregarding noise, the resolution R_{system} and blur B_{system} of the system can be approximated using the blur of each component [68]:

$$\frac{1}{R_{system}^2} = B_{system}^2 = B_{camera}^2 + B_{film}^2 + B_{scanning}^2 + B_{motion}^2 + B_{sampling}^2 \quad (4.17)$$

With a film resolution of 600 lp/mm (line pairs per mm):

$$B_{system} = \sqrt{23.7^2 + 1.7^2 + 4.8^2 + 3.3^2 + 2.5^2} = 24.6\mu\text{m} \quad (4.18)$$

Equation 4.18 gives the worst case resolution of the VisualAudio system. The picture-taking phase is obviously the limiting factor of the whole process, however if the disc is flatter than assumed, the DOF decreases and the resolution increases. The ideal resolution, considering minimal depth of field at photography (Table 4.1) and scanning (Equation 4.6), as well as a non-modulated groove producing no motion blur, will be:

$$B_{system} = \sqrt{16.9^2 + 1.7^2 + 3^2 + 2.5^2} = 17.4\mu m \quad (4.19)$$

The blur level may slightly differ between the two dimensions of the plan, as the sampling blur appears only in the radial direction (parallel to the sensors line) and the motion blur appears only in the tangential direction (perpendicular to the sensor). Since these two causes of blur are minor compared to the optical blur, this implies only a 0,1 μm difference in both direction.

4.2.8 Effect of the blur on the tangential direction

Blur in the tangential direction will produce a low pass filtering on the acquired image grey levels. High amplitude peaks will be slightly eroded, which will produce harmonics on the extracted sound. The blur will also attenuate the high frequencies, which wavelength are close or smaller to the blur size. Therefore the resolution rather than the sampling frequency is limiting the extracted sound bandwidth. Wavelength measures (Table 2.7) confirm that the resolution is sufficient for most of the 78 rpm sound extraction.

4.2.9 Effect of the blur on the radial direction

In order to be able to extract the sound, the blur B_{system} in the radial direction must be smaller than the features to extract, i.e. the groove bottom, the groove's walls and the land. This is the case according to Table 2.5.

In the radial direction, the blur spreads the edge over several pixels. But the information to locate the edge is still contained in the luminance profile and it is still possible to achieve subpixel accuracy location on a blurred edge [69].

4.2.10 Spatial variance of the system

The blur appearing in the VisualAudio process is not spatially invariant for several reasons:

1. The motion blur depends on the groove displacement, which is mainly driven by the disc audio content.
2. The record area integrated by a pixel is larger on the outer part of the record than on the inner part. The optical blur then affects more the acquired images of the inner part of the record.
3. The out-of-focus blur may be not constant, as it is produced by the disc warping, the depth of the emulsion layer on the film, the glass tray warping, the disc content (for stereo disc, where the depth of the groove depends on the sound content) ...
4. Due to the groove depth, the out-of-focus blur at photography is not equal at the top and bottom edges of the groove. Moreover, the VisualAudio imaging

chain contains two steps: the photography and the scanning. The visual features that are on the record (record groove's walls, scratches, dust...) are blurred at both steps; but there are also some objects that appear on the photographic film (film grain, pepper fog spots, dust or fiber pinholes... as described in Subsection 4.3.3), which are blurred only by the scanning optics.

Therefore the blur in the VisualAudio system is not constant and it can't be considered as a spatially invariant system, neither in the radial nor in the tangential direction.

4.3 Noise and local degradations analysis

The grey level pixels of the digital image either represent the groove's walls, the groove bottom or the land. If any spurious feature appears in the imaging chain, the corresponding grey levels can be considered as replacement noise, in the sense that they do not represent the groove or the land anymore. This image replacement noise produces impulsive noise on the extracted sound. Such image replacement noise can be produced at any stage of the VisualAudio process, which are detailed in the next three subsections: on the record, during the picture taking or at scanning.

4.3.1 Record

The quality of the record image depends on the record material and on the state of conservation of the disc. As exposed in Chapter 2, records surface may be contaminated by dust and other objects such as fibers or smoke deposits. The recording surface could also be damaged by scratches (Figure 4.3), crackles, locally broken or subject to mould or fungus attack (Figure 4.4). All of these produce either dark or light spots on the record image. The size, shape and grey level of these spots vary a lot, due to the large variety of spurious objects and existing damages. Several samples of such degradations are displayed on Figure 4.5.

The edges of the groove which appear on the record image correspond to the top and the bottom of the groove. These parts of the groove were not intended to be used for the record playback, since the turntable needle follow the walls of the groove. Thus there is no guarantee that the top edge was of a sharp and does represent accurately the groove's wall modulation. The groove edge sharpness couldn't be observed very accurately for direct cut discs. Thus, until now, it was not possible to observe and to determine whether the engraving process will lead to heaps of acetate on both sides of the groove and which are the consequences of these heaps for the VisualAudio process. Measurements of the different groove edges quality were performed (cf. Chapter 8, Evaluation), which do not show important quality variations between edges. However, groove edge sharpness certainly depends on the engraving stylus and the record material, and thus some noise may appear on the extracted sound.

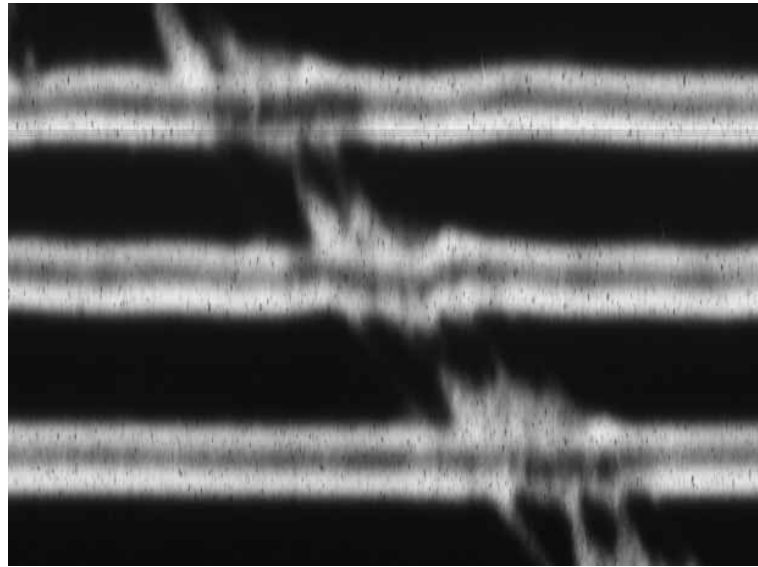


Figure 4.3: Pictures of records with a surface scratch.

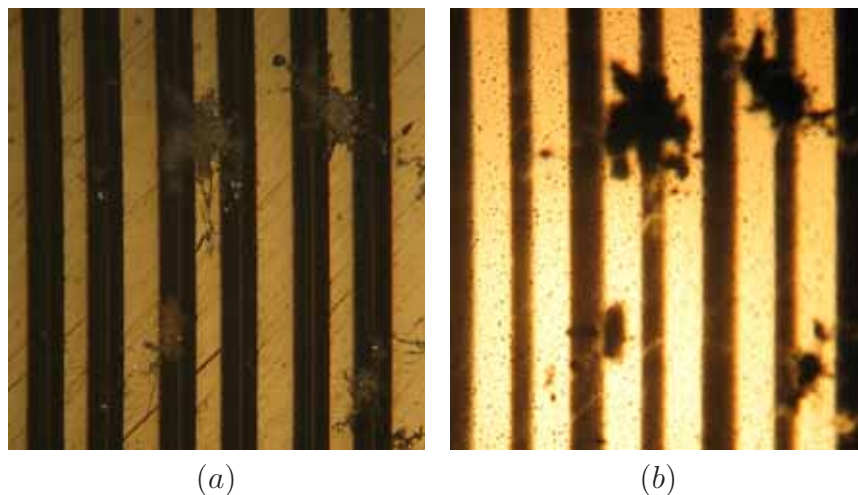


Figure 4.4: (a) Microscope view of fungus on an acetate record. (b) Microscope view of the same fungus on the film negative.

4.3.2 Records with shrinkage of the recording layer

Shrinkage of the recording layer is a particular case of degradations, as it produces large cracks over the record and shifts part of the recording layer and of the grooves. Thus part of the grooves are shifted up to several hundreds of micrometers. Figure 2.6 shows a disc with shrinkage of the recording layer and Figure 4.6 shows part of an acquisition performed on a disc with shrinkage of the recording layer. The matching of the groove over the cracks is then difficult and must be performed with a global analysis of the acquired ring, or even with a global image analysis over the whole record.

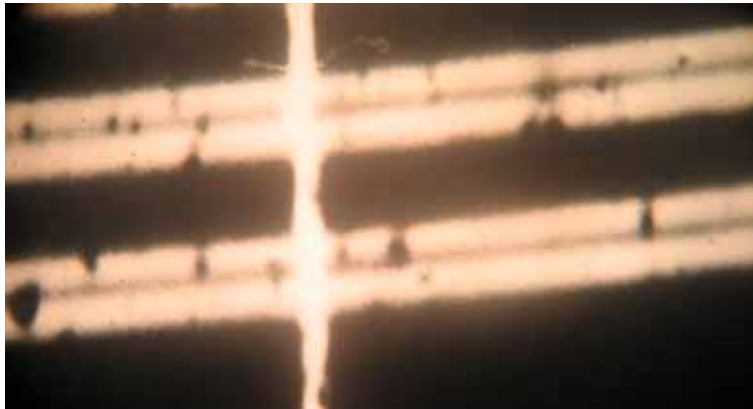


Figure 4.5: This film shows several samples of image degradations, which are due to the record: the black spots are produced by dust located in the groove; the white vertical line is a crackle. It should be noticed that the small hair, which appears above, was not on the record surface, but was located on the film plane of the camera during the picture taking, and therefore appear as a white spot (see "dust or fiber pinholes" in Subsection 4.3.3).



Figure 4.6: This acquisition has been performed on a disc with shrinkage of the recording layer. The black cracks are the visible part of the aluminium plate. At some cracks, the grooves (in white) present important shifts from one piece to the next.

4.3.3 Film and picture taking

During the film development process, silver halide grains which were exposed to a sufficient light intensity during the exposure, are converted to metallic silver grains. The size of these developed film grains is between 0.2 to 3 μm , depending on the kind of film, the exposure time and the development process. The relationship between the incident light during the picture taking and the density of silver grains on the developed film is non-linear: silver grains are randomly distributed in the emulsion,

and silver halide grains exposed to an equivalent light intensity do not necessarily produce the same silver grain size. Thus the shape as well as the randomness of the distribution and of the formation of silver grains contributes to produce noise on the photographic image, which is known as film grain noise. Film grain noise is often modeled as Poisson or Gaussian random process, and constitutes multiplicative noise on the image [62, 70]. If the film is processed in the linear region of the photographic response (where the slope of the curve equals to the gamma), the observed image $o(x, y)$ can be modeled as [70]:

$$o(x, y) = f(x, y) + \alpha \cdot [f(x, y)]^\beta \cdot n(x, y) \quad (4.20)$$

where $f(x, y)$ is the original image density, α is a proportionality constant, β is a constant value between 1/3 and 1/2 and $n(x, y)$ is white Gaussian noise with zero mean and unit variance. This model shows that the film grain noise is signal dependent, as it multiplies the original image density. For the VisualAudio process, we are interested in the audio signal, which is recorded in the displacement of the groove in the image and is independent of the film grain density. If we assume that the groove extraction process works on picture areas of similar density, we can consider the film grain noise as white noise on the extracted sound. Additional impulsive noise may be produced by films artifacts. These artifacts include mainly [20]:

- Pepper fog spots: spontaneous development of silver halide grains in emulsion results in dark round spots, 5 to 30 μm in diameter, in the light areas of the film. These are usually due to the development process: too high temperature, too long development time, developer contaminated by fixer... Pepper fog samples are displayed on Figure 4.7.
- Emulsion pickoff pinholes: white spots appear on the image, where the emulsion has been physically picked of the polyester base before or after the exposure.
- Dust or fiber pinholes: white spots occur when something prevents the exposure of the emulsion during the exposure: dust, fiber or scratch on the glass tray of the camera. Such a fiber pinhole is visible on the film presented on Figure 4.5.

4.3.4 CCD camera

The noise that may arise in a CCD image capture include [70]:

- Photon noise: refers to the natural variation of the incident photon flux. The photon noise is then proportional to the square root of the magnitude of the light signal.
- CCD noise: includes all kinds of noise that might occur on a CCD-sensor due to the electronic: dark charge, dark transfer, transfer noise...



Figure 4.7: Magnified view of a film: the small dark spots are the pepper fog spots. This picture has been taken with a very low exposure time. Thus we can also see the pepper fog spots in the exposed areas of the film and see that these spots are uniformly distributed over the light and dark areas of the picture.

- Readout noise: noise produced by the quantification of the electronic signal. Readout noise is important when exposures are short, but becoming less significant as exposure increases.
- Pattern noise: spatially fixed noise due to sensors inhomogeneities and to the presence of dust or other deposits on some sensors.

Dust, scratches or dirtiness, which may be present on the glass tray or on the film, cause dark spots on the acquired image, which will produce impulsive noise on the extracted sound. Dust or other spurious deposit on a camera sensor will lower the grey level of this sensor and produce impulsive noise on the extracted sound.

4.3.5 Signal to noise considerations

The film grain's shape and size limit the VisualAudio sound quality by producing noise on the extracted signal. The influence of the film grain on the final extracted sound does not only depend on the grain size, but also on many other parameters, such as the grains density, the sensitive layer depth, the exposure time, the pixel size and magnification, which define the integrated area. Therefore it is not possible to directly match the size of the film grain with the noise level of the extracted sound. However, the noise level can be estimated from the standard deviation of the extracted sound signal. Thus we can get an estimation of the signal to noise ratio (SNR) and of the resolution needed for the scanning process.

To estimate the SNR, we assume that the noise produced by the film and by the acquisition process has a uniform frequency distribution (white noise) and that the original signal cut on the disc is a sine wave:

$$s(t) = A_{ref} \sin(2\pi f_{ref} t) \quad (4.21)$$

where A_{ref} is the maximum amplitude at a given reference velocity v_{ref} for a frequency f_{ref} . In the next paragraphs, we will estimate the maximum noise standard

deviation on one edge σ_n allowed to reach a satisfying signal to noise ratio using the VisualAudio process. The following parameters were used:

- B_n the noise bandwidth: for white noise, B_n ranges from 0 Hz up to $f_s/2$ Hz, where f_s is the audio sampling frequency during scanning.
- B_s the sound bandwidth: B_s ranges from f_0 to f_1 , where $0 \leq f_0 < f_1 \leq f_s/2$.
- N the number of groove edges used to extract the signal, which is either two or four, depending on the shape of the groove.

The next three subsections develop the SNR formulae for the different recording modes: constant amplitude, constant velocity and equalized records. It should be noticed that the constant amplitude mode is not used by itself for disc recording, but it is used for some frequency ranges in equalized records. However the noise standard deviation has also been computed for constant amplitude recording, to demonstrate and understand the influence of noise on each recording mode. Finally, Subsection 4.3.5.4 matches the developed SNR formulae with the standard SNR performance of phonographic records (cf. Table 2.9) to determine the maximum noise level allowed for the VisualAudio system.

4.3.5.1 SNR for constant amplitude records

In constant amplitude mode, the peak to peak groove deviation A is constant, defining the signal variance σ_x^2 :

$$\sigma_x^2 = \frac{A^2}{8} \quad (4.22)$$

The noise is assumed to be uniformly distributed in the frequency band and independent on the N edges of the groove. Thus when averaging the signals extracted on all the edges, the power of the noise is decreased by a factor N . The out-of-band noise is also suppressed. We get then the resulting noise power:

$$P_n = \frac{\sigma_n^2}{NB_n/(f_1 - f_0)} \quad (4.23)$$

The signal to noise ratio for constant amplitude records is defined using Equations 4.22 and 4.23:

$$SNR = \frac{\sigma_x^2}{P_n} = \frac{A^2 NB_n}{8\sigma_n^2(f_1 - f_0)} \quad (4.24)$$

4.3.5.2 SNR for constant velocity records

In constant velocity mode, the derivative of the position represents the amplitude of the sound signal for a given frequency f_{ref} and peak amplitude A_{ref} :

$$v(t) = \frac{ds}{dt} = 2\pi f_{ref} A_{ref} \cos(2\pi f_{ref} t) \quad (4.25)$$

The reconstructed signal $x(t)$ is equal to the original signal $s(t)$ with an additional noise $n(t)$, with variance σ_n^2 :

$$x(t) = A_{ref} \sin(2\pi f_{ref} t) + n(t) \quad (4.26)$$

The velocity $y(t)$ of the reconstructed signal is then defined with an additional noise $n_1(t) = \frac{dn(t)}{dt}$:

$$y(t) = \frac{dx}{dt} = 2\pi f_{ref} A_{ref} \cos(2\pi f_{ref} t) + n_1(t) \quad (4.27)$$

The power spectral densities of the signal $\phi_y(f)$ and of the noise $\phi_n(f)$ and $\phi_{n1}(f)$ define the respective powers P_s and P_n and thus the SNR :

$$\phi_y(f) = \frac{4\pi^2 f_{ref}^2 A_{ref}^2}{4} (\delta(f + f_{ref}) + \delta(f - f_{ref})) \quad (4.28)$$

$$\phi_n(f) = \frac{\sigma_n^2}{2NB_n} \quad (4.29)$$

As $n_1(t)$ is the derivative of $n(t)$:

$$\phi_{n1}(f) = \frac{\sigma_n^2}{2NB_n} 4\pi^2 f^2 \quad (4.30)$$

$$P_s = 2\pi^2 f_{ref}^2 A_{ref}^2 \quad (4.31)$$

$$P_{n1} = 2 \int_{f_0}^{f_1} \frac{\sigma_n^2}{2NB_n} 4\pi^2 f^2 df = \frac{4\pi^2 \sigma_n^2 (f_1^3 - f_0^3)}{3NB_n} \quad (4.32)$$

$$SNR = \frac{3f_{ref}^2 A_{ref}^2 NB_n}{2\sigma_n^2 (f_1^3 - f_0^3)} \quad (4.33)$$

4.3.5.3 SNR for equalized records

The signal and the noise of an equalized record are filtered by the preamplifier transfer function $H(f)$. Compared to Equations 4.30 to 4.32, the filtered noise and signal powers are then modified as follows:

$$\phi_{n2}(f) = \phi_{n1}(f) |H(f)|^2 \quad (4.34)$$

$$P_{n2} = \frac{4\pi^2 \sigma_n^2}{2NB_n} 2 \int_{f_0}^{f_1} |H(f)|^2 f^2 df \quad (4.35)$$

$$P_s = 2\pi^2 f_{ref}^2 A_{ref}^2 |H(f_{ref})|^2 \quad (4.36)$$

$$SNR = \frac{P_s}{P_{n2}} \quad (4.37)$$

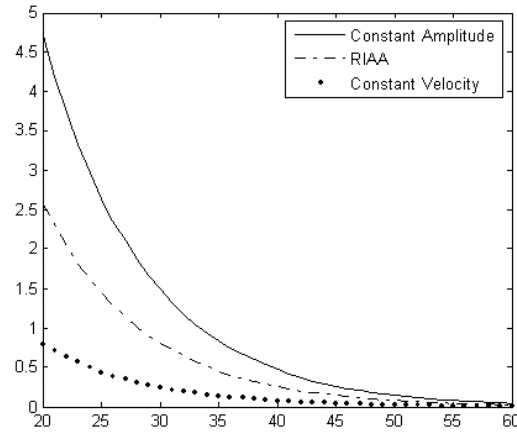


Figure 4.8: This figure shows the maximum allowed noise standard deviation σ_n (in mm) versus the SNR (in dB) for constant amplitude, constant velocity and RIAA equalized recordings.

4.3.5.4 Maximum noise level allowed

Using the formulae 4.24, 4.33 and 4.37, the maximum noise standard deviation σ_n has been computed for constant amplitude, constant velocity and equalized records (RIAA), with the following parameter values:

- The number of edges for a 78 rpm record is $N = 4$
- Image sampling frequency of 131 k-samples/rotation, which defines $B_n = 85$ kHz
- The sound bandwidth is bounded by $f_0 = 500$ Hz and $f_1 = 10$ kHz
- The reference level is defined as $f_{ref} = 1000$ Hz and $v_{ref} = 7$ cm/sec which determines $A_{ref} = 11.14 \mu\text{m}$

Figure 4.8 displays the maximum noise standard deviation σ_n allowed to reach specified SNR values. This figure shows clearly that constant amplitude recording reaches a better sound quality than constant velocity. Since constant amplitude recording has a too much limited frequency band, it is not a usable technique. Equalized recording, which stores some frequency ranges at constant amplitude and some at constant velocity, offers a good compromise to optimize the sound quality, while allowing recording a wide frequency band. However the SNR measurements are usually made on velocity recording, which is more universal and is not dependent on a specific implementation of on any equalization filter.

With the frequency responses, reference levels and frequency bands reported from Table 2.9, and using formulae 4.33, we can compute the maximum noise standard deviation allowed on one edge σ_n in order to reach the desired SNR. These values are given in Table 4.6.

	SNR (dB)	Max σ_n (μm) at 65k	Max σ_n (μm) at 131k
Shellac 78 rpm	17-37	0.17 - 1.7	0.24-2.41
Acetate / cellulose 78 rpm	37-47	0.02-0.08	0.03-0.1
NAB standard (1949) mono	40	0.06	0.08
NAB standard (1963) 33 rpm mono	55	0.0038	0.0054
NAB standard (1963) 33 rpm stereo	50	0.0048	0.0068

Table 4.6: Maximum noise standard deviation allowed to reach the desired SNR performances for a reference signal of 7 cm/sec @ 1kHz. Two image acquisition rates are considered: 65 and 131 k-samples per ring.

These maximum noise standard deviations give an estimate of the uncertainty level to reach in the image processing step, in order to extract good quality sound. These values seem pretty small compared to the radial size of a pixel (between 1 and 2.5 μm); but many subpixel methods achieve accuracy finer than a tenth of pixel in good conditions [71, 72].

4.4 Illumination variation

We consider that the lightening source of the photographic camera provides a homogeneous light over the record surface. But the light reflected by the record surface to the lens varies according to the shape and to the radial position of the groove. Thus the lightening system, together with the groove shape, has some effect on the record image. Subsection 4.4.1 studies the variation of the groove image according to the radial position on the record. The effect of light reflection inside the groove is then described in Subsection 4.4.2.

The light variation at scanning have less impact on the acquired image, as it works by transparency and as the surface to illuminate is flat. Thus scanning light inhomogeneities will mostly result in grey level difference, which have only limited implications on the extracted sound quality. Therefore the light variation will be only empirically studied later in the Chapter 8: Evaluation.

4.4.1 Radial variations

The basic illumination model considers that the record surface is illuminated perpendicularly by the lightening system and that flat parts of the record reflect the light to the lens, and that the groove's walls do not reflect the light to the lens. Such an illumination system is only possible with a perfectly perpendicular illumination provided by a small size light source. The light source of the VisualAudio camera is in fact an 83 cm diameter spiral luminescent tube, placed behind a diffusing glass. The lens mount has a diameter of 15 cm and is located in the center of the spiral tube. The model must then be enhanced to consider such illumination source.

A 78 rpm record groove is built of two walls having a $45^\circ \pm 5^\circ$ inclination to the record surface, and a groove bottom which curvature radius lies between 20 and 60 μm . We must also consider the groove depth and the illumination variation according to the groove position and modulation.

To simplify the illumination modeling, we will first consider a punctual lens surrounded by a circular light source. We assume that the groove has no depth variation and is not modulated. If we make a transversal cut of an unmodulated groove at a given radius r , we consider that all the normal to the surface lay in the same plane which passes through the lens and record center.

To enhance the model, we first consider the following hypotheses:

- The lens is in the middle of the light source, at the same height
- Working with a good diffusing glass, the spiral shape of the luminescent tube is no more visible, and the light source can be considered as homogeneous and circular with a diameter d
- The record is flat
- The record is well centered with respect to the lens and the light source
- The record is monophonic and presents then no depth variation
- The record is parallel to the light source and to the lens

For each point P located at a radial position r on the record surface, we can consider only a cut of the camera as shown on Figure 4.9: as the record, the light and lens are parallel, the light reflected by the record to the lens can only come from the cross section C of the light source which is in the plane normal to the lens bottom and which passes through the point P and the center of the light source.

Since we take picture with a 1:1 magnification, we can consider that the resulting image is equivalent to the projection of the groove light reflecting areas onto the record surface plane. Thus this projection will only affect the image of the two bottom edges e_{BO} and e_{BI} . Considering that the focus has been made on the record surface, the projection angle ϕ for each point is the angle between this point and the lens center. Such projections are shown on Figure 4.10. Since the lens is located above the center of the record, the visible effect of this projection on the record image is that the outer trace T_O is wider than the inner trace T_I by a few micrometers (depending on the groove bottom depth and curvature).

As stated in Section 3.2.2, the light will not only be reflected by the flat surfaces, but by all the surfaces having an inclination angle between α_1 and α_2 , which vary according to the radial groove position on the record. In other words, we can say that these angles α_1 and α_2 vary according to the variations of the distances c_1 and c_2 between the groove location and the further bounds of the light source on both sides. These distances c_1 and c_2 are shown on Figure 4.9.

Since these distances c_1 and c_2 (and therefore the angles α_1 and α_2) vary proportionally to the radial position of the groove, the position of the groove bottom areas which reflect the light to the lens will also change proportionally to the radial position of the groove. The effect on the groove image is a variation of the traces widths: as the radial position of the groove decreases, the width of the trace T_O decreases and the width of trace T_I increases. The effect of these shifts on the

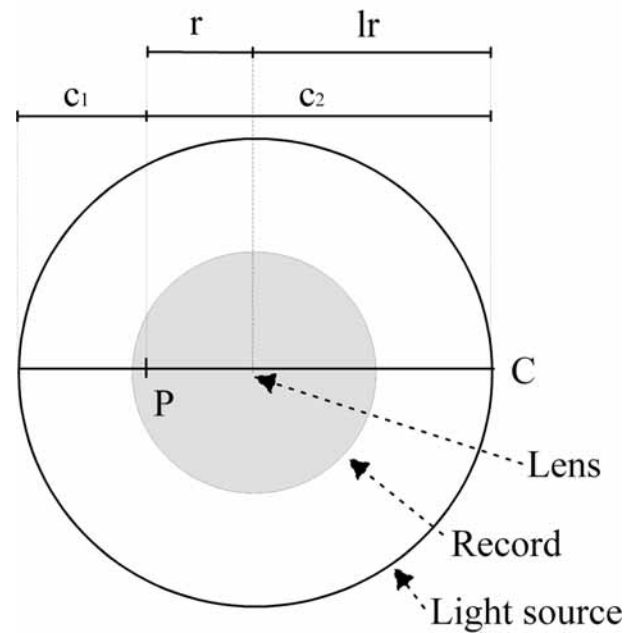


Figure 4.9: The light reflected to the lens by an unmodulated groove comes from a cross section C of the light source.

extracted sound signals will be an additional continuous component on both edges e_{BO} and e_{BI} . A sample of this variation is displayed on Figure 4.10.

Based on Equations 3.5 to 3.7, a groove simulation has been developed under Matlab to locate the light reflecting area of the groove over the record and to quantify their shifts and their effects on the extracted sound. Following parameters are considered:

- lens width
- light width
- light and lens vertical position (located on the same plane)
- groove radial position (from the center of the record)
- groove top width
- groove bottom width
- groove bottom radius
- groove angle between the surface and the walls

Considering the groove width and curvatures given in Table 2.5, the maximum shift of the reflecting areas in the groove bottom may be up to $10 \mu\text{m}$ between the outer and inner circumvolution of the same groove on a 30 cm record. The projected

shift on the groove image is lower, as the variation of the angle ϕ (between the groove bottom and the lens) compensates part of the shift. Thus on the image, the edges will be shifted up to $4 \mu\text{m}$ between the outer and inner part of the record. The effect of this shift on the extracted sound signal is negligible as it produces a very small amplitude constant component which is spread over the whole extracted sound. In comparison, the constant component amplitude which is produced by the record spiral on the extracted sound is of $150000 \mu\text{m}$.

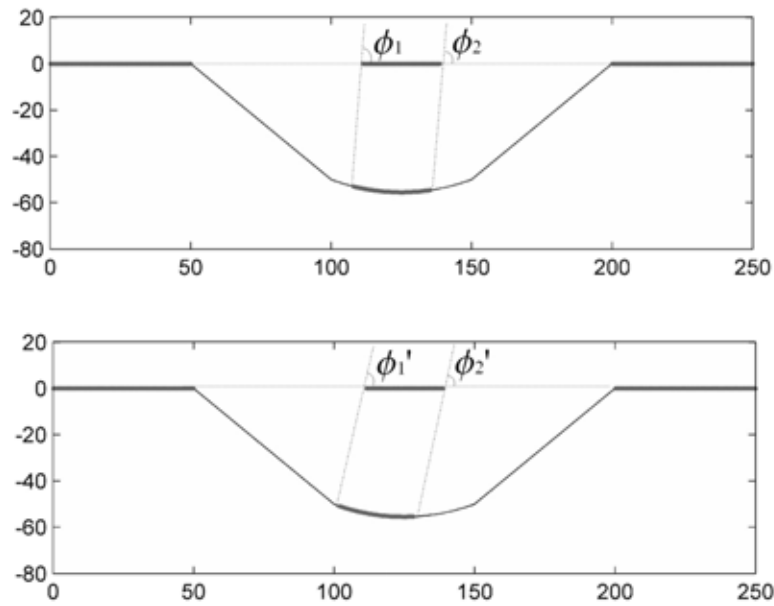


Figure 4.10: These two groove cut views are illuminated with an 83 cm diameter light source. Light reflecting areas are marked by a thick grey line. The upper groove is located at a radius $r = 5 \text{ cm}$ from the record center. The lower groove is located at a radius $r = 15 \text{ cm}$ from the record center. While the reflecting areas in the groove bottom has moved from $10 \mu\text{m}$ between the upper and lower cut view, their projection onto the record surface plane present only a small shift of around $2 \mu\text{m}$.

The radial position of a groove changes also due to the groove modulation. Therefore this will generate a small edge shift: for a $70 \mu\text{m}$ amplitude sinusoidal signal, the edge shift will be of $0.005 \mu\text{m}$. This shift will produce -82 dB harmonics, which is totally negligible.

The above calculations assumed an unmodulated groove. For a modulated groove, the groove orientation changes, which changes the distance to the outer part of the light, and will also produce an edge shift on the image. Thus the light cross section C' to consider for the point P illumination is no more equivalent to the light source diameter C . To insure a specular reflection to the lens, the cross section orientation angle φ must be twice the groove orientation, which depends on the local velocity of the recorded signal. On a modulated groove, c'_1 will increase and c'_2 will decrease compared to the initial cross section components c_1 and c_2 , as shown

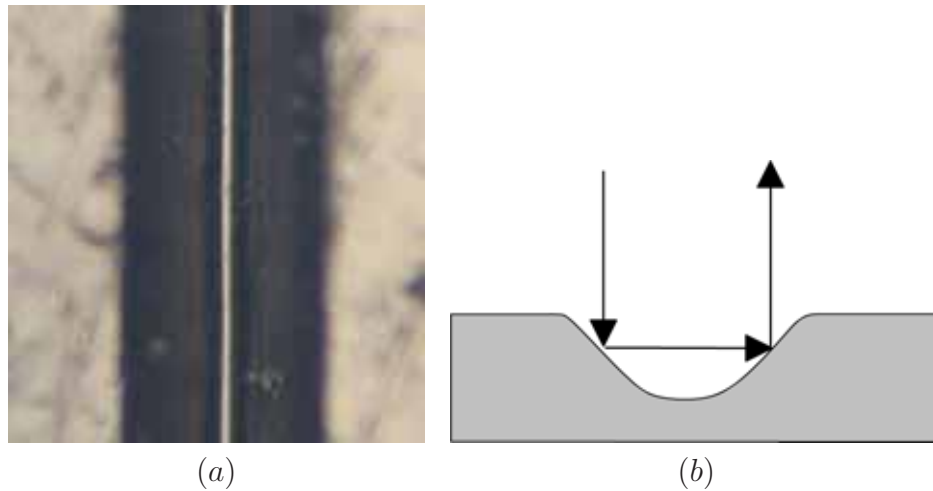


Figure 4.12: (a) Microscope top view of a 78 rpm record groove: with a perpendicular illumination, the land and bottom of the groove reflect the light back to the microscope lens. The walls of the groove reflect the light on each other as shown on (b) and thus look lighter than the bottom of the groove areas surrounding the white line.

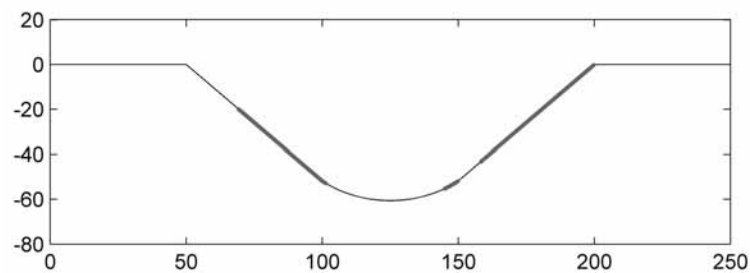


Figure 4.13: Indirect light simulation: this groove has a bottom radius of $40 \mu\text{m}$, walls inclination of 46° and is located at a radius $r = 15 \text{ cm}$ from the record center. The areas marked in black reflect indirect illumination to the lens and will then be darker on the record picture.

as the groove's walls inclination for example. But the groove simulation has been enhanced to evaluate also the indirect illumination effect. Figure 4.13 shows one situation where the groove indirect illumination will produce darker areas on the groove's walls of a record picture. Figure 4.14 shows a microscope view of a record photography made with the VisualAudio camera. The left trace of the groove looks darker: this is probably due to indirect lightening during the photography step.

The model above has considered a punctual lens, which is purely theoretical. Working with a circular lens will produce the following consequences on the image formation:

- Since the lens is located in the center of the light source, the light source cannot be considered as circular, but is a ring with an internal diameter which is equal or larger than the lens external diameter.
- The image of a point is the integration of the light cone coming from a section

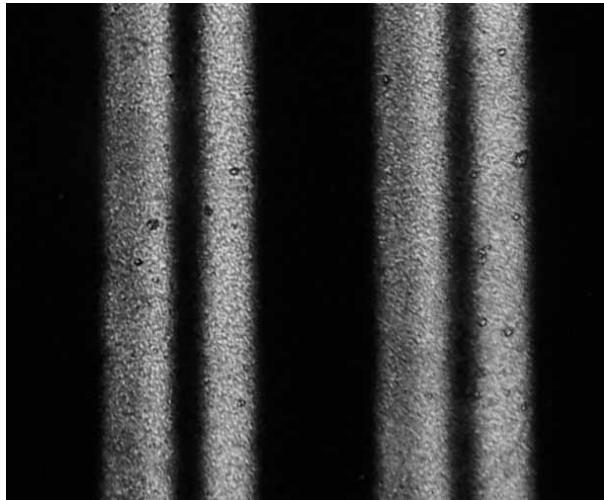


Figure 4.14: Microscope view of a 78 rpm photograph: the left trace of the groove looks darker, which is probably due to indirect lightening during the photography step.

of the light source which illuminates this point from the lens point of view. Figure 4.15 shows the section of the light source which could illuminate the groove at position r , depending on the groove surface inclination.

- We also see that the lens is placed in the middle of the considered light source section: this could slightly lower the light intensity of the reflected light for some groove surface inclinations, depending on the groove radial position and orientation.
- For groove radial positions smaller than the internal light source radius, the considered light section will be equal to the whole light source (both angles ψ will be equal to 360°). Thus there will be no variations of the distance c'_1 and c'_1 .

The impact of the circular lens model on the image and on the extracted sound hasn't been analytically modeled, because of the two following reasons:

1. The reflected light intensity variations due to the ring shape of the light on the image are pretty low, and without abrupt changes.
2. The variations of the maximum distances to the outer bound of the light source behave in the same way as the cross sections c_1 and c_2 in the punctual lens model. Thus the effects on the image will be similar, and the harmonics produced will be of the same order of magnitude (-40 dB).

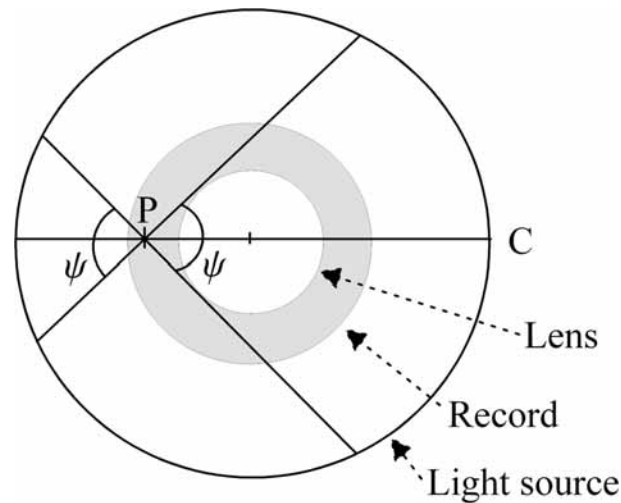


Figure 4.15: Depending on its inclination, the point P of an unmodulated groove could reflect the light from the two light source sections of angle ψ . Angle ψ is determined by the two tangents of the lens circle which pass through the point P .

4.5 Nonlinear distortions

4.5.1 Nonlinearity of recording media

Image nonlinearity is introduced by the nonlinear response of the recording media. In VisualAudio, this can occur at both the photography and the scanning steps. As presented in Section 3.1, the photographic film has a nonlinear response, which is partly described by its gamma. On the developed film, the silver grains density varies logarithmically with the incident light intensity and saturation occurs both in the black and white regions.

Saturation and non-linearity lead to some image information loss, as different exposure light intensity levels could lead to the same silver grain density on the developed film. On the other hand, saturated areas are less contaminated by the film grain noise, as the dark areas are opaque and the light saturated areas are completely transparent.

Saturation can also occur at both high and low light intensities at scanning, due to the limited sensors sensitivity. In practice, the scanner illumination system can easily be adjusted to keep the illumination level inside the sensors sensitivity for the whole record image scanning.

4.5.2 Geometrical distortions

The initial camera position and its displacement relatively to the record picture must be set accurately. The parallelism between the camera sensors and the scanned film must be guaranteed to get a well focused image. The sensors line must also stay in-line with the center of the groove spiral on the record picture to stay perpendicular to the path of an unmodulated groove.

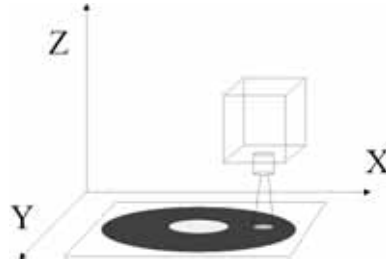


Figure 4.16: The record picture is on the XY-plane, and the CCD sensors line must be in line with the center of the groove spiral and parallel to the X-axis. A camera rotation in any of the three planes leads to geometrical distortions on the extracted signal.

No procedure has currently been defined on the current scanner prototype to measure and adjust the camera position relatively to the glass tray and record picture. But a small deviation in either direction could lead to distortions on the extracted sound. These distortions have limited effects on the extracted sound, as the camera misalignments stay relatively small.

To evaluate the geometrical distortions produced by a non-ideal camera positioning, we will study the effect of the camera rotation on itself in any of the three planes XY, XZ and YZ relatively to the glass tray (i.e. to the disc picture), as defined on Figure 4.16. The case of a camera shift along the Y-axis will also be considered, but shifts along the X-axis or the Z-axis are not of interest: the position along the Z-axis is manually adjusted prior to any acquisition to be in focus; and the initial position on the X-axis is chosen according to the beginning of the groove to scan. Any further displacement along the X-axis is performed relatively to this initial position.

These rotations and shifts will be first explained in Subsections 4.5.2.1 to 4.5.2.3, as well as their consequences on the acquired images. The subsection 4.5.2.4 will finally evaluate the effect of a camera misalignment on the sound signal sampling.

4.5.2.1 Rotation in the XY-plane and shift along the y-axis

As shown in Figure 4.17, the effect of a camera shift along the Y-axis and a camera rotation in the XY-plane is the same: the sensor line is no more in line with the center of the groove spiral. For a given local groove radius r and scanned area width s , a d_1 shift along the Y-axis produces equivalent geometrical distortions as a camera rotation of angle α in the XY-plane. The connection between these two cases is given by the following ratio:

$$\tan \alpha = \frac{d_0}{s} = \frac{d_1}{r} \quad (4.38)$$

Thus the only difference between the Y-axis shift and the XY-plane rotation is that the shift will have a constant d_1 over the whole record, and therefore the angle α variation and the shift d_0 will vary inversely proportionally to the local groove radius r . The rotation of the camera in the XY-plane will have a constant angle a

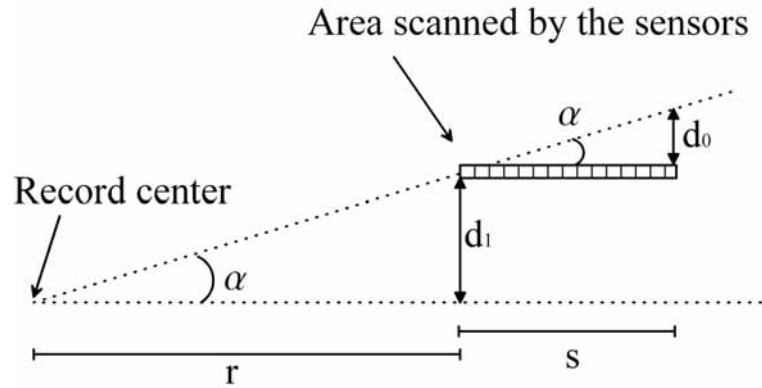


Figure 4.17: A camera shift of length d_1 along the Y-axis produces equivalent geometrical distortions as a camera rotation of angle α in the XY-plane.

during the complete acquisition, and thus a constant d_0 shift. The consequences of this shift on the signal sampling will be explained in the Subsection 4.5.2.4.

