

Lucas Conejero Takejame

Extração de áudio de discos de vinil por processamento de imagem

Brasil

2015

Lucas Conejero Takejame

Extração de áudio de discos de vinil por processamento de imagem

Texto apresentado à Escola Politécnica da
Universidade de São Paulo como requisito
para a conclusão do curso de graduação em
Engenharia Elétrica com ênfase em Compu-
tação, junto ao Departamento de Engenharia
de Computação e Sistemas Digitais (PCS)

Escola Politécnica da Universidade de São Paulo – EPUSP

Engenharia elétrica com ênfase em Computação

Graduação

Orientador: Edson Satoshi Gomi

Coorientador: Bruno Carvalho Albertini

Brasil

2015

Lucas Conejero Takejame

Extração de áudio de discos de vinil
por processamento de imagem/ Lucas Conejero Takejame. – Brasil, 2015-
57 p. : il. (algumas color.) ; 30 cm.

Orientador: Edson Satoshi Gomi

Trabalho de conclusão de curso (Graduação) – Escola Politécnica da Universidade de São Paulo –
EPUSP
Engenharia elétrica com ênfase em Computação
Graduação, 2015.

1. Digitalização. 2. Processamento de imagem. I. Orientador Edson Satoshi Gomi. II. Uni-
versidade de São Paulo. III. Escola Politécnica. IV. Extração do áudio por processamento de
imagem.

Lucas Conejero Takejame

Extração de áudio de discos de vinil por processamento de imagem

Texto apresentado à Escola Politécnica da
Universidade de São Paulo como requisito
para a conclusão do curso de graduação em
Engenharia Elétrica com ênfase em Compu-
tação, junto ao Departamento de Engenharia
de Computação e Sistemas Digitais (PCS)

Edson Satoshi Gomi
Orientador

Bruno Carvalho Albertini
Coorientador

Brasil
2015

Este trabalho é dedicado aos meus pais, que possibilitaram a conclusão desse curso.

Agradecimentos

Meus agradecimentos ao meu orientador Edson Satoshi Gomi e meu coorientador Bruno Carvalho Albertini pelo direcionamento ao longo do projeto e ao professor Carl Haber por ter fornecido as imagens necessárias para o teste do projeto.

Resumo

A mídia fonográfica era a única maneira de armazenamento de sons até a introdução da fita magnética no começo dos anos 1950. Há, portanto, uma vasta quantidade de discos gravados ainda hoje, indo de discos produzidos em larga escala a discos gravados em corte direto, que são cópias únicas. Muitos destes discos estão em estado delicado de deterioração onde ler o disco com uma agulha de toca-discos poderia danificar irreversivelmente o material.

Tendo em mente preservar o material armazenado nessas mídias, surge a necessidade de realizar a digitalização desses meios. Existem diversas técnicas de digitalização que podem ser aplicadas às mídias fonográficas, porém, dadas as condições em que se encontram muitos discos, seria necessário um método que não tivesse contato mecânico com o disco, para evitar maiores desgastes nas informações. O objetivo deste trabalho é construir um software que implemente um método de extração por processamento de imagem e o resultado obtido foi o áudio contido nas imagens do anel de um disco de vinil mono 78 RPM.

Keywords: Discos de vinil. digitalização. processamento de imagem.

Abstract

Phonographic media was the only way of keeping sounds until magnetic tape was born, in the beginning of 1950's. Therefore, there are a lot of records until today, going from large scale production records until single direct-cut copies. A lot of these records are in advanced state of deterioration, up to a level that even playing in a high quality turntable would damage the content irreversibly.

To save this material, comes the need of digitizing those medias. There are many ways of digitizing phonographic media, however, given the delicate conditions in which a lot of records find themselves, it would be necessary a method that didn't involve mechanical contact, so it wouldn't cause any damages in the content. The goal of this work is to build a software piece that implements an audio extraction via image processing method. The produced result detailed in this document is a 78 RPM mono vinyl's audio, obtained from 8 images of it's ring.

Keywords: Vinyl discs, digitization, image processing.

Lista de ilustrações

Figura 1 – Foto microscópica dos segmentos da trilha de um vinil mono. Fonte: [12]	21
Figura 2 – Foto microscópica do segmentos da trilha de um vinil estéreo. Fonte: [13]	21
Figura 3 – Ilustração das direções de modulação em cada caso de vinil. Fonte:[14]	22
Figura 4 – Desenho do perfil do sulco de nosso vinil. Fonte: [5](figura 4.1).	26
Figura 5 – Segmento de trilha de uma das imagens recebidas no projeto. Marcado em vermelho está um traço do segmento e delineado em verde está uma borda do traço. Fonte: Elaborado pelo autor.	27
Figura 6 – Fragmento de disco desconhecido. Fonte: Carl Haber.	30
Figura 7 – Modelo adotado para o perfil da trilha. Fonte: [5](figura 4.23).	32
Figura 8 – Detecção de borda por limiar local: os limiares locais τ_i são determinados por uma combinação linear das amplitudes A_i e valores de base B_i (linhas tracejadas). Fonte: [5] (figura 5.8).	33
Figura 9 – Diagrama do fluxo de informação que o programa executa. Fonte: Elaborado pelo autor.	39
Figura 10 – Imagens de um disco de vinil obtidas por câmera fotográfica. Fonte: Fotógrafo Thales Trigo.	40
Figura 11 – Imagens de um vinil 78 RPM obtidas com o microscópio USB. Fonte: Autor.	41
Figura 12 – Visualização do perfil de um segmento isolado da trilha contida na imagem usada no projeto. Fonte: Elaborado pelo autor.	42
Figura 13 – Visualização da derivada perfil da trilha contida na imagem usada no projeto. Fonte: Elaborado pelo autor.	42
Figura 14 – Região em que se realiza a média para obter um perfil suavizado, apontada pela seta. Fonte: Elaborado pelo autor.	44
Figura 15 – Perfil suavizado utilizado para obter a primeira estimativa de localização da borda. Note a influência da variação de iluminação na imagem expressa pela curvatura no perfil. Fonte: Elaborado pelo autor.	44
Figura 16 – Estimativa baseada no deslocamento das bordas vizinhas já definidas. Em vermelho, estão os pontos que são considerados estimativas na fatia de tempo e em azul, são os pontos candidatos coletados na etapa de detecção de borda. Os pontos rosa são as estimativas que caíram perto dos pontos candidatos. Fonte: Elaborado pelo autor.	45
Figura 17 – Estimativa baseada na média dos deslocamentos de todas as trilhas já definidas até então. Em vermelho, estão os pontos que são considerados estimativas na fatia de tempo e em azul, são os pontos candidatos coletados na etapa de detecção de borda. Fonte: Elaborado pelo autor.	46

Figura 18 – Caso em que há um ruído impulsivo e a linha de estimativas desvia da linha de candidatos. Em vermelho, estão os pontos que são considerados estimativas na fatia de tempo e em azul, são os pontos candidatos coletados na etapa de detecção de borda. Fonte: Elaborado pelo autor.	46
Figura 19 – Desvio no curso durante acompanhamento de traço causado por defeito. Os pixels azuis representam os candidatos a borda e os vermelhos, estimativas. Fonte: Elaborado pelo autor.	48
Figura 20 – Magnificação na imagem do vinil do projeto (não negativa), destacando os riscos perpendiculares a borda. Fonte: Elaborado pelo autor.	49
Figura 21 – Perfil do áudio obtido pela equipe do professor Carl Haber. Fonte: Elaborado pelo autor.	51
Figura 22 – Perfil do áudio obtido pelo software deste trabalho, sem borrão de movimento. Fonte: Elaborado pelo autor.	51
Figura 23 – Perfil do áudio obtido pelo software deste trabalho, com borrão de movimento. Fonte: Elaborado pelo autor.	51
Figura 24 – Espectrograma do áudio obtido pela equipe do professor Carl Haber. Imagem obtida pelo programa Audacity [11]. Fonte: Elaborado pelo autor.	52
Figura 25 – Espectrograma do áudio obtido pelo software deste trabalho, sem borrão de movimento. Imagem obtida pelo programa Audacity [11]. Fonte: Elaborado pelo autor.	52
Figura 26 – Espectrograma do áudio obtido pelo software deste trabalho, com borrão de movimento. Imagem obtida pelo programa Audacity [11]. Fonte: Elaborado pelo autor.	52

Lista de tabelas

Tabela 1 – Resultados do levantamento de dados realizado pela IASA. Fonte:	
[2](tabela 1.1)	20
Tabela 2 – Dados referentes ao vinil 78 rpm. Fonte:[2](tabela 2.4).	25

Sumário

1	INTRODUÇÃO	19
	Introdução	19
2	VINIL E MÍDIAS FONOGRÁFICAS	21
3	DIGITALIZAÇÃO POR PROCESSAMENTO DE IMAGEM	23
4	GEOMETRIA DO DISCO E TRILHA	25
4.1	Formação da imagem e definições	25
5	PROCESSO DE DIGITALIZAÇÃO	29
5.1	Obtenção de imagens	29
5.2	Processamento de imagens	31
5.3	Detecção grosseira de bordas	31
5.4	Detecção refinada de bordas	33
5.5	Acompanhamento de traço	33
5.5.1	Inicializar traços	34
5.5.2	Posições estimadas $\hat{r}_i(j)$ e $\hat{f}_i(j)$	34
5.5.3	Definição do alcance m	35
5.5.4	Seleção de candidatos	35
5.6	Exportação do áudio	37
5.6.1	Conexão das bordas	37
5.7	Recuperação do áudio	38
5.8	Tecnologias utilizadas	38
5.9	Projeto do software	39
5.10	Obtenção da imagem	39
5.11	Detecção de borda	41
5.12	Acompanhamento de traço	43
5.12.1	Inicialização das trilhas	43
5.12.2	Posições estimadas	43
5.12.3	Definição do alcance	45
5.12.4	Seleção de candidatos	45
5.12.5	Tratamento dos pontos indefinidos	47
5.12.6	União e derivada das posições da trilha	47
5.12.7	Re-amostragem	47
5.12.8	Mitigação de imperfeições	48

6	RESULTADOS	51
6.1	Avaliação	51
7	CONCLUSÃO	55
7.1	Dificuldades Encontradas	55
7.2	Aprendizado	55
8	BIBLIOGRAFIA	57

1 Introdução

O que nós percebemos como som são variações de pressão no ar que atingem nossos ouvidos. Partindo disso nosso cérebro associa esses estímulos a sensações diferentes, dependendo da intensidade e da frequência de variação dessas pressões. Mas como isso é traduzido numa peça sólida como um disco de vinil? Para isso, é necessário uma conversão da informação que presente no ar para algo traduzível em material sólido.

Hoje em dia existem diversas maneiras de fazer isso, sendo a mais popular o armazenamento digital, porém muitos registros de nossa cultura mundial estão armazenados de maneira analógica, que se traduz de maneira diferente. Num disco de vinil mono, por exemplo, o som é armazenado de maneira análoga à sua manifestação no ar: gira-se o disco em uma velocidade pré-determinada e corta-se, através de uma agulha, uma trilha espiral. Influenciamos essa agulha apenas na direção radial do disco, de modo que ela se mova analogamente a intensidade do som. Ao final do processo, obtemos uma representação material e visual do som em questão, e para lermos essa informação, deve-se fazer o caminho inverso: girar na mesma velocidade de gravação do disco e, com uma agulha que não corte dessa vez, ler as perturbações que obtivemos na direção radial e amplificá-las para que se propagem no ar novamente.

Arquivadores de som encontram uma grande quantidade de problemas com discos fonográficos. Para obter uma visão geral sobre quais são estes problemas, a IASA (International Association of Sound and Audiovisual Archives) realizou uma pesquisa sobre portadores audiovisuais em perigo, requisitada pela UNESCO (United Nations Educational, Scientific and Cultural Organization) em 2003 [1]. 2093 instituições em 184 países foram requisitadas para classificar suas mídias em 3 categorias: Boas condições, dando sinais de preocupação e claramente deteriorando. Os resultados dessa pesquisa são mostrados na tabela abaixo:

Como mostrado, há uma grande porcentagem de portadores de mídia classificados como dando sinais de preocupação e claramente deteriorando. Nesse cenário, é de grande necessidade um esforço na direção de um método que possa salvar esse conteúdo da extinção.

Desta necessidade surgiu em 2006 o projeto VisualAudio, no qual este trabalho se baseia. O objetivo do projeto VisualAudio era fazer um estudo sobre o método de extração de áudio de mídias fonográficas por meio de processamento de imagem, principalmente pela situação das mídias de acetato de corte direto que são cópias únicas e apenas 2,84% estão em boas condições.

