



UNIVERSIDADE FEDERAL DO AMAZONAS
FACULDADE DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

SEPARAÇÃO AUTOMÁTICA DE
INSTRUMENTOS DE PERCUSSÃO BRASILEIRA
A PARTIR DE MISTURA PRÉ-GRAVADA

Pedro Donadio de Tomaz Júnior

Manaus – Amazonas

Junho de 2016

Pedro Donadio de Tomaz Júnior

SEPARAÇÃO AUTOMÁTICA DE
INSTRUMENTOS DE PERCUSSÃO BRASILEIRA
A PARTIR DE MISTURA PRÉ-GRAVADA

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal do Amazonas, como requisito parcial para obtenção do título de Mestre em Engenharia Elétrica na área de concentração Controle e Automação de Sistemas.

Prof. D.Sc. Waldir Sabino da Silva Júnior (orientador)

Prof. D.Sc. Luiz Wagner Pereira Biscaíno (coorientador)

Pedro Donadio de Tomaz Júnior

**SEPARAÇÃO AUTOMÁTICA DE
INSTRUMENTOS DE PERCUSSÃO BRASILEIRA
A PARTIR DE MISTURA PRÉ-GRAVADA**

Banca Examinadora

Prof. D.Sc. Luiz Wagner Pereira Biscainho (Presidente)

Universidade Federal do Rio de Janeiro (UFRJ)

Prof. D.Sc. Wallace Alves Martins

Universidade Federal do Rio de Janeiro (UFRJ)

Prof. D.Sc. Diego Barreto Haddad

Centro de Educação Tecnológica Celso Suckow da Fonseca (CEFET/RJ)

Manaus – Amazonas

Junho de 2016

Ficha Catalográfica

Ficha catalográfica elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

T655s Tomaz Junior, Pedro Donadio de
Separação automática de instrumentos de percussão brasileira a partir de mistura pré-gravada / Pedro Donadio de Tomaz Junior.
2016
90 f.: il. color; 31 cm.

Orientador: Waldir Sabino da Silva Júnior
Orientador: Luiz Wagner Pereira Biscainho
Dissertação (Mestrado em Engenharia Elétrica) - Universidade Federal do Amazonas.

1. Separação de fontes. 2. NMF. 3. NMFD. 4. Filtragem do espectrograma por mediana. 5. Sub-bandas. I. Silva Júnior, Waldir Sabino da II. Universidade Federal do Amazonas III. Título

*Aos meus pais Pedro (in
memoriam) e Eneida.
A todos que, assim como eu,
respiram música.*

Agradecimentos

- A Deus;
- Aos meus pais, por terem me dado a oportunidade, apoio e consciência da importância que é seguir sempre aprendendo;
- À Thaísa, pelo apoio técnico e emocional desde o início;
- Aos meus orientadores, aqui listados em ordem alfabética:
 - Prof. Luiz Wagner, por aceitar a orientação e por ser essa fonte inesgotável de conhecimento e inspiração, tanto no processamento digital de sinais como na música. A dedicação aos orientandos surpreende e ratifica, todos os dias, seu compromisso com a educação.
 - Prof. Waldir Sabino, pela orientação acadêmica, profissional e amizade que vem desde os idos tempos de graduação. A maneira calma, serena e “pragmática” de lidar com os problemas é um exercício que eu procuro aprender sempre que possível.
 - Prof. Wallace Martins, pela orientação, paciência e competência técnica transferida a mim. Sem dúvida, uma das referências profissionais e intelectuais que admiro.
- Aos amigos do Laboratório de Processamento de Sinais UFAM. Alguns desde os tempos de graduação, outros de pós-graduação e outros da docência: Kenny, Wheidma, Luiz Eduardo, Igor, Rafael, Iury, Frutuoso, Hussama, Amoedo, Mauro, Mikail, Hiram, Cláudio, Vítor e Vitinho.
- Aos amigos do Laboratório de Sinais, Multimídia e Telecomunicações (SMT) Coppe/Poli/UFRJ que me receberam como se me conhecessem há anos: Mau-

rício, Isabela, Luís Felipe, Rafael, Barboza, Lucas, Matheus, Alan, Igor, Roberto, Camila, Iandra, Tadeu, Maestro, Curicica, Che, Zenam, Cláudio e Jonathan.

- Aos percussionistas que contribuíram com seu talento musical nas gravações da base de dados: Ticva Vinícius, Kayro Robson, Rinaldo Araújo (Batata), Eduardo Brasil e Joabe Santos.
- Ao técnico de som Shakal, pela sensibilidade e compromisso dedicados ao arranjo técnico de captação, organização e edição dos sinais adquiridos, fundamentais para a realização desta parte do trabalho.
- Ao estúdio du Valle, pela qualidade dos equipamentos colocados à disposição.
- Aos alunos, com os quais sigo aprendendo bastante a cada período;
- À Universidade Federal do Amazonas;
- À Universidade Federal do Rio de Janeiro;
- À CAPES, à FAPEAM à SAMSUNG pelo apoio financeiro.

Resumo da Dissertação apresentada à UFAM como parte dos requisitos necessários para a obtenção do grau de Mestre em Engenharia Elétrica

SEPARAÇÃO AUTOMÁTICA DE INSTRUMENTOS DE PERCUSSÃO
BRASILEIRA A PARTIR DE MISTURA PRÉ-GRAVADA

Pedro Donadio de Tomaz Júnior

Orientadores: Prof. D.Sc. Waldir Sabino da Silva Júnior

Prof. D.Sc. Luiz Wagner Pereira Biscainho

Programa: Pós-Graduação em Engenharia Elétrica

Apresenta-se, nesta dissertação, uma investigação sobre separação cega de fontes. O tema, atualmente bastante pesquisado, encontra aplicação em diversas áreas, sendo aqui explorado no contexto musical de instrumentos de percussão presentes na cultura brasileira. São descritas as técnicas de separação utilizadas: a fatoração de matrizes não-negativas (NMF, *non-negative matrix factorization*) e uma de suas extensões (NFMD *non-negative matrix factor deconvolution*), e a filtragem do espectrograma por mediana — que possuem naturezas distintas mas são reunidas neste trabalho para separar misturas percussivas. O termo mistura identifica uma faixa de áudio onde dois ou mais instrumentos tocam juntos. As duas contribuições principais desta dissertação são: a investigação das técnicas de separação de fontes aplicadas às sub-bandas do sinal e a geração de uma base de dados de instrumentos de percussão brasileira. São propostos, ainda, sistemas de separação envolvendo sub-bandas, NMF e filtragem por mediana em diversos arranjos, com o propósito de aprimorar a separação das misturas. Além disso, o efeito da qualidade da mistura sobre o sistema de separação é avaliado. Misturas com alto grau de reverberação são geradas através de *software* profissional de simulação para colocar o sistema à prova.

Palavras-chave: separação de fontes, NMF, NFMD, filtragem do espectrograma por mediana, instrumentos de percussão brasileira e sub-bandas.

Abstract of Dissertation presented to UFAM as a partial fulfillment of the requirements for the degree of Master in Electrical Engineering

AUTOMATIC SOURCE SEPARATION OF BRAZILIAN PERCUSSION
INSTRUMENTS FROM PRE-RECORDED MIXTURE

Pedro Donadio de Tomaz Júnior

Advisors: Prof. D.Sc. Waldir Sabino da Silva Júnior

Prof. D.Sc. Luiz Wagner Pereira Biscainho

Department: Postgraduate in Electrical Engineering

In this work, we present an investigation about blind source separation. Blind source separation, currently widely researched, is applied in many areas of knowledge, being here evaluated in the musical context, with Brazilian percussion instruments. The approached separation techniques are described: the NMF (non-negative matrix factorization) along with its extension NMFD (non-negative matrix factor deconvolution), and the spectrogram median filtering—which are essentially different, but combined in this thesis to separate percussive mixture signals. The term mixture identifies an audio track where two or more instruments are playing together. The two main contributions of this dissertation are: the investigation of blind source separation techniques applied to the signals' subbands and a recorded database of Brazilian percussion instruments. Other source separation systems are proposed; they combine the use of subbands, the NMF, and the median filter in various arrangements to enhance the mixture separations results. Furthermore, the effect of the mixture quality on the system performance are evaluated. Mixtures with high reverberation times have been produced with the aid of a professional simulation software to test the proposed systems.

Keywords: source separation, NMF, NMFD, median filter, brazilian percussion instruments and subbands.

Sumário

Sumário	ix
Lista de Figuras	xi
Lista de Tabelas	xiii
1 Introdução	1
1.1 Processamento digital de sinais de áudio	1
1.2 Separação de fontes	2
1.2.1 Separação cega de fontes	3
1.2.2 Discussão: o que é uma fonte?	4
1.3 Organização da dissertação	5
2 Análise tempo-frequencial	7
3 Separação de fontes: NMF e filtragem por mediana	12
3.1 NMF	12
3.1.1 Medidas de distorção	13
3.2 NMFD	15
3.3 NMF2D	16
3.4 Filtragem do espectrograma por mediana	17
3.5 Recuperação dos sinais separados	20
4 Separação em sub-bandas	22
4.1 Os bancos de filtros	22
4.1.1 Banco de filtros sem decimação	25
4.1.2 Especificações de projeto	26
5 Base de dados	30
5.1 Instrumentos	30
5.2 Ritmos musicais	33
5.3 Ambiente de gravação	34
5.3.1 Gravação das amostras individuais	34

5.3.2	Gravação das misturas acústicas	35
5.3.3	Misturas produzidas a partir de um simulador acústico	36
6	Experimentos	40
6.1	NMFD	41
6.2	Filtragem do espectrograma por mediana	42
6.3	NMFD aplicada nas sub-bandas do sinal	43
6.4	Filtragem do espectrograma por mediana aplicada nas sub-bandas do sinal	45
6.5	Filtragem do espectrograma por mediana, seguida de NMFD	45
6.6	Filtragem do espectrograma por mediana aplicada nas sub-bandas do sinal, seguida de NMFD	46
6.7	Misturas e resultados	48
6.7.1	Caso 1: mistura instantânea de reco-reco e surdo	49
6.7.2	Caso 2: mistura convolutiva de reco-reco e surdo	50
6.7.3	Caso 3: mistura instantânea de pandeiro de couro e tamborim	52
6.7.4	Caso 4: mistura convolutiva de pandeiro de couro e tamborim	53
7	Conclusões e trabalhos futuros	55
7.1	Contribuições	55
7.2	Trabalhos futuros	57
A	Base de dados completa	58
B	Ambiente simulado: materiais e características	66
C	Microfones	70
D	Medidas objetivas	71
	Referências Bibliográficas	74

Lista de Figuras

2.1	Representação no tempo discreto de um trinado tocado ao xilofone, seguido de uma nota tocada ao piano.	8
2.2	Representação na frequência de um trinado tocado ao xilofone, seguido de uma nota tocada ao piano	8
2.3	Exemplo de janelamento realizado em um sinal completo (superior). Abaixo, cada trecho janelado, em separado.	9
2.4	Espectrograma do mesmo sinal musical representado nas Figuras 2.1 e 2.2.	10
2.5	Espectrograma do sinal da Figura 2.4, utilizando uma janela de 4096 pontos.	11
2.6	Janelas de Hann e Hamming com 1024 pontos.	11
3.1	Representação no tempo discreto de duas notas tocadas ao xilofone.	17
3.2	Partes transitória e permanente das notas da Figura 3.1, obtidas pela filtragem do espectrograma por mediana.	18
3.3	Espectrograma do sinal apresentado na Figura 3.1.	18
3.4	Espectrograma das partes transitória e permanente do sinal da Figura 3.1.	20
3.5	Diagrama em blocos da filtragem do espectrograma por mediana.	20
4.1	Distribuição das notas nas teclas de um piano.	23
4.2	Notas grave e aguda tocadas ao piano.	23
4.3	Estrutura do banco de multitaxas em oitavas – análise.	24
4.4	Estrutura do banco de multitaxas em oitavas – síntese.	25
4.5	Estrutura em oitavas do banco de filtros sem decimação.	26
4.6	Representação da resposta em frequência dos filtros H_0 a H_4	28
4.7	Resposta em frequência dos filtros $H_0(z)$ e $H_1(z)$: magnitude e fase.	28
4.8	Resposta em frequência dos filtros $H_2(z)$ e $H_3(z)$: magnitude e fase.	29
4.9	Resposta em frequência do filtro $H_4(z)$: magnitude e fase.	29
5.1	Sala residencial.	38
5.2	Quadra de escola de samba.	38

5.3	Estúdio de gravação.	39
6.1	Diagrama em blocos da NMFD para 2 fontes.	41
6.2	Diagrama em blocos da filtragem por mediana.	42
6.3	Bloco representando a filtragem do sinal com as sub-bandas na saída.	43
6.4	NMFD aplicada às sub-bandas do sinal com 2 componentes por sub-banda.	43
6.5	Processo de síntese do sinal, separado por NMFD aplicada às sub-bandas, para 2 fontes presentes na mistura.	44
6.6	Filtragem do espectrograma por mediana aplicada nas sub-bandas do sinal.	46
6.7	Processo de síntese do sinal, separado por filtragem do espectrograma por mediana aplicada às sub-bandas.	47
6.8	Diagrama em bloco representando a filtragem do espectrograma por mediana, seguido de NMFD para 2 fontes.	47
6.9	Diagrama em bloco representando a filtragem do espectrograma por mediana, seguido de NMFD para 2 fontes.	48
B.1	Sala residencial.	66
B.2	Sala residencial.	67
B.3	Sala residencial.	67
B.4	Estúdio.	67
B.5	Estúdio.	68
B.6	Estúdio.	68
B.7	Quadra.	68
B.8	Quadra.	69
B.9	Quadra.	69

Lista de Tabelas

4.1	Ordem e faixa de transição dos filtros utilizados no banco sem de- cimação.	27
4.2	Limite das sub-bandas geradas.	28
5.1	Lista de instrumentos selecionados.	31
5.2	Agogô.	31
5.3	Caixa.	31
5.4	<i>Shake</i>	31
5.5	Cuíca.	31
5.6	Pandeiro.	32
5.7	Reco-reco.	32
5.8	Repique.	32
5.9	Surdo.	32
5.10	Tantã.	32
5.11	Tamborim.	32
5.12	Padrão de ritmos e andamentos utilizados na gravação da base, com duração de 20 segundos (tempo médio) para cada faixa.	33
5.13	Quantidade de gravações individuais por tipo de instrumento.	35
6.1	Misturas utilizadas.	48
A.1	Misturas com 4 instrumentos.	58
A.2	Misturas com 3 instrumentos.	59
A.3	Misturas com 2 instrumentos.	59
A.4	Gravações individuais - Pandeiro - músico 1.	60
A.5	Gravações individuais - Tamborim - músico 1.	60
A.6	Gravações individuais - Reco-reco - músico 1.	60
A.7	Gravações individuais - Caixa - músico 1.	60
A.8	Gravações individuais - Repique - músico 1.	61
A.9	Gravações individuais - Cuíca - músico 1.	61
A.10	Gravações individuais - Agogô - músico 1.	61
A.11	Gravações individuais - Shake - músico 1.	61

A.12 Gravações individuais - Tantã - músico 1.	61
A.13 Gravações individuais - Surdo - músico 1.	61
A.14 Gravações individuais - Pandeiro - músico 2.	62
A.15 Gravações individuais - Tamborim - músico 2.	62
A.16 Gravações individuais - Reco-reco - músico 2.	62
A.17 Gravações individuais - Caixa - músico 2.	62
A.18 Gravações individuais - Repique - músico 2.	63
A.19 Gravações individuais - Cuíca - músico 2.	63
A.20 Gravações individuais - Agogô - músico 2.	63
A.21 Gravações individuais - Shake - músico 2.	63
A.22 Gravações individuais - Tantã - músico 2.	63
A.23 Gravações individuais - Surdo - músico 2.	63
A.24 Gravações individuais - Pandeiro - músico 3.	64
A.25 Gravações individuais - Tamborim - músico 3.	64
A.26 Gravações individuais - Reco-reco - músico 3.	64
A.27 Gravações individuais - Caixa - músico 3.	64
A.28 Gravações individuais - Repique - músico 3.	64
A.29 Gravações individuais - Agogô - músico 3.	65
A.30 Gravações individuais - Cuíca - músico 3.	65
A.31 Gravações individuais - Tantã - músico 3.	65
A.32 Gravações individuais - Surdo - músico 3.	65
C.1 Microfones utilizados na captação dos instrumentos da base, a 20 centímetros de distância.	70
D.1 Caso 1: SDR, SIR e SAR.	72
D.2 Caso 2: SDR, SIR e SAR.	72
D.3 Caso 3: SDR, SIR e SAR.	72
D.4 Caso 4: SDR, SIR e SAR.	73

Capítulo 1

Introdução

As sensações e estímulos audiovisuais fazem parte da vida do ser humano, que é diariamente submetido a diversos eventos dessa natureza. As informações captadas pelos sensores do corpo e a forma como são interpretadas são tratadas por áreas como a psicologia e a arquitetura, expressam características das relações humanas, sendo formas de distinção entre outras espécies. O campo das artes revela a importância desses estímulos para a humanidade, quando observamos o teatro, a pintura, a escultura e a música como formas de expressão.

Com a evolução das comunicações, os conteúdos audiovisuais tais como imagens, vídeos e registros de fala ou música, armazenados em bancos de dados cada vez maiores, precisam ser acessados de maneira rápida e precisa. O tratamento desse tipo de informação é necessário não só pela sua importância intrínseca, mas também por necessidade de armazenamento e acesso.

Essa breve introdução reforça, de maneira ampla, a importância da área de processamento digital de sinais: o tratamento dado às informações reais no que hoje convencionou-se denominar mundo digital. No mundo real, um indivíduo é capaz de reconhecer visualmente uma pessoa famosa na multidão ou identificar o cantor de uma música somente ouvindo sua voz. Essa tarefa é tão natural que tendemos a subestimar sua complexidade. No entanto, realizar o mesmo trabalho através de um sistema digital nem sempre é tarefa fácil e exige organizar estratégias e algoritmos cada vez mais robustos. Neste trabalho, os sinais de áudio são de especial interesse.

1.1 Processamento digital de sinais de áudio

Nesta dissertação, são utilizados sinais musicais pré-gravados. Canções, arranjos e sinais de instrumentos ou voz têm sido objetos de trabalhos recentes em processamento digital de áudio [1]. Como temas principais, podem ser elencados a transcrição automática [2], que pretende extrair a informação rítmica e tonal dos sinais musicais, a fim de realizar a transcrição musical completa ou de instrumentos

específicos, a detecção de formas e padrões [3], que identifica estilos e temáticas musicais, e a recomendação musical [4], que sugere ao usuário um tipo de música e pode ser entendida como um passo além do processo de identificação de estilos.

No contexto de processamento de sinais de áudio, outro tema explorado pela comunidade científica é a *separação de fontes*. Esse tema é de abordagem complexa, começando pelo próprio conceito de fonte. De fato, tratar uma mistura¹ de sinais para realizar qualquer uma das atividades mencionadas anteriormente (como transcrição, detecção e recomendação musical) apresenta dificuldade superior à de uma outra que contemplasse tão somente sinais isolados. Assim, a separação de fontes serve à área de processamento de sinais de áudio como pré-processamento para muitas aplicações.

1.2 Separação de fontes

O objetivo principal deste trabalho é investigar as técnicas de separação de fontes em sinais musicais. Ocorre que a separação de fontes não é utilizada exclusivamente no contexto de sinais de áudio. Algumas aplicações em imagens são bastante comuns. Na área médica, interessa, por exemplo, separar o som do batimento cardíaco do bebê, ainda no ventre, do batimento cardíaco da mãe [5]. Na geologia, os métodos sísmicos de exploração de petróleo vêm sendo utilizados em conjunto com técnicas de separação de fontes, para avaliações com maior precisão [6].

Normalmente, a separação de fontes sonoras é apresentada a partir da descrição de um problema clássico conhecido por *cocktail party*: em um ambiente como uma festa onde várias pessoas conversam, um ouvinte ficar atento a apenas uma das falas, capturando a informação daquele discurso e dispensando as demais. O ouvido humano, em conjunto com o cérebro, consegue fazer esse trabalho de maneira eficiente e natural, por exemplo, quando o som do tráfego intenso de carros das grandes cidades, com todos os seus ruídos característicos, é ignorado e a música do *CD player* fica evidente; ou quando, simplesmente, é possível falar ao telefone dentro de um *shopping* lotado. Em geral, duas abordagens podem ser adotadas quando a separação das fontes é realizada: a primeira diz respeito à separação de um sinal para ser reproduzido; nesse caso, o sinal separado deve ser o mais próximo do sinal (da fonte) original. A segunda abordagem diz respeito à extração de informação sobre a(s) fonte(s) da mistura; nesse caso, a identificação de características presentes no sinal da fonte é importante, mas a reprodução/preservação do sinal não é necessária.

Os métodos usuais de separação de fontes baseiam-se em algum dos 3 paradigmas,

¹O termo *mistura* aparece ao longo do texto e, daqui para a frente, refere-se a uma faixa musical (sinal), de um canal, onde duas ou mais fontes tocam ao mesmo tempo.

independência estatística, esparsidade e não-negatividade [7].

No primeiro caso, destaca-se o método de ICA (*independent component analysis*), que pressupõe que as fontes presentes no sinal de mistura são independentes. Em seguida, motivado pelo teorema central do limite, considera-se que a soma de duas ou mais variáveis aleatórias não gaussianas apresenta comportamento mais próximo de uma distribuição gaussiana em comparação com a distribuição de cada variável individual; as fontes separadas são representadas pelas variáveis com distribuições de menor semelhança com a gaussiana. Uma barreira crítica encontrada por este método reside no fato de exigir, no mínimo, tantos sensores quantos forem as fontes. Representa o segundo caso o método *sparse coding*, que aproveita a esparsidade dos sinais, assumindo que cada componente permanece ativa somente em alguns instantes de tempo ou faixas de frequência. Em termos estatísticos, isso representa uma probabilidade alta de o ganho de uma componente ser igual a zero. Em outras palavras, a condição de esparsidade, aplicada geralmente sobre o ganho do sinal, significa que apenas um instrumento (fonte) está tocando em determinada região em tempo/frequência, favorecendo, assim, a separação.