4.5.2.2 Rotation in the YZ-plane

A camera rotation in the YZ-plane will also produce a shift of the scanned area, which will therefore no more be in line with the center of the spiral. Moreover, this YZ rotation will also change the magnification of the image. The magnification m of the lens is the ratio between the image distance i (between the camera and the lens) over the object distance o (between the lens and the object):

$$m = \frac{i}{o} \quad (4.39)$$

Since the image distance i is fixed by the optical tube length ($i = 16$ cm), a rotation of angle β in the YZ-plane will modify the object distance to o' , as displayed on Figure 4.18, leading to a new magnification m' :

$$m' = \frac{i}{o'} = \frac{i}{o} \cdot \cos \beta \quad (4.40)$$

We consider that $\beta = 1^\circ$ is a reasonable maximum for the rotation in the YZ-plane. Thus a rotation in the YZ-plane with an angle $\beta = 1^\circ$ produces a magnification change of 0.02%. As long as the focus has been correctly adjusted, this magnification change is not of high importance for the image acquisition process as it will be constant over the whole image. But it may produce small gaps between rings if the radial displacement hasn't been calibrated (see Section 8.2.2.1) since the camera rotation.

4.5.2.3 Rotation in the XZ-plane

A camera rotation in the XZ-plane will result in magnification differences between the inner and outer sensors of the camera, as the sensor line will no more be parallel

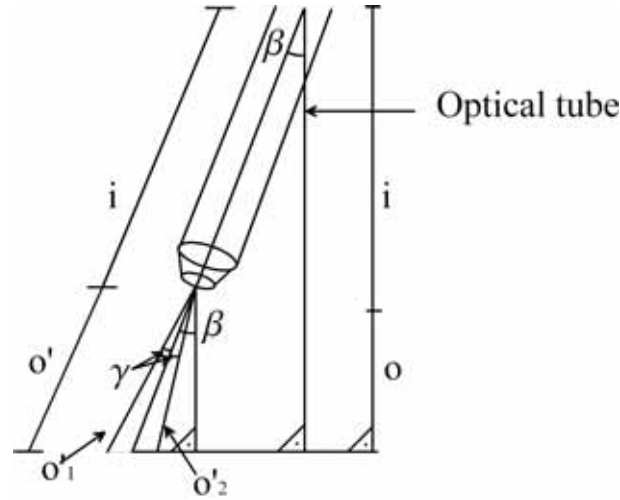


Figure 4.18: A rotation in the XZ-plane will change the magnification ratio between the inner and the outer sensors.

to the film (Figure 4.18). The magnification ratio m of the central sensor will be as stated in Equation 4.39, but for a camera rotation in the XZ-plane with an angle β , the difference of the magnifications m' and m'' for the inner and outer sensor will be:

$$|m' - m''| = |\cos(\beta + \gamma) - \cos(\beta - \gamma)| \quad (4.41)$$

With γ being the angle between the central sensor and the outer sensor optical path (see Figure 4.18). The depth difference between the object distance o'_1 and o'_2 is defined as follows:

$$|o'_1 - o'_2| = 2o' \cdot \sin \beta \cdot \sin \gamma \cdot \cos \beta \quad (4.42)$$

As $\gamma = 3.57^\circ$ for a 5x magnification lens and a 2 cm sensor width, then for a rotation angle $\beta = 1^\circ$ this magnification ratio difference $|m' - m''|$ will be 0.2%, leading to a $69.7 \mu\text{m}$ depth difference. Both sides of the acquired image will then be more out-of-focus and blurred.

4.5.2.4 Consequences of a misaligned camera on the signal sampling

If the sensor line is not in line with the center of the groove spiral, the digitized image will be slightly distorted, as displayed on Figure 4.19. This produces several effects on the extracted sound signal:

- Amplitude enhancement
- Gaps or overlaps at the joints between rings
- Sampling frequency modulation

- Phase shift between the signals extracted from the different edges of a same groove

All these effects are expanded in the following paragraphs.

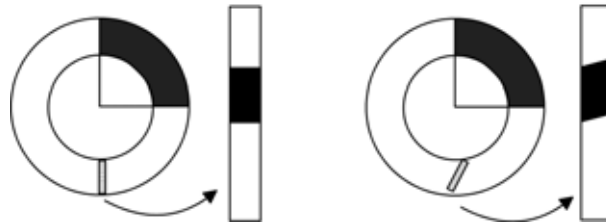


Figure 4.19: On the left, the ring acquisition is performed with a camera inline with the record center: the black section of the record corresponds to a rectangular portion of the acquired image. On the right acquisition, the camera is misaligned, leading to a distortion of the acquired image.

As shown on Figure 4.20, the amplitude A of the signal will be modified by the sampling process with a misaligned camera and transform into A' :

$$A' = \frac{A}{\cos \alpha} \quad (4.43)$$

With a camera rotation angle $\alpha = 1^\circ$, which is a reasonable maximum for an accurately adjusted camera, the amplitude of the extracted signal would be:

$$A' = 1.00015 \cdot A \quad (4.44)$$

Since the angle α is very small, this signal amplification is negligible and of no importance for the sound extraction process.

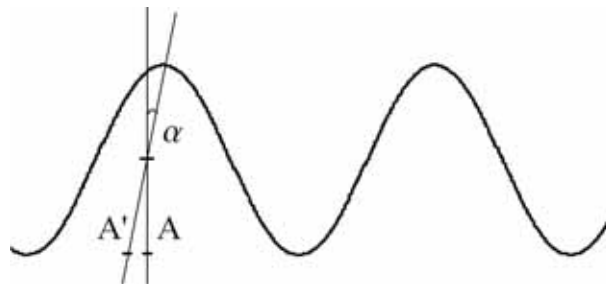


Figure 4.20: With a misaligned camera, the sound signal will be slightly amplified from amplitude A to A' . The sound signal sampling frequency will also be modulated by the signal itself.

The initial angular position of a ring acquisition depends on the local radius r . This radius is not the same at the beginning and the end of a ring. Thus consecutive rings signal will not exactly match, and there will be either a gap or an overlap at the joint between the signals extracted from two consecutive rings. Using Equation

4.38, considering a local radius $r = 100$ mm, $s = 4$ mm (for a sensor line width of 2 cm and a 5x magnification) and a rotation angle $a = 1^\circ$, corresponding to a shift $d_1 = 1.7$ mm, we get $d_0 = 70$ μm . This means that a 70 μm overlap (or gap depending on the rotation angle) will occur at the joint between two consecutive rings. This $d_0 = 70$ μm shift is equivalent to 5 to 10 pixels (or samples) for a 65K acquisition, depending on the local groove radius r .

Since the groove is engraved in spiral, the radius r varies at each circumvolution and this initial angular position varies also inside a ring from the beginning of a groove circumvolution to the following one. This leads to a higher signal sampling frequency (or lower depending on the sign of the rotation angle). The angular difference γ between the initial and final position of an unmodulated groove circumvolution is:

$$\gamma = \arctan\left(\frac{c \cdot \tan \alpha}{r}\right) \quad (4.45)$$

where r is the groove radius at the initial groove position, α is the camera rotation angle and c the radial displacement due to the spiral in one circumvolution. The sampling frequency f'_s of an unmodulated groove with a misaligned camera is:

$$f'_s = f_s \cdot \left(1 + \frac{\gamma}{2\pi}\right) \quad (4.46)$$

This signal sampling frequency varies with the radial position r of the groove, leading to smooth pitch variation of the extracted sound. We consider that an angle $\alpha = 1^\circ$ is a reasonable maximum for an accurately adjusted camera. Thus with $\alpha = 1^\circ$, a radius $r = 100$ mm, and a spiral move $c = 200$ μm , this sampling frequency vary by 0.03% which is lower than the standard turntable wow tolerance of 0.1% [55]. For a modulated groove, the sampling frequency changes in an even more complicated way, as the local radial groove position r varies according to the signal amplitude. Thus the sampling frequency will be signal dependent and it will produce harmonics on the extracted sound signal. Considering a sinusoidal signal of amplitude A and frequency f_0 , the sampled signal $x'[t]$ with a misaligned camera will be:

$$x'[t] = A' \cdot \sin(2\pi f_0 t + A' \cdot \sin(2\pi f_0 t) \cdot \sin(\alpha)) \quad (4.47)$$

It is difficult to quantify the harmonics level produced on the extracted sound, as it depends on many parameters, as for example the record rotation speed, the sound frequency and amplitude. Empirical tests have been performed on the sound extraction of a 300 Hz track. An acquisition has been performed with an ideal camera position (adjusted manually), and the test has been repeated after two successive camera rotations by 2° in the XY-plane. After a 4° camera rotation the *THD* (Total Harmonic Distortion) increased by 5 dB. Complete results are displayed in Table 4.7.

In the VisualAudio process, several signals are extracted from the edges representing the top and bottom of the groove. The sound signal is then produced with a weighted average of these signals. The initial angular position shift mentioned above

Estimated XY rotation angle β	SNR_1	THD_l
0°	37.63	-42.63
2°	36.93	-40.62
4°	36.09	-37.93

Table 4.7: Variation of the SNR and THD for different camera rotations in the XY-plane. The sampling variations due to the camera rotations increase the harmonics level. Section 8.1 provides a full description of the SNR_1 and THD_l measurements.

will produce small phase shifts between the signals representing a same groove. Small phase shifts are not audible in a sound signal if they are not correlated to the signal frequencies. But the addition of two phase shifted signals having the same frequency content produces a low pass filtering on the resulting signal. Moreover, it is important to avoid phase shift for advanced signal correction using signal correlations or autoregressive models.

4.5.3 Off-axis and pitch variation

When a sound is not played at a constant speed, it produces a sound modulation, which is called pitch variation defect. Pitch variation defects encompass wow effect, which has a once-per circumvolution period, and flutter for higher frequencies. These pitch variations can be caused by different factors in the VisualAudio system:

- Film off-axis (or eccentricity) during scanning: the center of the spiral is not exactly located at the center of rotation of the glass tray. The periodicity of this defect corresponds to one scanner circumvolution. A low frequency radial image shift is visible on the acquired ring (as displayed on Figure 4.21).
- Rotation speed variation at recording or at playback: not necessarily periodic: depends on the reason of the pitch variation: on a spring-driven recording turntable the speed may fade out, while on a motor-driven turntable, the variation can be produced by altered motor gears irregularities. There is no radial image shift on the acquired ring.
- Strong vibrations occurring around the scanner can produce vibrations of the glass plate or the camera: the modulation is not necessarily periodic. Shift will be perceptible on the acquired ring if the camera and the glass plate vibrate independently from each other.
- Deformed or warped record: a low frequency image shift is visible on the acquired ring for the deformation in the horizontal plane. Deformation in the vertical plane produces blur but no low frequencies. It should be noticed that the audio content of a warped record is not necessarily affected by wow effect: it depends whether the warping occurred before or after recording.

The major factor of wow in the VisualAudio acquisition process is the film eccentricity, considering that:

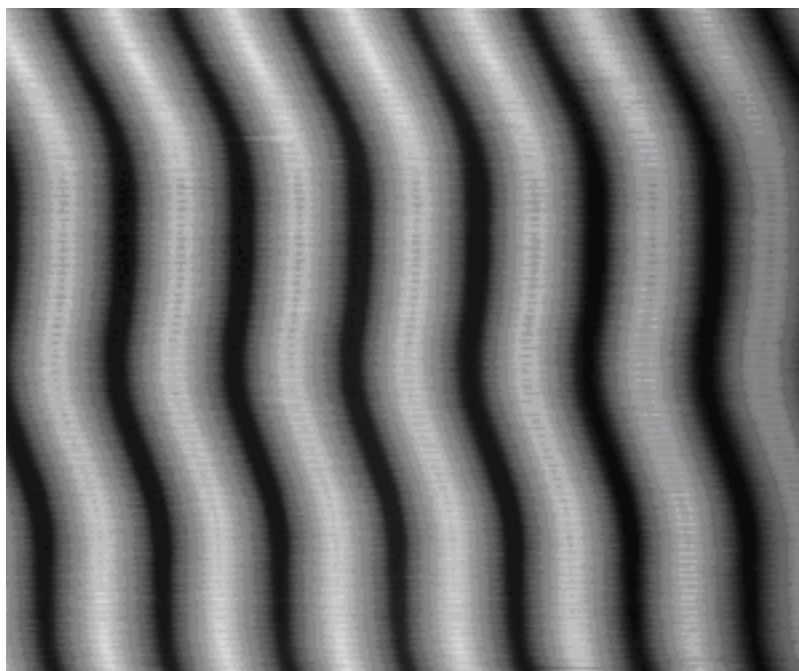


Figure 4.21: This complete ring image shows radial traces displacements. These displacements are due to the film off-axis on the scanner. It should be noticed that the image has been compressed in the tangential direction.

- The scanner rotation speed is constant, and matches the international standard of 0.1% wow variation for turntable (cf. DIN 45507 [55, 2]).
- Vibrations affecting the scanner are considerably lowered by the heavy mass of the scanner structure and can be considered as negligible when using a vibration isolation table.
- While warping occurs for vinyl, it is not a common damage for shellac and lacquer records, which are less flexible and then less subject to deformation.

Therefore we will concentrate on the film off-axis, as it is the major factor of wow defect in the VisualAudio acquisition process. When the film is not correctly centered on the scanner glass tray, the extracted sound is affected in several ways. At scanning, the off-axis film moves relatively to the camera in the horizontal plane, with a period corresponding to one scanner rotation. Thus the groove will move the same way during digitizing, generating additional low frequencies on the extracted sound signal, with a periodicity related to the number of seconds per rotation: for a 78 rpm, it produces additional frequencies of 1.3 Hz with harmonics, and for a 33 rpm it produces 0.55 Hz frequencies with harmonics. These low frequencies are non-audible and thus negligible. But these low frequencies displacements produce an uneven sampling of the groove image, which causes a modulation of the extracted sound, which is called pitch variation or wow effect.

During the scanning process, the image is sampled evenly in time at frequency f_s . Since the scanner rotation speed is constant, the image is sampled at constant

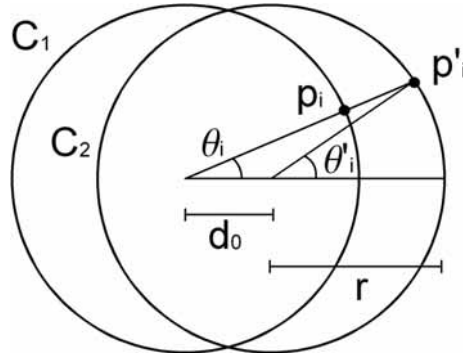


Figure 4.22: The sampling is performed evenly on C_1 . Since the record picture is rotating on the circle C_2 , sampling is uneven on the circle C_2 .

angular positions on the circle C_1 of radius r , corresponding to the current radial position on the record image. Thus the sampling angles θ_i are defined as follows:

$$\theta_i = \frac{2\pi i}{f_s} \quad (4.48)$$

We can define the sampling points p_i by their plane coordinates:

$$p_i = (x_i, y_i) = (r \cdot \cos(\theta_i), r \cdot \sin(\theta_i)) \quad (4.49)$$

If the film is not correctly centered, the sampling is performed on a circle C_2 , which center is at a distance d_0 from the center of the circle C_1 , as shown on Figure 4.22. Thus the sampling angles θ'_i and points p'_i on the circle C_2 are defined as:

$$\theta'_i = \theta_i + \xi_i \quad (4.50)$$

$$p'_i = (x'_i, y'_i) = (r \cdot \cos(\theta'_i), r \cdot \sin(\theta'_i)) \quad (4.51)$$

The angle ξ_i , which is produced by the off-axis, can be defined using the $\tan \theta_i$, which is determined on both circles:

$$\tan \theta_i = \frac{y_i}{x_i} = \frac{y'_i}{x'_i + d} \quad (4.52)$$

with

$$d = \frac{d_0}{r} \quad (4.53)$$

The angular shift ξ_i is therefore defined as:

$$\xi_i = \arcsin(d \cdot \sin \theta_i) \quad (4.54)$$

The sampling on the circle C_2 is then performed at the sampling angles θ'_i defined as follows:

$$\theta'_i = \theta_i + \arcsin(d \cdot \sin \theta_i) \quad (4.55)$$

The image sampling is then performed at angles θ'_i , which are not evenly spaced on C_2 . The extracted sound will thus suffer from a wow effect which is proportional to the off-axis displacement and inversely proportional to the radial position of the groove. Moreover, the local radius r of a modulated groove depends on the signal amplitude. Therefore the sampling variation of the extracted sound will in fact be signal dependent, producing harmonics on the extracted sound.

It should be noticed that a bad centering will also affect the scanning efficiency, as the overall surface to scan will be larger and as there will be fewer complete groove circumvolutions in each ring acquisition.

Some other phenomena could also cause pitch variation and irregular sampling: for example a varying turntable rotation speed at recording, a camera sampling time jitter, a disc warping or a non-homogeneous magnification during the picture taking phase (if the film and the record are not parallel on the camera). The main difference with the off-axis sound degradation, is that the latest-mentioned pitch variations modulate the sound signal, but they do not produce any low-frequency as the scanned image will not contain any visible move due to these phenomena (except for deformed records which present deformation in the horizontal plane).

4.6 Groove model

The camera light source is located on the lens plane and illuminates the disc from above. The groove's walls do not reflect the light to the lens and appear in light grey levels on the acquired image. Flat sections of the disc (land and bottom of the groove) reflect the light to the camera lens and thus appear black on the negative film.

Disregarding noise and blur, the ideal image of a 33 rpm microgroove is a white trace on a black background. While the walls of the coarse monophonic groove don't reflect light perpendicularly to the lens, the bottom of the groove does. Therefore there are two traces for each coarse groove on the negative film: one for each groove's wall, as displayed on Figure 4.23. Thus the ideal image of a coarse groove would be two white traces, called the outer trace T_O and inner trace T_I .

Since the sizes of the patterns to detect (land, bottom and walls of the groove) are larger than the blur size, we assume that neighboring patterns won't interfere in the edge detection process. Thus we may simplify the groove model and separate it in two or four distinct step edge models, having the same characteristics but not the same parameter values. Each edge can then be represented by a step function:

$$(A - B) \cdot u(x - x_0) + B \quad (4.56)$$

where A is the step amplitude, B the base grey level, x_0 the position of the edge and the unit step function $u(x)$ is defined as:

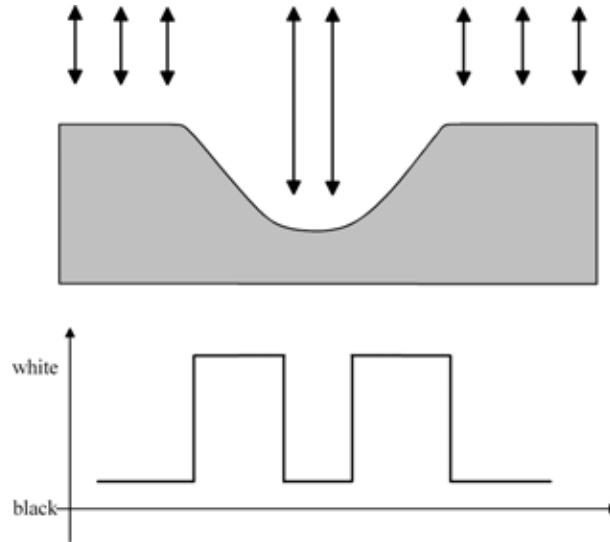


Figure 4.23: Basic luminance model for the photography of a 78 rpm groove: the land and the bottom of the groove reflect the light to the lens and appear as black on the negative film. The walls don't reflect the light to the lens and appear white on the negative film.

$$u(x) = \begin{cases} 1 & \text{if } 0 < x \\ 0 & \text{otherwise} \end{cases} \quad (4.57)$$

Blur on this edge is usually modeled by the Gaussian blurring kernel:

$$g(x, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2} \quad (4.58)$$

where σ is an unknown scale constant. We define σ_t and σ_b as the optical blur at the top respectively the bottom of the groove, and σ_m the motion blur. Optical and motion blurs are space-dependent, but they may be considered as locally space-invariant. We assume that the groove bottom curvature is regular, defining σ_s as the size of the shading blur due to the round shape of the coarse groove bottom (see 4.2.1). The step amplitude A and base level B may vary between the two traces, taking values A_1 , A_2 , B_1 and B_2 (as seen on Figure 4.24). These assumptions lead to the following 1-D edge models of a groove having its four edges e_1 , e_2 , e_3 and e_4 at positions x_1 , x_2 , x_3 and x_4 :

$$\begin{aligned} e_1(x) &= [(A_1 - B_1) \cdot u(x - x_1) + B_1] * g(x, \sigma_t) * g(x, \sigma_m) \\ e_2(x) &= [(A_1 - B_2) \cdot (1 - u(x - x_2)) + B_2] * g(x, \sigma_b) * g(x, \sigma_m) * g(x, \sigma_s) \\ e_3(x) &= [(A_2 - B_2) \cdot u(x - x_3) + B_2] * g(x, \sigma_b) * g(x, \sigma_m) * g(x, \sigma_s) \\ e_4(x) &= [(A_2 - B_1) \cdot (1 - u(x - x_4)) + B_1] * g(x, \sigma_t) * g(x, \sigma_m) \end{aligned} \quad (4.59)$$

If we rescale the step amplitudes to have a unique step amplitude value A , edges are then symmetric two by two: the top e_1 and e_4 , and the bottom of the groove e_2 and e_3 .

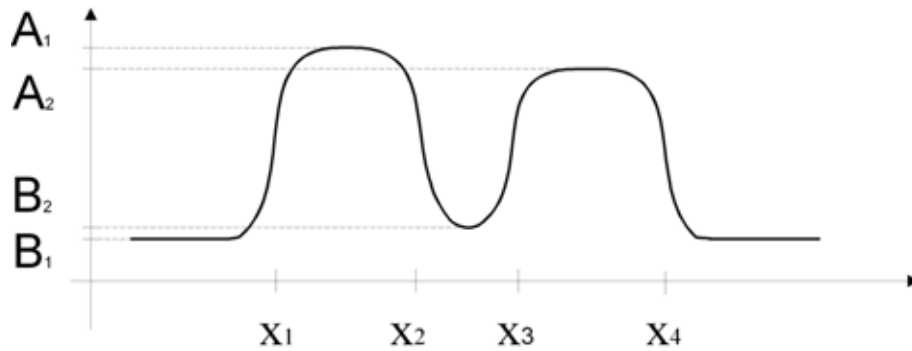


Figure 4.24: The groove model with the four edges located at x_1 , x_2 , x_3 and x_4 .

The base level B_1 corresponds to the groove bottom, and B_2 to the land of the record. Due to the difference in shape and illumination of these two regions, B_1 and B_2 may have different values. The land may also be smoother than the groove bottom, which has been engraved by the recording stylus, leading to different B_1 and B_2 values.

The step amplitudes A_1 and A_2 correspond to the two traces amplitudes. A_1 and A_2 may take different values, mainly due to illumination variations and the indirect light effect explained in Section 4.4.2.

Disregarding noise, the step amplitudes and base levels can be affected by the image acquisition process. Blur doesn't affect the step amplitudes, as the traces are much wider than the blur size. It shouldn't affect the base level B_1 , which corresponds to the land, as the grooves are usually separated by a distance greater than the blur size. But if the two bottom edges are too close, they blurred each other, resulting in an increased base level B_2 . Saturation could occur at both photography and scanning, and may affect any step amplitude or base level value.

The transition between the base level and the step amplitude is composed of a linear area L , which covers more than $15 \mu\text{m}$, which represent 6 pixels with the $4\times$ magnification and 15 pixels with the $10\times$. To measure the linear accuracy, we use a linear least means squares fitting of L . The coefficient of determination R^2 is then used to evaluate the quality of the fit: R^2 is a value between 0 and 1, where 1 means a perfect fitting [73]. By fitting the points which belongs to the linear area of the slope with a linear least squares method, we get $R^2 > 0.995$, for the non-damaged groove edges. This shows that the linear part is very close to a straight line. The linear and non-linear part of the transition are displayed on Figure 4.25.

4.7 Conclusion

This chapter described the image formation process in the VisualAudio system. It presented the various image degradation types: blur, noise and nonlinear distortions.

The total system resolution was presented in Subsection 4.2.7 and we showed that the optical blur B_{camera} produced at the photography (Subsection 4.2.2) is obviously

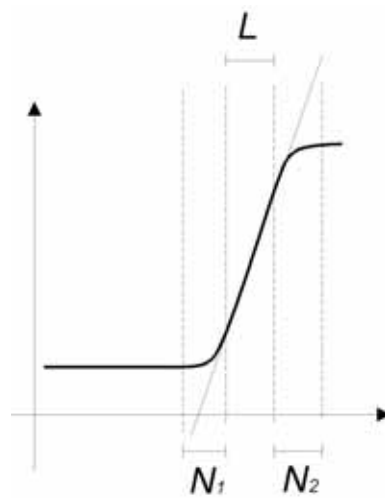


Figure 4.25: The transition between the base level and the step amplitude is composed of a linear area L and two non-linear areas N_1 and N_2 .

the limiting factor. The major consequence of the blur is a low pass filtering of the groove deviation and therefore of the signal. The blur does in fact filter away part of the out of band noise. A blur decrease would therefore not necessarily improve the inband signal to noise ratio, as the high frequency noise would increase.

The noise and local degradations can be produced by many facts and produced many different patterns on the groove image, depending on the illumination, on the type of record and on the type of film. It is then difficult to characterize all of the possible local degradations.

Most of the nonlinear distortions presented in Section 4.5 do not produce major degradations of the sound signal if the acquisition devices are correctly adjusted.

The properties of the groove image and the groove model developed in this chapter will be used for the groove extraction process, which will be expanded in Chapter 5. The corrections of the image degradations presented in this chapter will be described in Chapter 6.

Chapter 5

Groove extraction

The task of the groove extraction step is to estimate the precise groove position on the record images, based on the groove model and on the knowledge of the degradations that may appear on the acquired images, as described in Chapter 4.

Section 5.1 first gives an overview of the groove extraction algorithm. The three steps of this algorithm are then expanded: Section 5.2 describes the edge detection methods used in VisualAudio, Section 5.3 explains the groove reconstruction over one ring (as defined in Section 4.1) and Section 5.4 focuses on the groove reconstruction over the whole record. Section 5.5 finally summarizes the groove extraction algorithm.

This chapter assumes a relative good image quality, with low degradation level. Processing of more degraded images will be studied in the following Chapter 6.

5.1 Groove extraction algorithm

In order to understand the groove extraction process, Subsection 5.1.1 first explains the main properties of the groove images that can be used to extract the groove position. An algorithm overview is presented in Subsection 5.1.2.

The specific terms used to describe this algorithm (such as the trace, the ring or the radial and tangential direction), refer to the definitions presented in Section 4.1.

5.1.1 Basic properties of the groove image

The image analysis and groove model were presented in Chapter 4. In this analysis, it appears that the images are noisy and blurred, with several sources of blur, noise and degradations. Therefore there are many kinds of degradations and many possible combinations of these degradations. Thus the basic idea for the image processing is to look for the patterns which present the characteristics of the groove and reject other patterns. In Section 4.6, we have already modeled the groove in the radial dimension. Some other characteristics of the traces, the groove and the record picture can be used to accurately extract the groove displacement over the image. These basic assumptions are listed as follows:

- On the image, traces are characterized by lighter grey levels. Since the images are highly blurred, there are important transitions between the dark and the light grey areas, as presented in Sections 4.2.7 and 4.6.
- Smooth groove displacement: groove displacements between two consecutive samples are bounded and will not exceed a few micrometers, as demonstrated in Subsection 4.2.5.
- Groove width: the traces widths of an undamaged monophonic groove are almost constant, and the width changes are smooth. The width of the groove bottom then presents also only smooth variations.
- Traces don't overlap.
- Traces are continuous and cross the ring in the tangential direction.
- The spiral shape of the groove produces a smooth continuous displacement of the traces over the acquired image. Thus all the traces are shifted by the same amount between the beginning and the end of the ring.
- There are different levels of blurs applied on the different patterns of the image. Thus the image gradient analysis may provide interesting information about a specific pattern on the image.

Thus all of these properties will be used at different stages of the image processing in order to detect the edges, to follow the traces and to reconstruct the groove. Two of these properties will be further analyzed in the next two subsections in order to better understand how they can be used in the VisualAudio image processing: the trace width and the image gradient.

5.1.1.1 Trace width

The width of monophonic grooves is almost constant over the whole record. This means that, in theory, the width of the traces on the acquired images is also almost constant. Therefore the width of the traces, for both single and double trace grooves, could be used as a constraint to accurately extract the traces edges during the image processing step.

If the trace width is constant, this means that the edges are parallel on the acquired image, and that, disregarding noise, the extracted signals from both edges are synchronized and in phase. Thus when a width variation occurs on the image, it locates some edge degradation. Width variations can then be used to detect edges degradation.

Unfortunately, the width is in fact the distance between the two edge positions. Therefore it is the difference of two noisy signals and the noise standard deviation σ_w of the width is then equal to the noise standard deviation σ_e of a single edge multiplied by $\sqrt{2}$:

$$\sigma_w = \sigma_e \cdot \sqrt{2} \quad (5.1)$$

Then the width variation is more noise sensitive. Moreover, several kinds of width variations occur on records. In Section 4, we have already explained some variations between both traces of the same groove, as well as variations for the same trace between the outer and inner part of the record. Observations of many record acquisitions has lead us to the conclusion that the trace width shows important variations over the record, especially on direct cut record. Variations of 10% of the trace width are common on the VisualAudio acquired images even within small intervals (less than a few degrees). This can be due to the disc manufacturing, or due to a non-flat surface of the record, which will then change the local focus at photography. More important trace width variations due to the engraving process are also observed on some records.

Figure 5.1 shows an example of trace width variations occurring on all the traces. These variations do not affect the final sound quality, as it simply adds out of phase components to the signal extracted from both edges. These out of phase components are eliminated when combining the two signals into a final sound.



Figure 5.1: The trace and groove width show important variations.

Another problem arises with the width variation: a width variation detects the presence of an image degradation, but it doesn't provide any information on the kind of degradation and of the position of this degradation. Thus which edge should be corrected? The edge which shows the more or less variation? Figure 5.2 shows a sample of asymmetric traces: which edges or trace is the correct one? The local context on the image is not sufficient to answer this question, and this type of correction must be performed on the signal level as a post-processing.

Several kinds of damage detection based on the width variation have been tried up to now; but most of the detected image's local degradations were better detected by the impulsive noise detection scheme applied on a single edge (Section 6.3.5). Moreover, the width variations show a lack of precision, as it is much more difficult to determine which edge to correct. Thus, width variation does not appear as a reliable property for accurate detection of impulsive noise and spurious pixels. Therefore we decided to use the width variation neither for the edge detection, nor as a local degradations detection process.

The width is then used only in two cases:

- To define the scale of the operator used for the coarse edge detection: see Section 5.2.3.

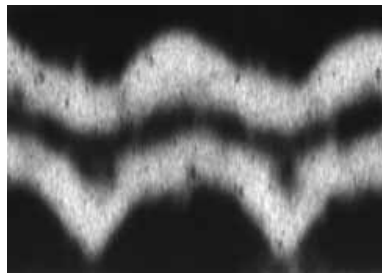


Figure 5.2: Asymmetric traces: which edges or trace is the correct one?

- To check the validity of the groove following process, when several edge points are available for the same trace at the same time. This processing will be further described in Subsection 5.3.1.4.

5.1.1.2 Image gradient

The gradient of the acquired images contains useful information for the groove extraction. Figure 5.3 (a) shows a sample acquired image of a 78 rpm, as well as its convolutions with a tangential (b) and a radial (c) 3×3 Sobel mask. These gradients present different properties in both directions:

1. Tangential gradient is very sensitive to noise and provides almost no information for the edge localization.
2. Radial gradient performs good edge localization, but with a poor accuracy, as the localized area is around $20 \mu\text{m}$ wide.

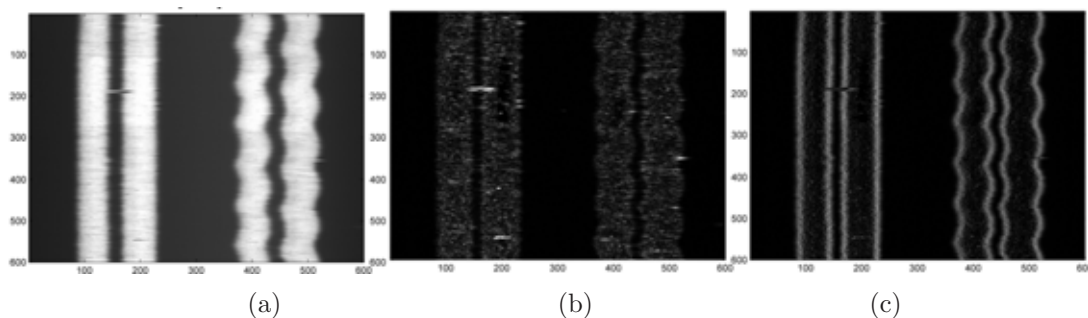


Figure 5.3: (a) Displays a sample of a 78 rpm acquisition image. The radial direction corresponds to the horizontal. The two images on the right result from the convolutions from the image (a) with Sobel masks in the tangential (b) and in the radial direction (c).

Thus the radial gradient is an interesting tool for a coarse edge detection, to locate roughly the traces on the acquired image. This coarse edge detection is described in Section 5.2.3. The tangential gradient is used for noise and damages localization, as described in Section 6.3.3.

5.1.2 Algorithm overview

The groove extraction algorithm is decomposed into three steps:

1. Detect all the possible edge points using edge detection methods. This step will be described in Section 5.2.
2. Using the detected edge points, build the traces (as defined in Section 4.1) that represent the groove over the ring. This step will be described in Section 5.3.
3. Using the traces, reconstruct the groove position over the whole record. This step will be described in Section 5.4.

Considering the groove model from Section 4.6 and the image properties described in Subsection 5.1.1, we decided to perform the image analysis in each axis separately. Therefore the edge detection is performed on the radial direction and the trace reconstruction in the tangential direction. Several reasons lead to a 1D edge detection process:

- We need to extract the radial displacement of the groove, which corresponds to a 1D displacement.
- For current record images and at small scales, the noise produces higher tangential gradient than the groove displacement. Use of larger scale operators in the tangential direction to smooth the noise would also smooth the extracted sound by applying a low-pass filter.
- Blur is not spatially invariant (as presented in Subsection 4.2.10) and may vary in the tangential direction.
- Due to the sampling process, pixels are non-isometric (see Section 3.3.4). Then we have much more accuracy for localizing edges in the radial direction than in the tangential one.
- In Subsection 5.1.1.2 we showed that it is easier to detect noise when working in the tangential direction.
- The groove is considered as perpendicular to the CCD camera during one sample acquisition and small displacements of the groove are considered as motion blur (see Section 4.2.5).
- The transition (as defined in 4.1) of an edge on one line may be corrupted and the corresponding edge point should then be rejected and marked as *undefined*. By working line by line, we just reject the transition information from the corrupted lines. The edge correction will be performed in a later stage with the surrounding edge information. By working with information provided by several acquired lines, we would need either to adapt the extraction algorithm for the edges surrounding a non valid transition, in order not to use the corrupted

information (i.e. make a local image correction), or we must reject also the valid transitions surrounding the corrupted one, in order to apply the same extraction process on all transitions.

Therefore the edge detection is applied line by line in the radial dimension. The tangential dimension is then used to accept or reject the detected edge points and to reconstruct the groove position over the ring (as presented in Section 5.3). The tangential direction information will be also used later to detect and correct the signal in a post-processing step (see Chapter 6).

Figure 5.4 sketches the groove extraction algorithm. The different steps of this algorithm are developed in the next two sections.

```

Initialize the traces
for each ring
  for each line
    Coarse edge detection: rough localization
    Fine edge detection: detect the edge point candidates
    for each trace
      Select the edge points candidates that best match the trace
    Groove reconstruction over the ring
  Groove reconstruction over the entire record

```

Figure 5.4: Pseudo-code of the groove extraction algorithm.

As presented in Chapter 4, a blurred edge can be modeled as the convolution of the unit step with a Gaussian kernel. The result of this convolution is a ramp edge, which is spread over several pixels and is called the transition. Based on many image observations, it appears that only the low grey levels area of the linear part of the transition is reliable to detect the position: the light areas of the traces are much more noisy and the non-linear part of the edges present important variations. Some examples of traces intensity profiles are shown on Figure 5.5.

5.2 Edge detection

Edge detection is a well documented topic of image processing and many methods and operators are presented in the literature. Subsection 5.2.1 briefly presents some of the existing methods proposed in the scientific literature. The choices for the VisualAudio edge detection are exposed in Subsection 5.2.2, and the two steps of the chosen method are then expanded in Subsections 5.2.3 and 5.2.4. An evaluation of the error produced by this edge detection process is finally presented in Subsection 5.2.5.

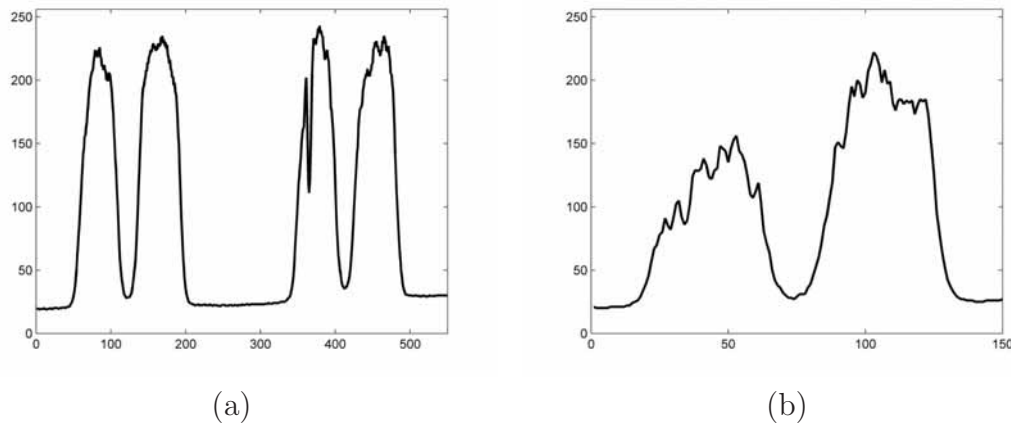


Figure 5.5: Samples of groove intensity profiles of (a) two grooves of a direct cut disc, (b) one groove of a shellac pressed record. The horizontal axis corresponds to the radial dimension, and the vertical axis represents the grey level values.

5.2.1 Literature review

This section does not present a complete review of the existing methods, but it briefly describes some of the interesting approaches used for edge detection.

According to Ye, accurate subpixel edge detection methods can be classified in three categories [69]:

- Moment based methods,
- Edge model fitting or least-squared error-based methods,
- Interpolation of the image data or their derivatives.

An overview of these three categories is presented in the next three subsections.

Several other methods can be used for edge detection, but are difficult to classify within the above mentioned categories, such as multiscale methods (Subsection 5.2.1.4) and snakes and active contours (Subsection 5.2.1.5).

5.2.1.1 Moment based methods

Tabatabai et al. have developed an ideal 1D step edge model which uses the first three gray level moments to solve three unknown model parameters. These parameters are the edge location, the background level and the contrast. The edges are thus located to subpixel accuracy [74]. Since the step edge model is not realistic for natural image processing, Shan and Boon proposed a blurred edge model for the 1D and 2D edge detection [75]. Then they used the first three spatial moments to determine the edge location, the blurring factor and the contrast.

Since numerical differentiation is used in these methods, noisy images may produce large errors in the edge localization.

5.2.1.2 Edge model fitting

An ideal blurred edge can be modeled as a one or two dimensional parameterized ramp signal. The edge locations are then estimated by fitting this model to the grey level values of the image. If the fit is sufficiently accurate, an edge is assumed to be located at the given position with the same parameters as the ideal edge model.

Hueckel, for example, has developed an algorithm to determine the presence of edges and lines by fitting a region of data to a Hilbert space with 9 parameters. Location reaches subpixel accuracy; but there is no mention of the effect of noise over the localization accuracy [76].

5.2.1.3 Interpolation

The interpolation based methods reach subpixel accuracy by interpolating the image data [77] or their derivatives [78, 79].

Baba et al. developed a one-dimensional subpixel edge detection method for measurement and object localization: they approximate the curvature of the first derivative of an image by a second order polynomial, and consider its maximum as the edge location [71]. This method reaches a precision of a $1/10^{th}$ of pixel in good conditions.

Truchetet et al. proposed a two steps method with interpolation: a first approximation of the edge position is determined by detecting the local maximum of the spatial gradient; a more precise determination of the maximum of the gradient is then obtained by a local computation of a spline interpolation of this gradient [78].

5.2.1.4 Multiscale methods

A step discontinuity in the grey levels of an image may either denote the presence of an edge or the presence of noise on the captured image. Moreover different physical boundaries from natural images produce different kinds of grey level transitions on captured images. Thus the choice of the scale for an edge detection application is of high importance. At large scale, only the coarser intensity changes are extracted, but edges are shifted from their original location. At smaller scale even the finer edges are detected but the results are very noise-sensitive. That is why multi-scale edge detection algorithms were introduced [80, 81, 82, 83]. These techniques propose to apply an edge detector at different scales and to combine the extracted edges. One of the main issues is to choose the correct scale corresponding to the type of edges to detect. If objects are widely spaced, accuracy of the edge detection increases at finer scales; but if the noise level is high, localization accuracy becomes poor at the finest scale.

5.2.1.5 Snakes

Snakes, or active contours, are deformable curves that are used to recognize objects from approximate models. The snake starts from an initial contour that is close to the true boundary. It is then deformed to adjust to the image features. The snake

evolves under the influence of internal forces coming from within the curve itself and external forces computed from the image data: the final detected boundary is obtained by minimizing the snake energy produced by these two forces [84]. Snakes are used in many applications, that include edge detection [85, 86], segmentation [87] and motion tracking [88].

Snakes boundary detection still present several problems: snakes are attracted by spurious edge points, they are subject to degeneration (shrinking and flattening), and they have difficulties to detect boundary concavities. Such concavities may appear on images of record, depending on the scanning frequency, as well as the recorded frequencies and amplitudes.