O objetivo deste projeto é implementar um software capaz de receber imagens de

Portadores	No instituições / No itens	Boas Con- dições	Com sinais de Preocupação	Em deterio- ração
Gravações em cilindro	20/43965	14,65 %	58,73 %	26,62 %
Discos Shellac	41/614935	95,06 %	4,93 %	0,02 %
Discos de corte direto	23/60332	2,84 %	35,03 %	62,13 %
discos de vinil	55/1855120	88,43 %	11,56 %	0,01 %
fitas magnéticas	49/2161941	76,94 %	21,28 %	1,78 %
CDs graváveis	48/193062	86,94 %	9,95 %	3,11 %
CDs de áudio	52/1128400	95,36 %	4,58 %	0,06 %
Fitas digitais R-DAT	29/198477	45,40 %	27,85 %	26,75 %

Tabela 1: Resultados do levantamento de dados realizado pela IASA. Fonte: [2](tabela 1.1)

um disco de vinil e extrair dessas o áudio contido na trilha do disco. O material vinil foi definido por ter sido o único material para o qual foi possível obter imagens adequadas, porém vale ressaltar que a metodologia foi desenvolvida para mídias fonográficas no geral, sendo necessários apenas retoques na implementação na mudança de ma mídia para outra.

Esta monografia esta estruturada em 8 capítulos apresentando como a informação do áudio é armazenada num disco no capítulo 2, porque é possível extrair o áudio a partir de imagens no capítulo 3, um detalhamento sobre o formato do disco e sua trilha no capítulo 4, uma visão geral do processo de digitalização pelo método visual, além de detalhes de sua implementação no capítulo 5. Ao fim, temos umas palavras sobre o resultado obtido no capítulo 6 e finalizando com as conclusões do projeto no capítulo 7 e a bibliografia no capítulo 8.

2 Vinil e mídias fonográficas

Antes de entrarmos no processo de digitalização em si, é necessário entender como é feita a transcrição de áudio em um disco de vinil. O vinil é um exemplo de mídia fonográfica, uma classe de portadores de conteúdo em que temos som armazenado de maneira física, por meio de modulações em sulcos cavados no material.

Essas modulações variam de acordo com o tipo do disco. No caso de um disco mono, a modulação se encontra apenas no plano horizontal, paralelo a superfície do disco, já no caso do disco estéreo, os canais direito e esquerdo são ambos modulados vertical e horizontalmente.

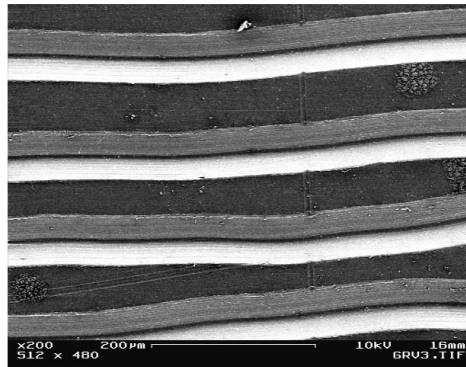


Figura 1: Foto microscópica dos segmentos da trilha de um vinil mono. Fonte: [12]

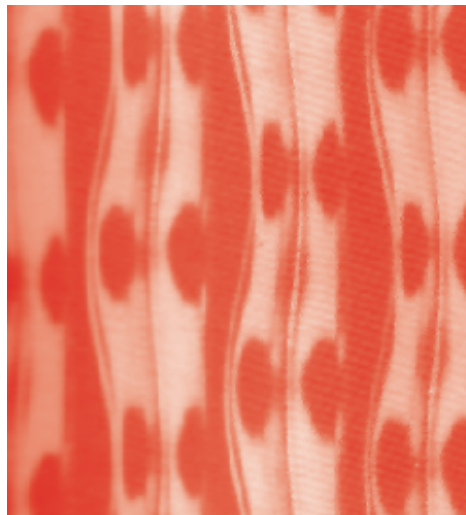


Figura 2: Foto microscópica do segmentos da trilha de um vinil estéreo. Fonte: [13]

Os canais do disco estéreo são modulados num ângulo de 45° graus com a superfície do disco e 90° graus entre si, sendo assim, quando um canal varia na intensidade, a parede da trilha se desloca perpendicularmente ao seu plano. Isso resulta numa variação de

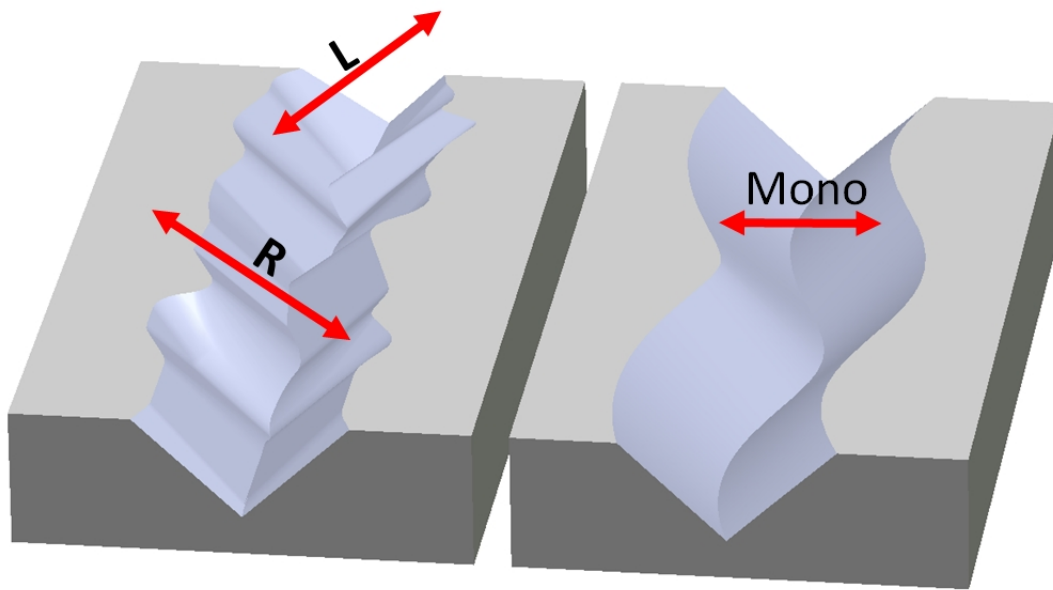


Figura 3: Ilustração das direções de modulação em cada caso de vinil. Fonte:[14]

profundidade e largura da trilha, o que causa a exclusão deste tipo de vinil do escopo de nosso projeto, visto que essa detecção da variação de profundidade da trilha com imagens 2D é ineficaz e, como veremos mais tarde, ter a largura da trilha fixa é de grande ajuda no processamento de imagem, já que pode ser usada para detecção de pontos espúrios na análise de borda.

No caso do disco mono, temos apenas a modulação horizontal, ou seja, a trilha quando esta sendo cavada no material sofre variações na direção radial do disco, de acordo as intensidades do som sendo armazenado ali. Essencialmente, o que um vinil armazena são as variações posicionais que, quando percorridas por uma agulha que gere sinais elétricos de acordo com sua movimentação radial, geram (salvo distorções e ruídos) o som original.

3 Digitalização por processamento de imagem

Como visto na sessão anterior, em uma mídia fonográfica, o áudio se manifesta graficamente no material do disco. Sendo assim, é possível uma recuperação deste áudio por meio da análise dessa imagem, desde que haja uma captura fiel das variações moduladas na direção radial da trilha.

Por captura fiel, entendamos que seja uma imagem em que as variações da trilha sejam discerníveis de maneira satisfatória apesar dos eventuais ruídos do processo de obtenção desta imagem. Tendo em mente que a informação essencial que queremos obter da imagem é o mapeamento completo da trilha, devemos simplificar a imagem o máximo possível para que a posição da trilha esteja em destaque desimpedida por quaisquer obstáculos que poderiam ser provenientes do processo de obtenção da imagem.

Sendo assim, em situação considerada ideal, teríamos uma imagem bicolor, contendo todo o comprimento da trilha em formato retangular, tendo as linhas de encontro da superfície do disco com a trilha demarcadas (as bordas) com uma largura mínima, sem borrões e em contraste com um fundo completamente na cor oposta. Porém, como estamos limitados a imagens imperfeitas, devemos restringir os efeitos que podem atrapalhar o mapeamento, buscando se aproximar da situação ideal à medida do possível.

É discutido no projeto de Sylvain Stotzer [2] os esforços empregados na obtenção das imagens que cabiam dentro das restrições, sendo dado um detalhamento maior na fase da fotografia analógica, e uma breve descrição dos escaneadores que foram construídos especificamente para digitalização dos filmes resultantes da etapa analógica. Na sessão de descrição do processo geral de digitalização 5 serão listadas as condições esperadas para que a imagem seja processada pelo software deste projeto com sucesso.

4 Geometria do disco e trilha

Um disco de vinil é gravado no sentido horário com uma espiral concêntrica correndo de fora para dentro. Alguns discos de corte direto para transcrição de rádio e alguns discos comerciais mais antigos seguem o caminho de dentro para fora, porém estes não serão cobertos nesse trabalho. Deve ser notado que os primeiros discos comerciais de 78 rpm não eram padronizados, portanto algumas medidas podem variar levemente das presentes na tabela abaixo:

diâmetro total(mm)	250	300
diâmetro até trilha mais externa(mm)	241.3	292.1
diâmetro até trilha mais interna(mm)	95	95
excentricidade máxima do buraco do eixo(um)	50	50
sulcos por mm	2.8 ~ 4.72	2.8 ~ 4.72
espaçamento de sulcos(um)	211 ~ 357	211 ~ 357
largura do sulco em cima(um)	150 ~ 200	150 ~ 200
profundidade do sulco(um)	40 ~ 80	40 ~ 80
raio no fundo do sulco(um)	20.32 ~ 63.5	20.32 ~ 63.5
ângulo da trilha(graus)	90 ± 8	90 ± 8
velocidade máxima de pico(cm/s)	28	28
amplitude máxima(um)	75	75

Tabela 2: Dados referentes ao vinil 78 rpm. Fonte:[2](tabela 2.4).

4.1 Formação da imagem e definições

A fotografia é uma projeção de um objeto 3D em um plano 2D. Nossa imagem é formada pela reflexão de luz da superfície do disco e pela falta de reflexão de outras partes.

A maior parte da reflexão na superfície num disco de vinil é especular (e.g. reflexo que ocorre num espelho) portanto nosso modelo leva esta reflexão em consideração. A iluminação do vinil no momento da foto deve ser perpendicular à sua superfície e coaxial, ou seja, a fonte deve estar o mais próximo possível do eixo da lente, de modo as regiões que refletirem mais luz de volta pra câmera produzirão uma imagem mais clara, e as regiões que não refletirem luz na direção da câmera, serão obscurecidas.

A imagem formada por uma foto adequada do vinil possui algumas características nas quais iremos nos basear para construção dos algoritmos. Como partimos do modelo baseado no projeto VisualAudio, consideramos os padrões do filme negativo que capturou a imagem do vinil, portanto o raciocínio empregado no parágrafo anterior é inverso: as regiões que refletirem mais luz na direção da câmera irão queimar mais o filme, deixando

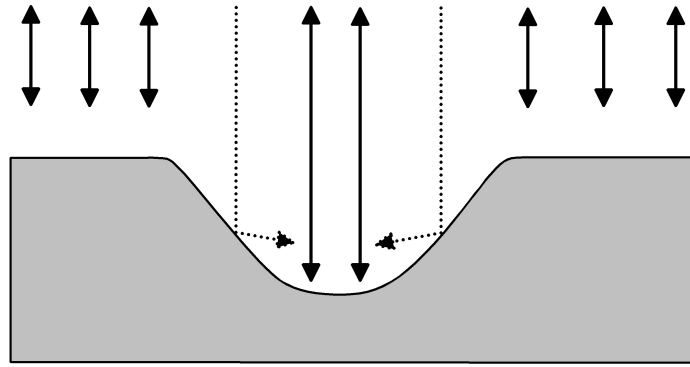


Figura 4: Desenho do perfil do sulco de nosso vinil. Fonte: [5](figura 4.1).

a região escura, e as regiões que dispersarem luz para fora do campo da lente aparecerão mais claras.

Como visto na figura 4, a luz que chega perpendicular à superfície do vinil será refletida de volta na direção da câmera (indicadas pelas setas sólidas) e a luz que chega na parede da trilha é dispersa para outra direção (há um pequeno reflexo causado pela luz que bate nas duas paredes e volta, mas para nosso escopo esse reflexos são negligenciáveis). Com este arranjo, podemos dividir a trilha em 3 partes:

- Traços claros (trace) : traços leves que representam a parede (wall) da trilha.
- Áreas escuras : superfície do disco (land) (entre trilhas) e fundo da trilha
- Transições cinzas: ficam entre as áreas escuras e as claras (maioria das vezes é borrado devido à aquisição de imagens)

Em uma situação geral, a trilha pode ser de um traço ou 2 traços, dependendo do tipo do vinil, porém, como nosso escopo se limita aos discos de 78 RPM, as trilhas no nosso trabalho serão sempre de 2 traços. Cada traço tem duas bordas, uma de subida (preto -> branco) e outra de descida (contrário). Em trilhas de 2 traços devemos distinguir as bordas de cima [top] (entre wall e land) e de baixo [bottom] (fundo da trilha). Posições relativas entre traços podem ser expressas por sua posição radial: interna (mais próxima do centro) e externa.