Atualmente, um método da terceira categoria em franco desenvolvimento é o método de **fatoração de matrizes não-negativas** (NMF, do inglês *non-negative matrix factorization*). Este será parte importante do presente trabalho.

1.2.1 Separação cega de fontes

Os tipos de abordagem de separação de fontes podem ser divididos em duas categorias: a separação supervisionada e a separação não supervisionada, também conhecida por separação cega (do inglês *blind source separation*). A separação supervisionada admite a utilização de informação a priori, permitindo que o sistema conheça algumas variáveis relativas às fontes envolvidas para auxiliar a separação. Esta informação a priori compreende um conjunto de dados (conhecido como conjunto de treinamento), de onde são extraídas as variáveis (ou características) que, para os sinais de musicais, podem representar informações espectral, temporal, harmônica ou sobre a energia. Logo, por similaridade, é possível, a partir das características conhecidas, identificar os instrumentos (ou fontes, de maneira mais ampla) presentes em determinada mistura. No caso da separação cega, pouca ou nenhuma informação a priori é utilizada, o que dificulta o processo de separação. A dificuldade é ainda maior em um contexto musical com um grande número de instrumentos. Além disso, a quantidade de microfones, a posição das fontes em relação aos microfones, a qualidade das misturas (convolutivas ou instantâneas), a reverberação da sala de gravação, entre outros, são fatores decisivos que determinam uma maior ou menor facilidade na separação. Músicos com certa experiência são capazes de transcre-

ver arranjos inteiros de cada instrumento presente em peças musicais consideradas de elevado grau de dificuldade, executadas por bandas ou orquestras com diversos naipes distintos de instrumentos. Esse fato exemplifica a capacidade do cérebro humano de interpretar separadamente os eventos sonoros percebidos. Nesse exemplo, ainda há a tarefa de transcrição musical, que não será abordada neste trabalho, mas que também é tema de pesquisas em processamento de sinais musicais. O desafio é treinar o sistema digital para separar instrumentos de um trecho musical, no caso do sistema supervisionado ou, como no caso deste trabalho, procurar definir um sistema que a partir de poucas informações sobre a mistura de entrada e com base nas técnicas propostas, consiga separar de maneira satisfatória as fontes envolvidas. O termo “satisfatória” diz respeito, principalmente, à avaliação subjetiva das faixas de áudio geradas na saída do sistema. Espera-se que o ouvinte possa afirmar que a fonte recuperada é igual à, ou próxima da fonte original. Obviamente, na falta dos sinais antes de serem misturados, a mistura original pode ser usada como referência para permitir a comparação sinal da mistura \times sinal separado.

Apesar de a separação tratada neste trabalho ser a separação cega de fontes, é preciso que haja uma delimitação nos sinais utilizados. A investigação desse trabalho é sobre a separação de instrumentos de percussão comuns na música popular brasileira, doravante descritos como instrumentos de percussão brasileira, por simplicidade. As faixas utilizadas (uma das contribuições desta dissertação) foram gravadas por músicos profissionais em estúdio profissional. Os diversos arranjos e misturas utilizados foram realizados para representação fiel de um conjunto percussivo tocando em ambientes comuns onde esse tipo de música é executado. Trabalhar a separação de fontes, nesse caso, é separar a sonoridade dos instrumentos envolvidos. Mas, o conceito de fonte é um tanto complexo. A Seção 1.2.2 sugere uma discussão pertinente ao tema.

1.2.2 Discussão: o que é uma fonte?

O conceito de fonte tem diversas interpretações. No exemplo clássico do *cocktail party*, a fonte é relacionada ao som (voz) emitido por uma pessoa. Independentemente de se esta pessoa varia ou não a forma de falar ao longo do tempo, para o ouvinte a voz de uma pessoa é tida como uma fonte. Este exemplo simples pode, a princípio, não deixar dúvidas sobre o conceito de fonte. No entanto, quando outras misturas sonoras são observadas, definir o que é uma fonte torna-se uma tarefa não trivial. Ao observar um violão, podemos considerar o instrumento inteiro como uma fonte única. Porém, o violão, composto tradicionalmente por 6 cordas, pode ser considerado como 6 fontes distintas, dependendo da interpretação dada. No caso do piano, o som é produzido através de várias cordas esticadas no corpo do

instrumento, percutidas por um pequeno martelo que é acionado por uma tecla. Com 88 teclas, é possível considerar cada tecla sua como uma fonte, de maneira análoga ao exemplo do violão. Mas algumas teclas, quando acionadas, atingem 2 ou 3 cordas simultaneamente. A percepção de uma nota para o ouvinte é exatamente a mistura do som produzido por 3 cordas distintas. Seriam três fontes emissoras, então? No caso da percussão, as mesmas situações são observadas. Um exemplo de instrumento utilizado na base de dados apresentada no Capítulo 5 é o pandeiro. O pandeiro, que é composto por uma pele que reveste toda parte superior do aro e por peças metálicas (platinelas) presas à lateral do instrumento, admite a mesma interpretação dos exemplos anteriores. Os sons de características distintas podem ser tratados como fontes distintas. Outro caso é do repique de anel, cujo som é produzido através do toque no revestimento de couro na parte superior do instrumento e do toque do anel na parte lateral.

Os exemplos citados deixam claro que é importante definir o que está sendo tratado como fonte. No decorrer deste trabalho, o termo fonte fica explícito para cada ocasião. Em geral, o interesse é representar, em separado, os instrumentos presentes na mistura. Logo, o pandeiro ou o repique de anel representam uma fonte cada um. Outras questões sobre como os algoritmos de separação lidam com esses eventos são tratadas em momentos oportunos ao longo do texto.

1.3 Organização da dissertação

Esta dissertação apresenta os temas investigados ao longo da pesquisa da seguinte forma.

No Capítulo 2 são apresentados conceitos e definições sobre o pré-processamento dos sinais de áudio. A representação do sinal num formato em que seja possível localizar seu conteúdo frequencial e sua ocorrência temporal é apresentada sob o título de representação tempo-frequencial. Ferramentas como a transformada de Fourier de tempo curto e a construção do espectrograma são destacadas como itens primordiais no tratamento de sinais musicais. Os passos posteriores a esta fase podem ser ajustados adequadamente a partir de uma escolha de parâmetros consentâneos à aplicação desejada. A configuração utilizada, como a escolha do janelamento e a quantidade de pontos da transformada, é destacada neste capítulo.

No Capítulo 3, as técnicas de separação de fontes estudadas são descritas em detalhes. A NMF e sua extensão, a NMFD (*non-negative matrix factorization deconvolution*) estão na base dos métodos atualmente mais utilizados em problemas de separação de fontes em sinais musicais. Uma alternativa engenhosa e simples a eles, a filtragem do espectrograma por mediana (*median filter*, em inglês) é utilizada quando se deseja distinguir as componentes transitórias das permanentes no

espectrograma da mistura.

No Capítulo 4, uma parte importante da proposta do trabalho é apresentada: a separação em sub-bandas. O emprego desta estratégia é justificado, e discute-se a melhor escolha de bancos de filtros para o trabalho.

No Capítulo 5, a segunda contribuição: a descrição da base de dados de percussão brasileira gerada. A escolha dos instrumentos e suas variações, baseadas nas gravações de ritmos tipicamente brasileiros, como o samba, a marcha e o samba de enredo, são cuidadosamente listados. As faixas gravadas dos instrumentos individualmente ou em conjunto e as misturas geradas a partir de um simulador acústico são contempladas neste capítulo.

No Capítulo 6, os experimentos realizados utilizando a base de dados e os métodos propostos são comparados. A condução dos experimentos visam à avaliação dos métodos existentes e da metodologia proposta.

No Capítulo 7, são descritos as conclusões e os trabalhos que podem ser gerados a partir dos resultados obtidos nessa dissertação.

No Apêndice A, a lista com todas as faixas gravadas para a base de dados e o modelo de microfone usado para gravar cada instrumento são mostrados. Também são relacionados a configuração do ambiente do *software* de simulação acústica para cada ambiente e os resultados objetivos obtidos para os diversos experimentos realizados.

Capítulo 2

Análise tempo-frequencial

Um passo anterior a qualquer tipo de processamento aplicado em sinais de áudio é encontrar uma representação que favoreça a análise dessa classe de sinais. As maneiras mais simples de representação desses sinais são a representação no domínio do tempo ou a representação no domínio da frequência. Cada um dos domínios carrega representações distintas da mesma informação. No domínio do tempo, por exemplo, os ataques dos instrumentos presentes em um sinal musical ficam nítidos. Entretanto, é difícil determinar que instrumento ou nota foi tocada em determinado instante de tempo. A Figura 2.1 ilustra a representação no domínio do tempo de um trecho de sinal de música.

Uma outra forma de se analisar um sinal consiste em verificar suas componentes de frequência. A transformada de Fourier¹ é a ferramenta matemática mais utilizada para levar um sinal no domínio do tempo para o domínio da frequência. Um exemplo de representação no domínio da frequência é observado na Figura 2.2.

Essa também não é uma representação completamente interpretável. As raias de frequência e suas respectivas amplitudes são observadas, mas o instante de tempo em que cada uma das notas foi efetivamente tocada não é evidenciado.

Uma representação do sinal que carrega ambas as informações é chamada de representação tempo-frequencial e pode ser obtida através da transformada de Fourier de tempo curto STFT (*short-time Fourier transform*). Para que as informações de tempo e de frequência possam coexistir na representação, o sinal no tempo é dividido em trechos menores, de igual tamanho (*frames*). Esse processo, conhecido como janelamento, é descrito matematicamente como

$$x_m^w[n] = x[n]w[n - m], \quad (2.1)$$

onde m é o deslocamento da janela $w[n]$, que percorre todo o sinal original $x[n]$. A

¹Em processamento digital, é utilizada a transformada discreta de Fourier (DFT - *discrete Fourier transform*).

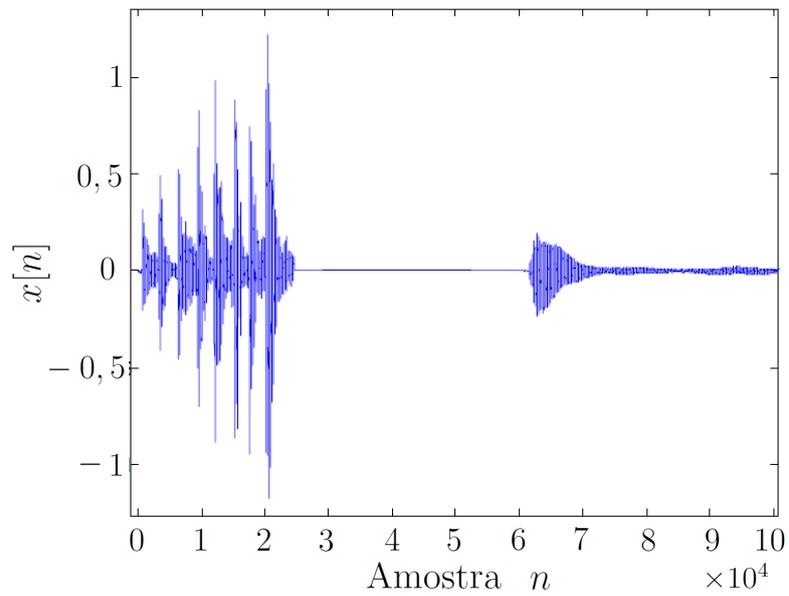


Figura 2.1: Representação no tempo discreto de um trinado tocado ao xilofone, seguido de uma nota tocada ao piano. As transições mais abruptas são os exatos momentos em que cada nota é tocada. O intervalo entre a última nota executada pelo xilofone e a nota executada pelo piano é de aproximadamente 1 segundo.

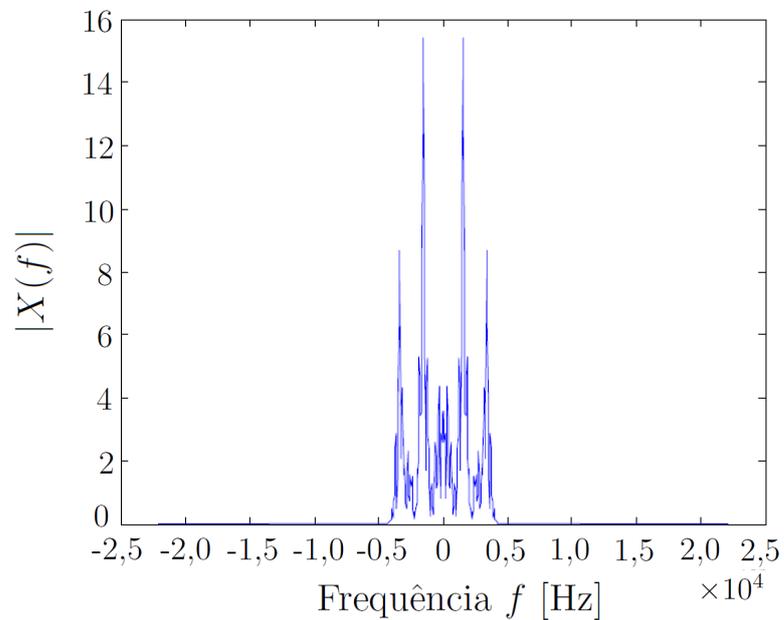


Figura 2.2: Representação na frequência de um trinado tocado ao xilofone, seguido de uma nota tocada ao piano (o mesmo sinal da Figura 2.1).

DFT (*discrete Fourier transform*) aplicada a cada trecho “janelado” do sinal original é a STFT $X_m[k]$, expressa por

$$X_m[k] = \sum_{n=0}^{N-1} x_m^w[n] e^{-j\frac{2\pi}{N}nk}, \quad (2.2)$$

onde N é o comprimento do sinal e k é o índice da frequência discreta, com $0 \leq k \leq N - 1$.

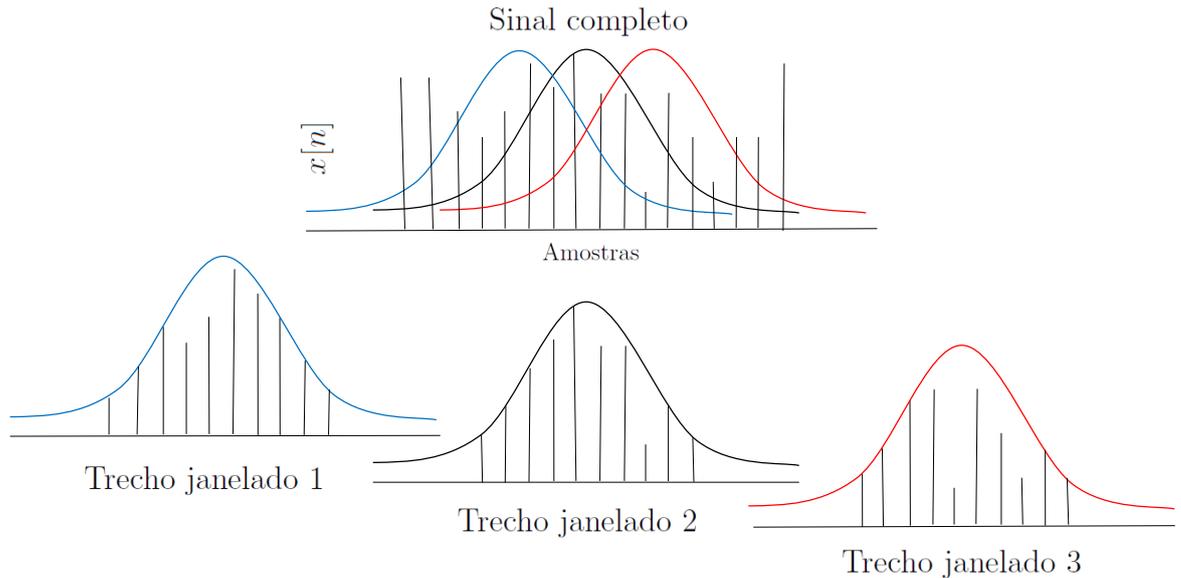


Figura 2.3: Exemplo de janelamento realizado em um sinal completo (superior). Abaixo, cada trecho janelado, em separado. As amostras e o formato do sinal são guiados pelo tipo de janela utilizada. A partir daí, é calculada a DFT de cada trecho.

Os resultados $X_m[k]$ podem ser organizados em uma matriz de dimensão $N \times M$, onde N representa a quantidade de pontos da DFT e M a quantidade de janelas utilizadas para dividir o sinal original. Matrizes com os valores de $|X_m[k]|$ em suas colunas são chamadas de **espectrogramas de magnitude** e matrizes com valores de $|X_m[k]|^2$ são chamadas de **espectrogramas de potência**. Uma representação do espectrograma de magnitude (o mais utilizado em processamento de áudio) é mostrada na Figura 2.4.

Assim, a representação do sinal de áudio fica mais coerente com a análise a ser feita: as raias de frequências ativas são, agora, observadas no intervalo de tempo em que ocorrem. Um fator importante dessa representação é o tamanho da janela utilizada para a segmentação do sinal. Assim como o sinal de interesse é visto no domínio do tempo e no domínio da frequência, a janela também influencia a representação do sinal em ambos os domínios. Isso porque uma janela grande no tempo implica

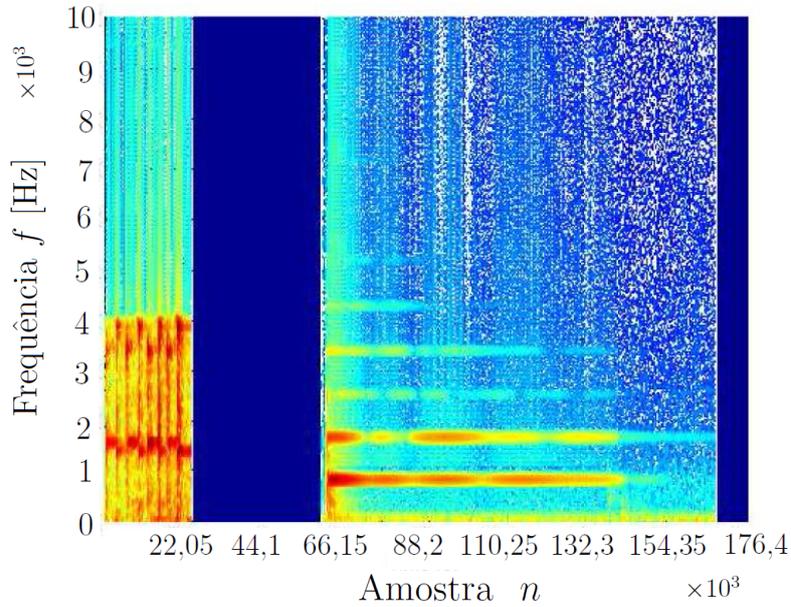


Figura 2.4: Espectrograma do mesmo sinal musical representado nas Figuras 2.1 e 2.2. As regiões de maior energia são representadas em vermelho. As faixas em azul escuro representam o intervalo (silêncio) entre a execução das notas.

uma boa resolução na frequência. No caso oposto, uma janela muito seletiva no tempo (janela muito pequena) tem, no domínio da frequência, uma resolução baixa. Esse critério de escolha do tamanho da janela pode ser exemplificado observando-se a transformada de Fourier das janelas propostas. Se a janela é tão pequena quanto um impulso, o que representa uma excelente resolução no tempo, no domínio da frequência essa janela compreende todas as frequências possíveis. Em contrapartida, uma janela de tamanho infinito no tempo (compreendendo o sinal inteiro), representa um impulso na frequência (máxima resolução frequencial). O efeito do tamanho da janela pode ser visto comparando-se os espectrogramas representados na Figura 2.4, onde foi utilizada uma janela de 512 pontos, e na Figura 2.5, onde foi utilizada uma janela de 4096 pontos.

Outro detalhe é a escolha do tipo da janela. As janelas retangulares, por conta das bordas abruptas, são evitadas em certos contextos, já que os cortes impostos aos sinais a representar tendem a introduzir conteúdo espúrio significativo em altas frequências. As janelas preferidas para as aplicações são as janelas de *Hann* ou de *Hamming* [8], mostradas na Figura 2.6. Para este trabalho, foram utilizadas as janelas de *Hamming*.

As janelas cujas bordas têm transições menos abruptas que as janelas retangulares geram, no domínio de frequência, uma atenuação maior nos lobos secundários, à esquerda e à direita do lobo principal. Se as janelas, quando aplicadas ao sinal inteiro, fossem do tipo retangular, o sinal janelado não apresentaria nenhum desvio

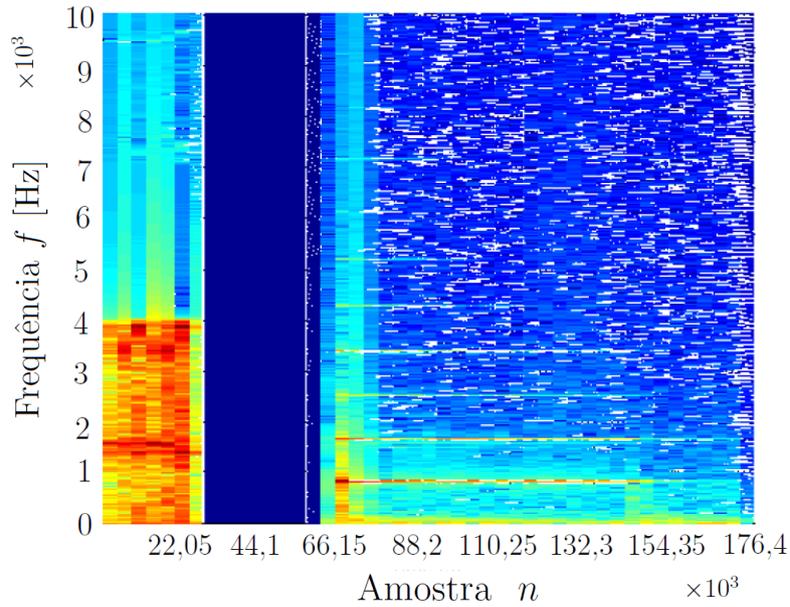


Figura 2.5: Espectrograma do sinal da Figura 2.4, utilizando uma janela de 4096 pontos. É possível notar as diferenças de resolução entre as duas figuras: com 4096 pontos a resolução na frequência é melhor, enquanto a resolução temporal fica comprometida.