5.2.2 Edge detection algorithm

As presented in Section 5.1.2, the edge detection is performed on each line of the acquired image. Each line is represented by a sequence L of grey level values. This sequence can be displayed using the intensity profile, as shown on Figure 5.6.

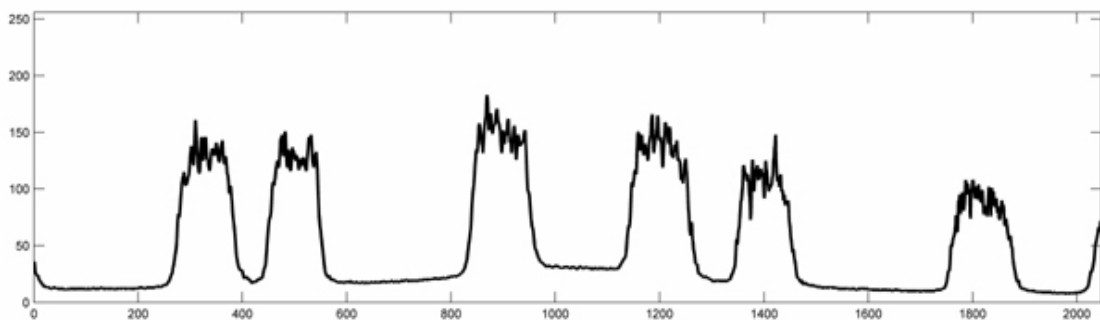


Figure 5.6: Intensity profile of a line: the horizontal axis corresponds to the radial position and the vertical axis to the grey levels value.

Among the methods presented in Subsection 5.2.1, moment based and gradient interpolation based methods do not seem to be relevant solutions to locate such edges accurately, as the linear part of the transition is spread over 6 to 12 pixels (see 4.6), which means that a local derivative does not locate the extremum accurately.

The VisualAudio edge detection uses two alternative methods. The first edge detection method is an interpolation based method, which uses a local threshold based on the local grey level extremum, as described in Subsection 5.2.4.1. The other two edge detection methods described in Subsections 5.2.4.2 and 5.2.4.3 use edge model fitting: the linear part of the transition can be accurately fitted by a straight line, as demonstrated in Section 4.6. Thus we used a least means squares fitting of this linear part.

Since the size of the patterns to extract is approximately known, the VisualAudio edge detection algorithm is inspired by the multiscale methods: first we detect the presence of the traces edges on the image, using a large scale operator (see Subsection 5.2.3). This operation locates the edges roughly by a coarse edge detection process.

A finer scale operator is then applied to get accurate positions of the edges (see Subsection 5.2.4).

5.2.3 Coarse edge detection

The purpose of the coarse edge detection step is to detect the traces roughly on a smooth de-noised image. Thus we detect the presence of all the traces and get approximate locations of the edges. Knowing these locations we can easily have a precise estimation of the local step amplitude A and base value B (as defined in the model presented in Section 4.6), even in case of large luminance variation over the image. This coarse edge detection skips also part of the false detections that are induced by dust or damages, as it locates only objects having the desired size (the size of the traces). The coarse edge detection is implemented by a convolution with a double box filter $b(x)$ defined by a $\lambda \times 1$ kernel:

$$b = [-1 \dots -1 \ 0 \ 1 \dots 1] \quad (5.2)$$

Since the groove width and trace width w are almost constant over the whole record, the scale of the double box filter is defined by:

$$\lambda = \alpha w \quad (5.3)$$

with α being usually between 0.2 and 0.8. The result of this convolution is a smooth approximation of the derivative of the intensity profile (as shown on Figure 5.7), which extremum roughly locate the transitions. The region between two extremum either locates a trace or a dark area. Therefore we can estimate the local step amplitude A and base value B by taking the maximal, respectively minimal, value in each of these regions. Several other methods have been tried to evaluate A : taking the maximum of a smoothed approximation of the acquired line, taking the median value over certain range or the median over a square area. But it appeared that using the maximum is more efficient, more reliable and results in the same or even better audio quality than the other above mentioned methods. The effectiveness of this method can be explained by the symmetry of the edge detection error, as described later in Subsection 5.2.5.

Another way to detect approximately the traces over the image would be to initialize the process as described above at the beginning of the image processing, and then to propagate this information in the tangential direction over time using the edges detected at the time i to get the location of the edges at time $i + 1$. However, such time propagated information also has the disadvantage to propagate errors. Due to scratches or other physical damage of the record, trace cuts may occur all over on the image and lead to the loss of some traces. Thus such coarse detection by propagation will lead to many errors for images of highly damaged record. Therefore the choice has been made to perform the coarse edge detection at each sampling time.

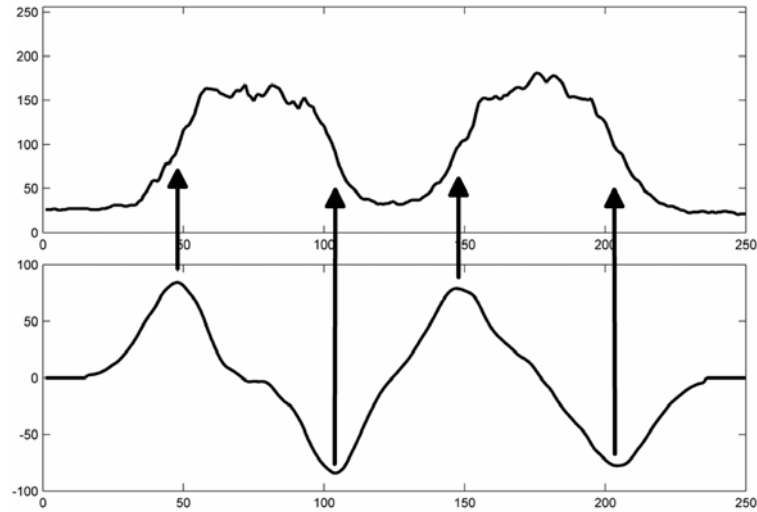


Figure 5.7: Intensity profile of a groove (above) and the smooth derivative approximation obtained by convolution with the double box filter (below). Minima and maxima roughly locate the edges (arrows).

5.2.4 Fine edge detection

The fine edge detection accurately locates edges on each acquired line. It then returns a collection of edge candidate couples (cr_k, cf_k) , each containing one raising cr_k and one falling cf_k edge candidate, with the raising edge being on the outer side of the falling edge. If a trace exceeds the limit of the ring in the radial dimension and produces only one edge candidate (the falling edge on the outer side or the inner edge on the inner side), it is considered as not interesting for the current ring processing and removed from the candidate collection.

Below, three alternative methods are presented for the fine edge detection. A first threshold based method is introduced in Subsection 5.2.4.1. A linear fitting method is then developed in Subsection 5.2.4.2 and enhanced in Subsection 5.2.4.3.

5.2.4.1 Local threshold

One of the constraints given by the model presented in Section 4.6 is to detect the position relatively to the local step amplitude A and base value B (which were determined by the coarse edge detection presented in Subsection 5.2.3) to keep the edges symmetric two by two. The simplest method to satisfy this constraint is to use an adaptive threshold τ defined as follows:

$$\tau = \beta \cdot (A - B) + B, \quad \beta \in [0, 1] \quad (5.4)$$

In practice β should be restricted to the $[0.2, 0.7]$ interval in order to be robust against noise, and to return positions located on the linear part of the edge slope. Since the light grey areas (high amplitude on the intensity profile) are noisier, a low β value produces results that are more robust to noise and degradations.

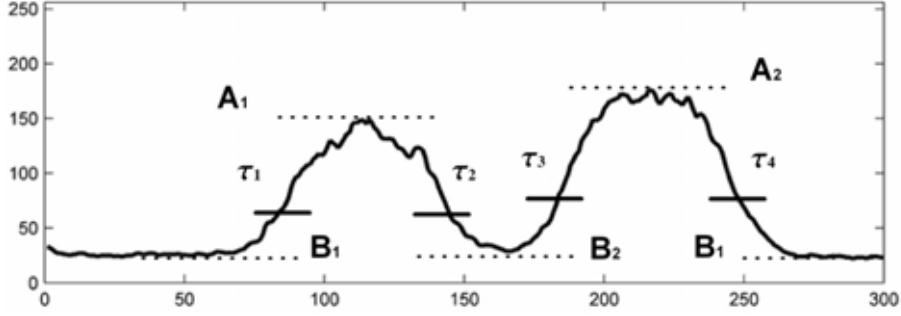


Figure 5.8: Edge detection by local thresholding: the local thresholds τ_i are determined using a linear combination of the amplitudes A_i and base levels B_i (dotted lines).

The thresholding is applied on the sequence L corresponding to one acquired line. Subpixel accuracy is then achieved by linear interpolation between the two consecutive pixels $L(i)$ and $L(i + 1)$ which satisfy:

$$\begin{aligned} L(i) \geq \tau \quad \text{and} \quad L(i + 1) < \tau \quad \text{for the raising edges} \\ \text{or} \quad L(i) < \tau \quad \text{and} \quad L(i + 1) \geq \tau \quad \text{for the falling edges} \end{aligned} \quad (5.5)$$

5.2.4.2 LMS edge detection

The LMS edge detection performs a linear fitting that minimizes the mean square error over the linear part of the transition. The LMS edge detection is performed in three steps:

1. Determine the line points, which will be used for the linear fitting.
2. Find the equation of the linear fitting.
3. Localize the edge point using the linear fitting.

These three steps are expanded in the next paragraphs.

The coarse edge detection roughly localized the trace edges. The points belonging to the linear part of the transition must then be located around each roughly detected edge. This localization process is similar to the coarse edge detection (see Subsection 5.2.3), but uses a smaller double box filter $b_2(x)$ of size $\lambda_2 \times 1$, where λ_2 is proportional to the size of the linear area of the transition:

$$\lambda_2 = \frac{m}{s} \cdot p \quad (5.6)$$

where m is the magnification ratio at scanning, p the approximate size of the linear part in micrometers and the s the physical size of one CCD sensor: $s = 10 \mu\text{m}$ for the current CCD camera and usual values of p range from 5 to 15, as empirically determined in Section 4.6. The convolution of the double box b_2 with the acquired line L returns the sequence D :

$$D = b_2 \otimes L \quad (5.7)$$

The index x_{ext} of each extremum from D roughly locates the middle of the linear part of a transition. The linear fitting must then be performed on the line points x_i located around the index extremum x_{ext} and which satisfy:

$$D(x_i) > \chi \cdot |D(x_{ext})| \quad (5.8)$$

where χ is defined between 0.8 and 0.95. Equation 5.8 determines the n pixel values which will be used for the linear fitting. It should be noticed that the number n of values may change from one fitting to the next. The selection of the points used for the linear fitting is shown on Figure 5.9, and the fitting is displayed on Figure 5.10.

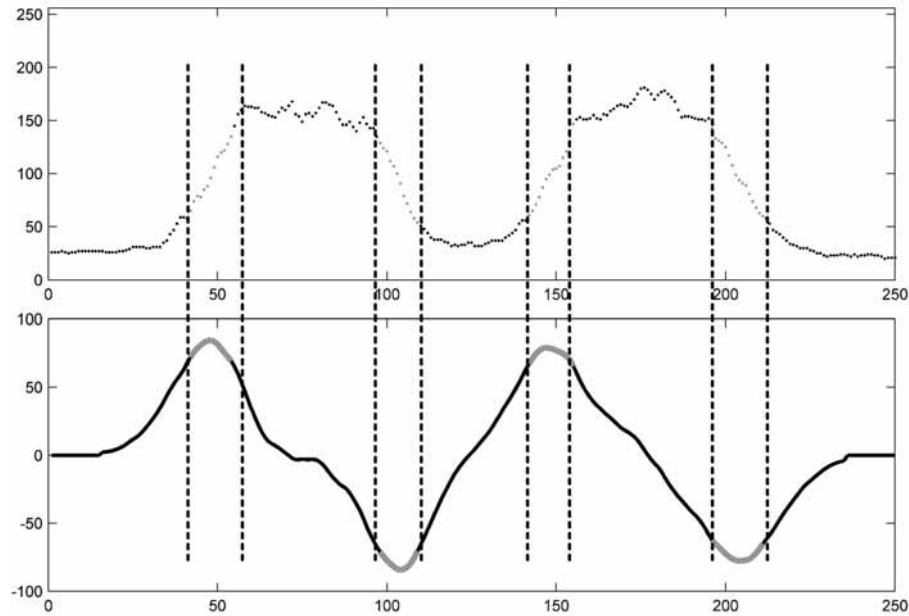


Figure 5.9: The upper graph shows the profile of a groove and the lower graph the smoothed derivative approximation of the profile. The points which are located around the extremum of the smoothed derivative are selected to be used for the linear fitting (see Figure 5.10). These selected points are located by the dashed lines.

The solution to the linear fitting can be written as a sequence V using the matrix U and the sequence Y defined with the current acquired line sequence L as follows:

$$U = \begin{bmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_n \end{bmatrix}^T \quad V = [b \ a]^T, \quad Y = [L(x_1) \dots L(x_n)]^T \quad (5.9)$$

The solution of the linear fitting is computed as follows [73]:

$$V = (U^T U)^{-1} (U^T Y) \quad (5.10)$$

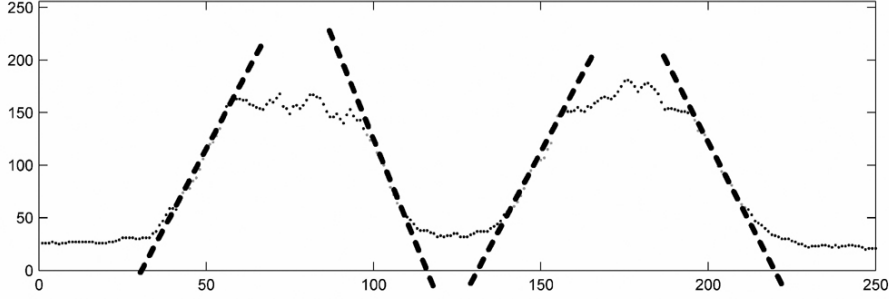


Figure 5.10: The linear fitting (dashed lines) are performed using the points selected in Figure 5.9.

The candidate edge point c is then the solution of the equation:

$$ac + b = \tau \quad (5.11)$$

where τ takes values in the interval $[0 \ 1]$.

The relevance of the candidate edge point c is estimated using the coefficient of determination R^2 of the linear fitting. The coefficient of determination R^2 is defined as the ratio between the regression sum of squares (SSR) and the total sum of squares (SST) [73]:

$$R^2 = \frac{SSR}{SST} \quad (5.12)$$

R^2 is a value between 0 and 1, where 1 means a perfect fitting [73].

In the LMS edge detection, the coefficient of determination of each fitting is compared to a predefined tolerance ζ in order to accept or reject the LMS estimation and the corresponding edge candidate c . If $R^2 > \zeta$ then the computed candidate edge point c computed by Equation 5.11 is accepted, otherwise it is rejected. Usually, ζ has a value of 0.95 or higher. If the LMS estimation is rejected, then there is no edge candidate c for the current extremum x_{ext} .

5.2.4.3 Weighted LMS

The weighted LMS edge detection is based on the iteratively reweighted least squares method (IRLS) described in [89] and [90]. The basic idea of the weighted LMS edge detection is to lower the impact of the spurious pixels on the linear fitting in order to get more accurate fitting and edge locations. Therefore a weight is attributed to each pixel and reevaluated at each step of the iterative process. The weighted LMS is applied similarly to the LMS edge detection (as presented in Subsection 5.2.4.2), but the solution of the linear fit presented in Equation 5.10 is replaced by the following iterative process:

1. The weighting matrix W is initialized as the identity matrix of size n .
2. $k = 0$.

3. $V(k) = (U^T W^T W U)^{-1} (U^T W^T W Y)$
4. $r_i(k) = (Y - UV(k))_i \cdot W_i$
5. $W_{jj}(k) = g(r_j(k))$ for the index j for which $r_j(k) = \max_i(r_i(k))$
6. $V(k+1) = (U^T W^T W U)^{-1} (U^T W^T W Y)$
7. $k = k + 1$
8. if $(R^2 > \zeta$ or $k > m)$ then stop otherwise go to step 4.

The parameter m is a maximum number of iterations, which is set to avoid oscillations. It should be noticed that the fitting convergence of this algorithm is not mandatory: if the iterated fitting is not stable it is still a good way to determine that the pixels in sequence Y do not represent a valid edge and to reject the candidate. Due to the iterative process, the tolerance ζ can be set higher than for the LMS method, as $\zeta = 0.99$ for example.

The weight functions $g(x)$ is either defined as inversely proportional to the residual $r_i(k)$ of each pixel, or as a non-linear function which removes the points having the higher residual $r_i(k)$ at each step:

$$g(x) = \begin{cases} 0 & \text{if } x = \max(r_i(k)) \\ 1 & \text{otherwise} \end{cases} \quad (5.13)$$

As most of the image local degradations produce replacement noise in the acquired images (see Section 4.3), the non-linear function presented in Equation 5.13 is more suitable than a weighting function inversely proportional to the residual of each pixel.

The weighted LMS edge detection is an interesting method to remove single spurious points (outliers). In our case, we have a blurred image, which means that when degradation occurs, several points are affected. Since the linear part of the slope is built of a few points, if we remove all the spurious points, we may get an insufficient number of points to get an accurate fitting.

5.2.5 Edge detection error

A few errors may occur during the edge localization, producing edge detection errors, which shift a detected edge e_i by Δx_i on the radial direction. If we assume the linear edge model of Subsection 4.6, and if the linear part of the edge is large enough to encompass the real and the detected edge position, the edge detection error is then also linearly related to the model parameters. The edge detection error can be formulated as the radial distance between the true edge position (\hat{x}, \hat{y}) , and the detected edge position (\bar{x}, \bar{y}) . For the adaptive threshold we can define \bar{y} as:

$$\bar{y} = \tau \cdot (A - B) + B \quad (5.14)$$

using a threshold τ , a step amplitude A and base level B . The edge detection error Δx_i of an edge i is defined using the slope p_i of the linear part of the transition:

$$\Delta x_i = \frac{\Delta y_i}{p_i} \quad (5.15)$$

where:

$$\Delta y_i = \hat{y}_i - \tau \cdot (A - B) - B \quad (5.16)$$

Thus the sum of the errors of the four edges is given by:

$$\Delta x_{tot} = \sum_i \Delta x_i = \frac{\Delta y_1}{p_1} + \frac{\Delta y_2}{p_2} + \frac{\Delta y_3}{p_3} + \frac{\Delta y_4}{p_4} \quad (5.17)$$

In an ideal case, all step amplitudes A are equal, implying that all Δy are equal and all slopes p_i are equal with opposite signs. Thus we have a perfect symmetry between the raising and the falling edges two by two and the edge detection errors cancel each other:

$$\Delta x_{tot} = \frac{\Delta y_1}{p_1} - \frac{\Delta y_1}{p_1} + \frac{\Delta y_1}{p_1} - \frac{\Delta y_1}{p_1} = 0 \quad (5.18)$$

In Chapter 4, different factors have been studied, which may influence either the slope, the base level or the step amplitude. This has a direct impact on the error Δx :

- More optical blur at photography on either the top or bottom edges decreases the corresponding slopes.
- A narrow bottom width increases the base level for the bottom of the groove, when both bottom edges blur each other.
- The indirect light and non-uniform illumination produce different step amplitudes A_1 and A_2 on the two traces.

In case there is only one kind of degradation at a time, which still preserves the symmetry of each trace, then the errors cancel each other. If we have some indirect light problem for example, it means that there are two different amplitude steps A_1 and A_2 , then:

$$\begin{aligned} \Delta x_{tot} &= \frac{\hat{y} - \tau \cdot (A_1 - B_1) - B_1}{p_1} + \frac{\hat{y} - \tau \cdot (A_1 - B_1) - B_1}{-p_1} \\ &\quad + \frac{\hat{y} - \tau \cdot (A_2 - B_1) - B_1}{p_2} + \frac{\hat{y} - \tau \cdot (A_2 - B_1) - B_1}{-p_2} \quad (5.19) \\ &= 0 \end{aligned}$$

In case we have several kinds of degradation at the same time, the symmetry of the traces, and thus of the errors, is no more preserved. If, for example, we have two different amplitude steps and a narrow bottom width where both edges blur each

other, and assuming that the slopes p_1 and p_2 stay symmetric for each trace, then the errors do not cancel each other anymore:

$$\begin{aligned} \Delta x_{tot} &= \frac{\hat{y} - \tau \cdot (A_1 - B_1) - B_1}{p_1} + \frac{\hat{y} - \tau \cdot (A_1 - B_2) - B_2}{-p_1} \\ &\quad + \frac{\hat{y} - \tau \cdot (A_2 - B_2) - B_2}{p_2} + \frac{\hat{y} - \tau \cdot (A_2 - B_1) - B_1}{-p_2} \quad (5.20) \\ &= \frac{(1 - \tau) \cdot (B_2 - B_1) \cdot (p_2 - p_1)}{p_1 p_2} \end{aligned}$$

Thus the edge detection may be corrupted by an error in case of blur or inhomogeneous illumination. This error is then correlated to the image degradation variations affecting the base level B , the step amplitude A or the slope p of the transition. Fortunately, most of the blur and illumination variations are smooth, and vary at a very low frequency: proportional to the scanner rotation or to the radial position on the record. Therefore the error results in very low, non-audible audio frequencies (lower than 2 Hz); and even the first harmonics would be inaudible.

5.3 Groove reconstruction

Once the edge points candidates are detected by the fine edge detection (presented in Subsection 5.2.4), they still must be put into a row to reconstruct the traces over the ring. Then the different traces must be combined to rebuild the groove position within the ring. Before joining traces, we must first know whether this acquisition contains a single or a double trace groove and what is the reading way. The extraction of this information is described in Subsection 5.3.2 and 5.3.3. The groove reconstruction step finally outputs two edge sequences for single trace grooves and four edge sequences for double trace grooves.

5.3.1 Trace following

The trace following step uses the sequence of candidate edge point couples, which are output from the edge detection (see Subsection 5.2.4), to iteratively construct a set of traces. Each trace T_i is built of two sequences: one raising edge r_i and one falling edge sequence f_i , as shown on Figure 5.11. For negative images (acquired on negative films), the raising edge is always located on the outer side and the falling edge on the inner side.

The trace following algorithm works as follows, and steps 1 to 4 are further developed in subsections 5.3.1.1 to 5.3.1.4:

1. Initialize the traces T_i .
2. Define the estimated edge positions $\hat{r}_i(j)$ and $\hat{f}_i(j)$ for each trace i at time j .

3. Define a range m around each $\hat{r}_i(j)$ and $\hat{f}_i(j)$, which is supposed to include $r_i(j)$ respectively $f_i(j)$. These ranges m might have a limited overlapping.
4. Select the best edge point candidate c_k inside the corresponding range m for each edge sequence r_i and f_i .
5. Increment j and go to step 2.

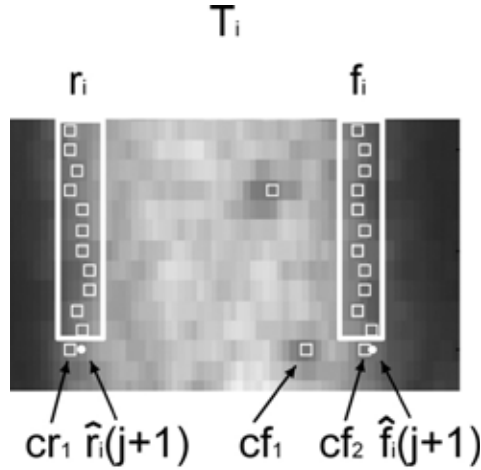


Figure 5.11: The trace T_i is built of a raising edge r_i and a falling edge f_i . The candidate selection is performed to minimize the distance between the candidates cr_k and the estimated positions \hat{r}_i , and the distance between the candidates cf_k and the estimated positions \hat{f}_i .

5.3.1.1 Initialization

The traces are initialized at the beginning of the acquired ring to define their initial locations. A sequence J is defined as the average of the n first and n last lines of the acquired ring, with n being usually between 5 and 30, depending on the record rotation speed (33 or 78 rpm), the image sampling frequency and the image degradation level. This sequence J is then processed by the edge processing method explained in Subsection 5.2: it generates a set of edge points couples, which define the estimated initial edges positions $\hat{r}_i(0)$ and $\hat{f}_i(0)$ of each trace i .

The initialization allows also estimating the trace width, by taking the median difference between the raising and falling edges of all the initially detected edge point couples. This estimated width w is then used in Equation 5.3 for the coarse edge detection step.

5.3.1.2 Estimated position

We use the estimated positions $\hat{r}_i(j)$ and $\hat{f}_i(j)$ to get an approximate position of each trace at each sampling time in any case, even when the image shows important degradations and when the trace is not visible or not correctly located on the image.

Since the groove radial displacement presents only smooth variations from a line to the next, the estimated edges positions $\hat{r}_i(j)$ and $\hat{f}_i(j)$ of each trace T_i at time j are defined by the extracted positions at the previous time instant $j - 1$:

$$\begin{aligned}\hat{r}_i(j) &= r_i(j - 1) \\ \hat{f}_i(j) &= f_i(j - 1)\end{aligned}\tag{5.21}$$

If an edge point could not be extracted at time $j - 1$ and is set to *undefined*. The estimated positions for *undefined* edges at time j are then defined as follows:

$$\begin{aligned}\hat{r}_i(j) &= \hat{r}_i(j - 1) + d \\ \hat{f}_i(j) &= \hat{f}_i(j - 1) + d\end{aligned}\tag{5.22}$$

using the previous estimated position at time $j - 1$ and a displacement d , which is defined either by the displacement of:

1. The other edge of the same trace, if it is not *undefined*.
2. The edges of the other trace of the same groove (for double trace groove), if it is not *undefined*.
3. The average of the displacement from all the other traces in the current image, otherwise.

The average of all the other traces displacements will result in an approximation of the low frequency components of the signal, which are produced by the groove spiral and the off-axis due to the bad centering of the film on the scanner. Thus it helps to approximately follow the trace position in case of trace cut or any other local image degradation.

5.3.1.3 Ranges definition

A range $\{m_{min,r,i}(j); m_{max,r,i}(j)\}$ and $\{m_{min,f,i}(j); m_{max,f,i}(j)\}$ is defined around each estimated edge position $\hat{r}_i(j)$ respectively $\hat{f}_i(j)$ for each trace i , as shown on Figure 5.12. This range is supposed to include $r_i(j)$, respectively $f_i(j)$, and is used to accept or reject the edge point candidates. These ranges are used to guarantee that the detected traces will not overlap each other. A limited overlapping is accepted between raising and falling edges, as these do not present the same characteristics. No overlapping is possible among raising edges or among falling edges.

At time j , the range is defined for each trace i using the last detected positions and some parameterized tolerance p :

$$m_{min,r,i}(j) = \hat{r}_i(j) - p \quad \text{and} \quad m_{max,r,i}(j) = \hat{r}_i(j) + p \tag{5.23}$$

$$m_{min,f,i}(j) = \hat{f}_i(j) - p \quad \text{and} \quad m_{max,f,i}(j) = \hat{f}_i(j) + p \tag{5.24}$$

The aim of this range is to separate the traces and avoid overlap in case of degradations and trace cuts; but it does not aim at removing small damages and noise. Thus the tolerance p shouldn't be too restrictive and must stay relatively

important compared to the edge displacement produced by the sound signal, as exposed in Equation 4.12. This way, it is also possible to work with the above defined estimated position, which presents only a limited accuracy. Usual values of p are then between 10 and 20 μm .

The purpose of these ranges is also a first rough denoising process. The ranges are then checked to ensure that they have at most a limited overlapping. In case of range overlapping, then the ranges are set to the neighboring estimated position:

$$\begin{aligned} \text{if } m_{\min,r,i+1}(j) < \hat{f}_i(j) & \text{ then } m_{\min,r,i+1}(j) = \hat{f}_i(j) \\ \text{if } m_{\max,f,i}(j) > \hat{r}_{i+1}(j) & \text{ then } m_{\max,f,i}(j) = \hat{r}_{i+1}(j) \end{aligned} \quad (5.25)$$

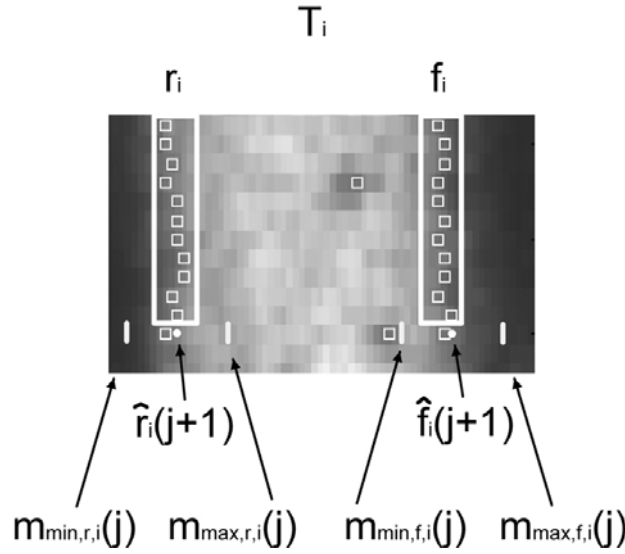


Figure 5.12: A range is defined around the estimated positions of the raising and falling edges of each trace.

5.3.1.4 Candidate selection

Each edge position $r_i(j)$ and $f_i(j)$ of each trace i , at time j are then defined by the most suitable candidate inside the corresponding range. Considering the case of edge r_i , there are three different possibilities at each time instant:

1. There is no edge point candidate inside the range, $r_i(j)$ is then set to *undefined*. The edge points marked as *undefined* will be handled later in Chapter 6.
2. There is a unique candidate cr_k , this candidate is then used to define the edge:

$$r_i(j) = cr_k \quad (5.26)$$

3. There are several candidates inside the range. As displayed on Figure 5.11, the edge position $r_i(j)$ is then selected among all these candidates cr_k to minimize:

$$|\hat{r}_i(j) - cr_k| \quad (5.27)$$

The candidate selection for $f_i(j)$ is strictly similar than for $r_i(j)$, the only difference is that we use the falling edge candidates cf_k instead of the raising cr_k : due to the blur level and the edge detection methods used, most of the time, candidates are unique inside a range. However, there may be some local image degradations, which produce another candidate inside the range. In such case, the mean width $\bar{w}_i(j)$ (which will be defined in Equation 5.30) of the trace i defined at time j can then be used to check the validity of the choice. This width check ensures that the trace will not follow a scratch or hair, or any other light grey pattern that may appear across the traces on the acquired image. Two conditions are estimated for this check:

$$|\hat{f}_i(j) - r_i(j) - \bar{w}_i(j)| > |\hat{f}_i(j) - cr_k - \bar{w}_i(j)| \quad (5.28)$$

$$|f_i(j) - \hat{r}_i(j) - \bar{w}_i(j)| > |cf_k - \hat{r}_i(j) - \bar{w}_i(j)| \quad (5.29)$$

If one of the rejected edge points candidates cr_k satisfies condition 5.28, then $r_i(j)$ must be set to *undefined* and a backtracking process must set some of the previous edges points of r_i to *undefined*.

If one of the rejected edge points candidates cf_k satisfies condition 5.29, then $f_i(j)$ must be set to *undefined* and a backtracking process must set some of the previous edges points of f_i to *undefined*.

If both conditions 5.28 and 5.29 are true at the same time, it means that it is difficult to take a decision at time j . Therefore the selected candidates are kept, and the correction will be performed on a next acquired line. Correction at time j will then be performed later by the backtracking process.

In case Equation 5.28 or 5.29 is true, the backtracking correction is then applied as follows, using the mean width $\bar{w}_i(j)$ of each trace i :

1. Set the value of the wrong edge sequence at time j to *undefined*
2. $j = j - 1$
3. Compute the width at time j : $w_i(j) = f_i(j) - r_i(j)$
4. If $(|\bar{w}_i(j) - w_i(j)| > \zeta)$ then go to step 1.

Since the width may suffer from important variations (cf. Subsection 5.1.1.1) ζ cannot be too straight: a value of ζ between 5% and 10% usually gives satisfying results. The mean width $\bar{w}_i(j)$ is updated at each time j to follow the smooth variations of the width over a circumvolution as follows:

$$\bar{w}_i(j) = \varsigma \cdot (\hat{f}_i(j) - \hat{r}_i(j)) + (1 - \varsigma) \cdot \bar{w}_i(j - 1) \quad (5.30)$$

with ς very close to 0, for example $\varsigma = \frac{1}{1000}$ or lower.

Since all the traces represent the same groove, the mean width, which is currently defined for each trace independently, could also be replaced by a unique mean width for single trace grooves and by two means for the double trace groove: one for the inner traces and one for the outer traces.

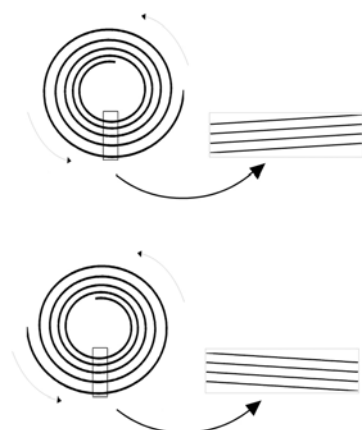


Figure 5.13: The orientation of the record spiral on the acquired image allows determining the reading way of the record.

5.3.2 Reading way

As presented in Section 2.9, discs are either recorded from outside to inside or from inside to outside. Since the turntables rotate clockwise for all the records, the spiral orientation depends on the reading way. The VisualAudio scanning process is the same for both kinds of discs: films are put with the sensitive layer on the top and rotate counterclockwise. Thus the traces on the rings have the same orientation than the original reading way: each trace is constantly shifted to the center for a disc recorded from outside to inside, and shifted to outside otherwise. To detect the reading way, we just need to get the sign of the average trace shifts after one circumvolution, as shown on Figure 5.13.

If, by mistake, the film is put back to front on the scanner glass plate, the traces orientation will change on the acquired images. Therefore the reading way will be interpreted as the reverse of the original one. This error can be manually corrected by reversing the sound file content using any sound editor. The reading way could also be automatically determined and corrected by a sound signal analysis to avoid error when the film is put back to front: musical instruments or speech consonants have some causal properties, as the attack or decay for example, which can be detected and used to determine the reading way of the extracted sound file [91, 92]. Such an automated correction processing is currently not implemented for the VisualAudio process as it is signal dependent and is also certainly more time consuming than the manual correction (to reverse the sound file) for what is, in fact, a minor and rare correction.

5.3.3 Single and double trace grooves

The single or double trace grooves can be distinguished by image processing. Once the trace reconstruction is finished, the easiest way to detect whether a groove is represented by one or two traces is to check the trace joint between the beginning

and end of a single ring acquisition. For single trace grooves, the end of each trace T_i matches the beginning of the following trace T_{i+1} , within a small range due to the groove displacement. If a groove is represented by two traces, the end of each trace T_i matches the beginning of the trace T_{i+2} , with the same range as before.

The distance between the end of trace and the beginning of the next one is also used as a check of the image processing quality, in order to detect whether the groove processing missed a trace or mixed several traces together.

For the case of double trace groove acquisition, we must still determine which traces must be linked together, or in other words: which traces represent the left respectively right walls of the groove. For a record which shows well distinguished traces, the easiest way is to number the traces, and to identify the odd traces as the left groove's wall and the even as the right wall. This solution is only partial and is not applicable for damaged records with large trace discontinuities, and for acquisitions which does not start at the beginning (and does not finish at the end) of the recorded area. Moreover, the lead-in, lead-out and the outermost and innermost trace often show some different visual aspect on the acquired image. Therefore this is not a reliable way to determine the trace linking over the whole record.

Another way to distinguish between the right and left walls traces is to use the correlation between the two traces of the same groove: we compute the distance between one trace and each of its two neighbors at many locations on the acquired image. The trace must then be linked to the neighbor, which distance is the more constant over the acquisition.

Unfortunately, the outer and inner circumvolutions of the groove are unmodulated. Thus it is sometimes difficult to determine the number of traces using this method at the beginning or end of a record. But contrary to the even/odd method, it is possible to use the correlation method on any traces at any location over the record, and to propagate then the information until the outer and inner traces. This guarantees a solution since normally there is at least one modulated trace somewhere on the record surface...

Some record acquisitions show a mix between single and double trace groove. In that case, it is possible to force the detection to work with only one trace, and to define a minimal trace width, which is then used to determine the box function size for the coarse edge detection step.

5.3.4 Trace joint

Once the number of traces per groove and the reading way are known, the groove can be reconstructed by putting the corresponding raising edge sequences r_i and falling edge sequences f_i of each trace i in a row, to form either two or four edge sequences, which define the complete groove displacement extracted within a ring. Prior to combining the edge sequences, several checks are performed, to remove some edges:

- The traces which go out of the ring, and do therefore not show a complete circumvolution on the current ring, are canceled: either they have been handled in the previous ring or will be handled in the next ring.

- double trace grooves are also checked to ensure that both traces are defined for the first and last groove circumvolutions. In case there is only one trace (with the other going out of the ring), the corresponding double trace groove circumvolution is canceled.
- First traces of the ring are compared to the position of the last trace of the previous ring to check if there is no overlap with the first traces of the current ring. The overlapping traces are then removed.

Thus, the combination of the edges returns two edge sequences for each ring j : the top outer $s_{TO,j}$ and the top inner edge $s_{TI,j}$:

$$\begin{aligned} s_{TO,j} &= [r_1^T \dots r_n^T]^T \\ s_{TI,j} &= [f_1^T \dots f_n^T]^T \end{aligned} \quad (5.31)$$

For double trace grooves, combination results in four edge sequences for each ring j : the top outer $s_{TO,j}$ and the bottom outer edge $s_{BO,j}$, the bottom inner $s_{BI,j}$ and the top inner edge $s_{TI,j}$:

$$\begin{aligned} s_{TO,j} &= [r_1^T, r_3^T, \dots, r_{n-1}^T]^T \\ s_{BO,j} &= [f_1^T, f_3^T, \dots, f_{n-1}^T]^T \\ s_{BI,j} &= [r_2^T, r_4^T, \dots, r_n^T]^T \\ s_{TI,j} &= [f_2^T, f_4^T, \dots, f_n^T]^T \end{aligned} \quad (5.32)$$

Each ring extraction then outputs two or four edges sequences and a shift value k_i corresponding to the radial displacement of the camera for the next ring acquisition, measured in pixels. This shift can be chosen in two different ways, corresponding to two different scanning modes:

- When scanning is performed independently from the image processing, the radial shift k is defined before to start and is a constant value over the whole record.
- When scanning and image processing are performed for each ring prior to any radial displacement, the shift k_j is defined as the minimum extracted position of the last extracted trace on the current ring j . This shift k_j takes then a different value for each ring j . This mode minimizes the number of rings acquisitions for a complete record scanning.

5.3.5 Bypassing the ring structure

The image processing is working on each ring independently, as the acquired images are provided ring by ring by the scanner. However, one may want to bypass this structure to work directly with a unique image that is built by the composition of

all the rings. This would present the advantage to avoid overlap between rings at scanning, and thus to acquire around 20% to 30% less ring images.

If we want to bypass the ring structure, and to consider all the rings as a unique picture, we need to satisfy the following constraints:

- Joints between rings must be perfect or corrected: geometrical deformations could lead to bad joints producing many clicks on the extracted sound. Otherwise, instead of one jump at each ring junction (which is easy to correct, as we know when the end of the ring is reached), we will get one jump every time the groove is crossing from a ring acquisition to the other.
- While a bad calibration of the radial displacement will produce a unique click when working with ring structure, it may produce multiple clicks if we consider all the rings as a unique image.
- Film illumination at scanning must not present high difference between both sides of the ring; otherwise, it may lower the edge detection accuracy on the traces which crosses several rings. The correction of such illumination inhomogeneity will be discussed in Section 6.1.1.

Since a record image is covered by a large amount of rings, which already represent large images, it is natural to split the whole image for image processing in order to work with a reasonable amount of data at a time. This split is well done by the ring structure, and we don't think that introducing another images composition or split will lead to significant improvements in terms of sound quality or efficiency.

5.4 Extraction of an entire record

The last step for the image processing is to rebuild the complete edge positions over the whole record. The edge sequences $s_{i,j}$ and shift values k_i (as defined in Subsection 5.3.4) are combined to rebuild either two or four sequences defining the edge positions on all the rings. Starting from the first acquisition, the radial shifts k from the preceding rings are added to each element of the current edge sequence to output the two or four edge sequences S :

$$\begin{aligned}
 S_{TO} &= [s_{TO,1}, s_{TO,2} + k_1, \dots, s_{TO,n} + \sum_{i=1}^{n-1} k_i] \\
 S_{BO} &= [s_{BO,1}, s_{BO,2} + k_1, \dots, s_{BO,n} + \sum_{i=1}^{n-1} k_i] \\
 S_{BI} &= [s_{BI,1}, s_{BI,2} + k_1, \dots, s_{BI,n} + \sum_{i=1}^{n-1} k_i] \\
 S_{TI} &= [s_{TI,1}, s_{TI,2} + k_1, \dots, s_{TI,n} + \sum_{i=1}^{n-1} k_i]
 \end{aligned} \tag{5.33}$$

The groove extraction is performed on all the acquired rings successively to get the complete sound content. However, if for any reason the scanned area exceeds

the recorded surface, the extraction stops at the first ring which doesn't contain any groove circumvolution. If a record contains several tracks, which grooves are physically distinct and where the land between the tracks is larger than the ring width, the VisualAudio processing must be restarted for each independent track, exactly the same as with a turntable.

It should be noticed that, at scanning, the camera and lightening system are limited in their radial displacement by the rotation axis. Thus it is not possible to scan the inner part of records at radius lower than 3.8 cm. This could be another reason for the extraction process to end; but this is not a limiting factor in practice, as the standard minimum inner record radius is more than 4.5 cm (Table 2.4), and that we were able to acquire the entire recorded surface of all the records we got until now, even for old non-standardized records.

5.5 Conclusion

The groove extraction algorithm processes each ring independently to extract the position of the groove at each sampling time. These positions are then combined to reconstruct the groove over each ring and then over all the rings to extract the groove position over the whole record.