Na figura 5, um exemplo da imagem recebida para o teste do programa, cedida por Carl Haber:

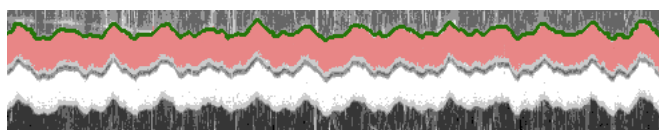


Figura 5: Segmento de trilha de uma das imagens recebidas no projeto. Marcado em vermelho está um traço do segmento e delineado em verde está uma borda do traço. Fonte: Elaborado pelo autor.

5 Processo de Digitalização

O processo de digitalização da imagem de um vinil em que esse trabalho se baseia pode ser dividida em 3 partes:

1. Obtenção das imagens
2. Processamento das imagens
3. Exportação do áudio.

5.1 Obtenção de imagens

Essa parte do processo é a única que não faz parte do escopo deste trabalho. A obtenção de imagens digitais apropriadas de um disco fonográfico requer equipamento especializado e poderia ser tema de um trabalho de conclusão de curso por si só.

Para citar uma maneira de obter as imagens, vale descrever o processo utilizado no projeto VisualAudio, em que se dividiu essa etapa em duas partes: a fotografia analógica e o escaneamento do filme. A causa desta quebra em duas etapas é justificada na tese [2] pelos principais motivos de livrar a mídia portadora do áudio o quanto antes do fardo de ser cópia única (já que o filme seria um novo portador válido daquele conteúdo) e ter garantir máxima resolução possível, visto que na época do projeto não havia consenso sobre a tecnologia de captura digital vigente ser suficiente para extração fiel do áudio por processamento de imagem.

Para a primeira etapa foi construída uma câmera analógica enorme, visando obter uma relação de 1 para 1 entre a imagem projetada no filme e o disco original. É descrito na tese de Stotzer [2] os possíveis ruídos inseridos na informação do disco causados por essa etapa, como o 'pepper fog spots' (grão de haleto de prata formados espontaneamente na emulsão do filme), grãos do filme, entre outros.

Da primeira etapa é obtido um filme do tamanho do vinil, o qual é escaneado por uma máquina construída especificamente para isso. O filme então é posto numa bandeja de vidro rotatória, imitando o propósito de uma bandeja de toca-discos, que gira numa velocidade apropriada, sincronizada com a câmera do escaneador, que mantendo uma taxa de amostragem constante e precisamente controlada, produzindo ao final deste processo uma série de imagens componentes de uma coroa do disco. De anel em anel a trilha do disco é coberta por completo, sendo que cada anel é uma coroa circular (concêntrica com

o disco) que cobre um pedaço da trilha, do qual alguns segmentos são descartados por estarem incompletos no anel. Desse modo é possível começar o processamento da imagem, já digitalizada.

Em nosso projeto, usaremos as imagens cedidas por Carl Haber, professor de física que trabalhou em um outro projeto, semelhante ao VisualAudio, chamado IRENE [3]. Neste projeto a etapa de digitalização era feita diretamente através do escaneamento do disco. Note que as cores da figura 6 são contrárias às esperadas pelo software baseado na imagem de negativos.

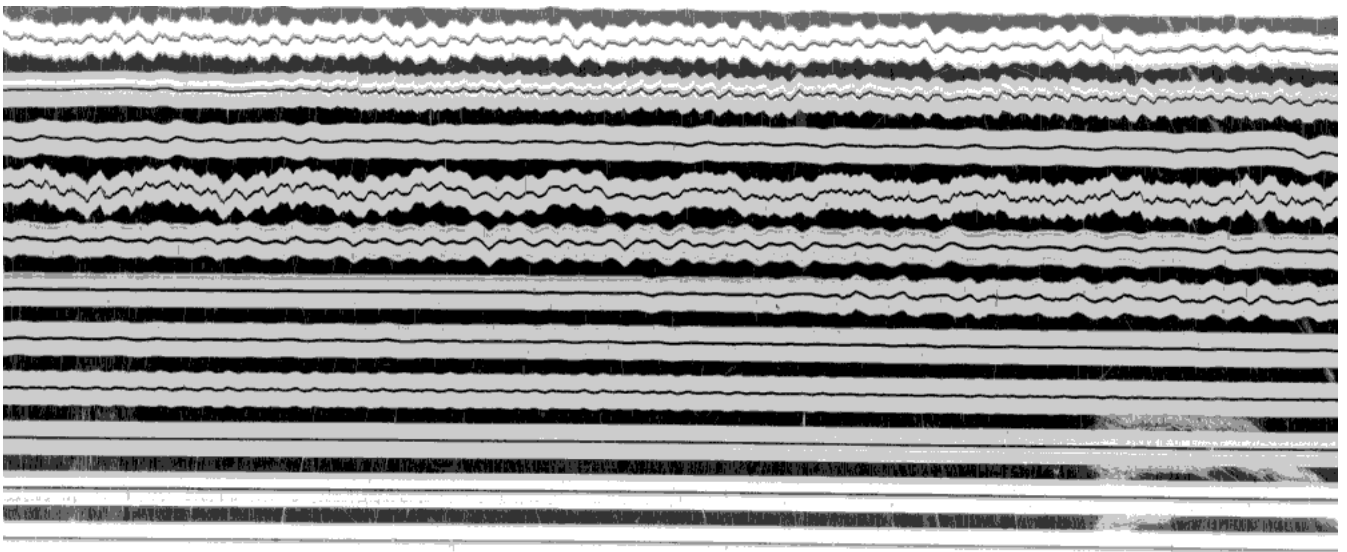


Figura 6: Fragmento de disco desconhecido. Fonte: Carl Haber.

O programa resultante deste trabalho recebe como entrada uma série de imagens que, juntas, correspondem à cobertura de uma coroa (ou anel) de um disco de vinil, 78 RPM mono. Com base nos processos descritos nos materiais de referência [2], foi sintetizada uma lista de requisitos para as imagens utilizadas pelo software deste projeto:

- Ter resolução tal que permita a diferenciação de distâncias de $2.5 \mu\text{m}$.
- A imagem deve estar em tons de cinza (valores de 0 a 255), no formato bmp ou tiff, sendo os valores mais altos (branco) correspondentes a parede da trilha e os mais escuros (preto) ao fundo da trilha e superfície do disco.
- As trilhas devem estar praticamente paralelas a borda da imagem (salvo o deslocamento linear que a trilha faz em direção ao centro), de modo que a curvatura do disco não seja relevante.
- As imagens, quando justapostas, não devem apresentar descontinuidade (como se a série de imagens pudesse ser tratada como uma única).

- A iluminação no momento da captura deve ser a mais próxima de homogênea possível, de modo que as paredes da trilha possuam uma intensidade semelhante de cinza ao longo de toda imagem.

5.2 Processamento de imagens

O processamento de imagem é parte integrante das funções do software deste projeto. Esse processamento cumpre os seguintes objetivos:

1. Reconhecer os segmentos de trilha do disco na imagem.
2. Amostrar esses segmentos.
3. Reconhecer e lidar com pontos de trilha indefinidos.
4. Unir os segmentos da trilha formando um segmento único correspondente ao anel usado de entrada para o programa.

O pseudo-código do algoritmo que é aplicado às imagens pode ser resumido da seguinte maneira:

- Inicializar traços
- Para cada anel
 - Para cada linha
 - * Detecção Grosseira: localização aproximada
 - * Detecção refinada: Detecta os candidatos a ponto de borda
 - * Para cada traço
 - Selecionar os candidatos que se encaixam melhor no traço
 - Reconstrução de trilha ao longo do anel
- Reconstrução de trilha ao longo do disco

5.3 Detecção grosseira de bordas

O algoritmo apresentado nesta seção é aplicado linha a linha da imagem adquirida, sendo cada linha correspondente a uma fatia de tempo. Cada linha é representada por uma sequência L de valores de cinza (como pode ser visto no perfil da figura 12).

O objetivo dessa etapa é detectar traços grosseiramente em uma imagem suave sem ruído. Desse modo obtemos a presença de todos traços e conseguimos uma localização

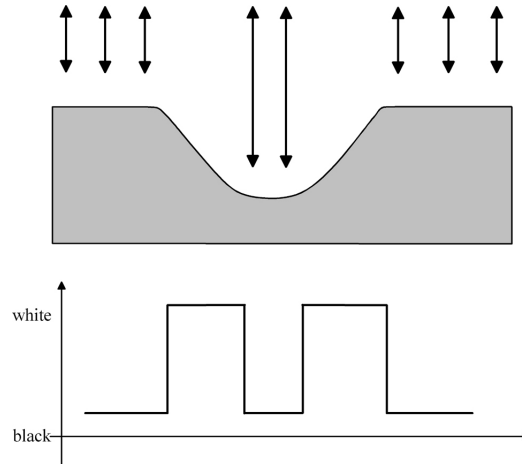


Figura 7: Modelo adotado para o perfil da trilha. Fonte: [5](figura 4.23).

(é definido um ponto que está na região de transição das cores) aproximada de todas bordas. Com esses locais, podemos estimar a amplitude de degrau A e valor base B (como é demarcado na figura 8), mesmo em casos de alta variação de iluminação. Essa etapa evita detecção de alguns pontos inválidos (induzidos por particulado ou danos) pois localiza apenas objetos com tamanho desejado.

A detecção é implementada através de uma convolução com um double box filter [2] $b(x)$ definido por um kernel $\lambda \times 1$:

$$b = [-1 \dots -101 \dots 1] \quad (5.1)$$

Como a largura w da trilha e do traço são praticamente constantes, a escala do filtro é definida por:

$$\lambda = \alpha * w \quad (5.2)$$

com α entre 0.2 e 0.8 geralmente. A região entre dois extremos (do resultado da convolução) localiza um traço ou uma área escura.

Daí podemos obter a amplitude A e o valor base B tirando, respectivamente, o valor máximo e o mínimo em cada uma dessas regiões. Outra maneira de detectar aproximadamente os traços seria iniciar o processo descrito acima no início do processamento e propagar essa informação na direção tangencial pelo tempo usando as bordas detectadas do tempo i para obter as bordas do tempo $i+1$, porém esse método também propaga erros e gera um resultado final ruim, principalmente em discos muito danificados.

5.4 Detecção refinada de bordas

Uma das restrições apresentadas pelo modelo é detectar a posição da borda relativa a amplitude do degrau local A e o valor base B (determinados anteriormente) para manter as bordas simétricas. O método mais simples para realizar isso é definir um limiar adaptativo τ definido na equação 5.3:

$$\tau = \beta(A - B) + B, \beta \in [0, 1] \quad (5.3)$$

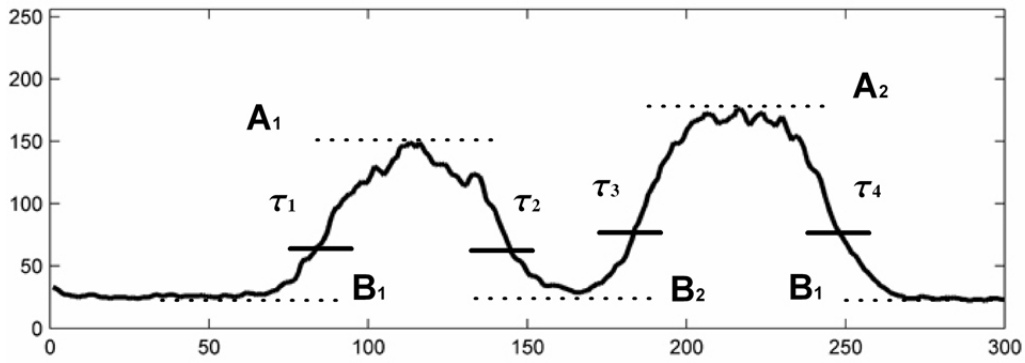


Figura 8: Detecção de borda por limiar local: os limiares locais τ_i são determinados por uma combinação linear das amplitudes A_i e valores de base B_i (linhas tracejadas). Fonte: [5] (figura 5.8).