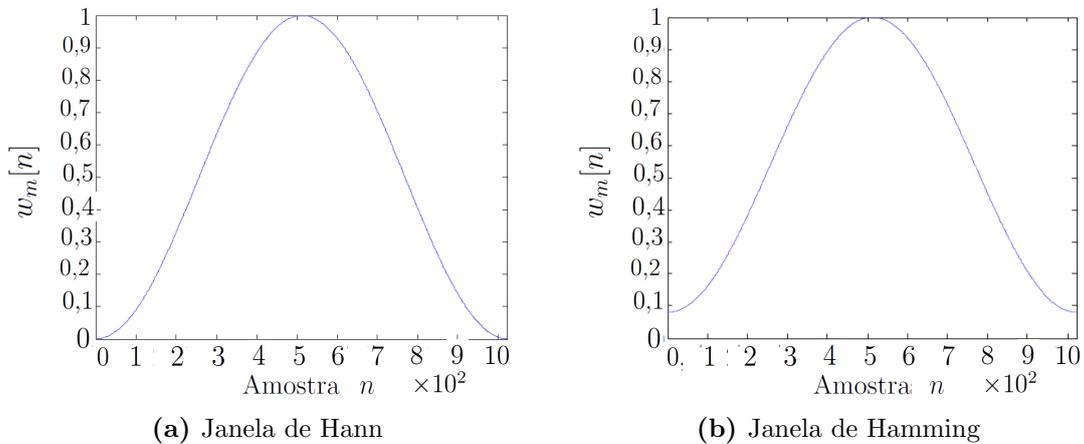


Figura 2.6: Janelas de Hann e Hamming com 1024 pontos.

em relação à sua magnitude. No entanto, quando é feita a opção por janelas com bordas suavizadas, pelos motivos expostos anteriormente, com exceção da parte central da janela, o restante do sinal fica atenuado. Para compensar esta distorção, as janelas devem ser sobrepostas de forma complementar, de modo que nenhuma parcela do sinal seja representada diferentemente das demais. No caso das janelas aqui consideradas, isso se consegue com uma escolha judiciosa do passo do deslocamento das janelas, normalmente variado entre 50 ou 75 por cento do tamanho da janela.

Capítulo 3

Separação de fontes: NMF e filtragem por mediana

Os métodos apresentados neste capítulo utilizam estratégias diferentes para separação das fontes. A NMF se presta a misturas musicais genéricas pela associação das diferentes fontes a padrões espectro-temporais que se repetem ao longo do espectrograma da mistura. Já a filtragem por mediana do espectrograma da mistura permite segregar porções predominantemente esparsas no tempo das porções predominantemente esparsas na frequência, permitindo com isso separar, por exemplo, fontes harmônicas de percussivas. A motivação da seleção dos métodos, bem como a descrição detalhada de cada um deles, são apresentadas a seguir.

3.1 NMF

A NMF (*non-negative matrix factorization*) é o método mais explorado na atualidade para tratar o problema de separação de fontes acústicas, principalmente por apresentar a capacidade de separar uma mistura captada por apenas um sensor (microfone). Alguns trabalhos recentes utilizam a NMF em problemas de classificação de instrumentos de percussão [9] e de separação de instrumentos musicais [10] e [11]. Em [12] e [13] a separação de instrumentos musicais é aprimorada adicionando-se restrições à NMF.

O método baseia-se na fatoração do espectrograma de magnitude do sinal da mistura em duas outras matrizes não-negativas. A característica da não-negatividade tem por objetivo dar sentido físico ao modelo matemático, representando magnitudes na frequência e ganhos positivos. Além disso, com os valores não-negativos garante-se que não haverá cancelamento de componentes por interferência destrutiva. O modelo para o sinal da mistura é formado pela combinação de uma matriz contendo uma base de padrões no domínio da frequência (\mathbf{B}) e de uma matriz com

os ganhos de cada vetor da base ao longo dos quadros, ou *frames* (\mathbf{G}), de forma que:

$$\mathbf{X} \approx \hat{\mathbf{X}} = \mathbf{B}\mathbf{G}, \quad (3.1)$$

onde \mathbf{X} é o espectrograma real da mistura e $\hat{\mathbf{X}}$ é a estimativa de \mathbf{X} . As matrizes correspondentes aos espectrogramas de magnitude possuem dimensão $N \times M$, onde N representa a quantidade de raias (ou canais) de frequência e M a quantidade de quadros (*frames*) de tempo utilizados para a segmentação do sinal. \mathbf{B} e \mathbf{G} têm dimensões $N \times Q$ e $Q \times M$, respectivamente, onde Q é o número de vetores da base frequencial.

Interpretando a matriz \mathbf{B} com mais cuidado, é possível identificar algumas peculiaridades. As colunas de \mathbf{B} representam os vetores da base frequencial \mathbf{b}_q (base frequencial) que aparecem no sinal ao longo do tempo. Para cada quadro do sinal, uma diferente combinação de vetores da base é produzida. Essa combinação é resultado da ponderação feita pela matriz de ganho \mathbf{G} . Em cada linha \mathbf{g}_q da matriz \mathbf{G} está representada a evolução do ganho de cada vetor \mathbf{b}_q ao longo do tempo.

A multiplicação de cada vetor \mathbf{b}_q por seu respectivo ganho em \mathbf{g}_q é chamada de componente. A soma de todas as componentes ao longo de todos os quadros de tempo gera o espectrograma $\hat{\mathbf{X}}$. Essa característica é interessante e válida o fato de a não-negatividade das matrizes ser a essência do método. A soma de todas as componentes (não-negativas por definição) produz o espectrograma como resultado final, sem que nenhum sinal negativo acabe por anular a contribuição de uma componente.

3.1.1 Medidas de distorção

A solução do problema reside em encontrar matrizes \mathbf{B} e \mathbf{G} , não-negativas, tais que a medida de distorção entre o espectrograma da mistura \mathbf{X} e sua estimativa $\hat{\mathbf{X}}$ seja mínima, ou seja, $\mathbf{B}\mathbf{G}$ deve ser o mais próximo possível de \mathbf{X} . Existem algumas medidas de distorção para essa abordagem. A mais simples é o quadrado da distância euclidiana, descrita por

$$D_e = \frac{1}{2} \left\| \mathbf{X} - \hat{\mathbf{X}} \right\|_F^2, \quad (3.2)$$

onde $\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^n \sum_{j=1}^k |a_{ij}|^2}$ é a norma de Frobenius. Outra medida bastante comum para aferir tal distância é a divergência de Kullback-Leibler [14], descrita por

$$D_{\text{KL}} = \left| \mathbf{X} \odot \ln \left(\frac{\mathbf{X}}{\hat{\mathbf{X}}} \right) - \mathbf{X} + \hat{\mathbf{X}} \right|, \quad (3.3)$$

onde \odot representa a multiplicação matricial ponto a ponto (produto de Hadamard) e $|\cdot|$ indica a soma de todos os elementos. A divisão entre matrizes também é ponto a ponto e $\ln(\cdot)$ atua sobre cada elemento da matriz. A divergência de Kullback-Leibler é, de fato, uma medida de divergência estatística entre a probabilidade “real” e a probabilidade que se quer aproximar. Por ser uma “medida” não simétrica, já que $D_{\text{KL}}(u||v) \neq D_{\text{KL}}(v||u)$, onde u e v são medidas de probabilidade, não é tratada como distância, mas como divergência. A divergência de Kullback-Leibler tem sido preferida para esse tipo de aplicação. O motivo é que, por ser uma função linear, diferentemente da distância euclidiana (quadrática), é mais sensível a baixas energias do sinal.

Com nenhuma informação *a priori* e para solucionar o problema de minimização da distorção, o algoritmo de otimização mais usualmente empregado é o do gradiente descendente [15], que, por atuar em um problema convexo, pode ser implementado mediante a atualização alternada das matrizes \mathbf{B} e \mathbf{G} . Para a distância euclidiana, as regras de atualização são [14]

$$\mathbf{B}_{i+1} = \mathbf{B}_i \odot \frac{\mathbf{X}\mathbf{G}_i^T}{\mathbf{B}_i\mathbf{G}_i\mathbf{G}_i^T} \quad (3.4)$$

$$\mathbf{G}_{i+1} = \mathbf{G}_i \odot \frac{\mathbf{B}_{i+1}^T\mathbf{X}}{\mathbf{B}_{i+1}^T\mathbf{B}_{i+1}\mathbf{G}_i^T}, \quad (3.5)$$

onde i indica a iteração de atualização do algoritmo.

Para a divergência de Kullback-Leibler, as regras de atualização são

$$\mathbf{B}_{i+1} = \mathbf{B}_i \odot \frac{\frac{\mathbf{x}}{\tilde{\mathbf{x}}_i}\mathbf{G}_i^T}{\mathbf{1}\mathbf{G}_i^T} \quad (3.6)$$

$$\mathbf{G}_{i+1} = \mathbf{G}_i \odot \frac{\mathbf{B}_{i+1}^T\frac{\mathbf{x}}{\tilde{\mathbf{x}}_i}}{\mathbf{B}_{i+1}^T\mathbf{1}}, \quad (3.7)$$

onde $\mathbf{1}$ é uma matriz $N \times M$ com todos os elementos iguais a 1. Assim, podemos utilizá-las alternadamente até a convergência das matrizes \mathbf{B} e \mathbf{G} ou até um máximo de iterações pré-estipulado para o algoritmo. O Algoritmo 1 sintetiza todas as etapas do método NMF pela divergência de Kullback-Leibler (adotada neste trabalho).

A NMF possui alguns desdobramentos que auxiliam na separação de fontes específicas, como é o caso da NMF_D (*non-negative matrix factor deconvolution*) [16] e da NMF_{2D} (*non-negative matrix factor 2-D deconvolution*) [17].

Algoritmo 1: Algoritmo NMF, utilizando a divergência de Kullback-Leibler.

Entrada: Matriz \mathbf{X} e número de instrumentos Q **Saída:** Matrizes \mathbf{B} e \mathbf{G} não-negativas**início**

Inicializar as matrizes com valores não-negativos;

 Atualizar a matriz \mathbf{G} , segundo equação (3.7); Atualizar a matriz \mathbf{B} , segundo equação (3.6);

Repetir os passos 2 e 3 até atingir a convergência ou o número máximo de iterações.

fim

3.2 NMFD

A NFMD (*non-negative matrix factorization deconvolution*) é uma extensão da NMF. A formulação da NMF pressupõe que um padrão espectral ocupa apenas um quadro do sinal. Mas nem sempre a evolução das frequências é invariante. Em alguns casos, os padrões frequenciais se comportam de maneira variável ao longo do tempo, e para serem mais bem descritos precisam ocupar mais de um quadro temporal. A ideia por trás do método é que seja possível a representação de uma nota de um instrumento. Por exemplo, quando uma nota de piano ou xilofone é tocada, seu padrão espectral varia no tempo. Ambas as notas possuem um ataque, seguido de uma sustentação (aproximadamente) harmônica. Essa evolução espectral é representada através de matrizes de bases espectrais, diferentemente da NMF, que representa os padrões como vetores. [18] aborda a transcrição de sinais de bateria através da NMFD.

A NMFD apresenta uma modificação na matriz de ganho \mathbf{G} , que sofre deslocamentos.

$$\mathbf{X} \approx \hat{\mathbf{X}} = \sum_{l=0}^{\tau-1} \mathbf{B}^l \overset{\rightarrow l}{\mathbf{G}}, \quad (3.8)$$

onde τ é o numero de quadros permitidos para cada componente e o operador \rightarrow representa a operação de deslocamento horizontal dos elementos de uma matriz. Matematicamente, há um deslocamento de colunas da matriz de ganhos \mathbf{G} na direção da seta \rightarrow , acrescentando tantas colunas de zeros quantos forem os deslocamentos. Exemplificando, temos

$$\overset{\rightarrow 0}{\mathbf{Y}} = \begin{bmatrix} 4 & 9 & 7 \\ 2 & 1 & 8 \\ 3 & 5 & 2 \end{bmatrix} \quad \overset{\rightarrow 1}{\mathbf{Y}} = \begin{bmatrix} 0 & 4 & 9 \\ 0 & 2 & 1 \\ 0 & 3 & 5 \end{bmatrix} \quad \overset{\rightarrow 2}{\mathbf{Y}} = \begin{bmatrix} 0 & 0 & 4 \\ 0 & 0 & 2 \\ 0 & 0 & 3 \end{bmatrix},$$

Na matriz à esquerda, o deslocamento igual a zero não altera a matriz, independente do sentido da seta. Na matriz ao centro, o deslocamento igual a 1 para a direita

faz com que a coluna 3 seja retirada da matriz: as colunas 1 e 2 passam a ocupar as colunas 2 e 3, respectivamente, e a coluna 1 é preenchida com zeros. A matriz à direita sofre um deslocamento de 2 para a direita, fazendo com que as colunas 2 e 3 sejam retiradas da matriz; a coluna 1 passa a ocupar a coluna 3 da matriz e as colunas 1 e 2 são preenchidas com zero. De maneira análoga acontece para o sentido oposto da seta.

A matriz \mathbf{B}^l possui dimensão $N \times M$ e representa os padrões espectrais das componentes para um quadro l . A matriz $\overset{\rightarrow}{\mathbf{G}}$ possui dimensão $M \times T$. Dessa forma, é possível em tese representar uma nota como uma componente, ainda que haja variação do seu padrão frequencial ao longo do tempo.

As restrições de não-negatividade permanecem para a NMFD e as equações de atualização de \mathbf{B} e \mathbf{G} são, agora, com os deslocamentos temporais, descritas como

$$\mathbf{B}_{i+1}^l = \mathbf{B}_i^l \odot \frac{\overset{\rightarrow}{\mathbf{x}}_i (\overset{\rightarrow}{\mathbf{G}}_i^T)}{\mathbf{1} (\overset{\rightarrow}{\mathbf{G}}_i^T)} \quad (3.9)$$

$$\mathbf{G}_{i+1}^l = \mathbf{G}_i^l \odot \frac{\mathbf{B}_{i+1}^{lT} (\overset{\leftarrow}{\mathbf{x}}_i)}{\mathbf{B}_{i+1}^{lT} \mathbf{1}}. \quad (3.10)$$

As equações são atualizadas alternadamente, como na NMF. A diferença é que na NMFD as matrizes precisam ser atualizadas para todos os valores de l . Isso porque, enquanto na NMF os padrões frequenciais eram representados por vetores \mathbf{b}_q (coluna da matriz \mathbf{B}) agrupados em uma matriz \mathbf{B} , na NMFD cada padrão frequencial passa a ser uma matriz, parte de um tensor \mathbf{B} . O ganho de cada matriz \mathbf{B}_l é variado ao longo do tempo deslocando-se a matriz \mathbf{G} , l vezes¹.

3.3 NMF2D

A NFM2D (*non-negative matrix factor 2-D deconvolution*) apresenta mais uma extensão da NMF, desta vez permitindo utilizar o mesmo padrão tempo-frequencial para as diferentes notas emitidas por um instrumento, introduzindo deslocamentos verticais na matriz \mathbf{B} . Isso é particularmente útil em instrumentos melódicos. Como o presente trabalho se debruça sobre percussão não melódica, não utilizará a NMF2D.

¹Para verificar a dedução das equações multiplicativas da NMFD para a divergência de Kullback-Leibler, consultar [19].

3.4 Filtragem do espectrograma por mediana

Um método bastante engenhoso e simples para separação de fontes é o de filtragem do espectrograma por mediana. Os trabalhos sobre o método de filtragem do espectrograma por mediana pretendem separar o sinal musical em harmonia e percussão, como em [20]. O ponto de partida, assim como os métodos apresentados anteriormente, é o espectrograma de magnitude do sinal de mistura. Essa representação tempo-frequencial do sinal musical é capaz de destacar as características dos instrumentos harmônicos e dos instrumentos percussivos. Antes, porém, de apresentar o funcionamento do método, é importante identificar dois conceitos fundamentais: as componentes transitórias e as componentes permanentes ou tonais (no sentido de apresentarem altura definida) do sinal. Na Figura 3.2 são mostradas as partes transitória e permanente de uma nota musical.

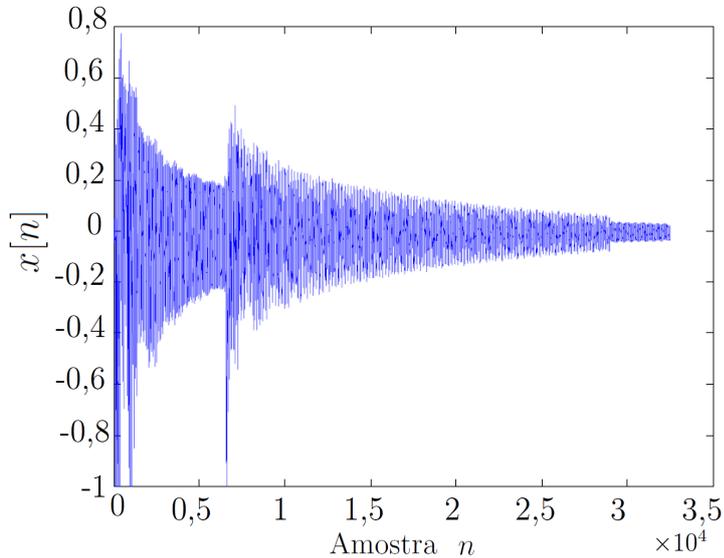


Figura 3.1: Representação no tempo discreto de duas notas tocadas ao xilofone.

A partir desse conceito, o espectrograma do sinal é modelado como a soma das partes transitória e permanente do sinal

$$|X(n, m)| = |X_t(n, m)| + |X_p(n, m)|, \quad (3.11)$$

onde a porção transitória $|X_t(n, m)|$ é esparsa no tempo e suave na frequência e a porção permanente $|X_p(n, m)|$ é esparsa na frequência e suave no tempo.

As regiões com grande concentração de energia no tempo (linhas verticais mais escuras no espectrograma) são classificadas como transitórias. As regiões que possuem a concentração de energia em uma raia de frequência (linhas horizontais mais escuras no espectrograma) são classificadas como permanentes. É comum denominar as partes transitória e permanente como percussiva e harmônica, respectivamente. De

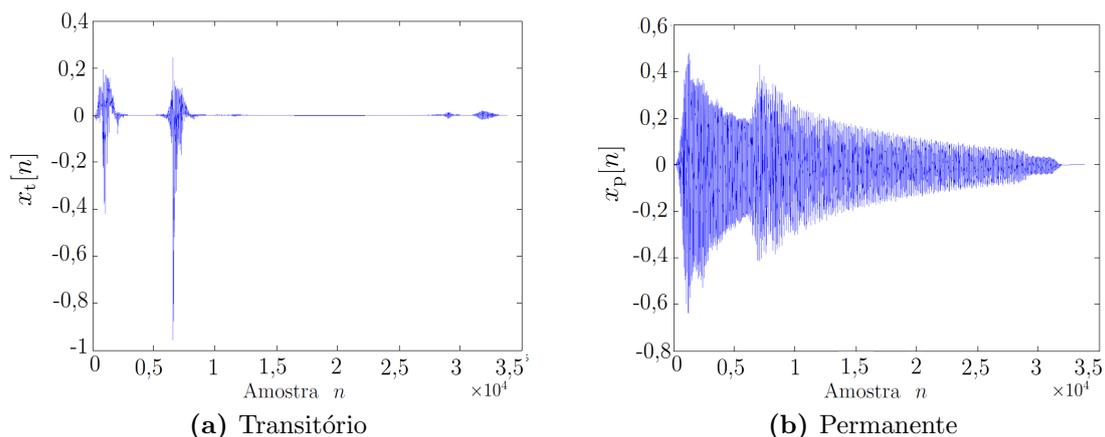


Figura 3.2: Partes transitória e permanente das notas da Figura 3.1, obtidas pela filtragem do espectrograma por mediana.

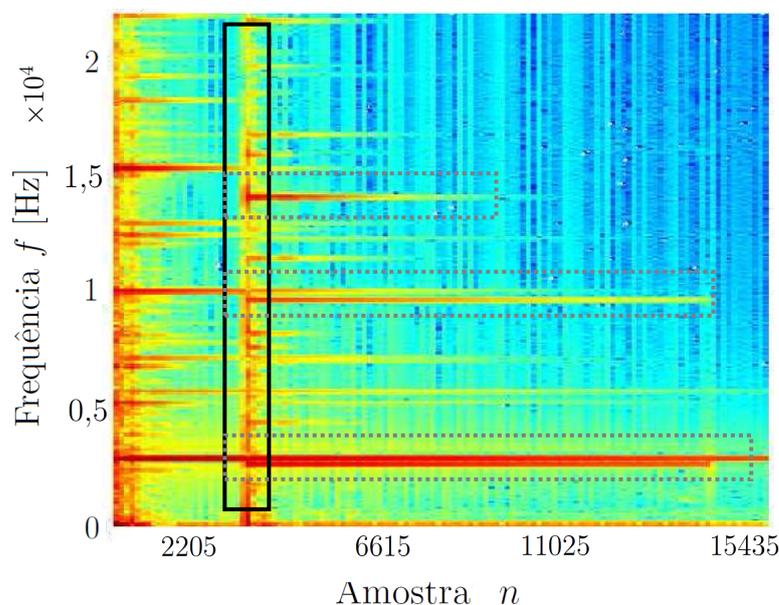


Figura 3.3: Espectrograma do sinal apresentado na Figura 3.1. Destacada em preto, as concentrações de energia no tempo, representando o ataque da segunda nota. Na linha pontilhada cinza, as concentrações de energia na frequência, representando a parte harmônica da mesma nota.

fato, em geral, instrumentos de percussão produzem linhas verticais bem definidas no espectrograma, enquanto os sons harmônicos tendem a produzir linhas horizontais. No entanto, deve-se ter cuidado com essa generalização, pois, por exemplo, em uma nota de piano (que se pretende separar como parte harmônica da mistura), o ataque é caracterizado verticalmente no espectrograma (transitório), enquanto que o restante do sinal é caracterizado como permanente. Da mesma forma, um instrumento de percussão como o surdo também pode possuir uma porção permanente

bastante longa.