Coarse edge detection is first applied to detect the presence of all the edges over the image, even in case of damages or non-homogeneous groove's wall illumination. Finer edge detection then locates accurately the edge candidates to subpixel precision. The groove reconstruction step selects the candidates that best match the existing traces.

This algorithm assumes that the image has low degradation level. Special cases and processing of more degraded images will be studied in the following Chapter 6. The performance of the groove extraction method presented in this chapter will be discussed in Chapter 8.

Chapter 6

Signal restoration

The specific degradations affecting the VisualAudio Acquisition rings were explained in details in Chapter 4. These degradations can be corrected by specific procedures. This chapter presents several possible corrections as well as the choices that have been made for the signal restoration.

Section 6.1 first focuses on the correction for various acquisition artifacts: inhomogeneous illumination, sensor inhomogeneities and pitch variations due to the off-axis centering of the film. Deblurring methods are discussed in Section 6.2 and Section 6.3 presents procedures to correct local image degradations. Discs with shrinkage of the recording layer (see Subsection 4.3.2) are currently not fully handled, however Section 6.4 proposes some ideas for the restoration of such discs. Finally, Section 6.5 presents concluding remarks about the signal restoration part of the VisualAudio system.

Set apart the off-axis correction which is fully discussed in Subsection 6.1.2, all the methods presented in this chapter will be evaluated in the last chapter, which introduces some metrics for the evaluation.

6.1 Acquisition artifacts

This section focuses on the corrections of the artifacts that affect the image during the acquisition. Subsection 6.1.1 proposes a correction procedure for the non-uniform illumination and CCD camera pattern noise (see Subsection 4.3.4). Subsection 6.1.2 presents a solution to correct the pitch variations produced by the film off-axis presented in Section 4.5.3.

However, the best way to correct this kind of image degradation is to avoid them, by performing mechanical corrections and defining precise acquisition procedures, free of artifacts.

6.1.1 Camera calibration

Since the area acquired by the CCD camera is wide compared to the light source size, it is difficult to guarantee a uniform illumination of the scanned film. Therefore the center of the acquired image has a stronger illumination and appears lighter than

the borders. The CCD camera also produces fixed pattern noise (FPN), which is the variation in output pixel values that arises from inhomogeneities of the camera sensors. Computational calibration or appropriate electronics can effectively correct the inhomogeneous illumination and remove FPN; but correction by calibration is possible only for monochromatic images, or images which spectral composition doesn't vary with time [93, 94, 95]. Since the VisualAudio scanner works with monochromatic light, a calibration of the image is a useful step to enhance the sound extraction quality. We assume that the image signal $f(m)$ is the product of a transmitted signal $a(m)$ with a space-dependent illumination function $b(m)$ and a space dependent sensor sensitivity $e(m)$:

$$f(m) = a(m) \cdot b(m) \cdot e(m) \quad (6.1)$$

The signal of interest is $a(m)$, and we need to eliminate the effect of $b(m)$ and $e(m)$ as much as possible. To enhance the quality of the image we must calibrate the output of the camera to correct the pattern noise. This can be done by multiplying the output image with some weighting factors. This process is performed in two steps:

1. Calibration initialization: an acquisition is performed without film, with unsaturated light. Using the output s_i from each sensor i and the maximum value of all the sensor output max_s , a weight factor w_i is then defined for each sensor i :

$$w_i = \frac{max_s}{s_i} \quad (6.2)$$

2. Calibration of the output image: the weighting factors w_i are then used to multiply the output of each sensor during the acquisition of the record image.

Figure 6.1 shows part of an image acquired with a non-homogeneous illumination, and Figure 6.2 shows the same image after the calibration process. This correction improves the edge detection in the case of non-uniform illumination, although the bit-depth is just stretched and not enhanced.

The presence of dust on camera sensors will produce darker lines along the tangential direction (as visible on Figure 6.1). This cannot be canceled with a simple multiplicative calibration process and a much more complicated model is needed. A simpler solution consists in cleaning the sensors when such spurious dust appears. Therefore a check is performed on the weighting factors w_i to detect the spurious w_i , which have a too high amplitude or a too large difference with their neighbors. The check function $sp(w_i)$ is defined in Equation 6.3. The weighting factors w_i for which $sp(w_i) = 1$ indicate either the presence of a dust on the sensor i or a deficient sensor. A warning is then generated to ask the operator to clean and verify the CCD sensor i .

$$sp(w_i) = \begin{cases} 1 & \text{if } (w_i > t_1) \text{ or } (w_i > t_2 \cdot w_{i-2} \text{ and } w_i > t_2 \cdot w_{i+2}) \\ 0 & \text{otherwise} \end{cases} \quad (6.3)$$

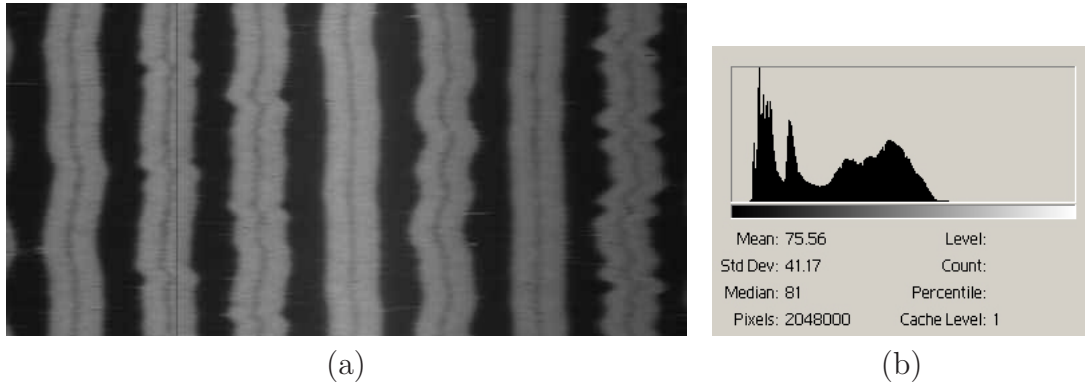


Figure 6.1: (a) shows part of an acquisition of a 78 rpm record: the illumination is not homogeneous and the image histogram (b) shows various spread peaks.

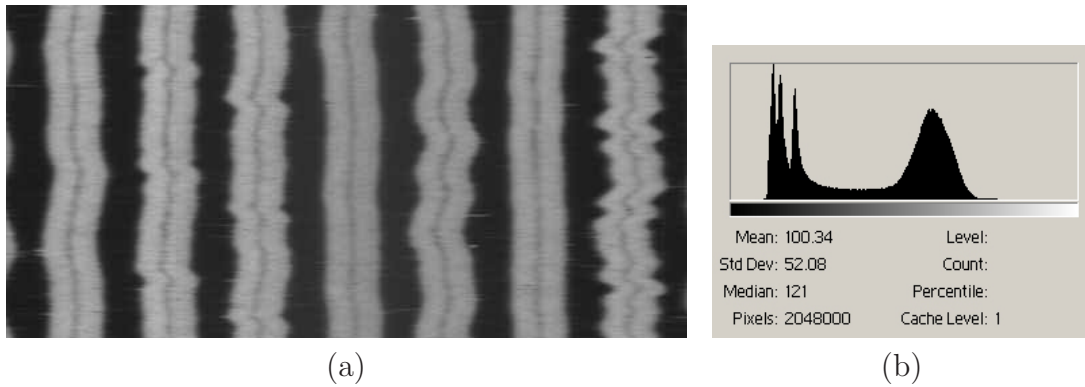


Figure 6.2: The figure (a) displays the same image area than Figure 6.1 after the calibration process. The histogram (b) of this calibrated image shows narrower peaks.

where the thresholds have been empirically fixed to:

$$t_1 = 2.5 \quad \text{and} \quad t_2 = 1.1 \quad (6.4)$$

The camera calibration process could still be enhanced to separate the image variance due to the sensor noise from the variance due to the non-uniform illumination. Healey et al. proposed a multistep algorithm to distinguish between dark current noise, fixed pattern noise and spatial variation caused by non-uniform illumination [95]. Dark current noise is the dominant noise factor at low light intensity; thus by working with dark reference images it is possible to estimate the variation of the dark current noise. The effect of fixed pattern noise and spatially non-uniform illumination can be distinguished by considering different imaging configurations and moving the light source between the image acquisitions. These different variations can then be combined in a more complex imaging model, which distinguish between dark current noise, fixed pattern noise and spatial variation caused by non-uniform illumination.

We have not developed such an enhanced model in the VisualAudio system, since the CCD noise and non-uniform illumination are currently not dominant components of the extracted sound deteriorations, and produce negligible effects on the extracted sound.

6.1.2 Pitch variation and wow correction

As presented in Section 2.8 and 4.5.3, the extracted sound pitch variations may have been produced by several causes: a deformed or warped record, film off-axis during scanning, rotation speed variation at recording or at playback. Godsill and Rayner propose a general pitch variation restoration [13], which is shortly explained in Section 6.1.2.1. However, as the main cause of wow in the VisualAudio project remains the film off-axis, and since the off-axis wow has some specific properties, a specific correction procedure has been developed to correct this specific case of wow effect using the image information and is independent from the recorded sound.

6.1.2.1 Restoration pitch variation defects

Godsill and Rayner propose a method to restore pitch variations, which can be used to remove wow and flutter (which were defined in Subsection 2.8). They separate the original sound in many blocks. For each block of sound signal, they first identify significant tonal components, using a time-varying spectral estimator, as the discrete Fourier transform (DFT) for example. Data blocks must be short enough to have nearly constant frequency components within the block. The main peaks from the DFT are extracted to build frequency tracks: peaks which are higher than a chosen threshold are split into two categories: those which fit the peaks from the previous block, and then belong to the same frequency track, and those which do not match existing tracks and will begin a new track. Thus in case of pitch variation the frequencies of the tracks evolve with time and enable to estimate a pitch variation curve. Using this pitch variation curve, the restoration is performed by non-uniform resampling of the distorted signal. Results show that wow can be completely removed from sounds having significant tonal components; but as this procedure depends on the sound content, some difficulties are expected for flutter restoration, when the recorded sound has only few tonal components or contains some tremolo or vibrato for example [13].

6.1.2.2 Film off-axis correction

As presented in Section 4.5.3, the film off-axis produces low frequencies on the extracted sound, as well as sound modulation. The low frequencies are not audible, but the wow effect produced by the sound modulation can be very disturbing for significant amplitude off-axis. Thus a correction is proposed, based on the information extracted from the image.

The frequency modulation produced by the film off-axis has the following interesting properties:

- It is driven by the low frequencies of the radial displacement of the film, relatively to the camera.
- The period of the radial displacement of the film corresponds to one circumvolution of the groove.
- The radial displacement of the film is independent from the radial position

The undistorted signal S can be defined by an uneven resampling of the extracted warped discrete signal S_w , using a time warping function f_w :

$$S(i) = S_w(f_w(i)) \quad (6.5)$$

Since the modulation is driven by the low frequency moves of the film, it is proportional to the low frequency components of the extracted sound. Thus the basic idea of the film off-axis correction is to determine a signal approximation containing only the low frequency component (LFC) of the warped signal S_w over an entire number of wow periods. S_w is then resampled using a time warping function f_w , which is defined using the LFC. For an off-axis driven pitch variation, the LFC is constant over the whole record, but f_w will vary according to the radial position of the current circumvolution, as stated in Equations 4.53 to 4.55. In a first step, f_w is built considering a constant radial position r for the whole groove circumvolution, and considering the radial motions due to the sound signal and to the spiral as negligible. Thus we can apply the following correction scheme, which is applied for each groove circumvolution:

1. Use the DFT (Discrete Fourier Transform) $F(k)$ to get the coefficients a_1 and b_1 of the Fourier series on N samples, corresponding to an entire number of periods of the wow degradation (and therefore to an entire number of groove circumvolutions):

$$F(k) = a_0 + \sum_{n=1}^{N-1} a_n \cdot \cos(n \cdot \omega \cdot k) + b_n \cdot \sin(n \cdot \omega \cdot k) \quad (6.6)$$

2. Build the low frequency component function LFC using the first coefficients a_1 and b_1 of the FFT, which correspond to the once per circumvolution period:

$$LFC(k) = a_1 \cdot \cos(\omega \cdot k) + b_1 \cdot \sin(\omega \cdot k) \quad (6.7)$$

3. Determine the index k_{max} of the maximum value of the LFC. At index k_{max} the angular positions $\theta_{k_{max}}$ and $\theta'_{k_{max}}$, corresponding to the ideal and the real sampling angular position, are equal (cf. Equation 4.55). The minimum of the LFC could also be used, as the centers of the circles C_1 and C_2 and the sample point p_k (as shown on Figure 4.22) are in line at the indices of both extremum of the LFC.

4. Based on Equation 4.55, define the resampling function f_w using the amplitude d as defined in Equation 4.53:

$$f_w(k) = k - \arcsin\left(d \cdot \sin\left(\frac{(k - k_{max}) \cdot 2\pi}{N}\right)\right) \quad (6.8)$$

with $d = \frac{2 \cdot \sqrt{a_1^2 + b_1^2}}{r}$

using the first coefficients a_1 and b_1 of the FFT, the number of samples N and the local radial position r on the record image.

5. f_w returns in fact a real value, which cannot be directly used as an index for the discrete sampled signal S_w . Thus an interpolation must be performed on the warped signal S_w to get the resampled signal at the integer index i between the closest $f_w(j)$ values: a linear interpolation between the two closest corrected indices $f_w(j)$ and $f_w(j + 1)$ is considered as sufficient, as the LFC is highly oversampled: it contains one period of a sinusoid sampled by at least 65 k-samples. Thus the resampled signal $S_r(i)$ is defined as:

$$S_r(i) = \frac{f_w(j + 1) - i}{f_w(j + 1) - f_w(j)} \cdot S_w(j) + \frac{i - f_w(j)}{f_w(j + 1) - f_w(j)} \cdot S_w(j + 1) \quad (6.9)$$

for integer indices j which satisfy:

$$f_w(j) < i < f_w(j + 1) \quad (6.10)$$

If the original record is warped or deformed, the low frequencies components of the LFC are not constant over the whole record; but this wow restoration procedure can still be used, while the LFC is determined at the inner circumvolutions of the groove, where the record deformations are negligible. For crackled records, the LFC can be defined on the groove circumvolutions that are not damaged and do not present discontinuities. If no undamaged groove circumvolution exists, the LFC can still be determined by dividing the record surface into distinct blocks corresponding to the same angular distance, which all encompass an undamaged section of the groove (at any radial position). The LFC is then approximated using the signal reconstituted over the consecutive blocks.

Figure 6.3 shows two spectra of a 300 Hz track acquired with a large film off-axis: first without any pitch variation correction, and then with the pitch variation correction presented in this section. The main peak width is considerably reduced to the same level as an ideal mechanical centering. This method works well to remove the wow effect produced by the film off-axis. However, if the wow effect is produced by several causes, the LFC must be built using more information than the extracted signal low frequencies, with a process similar to the method presented in Subsection 6.1.2.1.

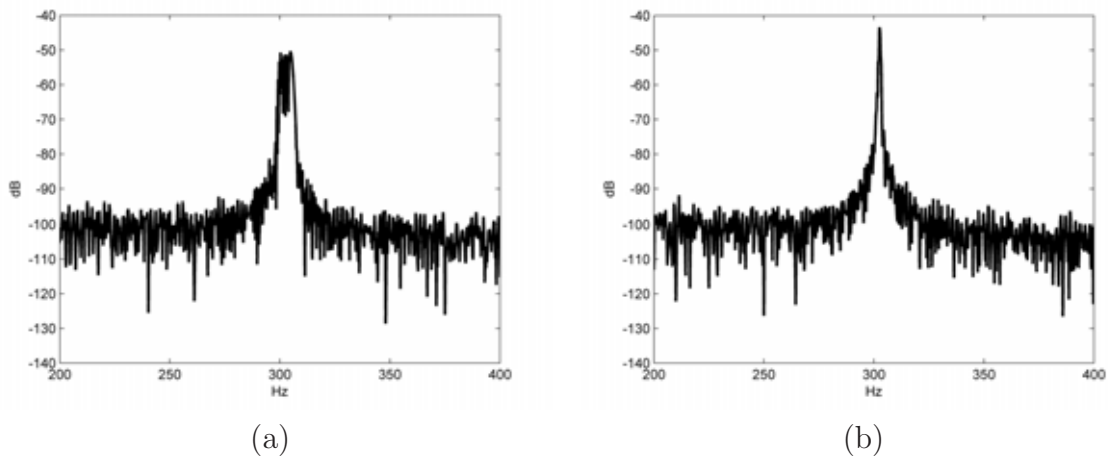


Figure 6.3: spectra of 300 Hz track acquisitions (a) without and (b) with pitch variation correction: the main peak width is considerably reduced by the correction.

6.1.2.3 Manual centering method

The initial centering of the film on the scanner glass plate is performed by a human operator. The acquired image is displayed live on a scrolling window. The operator must determine the position of either the minimum or maximum off-axis. Once the film reached this position, the operator stops the rotating scanner, and moves the film according to the estimated off-axis amplitude. Rough off-axis estimation is made at the first step; more precise off-axis estimations can be performed by looking at the groove displacement when the off-axis is smaller than the width of an acquired ring. Although this method has a limited precision, the resulting film off-axis after 2 or 3 centering steps can be less than $20 \mu\text{m}$.

6.1.2.4 Mechanical centering method

In the scope of the VisualAudio project, Simon Gremaud defined a mechanical correction for the VisualAudio film off-axis [96]. He first divides an acquired ring in a predefined number of radial blocks. Sobel masks are applied in the two directions on each block, to determine the average groove orientation inside each block. The orientation variation curve is then built with the reverse Fourier transform of the first two FFT coefficients of the average orientation curve. Once the position and amplitude are measured on the LFC, the mechanical correction is performed in four steps. The whole mechanical centering method works then as follows:

1. The scanner acquires the ring image.
2. The acquired ring is divided into block of constant size.
3. Sobel masks are applied in both direction on each block to determine the average orientation of each block.

4. The orientation variation curve is then built with the reverse Fourier transform of the first two FFT coefficients of the average orientation curve.
5. The maximum amplitude a of the orientation variation curve is measured, as well as the angular position p where it occurs.
6. The scanner is stopped at the given position p .
7. The operator puts a mechanical finger on the film.
8. The mechanical finger is pulled (or pushed) by the radial motor to perform a radial displacement a to center the film.
9. The mechanical finger is raised by the operator.
10. The whole process is repeated until the centering is satisfactory, i.e. the measured amplitude a is lower than a given threshold.

The algorithm correctly detects the values a and p . Unfortunately, this method suffers from a lack of precision due to the mechanical process: the film displacements are not sufficiently homogeneous, which is mainly due to the liquid used to stick the film by capillarity. Moreover, this correction requires several ring acquisitions and processing, and currently lasts several minutes; therefore the manual centering process is currently faster and more efficient than Gremaud's mechanical centering method.

6.2 Deblurring

In many image processing applications, an observed image $g(x, y)$ is modeled as the two-dimensional convolution of the true image $f(x, y)$ with a linear spatially-invariant blur $h(x, y)$, which is also known as the point-spread function (PSF):

$$g(x, y) = f(x, y) \otimes h(x, y) \quad (6.11)$$

The problem of recovering the original image $f(x, y)$ from Equation 6.11 requires the deconvolution of the PSF from the degraded image $g(x, y)$, which can be performed by a variety of well-known techniques such as inverse filtering, Wiener filtering or constrained iterative deconvolution methods [97, 98].

In most algorithms, the PSF $h(x, y)$ is supposed to be known, prior to the deconvolution procedure. Unfortunately, this is usually not the case in practice, and the true image $f(x, y)$ must be estimated by using only partial information about the blurring process. Such estimation process is called blind deconvolution [99].

Here are some important considerations about the VisualAudio images, regarding blur and deconvolution:

- The deconvolution methods are, in general, an ill-conditioned problem: a small perturbation of the input data produces large deviation in the estimated solution. Since there are several noise sources in the VisualAudio imaging chain,

and as we need to extract an accurate groove position, we don't want to introduce an additional uncertainty to the localization process.

- The PSF of the VisualAudio imaging chain is not linear, as it is a combination of optical blurs and of the non-linear response of the photographic film.
- Due to the groove depth, the out-of-focus blur is not equal at the top and bottom edges of the groove. Moreover, as explained in Chapter 4, the VisualAudio imaging chain contains mainly two blurring steps: the photography and the scanning. The patterns that appear on the record images (record groove's walls, scratches, dust...) are blurred at both steps; but there are also some patterns that appear only at scanning (film grain, pepper fog spots, dust, fiber pinholes), which are blurred only by the scanning optical blur. Therefore the PSF of the acquired images is not spatially invariant.

Based on the above considerations, we chose not to apply any deblurring method on the VisualAudio images: deblurring is not mandatory for accurate edge detection, as long as the algorithms are adapted to blurred edges. Moreover, the blur can be useful to detect spurious pixels: the pixels showing a too large contrast with their neighbors do presumably not belong to the groove edge, but to a spurious pattern which has been produced only at the scanning step and was then not blurred by the photographic lens. This property will be used in the corrupted pixel map system described in Section 6.3.3.

6.3 Local damage correction

As presented in Chapter 4, many image degradations could occur during the whole image acquisition process, which may produce impulsive noise on the extracted sound. For more efficient edge detection and groove extraction processes, it is desirable to have as little degradation as possible on the processed image.

Usual signal restoration procedures are built of two steps: detection and correction. In VisualAudio, there are two kinds of signals: the 2D image and the 1D extracted audio signals. Both signals have their own specificities and properties. Therefore the detection/correction steps can be either processed on the image, or on the sound, or on both, or split between the two levels, depending on the degradation type. In the scope of this work, this chapter focuses on the image processing procedures.

Thus the local damages detection and correction achieve two goals in the VisualAudio process:

1. By localizing large damages, we avoid some extraction errors. For example, as scratches and traces may look similar on the digitized image, when a scratch crosses some traces the edge detection process may be misled and follow the scratch instead of the trace. By locating the scratches on the image, it is then possible to reject the scratch edges points during the trace extraction process, and thus to avoid following the scratches instead of the grooves.

2. Some damages are clearly visible on the digitized image, but produce sound degradations which may be harder to detect on the sound signal. By using the contextual information on the image, such degradation can be localized prior to the sound extraction.

This section describes the methods we used to detect and correct the local image damages in the VisualAudio system. A short review of existing methods is first presented in Subsection 6.3.1, followed by the description of the chosen method in Subsections 6.3.2 to 6.3.6. Subsection 6.3.7 then presents the signal reconstruction methods.

6.3.1 Existing methods

6.3.1.1 Image smoothing

Mean filters, applied either with a square mask or only in the tangential direction, could be used to lower the effect of impulsive noise on the extracted sound; but it would also low-pass the whole extracted sound, which is not an interesting solution.

A combination of erosion and dilation could be used to remove spurious pixels. A large structuring element would be necessary to remove large defects such as dusts or scratches, resulting in sound deterioration. Moreover erosion/dilation is a blind process, in the sense that they are applied on the whole image and not only where necessary, thus it may degrade some correct image areas, and thus lower the quality of the extracted sound.

Median and median-like filters (Rank Order Mean [100], Peak-and-valley [101] or pseudomedian [102] for example) are well adapted to remove isolated spurious pixels in homogeneous region. Due to the blur, most damaged areas in the VisualAudio images are spread over several pixels. Therefore median-like filters would need large masks to handle such damages, which would lead to distortions and low-pass filtering of the extracted sound content.

6.3.1.2 Image reconstruction

Various image reconstruction methods are described in the literature, using frequency and spatial domain information [103, 104, 105, 106]. These methods focus on the reconstruction of completely lost blocks of image. Most of these image reconstruction techniques produce good visual corrections and perform good reconstruction of smoothed and textured areas, but are not adapted to reconstruct blurred edges at a subpixel accuracy. Moreover there is no description of the lost blocks detection: either the detection is supposed to be done earlier, or it is performed by user's interaction.

6.3.1.3 Frequency domain correction

A zero-mean stationary signal can be modeled by an autoregressive (AR) process. The signal x is then expressed by a finite linear combination of past values of the process and a white noise input e_n :

$$x_n = \sum_{i=1}^p a_i x_{i-1} + e_n \quad (6.12)$$

The AR model is commonly used for click detection and audio signal interpolation. A common interpolation implementation is the LSAR (Least Squares Autoregressive), which minimizes the predicted error over a given block of samples [13]. The AR+basis methods modify the AR-based interpolator to include some deterministic basis functions, which can be either sinusoids or wavelets [13, 107]. These methods are well adapted to signals having a non-zero mean or smooth underlying trends. This can be an interesting solution for the VisualAudio signal containing some low frequencies components produced by the groove spiral and the film off-axis.

The AR based interpolations give good results to correct gaps up to a few hundreds samples at a sampling frequency of 44.1 kHz, but they perform poorly for longer gaps. Esquef proposed then a frequency-warped version of Burg's method. This method estimates two AR models: one for the segment that precedes the gap and one for the segment that follows the gap. The interpolation is then obtained by cross-fading the two extrapolated sequences. This method gives satisfactory results on gap larger than 2000 samples, which represents 45 ms at 44.1 kHz [108].

More details on the audio interpolation methods can be found in [109, 12, 13].

Such methods have not been implemented in the VisualAudio system up to now since:

1. we wanted to concentrate more on the image processing part of this project.
2. a variety of denoising tools already exist on the market ([15, 16]) and it is still possible to use them as a post-processing on the VisualAudio extracted sounds.

6.3.2 Image smoothing

As presented in Subsection 6.3.1.1, a usual way to lower the noise level in an image is to smooth the image prior to process it.

In the VisualAudio system, smoothing methods are not applied in both dimensions, but are applied only in the tangential direction, for the following reasons:

- There is more resolution in the radial direction and the degradations are larger than a pixel. Thus if a pixel is corrupted, its neighbors in the radial directions are probably also corrupted, but its tangential neighbors are less probably corrupted.
- If the image is damaged on the transition areas (as defined in Section 4.1), the transition will then be more accurately reconstructed using the information of the surrounding lines.
- Image smoothing will lower the image and sound resolution: an image smoothing in the radial direction will lower the edge detection accuracy and smoothing in the tangential direction will lower the audio frequency bandwidth of the extracted sound.

The image smoothing in the tangential direction can be used, assuming that the groove is almost perpendicular to the tangential direction and that its modulations are smooth. The following methods have been implemented in VisualAudio:

1. Tangential mean: the ring image is tangentially convolved with a vector $V = [1/n, 1/n, \dots, 1/n]^T$ of size n . Image processing is then applied on the image resulting from this convolution.
2. Non-linear smoothing based on the mean, to replace only the corrupted pixels: the image I is convolved with a vector $V = [1/n, 1/n, \dots, 1/n]^T$ of size n :

$$M = I \otimes V \quad (6.13)$$

We then define a threshold τ , which will be used to detect the pixels to replace. The following replacement procedure is then applied for each pixel $I(r, t)$:

$$\text{if } |M(r, t) - I(r, t)| > \tau \quad \text{then} \quad I(r, t) = M(r, t) \quad (6.14)$$

This non-linear smoothing method is not fully satisfying, as it affects many pixels located in the transition areas.

3. Non-linear smoothing based on the mean before and after the current pixel, to replace only the corrupted pixels: nonlinear tangential mean with different checks before and after: images I are convolved with a vector $V = [1/n, 1/n, \dots, 1/n]^T$ of size n , as presented in Equation 6.13.

We then define a threshold τ , which is used to check each pixel by comparing it with the mean value of the pixels located before and after. The following algorithm is thus applied for each pixel $I(r, t)$:

$$\begin{aligned} \text{if} \quad & (|M(r, t - n/2) - I(r, t)| > \tau \\ & \text{and} \quad |M(r, t + n/2) - I(x, y)| > \tau \\ & \text{and} \quad |M(r, t - n/2) - M(r, t + n/2)| < \tau) \\ \text{then} \quad & I(r, t) = (M(r, t - n/2) + M(r, t + n/2))/2 \end{aligned} \quad (6.15)$$

This non-linear smoothing based on the mean before and after is more edge preserving, as it takes into account the context before and after the current pixel.

The size n of the smoothing masks ranges between 3 and 150, depending on the record degradation level.

It should be noticed that the acquired images represent rings from the original record, and they are therefore circular in the tangential direction. Thus the convolution is applied circularly and is fully defined at the image borders.

Due to the high resolution of the system, the nature of the picture (digitized photography with film grain, dust...) and the high blur level, most of the groove image degradations are larger than the pixel size. Therefore if we want to apply

a smoothing to remove this kind of damages, we would need a large mask, which will also affect the quality of the audio extraction by producing a low pass filtering. Image smoothing is therefore not a satisfying solution to be applied as an image pre-processing step in the VisualAudio system. However, image smoothing is well adapted as a first processing for the 2-passes method used for bad quality records, which will be presented in Section 6.3.4.

6.3.3 Corrupted pixel map

The basic assumption of the corrupted pixel map is that the groove image degradations produce replacement noise: a dust, a scratch or pepper fog will replace the value of the corresponding pixels on the digital image without any correlation to the original value. Due to the blur, the surrounding pixels are also corrupted by additive noise: blur produces a smoothed area where pixel values are a combination of the real pixel values and of the corrupted area values. This means that some pixels of the image do not represent the light reflection level of the groove, but they are due to some other spurious patterns.

The corrupted map system is built of two steps: detection and correction. The detection step consists in locating these spurious patterns, and building a map of the corrupted pixels they produced on the digitized image (see Figure 6.4). Once the map is complete, we must try to correct these corrupted pixels or avoid to use them during the image processing step.

Therefore a pre-processing step will detect the spurious pixel areas on the image, and mark them as corrupted in a corrupted pixel map. During the edge detection step, the corrupted pixel map is used to validate or eliminate the detected edges: edge points which are marked as corrupted will be rejected. If the neighborhood pixels used for the edge detection are marked as corrupted, the corresponding edge point will also be rejected.

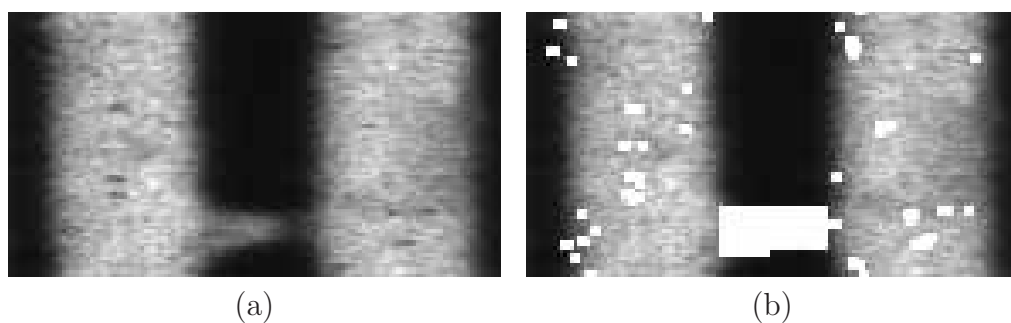


Figure 6.4: A sample of an original acquisition ring is displayed on (a). The detected corrupted pixels are marked in white in the image (b).

6.3.3.1 Corrupted pixels detection

Figure 5.3 shows the ability of the tangential gradient to detect corrupted pixels. This is mainly due to two reasons:

1. The traces are almost parallel to the tangential axis of the image, and their radial moves are very smooth.
2. Due to the non-isometric format of the pixel and the image, the spurious patterns (dust, fungus, pepper fog, scratches, hairs...) appear mainly as elongated spots in the radial direction on the acquired rings.

Therefore by applying a tangential gradient, we may detect these degradations. However, the process must avoid false detections. Thus the suspected corrupted pixels must also be strongly correlated in the radial direction to define a corrupted area and to avoid to detect traces edge as corrupted pixels.

Due to the nature of the noise (not related to the pixel size) and to the blur, the pixels in the neighborhood of the corrupted pixels must also be considered as corrupted. Thus once a pixel is detected as corrupted, the neighboring pixels are also marked as corrupted.

The images degradations occur in any of the three kinds of image features: traces, dark areas and transitions. It is relatively easy to detect damaged pixels in the traces and dark areas, since these patterns show homogeneous grey level; but unfortunately it is of little help, since these areas provide almost no information for the sound extraction process. On the other hand, the damaged pixels in the transitions are harder to detect, as the transitions present grey level variations as well as non-linearities. But as most of the damages are larger than the size of a pixel, it means that the probability is high that the damaged area is spread over radial neighboring pixels, which belong either to a trace or to a black area. Thus some of the damaged areas affecting the transitions can be detected in the neighboring homogeneous regions, and the corrupted areas can then be spread over near pixels.

Corrupted pixels are also blurred by the acquisition process, but they are not all affected by the same blur level, as some degradations appear before the photographic step on the record (dust or scratch for example) and some other appear only at the scanning step (dust or pepper fog).

Thus, since traces are tangential and damages appear as radial elongated objects, then small damaged areas could be located by a convolution with a tangential gradient mask. Due to the scanning blur (presented in Section 4.2.4), it is preferable to have a gradient mask larger than two pixels. On the other hand, using a too large gradient mask will increase the number of false detections. Thus it has been chosen to work with the following gradient masks M_1 and M_2 :

$$M_1 = [-1, -1, 2, 0, 0]^T \quad \text{and} \quad M_2 = [0, 0, 2, -1, -1]^T \quad (6.16)$$

Each tangential vector of the acquired image I is then convolved with the two gradient masks M_1 and M_2 :

$$G_1 = I \otimes M_1 \quad \text{and} \quad G_2 = I \otimes M_2 \quad (6.17)$$

Since the images represent rings, and therefore are circular in the tangential direction, the convolution is applied circularly and is fully defined at the image borders.

By using these two gradient approximations, we can separate two kinds of degradations:

- Small localized degradations, due to film grain or small dust and where $G_1(r, t)$ and $G_2(r, t)$ are of the same sign.
- Larger degradations due to scratches for example, where $G_1(r, t)$ and $G_2(r, t)$ are of opposite signs.

The difference for the processing is that a small degradation will require less radial correlation to be considered as corrupted than a large one, and that the larger damaged area must then be localized in an additional step. Thus the corrupted pixel map MAP is filled up first with the small degradations, and then with the large degradations. A temporary map $TEMPMAP$ is used to store the pixels state, according to the gradient mask convolution results G_1 and G_2 .

The small degradations are localized as follows (see Figure 6.5):

1. The maps $TEMPMAP$ and MAP are initialized with the value $CORRECT$
2. Detection of the possible corrupted pixels, by checking the results G_1 and G_2 of the convolutions with M_1 and M_2 (Equation 6.17):

for each pixel $I(r, t)$

if ($|G_1(r, t)| > \lambda_s$ and $|G_2(r, t)| > \lambda_s$)

if $sign(G_1(r, t)) = sign(G_2(r, t))$

$TEMPMAP(r, t) = SMALL_DEG$

3. Detection of the small corrupted pixels areas, by checking the radial correlation over r_s pixels. The corrupted area is then extended to the m_s neighbor pixels (in the radial direction).

for each pixel $I(r, t)$ where $TEMPMAP(r, t) = SMALL_DEG$

if ($TEMPMAP(r + i, t) = SMALL_DEG \forall i = 0, \dots, r_s$)

for $i = -m_s, \dots, +m_s$

$MAP(r + i, t) = CORRUPTED$

The large degradations are localized as follows:

1. The map $TEMPMAP$ is initialized with the value $CORRECT$ (MAP is not reinitialized, as it already contains the small degradations locations)
2. Detection of the possible corrupted pixels, by checking the results G_1 and G_2 of the convolutions with M_1 and M_2 (see Figure 6.6 (a)):

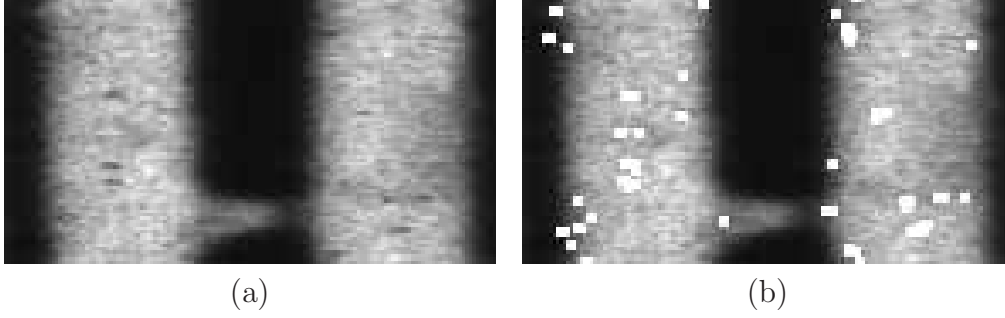


Figure 6.5: The original image is displayed on (a). The first step to build the corrupted pixel map is to localize the small degradations, which are displayed in white on (b).

for each pixel $I(r, t)$
 if $(|G_1(r, t)| > \lambda_l \text{ and } |G_2(r, t)| > \lambda_l)$
 if $sign(G_1(r, t)) \neq sign(G_2(r, t))$
 if $sign(G_1(r, t)) > 0$
 $TEMPMAP(r, t) = BLACKTOWHITE$
 else
 $TEMPMAP(r, t) = WHITETOBLACK$

3. Detection of the possible large corrupted pixels areas of interest, by checking the radial correlation over r_l pixels:

for each pixel $I(r, t)$ where $TEMPMAP(r, t) = BLACKTOWHITE$
 if $(TEMPMAP(r + i, t) \neq BLACKTOWHITE \text{ for any } i = 1, \dots, r_l)$
 $TEMPMAP(r, t) = CORRECT$
 for each pixel $I(r, t)$ where $TEMPMAP(r, t) = WHITETOBLACK$
 if $(TEMPMAP(r + i, t) \neq WHITETOBLACK \text{ for any } i = 1, \dots, r_l)$
 $TEMPMAP(r, t) = CORRECT$

4. Define the large damaged areas, by joining the *WHITETOBLACK* to the *BLACKTOWHITE* areas and by extending the area on m_l pixels radially and m_t pixels tangentially (see Figure 6.6 (b)):

for each pixel $I(r, t)$ where $TEMPMAP(r, t) = BLACKTOWHITE$
 if $(TEMPMAP(r, t + j) = WHITETOBLACK \text{ for any } j = -t_l, \dots, t_l)$
 $MAP(r + i, t + k) = CORRUPTED \forall i = -m_l, \dots, m_l$
 and $\forall k = -m_t, \dots, j + m_t$
 for each pixel $I(r, t)$ where $TEMPMAP(r, t) = WHITETOBLACK$
 if $(TEMPMAP(r, t + j) = BLACKTOWHITE \text{ for any } j = -t_l, \dots, t_l)$
 $MAP(r + i, t + k) = CORRUPTED \forall i = -m_l, \dots, m_l$
 and $\forall k = -m_t, \dots, j + m_t$

Usual values for the thresholds λ_s and λ_l , the margins m_s , m_l and m_t and the radial and tangential extensions r_s , r_l and t_l are currently empirically defined as follows:

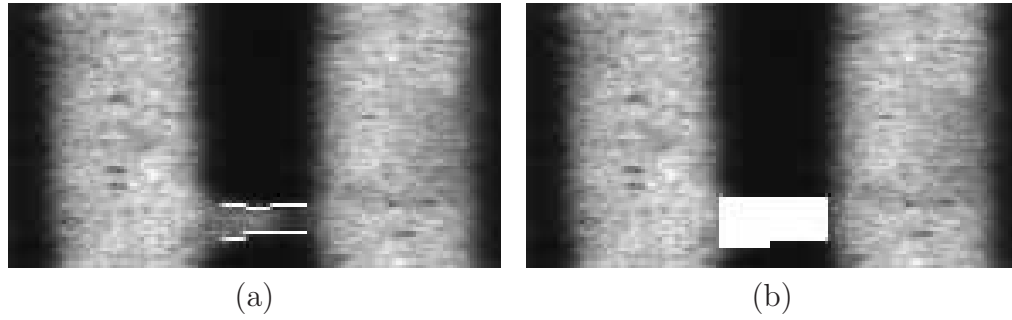


Figure 6.6: First, the bounds of the large degradations are detected and marked in white on (a). Then, the areas are extended to join the bounds and surrounding pixels (b).

- λ_s is dynamically defined for each pixel $I(r, t)$ as:

$$\lambda_s = 20 \cdot (2 + 4 \cdot (I(r, t - 2) + I(r, t - 1) + I(r, t + 1) + I(r, t + 2)))$$

- λ_l is dynamically defined for each pixel $I(r, t)$ as:

$$\lambda_l = 20 \cdot (2 + 4 \cdot (I(r, t - 2) + I(r, t - 1) + I(r, t + 1) + I(r, t + 2)))$$

- t_l : from 20 to 100, depending on the kind of degradations and on the sampling speed.
- $r_s = 2$
- $m_s = 2$
- $r_l = 5$
- $m_l = 10$
- $m_t = 2$

The thresholds λ_s and λ_l definitions contain a fix part (the term "2") and a dynamic part, which fluctuates according to the grey level. Thus, the degradation can be detected in both the light and the dark areas of the image.

Due to the radial margin neighborhood r_s and r_l , the corrupted pixels may not be detected at the radial borders of the ring. This is not important since the pixels closed to the border will not be used for the edge detection process.

It should be noticed that if the record picture is badly centered on the scanner, the traces will be less perpendicular to the sensors, resulting in a lower quality of the corrupted map corrections. If the record speed is low (for example 33 rpm) with high sound amplitudes, then this corrupted pixels detection scheme may not be well adapted.

6.3.3.2 Use of the corrupted pixel map

Once the corrupted pixel map is built, we must use the results to correct or enhance the extracted sound signal. There are two ways to use the corrupted pixel map:

- The map can be used to locate the corrupted pixels and to correct them on the acquired image, prior to the groove extraction process. However if one pixel is considered as corrupted in the transition, its close neighbors are also corrupted due to the high blur level. This means that there is a very low probability to still get reliable information in this transition to reconstruct the corrupted pixels correctly. By using the information of the tangential neighbors to correct small corrupted areas, the new transition pixel values will be an average of the surrounding transitions. Such a correction leads approximately to a linear interpolation of the extracted sound. Linear interpolation is not a very good correction for audio signal, however if this kind of correction is applied only on small gaps of oversampled signals it is satisfactory. The non-linear image smoothing methods explained in 6.3.2 may then be used to reconstruct the image using only the non-corrupted pixels.
- The map can be used in the groove extraction process, in order not to use the corrupted pixels to extract the edges. This will increase the number of undefined edge points during the extraction process, and the correction step will take place using only the edge points, which are then considered as correct.