Um limiar é aplicado a uma sequência L correspondente à uma linha adquirida. A estimativa sub-pixel, ou seja, contendo valor contínuo, é adquirida obtendo o valor da interpolação entre $L(i)$ e $L(i + 1)$ que satisfazem:

$$\begin{aligned} L(i) &\geq \tau \text{ e } L(i + 1) < \tau \text{ para bordas de descida ou} \\ L(i) &< \tau \text{ e } L(i + 1) \geq \tau \text{ para bordas de subida.} \end{aligned} \quad (5.4)$$

5.5 Acompanhamento de traço

Esse trecho do processamento irá definir o formato dos segmentos de trilha. Podemos dividir este processo nas seguintes partes:

1. Inicializar traços T_i .
2. Definir estimativas para cada borda do traço T_i , $\hat{r}_i(j)$ e $\hat{f}_i(j)$.
3. Definir o alcance de tolerância m para cada estimativa de borda.
4. Escolher o melhor candidato dentro do intervalo de tolerância.
5. Incrementar o valor da fatia de tempo j e voltar ao passo 2.

5.5.1 Inicializar traços

Os traços são inicializados a partir da primeira e da última imagem do anel. Tomam-se as n primeiras linhas da primeira imagem e as n últimas linhas da última imagem para calcular uma média ponto a ponto, com n podendo variar entre 5 e 30, dependendo da velocidade de rotação do disco (RPM), frequência de amostra da imagem e grau de degradação da imagem. O resultado da média (figura 14) dessas linhas é a sequência de pontos J que apresentará um perfil suavizado das trilhas no começo do anel (como apresentado na imagem 15 para o caso das imagens usadas no projeto). Essa sequência J agora é posta sob os processos de detecção de borda das seções 5.3 e 5.4 de onde obteremos as estimativas de candidatos a borda. Estes candidatos são as estimativas $\hat{r}_i(j)$ e $\hat{f}_i(j)$ de cada traço.

5.5.2 Posições estimadas $\hat{r}_i(j)$ e $\hat{f}_i(j)$

As estimativas $\hat{r}_i(j)$ e $\hat{f}_i(j)$ são usadas para se ter uma idéia da posição de cada traço na figura, mesmo quando o disco apresenta degradações relevantes e o traço não é visível ou está fora do lugar.

Como o deslocamento radial das trilhas apresenta apenas variações suaves de uma linha para a próxima, as estimativas $\hat{r}_i(j)$ e $\hat{f}_i(j)$ de cada traço T_i são obtidos com base nos candidatos escolhidos do tempo $j-1$:

$$\begin{aligned}\hat{r}_i(j) &= r_i(j-1) \\ \hat{f}_i(j) &= f_i(j-1)\end{aligned}\tag{5.5}$$

Se a borda do tempo $j-1$ não pode ser extraída ela é marcada como indefinida, nesse caso:

$$\begin{aligned}\hat{r}_i(j) &= r_i(j-1) + d \\ \hat{f}_i(j) &= f_i(j-1) + d\end{aligned}\tag{5.6}$$

onde d , que é definido por:

1. Deslocamento da outra borda do mesmo traço, se não for indefinida.
2. Deslocamento das bordas do outro traço da mesma trilha (caso 78 rpm), se não forem indefinidas.
3. Média dos deslocamentos de todos outros traços da imagem.

A média dos deslocamentos dos outros traços vai resultar numa aproximação dos componentes de baixa frequência no sinal, causados pela espiral da trilha e a descentralização do eixo do disco no momento de captura da imagem.

5.5.3 Definição do alcance m

O objetivo do alcance m é separar traços e impedir sobreposição (entre bordas de subida ou entre bordas de descida) em casos de degradação. Outro propósito também é um primeiro processo de tratamento de ruído.

$$m_{min,r,i}(j) = \hat{r}_i(j) - p \text{ e } m_{max,r,i}(j) = \hat{r}_i(j) + p \quad (5.7)$$

$$m_{min,f,i}(j) = \hat{f}_i(j) - p \text{ e } m_{max,f,i}(j) = \hat{f}_i(j) + p \quad (5.8)$$

onde p é o alcance e costuma variar entre 10 e 20 μm .

No caso do range m de uma trilha sobrepor o candidato a borda de outra trilha, o valor de m deve ser reajustado da seguinte maneira:

$$\text{Se } m_{min,r,i+1}(j) < \hat{f}_i(j) \text{ então } m_{min,r,i+1}(j) = \hat{f}_i(j) \quad (5.9)$$

$$\text{Se } m_{max,f,i}(j) > \hat{r}_{i+1}(j) \text{ então } m_{max,f,i}(j) = \hat{r}_{i+1}(j) \quad (5.10)$$

5.5.4 Seleção de candidatos

As posições de borda $r_i(j)$ e $f_i(j)$ são escolhidas como melhores candidatas dentro do alcance m definido em volta das estimativas $\hat{r}_i(j)$ e $\hat{f}_i(j)$. Existem 3 possibilidades no caso da avaliação de $\hat{r}_i(j)$ e são:

- Caso não haja candidato escolhido, $r_i(j)$ é marcado como indefinido e é resolvido em etapa posterior.
- Se houver apenas um candidato $cr_k(i)$ esse é usado para definir a borda

$$r_i(j) = cr_k(i). \quad (5.11)$$

- Se houverem vários candidatos, é escolhido o candidato $cr_k(i)$ retorna o menor valor possível para a operação:

$$|\hat{r}_i(j) - cr_k(i)| \quad (5.12)$$

A escolha de $f_i(j)$ é basicamente igual à de $r_i(j)$. Devido aos métodos utilizados e a suavização causada pelo *blur* na captura da imagem, na maior parte do tempo há apenas

um candidato à borda dentro do alcance m . Porém, há a possibilidade de degradações no disco, que produzem outros candidatos. Nesses casos, a média de largura $\bar{w}_i(j)$ do traço i no tempo j pode ser usada para testar a validade da escolha. Duas condições são estimadas para essa avaliação:

$$\begin{aligned} |\hat{f}_i(j) - r_i(j) - \bar{w}_i(j)| &> |\hat{f}_i(j) - cr_k(i) - \bar{w}_i(j)| \\ |f_i(j) - \hat{r}_i(j) - \bar{w}_i(j)| &> |cf_k(i) - \hat{r}_i(j) - \bar{w}_i(j)| \end{aligned} \quad (5.13)$$

O que as equações 5.13 estão fazendo (tomando a primeira como exemplo) é comparar a distância entre o ponto escolhido (no caso $r_i(j)$) e o ponto de estimativa da outra borda do traço ($\hat{f}_i(j)$) com o tamanho da média de largura $\bar{w}_i(j)$. Essa comparação é feita com outros candidatos no lado direito da inequação e, se algum candidato gerar uma distância melhor (mais próxima de $\bar{w}_i(j)$), esse ponto deve ser marcado como indefinido e um processo de análise na direção decrescente de j marcará os pontos necessários como indefinidos também. A primeira inequação avalia os candidatos de borda de subida e a outra os de descida.

Se ambas condições forem verdade, quer dizer que é difícil tomar uma decisão no tempo j . Assim, os candidatos são mantidos e uma análise será realizada posteriormente através do processo reverso. Esse processo segue os seguintes passos:

1. Alterar o ponto escolhido para indefinido.
2. Decrementar j .
3. Calcular a largura no tempo j : $w_i(j) = |f_i(j) - r_i(j)|$
4. Se $|\bar{w}_i(j) - w_i(j)| > \zeta$ então voltar ao passo 1.

Como a largura pode sofrer variações importantes, ζ não pode ser muito rígido: um valor de 5% e 10% da média $\bar{w}_i(j)$ geralmente gera resultados satisfatórios. A média $\bar{w}_i(j)$ é atualizada a cada tempo j para seguir as variações suaves de largura de cada circunvolução:

$$\bar{w}_i(j) = \varsigma \cdot (\hat{f}_i(j) - \hat{r}_i(j)) + (1 - \varsigma) \cdot \bar{w}_i(j - 1) \quad (5.14)$$

com ς sendo bem próximo de 0, e.g. $\varsigma = \frac{1}{1000}$ ou menor. Como todos traços fazem parte da mesma trilha (uma trilha só que percorre o disco em espiral, o que vemos nas imagens são segmentos dessa trilha), a largura média do traço, que agora é definida pra cada trecho, pode ser substituída por 2 médias (já que vinis de 78 rpm tem 2 traços por trilha) para a trilha completa: 1 para traços internos e outra para traços externos.

5.6 Exportação do áudio

Ao chegar nesta etapa, devemos ter uma estrutura de dados no programa correspondendo a um segmento de trilha arbitrário, sendo correspondente à um anel ou a todo o disco. Os objetivos desta etapa são:

1. Obter áudio à partir da sequência de amostras recebida.
2. Equalizar sequência de áudio.
3. Exportar um arquivo no formato WAV.

5.6.1 Conexão das bordas

A etapa anterior fornece uma estrutura de dados contendo n segmentos de trilha, para cada estrutura haverá 4 sequências de números, cada uma correspondente a uma borda dos dois traços que compõem a trilha. Essa estrutura ainda não está preparada para ser transformada em áudio, e para deixá-la apropriada, devemos unificar esses segmentos em uma sequência unificada que corresponderá ao segmento da trilha completo capturado pelo anel introduzido como entrada no programa.

Admitindo que as posições das bordas contidas na estrutura de dados recebida correspondem aos índices do pixel dentro da matriz que contém a imagem, o seguinte algoritmo em pseudo-código é usado para a junção de suas extremidades:

- Criar nova estrutura de dado NT para conter a união das trilhas
- $distancia_entre_trilhas = 0$
- Para cada trilha $T(i)$, i indo desde 0 até $n - 1$:
 - Para cada valor j da trilha $T(i)$

$$* NT[(i + 1) * Numero_de_valores_em_uma_trilha + j] = T(i)[j] + distancia_entre_trilhas$$
 - calcular diferença entre a posição de cada borda da última fatia de tempo da trilha $T(i)$ e a posição de cada borda da primeira fatia de tempo da trilha $T(i + 1)$
 - $media =$ media das 4 diferenças
 - $distancia_entre_trilhas = distancia_entre_trilhas + media$

Terminado este processo, unificam-se as 4 bordas contidas na estrutura NT em apenas uma sequência de valores, calculando a média das posições das 4 bordas.

5.7 Recuperação do áudio

Tendo em mãos agora uma estrutura contendo as posições de borda do segmento contido no anel, obtemos o áudio através da derivada dessas posições, visto que esta função representa o sinal que a agulha de um toca-discos gera por indução eletromagnética, causado pelo formato da trilha.

Como estamos tratando de um sinal discreto no tempo, que representa uma versão quantizada do que seria o sinal produzido por uma agulha, a simples diferença de um valor para o próximo dividida pelo período da amostragem não seria um bom representante da derivada, visto que a presença de um ponto espúrio ou mal coletado da imagem pode gerar um pico de valor nessa derivada, que seria percebido como ruído no resultado final. É recomendado então uma aproximação da derivada que leve em conta um contexto maior que apenas dois pontos consecutivos. Sugere-se em [2](7.1.2) uma convolução com um filtro digital diferenciativo. A resposta em frequência deste filtro idealmente é proporcional a frequência, porém, como não é possível ter um filtro ideal, o recomendado é a utilização de um filtro de diferenciação de tamanho 11, que diferencia o sinal de entrada em uma largura de banda de 0 Hz até $\frac{0.8*f_s}{2}$, onde f_s é a frequência de amostra.

5.8 Tecnologias utilizadas

Para a implementação do programa, foram selecionadas as seguintes tecnologias:

- Linguagem de programação C++
- Biblioteca OpenCV

A biblioteca OpenCV foi escolhida para manipulação das imagens de entrada do programa. Esta biblioteca é escrita em C++ e possui uma interface com C por razões de manutenção do legado. Além destas duas linguagens, ainda existem envolvimento com Python, Java, MATLAB/OCTAVE, C#, Perl, Ruby, e outras, apesar de, entre essas, apenas a interface com python (além de C e C++) é indicada na documentação. Com esse cenário foi escolhida a linguagem C++ para implementar o software do projeto, pois é a linguagem da implementação atual da biblioteca e possui uma interface mais sólida.

Além das tecnologias citadas acima, foram inclusos outro pequenos módulos durante a implementação do programa, sendo eles os programas *cvplot* e *make_wav* que resolvem os problemas de, respectivamente, imprimir gráfico para uma sequência de valores (usados para se ter feedback dos algoritmos) e gerar arquivos WAV partindo de uma sequência de valores [4] [5].

5.9 Projeto do software

A função deste software é receber um conjunto de imagens e exportar um arquivo de áudio. Sua complexidade reside principalmente nos algoritmos. O projeto deste software é mostrado na figura 9 e mais detalhadamente explicado através das descrições no capítulo 5.