A filtragem por mediana, do ponto de vista matemático, realiza a varredura do sinal nas duas direções: tempo (horizontal) e frequência (vertical). O algoritmo de filtragem por mediana estabelece duas janelas de tamanhos iguais a $2N_p + 1$ e $2N_t + 1$ que são movidas por todo o eixo do tempo e da frequência, respectivamente, de amostra em amostra, substituindo o valor central da janela pela mediana dos valores por ela abrangidos. Isso faz com que, ao final da varredura em ambos os eixos, o algoritmo retorne dois espectrogramas separados, representando, como exposto anteriormente, a parte permanente (que resultou da filtragem ao longo do tempo, a qual eliminou as linhas verticais) e a parte transitória (que resultou da filtragem ao longo da frequência, a qual eliminou as linhas horizontais) do sinal. O processo é descrito no Algoritmo 2 e no diagrama da Figura 3.5.

Algoritmo 2: Algoritmo *Median Filter*

Entrada: $x[n]$, N_p e N_t

Saída: $x_t[n]$ e $x_p[n]$

início

 Calcular $X[n, m] = \text{STFT}(x[n])$

para todo $i = 1$ até n_{\max} **faça**

para todo $j = 1$ até m_{\max} **faça**

$\hat{X}_p[i, j] \leftarrow \text{median}[|X[i - N_p : i + N_p, j]|]$

$\hat{X}_t[i, j] \leftarrow \text{median}[|X[i, j - N_t : j + N_t]|]$

fim

fim

$M_t \leftarrow \frac{X_t^2}{X_p^2 + X_t^2}$

$M_p \leftarrow \frac{X_p^2}{X_p^2 + X_t^2}$

$X_t(n, m) = M_t \odot X(n, m)$

$X_p(n, m) = M_p \odot X(n, m)$

$x_t[n] \leftarrow \text{STFT}^{-1}(X_t(n, m))$

$x_p[n] \leftarrow \text{STFT}^{-1}(X_p(n, m))$

fim

A Figura 3.4 apresenta dois espectrogramas separados pelo algoritmo de filtragem por mediana. A entrada do sistema é o sinal cujo espectrograma foi apresentado na Figura 3.3.

A princípio, este método é usado em [21] e [20] para separar sinais cuja mistura possui sons percussivos e harmônicos. A investigação da filtragem por mediana neste trabalho, que utiliza uma base de dados composta essencialmente de instrumentos de percussão, permite avaliar em especial os efeitos do método sobre misturas que contenham instrumentos como cuíca e agô, os quais possuem *pitch* definido.

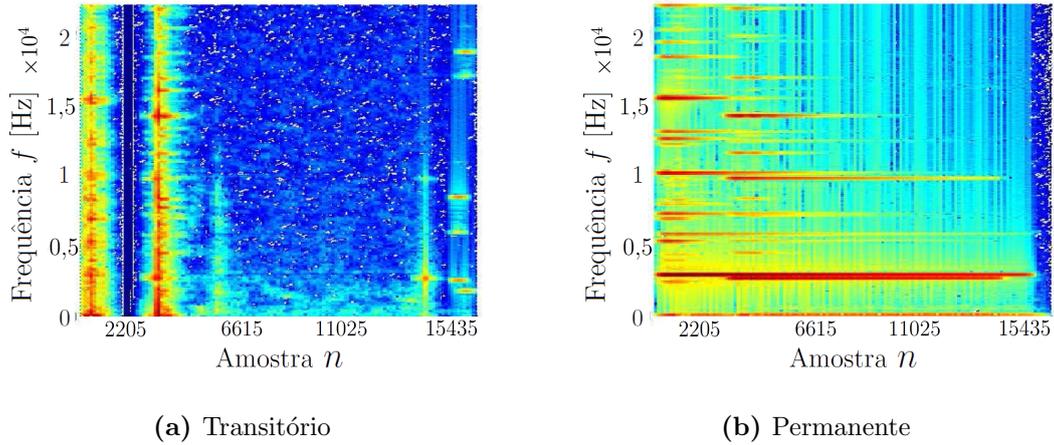


Figura 3.4: Espectrograma das partes transitória e permanente do sinal da Figura 3.1. Resultados obtidos após a filtragem do espectrograma do sinal original por mediana.

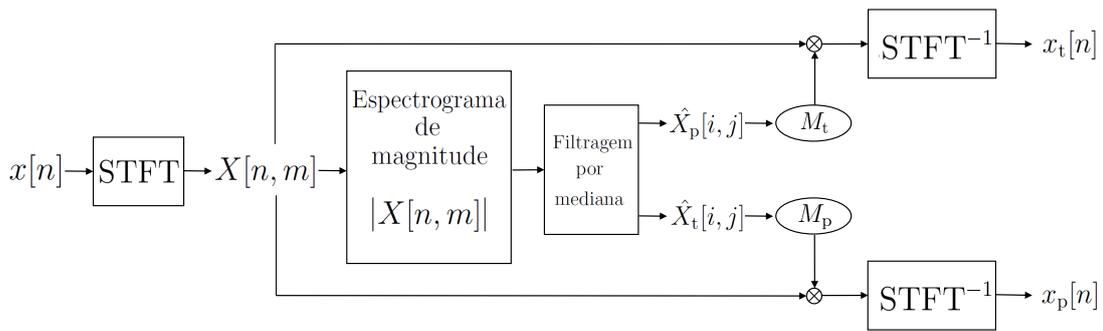


Figura 3.5: Diagrama em blocos da filtragem do espectrograma por mediana.

3.5 Recuperação dos sinais separados

Com exceção do método de filtragem do espectrograma por mediana, as demais técnicas apresentadas descartam a informação de fase do sinal. No caso da filtragem, a máscara calculada é aplicada sobre a mistura, da qual as informações de magnitude e fase da mistura são conservadas. Porém, para os métodos de fatoração, apenas o espectrograma de magnitude do sinal original \mathbf{X} é utilizado para o cálculo de \mathbf{B} e \mathbf{G} . Além disso, o espectrograma obtido após a fatoração é uma estimativa do espectrograma do sinal original.

Para superar este problema, algumas técnicas são utilizadas. Dentre elas, a utilização da fase da própria mistura original é uma solução possível. Entretanto, quando tratamos sinais musicais, esta solução simples pode não ser a mais indicada. Os ataques sucessivos dos instrumentos ficam mais imprecisos quanto maior for a

diferença entre o espectrograma original e o estimado.

Outro algoritmo utilizado na tentativa de recuperar a informação de fase é o algoritmo de Griffin e Lim (G&L) [22]. O algoritmo tenta encontrar um sinal real no domínio do tempo cujo espectrograma de magnitude seja mais próximo possível de $\hat{\mathbf{X}}$. O algoritmo processa o sinal por inteiro e utiliza, para recuperação da fase de um quadro, a informação de quadros passados e futuros, sendo inviável para aplicações em tempo real.

No sentido de contemplar implementações em tempo real, o algoritmo *real-time iterative spectrogram inversion* (RTISI) [23] foi desenvolvido. O sinal, agora, é estimado quadro a quadro, levando em consideração apenas a informação de quadros anteriores. O custo computacional é minimizado em relação ao algoritmo G&L .

O método utilizado neste trabalho é uma extensão do RTISI conhecida como RTISI-LA (*real-time iterative spectrogram inversion with look-ahead*) [24]. Sua principal vantagem em relação ao método RTISI é a versatilidade. Enquanto no RTISI somente o quadro atual e quadros passados influenciavam na estimação, no RTISI-LA quadros futuros (*look-ahead*) contribuem no cálculo. A versatilidade diz respeito exatamente à quantidade de quadros *look-ahead* utilizados. Se a aplicação demanda rapidez, poucos quadros (ou nenhum) são utilizados no *look-ahead*. Se a qualidade da reconstrução é priorizada, mais quadros futuros podem ser utilizados.

Capítulo 4

Separação em sub-bandas

Uma contribuição deste trabalho é a investigação da separação de fontes nas diferentes sub-bandas do sinal. A estratégia consiste em separar o sinal em faixas limitadas de frequência, utilizando as técnicas de separação em trechos com menor conteúdo espectral. Com isso, a ideia é que o conteúdo reduzido de cada sub-banda possa facilitar o trabalho da NMF ou da filtragem do espectrograma por mediana. Para separar o espectro do sinal em sub-bandas, existem na literatura diversas técnicas de filtragem, como exposto em [8]. Para este trabalho e com o objetivo de avaliar a separação de fontes principalmente de maneira perceptiva (subjativa), os bancos de filtros foram projetados especificamente para aplicação nos sinais musicais.

4.1 Os bancos de filtros

Sob os aspectos musicais, os critérios adotados, desde a organização da estrutura do banco no espectro até o tipo de filtro utilizado, priorizaram a avaliação subjetiva das fontes separadas. Nesse sentido, as sub-bandas foram escolhidas com tamanho variável, com faixas mais estreitas nas baixas frequências e faixas mais largas nas altas frequências. A explicação para esse tipo de escolha encontra resposta na percepção da altura (*pitch*) pelo ouvido humano: a partir de certa frequência, diferenças de altura percebidas como iguais correspondem a saltos geométricos na frequência.

Essa percepção está na raiz da organização das escalas musicais. A escala musical ocidental convencionada possui 12 notas (alturas) por oitava. Em um instrumento como o piano, é possível observar a distribuição das notas da escala ocidental entre as teclas brancas e pretas.

Matematicamente, o intervalo entre duas notas adjacentes (um semitom) da escala ocidental obedece a relação de $\sqrt[12]{2}$ isto é, o cálculo da frequência de uma nota k semitons acima da nota de f_n é dado por

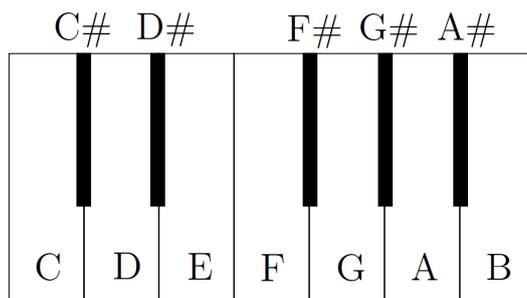


Figura 4.1: Distribuição das notas nas teclas de um piano: C=dó, C#=dó sus-
tenido, D=ré, D#=ré sus-tenido, E=mi, F=fá, F#=fá sus-tenido, G=sol, G#=sol
sus-tenido, A=lá, A#=lá sus-tenido e B=si.

$$f_{n+k} = f_n (\sqrt[12]{2})^k. \quad (4.1)$$

A afinação (frequência) de cada nota é, por convenção, estabelecida a partir do A4 (primeiro lá acima do dó central do piano), que corresponde a 440 Hz. Assim, ainda para o exemplo do piano, a frequência de qualquer nota pode ser calculada, considerando a numeração das teclas de -48 a 39, pela equação

$$f_n = 440 \times 2^{\frac{n}{12}}. \quad (4.2)$$

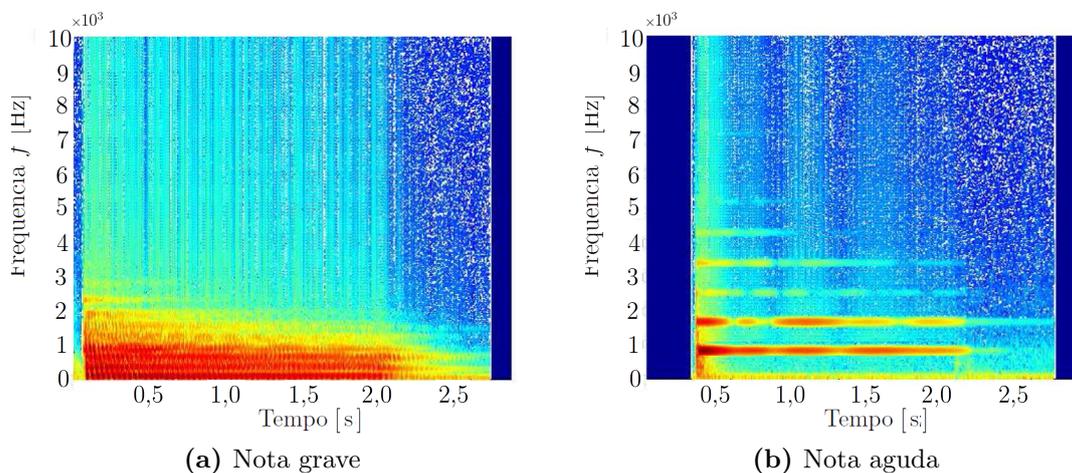


Figura 4.2: Notas grave e aguda tocadas ao piano. Notam-se em (a) as raias de frequência, fundamental e demais harmônicas, bem próximas umas das outras. Em (b) as raias aparecem mais espaçadas.

Toda essa discussão sugere uma organização de um banco de filtros apropriado para dividir o espectro de um sinal musical em bandas geometricamente crescentes com a frequência. Uma escolha extremamente simplificadora é separar sucessivamente a oitava superior do espectro. Por exemplo, um sinal cujo espectro ocupa até

20 kHz seria separado em faixas de 10 a 20 kHz, de 5 a 10 kHz e assim sucessivamente até a banda inferior desejada.

A decisão seguinte foi quanto a trabalhar com bancos de filtros multitaxa ou não.

Sabe-se que após ser separado em 2 meias bandas por um passa-baixas e um passa-altas, um sinal originalmente amostrado à taxa f_s pode ser representado por suas duas versões filtradas, cada uma à taxa $f_s/2$ [8]. Utilizando-se a operação de decimação por M (aproveitamento de apenas 1 em cada M amostras), representada por $\downarrow M$, é possível reduzir a taxa de amostragem de um sinal por um fator inteiro M . Utilizando-se a operação de interpolação por L (inserção de $L - 1$ zeros entre cada duas amostras consecutivas), representada por $\uparrow L$, é possível aumentar a taxa de amostragem de um sinal por um fator inteiro L . Acrescentando-se as filtragens correspondentes (idênticas a cada estágio) e seguindo essa filosofia (que minimiza a necessidade de processamento), os bancos de filtros de análise (em que se dá a redução e taxa) e de síntese (em que se dá a recondução à taxa original) organizados em oitavas assumiriam as formas das Figuras 4.3 e 4.4, respectivamente.

Ocorre que pela natureza dos experimentos a serem realizados no presente trabalho, será necessário a todo momento combinar faixas de frequências distintas. Com o uso de uma estrutura em multitaxa como a descrita anteriormente, isso geraria um *overhead* considerável na complexidade da organização do sistema. Por esse motivo, optou-se pelo uso de um banco de filtros na taxa original do sinal.

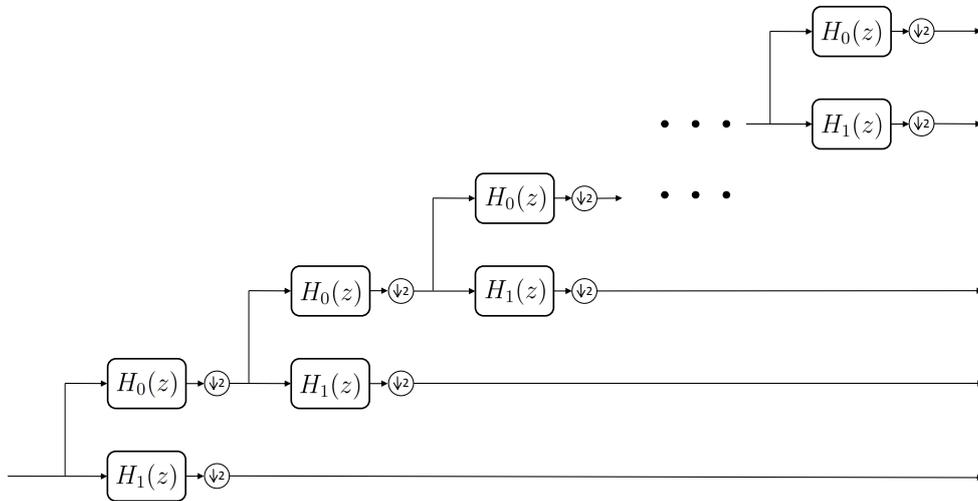


Figura 4.3: Estrutura do banco de QMF em oitavas – análise. $H_0(z)$ representa o filtro passa-baixas e $H_1(z)$ o filtro passa-altas.

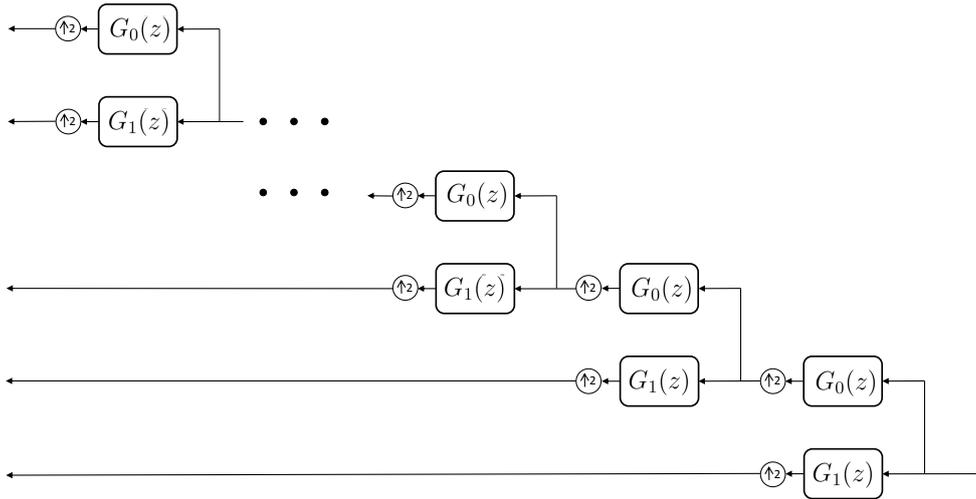


Figura 4.4: Estrutura do banco de QMF em oitavas – síntese. $G_0(z)$ representa o filtro passa-altas e $G_1(z)$ o filtro passa-baixas.

4.1.1 Banco de filtros sem decimação

Em cada estágio dos bancos de filtros em multitaxa, os filtros passa-baixas e passa-altas projetados podem ser os mesmos dos empregados no estágio anterior. Isso porque os filtros sempre separam a banda do sinal em duas metades. Em outras palavras, a ordem do filtro não precisa variar pois, apesar de separar faixas cada vez menores, a taxa de amostragem do sinal também é reduzida.

No caso em que se mantém a taxa de amostragem original em todas as bandas, o fato de a organização em oitavas gerar sub-bandas cada vez menores exige seletividade proporcionalmente maior dos filtros. Esta, por sua vez, se reflete na ordem do filtro projetado. Quanto mais seletivo o filtro projetado, maior a ordem requerida.

Outro fator a ser considerado sobre os tipos de filtros utilizados para compor o banco de filtros é a fase. Para aplicações de áudio, é essencial que estes possuam fase linear (ou seja, que possuam atraso de grupo constante). Para realizar a reconstrução perfeita com filtros com fase linear, optou-se pela utilização de filtros FIR (*finite-duration impulse response*).

Os filtros FIR são caracterizados pela equação de diferença

$$y[n] = \sum_{l=0}^{\alpha} h[l]x[n-l], \quad (4.3)$$

onde $h[l]$ é a resposta ao impulso do filtro. Aplicando a transformada z à equação (4.3), temos

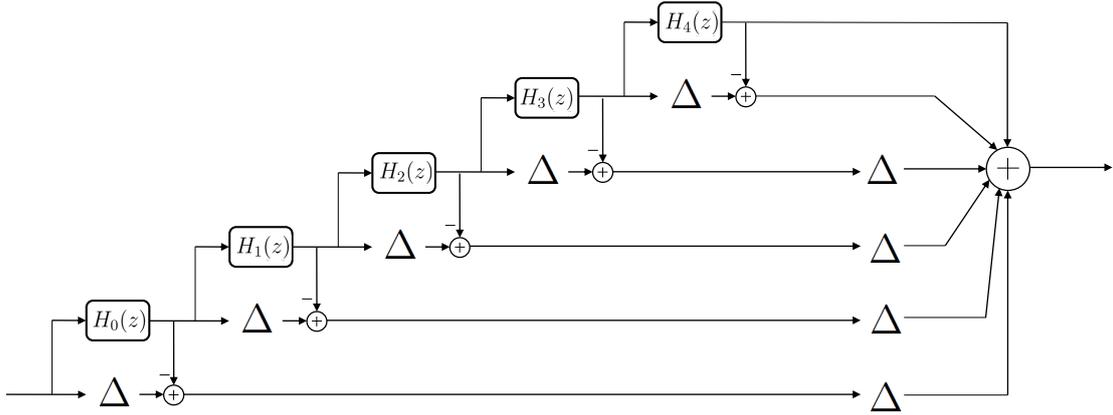


Figura 4.5: Estrutura em oitavas do banco de filtros (todos passa-baixas de meia banda) sem decimação. Δ representa os atrasos inseridos pela operação de filtragem.

$$H(z) = \sum_{l=0}^{\alpha} h[l]z^{-l}, \quad (4.4)$$

que é a forma usual de representação de filtros digitais. Para fase linear, a resposta em frequência do filtro deve ter a forma

$$H(e^{j\omega}) = B(\omega)e^{-j\omega\tau+j\phi}, \quad (4.5)$$

onde $B(\omega)$ é real e τ e ϕ são constantes. A resposta ao impulso, por sua vez, satisfaz

$$h[n] = \frac{e^{j\phi}}{2\pi} \int_{-\pi}^{\pi} B(\omega)e^{j\omega(n-\tau)}d\omega \quad (4.6)$$

Os filtros FIR com fase linear apresentam resposta ao impulso bem particulares [8], descritas pela equação

$$h[n] = (-1)^k h[M - n], \quad k \in \mathbb{Z}. \quad (4.7)$$

É possível observar em (4.7) que a resposta ao impulso é simétrica ou antisimétrica. A forma utilizada para o projeto do banco de filtros descrito na Subseção 4.1.2 é simétrica, de ordem par.