6.3.3.3 Corrupted pixel map enhancement

The corrupted pixel map concept can also be enhanced: instead of working with tangential gradients, another approach could be to work with local gradients, which follow the direction of the groove. This enhanced corrupted pixel map system can be sketched as follows:

1. Apply a first edge detection on a smoothed image (similar to the first pass of the 2-pass algorithm presented in Subsection 6.3.4).
2. Reconstruct the groove.
3. Compute the local gradient for each pixel *in the direction that is tangential to the closer edge*.
4. Build the corrupted map based on this local gradient.
5. Extract the groove position by edge detection.

This way, the corrupted map system will be more adapted to detect the corrupted areas located in the transitions.

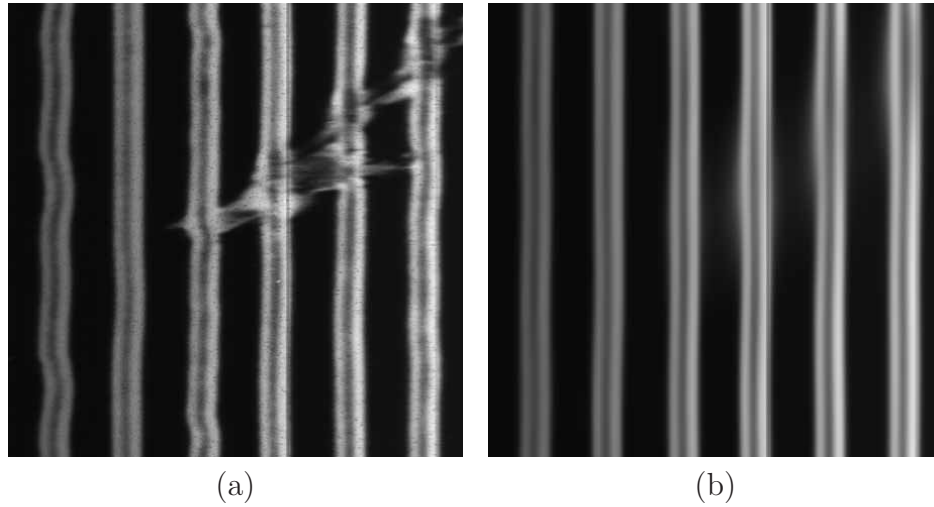


Figure 6.7: A sample of an original acquisition ring is displayed on (a). The degradation of the groove image is presumably due to a scratch on the record surface. (b) is the smooth version of the original image (a). The groove modulation and the image degradations are no more visible.

6.3.4 2-passes trace extraction

In case of large degraded areas with smooth edges, it is hard to localize accurately the damaged areas on the image with the corrupted pixel map. Any edge detected in these areas would be inaccurate. If the scratch width and radial displacement over time are similar to the ones of a trace, it may also produce false edge detections and lead to a trace cut or a merge with another trace.

A solution to separate the traces is to refer only to the low frequencies component of the groove displacement. These low frequencies are driven mainly by the spiral of the record and the off-axis of the film on the scanner. The proposed solution is then to apply a two passes trace extraction algorithm. The first pass edge detection is applied on images, which have been strongly averaged in the tangential direction, with the methods presented in Subsection 6.3.2. On these smoothed images, the traces look to be almost tangential with low frequencies modulations, as shown on Figure 6.7. The scratches and other large spots on the image have been smoothed and are no more visible. Thus they will no more mislead the trace following algorithm. The first pass trace extraction provides edges, which are driven only by the low frequencies due to the off-axis and the spiral. The true edges of the trace are located in a close neighborhood of the smoothed edge, at a distance bounded by a parametrized amplitude Γ , which is the maximum amplitude of the lower sound frequencies which have been removed on the smooth image.

The second pass edge detection is then applied on the original acquired image, but the ranges used for the trace following (Subsection 5.3.1) are then defined by the first pass trace extraction results: the smoothed edges ($\pm\Gamma$) build a corridor, and only the edges points detected inside this corridor will be considered for the second pass edge detection. Thus the trace extraction is non-causal and will be less sensitive to the local degradations.

The size of the mean filter is currently set manually. This size could be set automatically by an iterative process, which increases the mean filter size as long as the number of undefined edge points is too large, or as long as the traces do not join correctly between the beginning and end of the ring. Such an iterative process would increase the processing time (as the first pass is then repeated multiple times, with an image smoothing at each step); but it could be an interesting solution for records presenting a high degradation level.

Instead of working with a mean image, another possibility could be to work with an image that is downsampled in the tangential direction, for example by taking only one line out of five or ten. This would drastically lower the processing time of the first pass processing; but the ranges built this way would also be more sensitive to impulsive noise and to the higher frequencies groove modulations and thus lower the quality of the overall process. The results could even be very bad if the downsampling frequency used to build the reduced image is correlated to the sound frequency which modulated the groove. Therefore we kept the smoothing step to build the corridors at the first pass of the algorithm.

6.3.5 1D impulses detection

This detection scheme is based on the fact that the signal is oversampled and presents only smooth variations. In order to be independent of the sound signal, the corrupted samples detection is applied on the second derivative $\partial_\epsilon^2 S_i$ of the extracted signal S_i :

$$\partial_\epsilon^2 S_i(k) = |S_i(k - \epsilon) - 2 \cdot S_i(k) + S_i(k + \epsilon)| \quad (6.18)$$

where ϵ is used to check samples over a longer time interval, to be independent of the blur, which affects neighbor samples. The detection is then performed using a double-threshold method [110]. The high threshold τ_h is used to detect the impulses and the lower threshold τ_l to estimate the impulsion duration and to determine the samples to correct. Both τ_h and τ_l are defined using the standard deviation σ of $\partial_\epsilon^2 S_i$ defined on a local window, typically over 1000 samples:

$$\tau_h = a\sigma \quad \text{and} \quad \tau_l = b\sigma, \quad \text{with} \quad a > b > 1 \quad (6.19)$$

The factors are usually fixed as $a = 4$ and $b = 2$. The corrupted samples detection process is then applied as follows for the signal S_i :

1. $iter = 1$
2. Init the error map E_i for the signal S_i as false:

$$E_i(j) = false \quad \forall j \quad (6.20)$$

3. Detect the samples $S_i(j)$ with $\partial_\epsilon^2 S_i(j) > \tau_h \quad \forall \epsilon = 3, 6, 9, \dots, E$

4. Starting from $S_i(j)$, all the contiguous samples with $\partial_c^2 S_i(k) > \tau_l$ are marked as corrupted:

$$E_i(k) = true \quad (6.21)$$

5. $iter = iter + 1$
6. if $iter < max_{iter}$ go back to step 2 for the next iteration
7. The samples $S_i(j)$ are marked as *undefined* $\forall j$ where $E_i(j) = true$

The parameters max_{iter} and E are defined by the user and usually takes values ranging from 2 to 5 for max_{iter} and 9 or 12 for E .

6.3.6 LMS signal fitting

Since the signal is largely oversampled and should present only smooth variations over a small time interval, we can use a polynomial fitting to smooth the signal S_i and remove part of the impulsive noise.

Such a signal smoothing step has been implemented to lower the impulsive noise and respect the smooth variation of the groove edge.

1. An m order polynomial is fitted to the n points centered around the current point j .
2. All the points $S_i(j)$ which are too distant from the fitted value $\hat{S}_i(j)$ are removed from the fitting, using a threshold t :

$$|S_i(j) - \hat{S}_i(j)| < t \quad (6.22)$$

3. If too many points have been removed (more than $n/3$), $S_i(j)$ is marked as *undefined*.

This smoothing can be performed several times to get a smoother signal, with decreasing thresholds t . The order m of the polynomial shouldn't be larger than four to guarantee a fitting stability. The number of points n must also be set according to the polynomial order, being at least twice the value of m .

6.3.7 Signal reconstruction

Once they are detected, the undefined and corrupted samples must then be reconstructed with the available information. This correction can be based either on the spatial or on the frequency information.

The currently implemented methods are based on spatial information. Linear interpolation methods are presented in Subsections 6.3.7.1 and 6.3.7.2. These methods rely first on the fact that the images have a low degradation level, which produces only small gaps to correct. They rely also on the fact that the extracted signal is largely oversampled, and is thus almost linear when observed at small scales.

Subsection 6.3.7.3 presents another method, which considers a combination of the non-corrupted edges to correct the corrupted samples in a given time interval.

Such methods are not satisfactory for gaps longer than a few samples. But they were implemented in a first stage, in order to have some simple correction scheme available.

6.3.7.1 Linear interpolation

The first method is a linear interpolation, which can be used to correct very small gaps. Linear interpolation between the last valid sample $S_i(t_1)$ and the next valid sample $S_i(t_2)$ of the edge S_i , using the slope $l(S_i, t_1, t_2)$ between these points:

$$l(S_i, t_1, t_2) = \frac{S_i(t_2) - S_i(t_1)}{t_2 - t_1} \quad (6.23)$$

$$S_i(t + j) = S_i(t) + j \cdot l(S_i, t_1, t_2) \quad (6.24)$$

6.3.7.2 LMS interpolation

Each damaged area $[S_i(t_1), S_i(t_2)]$ is corrected by a low order interpolation, which is based on a least mean squares fitting of the neighboring samples $[S_i(t_1 - k), S_i(t_1)]$ and $[S_i(t_2), S_i(t_2 + k)]$. The number of samples on both sides of the gap must be sufficient to ensure correct curvature of the interpolation. The fitted value $\hat{S}_i(j)$ is then substituted to $S_i(j)$ if their difference is larger than a given threshold:

$$S_i(j) = \begin{cases} S_i(j) & \text{if } |S_i(j) - \hat{S}_i(j)| < t \\ \hat{S}_i(j) & \text{otherwise} \end{cases} \quad (6.25)$$

The second order interpolation approximates the curvature of the oversampled signal, and it can thus be used to correct slightly longer gaps. Higher order interpolations are not used as that may lead to unstable approximations and result in high noise level.

6.3.7.3 Edge copy

If the damage is correctly localized on an oversampled signal S_i , this signal can be corrected using the displacement of the other edges. For this correction, we then assume that there is no phase shift between the edges and that only one edge is damaged in a given time interval.

This correction assumes that all of the other edges do not contain any corrupted samples in the corresponding time interval between samples t_1 and t_2 . Thus this correction is applicable for small damaged areas affecting only one edge at a time. It is difficult to use this method for larger corrections, as large damages corrupt both edges of a trace, and even all the edges of a groove at the same time. Moreover, this correction does not take into account the symmetric edge detection error (see Subsection 5.2.5), which will then not be canceled by the combination of all the edges.

To correct n values of the edge S_i between time t_1 and t_2 , we use the values of the correlated edge S_j between the samples t_1 and t_2 . A linear term is added (as presented in Equation 6.23) in order to cancel the low frequencies components of the edge S_j :

$$S_i(t) = S_i(t-1) + \frac{\sum_{j \neq i}^m S_j(t) - S_j(t-1) - l(S_j, t_1, t_2)}{m-1} + l(S_i, t_1, t_2) \quad (6.26)$$

In order to be less sensitive to the edge detection error and to the scanning blur variations, we can use a copy of only one signal S_j to correct the corrupted signal S_i , with S_j and S_i being extracted from the same edge type: either raising or falling edge:

$$S_i(t) = S_i(t-1) + S_j(t) - S_j(t-1) + l(S_i, t_1, t_2) - l(S_j, t_1, t_2) \quad (6.27)$$

6.4 Shrinkage of the recording layer

In Subsection 4.3.2, we have presented the problem of the shrinkage of the recording layer for acetate records. The restoration of the discs with shrinkage of the recording layer is currently not fully implemented and is still an ongoing research topic.

6.4.1 Proposed correction

The skrinked records present many discontinuities, but contrary to the previously presented degradations, in the skrinked records image the cut traces may be radially shifted by several hundreds of micrometers (see Figure 6.8), which is much more than the space between the grooves, leading to ambiguities when we try to recompose the groove displacement over the whole record.

The current proposed restoration scheme consists in extracting the traces from the undamaged areas and reconstructing the signal like a puzzle using the traces fragments. Several kinds of information can be used to put the fragments together:

- Geometrical information: number the traces fragments from the inner and outer side of each trace, and recompose the puzzle by putting together the numbered traces from the various pieces of the puzzle: put together the fragments from the first trace from outside, from the second trace, and so on.
- Matching: define the distance between the fragments and try to find the matching combination which minimizes the total distance between the fragments.
- Frequency information: characterize the end of each fragment by an AR model, or by its frequency spectrum. Then we must find the best matching for the AR coefficients of each trace fragment over each crack.

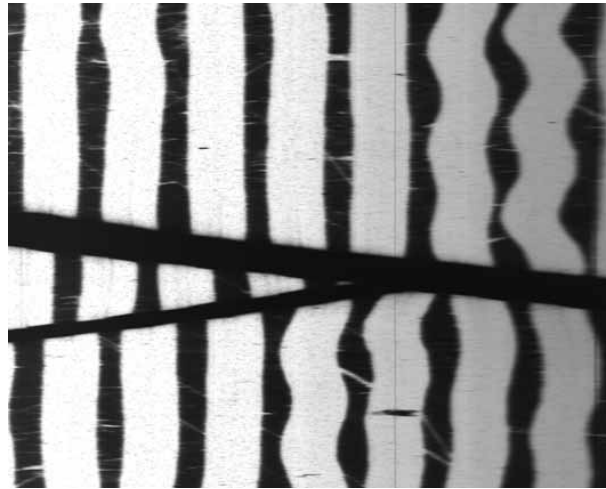


Figure 6.8: The shrunk records present shifts of several hundreds of micrometers over the crackles. The land (vertical dark areas) and the crackles (dark areas crossing the light traces) average grey levels are the same and it is thus difficult to accurately distinguish between them.

The trace fragments are built the same way as the traces, as explained in Section 5.3.1. Trace cuts are then defined when one of the following conditions is true:

- Both trace's edges are considered as *undefined* on a section longer than c_1 pixels.
- The edges from all the traces are considered as *undefined* on a section longer than c_2 pixels, with $c_2 < c_1$.

6.4.2 Lost points collection system

The lost points collection (LPC) handles all the detected edge point couples which were not used for the groove reconstruction step (see Section 5.3) and are therefore not part of an existing trace. The remaining edge point couples are put in the LPC with the same mechanism as the line processing (see Subsection 5.3.1): the edge point couples are matched to a trace candidate if it is close enough to this trace. If no trace matches this point couple, then a new trace candidate is created with this edge point couple. A trace candidate must contain at least t edge points to be considered as a trace fragment, with $t > 10$ to avoid false trace detection. A trace candidate that is not updated since x lines is considered as not being a trace and is then canceled. LPC works only with point couples and not with single edge points, to ensure that a local degradation will not be considered as a new trace.

6.4.3 Detect the cuts

Up to now, we considered two kinds of features on the acquired images: the traces, and the dark areas representing the land and the groove bottom. To handle shrunk records it is desirable to detect the cuts as a new feature on the acquired image. This

allows to easily identify the trace fragments. Unfortunately, this is currently not the case, and in many cases, the average grey levels representing cuts and dark areas are very close (less than 6 Digital Numbers or grey levels) and difficult to distinguish on noisy images as shown on Figure 6.8. Thus the cuts cannot be currently detected by using the grey level, but only using the knowledge of the trace properties, as mentioned in Subsection 5.1.1.

Several solutions exist to get increased contrast between the cuts and the dark areas. First, we can lower the exposure time at the picture taking step. This will result in a lower contrast between the traces and the land, and thus in a coarser graininess in the traces edges. The other solution is to use films with a lower gamma: this will ensure sufficient dynamic in the film to differentiate between the three kinds of features. Unfortunately, the lower gamma films are much grainier, resulting in a noisier extracted sound.

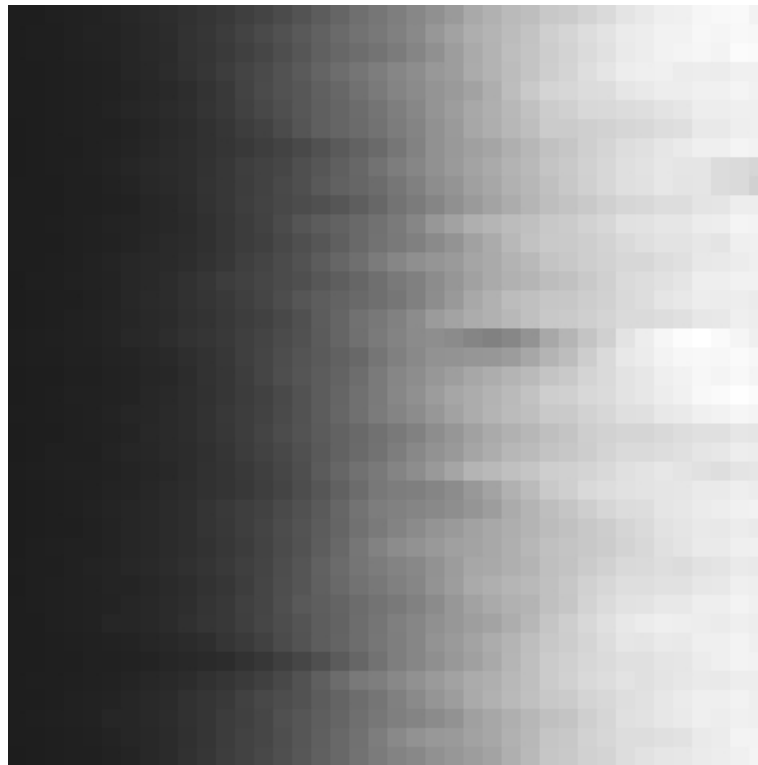


Figure 6.9: Magnified view of a 40×40 pixels ring image acquired on an unmodulated groove: the edge to extract should then be represented by a straight line.

6.5 Conclusion

This chapter described the restoration methods applied to correct the degradations affecting the signal in the VisualAudio system. In the VisualAudio project, we tried to get the best sound extraction using image processing techniques. We tried to

avoid the frequency domain sound corrections, which are not always well accepted by some sound archivists.

The off-axis correction based on the groove displacement on the acquired image (presented in Section 6.1.2.2) is only able to correct some of the pitch variation defects, and is then only a partial solution to the pitch variation problem. A frequency based correction would be more general and efficient.

The analysis of the local degradations and the various correction methods applied in the VisualAudio groove extraction process shows that it is difficult to accurately localize the image degradations. Some degradations are easily detectable in the homogeneous areas: dark spots in the traces or light spots in the dark areas; but these degradations do often not affect the sound signal, which is extracted from the transitions between the traces and dark areas. The degradations which are located on the transitions are much more difficult to locate and correct, as they mainly do not appear as a specific pattern that can be isolated, but as local cuts or spreads of the trace. Figure 6.9 shows a 40×40 pixels magnified view of a ring image acquired on an unmodulated groove and illustrates this problem. The edge to extract should then be a straight line, but the transition shows important uncertainties. These uncertainties are difficult to locate as specific patterns due to dusts, film grain or any other damage in the imaging chain. This shows the limitation of the image processing capabilities on the sound extraction process.

Therefore it is very difficult to use the knowledge of the image degradations to correct them. The more natural and efficient way to correct these degradations would then be to use the information of the groove modulation in a more accurate way. This shows the necessity of a frequency based model to perform detection and correction of the signal, using ARMA, LSAR or Warped linear prediction for example [13, 12, 111].

Chapter 7

Sound processing

Once the groove position has been extracted out of the image, it must still be transformed in a sound signal and output in a standard audio format. This is the purpose of the sound processing stage, which can be divided into three steps:

1. Recovering the sound signal from the groove position, which means to recover the sound as it was physically recorded on the disc.
2. Reconstituting the sound that has been effectively captured by the recording system. For this purpose we need some a priori knowledge on the recording process and pre-emphasis equalization.
3. Outputting the extracted signal in a standard file format without information loss.

If the first step is obviously mandatory in the VisualAudio process, the second and third steps must be flexible and parameterized as they depend on the needs and requirements of the final user. These three steps are developed in details in the next three sections.

7.1 Stored sound reconstitution

The edge extraction process results in edge positions sequences. These positions must be transformed into the sound as it has been recorded on the disc. First, we must combine the edge sequences produced by the groove extraction to recover each channel O_j of the output sound, which is described in Subsection 7.1.1. These channels O_j still correspond to position sequences; thus as the sound was stored in the needle's radial velocity, these O_j must then be derived to get the recorded sound. This derivation process is explained in Subsection 7.1.2. Subsection 7.1.3 then describes the filtering that is applied to remove the low frequencies produced by the scanning and the spiral shape of the groove.

7.1.1 Sound extraction in mono and stereo

The output of the groove extraction consists of two or four edge sequences S_{TO} , S_{BO} , S_{BI} and S_{TI} , depending on the number of traces that appeared for each groove circumvolution on the acquired images.

On monophonic records, the signal is encoded in the radial displacement of the groove. Thus the output monophonic audio signal O_M is defined by the average of the four extracted edges sequences for double traces grooves:

$$O_M = \frac{S_{TO} + S_{BO} + S_{BI} + S_{TI}}{4} \quad (7.1)$$

The single trace grooves have only two edges and O_M is then defined as:

$$O_M = \frac{S_{TO} + S_{TI}}{2} \quad (7.2)$$

In case of specific edge degradation or if the top and bottom edges present different levels of degradations and sound quality, then the output sound signal can be defined as any combination of the available edges sequences.

Stereophonic records are manufactured with the microgroove technology, which means that the groove bottom radius is very small (cf. Table 2.5) and thus there is only one visible trace per groove circumvolution on the acquired image. The modulations of the inner and outer stereo groove's walls encode the left O_L respectively right channel O_R (as explained in Section 2.2). The edge signals extracted by the groove extraction correspond to the surface displacement of the groove, not the wall displacement. Since these walls have an angle of 45° to the surface, the visible surface displacement d_s equals to the wall displacement d_w enhanced by a factor $\sqrt{2}$:

$$\sin(45^\circ) = \frac{d_w}{d_s} \quad \text{then} \quad d_s = \sqrt{2} \cdot d_w \quad (7.3)$$

Therefore the two extracted signal sequences S_{TI} and S_{TO} correspond linearly to the left O_L , respectively right O_R output channel:

$$O_L = \frac{S_{TI}}{\sqrt{2}} \quad \text{and} \quad O_R = \frac{S_{TO}}{\sqrt{2}} \quad (7.4)$$

7.1.2 Derivation

The extracted signal S is the groove position; but as stated in Chapter 2, the sound signal corresponds to the stylus radial velocity. To get the velocity of the needle, we need to differentiate the groove position. This is performed by a digital differentiator filter, which boosts the signal amplitude linearly to its frequency, resulting in a gain of 6 dB/octave. The frequency response of an ideal differentiator filter is proportional to the frequency [112]:

$$H_{DD}(j\omega) = j\omega \quad (7.5)$$

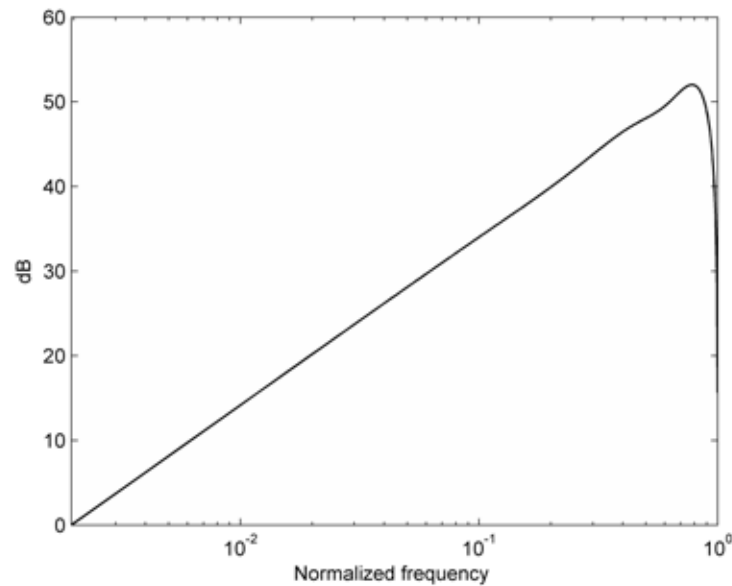


Figure 7.1: Frequency response of the digital differentiator filter on a normalized frequency range.

In practice, the ideal digital differentiator filter cannot be implemented, because the digital filter is periodic and the ideal impulse response would be of infinite length. Thus it must be approximated and a digital differentiator filter is then usually combined with a low-pass filter, which produces the derivation on a limited frequency band only.

In VisualAudio, the extracted signal is oversampled by factors three or higher. Therefore a digital differentiator working on a limited bandwidth is a satisfying solution. Such a filter can be computed using a filter design tool like the Matlab FDA tool [113]. The differentiator is implemented with a stable FIR (Finite Impulse Response) filter of length 11, which differentiates the input signal on a frequency band ranging from 0 Hz up to $0.8 \cdot f_s/2$ Hz, where f_s is the sampling frequency. Figure 7.1 shows the frequency response of this differentiator FIR filter, which difference with the ideal differentiator (Equation 7.5) is lower than 0.4 dB in the audible frequency band.

If de-emphasis equalization is applied on the extracted sound, the differentiator filter is combined directly with the de-emphasis, in order to reduce the number of filters and the computing time. This filter combination will be later explained in Subsection 7.2.3.

7.1.3 Low frequencies removal

As explained in Chapter 4, the low frequencies (below 10 Hz) are mainly due to the record spiral and to the film off-axis. Even if these frequencies are not audible, they are not part of the original sound and must be removed, because these low frequencies have large amplitudes, which may limit the dynamic of the whole sound

when stored in a digital file.

These low frequencies can be removed by applying a high-pass IIR (Infinite Impulse Response) filter removing the low frequency component under 20 Hz, which correspond to the lowest audible frequency [114].

7.2 Recorded sound recovering

The extracted sound could also be processed using some a priori knowledge on the sound and on the way it was recorded. The recorded sound may contain information of a limited bandwidth only. The higher frequencies of the human voice, for example, are around 4 kHz. Thus in some cases, it could be interesting to reduce the bandwidth of spoken sound document down to 4 kHz (instead of the 22.05 kHz of an audio CD for example), in order to remove higher frequencies, which then contain only noise. Additionally, the sound could possibly have been filtered at recording in two ways:

- By the recording material, which has a limited bandwidth due to mechanical and physical limitations. For further details, refer to [51, 38], which explain some of these limitations for early recording systems.
- By a pre-emphasis equalization (see Subsection 2.7.3).

If the bandwidth of the recording material is known, a low-pass filter can be applied with the same cut frequency, as explained in Subsection 7.2.1. De-emphasis is used to reverse the pre-emphasis equalization, as explained in Subsections 7.2.2 and 7.2.3. However the following considerations should be taken into account prior to applying any such processing:

- Background noise is part of the recording. So even if spoken words, for example, have a limited bandwidth, the background noise may contain some information, and it could be interesting to keep it.
- Human ear can perceive sounds from 20 Hz up to 20 kHz [114]. The discussion is still open to know whether higher frequencies and harmonics are also perceptible by other means than the ear. Interested readers are referred to [115, 116] for further details on the topic. Thus it is not advised to reduce the frequency band, as it may contain some information, even if it is not audible.
- De-emphasis is useful to reverse the recording mode, but the used pre-emphasis curve is not always known. In such a case, it is advised to use a flat transfer, which means to apply no de-emphasis [2].

These considerations overtake the scope of this work. The practical consequence is that the low-pass and de-emphasis processing depend strongly on the user needs and requirements. If the extracted sound is aimed to be published, for example, it should be processed to sound as close as possible to the original, even if the recording

mechanical limitations and pre-emphasis curve are unknown. On the other hand, if the sound is extracted to be stored in an archive database in a digital format, it should be as complete as possible (without any low-pass filtering) and the de-emphasis should be applied only when the original curve is precisely known. This way the digital file can be processed later, with newer technology or additional a priori information.

7.2.1 Removal of high frequencies

The higher out-of-band frequencies can be removed using a low-pass filter. The choice of the cut frequency f_c is defined by the user for each record independently.

In VisualAudio, this is implemented by a FIR filter with Hamming smoothing window. This filter is invoked multiple times for a steeper cut.

7.2.2 Pre-amplification equalization

As presented in Equation 7.6, the pre-amplification filtering transfer function is usually defined by three time constants τ_1 , τ_2 , τ_3 and a complex transfer function H_p :

$$H_p(j\omega) = \frac{1 + j\omega\tau_2}{(1 + j\omega\tau_1)(1 + j\omega\tau_3)} \quad (7.6)$$

The pre-emphasis equalization curves that are defined by two turnovers only, can be defined with the same transfer function. In this case, the first time constant must be set to at a high value corresponding to a very low frequency (under 20 Hz), which is inaudible and out of the recorded frequency band. The resulting magnitude response will be correct, with possibly a small uniform gain factor over the whole frequency band, which is negligible.

This analog filter H_p can be implemented by a second-order digital filter $G_p(z)$ of the general form:

$$G_p(z) = \frac{\sum_{m=0}^2 b_m z^{-m}}{\sum_{m=0}^2 a_m z^{-m}} \quad (7.7)$$

The coefficient a_m and b_m of the digital filter can be determined using the bilinear transform, which substitutes s to z , and maps the imaginary axis around the unit circle [117]:

$$G_p(s) = \frac{1 + s\tau_2}{(1 + s\tau_1)(1 + s\tau_3)} \quad (7.8)$$

where

$$s = \frac{2}{T_e} \cdot \frac{1 - z^{-1}}{1 + z^{-1}}, \quad T_e = \frac{1}{f_s} \quad \text{and} \quad T_i = f_s \cdot \tau_i \quad (7.9)$$

The drawback of the bilinear transform is that it introduces a non-linear correspondence between the imaginary axis and the unit circle. The frequency axis is then warped, and compressed more and more when approaching half the sampling

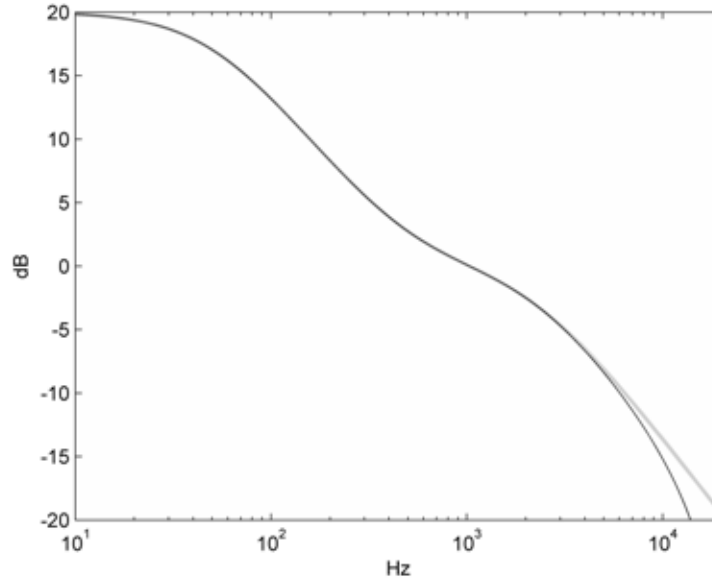


Figure 7.2: The analog (black) and digital (grey) RIAA de-emphasis filter responses are exactly superimposed up to 5 kHz, then the digital frequency response falls down faster than the analog.

frequency f_s . To compensate this frequency axis warping, the time constants must be warped the same way, so that the digital filter has the same frequency response as the analog filter for the audible bandwidth [117]. Equation 7.8 is then replaced by:

$$G_p(s) = \frac{1 + sT_2}{(1 + sT_1)(1 + sT_3)} \quad \text{where} \quad T_i = \frac{f_s \cdot \tan\left(\frac{\pi T_i}{f_s}\right)}{\pi} \quad (7.10)$$

By factorizing Equation 7.10, we get the following IIR filter coefficients:

$$b_0 = 1 + \frac{2T_2}{T_e} \quad a_0 = 1 + \frac{2T_1}{T_e} + \frac{2T_3}{T_e} + \frac{4T_1T_3}{T_e^2} \quad (7.11)$$

$$b_1 = 2 \quad a_1 = 2 - \frac{8T_1T_3}{T_e^2} \quad (7.12)$$

$$b_2 = 1 - \frac{2T_2}{T_e} \quad a_2 = 1 - \frac{2T_1}{T_e} - \frac{2T_3}{T_e} + \frac{4T_1T_3}{T_e^2} \quad (7.13)$$

These coefficients must still be divided by a_0 to normalize the coefficient $a_0 = 1$. The response of this digital filter $G_p(z)$ matches exactly the analog response of $H_p(j\omega)$ up to 5 kHz, as displayed on Figure 7.2. At 10 kHz, the digital filter response presents a 1.5 dB loss compared to the analog version. As mentioned in Subsection 2.7.3, this digital implementation could then be enhanced with a higher frequency turnover, which will attenuate the higher frequencies response difference. But the used high turnovers are not standardized and rarely documented.

7.2.3 Equalization combined with derivation

The derivation (Subsection 7.1.2) and equalization (Subsection 7.2.2) can be combined in a single filter, for simplification. Thus the pre-amplification transfer function H_p (Equation 7.6) can be combined with a differentiator filter (Equation 7.5), which defines the combined derivation and pre-amplification transfer function H_d :

$$H_d(j\omega) = \frac{j\omega\tau_1(1 + j\omega\tau_2)}{(1 + j\omega\tau_1)(1 + j\omega\tau_3)} \quad (7.14)$$

The additional τ_1 factor on the numerator is the relative gain, which is necessary to get the 0 dB level on the middle frequency band (between the frequencies f_1 and f_2 corresponding to the time constants τ_1 and τ_2).

The implementation of $H_d(j\omega)$ has the same general form $G_d(z)$ as in Equation 7.7. By using the bilinear transform on Equation 7.14, we get the following coefficients values for the IIR second-order filter $G_d(z)$:

$$b_0 = \frac{2T_1}{T_e} \left(\frac{2T_2}{T_e} + 1 \right) \quad a_0 = 1 + \frac{2T_1}{T_e} + \frac{2T_3}{T_e} + \frac{4T_1T_3}{T_e^2} \quad (7.15)$$

$$b_1 = -\frac{8T_1T_2}{T_e^2} \quad a_1 = 2 - \frac{8T_1T_3}{T_e^2} \quad (7.16)$$

$$b_2 = \frac{2T_1}{T_e} \left(\frac{2T_2}{T_e} - 1 \right) \quad a_2 = 1 - \frac{2T_1}{T_e} - \frac{2T_3}{T_e} + \frac{4T_1T_3}{T_e^2} \quad (7.17)$$

These coefficients must still be divided by a_0 to normalize the coefficient $a_0 = 1$.

As displayed on Figure 7.3, the analog and digital combined filter responses are almost equal on the whole frequency band. At 10 kHz, the digital filter response presents a 0.1 dB loss compared to the analog version, and this difference is then almost constant for higher frequencies. Thus this combined differentiator and RIAA de-emphasis is more accurate than the use of two separate digital filters as presented in Subsections 7.1.2 and 7.2.2. The reason is that the two infinite asymptotes of the de-emphasis and the differentiator filters compensate each other, leading to a horizontal asymptote (Figure 6.2), which is much less sensitive to the time warping deformation at high frequencies.

However, the comparison with the analog version is slightly biased, as the real pre-emphasis filter usually uses an additional high frequency turnover, which changes the higher frequencies response (cf. Subsection 2.7.3 and [56]).

7.3 Output file format

The output audio content will be saved in a wave file, which stores digital sounds in an uncompressed lossless format. This format supports a variety of samples rates, bit resolutions and number of channels. Thus the output bit depth and sampling rate can be adapted to suit the various standards in use for audio archiving, mainly 16 or 24 bits encoding sampled at 44.1 kHz, 48 kHz or 96 kHz. The number of channels

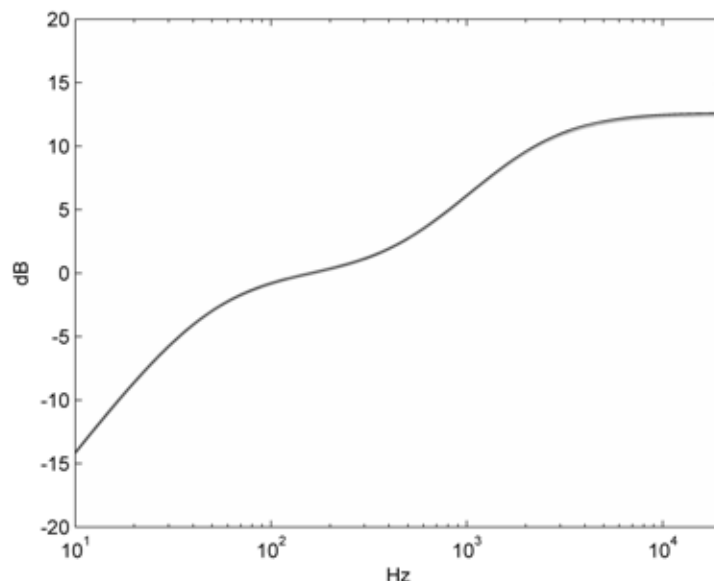


Figure 7.3: The digital (grey) combined (derivative and RIAA de-emphasis) filter response is almost perfectly superimposed to the analog (black) response on the whole frequency band.

has already been defined in Section 7.1.1, and it is then not useful to further explore this topic.

7.3.1 Sampling rate

The scanner captures the image at a line rate (LPR) of either 65k-lines per rotation or 131k-lines per rotation. These image sampling rates combined with the original record speeds RPM (rotation per minute) define the audio sampling frequency f_{as} as follows:

$$f_{as} = \frac{RPM \cdot LPR}{60} \quad (7.18)$$

This results in the audio sampling rates presented in Table 7.1.

LPR (acquired lines per scanner rotation)	65536	131072
Output audio sampling frequency for 78 rpm	85'196 Hz	170'393 Hz
Output audio sampling frequency for 33 rpm	36'408 Hz	72'817 Hz

Table 7.1: Effective audio sampling frequencies for various records rotation and scanner acquisition speeds.

The main standard audio sampling rates used for recording and archiving are 44.1 kHz (audio CD), 48 kHz (DAT and DVD) and 96 kHz (DVD). The audio output file from the VisualAudio process must then be resampled to match one of these standard sampling rates. All these sampling rates provide enough bandwidth to store sounds extracted from old records at either 33 and 78 rpm (see Table 7.1).

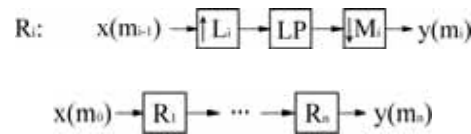


Figure 7.4: Multistage resampling: each resampling stage R_i is defined by an interpolation of factor L_i , a low-pass filter LP and a decimation of factor M_i . The complete L/M resampling is performed by the sequence of R_i .

However, it should be noticed that the output audio sampling frequency for a 33 rpm resulting from an acquisition performed with 65k-lines/rotation (36'408 Hz) may be insufficient to cover the whole bandwidth of modern 33 rpm and may produce aliasing. Therefore it is advised to acquire 33 rpm images with 131k-lines/rotation. To perform a sampling rate conversion of any rational factor L/M , the easiest way is to first interpolate the signal by an integer factor L , and then to decimate it by an integer factor M . A digital low-pass filter, with a cutoff frequency depending on the maximum between L and M , must be applied between the two operations to avoid aliasing. If we want to reach a final sampling frequency of 44.1 kHz with an audio extraction having $LPR = 65536$ for example, the sampling rate conversion must then use a rational factor of 44100/85196. These numerator and denominator are relatively big, which will lead to high computational and storage needs. To reduce the storage needs and the computational complexity of the resampling process, we use a multistage scheme, which consists in a cascade of decimations and interpolations. Crochiere advises to use multistage implementation for resampling in such cases, when $L \gg 1$, $M \gg 1$ and when the L/M ratio is close to 1 [118]. The L/M factor is then decomposed into smaller rational factors $L1/M1, L2/M2...$ where:

$$L = \prod_i L_i \quad \text{and} \quad M = \prod_i M_i \quad (7.19)$$

These factors are displayed in Tables 7.2 and 7.3 for varying disc speeds, scanning speeds and output frequencies. The resampling is then applied as shown on Figure 7.4.

Resampling must first result in higher temporary sampling rates to avoid aliasing and keep the whole frequency bandwidth content; therefore the rational factors L_i/M_i bigger than one are used at the first resampling stages. By ordering the rational factors this way, the intermediate low-pass filter can be performed using the same cut-off frequency f_c as defined in Subsection 7.2.1. The resulting resampled signal is weakened by the downsampling process and it must then be enhanced by a gain corresponding to the denominator $M(= M1 \cdot M2 \cdot M3 \cdot \dots)$ of this resampling process.

7.3.2 Bit depth

Wav audio files are usually encoded with a dynamic of either 16 or 24 bits, which corresponds to a theoretical dynamic range of 96 dB, respectively 144 dB. The real

	33 rpm	78 rpm
L / M	(65536 * SSF) / 79380	(65536 * SSF) / 33792
L1 / M1	7 / 4	3 / (4*SSF)
L2 / M2	7 / 4	11 / 16
L3 / M3	15 / 16	-
L4 / M4	9 / 16	-
L5 / M5	3 / (4*SSF)	-

Table 7.2: Resampling factors to reach 44.1 kHz: the scan speed factor is defined to $SSF = 1$ for a sampling rate of 65K-lines/rotation and $SSF = 2$ for a sampling rate of 131k-lines/rotation.