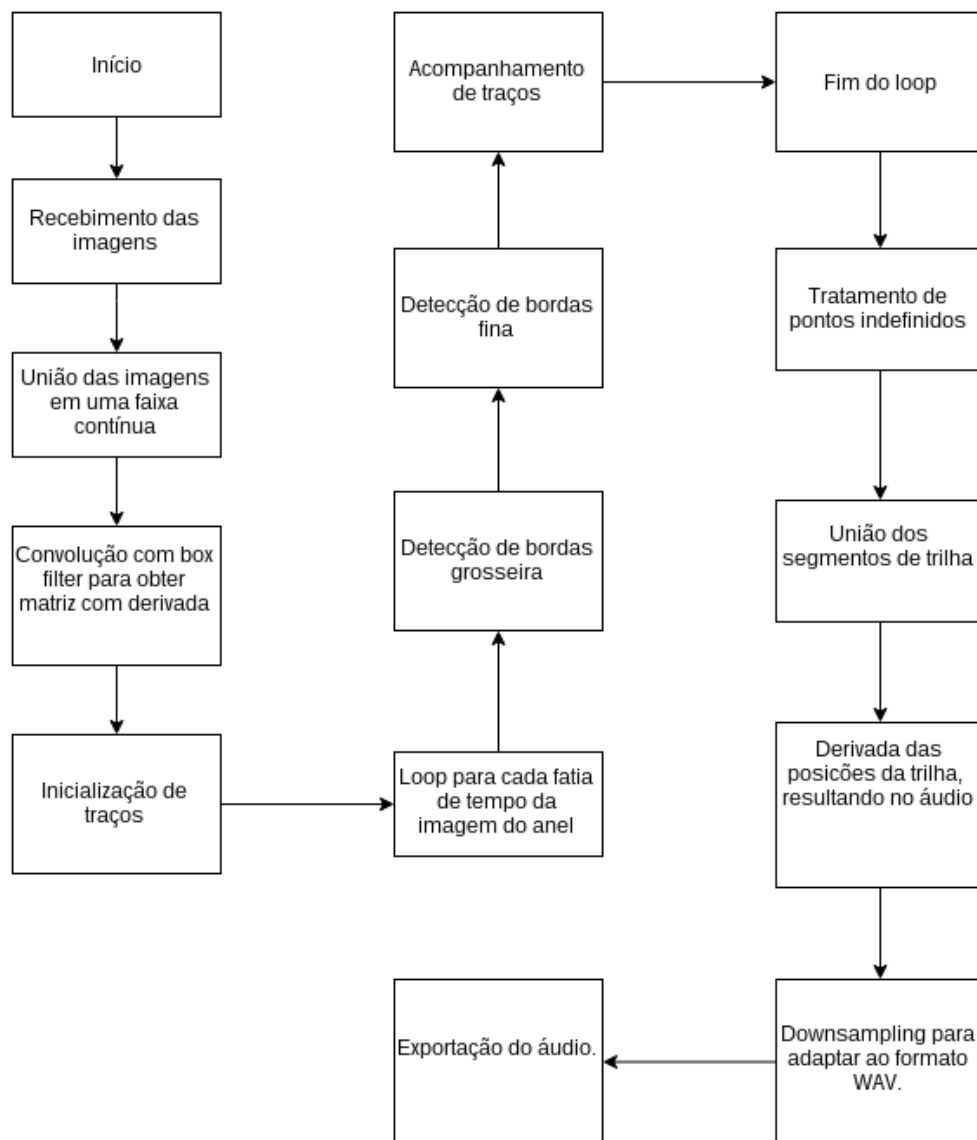


Figura 9: Diagrama do fluxo de informação que o programa executa. Fonte: Elaborado pelo autor.

5.10 Obtenção da imagem

A ideia no início do projeto era fotografar o disco de vinil por meio de uma câmera fotográfica, esta empreitada tomou lugar no estúdio Full Frame, com o fotógrafo e

diretor Thales Trigo. Porém, conforme foram obtidas mais informações sobre os projetos relacionados [2][3], ficou evidente que o grau de detalhe das fotos (10) não era satisfatório.

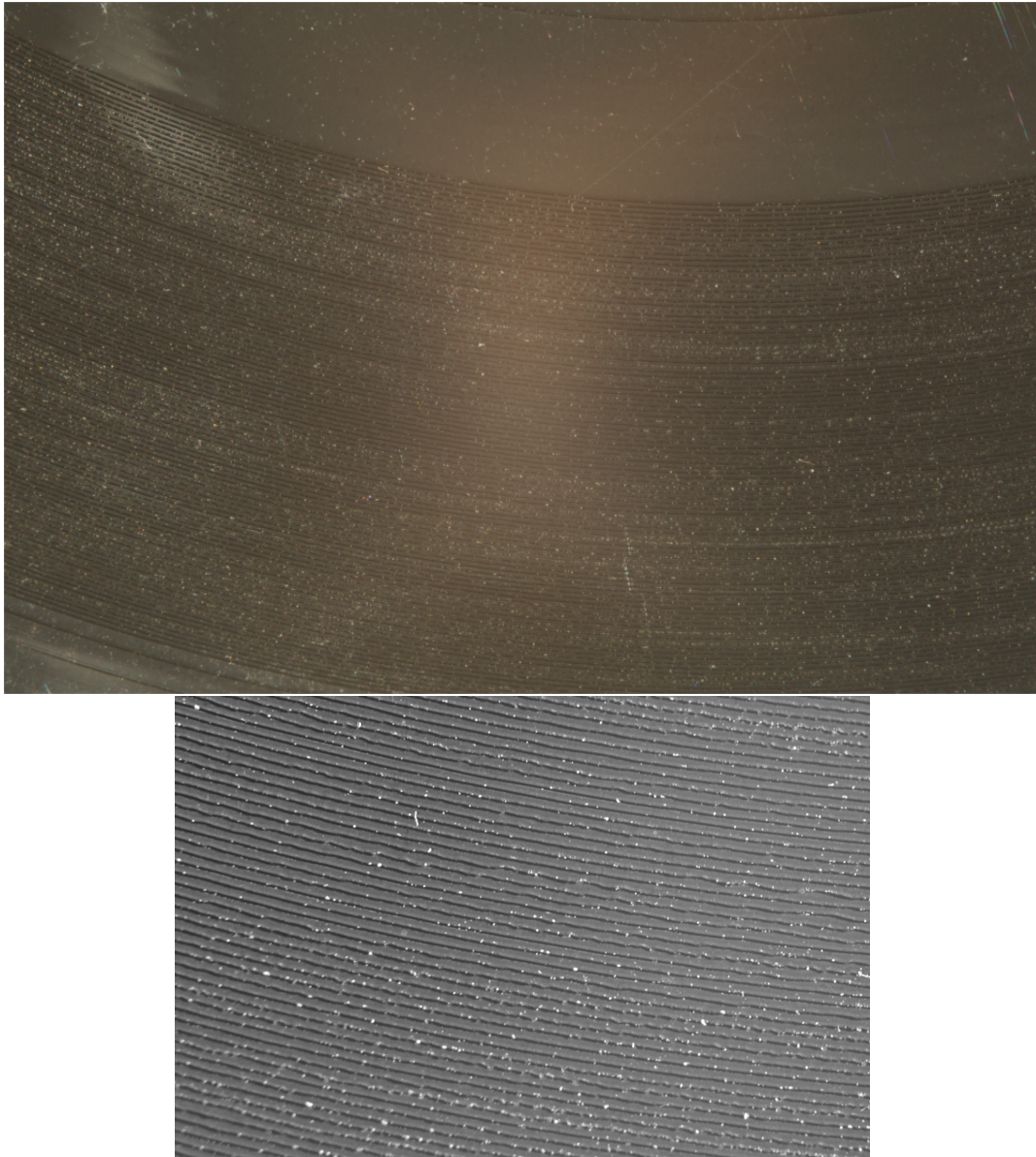


Figura 10: Imagens de um disco de vinil obtidas por câmera fotográfica. Fonte: Fotografia Thales Trigo.

A segunda ideia foi fotografar o disco usando um microscópio USB. Para isso, compramos um microscópio USB com o qual foram feitas imagens teste para a tentativa de extração do áudio. As imagens obtidas são ilustradas em 11.

Pode se notar que a diferença de qualidade entre essas imagens e as imagens utilizadas (exemplo na figura 6) por fim no projeto é enorme. Além disso ficou evidente a dificuldade de se obter estas imagens sem a utilização de um equipamento adequado, com a precisão necessária. Compreendido que a construção de um equipamento de obtenção de imagens aumentaria exponencialmente a complexidade do projeto, excluiu-se esta etapa

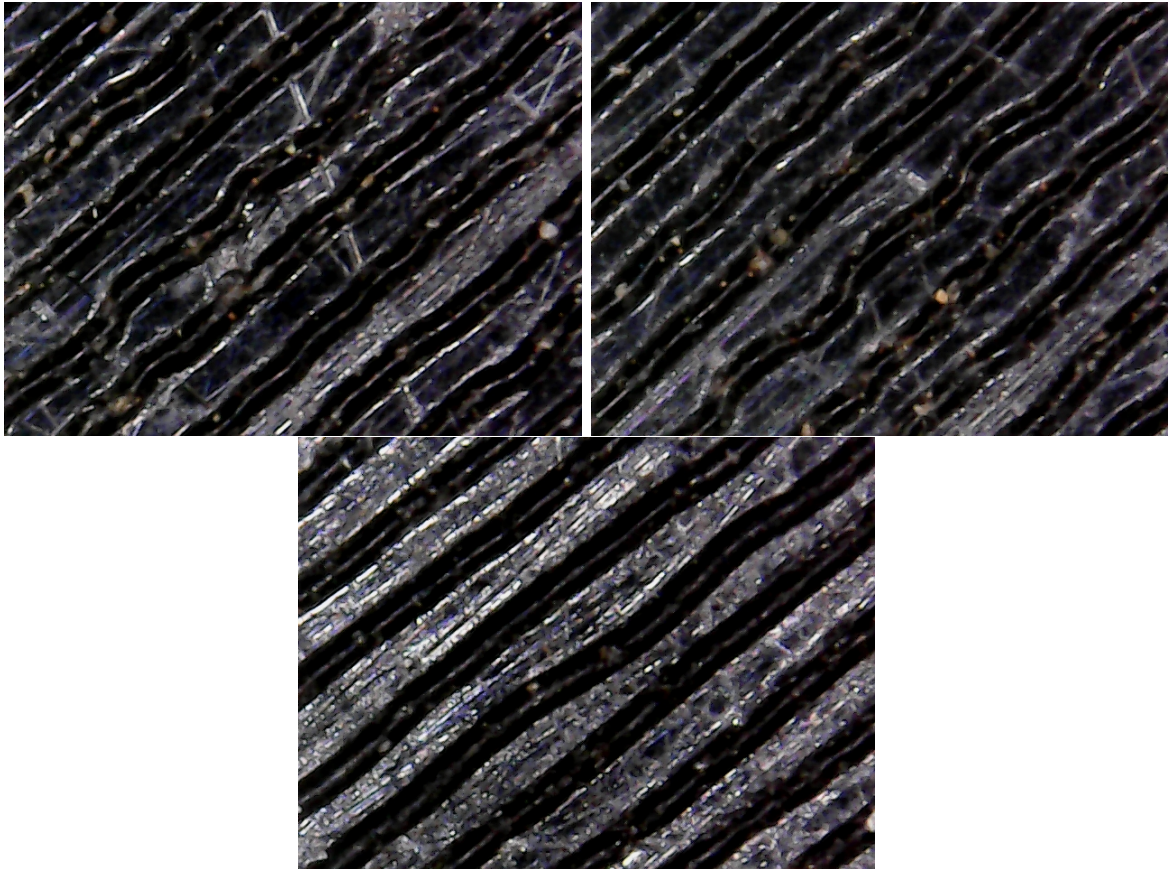


Figura 11: Imagens de um vinil 78 RPM obtidas com o microscópio USB. Fonte: Autor.

do escopo deste trabalho.

Sendo assim, foi preciso obter imagens que fossem coerentes com os requisitos deste trabalho com a ajuda de terceiros. Foram enviados e-mails para os autores dos artigos encontrados relacionados ao assunto [6][7] e, de todos, apenas o professor Carl Haber respondeu. Após algumas semanas do início da conversa, o professor enviou um link para uma página com as imagens que foram incorporadas ao nosso projeto. As informações obtidas sobre estas imagens limitam-se ao que está na página na referência [8].

5.11 Detecção de borda

Esse foi um dos primeiros passos a serem aplicados, sendo um dos mais importantes no projeto. Foi implementado o algoritmo de detecção grosseira explicado na Seção 5.3.

Neste capítulo explicaremos a utilização da biblioteca OpenCV, da qual foi utilizada a estrutura *Mat*, armazenadora da matriz da imagem e sua função de filtro, que foi utilizada para realizar a convolução.

Visto que os valores adotados para α no algoritmo eram determinados empiricamente, foram feitos testes variando α para avaliar o que seria mais adequado no caso.