4.1.2 Especificações de projeto

A quantidade de sub-bandas geradas para o projeto do banco de filtros sem decimação é limitada pela quantidade de coeficientes do filtro projetado. As características de *ripple* R_p na banda de passagem que se estende de 0 à frequência ω_i ,

ripple R_r na banda de rejeição que se estende da frequência ω_s até π e transição entre as duas são as mesmas que seriam exigidas no projeto dos filtros não decimados. Assim, o ponto de partida para o banco de filtros sem decimação é o filtro passa-baixas H_0 , cuja ordem mínima é $N = 54$, para $R_p = 0,1$ dB, $R_r = -60$ dB, com $\omega_i = (1 - 0,1)2\pi$ e $\omega_s = (1 + 0,1)2\pi$. O projeto dos filtros subsequentes explorou maximamente o limite computacional de precisão realizável pela plataforma *Matlab*.

Tabela 4.1: Ordem e faixa de transição dos filtros utilizados no banco sem decimação.

Filtro	Ordem (N)	ω_i	ω_s
$H_0(z)$	54	$(1 - 0,1)2\pi$	$(1 + 0,1)2\pi$
$H_1(z)$	218	$(1 - 0,05)2\pi$	$(1 + 0,05)2\pi$
$H_2(z)$	870	$(1 - 0,025)2\pi$	$(1 + 0,025)2\pi$
$H_3(z)$	3478	$(1 - 0,0125)2\pi$	$(1 + 0,0125)2\pi$
$H_4(z)$	13906	$(1 - 0,00625)2\pi$	$(1 + 0,00625)2\pi$

O interessante do projeto de bancos de filtros dessa natureza é que a reconstrução do sinal não depende de filtros de síntese, pois a taxa de amostragem em cada sub-banda é a mesma. Com a taxa de amostragem invariável, a reconstrução é feita a partir da soma direta da saída de cada uma das bandas, após compensados os atrasos relativos à operação de filtragem.

A desvantagem de trabalhar com bancos de filtros sem decimação é o limite imposto pelo processamento computacional. A decisão tomada em trabalhar com o banco nesses moldes levou em consideração, ainda, duas questões: utilizar um banco de filtros com decimação e explorar um número maior de sub-bandas, aumentando a capacidade de investigação de separação, ou utilizar banco de filtros sem decimação e trabalhar em um número de bandas limitado. Dessas questões, a segunda foi escolhida sobretudo porque, como exposto anteriormente, a vantagem de um banco de filtros trabalhando na mesma taxa (por ser menos complexo) predomina sobre a primeira opção.

As sub-bandas obtidas a partir do projeto do banco de filtros utilizado são descritas na Tabela 4.2. A resposta em frequência do banco de filtros é ilustrada na Figura 4.6.

As Figuras 4.7 a 4.9 apresentam a resposta em frequência (fase e magnitude) real dos filtros projetados. $|H(e^{j\omega})|$ [dB].

Tabela 4.2: Limite das sub-bandas geradas.

Sub-banda	Sub-banda	
	lim. inferior (Hz)	lim. superior (Hz)
1	11025	22050
2	5512,5	11025
3	2756,25	5512,5
4	1378,12	2756,25
5	689,06	1378,12
6	0	689,06

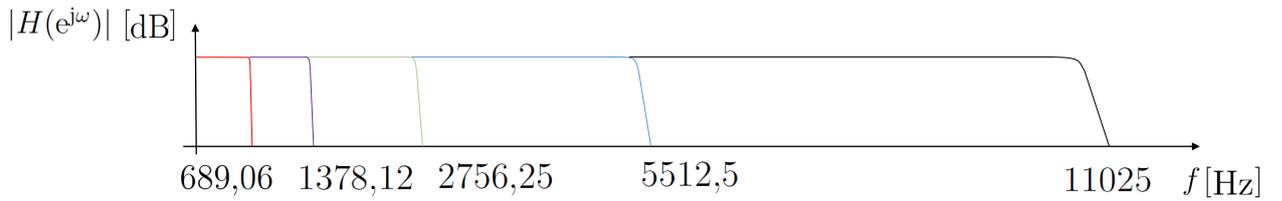
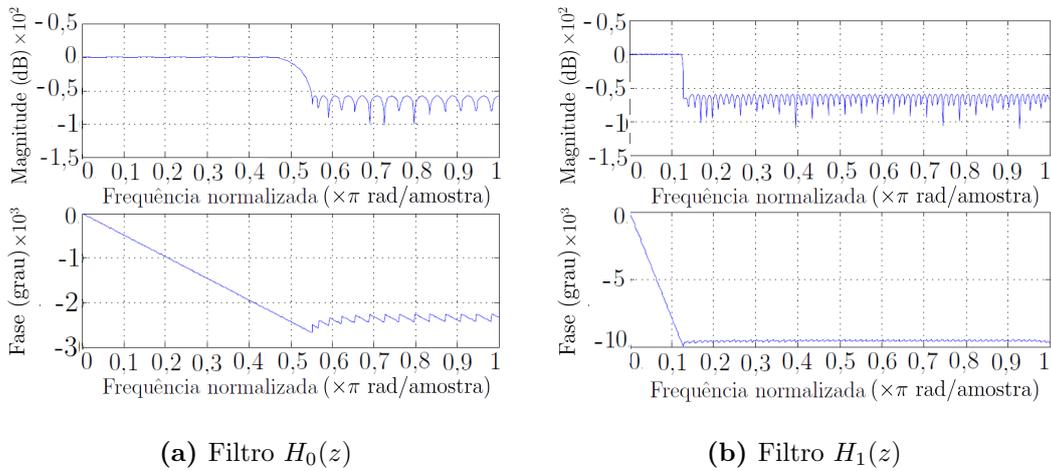


Figura 4.6: Representação da resposta em frequência dos filtros H_0 a H_4 .



(a) Filtro $H_0(z)$

(b) Filtro $H_1(z)$

Figura 4.7: Resposta em frequência dos filtros $H_0(z)$ e $H_1(z)$: magnitude e fase.

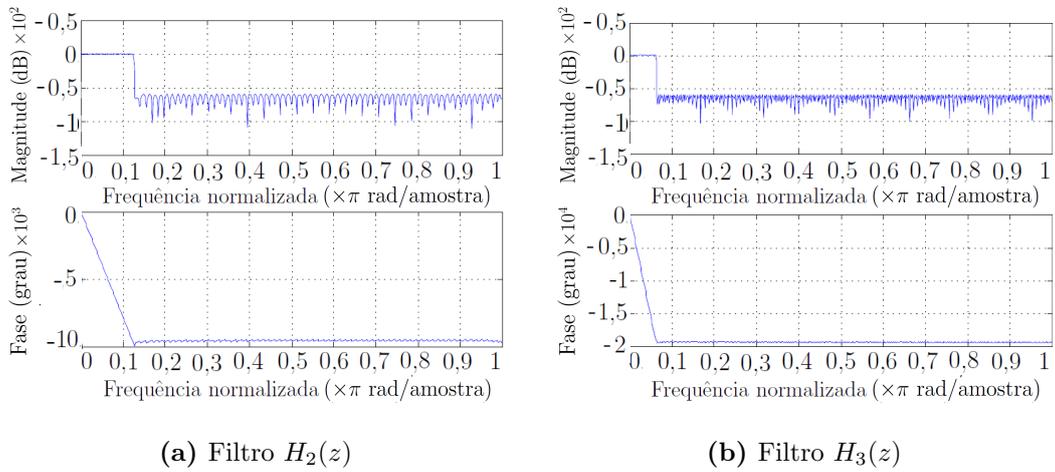


Figura 4.8: Resposta em frequência dos filtros $H_2(z)$ e $H_3(z)$: magnitude e fase.

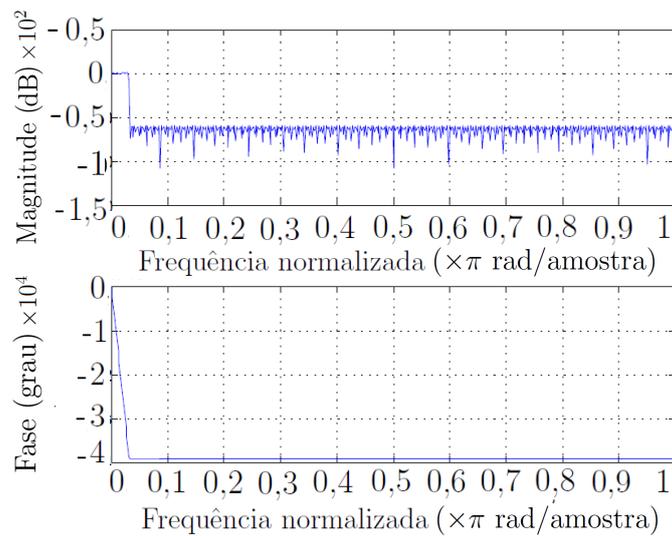


Figura 4.9: Resposta em frequência do filtro $H_4(z)$: magnitude e fase.

Capítulo 5

Base de dados

O objetivo de avaliar as técnicas atuais de separação de fontes em sinais musicais de percussão e a riqueza de instrumentos e ritmos presentes na música brasileira motivaram uma das contribuições deste trabalho: a produção de uma base de dados de instrumentos de percussão brasileira.

É importante ressaltar que alguns dos instrumentos utilizados na base de dados não têm origem brasileira, como é o caso do pandeiro, que é de origem indiana mas largamente empregado na música popular brasileira e item quase obrigatório nas rodas de samba e choro.

Além dos instrumentos tradicionais, as levadas (células rítmicas executadas por um instrumento) e variações rítmicas dos instrumentos são aquelas utilizadas comumente em gravações de cada gênero musical considerado.

5.1 Instrumentos

Os instrumentos escolhidos para a gravação da base de dados foram, dentre um extenso leque de opções, os mais representativos da cultura musical brasileira, sobretudo analisando os instrumentos que caracterizam aquele tido como o mais brasileiro de todos os ritmos: o samba. Ainda, alguns critérios foram levados em consideração para que a base fosse suficientemente genérica, como a variação de cada instrumento de um mesmo tipo (tamanho e formato), da forma de produção do som (mão, baqueta de madeira, baqueta de material plástico e baqueta metálica), da classe do instrumento (membranofones ou idiofones), do material (revestimento de couro, napa, náilon e a estrutura metálica, de madeira ou plástica) e do *pitch* dos instrumentos que possuem uma afinação natural (cuíca e agogô) ou que dependem de afinação relativa quando tocados em conjunto (como funcionam os surdos de escolas de samba, que possuem, em geral, três afinações distintas). A partir dos critérios propostos, dez tipos de instrumentos foram selecionados, sendo listados na Tabela 5.1:

Tabela 5.1: Lista de instrumentos selecionados.

1	Agogô
2	Caixa
3	Chocalho ou <i>shake</i>
4	Cuíca
5	Pandeiro
6	Reco-reco
7	Repique
8	Surdo
9	Tantã
10	Tamborim

Para cada tipo de instrumento selecionado, as variações encontradas são inúmeras. As variações de cada tipo de instrumento utilizadas para a gravação em estúdio aparecem listadas nas Tabelas 5.2 a 5.11:

Tabela 5.2: Agogô.

	Instrumento	Baqueta	Material	Formato
1	Agogô	Madeira	Metal	2 notas

Tabela 5.3: Caixa.

	Instrumento	Baqueta	Tamanho (polegada)	Revestimento	Bordões
2	Caixa	Madeira	12	Náilon	4
3	Caixa	Madeira	12	Náilon	6

Tabela 5.4: *Shake*.

	Instrumento	Formato	Material
4	<i>Shake</i>	Cilíndrico	Madeira
5	<i>Shake</i>	Cilíndrico	Metal
6	<i>Shake</i>	Oval	Plástico
7	<i>Shake</i>	Cilíndrico (duplo)	Plástico

Tabela 5.5: Cuíca.

	Instrumento	Tamanho (polegada)
8	Cuíca	8
9	Cuíca	9,5

Tabela 5.6: Pandeiro.

	Instrumento	Tamanho (polegada)	Revestimento
10	Pandeiro	10	Náilon
11	Pandeiro	10	Couro
12	Pandeiro	11	Náilon
13	Pandeiro	11	Couro
14	Pandeiro	12	Náilon
15	Pandeiro	13	Couro

Tabela 5.7: Reco-reco.

	Instrumento	Tamanho (polegada)	Revestimento	Mola (Quantidade)
16	Reco-reco	10	Náilon	1
17	Reco-reco	10	Couro	2
18	Reco-reco	11	Náilon	3
19	Reco-reco	11	Couro	—

Tabela 5.8: Repique.

	Instrumento	Tamanho (polegada)	Revestimento	Baqueta
20	Repique	10	Náilon	—
21	Repique de Baqueta (Repinique)	12	Náilon	Madeira
22	Repique de anel	12	Couro	—

Tabela 5.9: Surdo.

	Instrumento	Tamanho (polegada)	Revestimento
23	Surdo	18	Couro
24	Surdo	20	Couro

Tabela 5.10: Tantã.

	Instrumento	Tamanho (polegada)	Revestimento
25	Tantã corte	10	Couro
26	Tantã corte	12	Couro
27	Tantã marcação	14	Napa/Náilon

Tabela 5.11: Tamborim.

	Instrumento	Baqueta	Revestimento
28	Tamborim	Madeira	Couro
29	Tamborim	Madeira	Náilon
30	Tamborim	Plástica	Náilon

5.2 Ritmos musicais

Quanto à execução instrumental, os ritmos ou levadas estão centrados no samba e nas suas variações mais características: samba tradicional, partido alto e samba de enredo. As outras levadas apresentadas são a da capoeira (principalmente para a execução do pandeiro), da marcha (ou marchinha, presente em músicas carnavalescas e hinos), baião e maxixe.

Para cada um dos instrumentos, há, pelo menos, uma gravação “livre”. A gravação livre foi proposta aos músicos para permitir capturar o improviso, uma característica muito peculiar da roda de samba. As músicas apresentadas nas rodas de samba informais possuem convenções, em momentos específicos ou não, baseadas em gravações ou por acordo entre os músicos, onde determinado instrumento foge à célula original e subdivide a pulsação rítmica à sua maneira. Por isso, a gravação “livre” tem o modo do improviso de cada músico participante.

As várias gravações dos estilos propostos sugerem ainda mais especificidade em relação aos ritmos. Diferenças pequenas de andamento e execução são classificadas de maneiras distintas dentro do mesmo gênero. A exemplo disso, dentro do próprio samba existem classificações como samba de roda, samba-canção, bossa-nova e muitas outras. A especificação do andamento que foi utilizado em cada ritmo musical é mostrada na Tabela 5.12:

Tabela 5.12: Padrão de ritmos e andamentos utilizados na gravação da base, com duração de 20 segundos (tempo médio) para cada faixa.

Ritmo	Batidas por minuto (bpm)
Samba	80
Partido alto	100
Samba de enredo	130
Capoeira	65
Marcha	120
Livre	variável

As gravações livres variam, em termos de bpm, em cada amostra. O compromisso com a naturalidade permitiu que o andamento fosse selecionado pelo próprio músico, deixando-o à vontade para o improviso. Todas as faixas gravadas têm entre 20 e 25 s de duração. Os ritmos em que não se costuma utilizar um dado instrumento (por exemplo, caixa ou repique tocando capoeira) não foram explorados. Foi valorizada a fidelidade ao estilo musical em detrimento da simetria da base.

5.3 Ambiente de gravação

As gravações foram realizadas entre os meses de outubro e dezembro de 2015, no estúdio duValle, localizado em Manaus-AM. A sala de gravação possui dimensões $4,3\text{ m} \times 3,4\text{ m} \times 2,3\text{ m}$. O revestimento acústico interno apresenta madeira do piso até 1 (um) metro de altura e almofada acústica no restante da parede, teto e porta. A distribuição, bem como a escolha do microfone para cada instrumento, foram previamente testadas e definidas, dadas as características acústicas da sala e as características sonoras de cada instrumento. Assim, foi possível realizar a gravação com microfones direcionais, a vinte centímetros de distância, para a maioria dos instrumentos. As características de gravação em relação ao modelo do microfone utilizado para cada instrumento estão registradas na Tabela C.1. As gravações foram feitas com frequência de amostragem de 44100 Hz e 16 bits de resolução, no *software* de edição de áudio Nuendo-Steinberg, e estão disponíveis em <http://www.smt.ufrj.br/~pedro.donadio/>. O roteiro de gravação seguiu duas partes distintas e fundamentais para a produção da base. A primeira contempla as gravações dos instrumentos solo, gravados separadamente com apenas um músico executando um instrumento, em diferentes ritmos. A segunda parte contempla a gravação das misturas acústicas, com arranjos envolvendo dois, três e até quatro músicos tocando ao mesmo tempo dentro da sala de gravação, também executando ritmos distintos.

5.3.1 Gravação das amostras individuais

A primeira parte contou com três músicos gravando as amostras individuais, em dias diferentes. Cada percussionista gravou toda a gama de instrumentos em todos os ritmos propostos. Exemplo: o músico 1 grava uma faixa de vinte segundos, utilizando um pandeiro de couro, dez polegadas, no ritmo de samba (80 bpm). Em seguida, usa o mesmo pandeiro de couro, dez polegadas, para gravar o ritmo partido alto (100 bpm). Depois, segue para a gravação, com o mesmo pandeiro, do ritmo samba de enredo (130 bpm). O mesmo acontece para a marcha e para a gravação livre. Encerrados os ritmos, o músico 1 troca de instrumento, prepara o pandeiro de Náilon, dez polegadas, e grava os mesmos ritmos anteriormente executados com o pandeiro de couro de dez polegadas. Esse procedimento acontece para todos os instrumentos disponíveis para o músico 1. Generalizando, cada músico grava todos os ritmos com todos os instrumentos (a sequência detalhada dos instrumentos gravados e suas variações são apresentadas na Seção 5.1). A lista completa das gravações das misturas acústicas, com o conteúdo rítmico e instrumental de todas as faixas apresentadas, é mostrada nas Tabelas A.1 a A.3 do Apêndice.

Outro aspecto importante é que os instrumentos utilizados são de propriedade

de cada músico. Logo, a gravação do pandeiro de couro de dez polegadas do músico 2 foi feita com um pandeiro de couro de dez polegadas diferente daquele utilizado pelo músico 1. Apesar de características similares, os instrumentos representam diferenças sutis porém importantes quando se trata de uma base de dados. As gravações individuais foram realizadas, em sua maioria, com instrumentos diferentes, salvo raríssimas repetições. Um resumo da quantidade de gravações por instrumentos e a quantidade total de faixas gravadas são apresentados na Tabela 5.13.

Tabela 5.13: Quantidade de gravações individuais por tipo de instrumento.

	Instrumento	Característica	Revestimento	Faixas Gravadas
1	Pandeiro	10"	Náilon	7
2	Pandeiro	11"	Náilon	21
3	Pandeiro	12"	Náilon	21
4	Pandeiro	10"	Couro	21
5	Pandeiro	11"	Couro	7
6	Pandeiro	12"	Couro	7
7	Tamborim	6"	Náilon	17
8	Tamborim	6"	Couro	3
9	Surdo	18"	Couro	10
10	Surdo	20"	Couro	16
11	Tantã	Corte	Couro	13
12	Tantã	Marcação	Napa/Náilon	10
13	Repique	Baqueta	Náilon	13
14	Repique	Mão	Náilon	12
15	Repique	Anel	Couro	12
16	Caixa	12"	Náilon	17
17	Shake	Metal,madeira ou plástico	-	22
18	Agogô	2 notas	-	12
19	Reco-reco	Madeira	-	8
20	Reco-rerco	Metal	-	16
21	Cuíca	8" ou 9,5"	Couro	8
			TOTAL	273

A partir das faixas individuais de cada instrumento, é possível gerar misturas artificiais de quantos instrumentos forem necessários, via *software* de edição de áudio, como o *Audacity*¹, e simulação de áudio, como o *CATT-Acoustic* (utilizados neste trabalho).

5.3.2 Gravação das misturas acústicas

A segunda parte da gravação ocorreu em um único dia, em que quatro músicos, formando diferentes arranjos, gravaram as misturas acústicas. Nesse cenário de

¹Disponível para *download* na página do fabricante: <http://www.audacityteam.org/>

gravação, as disposições dos microfones foram escolhidas de modo que, no caso extremo dos quatro percussionistas ao mesmo tempo no estúdio, o vazamento de áudio no microfone de um instrumento fosse o menor possível. Ainda para a mistura acústica, além dos microfones direcionais a vinte centímetros do instrumento, dois microfones omnidirecionais, um na parte da frente da sala, centralizado, e outro na parte de trás da sala, captavam o som ambiente. Um esboço da sala e das disposições de microfones/instrumentos pode ser visto na Figura 5.3.

As misturas definidas previamente no roteiro de gravação seguiram, basicamente, os critérios de arranjos musicais presentes em gravações comuns do gênero. Assim, para o arranjo com quatro instrumentos de percussão foram geradas 19 gravações, para três instrumentos 29 gravações, e para dois instrumentos 45 gravações, cada uma com 20 segundos de duração (em média). A lista completa das gravações individuais, com o conteúdo rítmico e instrumental de todas as faixas apresentadas, é mostrada nas Tabelas A.4 a A.32.