	33 rpm	78 rpm
L / M	(65536 * SSF) / (86400*OFF)	(65536 * SSF) / (36800 * OFF)
L1 / M1	5 / 4	23 / 16
L2 / M2	5 / 4	(5*OFF) / 8
L3 / M3	9 / 8	5 / (8*SSF)
L4 / M4	(3*OFF) / (4*SSF)	-

Table 7.3: Resampling factors to reach 48 kHz or 96 kHz: the scan speed factor is defined to $SSF = 1$ for 65K-lines/rotation and $SSF = 1$ for 131k-lines/rotation. The output frequency factor is defined to $OFF = 1$ for an output sampling frequency of 48 kHz and $OFF = 2$ for 96 kHz.

dynamic of an audio file is determined by the ratio between its lower and higher amplitude components. The lower bound of the dynamic depends on the system noise (white noise), which limits the ability to store low amplitude components of the sound. The higher amplitude components are usually produced by the clicks and other local sound degradations, which may have very large amplitude compared to the sound. Therefore the "Guidelines for Audio Preservation" advises to work with 24 bits encoding, to ensure sufficient dynamic in case of local defects and clicks [2].

The clicks amplitude in the VisualAudio system is limited by the groove following stage of the image processing (see Chapter 5). On the other side, the white noise is relatively high, as explained in the resolution analysis in Chapter 4. Therefore the real output audio bit depth is correctly encoded with 16 bits for records without major degradations.

All the signal processing parts of the VisualAudio program work with double precision 64 bits floating point numbers. This ensures sufficient accuracy for the intermediate storage and processing of the sound information and enough dynamic for the output audio content. The output audio must then be normalized to fit in the reduced 16 or 24 bits depth. This normalization is straightforward: as we know the order of magnitude of the maximum groove modulation amplitude, the photography magnification ratio (usually 1:1) and the scanning optics magnification used ($4\times$ or $10\times$), it is easy to determine the expected maximum amplitude of the resulting file and to normalize the output audio bit depth in order to reach the desired standard of 16 or 24 bits. This normalization has been empirically defined by a multiplication with the following factor η :

$$\eta = 10^{g/20} \cdot \frac{10}{m \cdot 512} \quad (7.20)$$

This factor η is defined using the magnification ratio m of the scanning optics and a parameterized gain factor g in dB, which allows calibrating the audio output

as desired. In case of highly degraded sound extraction, the normalization can be performed automatically, using the maximum detected amplitude contained in the file. Unfortunately, this will also lower the overall volume, which will no more be standardized and could then vary from an extraction to the next.

Chapter 8

Evaluation

This chapter evaluates the processing and the sound quality of the VisualAudio system. All the tests and quality measurements presented in the next sections have been performed on sounds extracted with the VisualAudio system. This way we work with real data, which have passed through the whole imaging chain.

Section 8.1 first introduces the measures and tests protocols used for the evaluation. The tests of the acquisition process are presented in Section 8.2, in order to evaluate the impact of the various sampling and image acquisition parameters. The groove extraction and image processing methods are then evaluated in Section 8.3. Section 8.4 presents the mass testing performed on different kinds of records and discusses the results. Section 8.5 focuses on the processing time which is needed to extract sound with the VisualAudio system and Section 8.6 evaluates the VisualAudio performance and compares the output sound quality with the performance of standard turntables. Conclusions about the tests and evaluations are drawn in Section 8.7.

8.1 Measures and tests protocols

We used specific records and measurements to evaluate the quality of the VisualAudio system. Subsection 8.1.1 presents the records that have been used to perform the tests. The conventions used for the acquisition naming are then detailed in 8.1.2. There are mainly three measures, which were used for these tests: the Signal to Noise Ratio (*SNR*), the Total Harmonic Distortion (*THD*) and the Standard Deviation (*STD*), which are defined in Subsections 8.1.3 to 8.1.5. Subsection 8.1.6 then defines how and where to apply these measures.

8.1.1 Tests records

It is difficult to accurately measure the noise and the distortions on natural speech or music signals. Thus we performed most of our measurements on test records containing specific tracks: tracks with a reference signal which is a pure tone (single frequency without harmonics) or silent tracks.

The main focus of VisualAudio is to extract sound from 78 rpm direct cut records. Unfortunately it is much harder to find 78 rpm direct cut test records than pressed 33 rpm tests records. Thus our tests are partly performed on 33 rpm. The advantage of using 33 rpm pressed records is that the discs are newer and produced with high technology recording machine, thus the visual aspect and the audio content are of high quality leading to accurate measurements.

78 rpm rotates 2.3 faster than 33 rpm, leading to the same ratio for the audio sampling frequencies: a 1 kHz tone on a 33 rpm has the same period length than a 2.3 kHz on a 78 rpm. However, the amplitude of the signals used for records and turntable evaluations are the same, defined by a 7 cm/sec maximum velocity at 1 kHz (as presented in Section 2.10).

Table 8.1 presents the four records used for these tests and Table 8.2 displays the content of the tracks used.

Record ID	Label	Speed	Type
Bur5	Columbia CZ 986	78 rpm	Shellac pressed record
U2218	Radio Genossenschaft Basel No U2218	78 rpm	Acetate direct cut record
HFN 001	Hi-fi news test record HFN 001	33 rpm	Vinyl pressed record
PS	Test record Popular Science	33 rpm	Vinyl pressed record

Table 8.1: Description and references of the records used for the evaluation.

Record ID	Track	Mean radius	Sound content
Bur5	-	11.9 cm	Beginning of record with low level sound content
Bur5	-	5.4 cm	End of record with low level sound content
U2218	-	13.8 cm	Beginning of record with low level sound content
U2218	-	6.4 cm	End of record with low level sound content
HFN 001	1	14.1 cm	300 Hz frequency with 15 dB gain
HFN 001	6	7.7 cm	Silent unmodulated groove
HFN 001	7	6.8 cm	300 Hz frequency with 15 dB gain
PS	1	14.2 cm	1000 Hz, peak velocity 7cm/sec
PS	2	13.4 cm	Silent unmodulated groove
PS	6	10.5 cm	maximum level glide tone 20 kHz to 20 Hz (sweep)
PS	9	6.4 cm	1000 Hz, peak velocity 7cm/sec

Table 8.2: Description of the various tracks used for the evaluation, where the record ID refers to Table 8.1.

Since the signal frequencies and levels are not the same on all the tracks and test records, the measurement results will not be comparable from one record to the other. The 300 Hz tracks have a gain of +15 dB, and the signal to noise results presented in this chapter are therefore around 15 dB higher than the signal to noise ratio for the PS record (see Tables 8.1 and 8.2).

8.1.2 Acquisition naming conventions

All the acquisitions used in this chapter are named with a compound string containing the following information:

- Record ID (as specified in Table 8.1).
- Track No: available only for 33 rpm records (as specified in Table 8.2).

- Opening time of the CCD camera in microseconds.
- Acquisitions performed at a scanning frequency of 131 k-lines per scanner rotation are additionally labeled "131K". The default sampling frequency of 65 k-lines is not mentioned in the acquisition name.
- In some specific cases, the acquisitions are numbered to distinguish between several acquisitions performed in the same conditions on the same record.

Thus, the acquisition "HFN001_track1_OT20", for example, was performed on the first track of the HI-HI news test record HFN 001, using a camera opening time of 20 μ s, and sampled with 65 k-lines per scanner rotation.

8.1.3 SNR (Signal to Noise Ratio)

The *SNR* (signal to noise ratio) is the power ratio between a signal and the background noise. In the scope of VisualAudio, the *SNR* is computed on sounds extracted using a Hanning window on tracks containing a unique frequency f_0 . *SNR* is expressed in dB and is computed as follows:

$$SNR = 10 \cdot \log_{10} \left(\frac{P_{sig}}{P_{noise}} \right) \quad (8.1)$$

where P_{sig} is the signal power at f_0 and P_{noise} the noise power. P_{sig} is measured on a frequency range width of $\pm f_{margin}$ around the main peak frequency. Usual frequency range is fixed by $f_{margin} = 5$ Hz. P_{noise} is the power at the remaining frequencies out of the frequency range $f_0 \pm f_{margin}$, from 100 Hz up to 10 kHz.

SNR_1 is the value of the best *SNR* measure over one consecutive second of audio content. SNR_{total} is the measure of the *SNR* over a complete ring extraction.

8.1.4 THD (Total Harmonic Distortion)

The *THD* (total harmonic distortion) of a signal is the ratio of the sum of the powers of all harmonic frequencies above the fundamental frequency to the power of the fundamental. In the scope of VisualAudio, the *THD* is computed on sounds extracted using a Hanning window on tracks containing a unique frequency f_0 . *THD* is expressed in dB and is computed as follows:

$$THD = 10 \cdot \log_{10} \left(\frac{P_{harmonics}}{P_{sig}} \right) \quad (8.2)$$

where P_{sig} is the signal power at f_0 and is measured on a frequency range width of $\pm f_{margin}$, with $f_{margin} = 5$ Hz. $P_{harmonics}$ is the power of the harmonics ($2f_0, 3f_0, \dots$) up to $f_{max} = 10$ kHz. Harmonic peaks are also considered on a frequency range width of $\pm f_{margin}$ around the main peak frequency, with the usual frequency margin fixed by $f_{margin} = 5$ Hz.

The *THD* is a way to measure the distortion level, which is due to the acquisition process (geometrical distortion) or the groove extraction processing (edge detection error) for example.

The *THD* measure may be biased by the white noise level, since the highest harmonics are embedded in white noise. THD_1 is the value of the best *THD* measure over one consecutive second of audio content. THD_{total} is the measure of the *THD* over a complete ring extraction.

8.1.5 STD (Standard Deviation)

The displacement between consecutive samples of a signal extracted from an unmodulated groove is mainly driven by the noise, but is also affected to a much lesser extent by the film off-axis and the spiral. Thus the standard deviation of such an unmodulated signal is a good measure of the white noise level. Unfortunately, this measure is affected by blur, which produces low-pass filtering on the signal and lowers the standard deviation measurements over close samples. Thus, in the scope of VisualAudio, the *STD* (standard deviation) is the measure in μm of the standard deviation σ_n of D , which is the displacement of an edge e between d samples:

$$D(i) = \frac{e(i) - e(i + d)}{\sqrt{2}} \quad (8.3)$$

where the $\sqrt{2}$ is used to get the standard deviation of a single sample. By measuring the $STD(d)$ with different d values, the *STD* can be evaluated without the bias produced by the blur low-pass. Figure 8.1 shows σ_n for $d = 1$ to 15. The $STD(d)$ for the smaller d values up to a point d_0 is affected by blur, and thus are lowered. For $d \geq d_0 + 1$, the $STD(d)$ can be approximated by a straight line of equation:

$$ax + b = y \quad (8.4)$$

where the slope a is mainly driven by the lower frequencies component of the signal. In the present case, these lower components are the sound (for a modulated groove), the film off-axis and the spiral of the record. The b parameter of Equation 8.4 is then a good approximation of the white noise standard deviation to define the *STD*:

$$STD = b \quad (8.5)$$

Since this signal is affected by blur, this measured *STD* is not a measure of the white noise over the whole frequency bandwidth until $f_s/2$, but it approximates the white noise level on the frequency band that still presents white noise characteristics after the low-pass filter, which is up to a cut frequency f_c . f_c can be defined by:

$$f_c = \frac{f_s}{2 \cdot d_0} \quad (8.6)$$

Thus if we want to use this value for the signal to noise calculations as presented on Equations 4.24, 4.33 and 4.37, we must consider only the limited frequency bandwidth and not the whole bandwidth until $f_s/2$.

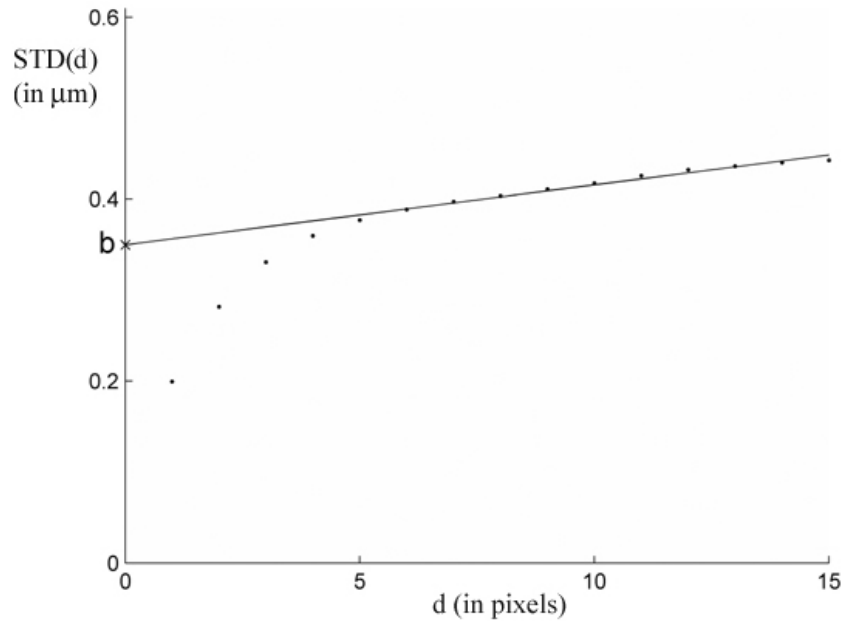


Figure 8.1: Measures of the $STD(d)$ for $d = 1$ to 15. The b value is a good estimation of the STD without the blur bias.

The advantage of the STD measure is that it can be performed on almost any kind of record as there are unmodulated circumvolutions of the groove at the beginning and end of almost each track. Complete silent tracks can also be found on test records.

STD_1 is the value of the best STD measure over one consecutive second of audio content. STD_{total} is the measure of the STD over a complete ring extraction.

8.1.6 Measurement domain

The SNR , THD and STD measurements are defined over several distinct measurement domains.

The measurement over a complete ring is sensitive to light variations and local degradations; but it is a good indicator of the stability of an acquisition. However measures on a complete image ring cannot be used to compare acquisitions performed using different magnification lenses, as the sound extraction duration varies according to the acquired surface on the record. It cannot also be used to compare silent sections at the beginning and end of a record, as these do not necessarily cover an entire acquisition ring. Even the pure frequency tracks used for the signal to noise measurements do not cover an entire ring image. Therefore, we introduced measurements on limited sections of the signal: over one second, over one groove circumvolution, and over one quarter groove circumvolution. Thus SNR_{total} , THD_{total} and STD_{total} are performed over a complete ring extraction, SNR_1 , THD_1 and STD_1 are measured over one consecutive second of audio con-

tent, SNR_c , THD_c , STD_c over one circumvolution, and finally SNR_q , THD_q , STD_q over one quarter groove circumvolution. Table 8.3 summarizes these measurements.

	SNR	THD	STD
Over whole ring	SNR_{total}	THD_{total}	STD_{total}
Over one second	SNR_1	THD_1	STD_1
Over one groove circumvolution	SNR_c	THD_c	STD_c
Over one quarter groove circumvolution	SNR_q	THD_q	STD_q

Table 8.3: Measurement domains for the SNR , THD and STD .

The measurements on one second of audio content, which are presented in this section, are the best values over one consecutive second within one ring sound extraction. This measure is more representative of the best achievable quality. It is also much more independent from local degradations of the image or any geometrical, light or blur variation that may occur between the border and the center of the digitized image.

In Section 4.3, we considered that noise is mainly due to the film graininess, which can be considered as white noise on a constant amplitude signal. Thus the SNR , THD and STD measures are performed on the extracted groove positions, which is a signal considered as having the constant amplitude characteristics, and which conserves the white noise hypothesis.

Thus it is also possible to compare different sounds from different kinds of records, even if they are not equalized with the same de-emphasis.

It should be noticed that the signal to noise measurements to evaluate records and turntables performance (see Section 2.10) apply on signals having constant velocity properties. Thus the SNR results cannot be directly compared to the values in Table 2.9. However, the correspondence between the constant amplitude, constant velocity and equalized signal to noise measurements was given in Section 4.3.5, and comparison between VisualAudio and turntables will be presented later in Section 8.6.

8.1.7 Parameters

The values presented in these tests may slightly vary for a same track from a table to the other, depending on the acquisition conditions and processing parameters. But for each set of acquisitions in each table and each spectrum display, there is always only one parameter change: the variation of the performance is then directly related to the unique parameter change.

Unless otherwise specified, the main parameters which were used for all the acquisitions presented in this chapter are fixed as follows:

- Sampling frequency: 65k-lines per rotation.
- Edge detection method is Threshold02: local threshold $\beta = 0.2$ (as defined in Subsection 5.2.4.1).

- Corrections: 1D impulses detection, with the number of iterations $max_{iter} = 2$, the maximum distance between the checked pixels $E = 9$, and the high and low thresholds $\tau_h = 4$ and $\tau_l = 2$ (as defined in Subsection 6.3.5).

8.2 Evaluation of the acquisition process

The image acquisition process has some influence on the final quality of the sound extraction. This section evaluates the conditions of the image acquisitions, in order to evaluate their effect on the extracted sound. This includes the acquisition regularity or reproducibility (see Subsection 8.2.1), the optics (Subsection 8.2.2), the sampling process (Subsections 8.2.3 and 8.2.4), the illumination (Subsection 8.2.5) and the acquisition of tracks located at various radial positions over the record (Subsection 8.2.6).

8.2.1 Acquisition regularity

Acquisitions performed with the VisualAudio scanner are stable and consecutive acquisitions of the same film in the same conditions lead to very similar images. To measure this regularity, we acquired several ring images I of the same film, at the same radial position where the sound modulation presents high amplitude. The scanning parameters (camera opening time, light level, scanning frequency...) were strictly the same. The image difference D_{ij} between $I_i - I_j$ is then defined as follows:

$$D_{ij} = I_i - I_j \quad (8.7)$$

The standard deviation of D_{ij} measured several times for several acquired rings i, j was around 2, measured in grey level for eight bits images with 256 grey levels. This is close to the CCD camera random noise RMS (Root Mean Square), which is of 0.8 DN (Digital Numbers or grey levels) [65]. If we normalize D_{ij} to be able to visualize the image difference in grey levels, we get the image as partially shown on Figure 8.2. In fact, the normalized difference image D_{ij} shows only a small shift between I_i and I_j that is smaller than the size of a pixel and was presumably produced by a small jitter of the sampling starting point, which can be due either to the electronic or to the belt which drives the scanner rotation axis.

This demonstrates that the rotation speed and the illumination during the acquisition are very stable, and that the differences in the sound extraction quality, which appear between two extractions in the next sections, are not due to scanning or sampling irregularities.

8.2.2 Optics

Two magnification lenses have been used for the scanner. They are both achromatic (with chromatic aberration correction) and work with a tube length of 160 mm:

1. 10× magnification with $NA = 0.25$

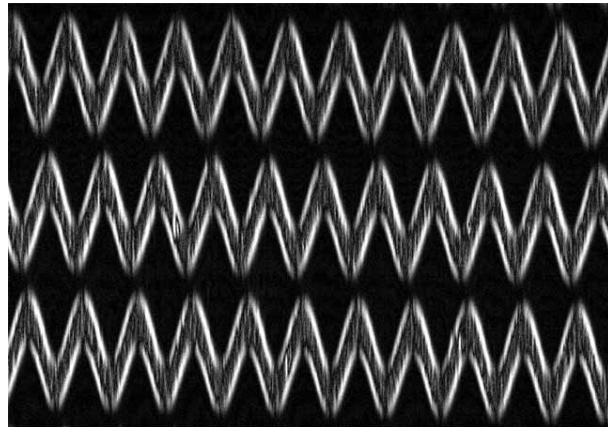


Figure 8.2: Sample of the normalized difference D_{ij} between two consecutive acquisitions: this puts in evidence a small regular shift between the two acquisitions.

2. $4\times$ magnification with $NA = 0.1$

The $10\times$ lens was used first and has been later replaced by the $4\times$ lens, as it matches perfectly our resolution needs as stated in Section 4.2.4. Some early tests have also been performed with a $20\times$ magnification lens; but the reduced depth of field of this lens was not sufficient to encompass the glass tray warping and the depth of the film emulsion layer. Thus the $20\times$ magnification is not considered in the current section.

8.2.2.1 Magnification

The exact magnification of a lens can change depending on the lens manufacturing and on the exact working distances that are used in a specific system. It is not of high importance to know the exact magnification ratio for the sound extraction, since the sound is encoded in the relative radial displacement of the groove and not absolute values. But we must know the exact radial size of a pixel, as the radial displacement of the camera between consecutive rings is driven by a number of pixels. Knowledge of the exact magnification ratio is also useful for testing purposes, since noise size measurements can then be given in μm for example.

Thus both lenses have been calibrated in order to know the exact magnification. This calibration has been performed in two steps. First the magnification has been evaluated by scanning a transparent micro-scale ruler. The second calibration measurement has been performed using the radial motor, to know the exact camera displacement, according to a given position in pixels. For each lens, the two calibration measurements lead to the same magnification ratio, which are displayed in Table 8.4. A correction factor has then been defined in order to match the number of pixels for radial displacements and the number of mm for the radial motor move.

Optics	Real magnification	Correction factor
10×	10.75×	0.935
4×	5.26×	0.76

Table 8.4: Real magnification ratio of the two lenses as used on the VisualAudio scanner.

8.2.2.2 Extraction quality

The extracted sound quality has been measured using a 300 Hz track and the results are shown on Table 8.5. The results obtained with the 4× lens are slightly better. This confirms the resolution analysis of Section 4.2.4, which stated that the best resolution would be reached with a 4× magnification lens.

Acquisition	Lens	SNR_{total}	SNR_1	THD_{total}	THD_1
HFN001_track7_OT70_1	4×	32.8	35.8	-38.8	-40.2
HFN001_track7_OT70_2	4×	32.9	34.8	-39.0	-40.1
HFN001_track7_OT70_3	10×	30.5	35.1	-39.2	-40.0
HFN001_track7_OT70_4	10×	29.2	33.3	-40.3	-41.4

Table 8.5: SNR and THD measured on acquisition of the same track performed with the 4× and the 10× lens.

Harmonics have been compared on acquisitions with both lenses and the results are displayed on Table 8.6. The sound extracted with the 4× magnification has been enhanced by 6 dB, to reach the same power at the 300 Hz main peak than the 10× magnification acquisition, in order to compare the level of the other harmonics.

Peak	Lens 4×	Lens 10×
Main peak 300 Hz	-63 dB	-63 dB
2nd harmonic 600 Hz	-107 dB	-107 dB
3rd harmonic 900 Hz	-109 dB	-110 dB
4th harmonic 1200 Hz	-115 dB	-116 dB
5th harmonic 1500 Hz	-126 dB	-126 dB

Table 8.6: Peak and harmonics power: the harmonics are similar with both lenses.

The 4× magnification lens matches the resolution needs stated in Subsection 4.2.4. Moreover, it presents the advantage that a full record is scanned with less than half the number of rings that would be necessary with the 10× lens: the acquisition time is then divided by two. The drawback of the 4× lens is that the scanned area is larger, and that it is harder to have a homogeneous illumination over the whole scanned area. The effect of such inhomogeneous light will be further studied in Subsection 8.2.5. The 4× lens increases also the scanning blur (see Section 4.2.4). This blur slightly increases the lowpass filter effect and is visible on the higher frequencies of the spectrum in Figure 8.3.

8.2.3 Opening time

The sensors integrate a film area during a given opening time. When working with a longer opening time, the camera integrates a larger area having then a higher density of film grains. Therefore working with a longer opening time will lower

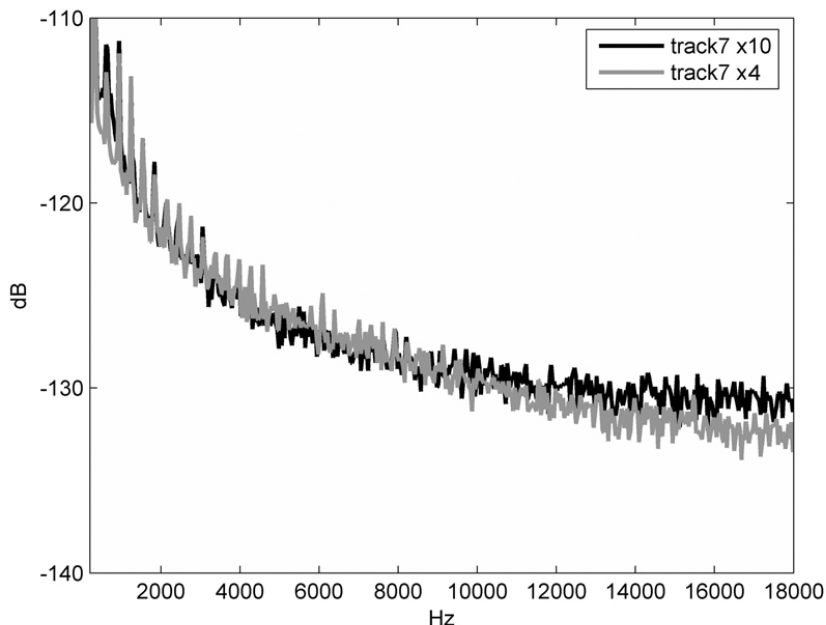


Figure 8.3: Superimposed spectra of the sound extracted from the track 7 acquired with the 4 \times and 10 \times magnification lenses. The 4 \times extracted sound has been increased by 6 dB to normalize the 300 Hz peak on both spectra. The higher frequencies of the 4 \times lens extraction are attenuated by a higher scanning blur level.

the grey level uncertainty which produces noise. However, a longer opening time increases the scanning blur, which attenuates the higher frequencies of the extracted sound.

With the camera sampling rate of 13 kHz, the maximum opening time is 76 μs for each sample. For test purposes, we performed acquisitions on the same film with opening times ranging from 20 to 70 μs . Figure 8.4 shows the STD_1 of an unmodulated groove, which correspond to the acquisition noise, for varying opening times. The effect of the opening time variation is clearly visible: when the opening time triples, the STD_1 decreases by a factor 1.25.

Figure 8.5 shows the effect of the opening time variation: when the opening time increases, the blur increases but the noise decreases. Figure 8.6 shows the spectrum of three acquisitions performed with opening times of 20, 40 and 60 μs . Thus the high frequencies attenuation due to the longer integration time and to the scanning blur is visible. Frequencies are then attenuated above 5 kHz: this means that the sound content above 5 kHz would also be affected. Thus it is not desirable to have a too long opening time, in order to keep the same output level for all frequencies. This 5 kHz cut frequency correspond to 11.8 kHz for a 78 rpm ($5 \cdot 78 / 33 = 11.8 \text{ kHz}$)

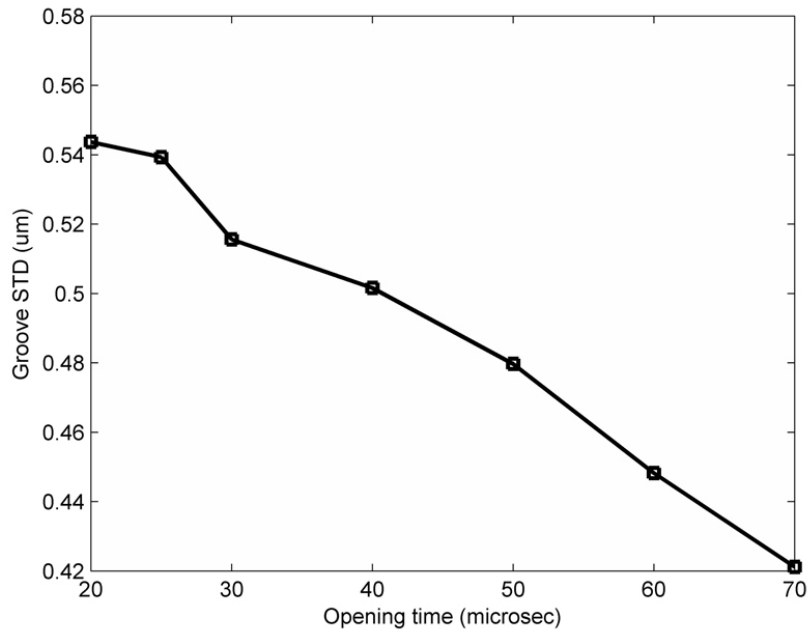


Figure 8.4: STD_1 of a signal extracted from an unmodulated groove in μm , for varying opening times in μs .

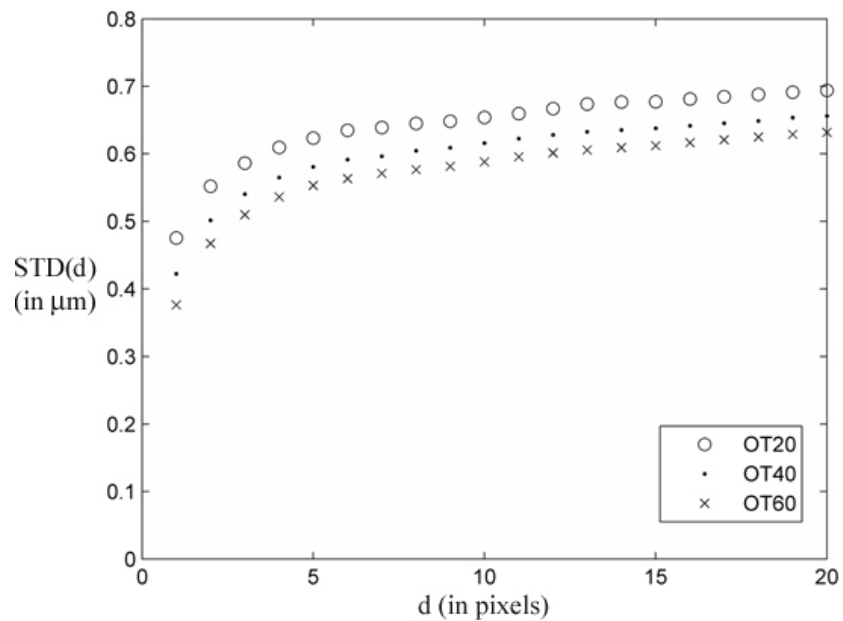


Figure 8.5: $STD(d)$ of an unmodulated groove acquired with camera opening times of 20, 40 and 60 μs . When the opening time increases, the blur increases but the noise decreases.

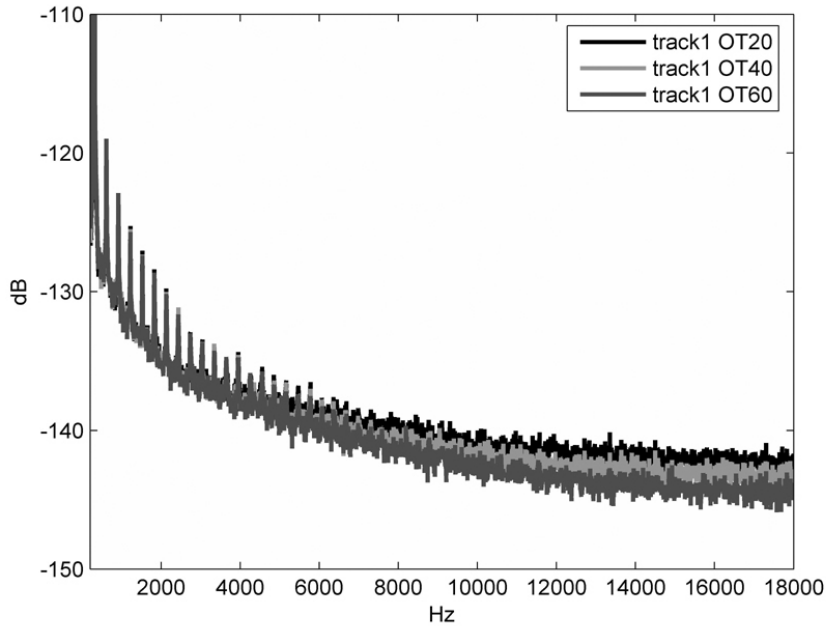


Figure 8.6: Spectrum of three acquisitions performed with camera opening times of 20, 40 and 60 μs .

8.2.4 Oversampling

The minimum image sampling rate is given by the Shannon sampling theorem, which states that the sampling frequency must be at least twice the signal bandwidth [117]. For example, if we would like to get a sound extraction up to 22 kHz out of the image of a 78 rpm record, the minimum image sampling rate should be of 33.9 k-lines/ring ($= 44.1 \times 60/78$). Working with higher image sampling rate will spread the white noise on a wider bandwidth, and thus lower the noise level in the audible bandwidth, which ranges from 20 Hz to 20 kHz. This process is called oversampling.

During the image acquisition, each sensor integrates the light reflected by a rotating area of the record surface at every sampling period. Then if we increase the sampling rate, we increase the number of samples to acquire the same surface, which means that we will integrate a smaller surface for each sensor at each sampling time. This means that the noise, which is present on the record picture (film grain, dust...), will be less averaged and that the noise level on the extracted sound will increase. On the other hand, the size of the film area integrated by a sensor at each sampling time decreases inversely proportionally to the sampling frequency. Therefore the optical blur affects more samples at higher sampling frequency, and the low pass filter produced by the optical blur has a lower cut frequency. So what is the real gain when working with such oversampled images?

Acquisitions have been performed with the two available image sampling frequencies of 65 and 131 k-lines/ring. We have measured the SNR on such sounds, and Table 8.7 displays the SNR_1 results. The resulting SNR and THD were

almost identical at both scanning frequencies, showing that the oversampling gain is very limited.

Acquisition	65k	131k
HFN001_track1_OT20	34.64	35.41
HFN001_track7_OT60	36.94	37.09
PS_track1_OT60	19.15	19.82
PS_track9_OT20	20.27	20.63

Table 8.7: SNR_1 on 1 second of sound content on a 10 kHz bandwidth: the SNR_1 results on 131k acquisition are better, due to the lower noise level above 10 kHz, which is due to the low-pass filtering produced by the acquisition blur.

Figures 8.7 and 8.8 show comparative spectra of sounds extracted from 65 and 131 k-lines/ring acquisitions. These spectra show that there are less high frequency noise for the 131 k-lines/ring acquisition. Thus the oversampling gain is only due to a high frequency attenuation, and not to a lower white noise level on the whole frequency band.

The STD has been measured to evaluate the blur level. Figure 8.9 shows the $STD(d)$ and puts in evidence the effect of the sampling blur while oversampling: the STD at 131 k-lines/ring is lower for close pixels, when $d < 4$, and higher when $d > 5$. The number of blurred pixels d_0 is also around two times higher for the 131 k-lines/ring acquisition than for 65 k-lines/ring.

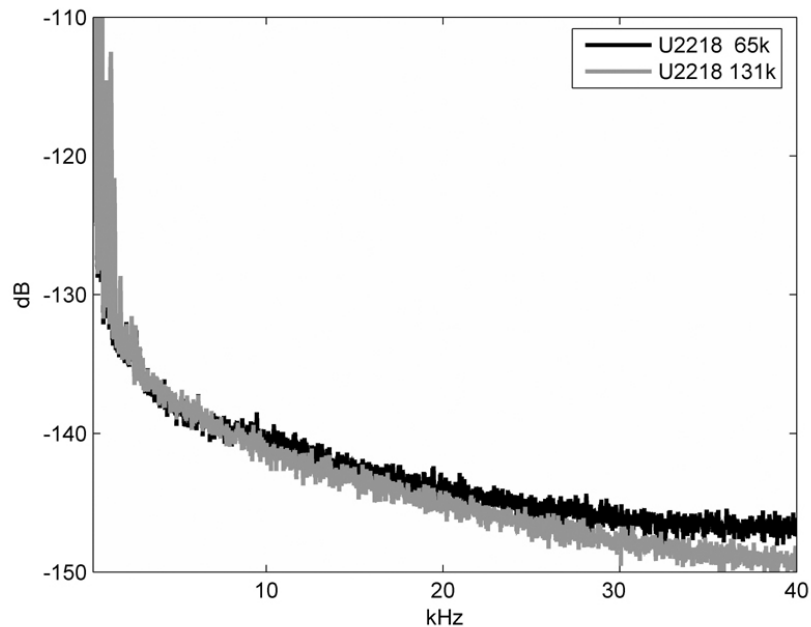


Figure 8.7: Spectrum of a signals acquired at 65 k-lines/ring and 131 k-lines/ring on an unmodulated area of a 78 rpm record.

Oversampling does not significantly enhance the extracted sound quality. This is explained by two reasons:

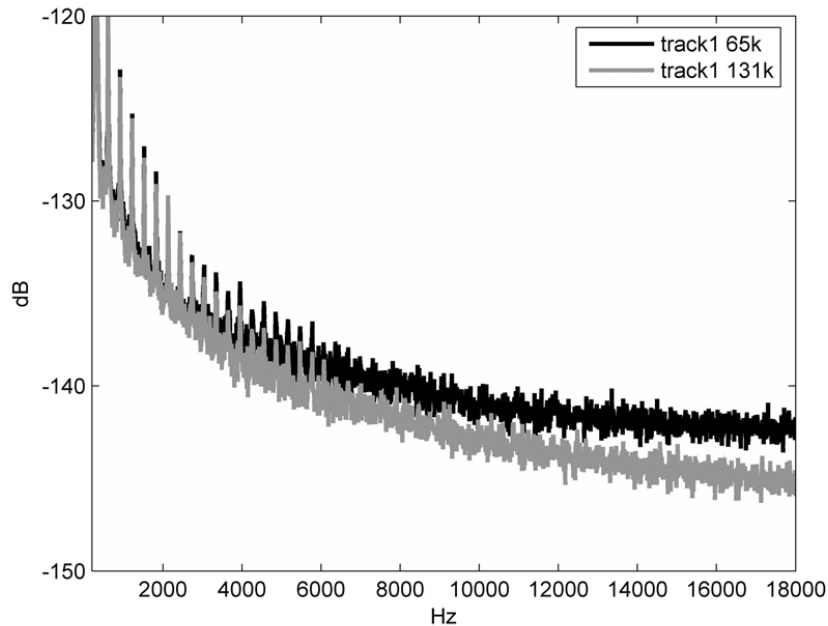


Figure 8.8: Spectrum of signals acquired at 65 k-lines/ring and 131 k-lines/ring on a 300 Hz track of a 33 rpm.

1. The optical blur size is the same for both sampling frequencies. When we increase the sampling frequency, the size of the samples decreases. Therefore, the blur affects twice the number of samples at 131k than at 65k, which means that the cut frequency of the low-pass filter due to the blur is almost the same at 65k and 131k.
2. Noise is not uniformly distributed, but has some low frequency components which are not modified by an oversampling process.

8.2.5 Light intensity and grey levels dynamic

The scanning light level can be changed to be adapted to the film darkness and camera opening time. The grey levels of the acquired images are related to the chosen light level, and it is interesting to evaluate the effect of the light level variation on the final sound quality. This way we can also analyze whether the image dynamic (or bit depth, which defines the number of grey levels) is sufficient or whether it is a limiting factor, which lowers the sound quality.

We performed several acquisitions of the same film (track 7 of the HFN record, containing a 300 Hz frequency), by varying the light intensity. The light levels are numbered from 0 (stronger) to 255 (lower). Acquisitions and measurements were performed at every 5 light levels between 0 and 90. The images acquired with level 95 and above are not considered as they show a grey level dynamic of at most 1 grey level between the lighter and darker pixel, which obviously gives insufficient

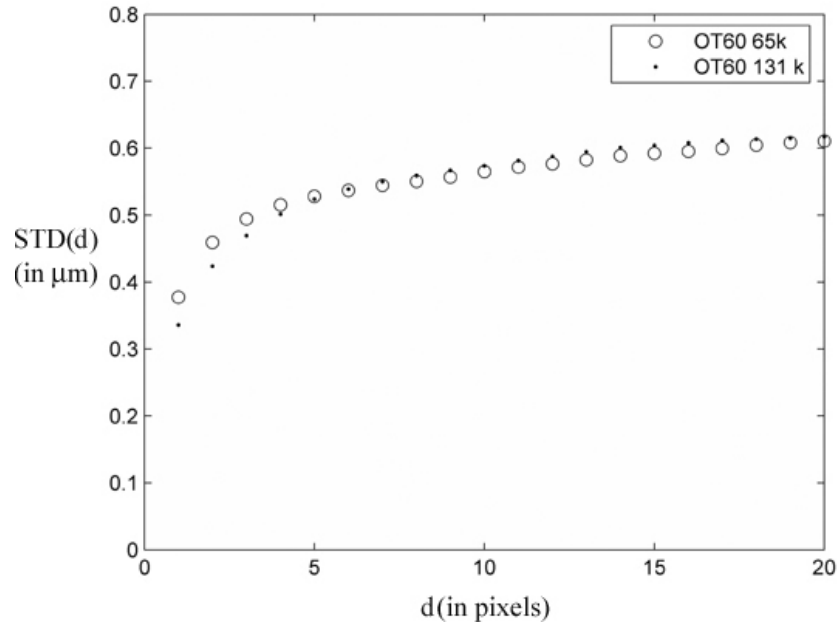


Figure 8.9: $STD(d)$ of two unmodulated grooves acquired at 65 k-lines/ring and 131 k-lines/ring: due to the higher blur level at 131k, the 131k acquisition has a lower STD for $d < 5$ and a higher STD for $d > 8$. The size of the blur measured in pixels d_0 can be approximated as $d_0 = 4$ at 65k and $d_0 = 7$ at 131k.

contrast for image processing.

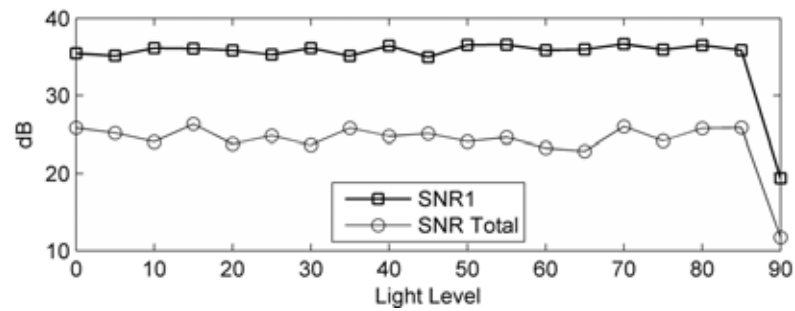
We measured the SNR_1 , SNR_{total} , THD_1 and THD_{total} and put them into relation with the minimum, mean and maximum grey level of the acquired image for each light level. Results are displayed on Figure 8.10. While the minimum grey level is almost constant at 4 or 5 DN (Digital Numbers or grey levels) for all the acquisitions, the maximum grey level is saturated at 255 DN up to light level 60, and decreases linearly since then up to light level 90. SNR_1 , SNR_{total} , THD_1 and THD_{total} show almost constant values up to light level 85, where the grey level dynamic is only of 43 DN. This demonstrates first that the light saturation at scanning does not lower the extracted sound quality. The second observation is that when the dynamic is lowered by a factor 5 from 255 DN down to 43 DN (at light level 85), the SNR and THD measurements are not affected. Thus the light level and grey level bit depth are not limiting factors for the current sound extraction process.