A variação deste parâmetro influencia no quão suave o perfil da trilha ficará após a convolução, porém adotar um valor muito grande implica na anulação de alguns pontos de borda da trilha e um valor muito baixo resulta no destaque de pontos de borda falsos considerando o escopo do nosso trabalho. O valor escolhido para α foi 0.2 pois mostrou um resultado aceitável no processo de convolução.

Após a convolução, há ainda a normalização deste perfil, visto que os valores de seus picos variam muito de acordo com o tamanho da imagem utilizada. Para isso foi realizado um laço em cada fatia de tempo da matriz recipiente do perfil, buscando o módulo do valor máximo encontrado e dividindo os valores da fatia pelo máximo encontrado.

A implementação deste algoritmo ainda encontra um obstáculo, que é a detecção dos picos reais na derivada do perfil (figura 13), visto que como o perfil da derivada é composto por muitas micro-variações (mesmo com a suavização causada pelo tamanho da matriz de convolução utilizada na derivada). A técnica de encontrar a tangente de inclinação 0 neste perfil (da derivada, lembrando) para localizar os picos não funciona muito bem, visto que surgem diversos pontos no perfil cuja textura cria pontos de inclinação horizontal.

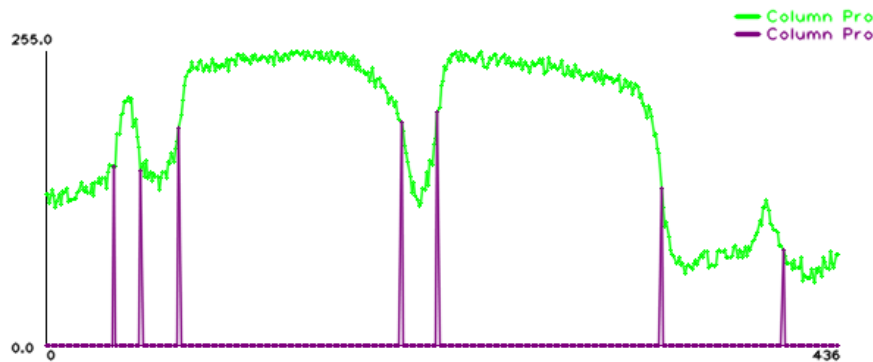


Figura 12: Visualização do perfil de um segmento isolado da trilha contida na imagem usada no projeto. Fonte: Elaborado pelo autor.

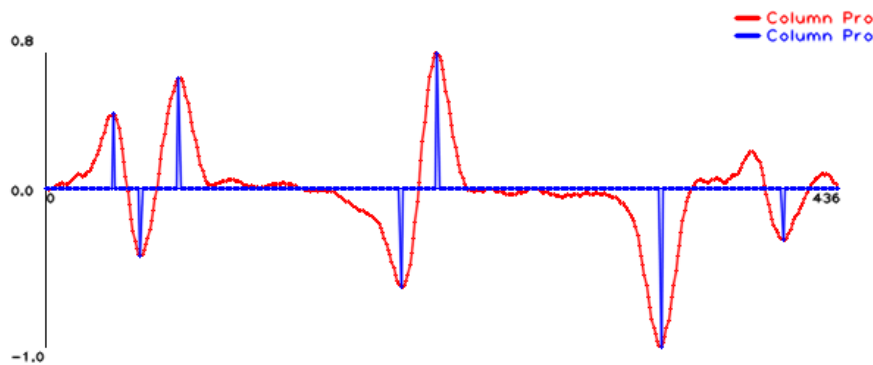


Figura 13: Visualização da derivada perfil da trilha contida na imagem usada no projeto. Fonte: Elaborado pelo autor.

Para enfrentar esse problema foi configurada uma altura mínima em valor absoluto para o pico ser considerado pico (iniciando em 0.3) e além disso, foi determinada uma distância depois do pico encontrado em que não se realizaria o teste, já que é esperado que não hajam 2 picos muito próximos.

Por conta da variação do formato dos picos na derivada do perfil pode haver casos em que um pico ficaria abaixo da altura mínima, perdendo informação valiosa. Com o fim de conseguir detectar alguns picos que passariam despercebidos deixando aquela fatia de tempo com menos bordas do que a média, foi criado um mecanismo de recursão no código em que, caso o número de bordas detectadas naquela fatia de tempo seja menor que a média, é realizada a detecção novamente com uma altura mínima reduzida.

Quanto ao algoritmo de detecção refinada de borda, seu intuito é uma normalização da posição declarada da borda com precisão sub-pixel, ou seja, tendo em mente que o modelo do perfil da trilha é descrito como dois montes, é definido em que parte da inclinação que é correspondente a borda deve ser a referência de sua posição.

Para este algoritmo também é utilizada uma variável empírica, sendo ela o *threshold* que da o nome da técnica na tese referência [2]. No nosso caso, foi escolhido o valor 0.7 pois nos testes realizados com a implementação de seleção de candidato, foi o valor que apresentou menor número de pontos indefinidos.

5.12 Acompanhamento de traço

A fase de acompanhamento de traço tem por finalidade a reunião dos candidatos a ponto de borda mais apropriados em uma estrutura correspondente à trilha do vinil.

5.12.1 Inicialização das trilhas

A inicialização das trilhas começa com o passo de inicialização dos traços, no qual se estabelece uma estimativa dos pontos iniciais de cada borda de cada traço da trilha, isso é feito a partir da média descrita na Seção 5.5.1 e ilustrada na Figura 14. O resultado é o perfil visualizado na Figura 15.

5.12.2 Posições estimadas

As posições estimadas são as referências pelas quais se escolherá o candidato mais apropriado a ponto de borda. A situação esperada é tomar o ponto definido na fatia de tempo anterior para usar como estimativa, porém muitas vezes os defeitos encontrados na superfície do vinil levam a sair desta situação, levando assim a obter a estimativa com base em outras informações. Na Figura 16 (gerada pelo programa, demarcando os pontos relevantes durante o processo) temos um caso em que foi necessário obter a estimativa de

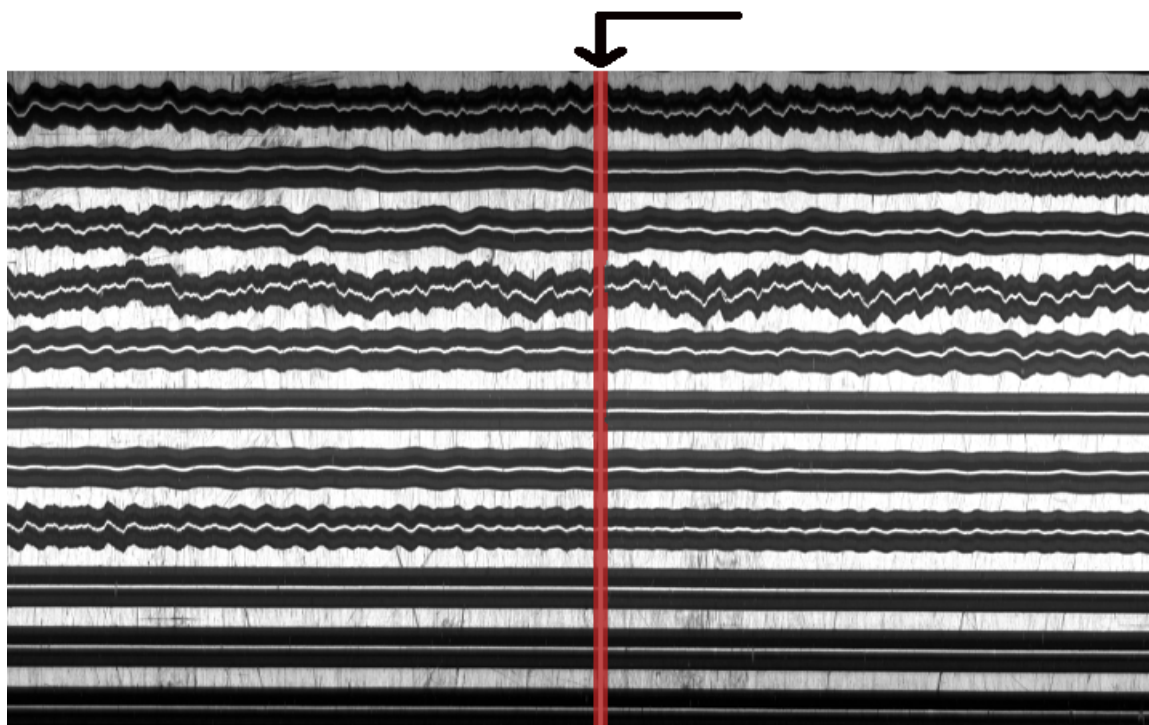


Figura 14: Região em que se realiza a média para obter um perfil suavizado, apontada pela seta. Fonte: Elaborado pelo autor.

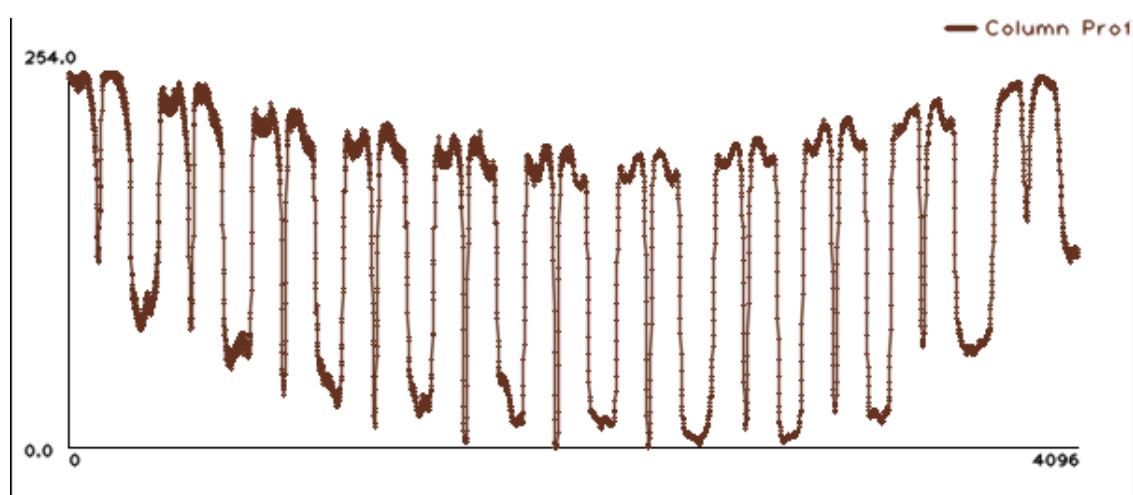


Figura 15: Perfil suavizado utilizado para obter a primeira estimativa de localização da borda. Note a influência da variação de iluminação na imagem expressa pela curvatura no perfil. Fonte: Elaborado pelo autor.

uma borda baseada em outra borda da mesma trilha e na Figura 17 temos um caso em que foi necessário obter uma média dos deslocamentos das trilhas para obter a estimativa.

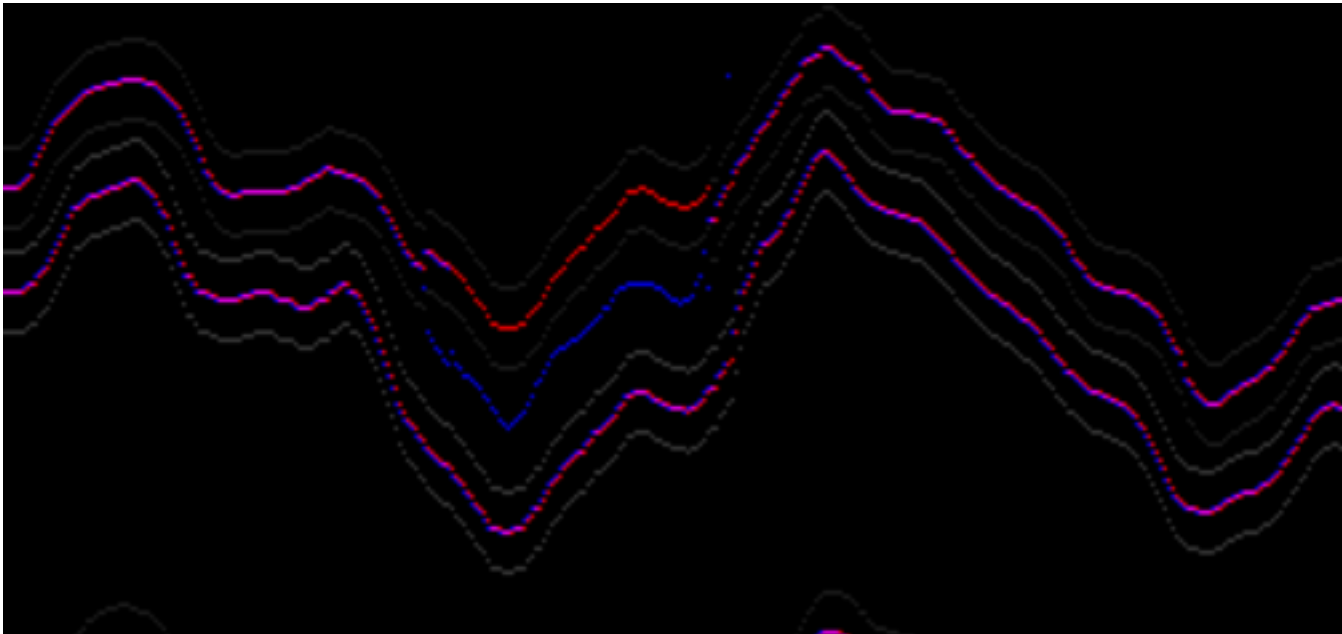


Figura 16: Estimativa baseada no deslocamento das bordas vizinhas já definidas. Em vermelho, estão os pontos que são considerados estimativas na fatia de tempo e em azul, são os pontos candidatos coletados na etapa de detecção de borda. Os pontos rosa são as estimativas que caíram perto dos pontos candidatos. Fonte: Elaborado pelo autor.