5.3.3 Misturas produzidas a partir de um simulador acústico

Para permitir a verificação dos efeitos de separação de fontes de percussão em condições mais controladas e variáveis, foram criadas misturas geradas a partir de um simulador acústico.

O *software* de simulação CATT-*Acoustic* TM v9 (CATT-A v9.0a e TUCTv1.0g)² permite a configuração/recriação detalhada dos ambientes, como dimensão real, revestimento de cada parede, e quantidade de pessoas, estimando medidas acústicas (T30, por exemplo) de cada planta. Os ambientes criados via CATT visam a aproximar as simulações das situações reais, em que os instrumentos são tocados e o som produzido é influenciado pelo ambiente, sendo absorvido e refletido de maneira diferente em cada superfície ou obstáculo. As configurações permitem, ainda, a seleção do tipo e posição das fontes e o tipo e posição dos microfones presentes.

Reunidas as configurações do ambiente e as posições de fontes e microfones, a simulação no CATT retorna a resposta ao impulso do trajeto fonte-sensor. As respostas podem ser geradas em diversos formatos mas, por compatibilidade com os demais experimentos realizados, foram salvas no formato MATLAB. A partir deste resultado foi possível gerar uma simulação acústica do som de cada fonte, convoluindo a amostra gravada individualmente com a resposta ao impulso do ambiente. É importante notar que a resposta do ambiente varia de acordo com as posições da fonte e do microfone escolhidas. Assim, o posicionamento fonte-microfone configurado para a simulação diz respeito ao lugar usual no ambiente descrito (palco e

²Página do fabricante: <http://www.catt.se/>

plateia). Matematicamente, tem-se

$$s_{\text{as}}[n] = (s_o * h)[n], \quad (5.1)$$

onde $s_{\text{as}}[n]$ é o sinal acústico simulado e $s_o[n]$ é o registro original do instrumento.

Com as amostras simuladas, as misturas desejadas foram editadas através do *Audacity*. Enquanto o *software* de simulação gera a resposta do ambiente, na edição são ajustados o ganho de cada instrumento e o tamanho (segundos) da faixa de áudio, além de realizar a mixagem dos instrumentos em 1 canal. O processo se inicia no *software* de edição, sincronizando musicalmente as faixas individuais. Em seguida, cada faixa, individualmente, é convoluída com a resposta ao impulso correspondente à posição da fonte daquele instrumento (configuração descrita nas Seções 5.3.3 à 5.3.3). Após isso, o *software* de edição é utilizado, novamente, para a mixagem dos canais. A proposta da utilização do CATT consistiu em simular três ambientes distintos: a sala de uma casa, uma quadra de escola de samba e um estúdio de gravação (nesse caso, uma aproximação do estúdio onde as misturas acústicas foram gravadas, a fim de comparar os cenários real e simulado).

As configurações dos ambientes simulados são apresentadas a seguir.

Sala residencial

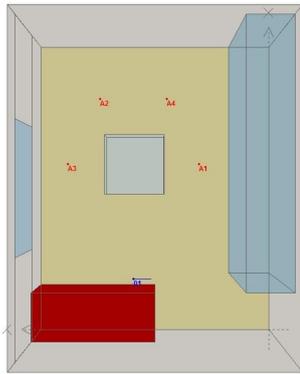
As dimensões da sala são 4 m × 5 m × 2,8 m, tamanho aproximado de uma sala comum. Na planta mostrada na Figura 5.1 é possível observar os detalhes inseridos para a aproximação do ambiente residencial: mesa de centro, janela, armário e sofá.

A possível configuração adotada prevê os instrumentos de um lado da mesa (pontos vermelhos, A1, A2, A3 e A4) e o microfone (ouvinte, representado pelo ponto azul, 01) ao fundo da sala, próximo ao sofá.

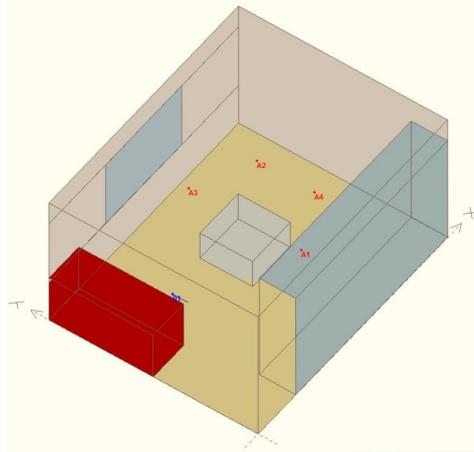
Quadra de escola de samba

A quadra possui dimensões 30 m × 50 m × 15 m. Normalmente, um ambiente como este possui dimensões grandes e pouco tratamento acústico. A configuração adotada pode ser vista na Figura 5.2.

As fontes ficam dispostas representando cada músico tocando de pé, sobre o palco.

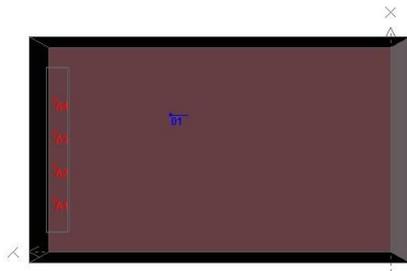


(a) Vista superior

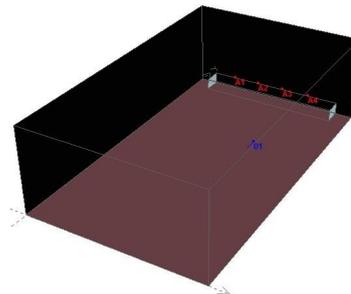


(b) Perspectiva

Figura 5.1: Sala residencial. Os objetos caracterizam uma sala comum, com um armário do lado direito (azul escuro), um sofá na parte de trás (em vermelho), uma mesa de centro (azul mais claro) e uma janela na parede esquerda. Os instrumentos (fontes) estão dispostos em volta da mesa (A1-A4) e o ouvinte (sensor), próximo ao sofá (01).



(a) Vista superior



(b) Perspectiva

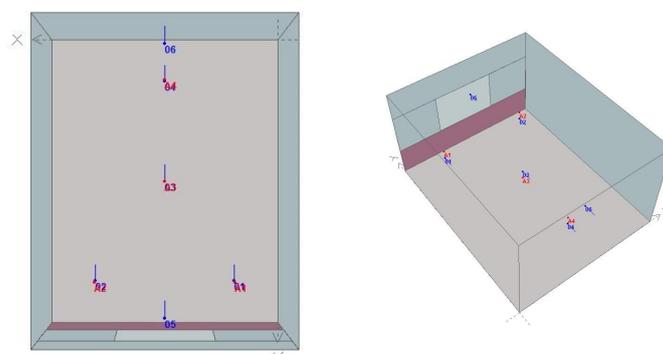
Figura 5.2: Quadra de escola de samba. Sobre o palco (do lado esquerdo da vista superior), quatro instrumentos, dispostos lateralmente a 1 m de altura do chão do palco.

Estúdio de gravação

O estúdio de gravação simulado tem as mesmas medidas e revestimentos do estúdio real onde aconteceram as gravações da base de dados (Figura 5.3).

Para esse caso, os microfones à frente e ao fundo da sala foram usados como referências, e a posição dos músicos no estúdio na gravação acústica foi mantida.

As misturas feitas a partir das gravações individuais, as misturas acústicas realizadas em estúdio e as misturas geradas a partir de um simulador acústico resultam em sinais distintos. A investigação das técnicas de separação de fontes sobre os diversos tipos de mistura também é objeto deste trabalho.



(a) Vista superior

(b) Perspectiva

Figura 5.3: Estúdio de gravação. Simulação do ambiente real de gravação: 4 fontes (A1-A4) e 6 microfones (01-06) ao redor da sala. Os microfones 05 e 06 captam o som ambiente, enquanto os microfones A1-A4 são posicionados a 20 cm dos instrumentos. Na parte da frente, vê-se a janela que separa o ambiente de gravação da sala de operação técnica.

Capítulo 6

Experimentos

Com os métodos apresentados, os experimentos foram conduzidos com o fim de que a estimativa das fontes separadas fosse a mais próxima possível das fontes originais, sobretudo sob a avaliação subjetiva. A avaliação precisa responder a 3 perguntas principais:

1. O sinal separado contém todo o conteúdo da fonte alvo?
2. O sinal separado contém resíduo das demais fontes?
3. O sinal soa natural?

Para realizar a comparação entre os métodos apresentados, uma mesma mistura é separada por diversos métodos ou combinações de métodos diferentes. Os métodos utilizados são:

1. NMFD;
2. Filtragem do espectrograma por mediana;
3. NMFD aplicada nas sub-bandas do sinal;
4. Filtragem do espectrograma por mediana aplicada nas sub-bandas do sinal;
5. Filtragem do espectrograma por mediana, seguido de NMFD;
6. Filtragem do espectrograma por mediana aplicada nas sub-bandas do sinal, seguido de NMFD.

A motivação para os experimentos é, em 3 e 4, utilizar as técnicas de separação estudadas em porções menores (em relação ao conteúdo espectral) do sinal, com o objetivo de facilitar a separação. Em 5, separar uma mistura pela aplicação sequencial de dois métodos diferentes. A sequência da filtragem do espectrograma seguida de NMFD pretende priorizar as misturas em que um ou mais instrumentos

possuam uma sustentação aproximadamente harmônica. Em 6, um sistema completo, utilizando todas as técnicas estudadas, é testado — estendendo a proposta 4.

As misturas escolhidas para ilustrar os resultados da separação obtida com os métodos propostos têm características distintas. Para experimentar o sistema de maneira abrangente, as misturas são testadas para os casos em que os instrumentos: ocupam a mesma banda de frequência ou ocupam bandas distintas; tocam ao mesmo tempo (ataque) ou em tempos diferentes. Cada particularidade interfere na qualidade de separação. Além dos aspectos musicais, relacionados ao timbre do instrumento e ao ritmo executado, as características acústicas também são colocadas à prova: as misturas podem ter sido geradas artificialmente, simulando a execução de um conjunto percussivo em ambientes como casa, quadra e estúdio, ou serem misturas instantâneas ou acústicas, como descrito no Capítulo 5.

Antes, porém, são apresentados os sistemas de separação propostos neste trabalho.

6.1 NMFD

O bloco que representa a NMFD compreende o algoritmo apresentado no Capítulo 3. O algoritmo, basicamente, deve ser parametrizado de acordo com a mistura utilizada, isto é, número de fontes e número de deslocamentos permitidos. Nesse ponto, cabe uma justificativa ao leitor sobre a preferência da NMFD em relação à NMF. Em experimentos prévios, para os instrumentos de percussão, a utilização da NMFD já demonstrava desempenho superior quanto à capacidade e à qualidade de separação dos instrumentos. A explicação é bastante intuitiva, visto que os padrões frequenciais dos instrumentos de percussão, em geral, têm comportamento variante ao longo do tempo. Logo, um padrão frequencial de um instrumento percussivo, quando representado por alguns deslocamentos da NMFD, gera resultados melhores do que a NMF.

O sistema básico utilizado na NMFD para duas fontes é apresentado na Figura 6.1.



Figura 6.1: Diagrama em blocos da NMFD para 2 fontes.

A quantidade de faixas geradas na saída do sistema é igual ao número de fontes passadas como parâmetro para o algoritmo. As fontes, para o caso apresentado, são exatamente os instrumentos presentes na mistura. É importante dizer que, um mesmo instrumento pode possuir padrões espectrais bastante diferentes, que soam em instantes de tempo diferentes. Sendo assim, passar para o algoritmo um número de fontes maior que o número de instrumentos presentes em uma mistura pode representar, nesses casos, uma boa escolha. Um caso que ilustra essa questão é o do pandeiro, que possui um revestimento que pode ser de couro ou náilon, cercado por platinelas. O som produzido por ambas as partes do pandeiro são completamente diferentes do ponto de vista espectral. Dada uma mistura onde o pandeiro está presente, o resultado da opção de passar um número maior de fontes para o algoritmo é, na saída do sistema, o instrumento pandeiro ficar dividido em duas subfontes (ou componentes, dependendo da interpretação dada, conforme discussão do Capítulo 1). Estas podem ser posteriormente agrupadas convenientemente para se chegar ao instrumento (completo) original.

6.2 Filtragem do espectrograma por mediana

O diagrama apresentado na Figura 6.2 relaciona a entrada e saídas do algoritmo apresentado. Independente da quantidade de fontes presentes, a saída sempre é dividida em parte transitória e permanente. O parâmetro passado para o algoritmo é o tamanho da janela utilizada para o cálculo da mediana. O tamanho das janelas pode ser independente em ambos os eixos, isto é, é possível escolher uma janela de tamanho N_p na filtragem do espectrograma no eixo do tempo, e uma janela de tamanho N_t na filtragem do espectrograma no eixo da frequência.

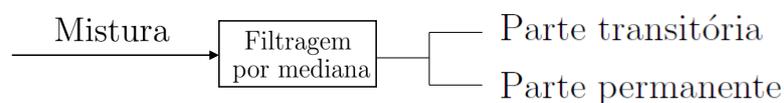


Figura 6.2: Diagrama em blocos da filtragem por mediana.

A saída do sistema é a mistura em suas porções recuperadas da parte permanente e a parte transitória. Ainda, uma outra saída possível ocorre se mais de uma iteração do algoritmo for realizada. O processo de refiltragem das partes gera uma saída *residual*, que não poderia ser classificada estritamente nem como transitória nem como permanente.

6.3 NMFD aplicada nas sub-bandas do sinal

O processo da divisão do sinal em sub-bandas apresentado no Capítulo 4 permite a investigação, proposta por este trabalho, da separação de fontes em cada banda de frequência do sinal. Para esta configuração, a saída do sistema aparece representada nas Figuras 6.3 e 6.4.

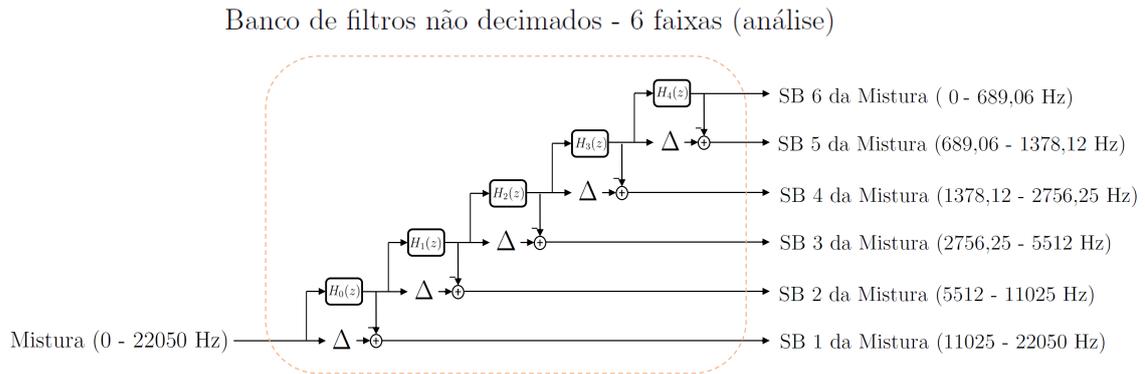


Figura 6.3: Bloco representando a filtragem do sinal com as sub-bandas na saída. SB r identifica a r -ésima sub-banda do sinal.

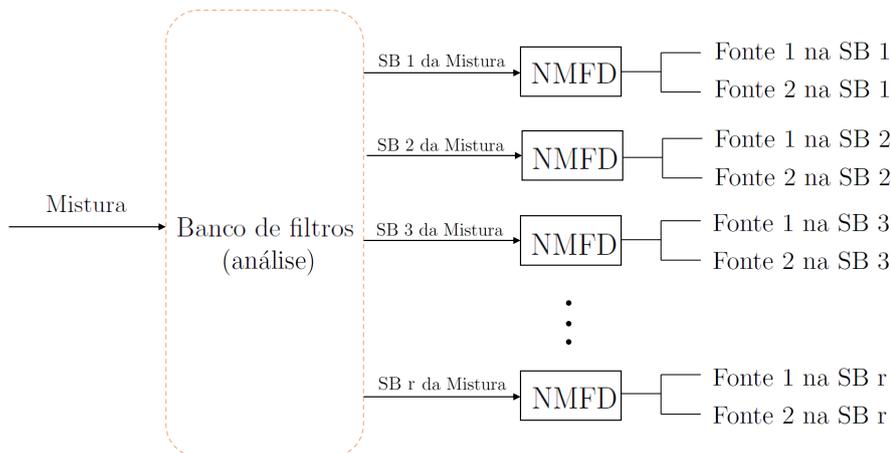
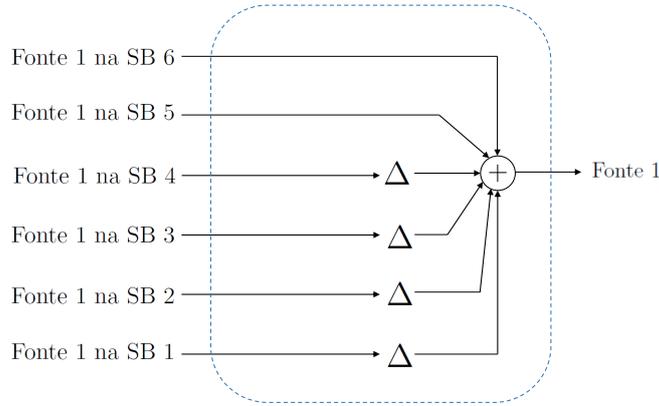


Figura 6.4: NMFD aplicada às sub-bandas do sinal com 2 componentes por sub-banda.

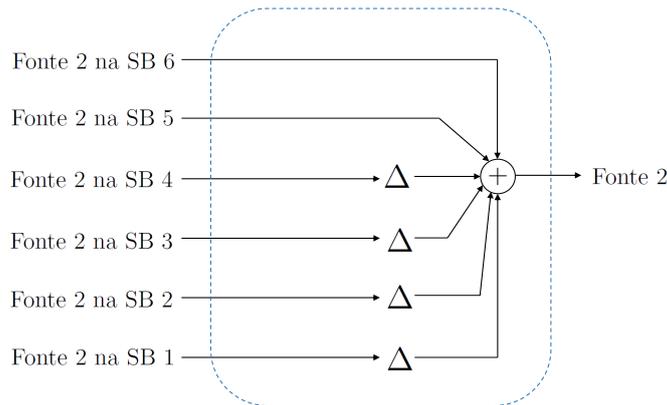
A Figura 6.5 ilustra o caso em que o resultado da NMFD nas sub-bandas é, para a saída 1 da NMFD, conteúdo pertencente à fonte 1, e para a saída 2, conteúdo

Banco de filtros não decimados - 6 faixas (síntese)



(a) Síntese do sinal (referente à fonte 1).

Banco de filtros não decimados - 6 faixas (síntese)



(b) Síntese do sinal (referente à fonte 2) .

Figura 6.5: Processo de síntese do sinal, separado por NMFD aplicada às sub-bandas, para 2 fontes presentes na mistura.

pertencente à fonte 2. Entretanto, há a possibilidade de em determinada sub-banda o resultado da NMFD compreender partes de uma mesma fonte.

Para cada sub-banda do sinal, a saída da NMFD gera Z fontes separadas, conforme parâmetro passado ao algoritmo. A recuperação da fonte original, mantendo o timbre do instrumento, requer agrupar a separação gerada pela NMFD em cada sub-banda e proceder com a síntese do sinal nos moldes descritos na Seção 3.2. Realizar esse agrupamento é um desafio. É possível exemplificar o caso quando são tratados sinais que ocupam faixas distintas de frequência, por exemplo, um surdo e um tamborim. Nas faixas mais agudas do sinal, não há presença do surdo, que pos-

sui um som essencialmente grave, em torno de 100 Hz. Isso significa que, na maior parte das sub-bandas não há presença do surdo. De maneira complementar, nas bandas mais graves, o sinal do tamborim praticamente não aparece. Entretanto, a NMFD é aplicada a todas as sub-bandas. Um outro detalhe é que, como a separação é realizada de maneira independente em cada sub-banda, sem nenhuma informação que relacione o padrão frequencial com o instante do tempo em que é ativado, não necessariamente as fontes em cada sub-banda são separadas da mesma forma: os sinais separados na sub-banda mais aguda podem não ter a mesma característica ou pertencer à mesma fonte dos sinais separados na sub-banda seguinte. É fácil perceber que a complexidade da resolução desse problema experimenta um crescimento à medida que o número dos sinais a serem separados aumenta. Neste trabalho, foi realizado o agrupamento manual de maneira perceptiva, a fim de relacionar componentes de sub-bandas diferentes que fazem parte do mesmo instrumento. Apesar de gerar mais saídas que precisam ser reorganizadas antes da síntese do sinal, a vantagem da proposta é permitir um maior refinamento na separação das fontes. Escolher determinadas componentes e/ou excluir outras contribui no processo de separação, dependendo do objetivo desejado quanto à qualidade da separação ou integridade dos instrumentos, por exemplo.

6.4 Filtragem do espectrograma por mediana aplicada nas sub-bandas do sinal

De maneira similar ao que acontece na NMFD aplicada nas sub-bandas, a filtragem por mediana separa o sinal segundo o diagrama representado na Figura 6.6.

Em geral, o problema de relacionar as fontes separadas em cada sub-banda por esse método é ligeiramente mais simples do que o problema abordado na separação da Seção 6.3. Independente da quantidade de instrumentos presentes, o método de filtragem do espectrograma por mediana gera sempre duas saídas, uma transitória e outra permanente.