It should be noticed that these measurements were performed on uncalibrated acquisitions; but calibrated images provide similar results.

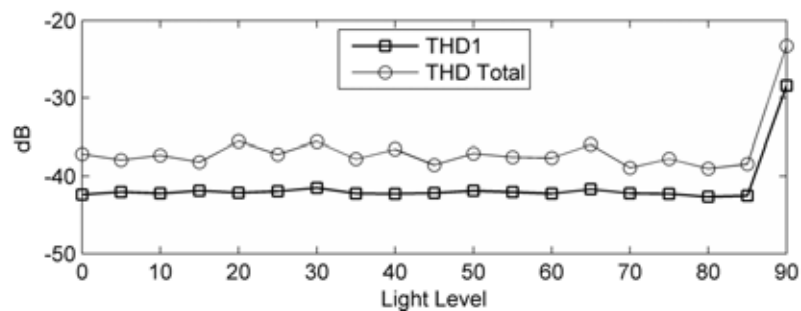
8.2.6 Quality variation over a record acquisition

The image acquisition, and thus also the sound extraction, is not always of the same quality over a ring and over a whole record. Thus we measured some variations that occur, in order to understand them, and to be able to limit these variations in future

(a)



(b)



(c)

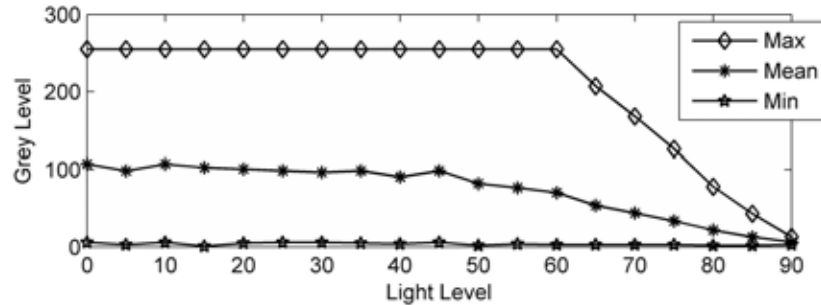


Figure 8.10: SNR_1 (a) and THD_1 (b) were measured on acquisitions with light levels ranging from 0 to 90. The graph (c) shows the maximum, mean and minimum grey levels of the acquired image for each light level acquisition.

versions of the VisualAudio scanner.

8.2.6.1 Radial variations

Lenses are corrected for geometrical aberrations; but there may be still some focus differences between the center and the border of the observed area. Thus we measured the SNR_c of all the groove circumvolutions on the same acquisition ring. Variations ranging from 1 to 3 dB occurred among the sound extractions of the different groove circumvolutions of the same acquired ring. This is due to the focus

variations between the center and the border of the acquired ring, which causes blur, lowers the high frequency noise and thus increases the SNR . Figure 8.11 shows the SNR_c measured on the first and last track of the test record, both containing a 300 Hz tone. These acquisitions were done on the same film and the same focus adjustment with the $4\times$ optics. The variations are clearly visible between the center and border SNR_c values. Another interesting fact is that the acquisition was in focus at the center of the acquisition on track 1 and on the border of the acquisition on track 7, which means that the glass tray is not perfectly flat and presents radial warping. It should be noticed that the blur variations affecting the SNR_c are also visible on the acquired images.

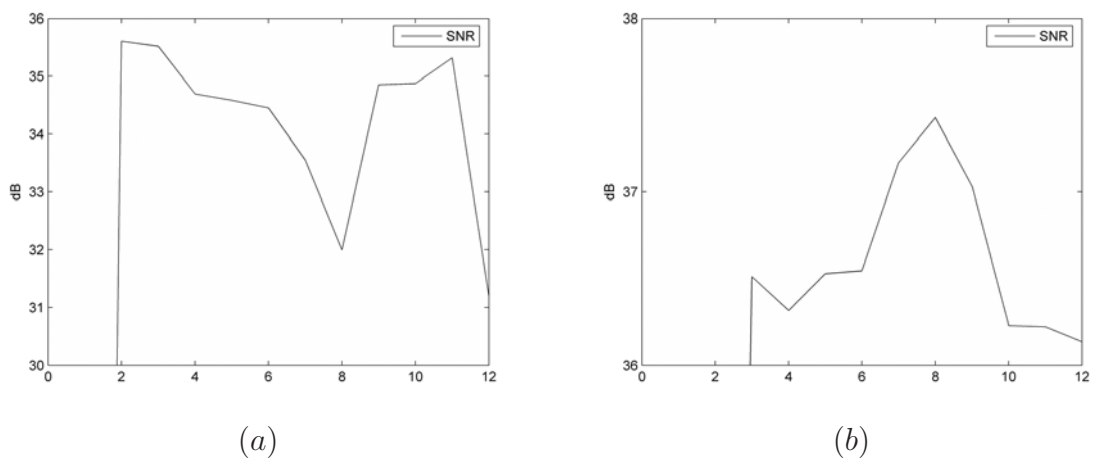


Figure 8.11: SNR_c measured for each of the twelve groove circumvolutions inside the same ring image: (a) on the first track, (b) on the last track of the test record. It should be noticed that the beginning and end of the tracks contain no sound, resulting in an irrelevantly lower SNR_c .

These radial variations among a single ring are more perceptible with the $4\times$ magnification lens than with the $10\times$ optics. This is because the $4\times$ magnification lens acquires a larger surface than the $10\times$ optics, producing more focus differences between the center and the sides of the ring image. It could also possibly be caused by a poorer lens manufacturing.

The easiest way not to be affected from the radial variation inside each acquired ring is to work only with the central part of the acquired ring and to reject the borders. The variation due to the glass tray should be corrected either by working with a flatter glass try, or by adjusting the focus several times during the whole record acquisition.

8.2.6.2 Tangential variations

If the focus varies around a circumvolution of the scanner, the sound extraction quality could also vary. Thus we measured the SNR_q over the whole acquisition using a sliding window representing one quarter of the ring length (16384 samples). The evolution of this SNR_q over six consecutive circumvolutions is shown on Figure

8.12. The local variations of the SNR_q show a periodicity that is clearly related with the scanner rotation: there are six periods corresponding to the six groove circumvolutions. This shows that focus is not constant over a circumvolution and evolves with the angular position of the film on the scanner. These variations range between 0 and 3 dB. These tangential focus variations can be also corrected using a flatter glass tray at scanning, or using an autofocus system.

A warped record may also produce some irregular focus over the acquired ring, producing the same kind of SNR_q variations with a periodicity related to a scanner circumvolution. But the depth of field at photography is much deeper than at scanning, and a record warping must be much more important to be perceptible on the acquired ring. Since the small image degradations at the film level (pepper fog for example) are also affected by varying blur, it means that this blur variation is produced at the scanning and not at the photography step.

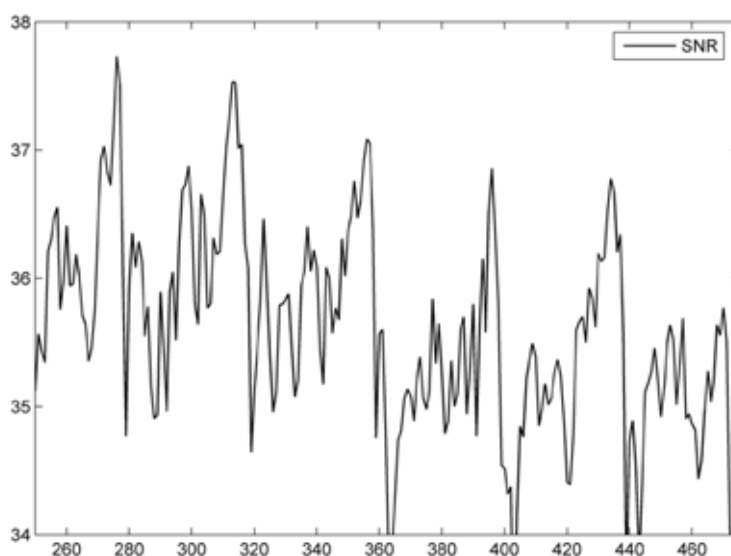


Figure 8.12: Evolution of the SNR_q measured on a sliding window over six consecutive groove circumvolutions. The variation periodicity is related to the angular position on the scanner.

8.2.6.3 Top, bottom, outer and inner edges

On the acquired record images, the top edges of the groove are produced by a presumably sharp cut and the bottom edges by the rounded groove bottom. Thus these two kinds of edges may lead to different audio quality on the extracted sound. The quality of the top and bottom edges of the groove varies also according to the state of conservation of the disc: a record with scratches and surface wear will have lower quality top edges. On the other hand, a record containing dust and fungus inside the groove will have lower quality bottom edges. The cutting stylus shape may influence the edge quality for direct cut records. The acquisitions of

such instantaneous records may also present less accurate top edges if the excavated record material pills up on both sides of the groove during the engraving process.

The SNR , THD could not be used since we did not have coarse groove record containing a track with a single frequency. Thus we measured the STD on each edge. Table 8.8 shows some STD_1 measurements on the four edges of various 78 rpm records. These STD tests show that the groove extraction reaches subpixel and submicron accuracy for all edges, but it did not allow determining which edges lead to the best sound extraction quality. There is no general rule and the edges quality changes from a record to the next. This is especially the case for direct cut discs, which show important variations of surface damages, engraving stylus and record manufacturing materials. Figure 8.13 shows two extreme cases of direct cut records acquisition: one showing better bottom edges, and the second with better top edges.

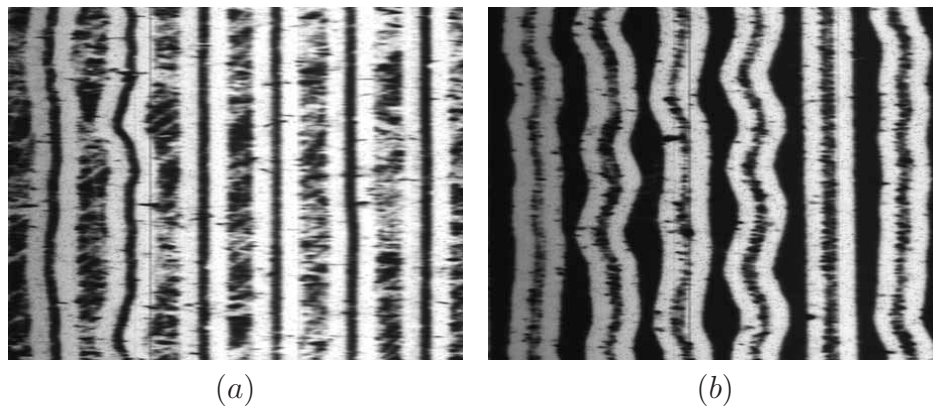


Figure 8.13: The quality of the top and bottom edges depends on the record state of conservation. These acquisitions of two different direct cut discs show two extreme cases: (a) the bottom edges are more reliable, (b) the top edges are less degraded.

A coarse groove test record with higher frequencies could help to determine the quality of the edges, especially for high frequencies, which could lead to different quality over the top and bottom edges. The resulting SNR will still depend on the cutting stylus shape.

Quality of the inner and outer top edges of 33 rpm were also compared on the same record acquisition and lead to similar results, as shown on Table 8.9 and 8.10. Variations appear between the edges, but it is not possible to relate these variations with the edge type (inner or outer).

Acquisition	Top outer	Bottom outer	Bottom inner	Top inner
bur5_ext_OT70_131k	0.34	0.45	0.34	0.37
bur5_int_OT70	0.42	0.38	0.38	0.30
U2218_ext_OT40	0.39	0.27	0.28	0.36
U2218_int_OT40	0.40	0.24	0.30	0.36

Table 8.8: STD_1 measured in μm for the four edges of various 78 rpm records.

Acquisition	Outer edge	Inner edge
PS_track1_OT60	15.52	16.44
PS_track9_OT20	20.65	18.12
HFN001_track1_OT20	28.88	31.86
HFN001_track7_OT40_131k	29.02	32.01

Table 8.9: Comparison of the SNR_{total} measured in dB for the outer and inner edges of 33 rpm test records.

Acquisition	Outer edge	Inner edge
PS_track2_OT20	0.50	0.49
PS_track2_OT40_131k	0.42	0.40
HFN001_track6_OT40	0.34	0.31
HFN001_track6_OT60_131k	0.31	0.28

Table 8.10: Comparison of the STD_1 measured in μm for the outer and inner edges of 33 rpm test records.

The amplitude of the second harmonic measured on the different edges and on the sum of the edges shows also interesting information. As displayed on Table 8.11, the sum of the inner and outer edges shows a 6 dB improvement of the main peak over each the same peak for each edge individually, which corresponds to the expected gain for amplification by a factor 2. On the other hand, the sum of both edges reduces the 2nd harmonic by 16 dB compared to the average of each edge. This means that part of the 2nd harmonic is due to a symmetric phenomenon that appears on both edges and which is canceled by summing both signals, mainly the edge detection error (cf. Section 5.2.5) and some effects of the scanning blur (see Section 4.2.4) for example.

Peak	Outer edge	Inner edge	Sum of both edges
Main peak 300 Hz	-63 dB	-64 dB	-57 dB
2nd harmonic 600 Hz	-93 dB	-92 dB	-109 dB

Table 8.11: Acquisition performed on acquisition HFN001_track7_OT70. Peak and 2nd harmonic power: the sum of both edges increases the main peak but lowers the harmonics level.

8.2.6.4 Radial position of the groove

On classical turntables, when the radial position of the needle on the record decreases, the noise produced by the graininess of the record material increases. Thus there may be signal to noise difference of more than 10 dB between the outer and inner groove circumvolutions on lacquer records [38].

The signal to noise variations due to the radial position on the VisualAudio system is slightly more complicated. There are mainly three phenomena which affect the noise level relatively to the radial position in opposite ways:

- Grain of the film: the size of the film area which is integrated by a sensor varies according to the radial position on the record. Thus the film grain density produces more uncertainty on the inner part of the record image, where the integrated area is smaller.

- Pepper fog (see Section 4.3.3): these grain clumps are evenly distributed on the film. Thus the samples acquired on the outer part of the record have a higher probability to be affected by such degradations, producing more noise on the outer groove circumvolutions. Figure 8.14 shows the effect of the pepper fog on the outer and inner acquisitions of the same film.
- Blur: increases inversely proportionally to the radial position, and thus lower the noise level on the inner groove circumvolutions.

Table 8.12, Table 8.13 and Table 8.14 present the SNR_1 , THD_1 and STD_1 measured on the outer and inner part of record pictures. The SNR_1 increases inversely proportionally to the radial position over the record and to the STD_1 . This is mainly due to the blur, which lowers the high frequency noise. This assumption is confirmed by the spectrum displayed on Figure 8.15 and 8.17. On the other hand, the harmonics level decreases on the inner part of the record. This is partly due to the blur, which attenuates the high frequencies, and partly due to the geometrical distortions which are more important on the outer part of the record image. Figure 8.16 shows the variation of the $STD(d)$ on the outer and inner part of a direct cut record. The inner acquisition is more affected by the blur, as the STD must be measured on more distant samples to reach a linear progression.

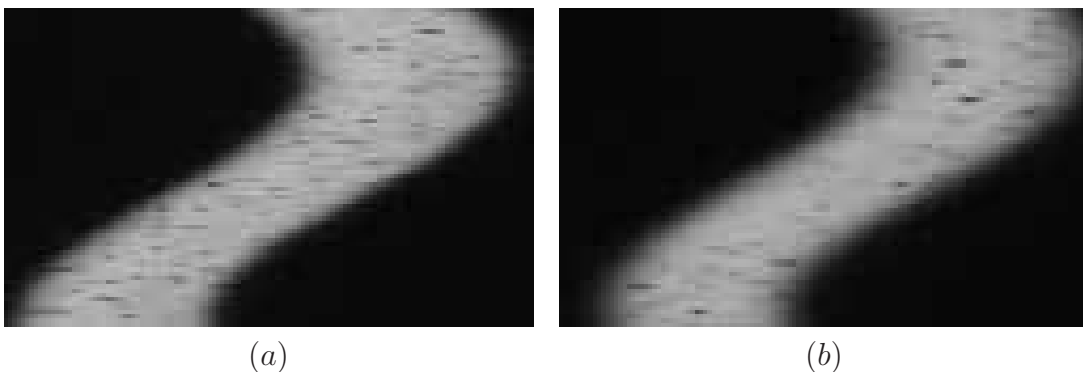


Figure 8.14: Views of acquisitions performed on track 1 (a) and track 7 (b) of the same film: there are more samples affected by the pepper fog on the outer track than on the inner track.

Acquisition	Outer track	Inner track
PS_OT20 1000 Hz	19.26	20.27
HFN001_OT60 300 Hz	34.56	36.94
HFN001_OT40_131k 300 Hz	34.68	36.97

Table 8.12: SNR_1 measured on similar tracks containing the same unique frequency on the outer and inner parts of the same record picture.

8.3 Image processing evaluation

This section evaluates the effects of the image processing on the extracted sound quality. Subsection 8.3.1 evaluates the camera calibration process, which has been

Acquisition	Outer track	Inner track
PS_OT20 1000 Hz	-30.09	-32.66
HFN001_OT60 300 Hz	-41.69	-42.00
HFN001_OT40_131k 300 Hz	-42.02	-42.17

Table 8.13: THD_1 measured on similar tracks containing the same unique frequency on the outer and inner parts of the same record picture.

Acquisition	Outer track	Inner track
HFN001_OT20	0.45	0.27
Bur5_OT70	0.53	0.39
U2218_int_OT40	0.42	0.32

Table 8.14: STD_1 measured on unmodulated groove circumvolutions on the outer and inner parts of the same record picture.

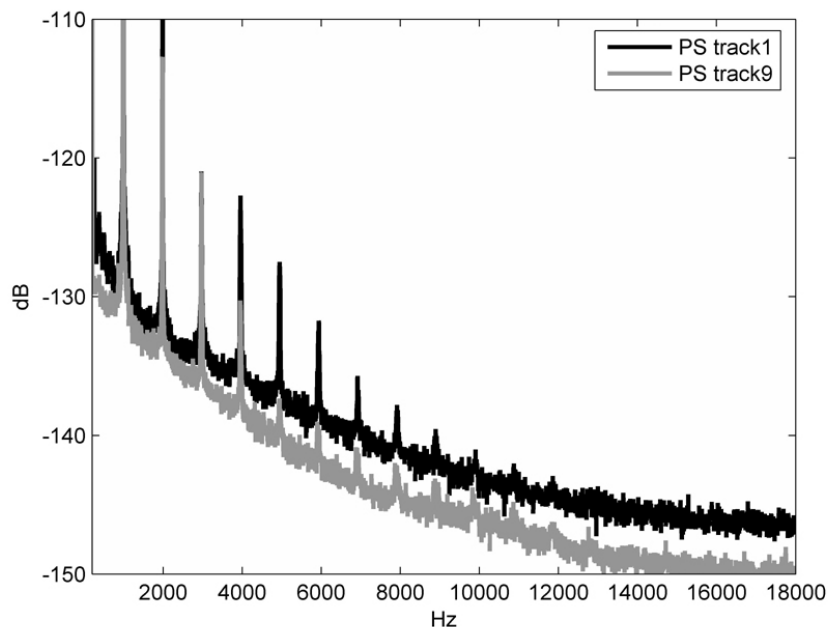


Figure 8.15: Spectrum of the tracks 1 and 9 of the PS test record, both containing a 1000 Hz frequency.

explained in Subsection 6.1.1. The edge detection methods developed in Chapter 5 are then compared in Subsection 8.3.2. Subsection 8.3.3 finally evaluates the image correction methods, which were presented in Chapter 6.

8.3.1 Camera calibration

The camera calibration method presented in Subsection 6.1.1 has been evaluated on several acquisitions. The results of the SNR measurements are displayed in table 8.15. These results show no difference on the output sound quality between the calibrated and uncalibrated acquisitions. This is explained by several reasons:

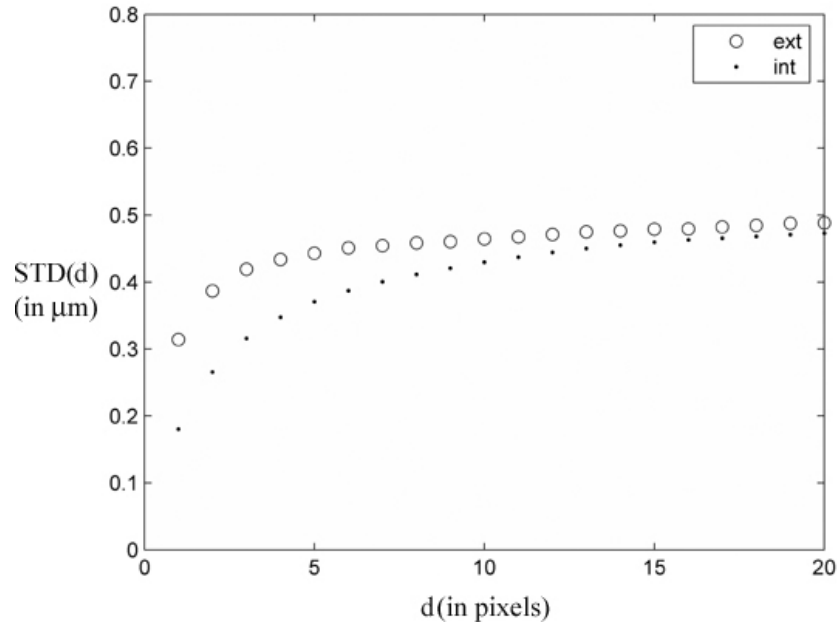


Figure 8.16: This Figure shows the $STD(d)$ measured on two unmodulated groove sections acquired on the outer (ext) and inner (int) parts of a direct cut record. The outer acquired signal is less affected by the blur.

- The fixed pattern noise of the CCD camera is only a minor source of noise in the VisualAudio imaging chain, and thus the correction of this pattern noise does not significantly increase the SNR .
- The non-homogeneities in the illumination do not lead to a major loss of quality, as already shown in Subsection 8.2.5.
- The edge detection process is already well adapted to some illumination variations among the traces.

Acquisition	Calibration	SNR_{total}	SNR_1
HFN001_track7_OT70	No	35.0	36.6
HFN001_track7_OT70	Yes	34.7	36.5
HFN001_track4_OT70	No	34.7	35.8
HFN001_track4_OT70	Yes	35.1	36.2

Table 8.15: SNR measured on acquisitions with and without the camera calibration.

8.3.2 Edge detection

The edge detection methods presented in Section 5.2 have been tested with several parameter configurations, which are summarized in Table 8.16. The LMS method (Subsection 5.2.4.2) is not presented in this evaluation, since it presents results similar to the weighted LMS method (Subsection 5.2.4.3). The four methods called

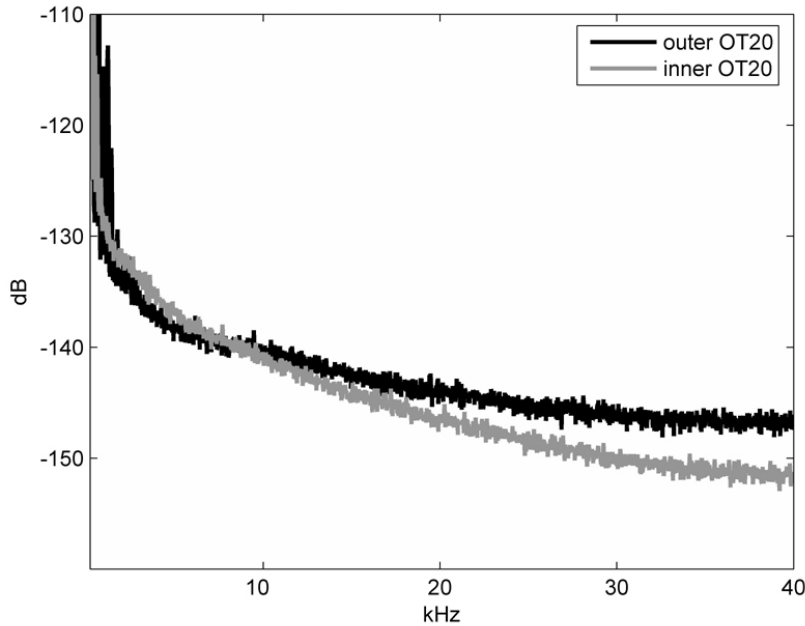


Figure 8.17: Spectrum of the sound extracted from the outermost and innermost groove circumvolutions of the U2218 78 rpm record. The audio content is not exactly the same on both extractions, but it clearly shows the low-pass effect produced by the blur on the inner groove extraction.

”box” correspond in fact to the coarse edge detection, which is applied without the fine edge detection: edges are localized at the extremum of the convolution with a box function. These methods are applied with large size masks, which smooth the image radially. Thus their results are given as a comparison basis to evaluate the other proposed methods: the weighted LMS and the local threshold.

Table 8.17 to Table 8.20 present the SNR_{total} , SNR_1 , THD_{total} and STD_1 tests results for the 10 configurations of edge detection methods. The Threshold02 obtains the best results for almost all these tests and all the acquisitions.

Another way to evaluate the edge detection method is to measure their reliability. For this purpose, we can look at the number of samples (or edge points) which were considered as *undefined* during the groove reconstruction process and the number of samples which necessitated interpolation after the 1D impulse detection. Table 8.21 shows such statistics for one acquisition. Once more, the Threshold02 method has the lowest rate of undefined samples, and almost the lowest rate of interpolated samples. It appears then that the simple threshold method is less affected by all the degradations which are present on these acquired images. These degradations require more corrections when we take into account a larger context for the edge detection, since there are more undefined edge points.

To tests the performance of the edge detection methods at higher frequencies, we used also a track containing a sweep. The ring acquisition was performed on the part of the track containing the sweep between 12 kHz and 3 kHz. The SNR_q has

Box20Pix	Convolution with a box function of size $\lambda = 20$ pixels
Box50Pix	Convolution with a box function of size $\lambda = 50$ pixels
Box50percent	Convolution with a box function of size $\lambda = \alpha w$, and $\alpha = 50\%$
Box80Percent	Convolution with a box function of size $\lambda = \alpha w$, and $\alpha = 80\%$
LMS5	Weighted LMS with $\lambda_2 = 5$ and $\chi = 0.8$
LMs10	Weighted LMS with $\lambda_2 = 10$ and $\chi = 0.9$
Threshold01	Local threshold with $\beta = 0.1$
Threshold02	Local threshold with $\beta = 0.2$
Threshold03	Local threshold with $\beta = 0.3$
Threshold04	Local threshold with $\beta = 0.4$

Table 8.16: Edge detection methods used in the current evaluation test. The parameters refer to the explanations of Section 5.2.

ED method	HFN001_track1_OT20	HFN001_track1_OT60	HFN001_track7_OT20_131k
Box20Pix	27.95	28.9	33.11
Box50Pix	13.39	9.89	19.49
Box50percent	28.25	29.34	33.13
Box80Percent	27.94	28.36	33.12
LMS5	28.92	29.01	33.66
LMs10	28.23	28.57	33.02
Threshold01	29.28	29.31	33.82
Threshold02	31.71	31.73	34.45
Threshold03	28.91	29.50	34.32
Threshold04	28.47	29.12	33.88

Table 8.17: SNR_{total} measured on several acquisitions. Acquisition PS_track9_OT40 is not presented in this table as the track is smaller than the acquired image, and the SNR_{total} measure is therefore not relevant.

ED method	PS_track9_OT40	HFN001_track1_OT20	HFN001_track7_OT20_131k
Box20Pix	18.42	30.45	35.43
Box50Pix	11.40	19.20	20.68
Box50percent	18.77	30.46	35.50
Box80Percent	17.65	30.72	34.93
LMS5	18.65	31.06	36.25
LMs10	18.87	30.87	36.40
Threshold01	21.17	31.89	36.65
Threshold02	20.58	34.64	37.12
Threshold03	19.49	31.36	36.93
Threshold04	18.78	30.91	36.49

Table 8.18: SNR_1 measured on several acquisitions.

ED method	PS_track9_OT40	HFN001_track1_OT20	HFN001_track7_OT20_131k
Box20Pix	-29.33	-36.52	-40.36
Box50Pix	-18.49	-20.35	-21.53
Box50percent	-30.96	-36.58	-40.37
Box80Percent	-31.13	-33.47	-39.13
LMS5	-29.50	-36.63	-40.28
LMs10	-29.16	-36.57	-39.87
Threshold01	-29.99	-36.65	-40.85
Threshold02	-30.45	-36.56	-40.83
Threshold03	-30.65	-36.92	-40.69
Threshold04	-30.61	-36.95	-40.58

Table 8.19: THD_{total} measured on several acquisitions.

been measured at regular intervals using several edge detection methods.

The frequency range used for this SNR_q calculation has been enhanced to $f_{margin} = \pm 50$ Hz (cf. Section 8.1.3), in order to encompass the larger signal peak produced by the frequency varying signal. This measurement method is not very

ED method	PS_track2_OT20	HFN001_track6_OT20	U2218_int_OT40
Box20Pix	0.74	0.48	0.42
Box50Pix	0.71	0.47	0.54
Box50percent	0.66	0.49	0.41
Box80Percent	0.65	0.49	0.41
LMS5	0.56	0.45	0.50
LMS10	0.56	0.60	0.59
Threshold01	0.49	0.34	0.33
Threshold02	0.49	0.32	0.32
Threshold03	0.50	0.34	0.33
Threshold04	0.50	0.36	0.34

Table 8.20: STD_1 measured on several acquisitions.

ED method	Undefined samples	Interpolated samples
Box20Pix	0.01%	4.62%
Box50Pix	6.06%	12.36%
Box50percent	0.01%	3.66%
Box80Percent	0.00%	3.20%
LMS5	3.82%	8.02%
LMS10	7.00%	12.45%
Threshold01	0.05%	2.15%
Threshold02	0.01%	2.28%
Threshold03	0.02%	2.66%
Threshold04	0.01%	3.14%

Table 8.21: Statistics of the samples considered as undefined during the edge detection and the interpolated samples for the groove extraction over one ring acquisition.

accurate, but it is sufficient to compare the frequency responses obtained with various edge detection methods, since these tests are all performed using the same acquired ring image. The high frequencies attenuation is similar for all edge detection methods with a different overall gain, and the best results at all frequencies are still reached by the Threshold02 method. Figure 8.18 displays these frequency responses.

The superiority of the Threshold02 method can be explained by several reasons. First it detects the edge position using the lower part of the linear grey level transition, and is therefore less sensitive to the noise than the Threshold03 or Threshold04, which detect the position at lighter grey levels. On the opposite, the Threshold01 may sometimes detect the edge in the non-linear part of the transition, leading to some non-linear errors, which increase the noise level and are not necessarily cancelled by summing the different edges. The LMS methods use a larger context (larger number of samples) to extract the edge position. Therefore they are more to the image degradations which occur in the grey level transition. The LMS methods use the information from the pixels that are in the middle of the transition, and which are therefore noisier than the pixels on the lower part of the linear grey level transition used for the Threshold02 method. Even the weighted version of the LMS does not increase the edge detection accuracy: as the linear part of the transition covers 6 to 10 pixels with a $4\times$ magnification, an optical blur of $5\mu\text{m}$ will spread the local image degradations over a large part of the transition, leaving at most a few uncorrupted pixels.

The results obtained with the acquisition HFN001_track7_OT20_131k shows that the differences between the various edge detection methods tends to vanish when

the blur level increases (due to the radial position and the higher sampling rate). However, the quality of the sound output can also be related to the number of corrections and interpolations performed. The current corrections are adapted for sparse small gaps. The edge detection methods which lead to correction of more than 10% of the samples are then penalized by the corrections methods.

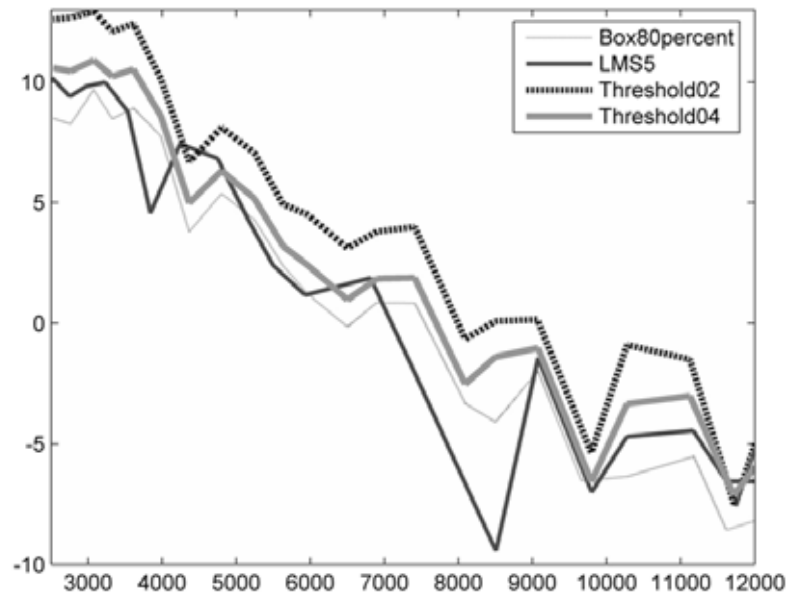


Figure 8.18: SNR_q measured on a sweep with various edge detection methods: the frequency response of all the methods are similar, with a different gain factor.

8.3.3 Corrections

Corrupted samples correction methods and their applications were discussed in Chapter 6. These methods affect also the output sound quality. We extract the sound from several acquisitions with four different correction schemes, defined as follows:

- No corrections: interpolations are performed only for the undefined samples.
- 2 iterations: 2 iterations of the 1D impulses detection, with $max_{iter} = 2$, $E = 9$, $\tau_h = 4$ and $\tau_l = 2$ (see Section 6.3.5).
- Corrupted pixel map (see Section 6.3.3).
- LMS4 (see Section 6.3.6).

The acquisition rings used for these tests did not present important degradations, with less than 0.5% of undefined samples.

Tables 8.22 to 8.24 display the SNR and THD results for these four correction schemes.

As expected, the THD_{total} levels are not lowered by the proposed corrections schemes, as the proposed schemes are intended to correct noise and not harmonic distortions. However, it shows that these corrections are sufficiently accurate to not increase the harmonics level.

The proposed corrections schemes increase both the SNR_{total} and SNR_1 . The 2 iterations and corrupted pixel map lead to almost similar sound quality.

One of the hypotheses for the corrupted pixel map is that the groove is almost parallel to the tangential direction, with only smooth variations. This assumption is not true for the test record used since these are 33 rpm with high amplitude signal. Therefore the corrupted pixel map correction is not well adapted in the current case and does not lead to much higher sound quality. The corrupted pixel map correction are more effective on natural sound recorded on 78 rpm, where the groove presents only smooth variations.

The LMS4 correction reaches better SNR results. This is mainly due to the signal smoothing, which perform a low-pass filter and attenuates strongly the high frequencies. This attenuation is clearly visible on Figure 8.19. Low pass filtering is not always desirable, as the sound itself may contain much higher frequencies than in the test record used in this case.

Acquisition	No Correction	2 iterations	Corrupted pixel map	LMS4
PS_track1_OT60	17.06	18.37	19.18	19.65
PS_track9_OT20	21.09	21.89	21.77	22.65
HFN001_track7_OT40	32.59	34.04	34.07	34.18

Table 8.22: Comparison of SNR_{total} with different methods of signal correction.

Acquisition	No Correction	2 iterations	Corrupted pixel map	LMS4
PS_track1_OT60	21.58	22.57	22.59	23.85
PS_track9_OT20	23.53	23.94	23.94	24.98
HFN001_track7_OT40	34.70	36.88	36.86	37.06

Table 8.23: Comparison of SNR_1 with different methods of signal correction.

Acquisition	No Correction	2 iterations	Corrupted pixel map	LMS4
PS_track1_OT60	-23.73	-23.99	-24.04	-25.27
PS_track9_OT20	-26.97	-27.23	-27.24	-27.52
HFN001_track7_OT40	-40.59	-40.58	-40.60	-40.59

Table 8.24: Comparison of THD_{total} with different methods of signal correction.

8.4 Mass testing

The VisualAudio system has already been used to extract sound from around 100 direct cut discs and more than 20 pressed records, provided by four sound archives from different countries. Thus these tests were performed on a large palette of many

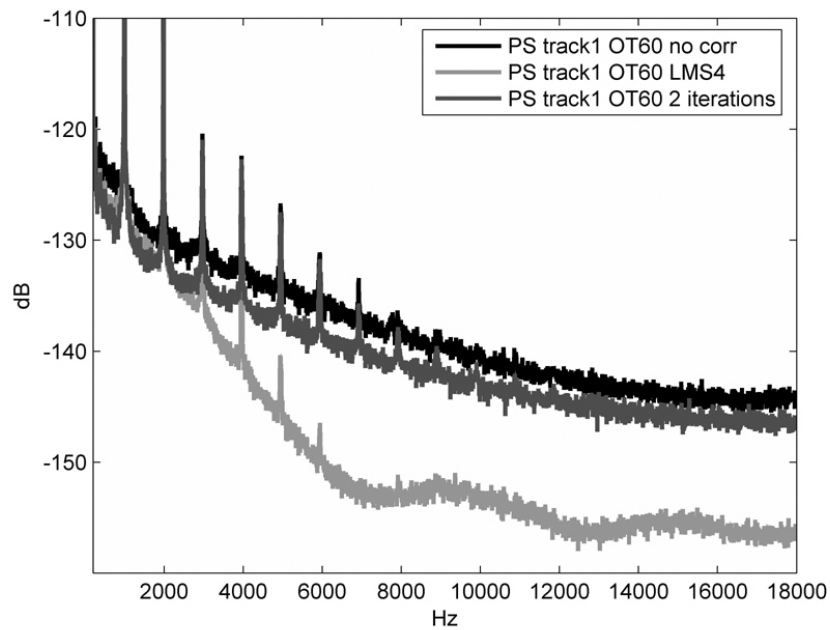


Figure 8.19: Spectrum of the PS_track1_OT60 sound extractions performed 1) without correction 2) with the LMS4 and 3) with 2 iterations of the 1D impulses detection.

different records, including coarse grooves and microgrooves, 33 rpm, 45 rpm, 60 rpm and 78 rpm, which diameter ranged between 17 cm and 40 cm and showing many different visual aspects. The records also presented many kinds of physical degradations: mould, scratches, shrinkage of the recording layer. As mentioned in Section 6.4, the correction of the records with shrinkage of the recording layer is not fully implemented; but we were able to extract audible sounds from some shrunk records, which presented only limited cracks, as the disc presented on Figure 8.20.



Figure 8.20: Audible sound could be extracted from this record, which shows limited shrinkage of the recording layer.

The main result from these tests is that we were able to extract sound from some records with mould that are currently considered as unreadable with standard turntables. Figure 8.21 shows one of these records, and Figure 8.22 shows the part of the edge extraction performed with the VisualAudio system. The edges were successfully extracted from the image using the 2-passes trace extraction, which allows to follow the low frequencies component of the groove modulations and avoid to follow the spurious patterns produced by the physical degradations.

The quality of the VisualAudio extraction will be discussed in details in the Section 8.6, however, we may point out a few additional comments out of these tests:

- The sound extractions are of reasonable quality for records with a low degradation level; but an important background noise is still present.
- Almost each new set of records introduced some new kind of physical degradations. Thus it is difficult to classify the degradations individually.
- One of the initial assumption was that we would work with black records. However, some of the records used for this mass testing were colored and translucent. Thus we also tested the VisualAudio sound extraction process on these records. They posed many problems at the photographic step and the resulting pictures are very noisy due to the light reflection on and inside the record surface. The tests to get good quality pictures of such records are still ongoing.



Figure 8.21: Record with mould produced by the growth of fungus: this kind of record is currently considered as unplayable with any standard turntable; but it is possible to extract sound of such records using the VisualAudio process.

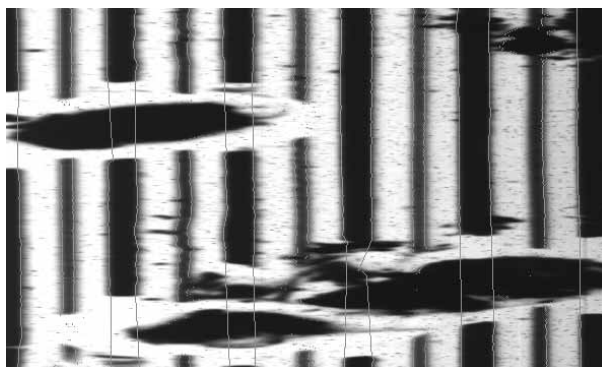


Figure 8.22: Local view of the acquisition of a record affect by mould. The light grey lines show the edges extracted with the VisualAudio system.

8.5 Processing time

The processing time is always an important factor in an automated system. However, it has more or less impact, depending on the use of the system: if the VisualAudio system is intended to be used as an archiving mean, the film scanning and processing is made on demand, and the processing time is then not a major bottleneck in the mass saving process. If the VisualAudio system is used as a transfer system (to transfer the sound from the records to digital files), then the processing time is more important, since all the records must be processed to generate digital files.

The VisualAudio scanning and processing times are proportional to the size of the scanned surface on the disc. For a 33 rpm record in good condition, which does not require any specific correction, the scanning and processing time is around 22 minutes with a 3,4 GHz Pentium 4 processor. This processing time is then almost equal to the recorded audio duration. Since the 78 rpm records have around 4 times less audio content than the 33 rpm for the same recorded surface, the ratio of the scanning and processing time to the audio duration is of 4:1. It should be noticed that this time duration encompasses neither the photography time, nor the handling (film fixation and centering, glass tray cleaning,...)

To optimize the processing time, we can separate the scanning and the image processing stages. This way the image processing of several extractions can be parallelized on several computers and performed remotely during the night for example. Therefore we limit the time using the scanner, which is the most critical resource of the system. The scanning time is then equal to 1.5 times the recorded sound duration.