5.12.3 Definição do alcance

O alcance da tolerância da estimativa é outro parâmetro que é melhor definido empiricamente. Sua escolha causa grande variação número de pontos indefinidos, isso ocorre porque no acompanhamento da borda ocorrem diversos casos causados por ruído em que o candidato a ponto de uma fatia de tempo possui uma distância consideravelmente maior que a esperada do candidato anterior. Esses ruídos impulsivos causam um desvio do caminho que a trilha seguia e pode comprometer o resto do processo. Nesse trabalho o alcance da tolerância escolhida foi de 8 pixels em volta da estimativa, foi tomado com base em testes com valores diferentes e a avaliação do número de pontos indefinidos e é ilustrado na figura 18.

5.12.4 Seleção de candidatos

Esse é o passo que define (ou marca como indefinido) os pontos de borda do traço. A princípio foi adotado o algoritmo descrito na seção 5.5.4, porém, a função de cálculo da estimativa $j + 1$ quando não há ponto definido no momento j estava tomando um tempo muito grande, o que tornou lento o processo de teste do programa e,

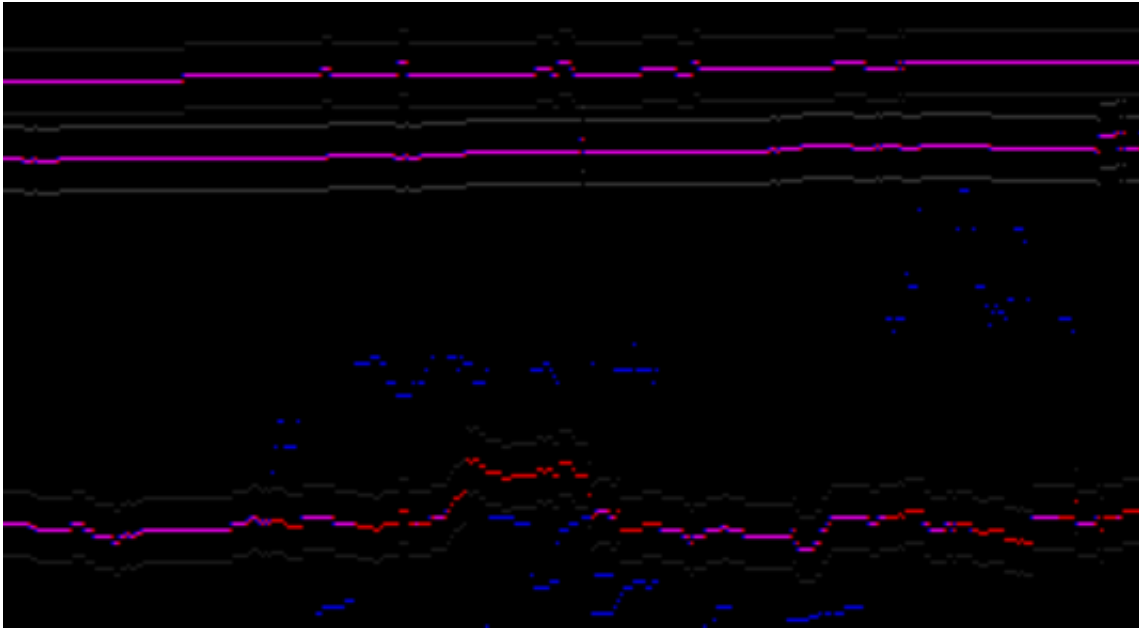


Figura 17: Estimativa baseada na média dos deslocamentos de todas as trilhas já definidas até então. Em vermelho, estão os pontos que são considerados estimativas na fatia de tempo e em azul, são os pontos candidatos coletados na etapa de detecção de borda. Fonte: Elaborado pelo autor.

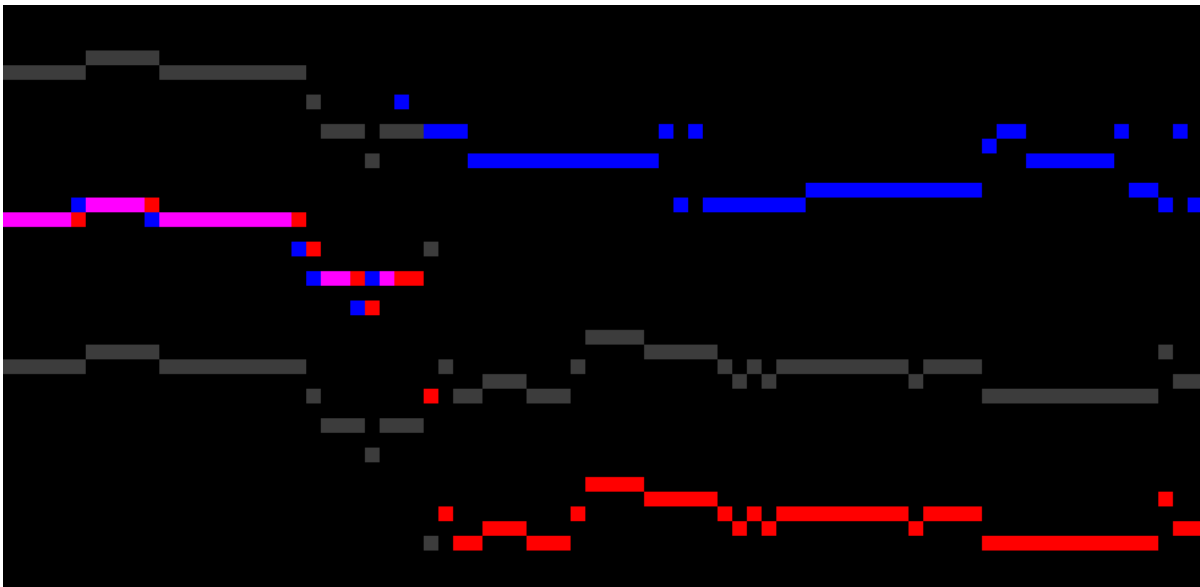


Figura 18: Caso em que há um ruído impulsivo e a linha de estimativas desvia da linha de candidatos. Em vermelho, estão os pontos que são considerados estimativas na fatia de tempo e em azul, são os pontos candidatos coletados na etapa de detecção de borda. Fonte: Elaborado pelo autor.

consequentemente, alongou seu desenvolvimento. Por conta disso esse trecho do processo foi ignorado temporariamente e foi observado que sua presença não era essencial. Os sons gerados nos resultados do projeto não realizaram esta etapa.

5.12.5 Tratamento dos pontos indefinidos

O conjunto das imagens recebidas é a principal referência que guia a eficácia dos algoritmos utilizados. Foi observado ao longo do processo que na maior parte do resultado de acompanhamento de traços, os defeitos que afetavam o curso do mapeamento não afetavam todas as bordas da trilha. Sendo assim e, considerando a premissa de que os discos de vinil monofônicos possuem uma trilha de tamanho aproximadamente fixo (o que se mostrou verdadeiro para o nosso caso), foi adotada a estratégia de avaliar as bordas definidas para o tempo J para se estimar a posição de um ponto que foi marcado como indefinido. Assim, por exemplo, enumerando as bordas de uma trilha de 0 a 3, se um ponto é marcado como indefinido na fatia de tempo J na borda 2, são avaliadas as bordas 3, 0 e 1 (0 e 1 apenas se não tiverem sido redefinidas nesse processo) na procura de um ponto definido durante o processo de seleção de candidatos. No caso de não haver outros pontos definidos para o tempo J , é adotado o ponto definido no tempo $J-1$, porém essas situações ocorreram em pequenos trechos do processo, produzindo um resultado aceitável dentro do escopo do projeto.

5.12.6 União e derivada das posições da trilha

O processo de união dos segmentos de trilha descrito na seção 5.6.1 foi seguido aproximadamente da mesma maneira, mudando a ordem de algumas operações que não necessitam ser comentadas. Já a derivada no caso foi implementada uma versão simples e rápida, devido ao curto espaço de tempo até o fim do projeto. Foi implementada uma derivada que consiste na simples diferença de valores entre uma amostra e sua próxima. Esse processo pode gerar impulsos notáveis no resultado final, porém, visto o tempo limitado, a solução foi muito bem vinda e cumpriu o seu papel, permitindo o reconhecimento do áudio.

5.12.7 Re-amostragem

A imagem recebida, e as imagens resultantes de trabalhos semelhantes a esse também, apresentam uma frequência de amostragem acima das utilizadas para áudio em CD's (44100 KHz). Isso possibilita a utilização do filtro derivativo comentado na seção 5.7, porém, o formato WAV não permite qualquer frequência de amostra para seus conteúdos. Logo, foi necessária uma etapa de re-amostragem do sinal, em que implementou-se o simples algoritmo: $tempo_amostrado = 0$;
 $taxa_velha = 104000$;

```

taxa_nova = 44100;
periodo_velho =  $\frac{1}{taxa_{velha}}$ ;
periodo_novo =  $\frac{1}{taxa_{nova}}$ ;
tempo_total = N_de_amostras * periodo_velho;
groove; // vetor que contém áudio antes da re-amostragem.
reamostrado; // vetor novo
Enquanto(tempo_amostrado < tempo_total)
{
index_continuo = tempo_amostrado/periodo_velho;
index_inteiro = TRANSFORMA_EM_INTEIRO(decimal_index);
delta = index_continuo - index_inteiro;
nova_amostra = groove[index_inteiro] + (groove[index_inteiro+1] - groove[index_inteiro]) *
delta;
reamostrado.adiciona_no_final(nova_amostra);
tempo_amostrado = tempo_amostrado + periodo_novo;
}

```

5.12.8 Mitigação de imperfeições

As imagens utilizadas no projeto são coerentes com os requisitos anteriormente descritos. Porém, existem imperfeições no vinil descrito na imagem que interferem no mapeamento das bordas da trilha. Como se pode ver na figura 20, existem riscos na direção oposta a direção que as trilhas seguem, esses riscos podem desviar o curso durante o acompanhamento de traço, como é ilustrado na Figura 19.

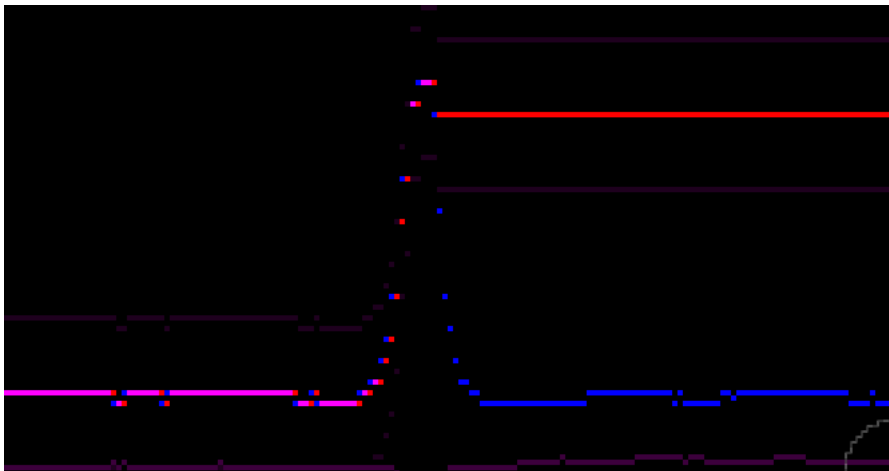


Figura 19: Desvio no curso durante acompanhamento de traço causado por defeito. Os pixels azuis representam os candidatos a borda e os vermelhos, estimativas.
Fonte: Elaborado pelo autor.