6.5 Filtragem do espectrograma por mediana, seguida de NMFD

Nessa proposta, as duas técnicas são organizadas em cascata. O objetivo desse esquema de separação é análogo ao utilizado no esquema de sub-bandas seguido de NMFD: utilizar a NMFD em parte do sinal com menos informação que o sinal original completo. Obviamente, essa “parte do sinal” refere-se a elementos totalmente diferentes. Enquanto no caso das sub-bandas a “parte do sinal” diz respeito

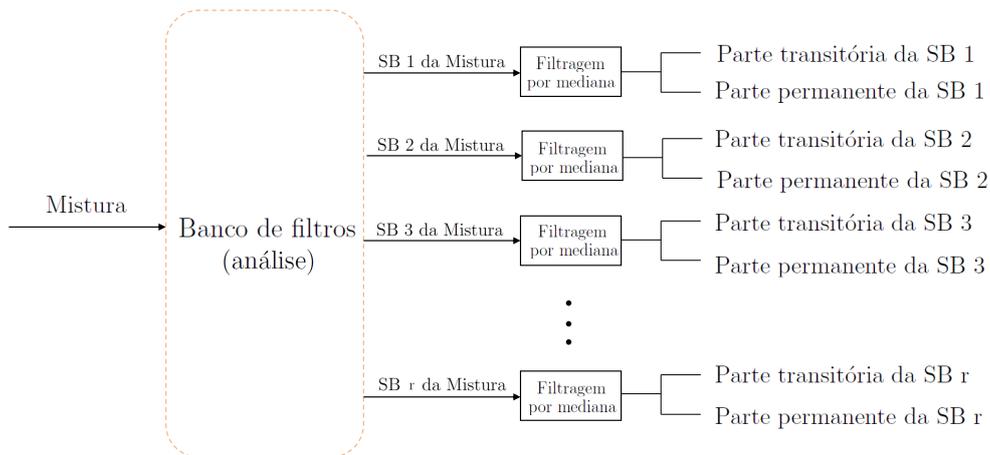


Figura 6.6: Filtragem do espectrograma por mediana aplicada nas sub-bandas do sinal.

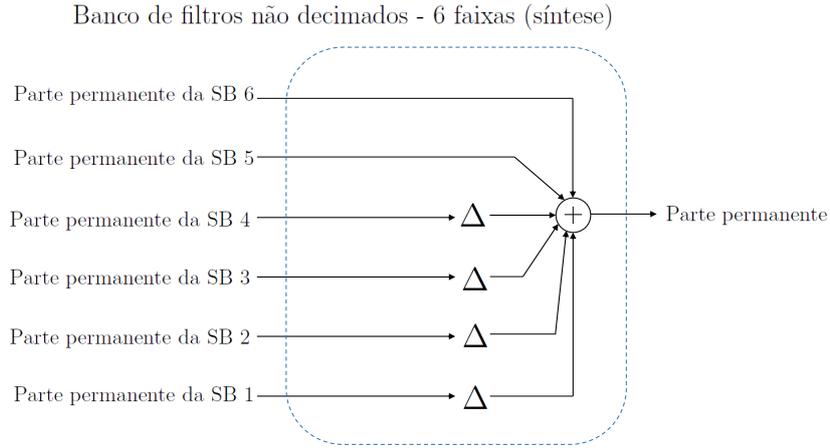
ao conteúdo espectral presente, na filtragem por mediana essa “parte do sinal” tem a ver com os conceitos de transitório e permanente. Um esquema de separação é descrito na Figura 6.8.

As fontes, para esse esquema da separação em cascata, podem ser reconstruídas através da soma das saídas representadas no diagrama.

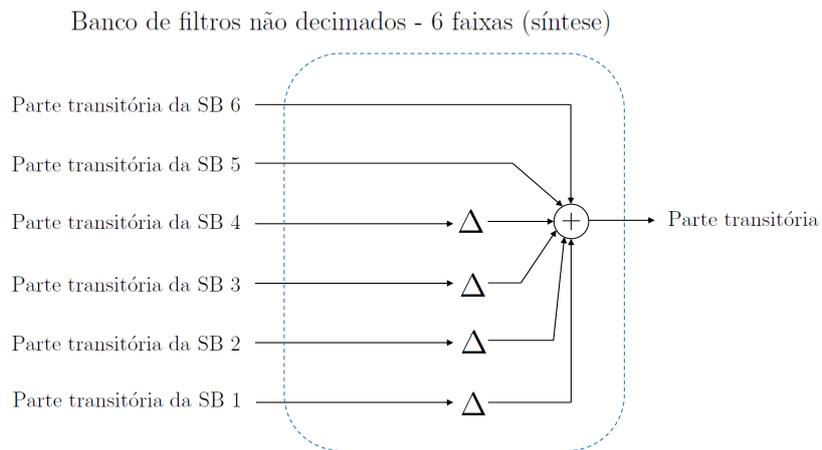
6.6 Filtragem do espectrograma por mediana aplicada nas sub-bandas do sinal, seguida de NMFD

Esse esquema de separação, apresentado na Figura 6.9, é o mais completo apanhado das técnicas apresentadas. A quantidade de variáveis pode ser manipulada de acordo com o objetivo de separação (qualidade de timbre, separação de artefatos, etc.).

O grande inconveniente desse tipo de combinação é exatamente o que faz dela a mais versátil: a quantidade de variáveis geradas na saída. Essa quantidade, proporcional aos parâmetros utilizados nos algoritmos NMF, impacta diretamente na síntese dos sinais. Com a necessidade de uma avaliação subjetiva das saídas parciais, o tempo de cada experimento aumenta. Para exemplificar: supondo uma mistura de 4 fontes, são geradas 6 sub-bandas após o processo de filtragem do sinal de entrada. A saída de cada sub-banda, ao passar pelo algoritmo de filtragem do espectrograma



(a) Síntese do sinal (referente à parte permanente).



(b) Síntese do sinal (referente à parte transitória) .

Figura 6.7: Processo de síntese do sinal, separado por filtragem do espectrograma por mediana aplicada às sub-bandas.

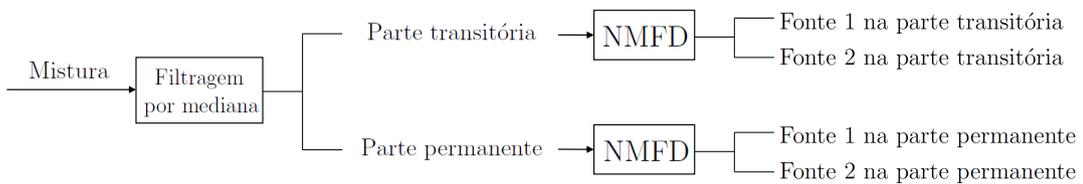


Figura 6.8: Diagrama em bloco representando a filtragem do espectrograma por mediana, seguido de NMFD para 2 fontes.

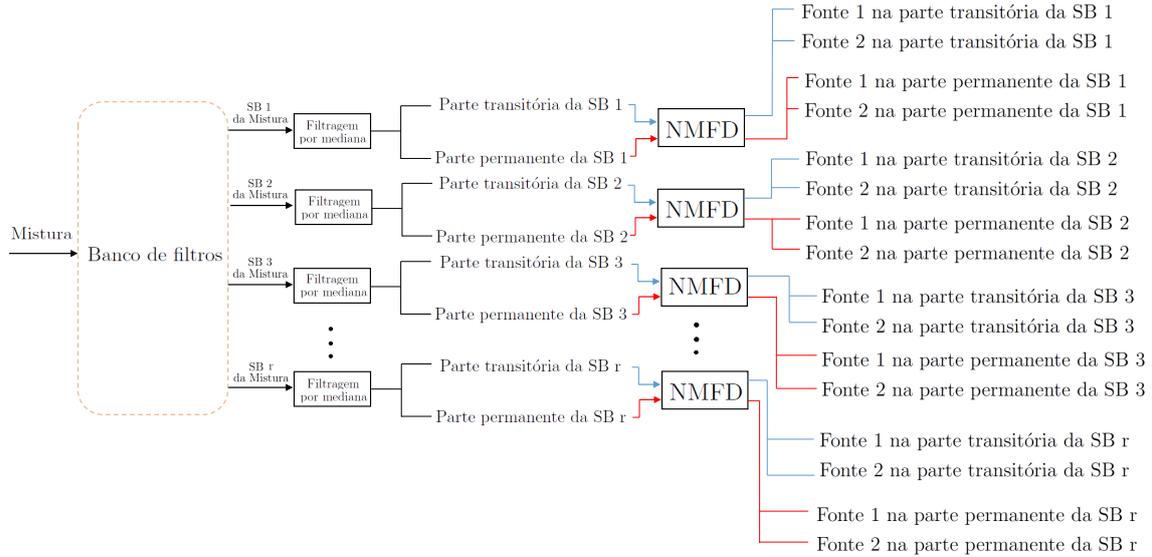


Figura 6.9: Diagrama em bloco representando a filtragem do espectrograma por mediana, seguido de NMF para 2 fontes.

por mediana, divide o sinal em mais 2 saídas, uma transitória e uma permanente. Por fim, a NMF (utilizando 4 fontes como parâmetro, no mínimo) aplicada a cada uma das saídas do estágio anterior acumula, na saída global do sistema, 48 faixas que precisarão ser combinadas.

6.7 Misturas e resultados

Como dito no início do capítulo, os experimentos pretendem abranger misturas com execuções de instrumentos e ritmos que sejam representativos do ponto de vista tempo-frequencial. Assim, as misturas escolhidas são descritas na Tabela 6.1

Tabela 6.1: Misturas utilizadas.

	Instrumento	Ritmo	bpm	Tipo de mistura	Faixa
1	Reco-reco	Samba	80	Instantânea	236
	Surdo				286
2	Reco-reco	Samba	80	Convolutiva	236
	Surdo				286
3	Pandeiro	Partido alto	100	Instantânea	202
	Tamborim				231
4	Pandeiro	Partido alto	100	Convolutiva	202
	Tamborim				231

6.7.1 Caso 1: mistura instantânea de reco-reco e surdo

A mistura 1, composta de reco-reco e surdo, é um exemplo de mistura com instrumentos cujo conteúdo espectral é bastante separado: o surdo com som grave, concentrado em baixas frequências, e o reco-reco com som agudo e timbre metálico. Além disso, o surdo toca marcando o tempo fraco do compasso binário¹ e o reco-reco toca, a maior parte do tempo, 8 vezes (semicolcheias²) por compasso. Isso mostra a diferença frequencial e temporal dos instrumentos.

A outra característica do caso 1 é a natureza da mistura. A mistura foi produzida a partir das faixas da base de dados gravadas individualmente, num processo similar às gravações comerciais de música popular. Após a aquisição dos sinais, um *software* de edição é utilizado para sincronizar musicalmente os sinais e ajustar valores de ganhos de acordo com a função de cada instrumento em um arranjo de percussão.

Resultados

A mistura linear reco-reco e surdo foi investigada por todos os sistemas de separação abordados. Os resultados mais relevantes são:

A NMFD, utilizada com 2 fontes e 10 deslocamentos permitidos separou razoavelmente bem a mistura. É possível perceber na saída de ambos os instrumentos um leve resíduo do outro. Em outras palavras, na separação do reco-reco são encontrados resquícios do surdo, principalmente no instante do ataque do instrumento mais grave. No caso do surdo são perceptíveis os indícios do reco-reco. Esse é um problema que decorre da dificuldade de separar os instrumentos quando dois ou mais tocam ao mesmo tempo. Apesar dos vestígios, os timbres permanecem naturais e pouco afetados pela estimação do espectro (www.smt.ufrj/~pedro.donadio - *Tabela Caso 1 - NMFD 1 e NMFD 2*).

A filtragem do espectrograma por mediana gera, nesse caso particular, um resultado interessante. A aposta da utilização do método, que, como já discutido previamente, foi concebido para a separação da parte harmônica da parte percussiva da música, apresenta-se favorável com instrumentos apenas percussivos. No caso abordado, a separação do surdo e do reco-reco tem características diferentes. Ao observar a faixa referente ao surdo, esperado na parte permanente da separação, é encontrado, de maneira muito evidente, o reco-reco (com o timbre pouco distorcido). A separação (no caso do surdo) é prejudicada, principalmente, pelo fato de o vestígio do reco-reco permanecer ao longo de toda a execução. Entretanto, como o surdo tem grande parte do seu som concentrado na parte permanente da separação, na parte transitória é esperado um bom resultado em relação ao reco-

¹O samba tem compasso binário, em geral notado como 2/4 (2 semínimas por compasso). O compasso binário é dividido em tempos 1 e 2, ou tempo forte e tempo fraco.

²Semicolcheia é a figura musical que equivale a 1/4 da semínima.

reco. E, de fato, é o que ocorre. A característica transitória, de ataques sucessivos, confere ao reco-reco uma separação satisfatória, com timbre natural. O ponto negativo é a presença do ataque do surdo, ocorrendo em curtos instantes de tempo. O procedimento da filtragem por mediana nas sub-bandas levou a resultados similares ao da filtragem por mediana aplicada aos espectrograma inteiro da mistura. A partir daí, duas alternativas foram experimentadas para aprimorar os resultados (www.smt.ufrj/~pedro.donadio - *Tabela Caso 1 - Mfilter 1 e Mfilter 2*).

A primeira delas foi a utilização das sub-bandas, a fim de atingir um alto grau de separação, o que gerou bom resultado. O aproveitamento de apenas algumas sub-bandas para a reconstrução é uma opção aceitável quando é desejada separação (apenas) dos sinais. O motivo é que apesar da dinâmica e dos ataques ficarem claramente caracterizados, a supressão de uma ou mais sub-bandas na reconstrução do sinal acaba por eliminar alguns harmônicos, comprometendo o timbre original do instrumento. Diante disso, a retirada de uma das sub-bandas removeu o ataque do surdo, antes presente no reco-reco. A segunda estratégia foi a utilização da combinação completa: sub-bandas, filtragem por mediana e NMFD. A faixa reconstruída do reco-reco com a presença do ataque do surdo foi passada pela NMFD. O resultado dessa separação manteve, de um lado, apenas o ataque do surdo e, do outro, o reco-reco com timbre íntegro. Do ponto de vista da separação, essa estratégia obteve bom resultado em ambos quesitos de separação e naturalidade do instrumento.

A aplicação da filtragem do espectrograma por mediana aplicada às sub-bandas do sinal seguida de NMFD obteve melhor resultado que as técnicas anteriores, principalmente para o reco-reco. A estratégia atingiu dois pontos interessantes: preservação do timbre do instrumento e o alto grau de separação entre eles (www.smt.ufrj/~pedro.donadio - *Tabela Caso 1 - SB MFILTER NMFD 1 e SB MFILTER NMFD 2*).

6.7.2 Caso 2: mistura convolutiva de reco-reco e surdo

No caso 2, as faixas utilizadas são as mesmas do caso 1. A diferença é a mistura realizada entre as duas faixas. As faixas individuais são convoluídas com a resposta ao impulso do ambiente “quadra”, descrito no Capítulo 5. O alto grau de reverberação inserido pelo ambiente, com dimensões grandes e revestimento interno não tratado é responsável pela degradação da qualidade sonora. Enquanto na mistura instantânea é possível escutar com clareza as nuances da dinâmica imposta pelo músico, no ambiente do caso 2 a reverberação mascara essa condição.

O objetivo desse tipo de mistura é a avaliação da separação de fontes sob o efeito de reverberação, verificando como se comportam os diferentes métodos de separação

em misturas de instrumentos iguais e de naturezas diferentes.

Resultados

O ambiente quadra é o pior caso, pois a reverberação atinge o maior nível dentre as misturas simuladas.

A NMFD, com apenas duas fontes, obteve um dos melhores resultados de separação. O timbre natural mantido após a separação, em ambos os instrumentos, foi o ponto alto do método. Ainda assim, é possível notar alguns artefatos nas fontes. Na faixa do surdo, a definição do ataque do instrumento fica comprometida e parte do reco-reco aparece nesse instante. No caso do reco-reco, o vestígio do ataque do surdo, ausente na faixa do surdo, é encontrado na faixa do reco-reco (www.smt.ufrj/~pedro.donadio - *Tabela Caso 2 - NMFD 1 e NMFD 2*).

A filtragem do espectrograma por mediana tem sua principal vantagem alterada pela natureza da mistura. No caso anterior da mistura instantânea, o timbre natural mantido após a separação é descaracterizado quando a reverberação está presente. O som refletido faz com que grande parte do reco-reco, antes separado como transitório, seja separado como permanente. Esse detalhe faz com que a qualidade e a fidelidade do timbre sejam perdidas (www.smt.ufrj/~pedro.donadio - *Tabela Caso 2 - MFilter 1 e MFilter 2*).

O sistema completo, com sub-bandas, filtragem por mediana e NMFD, consegue atingir duas fontes mais separadas do que os sistemas anteriores. Na busca por uma estratégia ideal, deve-se lembrar sempre do objetivo desejado: separação total dos instrumentos ou timbre natural dos instrumentos. O fato de sucessivas operações serem realizadas pode prejudicar uma ou outra meta. A separação em sub-bandas seguida pela filtragem por mediana resulta na faixa do reco-reco com o surdo ainda evidente. Nesse ponto, a NMFD é o caminho escolhido, na tentativa de separar ainda mais as duas fontes. O que ocorre é que, após a aplicação da NMFD, a faixa do reco-reco segue sem nenhum vestígio do surdo. Entretanto, parte do reco-reco também é extraída do sinal original, deixando o timbre menos realista.

O surdo, após a separação em sub-bandas e a filtragem por mediana, permanece com vestígio do reco-reco. Após a NMFD, em uma das saídas aparece grande parte do surdo, porém sem o ataque definido. Na outra saída, o vestígio do reco-reco e o ataque do surdo ficam caracterizados. Percebendo a possibilidade da recuperação da naturalidade do som do instrumento, a NMFD foi aplicada, mais uma vez. O intuito foi de separar os artefatos do reco-reco do ataque do surdo, recuperando o sinal a partir da soma direta da saída do surdo da primeira NMFD e da saída do ataque da segunda NMFD. O resultado disso é um som mais natural, completo, do surdo, mas com alguns ruídos (www.smt.ufrj/~pedro.donadio - *Tabela Caso 2 - SB MFILTER NMFD 1 e SB MFILTER NMFD 2*).

6.7.3 Caso 3: mistura instantânea de pandeiro de couro e tamborim

Para este caso, a mistura gerada utiliza instrumentos que se confundem quanto ao conteúdo espectral. O tamborim, com revestimento de náilon e baqueta de madeira, tem um som agudo. O pandeiro, por sua vez, possui 2 grupos de sons com características diferentes: um som mais grave, proveniente do toque no revestimento de couro, esticado sobre o aro, e um som mais agudo, das platinelas, que são pequenos discos metálicos presos na lateral do instrumento. O pandeiro toca, na maioria do tempo, em semicolcheias, enquanto o tamborim toca acentuando o contratempo do compasso binário.

Ambos os instrumentos tocam ao mesmo tempo em diversos momentos e possuem, na frequência, o som do tamborim e das platinelas sobrepostos. Aqui, a análise da estratégia de separação é um pouco mais complexa. É possível considerar parametrizar os algoritmos com um número maior de fontes, para que, ao final, as componentes de um único instrumento sejam agrupadas e o som original da fonte real (leia-se um instrumento) seja recuperado.

Resultados

A sobreposição espectral dos instrumentos é o fator a ser considerado na mistura do pandeiro e tamborim. Isso é mostrado no experimento com a NMFD. O pandeiro de couro é separado e seu timbre é mantido. Nenhum resíduo do tamborim é percebido. Em contrapartida, na faixa do tamborim, parte do som das platinelas fica misturado. Esse cenário é esperado e requer outras alternativas para separação.

A NFMD nas sub-bandas não aparece como estratégia eficiente. A tentativa de eliminar algumas bandas, a fim de separar o som das platinelas do som do tamborim, similar ao realizado nos casos anteriores, é extremamente prejudicial ao timbre do instrumento. Isso acontece exatamente pela sobreposição espectral das duas componentes (www.smt.ufrj/~pedro.donadio - *Tabela Caso 3 - SB NMFD 1 e SB NMFD 2*).

A sequência sub-banda, filtragem por mediana e NMFD consegue saída similar àquela obtida no experimento somente com a NMFD. A diferença é que o resíduo das platinelas aparece um pouco mais sutil do que no experimento com NMFD apenas. Colocando o tempo de simulação e a complexidade do sistema como critérios de escolha, o sistema com sub-bandas, filtragem por mediana e NMFD não é vantajoso para a mistura proposta.

Diante do problema de separação, uma alternativa encontrada foi utilizar o resultado da NMFD como ponto de partida para uma nova estratégia, que já apresentava um resultado bom, com presença de resíduos em apenas uma das fontes. Como os

resultados para as combinações com sub-bandas e para a filtragem por mediana não favoreceram a separação dos instrumentos, a solução encontrada foi a realização de nova NMFD sobre o resultado com resíduos. Assim, a combinação de métodos escolhida para a separação dos instrumentos do caso 3 foi a utilização da NMFD em cascata, separando os artefatos das platinelas do pandeiro presentes na faixa do tamborim.

Outro teste realizado a partir do sucesso da NMFD em cascata foi utilizar a NMFD uma vez apenas, selecionando um número de fontes maior do que a quantidade real de instrumentos presentes na mistura. Dessa forma, foram passadas 3 fontes para o algoritmo. As três faixas que aparecem na saída são o tamborim (sem artefatos), a parte grave do pandeiro (som do revestimento de couro) e as platinelas (parte aguda do pandeiro). A recuperação dos instrumentos é feita a partir da soma direta das partes pertencentes ao pandeiro. Além de demandar um tempo menor para recuperar 3 fontes (comparado com a estratégia em cascata, onde o algoritmo é realizado 2 vezes seguidas), em termos de qualidade de timbre, o resultado da NMFD com 3 fontes é superior aos dos outros esquemas de separação (www.smt.ufrj/~pedro.donadio - *Tabela Caso 3 - NMFD 1 e NMFD 2*).