The scanning and sound extraction times are displayed on Table 8.25 for coarse groove 78 rpm and microgroove 33 rpm records. These times are expressed in ratio to the recorded sound duration.

	Coarse groove 78 rpm	Microgroove 33 rpm
Scanning time	3:2	1:3
Processing time	5:2	2:3
Total time	4:1	1:1

Table 8.25: Ratio of the processing and scanning times to the recorded sound duration.

8.5.1 Time optimization

The processing time was not the main focus of this work, and the VisualAudio processing time is currently not fully optimized and can be decreased.

Some simple hardware optimizations could be performed, like using a faster CCD camera to decrease the scanning time or to parallelize the sound extraction processing of different records. There is no need to parallelize the record processing: since the system is intended to extract many records, the most natural way to parallelize is to distribute the extraction of n records on n computers at a time.

Three software optimizations are currently proposed to reduce the processing time:

1. The different stage of the processing are currently performed sequentially, roughly:
 - The image acquisition, performed either with the CCD camera (on-line acquisition) or loaded from a file (off-line acquisition)
 - Image pre-processing (corrupted pixel map, smoothing)
 - Groove extraction pass one
 - Possibly: second pass of the groove extraction
 - Signal correction

This way the various processes are independent. A possible time optimization consists in interleaving all the image processing steps in order to parse the image only once.

2. The acquired images are currently fully loaded into the memory before the image processing. Thus there is no difference between the processing of the acquired images provided by a file or by the CCD camera. A possible optimization consists in using a sliding window to read only a part of the image at a time. This will limit the memory needs but will increase the disc accesses for the off-line acquisition. This solution is thus much more interesting when there is only one image parsing.
3. The acquisition process currently requires three scanner rotations to get a ring image. The acquisition is performed during the first rotation, and the two following rotations are currently required to 1) transfer the image data to the hard disk drive 2) to perform the radial move of the camera for the next ring acquisition and 3) to synchronize the clock of the camera to start only once the computer is ready. These 3 steps could be optimized to reduce the acquisition

process to only two scanner rotations, which will lower the acquisition time by a third.

The first two proposed optimization will be time consuming in terms of development, for a low effective gain. It would be much more easy and efficient to parallelize the processing on several computers when needed. The gain on the acquisition process is much more interesting, as the scanner is one of the two critical resources of the system with the photographic camera. Therefore lowering the acquisition time by a third would speed up the whole processing. The scanning time for a coarse groove 78 rpm record would then drop down to a ratio 1:1, compared to the recorded sound duration.

8.6 Sound quality evaluation

In this section, the sound quality is evaluated in terms of bit depth (Subsection 8.6.1) and signal to noise ratio (Subsection 8.6.2).

8.6.1 Bit depth

The digitizing process does not occur on the sound, but on the image. Therefore it is difficult to define the extracted sound file in term of bit depth. The bit depth of the extracted sound was estimated by saving an extracted sound with variable bit depth, ranging from 1 to 24 bits. This sound is a pure frequency of 300 Hz, amplified by 15 dB and recorded on a microgroove showing 1 trace (2 edges) on the VisualAudio acquisition. The sound extraction has been performed without de-emphasis equalization (flat equalization). The extracted sound does not contain large clicks, which could limit the dynamic and could bias the quality estimation of the lower bit depth files. The sound content has been amplified at the maximum possible level, without saturation, in order to maximize the dynamic. The bit depth reduction was then performed by saving only the higher bits of the file. The SNR_{total} measured on these files are displayed on Table 8.26.

Nb of bits	from 24 to 10	9	8	7	6	5	4	3
SNR	18.41	18.40	18.35	18.17	17.52	15.59	11.73	6.44

Table 8.26: SNR_1 measured on acquisitions with different bit depth.

The SNR , measured independently on each file, show absolutely no variation for all files between 11 to 24 bits. The SNR difference between the 8 and 11 bits files shows a very small variation of 0.06 dB. Decreasing the bit depth below 7 bits results in an important loss of quality.

This shows that 16 or 24 bits files result in the same quality for the current VisualAudio extractions.

8.6.2 Signal to noise ratio evaluation

The extraction quality of the VisualAudio system can be measured with the signal to noise ratio using the standards defined for records and turntables which were presented in Section 2.10. The NAB standard (1949) for monophonic records specifies the measurement conditions as follows: the reference signal is a 1 kHz tone with a peak velocity of 7 cm/sec, the signal is measured on a flat frequency response (constant velocity) and the considered bandwidth ranges from 500 Hz to 10 kHz. The peak amplitude A_{ref} of such reference signal is then:

$$A_{ref} = \frac{7cm/sec}{2 \cdot \pi \cdot 1000Hz} = 11.1\mu m \quad (8.8)$$

The PS record (Popular Science test record presented in Tables 8.1 and 8.2) contains tracks with a 1 kHz tone and a peak velocity of 7 cm/sec. On the acquired images, the peak amplitude of this signal corresponds to 16 pixels with a $4\times$ magnification. Thus we must transform the pixels in micrometers by multiplying by the optical correction factor and by the size of the pixel in micrometers:

$$A_0 = \frac{16}{2} \cdot \frac{2.3}{3} \cdot \frac{10}{4} = 15.33\mu m \quad (8.9)$$

The ratio between A_0 and A_{ref} is of $\sqrt{2}$, which means that the peak velocity that is mentioned for these tracks refers to the speed of the groove's walls displacement, and not to the radial surface speed. This ratio between A_0 and A_{ref} corresponds then to a 3 dB loss and this gain must be subtracted from the measured signal to noise ratio in order to get the value as specified by the standard measurement. Table 8.27 shows the signal to noise ratio measured according to the 1949 NAB standard for monophonic records, including the 3 dB correction.

Acquisition	Correction level	SNR_{total}	SNR_1
PS_track1_OT60	2 iterations	10.37	12.62
PS_track9_OT20	2 iterations	13.54	14.38
PS_track1_OT60	LMS4	15.31	18.72

Table 8.27: Signal to noise ratio measured according to the NAB standard (1949) for monophonic records.

As stated in Subsection 8.1.5, the STD is the other way to measure the noise level. Using the signal to noise equations developed in Subsection 4.3.5, we can check the correctness of the measures presented in Table 8.27. Based on Tables 8.8 and 8.10, the STD is estimated around $0.4 \mu m$, with an approximate blur size of $d_0 = 4$ pixels. This gives a good estimation of the standard deviation σ_n of the noise for one edge extraction. According to Equation 8.6, the blur size $d_0 = 4$ and the sampling frequency $f_s = 85197$ Hz (for a scanning frequency of 65 k-lines) define a cut frequency $f_c = 10650$ Hz, which must be used as the noise bandwidth in Equation 4.33. This equation can then be rewritten as follows:

$$SNR = \frac{3f_{ref}^2 A_{ref}^2 N f_c}{2\sigma_n^2 (f_1^3 - f_0^3)} = \frac{3 \cdot 15.33^2 \cdot 1000^2 \cdot 2 \cdot 10650}{2 \cdot 0.4^2 \cdot (10000^3 - 500^3)} = 16.71dB \quad (8.10)$$

This calculation based on the *STD* confirms the order of magnitude of the signal to noise ratios presented in Table 8.27. The variations between the signal to noise ratios presented in 8.27 and in Equation 8.10 are due to the fact that the noise is not only white, but also presents some low frequency components.

The VisualAudio reaches then a signal to noise ratio of around 16 dB for microgroove 33 rpm records. This signal to noise ratio is increased by 3 dB when extracting sound from coarse groove records, as these present 4 edges on the acquired images. Table 8.28 summarizes the signal to noise ratios obtained with turntables, with the VisualAudio system as well as the signal to noise ratio required by the 1949 NAB standard.

	Frequency band (Hz)	S/N (dB)
Shellac 78 rpm	500-6000	17-37
Acetate / cellulose 78 rpm	500-10000	37-47
NAB standard (1949) mono	500-10000	40
VisualAudio 33 rpm	500-10000	16
VisualAudio 78 rpm	500-10000	19

Table 8.28: Comparison of the signal to noise level obtained with turntables, defined by the NAB standard and obtained with the VisualAudio system.

8.7 Global evaluation

Small variations of the acquisition conditions have a direct influence on the extracted sound quality. Since most of the blur is produced at the photography step and as there are no significant *STD* or *SNR* variations between the top and bottom edges of a groove, it means that most of the white noise is introduced between the picture taking and the scanning steps. The two main noise sources which affect the signal between the picture taking and the scanning steps are the film and the CCD noise. Since the CCD noise is relatively low and showed a low impact on the extracted sound, this means that the main noise source is the photographic film, which has important graininess as well as pepper fog. This is confirmed by the tests presented in this work: we changed many acquisition and image processing parameters and compared the results. Among all the tests performed, the film tests (see Section 3.1) show the larger *STD* variations, which means that the film is a very important source of noise in this system. The effect of the film graininess on the final sound could be lowered by working with magnified photographs. Working with a magnification ratio of 2:1 at the photography step would increase the signal to noise ratio by 3 dB; but it would also require to increase the size of the photographic camera from 20 cm in height (which is already of 2,2 meters) and to work with a glass tray of twice the current size, which would increase the optical blur at scanning. A high quality glass or an autofocus would then be necessary at scanning. Moreover, magnification of the record image will also pose some problem for the illumination of the inner part of the record, due to the position of the lens in the middle of the illumination system of the camera. Therefore the magnification does not appear as a satisfying solution.

Our tests show that the current resolution is optimal, and that some small variations occur, due to the glass tray warping or to the radial position of the groove on the record. The use of an autofocus system could lower these variations, increase the resolution locally and produce even more homogeneous images. These tests also point out the limit of the oversampling process: more oversampling does not increase the resolution, but it only spreads the existing information on more pixels.

Based on this evaluation, we can say that the VisualAudio sound extraction results in lower sound quality than a standard turntable playback, but it allows reading some records that are currently considered as unplayable on turntables.

Chapter 9

Conclusion, perspectives and discussion

This thesis focuses on the image analysis and processing parts of the sound extraction process called VisualAudio. The idea of the VisualAudio process is to take pictures of phonographic records in order to keep the sound on a long term analog storage medium: a photographic film. When needed, the record picture is scanned to extract the sound from the picture by image processing, using a dedicated scanner and algorithms.

The contributions of this work are summarized in Section 9.1. Perspectives and further developments are proposed in Section 9.2. Based on the analysis and evaluations developed in this thesis, the VisualAudio project is then discussed in Section 9.3.

9.1 Contributions

The VisualAudio system is a new and original approach to archive and read phonographic records. Thus the first contribution of this thesis was to provide an in-depth analysis of the various components and different technologies used in this system. The first two chapters presented the phonographic record technology and the VisualAudio acquisition system, in order to understand the key elements that may affect a record picture. Based on these descriptions, a thorough analysis of the imaging chain was presented: this analysis pointed out the different kinds of degradations that occur in the VisualAudio process: blur, noise and local degradations, illumination variations and nonlinear distortions due to acquisition artifacts. A model was then proposed to understand how these degradations affect the groove image.

The model and the imaging chain analysis were used to develop specific algorithms to extract the groove from the acquired image, to extract the sound and to restore the signal, which may be locally damaged by spurious patterns.

We have evaluated the groove extraction process and the VisualAudio system by measuring the noise level obtained on the extracted sound and by comparing it when varying several acquisition and processing parameters. This evaluation has

confirmed the initial resolution analysis. We also have showed that it is possible to extract the groove edges with submicron accuracy (see Tables 8.8 and 8.10), even on noisy blurred images. This evaluation has also pointed out some of the system limitations, such as the lack of oversampling gain on sound extracted from blurred images and the important noise level produced by the film graininess and artifacts (mainly the pepper fog).

Using these image processing methods, we demonstrated that it is possible to extract reasonable quality sound from various phonographic record pictures, including 33 rpm and 78 rpm, direct cut and pressed records. We also extracted the sound from some severely damaged records with mould or shrinkage of the lacquer coating. Such records are currently considered as unreadable with any existing turntable and any other optical reading method to the best of our knowledge.

Thus we hope that this system can be used to restore sounds from many records to save our sound heritage.

9.2 Perspectives

Even if the VisualAudio system is already in use to extract sound from records provided by different sound archives, there are still some ways for this project to evolve. This section proposes some direction for further research and enhancements of the VisualAudio system.

9.2.1 Frequency and time domain detection and correction

Sound correction by image processing is limited due to the resolution of the system: it is very difficult to detect and to correct the acquired images at submicron accuracy, especially when we rely on noisy blurred images.

There are many kinds of noise and local damages. Due to the nature of the acquired images, these degradations do often not appear as specific patterns, but are visible as damages on the traces (either cuts or trace spreads). Thus it is difficult to use the knowledge of the degradation to correct them, and it is much more natural to use the trace information to detect and correct the trace damages. The main information on the trace displacement is provided by the recorded signal: thus some more accurate corrections are achievable with the signal information, which has been extracted from the acquired image, using frequency based models such as ARMA, LSAR or Warped linear prediction for example ([13, 12, 111]).

The best restoration procedure for VisualAudio would be to perform a first detection on the image level, and thus to locate the degradations which are clearly visible on the image. This also ensures that the edge detection and groove reconstruction processes won't be misled by spurious patterns. An error map similar to the map used in Subsection 6.3.5 must then be constructed and transmitted to the frequency domain correction system. A second stage detection and correction must be performed on the frequency and time domain, working only with the samples which have not been defined as corrupted during the image processing.

9.2.2 Multichannel correction

The frequency domain correction exposed in Subsection 9.2.1 can be extended to a multichannel scheme. A multichannel approach could be very efficient to detect the groove degradations, using the two or four edges of a groove as different channels. A single channel model could then be used for an accurate signal reconstruction, based on a single edge. Working with a single edge at a time, the reconstruction would also encompass the edge symmetric errors (such as presented in Subsections 5.2.5 and 8.2.6.3) and these errors will then be correctly removed by the combination of the different edges. Working with a multichannel model for the signal reconstruction would remove part of the symmetrical errors on the reconstructed areas. Thus these symmetrical errors would not be fully canceled by the combination of the different edges.

9.2.3 Discs with shrinkage of the recording layer

Section 6.4 introduced some ideas for the processing of discs with shrinkage of the recording layer. Some shrunk records can already be processed, but larger cracks are currently not handled. The development of image processing for such discs is the next step in the VisualAudio project, and it will provide a very useful and appreciated tool, as these discs are currently considered as lost.

9.2.4 Film

The photographic films must be further studied. The film graininess is currently high and has an important impact on the extracted sound.

The pepper fog (see Section 4.3.3) may be produced by many different causes. This problem could be solved by changing some parameters in the photographic process, for example the products and the mix used for the development [20].

9.2.5 Direct reading

As discussed later in Subsection 9.3.3, some comparison must be performed with a direct optical playback system, which directly scans the record without the photographic step.

The main problem for a direct reading system would be to dynamically adjust the autofocus for the direct digitizing of the records, such as the Predictive Focus Tracking System which maintains continuous focus on a moving subject [119]. The use of the autofocus must be especially studied in the case of the records with shrinkage of the recording layer, where the recording layer lifts locally and may mislead the dynamic autofocus.

9.3 Discussion on the VisualAudio project

The objectives of the VisualAudio project are to use a long term storage medium for old records and to provide a contactless reading method, which can be used to playback all kinds of records, even severely damaged and shrunk records. Since the groove modulations are visible on the record surface, the VisualAudio system proposes to take a photograph of a disc, in order to keep an analog copy of the recorded sound. The photograph can later be playback using a dedicated scanner.

The main advantages of the VisualAudio system are:

- The record is preserved with an analog copy (the film).
- The analog copy can be stored for long term archiving.
- No need to use an autofocus system as the depth of field of the photography is much larger than the depth of field of the microscope lens.
- The archiving time (time to take the photograph) is low and thus well adapted for mass archiving.
- The archiving time is considerably lowered for damaged records.
- The optical playback system is not intrusive and does not damage records.
- The optical playback system is able to read records that are currently considered as unreadable.

These advantages mix two aspects of the VisualAudio system: the archiving system and the capacity of the optical playback system to read damaged records. These two aspects must then be considered separately. The common point that connects these two aspects is the photographic film, which is used as the analog sound storage medium and for the optical playback. Thus the above mentioned advantages are discussed in the next three subsections which are about the film, the archiving and the optical playback.

9.3.1 Photographic film

One of the main arguments to store an analog copy of the records is that the analog to digital transfer is lossy and that some people pretend that even a 24 bits/98 kHz digital file does not always contain the complete original recorded sound information. The evaluation of the VisualAudio system showed that the bit depth of the recorded sound is around 10 bits and that the blur produces a low-pass filter on the sound which attenuates the high frequencies starting around 10 kHz (see Section 8.6). Thus even if we may consider the photographic film as analog, it is of lower quality than a normal digital file.

Many different films have been tested, and we couldn't find films with finer graininess up to now. The graininess and film artifacts appear as an important problem, since they affect the storage medium used for archiving. Moreover it seems

difficult to drastically lower the impact of the film graininess on the sound, as the film grain size is between 0.2 and 3 μm (see 3.1.2) and the desired resolution (which was computed in Section 4.3.5.4) showed the need for noise structure finer than 0.1 μm . Since the photographic film market is strongly decreasing, the research and development efforts are currently very low to get new products with finer graininess.

9.3.2 Archiving

Processing time is also an important factor for mass archiving. If the VisualAudio system is intended to be used as a transfer system to a digital file, then the picture of the record must be taken and the film must be scanned. As presented in Section 8.5, the scanning time is currently comparable to the recorded sound duration and thus to the turntable playback time for records in a good stage of conservation. But in addition to the scanning time, the complete VisualAudio sound extraction duration encompasses also the picture taking time, the image processing time, as well as the handling of the various devices. We won't go into details in the whole processing time evaluation, as it overtakes the scope of this work; but based on these observations, we estimate that the VisualAudio sound extraction is slower for records in good state of conservation than standard turntable playback. But the turntable playback time is increased when the operator must study each record independently to select the correct reading stylus and speed settings. The turntable playback time even increases exponentially for records with severe damages (moulds or shrinkage of the recording layer for example), since the operator must read each piece of the groove independently, which may require more than 20 hours to extract 3 minutes of sound [2]. The VisualAudio photography and scanning times are constant for any record, in any stage of degradation, except the camera exposure time, which may vary from 15 to 60 seconds and is thus relatively short compared to the whole process. Thus the VisualAudio system provides an interesting solution in terms of archiving time for such records.

If the VisualAudio is used as an archiving system, then we only take pictures of the records for storage. Thus the archiving (or photography) time is constant: it is then two times faster than any digitizing system for records in good stage of conservation, and many times faster for damaged records. But the sound file is not directly accessible and the time to extract the sound may still be important. Moreover, there is currently only one way to evaluate the quality of the film and of the stored sound: by scanning the film.

As explained in the evaluation (see Section 8.7), the current VisualAudio sound quality is currently lower than a turntable playback. Therefore the proposed archiving system is mainly useful for records which are currently considered as unreadable or very difficult to read with a standard turntable. This means that this archiving system is devoted to only a part of the direct cut records that represent only a part of all the phonographic records. Moreover, one of the main dangers for the audiovisual carrier archiving cited in the UNESCO survey is the obsolescence of the reading equipment [1]. Thus it appears difficult to introduce a new storage medium on the market, which requires a new playback system, to archive only a limited

amount of records. The transfer of the optically extracted sounds in digital files seems much more efficient for future uses, as it doesn't imply to keep a scanning device in working order for several tens of years.

Acetate records may be destroyed by classical turntable playback, but they may be subject to shrinkage after the change of ambient conditions when they are taken out their storage room and sleeves. Thus, is it worth to take the risk to get some records shrunk and to keep then only a lower quality copy? In such case, the best solution would be to make a photo of the record, to get a safety copy, and then to extract sound with a turntable. If the turntable playback damages the record, then we still get the safety photograph copy. But such alternative is costly, and it is already difficult to raise funds for archiving and storage of the sound heritage.

9.3.3 Optical playback

The optical playback shows interesting capabilities to extract sound from degraded records. Unfortunately, the optical extraction from the photograph of a record leads to a low quality compared to a turntable playback. We showed that the photographic step is the main cause of blur, due to the optical blur, and the main cause of noise, due to the film graininess and other film defects such as the pepper fog.

As presented in Subsection 9.3.2, the archiving component of the VisualAudio system is debatable. Thus as the archiving is one of the main reasons to work with the photographic film, and as the photographic step is obviously the limiting stage of the VisualAudio system, the photography must be reconsidered. We haven't tested an autofocus system to directly read records up to now (such as the system proposed by Haber and Fedeyev [35]). Such a direct reading system would eliminate the source of most of the noise and of the blur in the digitized record images. We would propose to compare the VisualAudio system with a direct reading system to evaluate whether the film is suitable enough to be used as an intermediate sound storage medium, and whether the autofocus is able to handle any kind of records, even severely damaged or shrunk records. If the superiority of the direct reading is shown, we may follow the proposition of R. Lagendijk et al. concerning noisy blurred images: "Probably the best approach to the restoration of noisy blurred images would be to prevent the degradations from occurring at all" ([120]) and bypass the photographic step in the VisualAudio process. Thus by direct scanning, we could get better record images and extract better sound quality from the damaged records.

Bibliography

- [1] G. Boston. *Survey of Endangered Audiovisual Carriers, 2003*. UNESCO, Paris, 2003.
- [2] K. Bradley, editor. *Guidelines on the Production and Preservation of Digital Audio Objects*. IASA TC-04, 2004.
- [3] J. R. Stuart. A Proposal for the High-Quality Audio Application of High-Density CD Carriers. *Technical Subcommittee Acoustic Renaissance for Audio*, pages 1–26, June 1995.
- [4] S. Cavaglieri, O. Johnsen, and F. Bapst. Optical Retrieval and Storage of Analog Sound Recordings. In *AES 20th International Conference, Budapest*, 2001.
- [5] S. Stotzer, O. Johnsen, F. Bapst, C. Sudan, and R. Ingold. Phonographic Sound Extraction Using Image and Signal Processing. *Proceedings IEEE ICASSP*, May 2004.
- [6] S. Stotzer, O. Johnsen, F. Bapst, and R. Ingold. Groove Extraction of Phonographic Records. In H. Bunke and A. L. Spitz, editors, *Document Analysis Systems*, volume 3872 of *Lecture Notes in Computer Science*, pages 529–540. Springer, 2006.
- [7] N. Bosi. *VisualAudio : Acquisition et numérisation des images*. Travail de diplôme, Ecole d'ingénieurs et d'architectes de Fribourg, 2000.
- [8] O. Chassot and C. Miauton. *Reconstitution de la musique à partir de l'image numérique*. Travail de diplôme, Ecole d'ingénieurs et d'architectes de Fribourg, 2000.
- [9] Y. Pauchard and C. Sudan. *Die Technik um alte Schallplatten auf Filmnegativ zu archivieren und später ab Film zu hören*. Travail de diplôme, Ecole d'ingénieurs et d'architectes de Fribourg, 2001.
- [10] C. Milan. *Rapport d'activités: Optique - Mécanique - Electronique*. Rapport interne, Ecole d'ingénieurs et d'architectes de Fribourg, 2005.
- [11] O. Cappé. *Technique de réduction du bruit pour la restauration d'enregistrements musicaux*. PhD thesis, Ecole Nationale Supérieure des Télécommunications, Paris, 1993.

- [12] S.V. Vaseghi. *Advanced Digital Signal Processing and Noise Reduction*. Wiley, New York, 1996.
- [13] S. J. Godsill and P. J. W. Rayner. *Digital Audio Restoration*. Springer-Verlag London Limited, 1998.
- [14] P. Esquef. *Model-Based Analysis of Noisy Musical Recordings with Application to Audio Restoration*. PhD thesis, Helsinki University of Technology, April 2004.
- [15] CEDAR ToolsTM. <http://www.cedar-audio.com/>, last accessed 01.06.2006.
- [16] Weiss DNA1. <http://www.weiss.ch/>, last accessed 01.06.2006.
- [17] M. Crawford. Pioneer Experiments of Eugène Lauste in Recording Sound. *Journal of the Society of Motion Picture Engineers*, 17(4), October 1931.
- [18] E.W. Kellogg. History of Sound Motion Pictures. *Journal of the Society of Motion Picture Engineers*, 64, 1955.
- [19] E. I. Sponable. Historical Development of Sound Films. *Journal of the Society of Motion Picture Engineers*, 48(4), April 1947.
- [20] Theatrical Sound. <http://www.kodak.com/country/US/en/motion/support/h1/soundP.shtml>, last accessed 25.06.2006.
- [21] J. Valenzuela. Digital Audio Image Restoration: Introducing a New Approach to the Reproduction and Restoration of Analog Optical Soundtracks for Motion Picture Film. *IBC*, 2003.
- [22] T. Wittman. Lost in the Supermarket: Decoding Blurry Barcodes. *SIAM News*, 37(7), September 2004.
- [23] S. Esedoglu. Blind Deconvolution of Barcode Signals. *Inverse Problems*, 20:121–135, 2004.
- [24] E. Joseph and T. Pavlidis. Bar Code Waveform Recognition Using Peak Locations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(6):630–640, 1994.
- [25] J-C. Di Martino and S. Tabbone. An Approach to detect Lofar Lines. *Pattern Recognition Letters*, 17:37–46, 1996.
- [26] J. Poliak, P. Robert, and J. Goy. Optical Fibre Turntable for Archive Records. *Proceedings of the 92nd Convention AES, Vienna, Austria*, 1992.
- [27] V. V. Petrov, A. A. Kryuchin, S. M. Shanoylo, I. P. Ryabokon, and O. S. Onyshchenko. Optomechanical Method of Edison Cylinders Sound Reproduction. *Audio Eng. Soc. Preprint (Preprint 4491)*, Munich, March 1997.

- [28] W. Penn and M. J. Hanson. The Syracuse University Library Radius Project: Development of a non-destructive Playback System for Cylinder Recordings. *First Monday*, 8(5), May 2003.
- [29] ELP Laser Turntable. <http://www.laserturntable.com/>, last accessed 25.06.2006.
- [30] O. Springer. Digital Needle - A Virtual Gramophone, 2002. <http://www.cs.huji.ac.il/springer/>, last accessed 25.06.2006.
- [31] U. Kalla, N. Jalden, N. Lithhammer, M. Eriksson, and E. Perez. *The Digital Needle Project - Group Light Green*. Semester project, KTH Royal Institute of Technology, Stockholm, Sweden, 2003. <http://www.s3.kth.se/signal/edu/projekt/students/03/lightgreen/>, last accessed 25.06.2006.
- [32] P. Olsson, D. Öhlin, R. Olofsson, R. Vaerlien, and C. Ayrault. *The Digital Needle Project - Group Light Blue*. Semester project, KTH Royal Institute of Technology, Stockholm, Sweden, 2003. www.s3.kth.se/signal/edu/projekt/students/03/lightblue/, last accessed 25.06.2006.
- [33] M. McCann, P. Calamia, and N. Ailon. *Audio Extraction from Optical Scans of Records*. Final project, Princeton University, 2004. <http://saturn.vfx.com/mccann/visionfinal/doc/>, last accessed 25.06.2006.
- [34] PrestoSpace Project. <http://www.prestospace.org>, last accessed 01.06.2006.
- [35] V. Fadeyev and C. Haber. Reconstruction of Mechanically Recorded Sound by Image Processing. *J. Audio Eng. Soc.*, 51(12):1172–1185, December 2003.
- [36] 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(6):485–508, June 2005.
- [37] N. Lutz and M. Yerly. *Studies of Mechanical Recording Media with 3D Surface Profiling Methods Data Collection and Analysis*. Travail de diplôme, Ecole d'ingénieurs et d'architectes de Fribourg, 2005.
- [38] F. Langford-Smith. *Radiotron Designer's Handbook*. Wireless Press, 4th edition, 1953.
- [39] M. Rossi. *Traité d'Electricité, Vol XXI: électroacoustique*. Presses Polytechniques Romandes, 1986.
- [40] S. E. Schoenherr. *Recording Technology History*, 2005.
- [41] A History of Vinyl. <http://www.bbc.co.uk/music/features/vinyl/>, last accessed 25.06.2006.

- [42] T. Inoue, N. Takahashi, and I. Owaki. A Discrete Four-Channel Disc and its Reproducing System (CD-4 System). *J. Audio Eng. Soc.*, 19(7):576–583, July/August 1971.
- [43] A. G. Webster. Acoustical Impedance, and the Theory of Horns and of the Phonograph. *PNAS*, 5:275–282, 1919.
- [44] E. T. Canby. *Saturday Review Home Book of Recorded Music and Sound Reproduction*. Prentice - Hall, Inc. New York, 1952.
- [45] A. M. Max. Record Quality and Its Relation to Manufacturing. *J. Audio Eng. Soc.*, 3(1):19–25, January 1955.
- [46] G. St-Laurent. The Care and Handling of Recorded Sound Materials. *Music Division National Library Of Canada*, January 1996.
- [47] J. C. Ruda. Record Manufacturing: Making the Sound for Everyone. *J. Audio Eng. Soc.*, 25(10/11):702–711, October/November 1977.
- [48] W. R. Isom. Before the fine Groove and Stereo Record and other Innovations. *J. Audio Eng. Soc.*, 25(10/11):815–820, October/November 1977.
- [49] Vintage 78's - Getting in the Groove, 2006. <http://www.videointerchange.com/vintage.78s.htm>, last accessed 25.06.2006.
- [50] T. Martini, R. S. Firestone, A.G. Wilmore, and J. Miller. Last Word Archive: A different Spin. *New Scientist*, (2186), 1999.
- [51] J. P. Maxfield and H. C. Harrison. Methods of High Quality Recording and Reproduction of Music and Speech Based on Telephone Research. *The Bell System Technical Journal*, V(3):493–523, June 1926.
- [52] J. G. Woodward and E. C. Fox. A Study of Program-Level Overloading in Phonograph Recording. *J. Audio Eng. Soc.*, 11(1):16–23, January 1963.
- [53] R. Wilmut. Reproduction of 78 rpm records, Technical notes. <http://www.rfwilmut.clara.net/repro78/repro.html>, last accessed 25.06.2006.
- [54] ELBERG MD12, Multicurve Disc preamp 78 RPM and RIAA Record Preamplifier. http://www.vadlyd.dk/English/RIAA_and_78_RPM_preamp.html, last accessed 25.06.2006.
- [55] *Phonotechnik, DIN-Taschenbuch 523*. Deutsch Institut für Normung, Beuth Verlag, 1991.
- [56] A. Wright. The Tube Preamp Cookbook. *Vacuum State Electronics*, 1995.
- [57] Dimensional Standards Disc Phonograph Records for Home Use. *Record Industry Association of America, Inc. Bulletin No. E 4*, 1963.

- [58] J. J. Bubbers. A Report on the Proposed NAB Disc Playback Standard. *J. Audio Eng. Soc.*, 12(1):51–54, January 1964.
- [59] W. R. Isom. Record Materials, Part II: Evolution of the Disc Talking Machine. *J. Audio Eng. Soc.*, 25(10/11):718–723, October/November 1977.
- [60] A. C. Keller. Early Hi-Fi and Stereo Recording at Bell Laboratories (1931–1932). *J. Audio Eng. Soc.*, 29(4):273–280, April 1981.
- [61] T. H. James, editor. *The Theory of the Photographic Process*. New York, Macmillan, 4th edition, 1977.
- [62] C. N. Proudfoot, editor. *Handbook of Photographic Science and Engineering*. Wiley, New York, 2nd edition, 1973.
- [63] P. Z. Adelstein. ISO 18911 Imaging Materials - Processed Safety Photographic Films - Storage Practices. *Geneva: International Organization for Standardization*, 2000.
- [64] P. Z. Adelstein. IPI Media Storage Quick Reference. *Image Permanence Institute, Rochester Institute of Technology (USA)*, 2004.
- [65] Piranha CT-P1, CL-P1, Camera User’s Manual, C3-32-00253, rev. 10, 1999.
- [66] J. Elder. *The Visual Computation of Bounding Contours*. PhD thesis, Mc Gill University, Dept. of Electrical Engineering, 1995.
- [67] M. Born and E. Wolf. *Principles of Optics*. Oxford Pergamon Press, 6th edition, 1980.
- [68] Ctein. *Post Exposure: advanced Techniques for the photographic Printer*. Focal Press, 1997.
- [69] J. Ye, G. Fu, and U. P. Poudel. High-Accuracy Edge Detection with blurred Edge Model. *Image and Vision Computing*, 23(5):453–467, May 2005.
- [70] W. K. Pratt. *Digital Image Processing*. Wiley, New York, 1978.
- [71] M. Baba and K. Othani. A novel subpixel Edge Detection System for Dimension Measurement and Object Localization Using an analogue-based Approach. *Journal of Optics A: Pure and Applied Optics*, 3(4):276–283, 2001.
- [72] C. Steger. Evaluation of Subpixel Line and Edge Detection Precision and Accuracy. *International Archives of Photogrammetry and Remote Sensing*, XXXII:256–264, 1998. Part 3/1.
- [73] A. C. Rencher. *Linear Models in Statistics*. Wiley, New York, 2000.
- [74] A. J. Tabatabai and O. R. Mitchell. Edge Location to Subpixel Values in Digital Imagery. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:188–201, March 1984.

- [75] Y. Shan and G. W. Boon. Sub-pixel Location of Edges with Non-uniform Blurring: a finite Closed-form Approach. *Image and Vision Computing*, 18:1015–1023, 2000.
- [76] M. H. Hueckel. An Operator which Locates Edges in Digitized Pictures. *Journal of the ACM*, 18:113–125, January 1971.
- [77] F. Pedersini, A. Sarti, and S. Tubaro. Estimation and Compensation of Sub-pixel Edge Localization Error. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(11):1278–1284, November 1997.
- [78] F. Truchetet, F. Nicolier, and O. Laligant. Subpixel Edge Detection for Dimensional Control by Artificial Vision. *Journal of Electronic Imaging*, 10(1):234–239, January 2001.
- [79] C. Steger. Analytical and Empirical Performance Evaluation of Subpixel Line and Edge Detection. *Empirical Evaluation Methods in Computer Vision*, pages 188–210, 1998.
- [80] A. Rosenfeld and M. Thurston. Edge and Curve Detection for visual Scene Analysis. *IEEE Transactions on Computers*, 20(5):562–569, May 1971.
- [81] D. Marr and E. Hildreth. Theory of Edge Detection. *Proceedings of the Royal Society of London, Series B, Biological Sciences*, pages 187–217, February 1980.
- [82] S. Mallat and S. Zhong. Characterization of Signals from Multiscale Edges. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(7), July 1992.
- [83] T. Lindberg. *Scale Space Theory in Computer Vision*. Kluwer Academic Publishers, 1994.
- [84] C. Xu and J. L. Prince. Snakes, Shapes, and Gradient Vector Flow. *IEEE Transactions on Image Processing*, 7(3):359–369, March 1998.
- [85] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active Contour Models. *International Journal of Computer Vision*, 1:321331, 1987.
- [86] S. Venkatesh, M. Kisworo, and G. A. W. West. Detection of Curved edges at Subpixel Accuracy Using Deformable Models. *IEE Proceedings - Vision, Image and Signal Processing*, 142(5):304312, October 1995.
- [87] R. Durikovic, K. Kaneda, and H. Yamashita. Dynamic Contour: A Texture Approach and Contour Operations. *The Visual Computer*, 11:277289, 1995.
- [88] D. Terzopoulos and R. Szeliski. *Tracking with Kalman Snakes*. in Active Vision, Cambridge, MA: MIT Press, 1992.
- [89] P. J. Huber. *Robust Statistics*. Wiley, New York, 1981.

- [90] E. E. Kuruoglu, P. J. W. Rayner, and W. J. Fitzgerald. Impulsive Noise Elimination Using Polynomial Iteratively Reweighted Least Squares. *IEEE Digital Signal Processing Workshop Proceedings*, 13(5):347–350, September 1996.
- [91] J. P. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and M. B. Sandler. A Tutorial on Onset Detection in Music Signals. *IEEE Transactions on Speech and Audio Processing*, 13(5):1035–1047, September 2005.
- [92] X. Rodet and F. Jaillet. Detection and Modeling of Fast Attack Transients. *in Proceedings of the International Computer Music Conference*, 2001.
- [93] F. M. Wahl. *Digital Image Processing*. Artech House, Boston, 1987.
- [94] M. Kunt et al. *Traitement numériques des images*. Presses Polytechniques romandes, 1993.
- [95] G. Healey and R. Kondepudy. Radiometric CCD Camera Calibration and Noise Estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(3):267–276, March 1994.
- [96] S. Gremaud. *Centrage du film pour VisualAudio*. Travail de semestre, Ecole d’ingénieurs et d’architectes de Fribourg, 2003.
- [97] N. S. Kopeika. *A System Engineering Approach to Imaging*. SPIE Optical Engineering Press, 1998.
- [98] J. Biemond, R. L. Lagendijk, and R. M. Mersereau. Iterative Methods for Image Deblurring. *Proc IEEE*, 78(5):856–883, May 1990.
- [99] D. Kundur and D. Hatzinakos. Blind Image Deconvolution. *IEEE Signal Processing Magazine*, 13(3):43–64, May 1996.
- [100] E. Abreu, M. Lightstone, S.K. Mitra, and K. Arakawa. A New Efficient Approach for the Removal of Impulse Noise from Highly Corrupted Images. *ACM Multimedia*, 5(6):1012–1025, 1996.
- [101] P. S. Windyga. Fast Impulsive Noise Removal. *IEEE Transactions on Image Processing*, 10(1):173–179, January 2001.
- [102] W. K. Pratt, T. J. Cooper, and I. Kabir. Pseudomedian Filter. *Proc SPIE*, 534:34–43, 1985.
- [103] R. Bornard, E. Lecan, L. Laborelli, and J. H. Chenot. Missing Data Correction in Still Images and Image Sequences. *ACM Multimedia*, pages 355–361, 2002.
- [104] A. N. Hirani and T. Totsuka. Dual Domain Interactive Image Restoration: Basic Algorithm. *Proceedings of the IEEE International Conference on Image Processing*, 1:797–800, 1996.

- [105] O. G. Guleryuz. Iterated Denoising for Image Recovery. *IEEE Data Compression Conference*, 12(1):3–12, April 2002.
- [106] M. Belge and E. L. Miller. Wavelet Domain Image Restoration Using Edge Preserving Prior Models. *Proceedings of the IEEE International Conference on Image Processing*, 2:103–107, 1998.
- [107] X. Rodet. Musical Sound Signals Analysis/Synthesis: Sinusoidal+Residual and Elementary Waveform Models. *IEEE UK Symposium on applications of Time-Frequency and Time-Scale methods*, pages 111–120, August 1997.
- [108] P. Esquef, K. Roth, and I. Kauppinen. Interpolation of Long Gaps in Audio Signals Using Warped Burg’s Method. *Proceedings of the 6th Int. Conference on Digital Audio Effects (DAFx-03), London, UK*, pages 18–23, September 2003.
- [109] R. Veldhuis. *Restoration of Lost Samples in Digital Signals*. Prentice-Hall, Inc., Upper Saddle River, NJ, 1992.
- [110] P. Esquef, L. W. P. Biscainho, P. S. R. Diniz, and F. P. Freeland. A Double Threshold-Based Approach to Impulsive Noise Detection in Audio Signals. *Proc. X European Signal Processing Conf. (EUSIPCO 2000), Tampere, Finland*, pages 2041–2044, September 2000.
- [111] P. Esquef and M. Karjalainen. Detection of Clicks in Audio Signals Using Warped Linear Prediction. *Proceedings DSP 2002*, 2:1085–1088, 2002.
- [112] P. A. Lynn and W. F. Fuerst. *Introductory Digital Signal Processing with Computer Applications*. Chichester, UK: John Wiley & Sons, 1994.
- [113] M. D. Lutovac, D. V. Tošić, and B. L. Evans. *Filter Design for Signal Processing Using MATLAB and Mathematica*. Prentice Hall Upper Saddle River, New Jersey, 2001.
- [114] D. J. Benson. *A Mathematical Offering*. Cambridge University Press, September 2006.
- [115] T. Oohashi, E. Nishina, M. Honda, Y. Yonekura, Yoshitaka Fuwamoto, N. Kawai, T. Maekawa, S. Nakamura, H. Fukuyama, and H. Shibasaki. Inaudible High-Frequency Sounds Affect Brain Activity: Hypersonic Effect. *The Journal of Neurophysiology*, 83(6):3548–3558, June 2000.
- [116] M. L. Lenhardt, R. Skellett, P. Wang, and A. M. Clarke. Human Ultrasonic Speech Perception. *Science*, 253:82–85, July 1991.
- [117] M. Kunt. *Traité d’Electricité, Vol XX: Traitement numérique des signaux*. Presses Polytechniques Romandes, 1984.
- [118] R. E. Crochiere and L. R. Rabiner. *Multirate Digital Signal Processing*. Prentice-Hall, Englewood-Cliffs, 1983.

-
- [119] G. Surya and M. Subbarao. Continuous Focusing of Moving Objects Using Image Defocus. *Proceedings of SPIE's International Symposium, Machine Vision Applications, Architectures, and Systems Integration III*, 2347, July 1994.
- [120] R. L. Lagendijk, J. B. Biemond, and D.E.Boeke. Identification and Restoration of Noisy Blurred Images Using the Expectation-Maximization Algorithm. *IEEE Transactions on Acoustics Speech and Signal Processing*, 38(7), July 1990.

Curriculum vitae

I was born on January 18th, 1972 in Bern Switzerland. After seven years of primary and secondary school at the Ecole cantonale de Langue Française in Bern, I came to Fribourg, where I attended the Cycle d'Orientation du Belluard for two years. Then I attended the Collège Ste-Croix during four years and obtained the federal maturity type C (scientific) in 1991.

From 1991 to 1995 I studied at the University of Fribourg, where I graduated in Computer Science, with Mathematics as minor subject.

In 1996, I went to Canada to work as a research assistant at Laval University in Québec. Since 1998 I worked four years as a software engineer, developing banking and document management systems.

Since 2002 I have worked as a research assistant at both the Ecole d'ingénieurs et d'architectes de Fribourg and the University of Fribourg, where I wrote this PhD thesis.