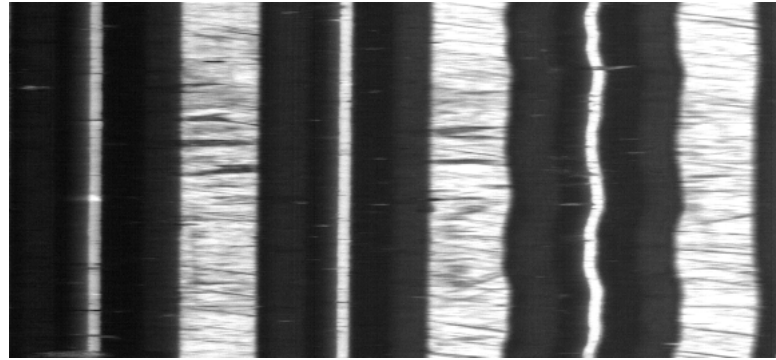


Figura 20: Magnificação na imagem do vinil do projeto (não negativa), destacando os riscos perpendiculares a borda. Fonte: Elaborado pelo autor.

Na expectativa de diminuir os efeitos destas imperfeições em específico, foi utilizado um filtro de borrão de movimento [10] no sentido paralelo ao que a trilha segue. O borrão de movimento atenua a diferença de valores entre o preto dos riscos e o branco da superfície do vinil, de modo que essa variação não seja mais relevante à detecção de bordas utilizada no mapeamento. Há, nos resultados do projeto, uma comparação do perfil do arquivo gerado pela imagem com borrão e sem borrão e, o áudio gerado em cima da imagem tratada com o borrão é consideravelmente menos ruidoso.

O filtro de borrão foi aplicado utilizando o programa *GIMP* [9], por questão de falta de tempo e de materiais de teste para uma automatização do processo.

6 Resultados

O software obteve êxito na exportação de áudio porém, a qualidade ficou comprometida devido as imperfeições contidas na imagem utilizada. Alguns dos defeitos que a imagem possui foram mitigados com um simples filtro de borrão de movimento no sentido paralelo à trilha, porém os problemas mais severos requerem um tratamento mais custoso que o tempo determinado para o trabalho não permitiu explorar.

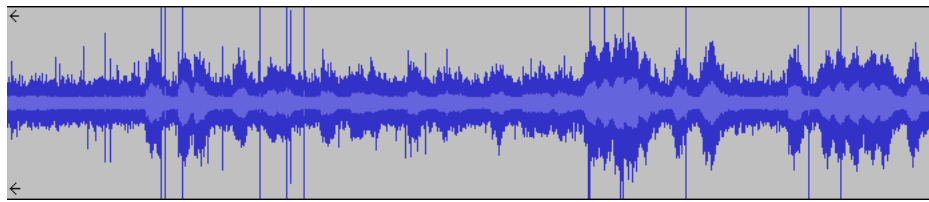


Figura 21: Perfil do áudio obtido pela equipe do professor Carl Haber. Fonte: Elaborado pelo autor.

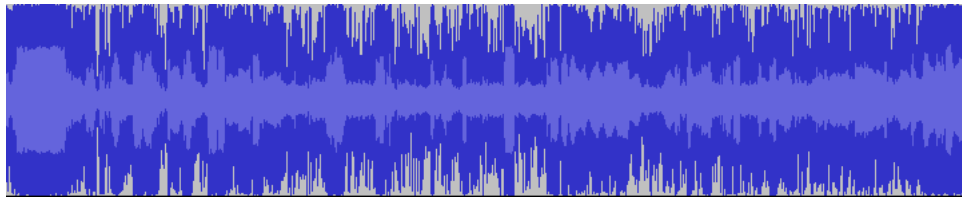


Figura 22: Perfil do áudio obtido pelo software deste trabalho, sem borrão de movimento. Fonte: Elaborado pelo autor.

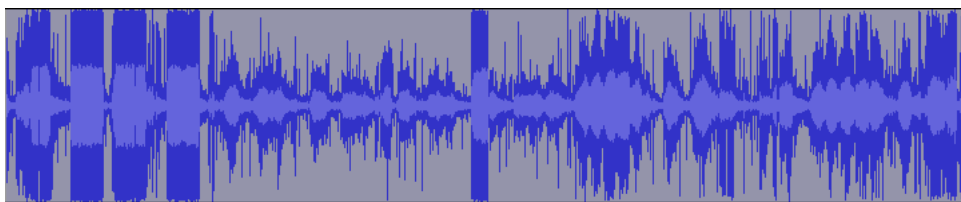


Figura 23: Perfil do áudio obtido pelo software deste trabalho, com borrão de movimento. Fonte: Elaborado pelo autor.

Como podemos observar através dos espectrogramas a seguir, nota-se uma diminuição evidente nas áreas de ruído na imagem em que o borrão foi aplicado.

6.1 Avaliação

A escassez de imagens de testes acabou por limitar o escopo inicial do projeto, não possibilitando almejar a extração de um vinil completo. Porém, espera-se que para atingir

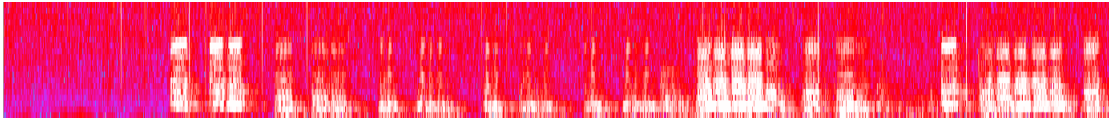


Figura 24: Espectrograma do áudio obtido pela equipe do professor Carl Haber. Imagem obtida pelo programa Audacity [11].Fonte: Elaborado pelo autor.

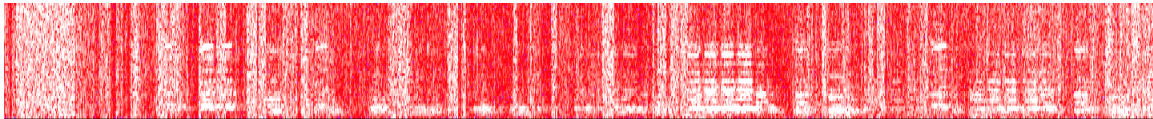


Figura 25: Espectrograma do áudio obtido pelo software deste trabalho, sem borrão de movimento. Imagem obtida pelo programa Audacity [11].Fonte: Elaborado pelo autor.

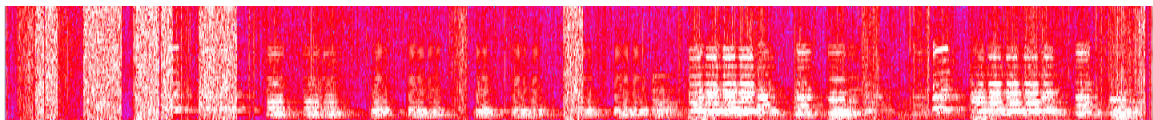


Figura 26: Espectrograma do áudio obtido pelo software deste trabalho, com borrão de movimento. Imagem obtida pelo programa Audacity [11]. Fonte: Elaborado pelo autor.

tal não seja necessário um esforço muito discrepante do investido neste quesito, mas um acréscimo na complexidade logística de manipular uma quantidade grande de dados e um acréscimo enorme no tempo de processamento.

O resultado obtido pelo software é reconhecível quando comparado com o resultado que a equipe do professor Carl Haber obteve. Apesar de serem notáveis as regiões em que as imperfeições do disco comprometeram o áudio original, sua maior parte é comparável com o áudio referência e há a expectativa de que se possa melhorar o processo incluindo umas etapas e melhorando outras.

Uma das etapas que foi deliberadamente escolhida por ser mais simples foi, por exemplo, o processo de derivação das posições da trilha mapeada. Era sabido que seria mais apropriado (como foi descrito na própria Seção 5.7) uma técnica de derivada que tivesse como entrada um contexto mais abrangente do que duas amostras por vez, porém esta decisão foi, como outras foram, primada pela gestão de riscos. A experiência tida com implementações com certo grau de complexidade e tempo reduzido em C++ dita que os problemas vão aparecer e consumirão tempo indeterminado para sua resolução, principalmente referentes ao gerenciamento de memória.

A tese do professor Sylvain [2], possui um estudo muito mais detalhado de mídias fonográficas e diversos ataques aos problemas que as degenerações dos discos podem causar no processo de digitalização. Uma possível solução para os defeitos contingentes contidos

nas imagens enviadas pelo professor Haber seria a adoção do mapa de pixels corrompidos [2](6.3.3), através do qual teria-se uma estimativa das regiões com defeitos entre as trilhas e dentro dos traços, e , tendo esta estimativa, poder-se-ia desconsiderar as informações dos pixels nas regiões identificadas ou aplicar uma correção anteriormente ao mapeamento das bordas.

7 Conclusão

Os resultados obtidos mostram que o objetivo do projeto foi atingido. A extração de áudio por processamento de imagem foi realizada. Claro, não na qualidade que poderia ter atingido caso houvesse um tempo maior, porém os conceitos aplicados foram validados e foi demonstrado através do software desenvolvido que a digitalização por processamento de imagem é viável, possuindo os insumos adequados. Espera-se que este trabalho sirva como base para um entendimento do funcionamento do processo e, quem sabe, sirva de inspiração para projetos mais sofisticados no futuro.

7.1 Dificuldades Encontradas

Sem dúvida a maior dificuldade do projeto foi a obtenção de imagens. A dificuldade dessa etapa foi subestimada no início do trabalho e se perdeu um tempo considerável buscando cumpri-la por conta própria. A sorte foi ter sido atendido pelo professor Carl Haber que gentilmente cedeu as imagens tanto aqui comentadas. Esse evento ocorreu apenas em 23 de julho e sem ele, o escopo do projeto estaria fadado a ser alterado. Passado esse problema, os problemas que mais sugaram esforços ao longo do processo estão vinculados a erros de implementação e os esforços para encontrá-los e corrigi-los e, também, algumas situações em que o entendimento do material fonte se mostrou errado após a aplicação do material no código, um exemplo disso seria o algoritmo de acompanhamento de traço, em que a ordem de algumas etapas não ficou clara no texto original.

7.2 Aprendizado

Com o esforço empregado na realização deste trabalho, foram adquiridos diversos conhecimentos, especialmente sobre o funcionamento das mídias fonográficas, sua variedade e importância. Foi, também, expandida a interface com processamento digital de imagens, com filtros digitais e um pouco de tratamento de áudio, que acabou não sendo aplicado no caso. Mas principalmente, a experiência mais relevante foi a de gestão de tempo exigida para se manter ativo nas outras atividades da escola e dividir esse tempo com a vida pessoal.

8 Bibliografia

- [1] G. Boston. Survey of Endangered Audiovisual Carriers, 2003. UNESCO, Paris, 2003.
- [2] S. Stotzer. Phonographic Record Sound Extraction By Image Processing. PhD Thesis, University of Fribourg, Setembro 2006
- [3] Projeto IRENE, Disponível em: <<http://irene.lbl.gov/>> Acesso em: 30 nov. 2015, 03:42:30.
- [4] Programa Cvplot, Disponível em: <<https://code.google.com/p/cvplot/>> Acesso em: 30 nov. 2015, 03:42:30.
- [5] Programa Makewav, Disponível em: <<https://gasstationwithoutpumps.wordpress.com/2011/10/08/making-wav-files-from-c-programs/>> Acesso em: 30 nov. 2015, 03:42:30.
- [6] V. Fadeyev and C. Haber Reconstruction of Mechanically Recorded Sound by Image Processing, Lawrence Berkeley National Laboratory, Março 2003
- [7] A. Hensman and K. Casey. Optical Reading and Playing of Sound Signals from Vinyl Records, 2007
- [8] Site com as imagens do projeto. Disponível em: <<http://bio16p.lbl.gov/Brazil.html>> Acesso em: 30 nov. 2015, 03:42:30.
- [9] Programa GIMP. Disponível em: <<https://www.gimp.org/>> Acesso em: 30 nov. 2015, 03:42:30.
- [10] Algoritmo de Borrão de movimento. Disponível em: <https://git.gnome.org/browse/gimp/tree/plugin-ins/common/blur-motion\protect\T1\textbracerightc?id=GIMP_2_8_0> Acesso em: 30 nov. 2015, 03:42:30.
- [11] Programa Audacity. Disponível em: <<http://sourceforge.net/projects/audacity/>> Acesso em: 30 nov. 2015, 03:42:30.
- [12] C. Supranowitz, Institute of Optics, University of Rochester. Disponível em: <<http://www.synthgear.com/2014/audio-gear/record-grooves-electron-microscope>>
- [13] Imagem, R. Harley, An LP Primer, HOW THE LP WORKS, 2007
- [14] Imagem do site: <http://www.filmaker.com/papers/pap_RM-Phono%20C-load%20&%20balance.pdf>