6.7.4 Caso 4: mistura convolutiva de pandeiro de couro e tamborim

De maneira análoga ao caso 2, no caso 4, os mesmos instrumentos do caso 3 são utilizados. E, novamente, a mistura é realizada artificialmente, convoluindo os sons das fontes originais com a resposta ao impulso da quadra. A qualidade, no caso 4, fica ainda pior, havendo dificuldade, inclusive, de identificar quantos e quais são os instrumentos presentes.

Resultados

A condução dos experimentos do caso 4 foi bastante complexa, principalmente pela natureza da mistura. Mesmo após qualquer método de separação, a dificuldade de identificar as partes pertencentes a cada fonte, principalmente na utilização das sub-bandas, foi de difícil execução.

A NMFD com 2 fontes e 10 deslocamentos não teve sucesso em separar os instrumentos. As reverberações, de fato, dificultam o trabalho da NMFD, com o agravante, no caso 4, da sobreposição espectral. A NMFD com 3 fontes também não resolve a separação de maneira satisfatória. (www.smt.ufrj/~pedro.donadio - *Tabela Caso 4 - NMFD 1 e NMFD 2*)

Sub-bandas, seguido de NMFD (3 fontes) foi o sistema de melhor desempenho para o caso 4. As saídas recompostas a partir das 18 faixas da saída global do

sistema são capazes de separar ambos os instrumentos de maneira mais eficiente que os demais métodos. O fato é que para este caso crítico em que a resposta do ambiente degrada em demasia os sons do conjunto percussivo, a separação ocorre sempre com pequenos vestígios. Das duas fontes, a de melhor recuperação foi o pandeiro, com artefatos quase imperceptíveis do tamborim. Em sua faixa, o tamborim mantém o seu timbre original, mas o efeito do pandeiro (principalmente das platinelas) é percebido ao longo de toda a faixa (www.smt.ufrj/~pedro.donadio - *Tabela Caso 4 - MF NMFD 1 e MF NMFD 2*).

O áudio dos resultados completos aqui discutidos e dos demais testes realizados podem ser acessados em www.smt.ufrj/~pedro.donadio. Para fins de comparação objetiva, cada caso é descrito pelas Tabelas D.1 a D.4 no Apêndice D com as medidas de distorção, interferência e artefatos dos sinais separados.

Capítulo 7

Conclusões e trabalhos futuros

7.1 Contribuições

Esta dissertação investigou o tema de separação cega de fontes, aplicado a sinais musicais —mais especificamente, a sinais musicais de percussão brasileira. A abordagem apresentada envolve a descrição das técnicas atualmente utilizadas para separação cega de fontes, como a NMF e suas variantes e a filtragem do espectrograma por mediana. A essas técnicas são adicionadas combinações vantajosas, do ponto de vista da separação de percussão. Parte da contribuição desse trabalho foi a exploração da filtragem do espectrograma por mediana, no contexto desses instrumentos percussivos. A literatura registra a utilização do método como ferramenta de separação da parte harmônica da parte percussiva da música. Ainda que de naturezas diferentes, a técnica de filtragem do espectrograma por mediana seguida da fatoração das matrizes não-negativas apresentou resultados bastante promissores.

Outra contribuição desta dissertação foi a aplicação das técnicas de separação de fontes nas sub-bandas do sinal. O projeto de banco de filtros, organizado em oitavas, favoreceu a aplicação escolhida e gerou os melhores resultados em alguns casos. A exploração conjunta de todos os métodos estudados de sub-bandas, filtragem por mediana e NMF é especialmente favorável por permitir ajustes minuciosos para extração de resíduos e artefatos, por exemplo.

Por fim, a última contribuição desta dissertação foi a geração de uma base de dados de instrumentos de percussão brasileira. A gravação, feita em estúdio profissional e por músicos profissionais, teve o cuidado de registrar alguns dos instrumentos mais representativos da cultura musical brasileira. Os ritmos executados tiveram como referência o samba, atingindo também as suas variações como o partido alto e o samba de enredo, e outros, como a marcha e a capoeira. As gravações individuais e em conjunto dos sinais seguem disponíveis para a extensão da pesquisa.

Importante reiterar que o trabalho ainda investigou o desempenho dos métodos

em relação ao tipo de mistura. Parte da base de dados gravada foi modificada, artificialmente, através de *software* específico para simular a execução do conjunto percussivo em ambientes como uma casa, uma quadra de escola de samba e um estúdio de gravação. A quantidade de reverberação provocada por cada ambiente também foi determinante para a escolha de combinações de métodos específicos.

Em linhas gerais, o método escolhido depende, sobretudo, do tipo de mistura e da característica dos instrumentos envolvidos. Os casos apresentados demonstram a dificuldade da concepção de um modelo único, genérico suficiente para atender todas as variantes do sinal de entrada. Porém, algumas conclusões podem ser tiradas sobre a abrangência dos métodos.

NMFD: tem resultados melhores, sob as mesmas condições, que a NMF e a NMFD2D. Em comparação com os demais métodos abordados, atendeu essencialmente os requisitos para separar misturas geradas de maneira instantânea. Entretanto, principalmente para os casos com alto grau de reverberação (que tende a dificultar a recuperação dos instrumentos), o número de fontes passado ao algoritmo deve ser maior que a quantidade de instrumentos reais presentes na mistura.

Filtragem do espectrograma por mediana: funciona satisfatoriamente para as misturas onde os instrumentos possuem características bem distintas. O método lida bem com misturas entre instrumentos como o surdo, que tem claramente uma parcela permanente, e o reco-reco, que possui características transitórias, por exemplo.

NMFD aplicada às sub-bandas do sinal: a dificuldade de reconstrução é compensada pela versatilidade do resultado final. Isso ocorre porque, ao se inserir mais graus de liberdade para tratar o problema (como é o caso das sub-bandas), é possível combinar os resultados obtidos da separação nas diversas sub-bandas, dependendo do objetivo do experimento. Pode-se obter uma boa separação sem se buscar, por exemplo, uma preservação fiel do timbre do instrumento.

Filtragem do espectrograma por mediana aplicada às sub-bandas do sinal: em geral, os resultados subjetivos obtidos se assemelham aos da filtragem do espectrograma completo do sinal. A diferença principal é que trabalhar nas sub-bandas pode ser útil para eliminar alguns artefatos, principalmente de instrumentos concentrados em uma sub-banda.

Filtragem do espectrograma por mediana, seguida de NMFD: a combinação direta das técnicas de filtragem e fatoração facilita a separação dos instrumentos, ainda que haja sobreposição espectral. Esse é o incremento em relação à resposta gerada pela filtragem apenas. O controle em relação ao número de fontes passado como parâmetro para a NMFD é fundamental, variando de acordo com a necessidade da mistura.

Filtragem do espectrograma por mediana aplicada às sub-bandas do sinal, seguido de NMFD: é a combinação mais versátil, permitindo trabalhar com diferentes tipos

de sinais e misturas. Outro ponto positivo é a possibilidade de evitar artefatos, que foi aumentada. Entretanto, a quantidade de saídas que devem ser combinadas para a recuperação do sinal é maior e pode ficar impraticável, dependendo da quantidade de fontes passadas como parâmetro da NMFD.

7.2 Trabalhos futuros

A motivação para os trabalhos futuros é combinação de partes não exploradas no escopo deste trabalho e de questões que foram vislumbradas ao longo do desenvolvimento na busca de resultados mais aprimorados das técnicas propostas. Apesar de um objetivo inicial atingido, no que diz respeito à separação de fontes percussivas, algumas questões ficam abertas, a saber:

- *Incorporação de restrições à NMF*, visto que o caso abordado comparou apenas a versão original (NMF) e as versões NMFD e NMF2D. Podem ser impostas restrições adicionais à NMF, como esparsidade e continuidade temporal. Sua escolha deve ter como meta favorecer a separação de alguns instrumentos percussivos.
- A fim de *automatizar* a recuperação dos sinais separados, tarefa realizada manualmente neste trabalho, a ideia é o algoritmo fazer esse trabalho de maneira independente. Uma possível solução testada foi a correlação entre os ganhos de cada fonte, que não gerou resultados satisfatórios. Aprimorar essa etapa pode contribuir para maior rapidez na finalização dos experimentos. Na mesma filosofia, principalmente quando da utilização da separação em sub-bandas, relacionar as componentes separadas em uma sub-banda com as componentes separadas nas outras de maneira automática contribui na redução do tempo de simulação. Até uma possível avaliação prévia dos padrões presentes, não aplicando separação de fontes nas sub-bandas em que somente um instrumento estiver presente, pode diminuir o custo computacional.
- *Utilizar filtragem multitaxa* para gerar um maior número de sub-bandas, permitindo subdividir as frequências mais baixas. Como observado neste trabalho, dividir um sinal em sub-bandas tem resultado positivo para a separação das fontes da mistura, o que leva a crer que aumentar o número de sub-bandas analisadas pode melhorar os resultados. Talvez, um passo importante seja a combinação adequada da filtragem multitaxa com a recuperação automatizada, citada no item anterior. Alguns testes realizados com bancos de filtros multitaxas mostraram a impossibilidade de identificar (perceptivamente) a fonte, principalmente nas sub-bandas dos extremos.

Apêndice A

Base de dados completa

As tabelas a seguir apresentam informações sobre as faixas gravadas e disponíveis da base de dados gerada para este trabalho. Nas Tabelas A.1, A.2 e A.3 são mostradas as gravações acústicas com 4, 3 e 2 instrumentos, respectivamente. Nas Tabelas A.4 a A.32 são mostradas as gravações individuais de cada instrumento. Para este caso, as faixas são numeradas com 3 algarismos, sendo o primeiro o algarismo que identifica o músico e os dois restantes a numeração ordinária das faixas gravadas por ele.

Tabela A.1: Misturas com 4 instrumentos.

Faixa no.	Instrumentos da mistura			
	1	2	3	4
1	Pandeiro de couro 10"	Tantã corte 12"	Surdo 18"	Tamborim (baqueta de madeira)
2	Pandeiro de náilon 11"	Tantã corte 12"	Surdo 18"	Tamborim (baqueta de madeira)
3	Pandeiro de couro 10"	Tantã corte 12"	Tantã marcação 14"	Tamborim (baqueta de madeira)
4	Pandeiro de náilon 11"	Tantã corte 12"	Tantã marcação 14"	Tamborim (baqueta de madeira)
5	Pandeiro de náilon 11"	Tantã marcação 14"	Shake	Tamborim (baqueta de madeira)
6	Pandeiro de náilon 12"	Tantã corte 12"	Surdo 18"	Shake
7	Pandeiro de couro 10"	Surdo 18"	Reco-reco metálico 2 molas	Repique de mão 10"
8	Pandeiro de couro 10"	Surdo 18"	Reco-reco metálico 2 molas	Repique de anel 12"
9	Pandeiro de couro 10"	Tantã corte 12"	Surdo 18"	Agogô
10	Pandeiro de náilon 11"	Surdo 18"	Reco-reco metálico 2 molas	Caixa 12" 6 bordões
11	Tantã marcação 14"	Agogô	Tamborim (baqueta de náilon)	Caixa 12" 6 bordões
12	Surdo 18"	Caixa 12" 6 bordões	Cuíca 9,5"	Repinique 12"
13	Surdo 18"	Caixa 12" 6 bordões	Cuíca 9,5"	Tamborim (baqueta de náilon)
14	Surdo 18"	Cuíca 9,5"	Repinique 12"	Tamborim (baqueta de náilon)
15	Tantã marcação 14"	Agogô	Cuíca 9,5"	Shake (metálico)
16	Pandeiro de couro 10"	Pandeiro de náilon 11"	Tantã corte 12"	Surdo 18"
17	Pandeiro de couro 10"	Pandeiro de náilon 11"	Tantã marcação 14"	Repique de mão 10"
18	Pandeiro de náilon 11"	Tantã corte 12"	Repique de mão 10"	Shake
19	Agogô	Cuíca 9,5"	Repique de anel 12"	Shake

Tabela A.2: Misturas com 3 instrumentos.

Faixa no.	Instrumentos da mistura		
	1	2	3
20	Padeiro couro 10"	Padeiro náilon 12"	Tantã marcação 14"
21	Padeiro couro 10"	Tantã corte 12"	Tamborim (baqueta de madeira)
22	Padeiro nylon 12"	Surdo 18"	Tamborim (baqueta de madeira)
23	Padeiro couro 10"	Surdo 18"	Tamborim (baqueta de madeira)
24	Padeiro nylon 12"	Tantã corte 12"	Shake
25	Caixa 12" 6 bordões	Agogô	Repinique 12"
26	Caixa 12" 6 bordões	Cuíca 9,5"	Repinique 12"
27	Caixa 12" 6 bordões	Repinique 12"	Tamborim (baqueta de náilon)
28	Caixa 12" 6 bordões	Padeiro couro 10"	Surdo 18"
29	Tantã marcação 14"	Reco-reco metálico (2 molas)	Cuíca 9,5"
30	Surdo 18"	Agogô	Repique 10"
31	Padeiro couro 10"	Tantã marcação 14"	Repique 10"
32	Reco-reco metálico (2 molas)	Repique de anel 12"	Tamborim (baqueta de madeira)
33	Agogô	Repique de anel 12"	Shake
34	Padeiro nylon 10"	Surdo 18"	Reco-reco metálico (2 molas)
35	Tantã marcação 14"	Reco-reco metálico (2 molas)	Caixa 12" 6 bordões
36	Tantã marcação 14"	Repique 10"	Tamborim (baqueta de madeira)
37	Surdo 18"	Agogô	Shake
38	Padeiro nylon 12"	Tantã marcação 14"	Caixa 12" 6 bordões
39	Tantã corte 12"	Caixa 12" 6 bordões	Tamborim (baqueta de náilon)
40	Padeiro nylon 11"	Repique de anel 12"	Tamborim (baqueta de madeira)
41	Padeiro couro 10"	Surdo 18"	Surdo 16"
42	Padeiro nylon 11"	Surdo 18"	Surdo 16"
43	Reco-reco metálico (2 molas)	Surdo 18"	Surdo 16"
44	Agogô	Surdo 18"	Surdo 16"
45	Cuíca 9,5"	Surdo 18"	Surdo 16"
46	Tamborim (baqueta de náilon)	Surdo 18"	Surdo 16"
47	Caixa 12" 6 bordões	Surdo 18"	Surdo 16"
48	Repinique 12"	Surdo 18"	Surdo 16"

Tabela A.3: Misturas com 2 instrumentos.

Faixa no.	Instrumentos da mistura	
	1	2
49	Padeiro couro 10"	Padeiro náilon 11"
50	Tantã marcação 14"	Tantã corte 12"
51	Tantã corte 12"	Surdo 18"
52	Padeiro couro 10"	Surdo 18"
53	Padeiro náilon 11"	Surdo 18"
54	Padeiro náilon 11"	Tantã corte 12"
55	Padeiro náilon 11"	Tantã marcação 14"
56	Padeiro couro 10"	Tantã marcação 14"
57	Padeiro náilon 11"	Reco-reco metálico (2 molas)
58	Padeiro couro 10"	Reco-reco metálico (2 molas)
59	Tantã corte 12"	Reco-reco metálico (2 molas)
60	Surdo 18"	Reco-reco metálico (2 molas)
61	Caixa 12" 6 bordões	Agogô
62	Caixa 12" 6 bordões	Surdo 18"
63	Padeiro náilon 11"	Caixa 12" 6 bordões
64	Cuíca 9,5"	Repique 10"
65	Tamborim (baqueta de madeira)	Repique 10"
66	Caixa 12" 6 bordões	Shake
67	Agogô	Cuíca 6"
68	Agogô	Tamborim (baqueta de madeira)
69	Padeiro couro 10"	Tamborim (baqueta de madeira)
70	Tantã marcação 14"	Tamborim (baqueta de madeira)
71	Padeiro náilon 11"	Tamborim (baqueta de madeira)
72	Tantã marcação 14"	Cuíca 6"
73	Surdo 18"	Shake
74	Reco-reco metálico (2 molas)	Repique de anel 12"
75	Repique de anel 12"	Tamborim (baqueta de madeira)
76	Tantã corte 12"	Repinique 12"
77	Surdo 18"	Repinique 12"
78	Agogô	Repinique 12"
79	Reco-reco metálico (2 molas)	Repinique 12"
80	Reco-reco de madeira	Repinique 12"
81	Repique 10"	Repique de anel 12"
82	Repique 10"	Repinique 12"
83	Repique de anel 12"	Repinique 12"
84	Tantã marcação 14"	Repique 10"
85	Tantã marcação 14"	Repinique 12"
86	Tantã marcação 14"	Repique de anel 12"
87	Padeiro couro 10"	Caixa 12" 6 bordões
88	Padeiro couro 10"	Shake
89	Padeiro náilon 11"	Shake
90	Tantã corte 12"	Agogô
91	Surdo 18"	Tamborim (baqueta de madeira)
92	Surdo 18"	Reco-reco de madeira
93	Padeiro couro 10"	Tantã corte 12"

Tabela A.4: Gravações individuais - Pandeiro - músico 1.

Faixa no.	Instrumento	Tamanho	Revestimento	Ritmo	Bpm
101	PANDEIRO 1	10	SINTÉTICO	SAMBA	80
102	PANDEIRO 1	10	SINTÉTICO	PARTIDO ALTO	100
103	PANDEIRO 1	10	SINTÉTICO	SAMBA DE ENREDO	130
104	PANDEIRO 1	10	SINTÉTICO	MARCHA	120
105	PANDEIRO 1	10	SINTÉTICO	CAPOEIRA	65
106	PANDEIRO 1	10	SINTÉTICO	VIRADA I	100
107	PANDEIRO 1	10	SINTÉTICO	VIRADA II	120
108	PANDEIRO 2	10	COURO	SAMBA	80
109	PANDEIRO 2	10	COURO	PARTIDO ALTO	100
110	PANDEIRO 2	10	COURO	SAMBA DE ENREDO	130
111	PANDEIRO 2	10	COURO	MARCHA	120
112	PANDEIRO 2	10	COURO	CAPOEIRA	65
113	PANDEIRO 2	10	COURO	VIRADA I	100
114	PANDEIRO 2	10	COURO	VIRADA II	120
115	PANDEIRO 3	11	SINTÉTICO	SAMBA	80
116	PANDEIRO 3	11	SINTÉTICO	PARTIDO ALTO	100
117	PANDEIRO 3	11	SINTÉTICO	SAMBA DE ENREDO	130
118	PANDEIRO 3	11	SINTÉTICO	MARCHA	120
119	PANDEIRO 3	11	SINTÉTICO	CAPOEIRA	65
120	PANDEIRO 3	11	SINTÉTICO	VIRADA I	100
121	PANDEIRO 3	11	SINTÉTICO	VIRADA II	75
122	PANDEIRO 4	11	COURO	SAMBA	80
123	PANDEIRO 4	11	COURO	PARTIDO ALTO	100
124	PANDEIRO 4	11	COURO	SAMBA DE ENREDO	130
125	PANDEIRO 4	11	COURO	MARCHA	120
126	PANDEIRO 4	11	COURO	CAPOEIRA	65
127	PANDEIRO 4	11	COURO	VIRADA I	100
128	PANDEIRO 4	11	COURO	VIRADA II	75
129	PANDEIRO 5	12	SINTÉTICO	SAMBA	80
130	PANDEIRO 5	12	SINTÉTICO	PARTIDO ALTO	100
131	PANDEIRO 5	12	SINTÉTICO	SAMBA DE ENREDO	130
132	PANDEIRO 5	12	SINTÉTICO	MARCHA	120
133	PANDEIRO 5	12	SINTÉTICO	CAPOEIRA	65
134	PANDEIRO 5	12	SINTÉTICO	VIRADA I	100
135	PANDEIRO 5	12	SINTÉTICO	VIRADA II	75

Tabela A.5: Gravações individuais - Tamborim - músico 1.

Faixa no.	Instrumento	Tamanho	Revestimento	Baqueta	Ritmo	Bpm
136	TAMBORIM 1	6	SINTÉTICO	BAQUETA RÍGIDA	SAMBA	80
137	TAMBORIM 1	6	SINTÉTICO	BAQUETA RÍGIDA	PARTIDO ALTO	100
138	TAMBORIM 1	6	SINTÉTICO	BAQUETA RÍGIDA	SAMBA DE ENREDO	130
139	TAMBORIM 1	6	SINTÉTICO	BAQUETA MALEÁVEL	SAMBA DE ENREDO	130
140	TAMBORIM 1	6	SINTÉTICO	BAQUETA MALEÁVEL	VIRADA	130
141	TAMBORIM 2	6	COURO	BAQUETA RÍGIDA	SAMBA	80
142	TAMBORIM 2	6	COURO	BAQUETA RÍGIDA	PARTIDO ALTO	100
143	TAMBORIM 2	6	COURO	BAQUETA RÍGIDA	SAMBA DE ENREDO	130

Tabela A.6: Gravações individuais - Reco-reco - músico 1.

Faixa no.	Instrumento	Material	Mola	Ritmo	Bpm
144	RECO 1	METAL	2	SAMBA	80
145	RECO 1	METAL	2	PARTIDO ALTO	100
146	RECO 1	METAL	2	SAMBA DE ENREDO	130
147	RECO 1	METAL	2	VIRADA	100
148	RECO 2	MADEIRA	2	SAMBA	80
149	RECO 2	MADEIRA	2	PARTIDO ALTO	100
150	RECO 2	MADEIRA	2	SAMBA DE ENREDO	130
151	RECO 2	MADEIRA	2	VIRADA	100

Tabela A.7: Gravações individuais - Caixa - músico 1.

Faixa no.	Instrumento	Tamanho	Revestimento	Ritmo
152	CAIXA 1	12	SINTÉTICA	SAMBA
153	CAIXA 1	12	SINTÉTICA	PARTIDO ALTO
154	CAIXA 1	12	SINTÉTICA	SAMBA DE ENREDO
155	CAIXA 1	12	SINTÉTICA	MARCHA
156	CAIXA 1	12	SINTÉTICA	VIRADA

Tabela A.8: Gravações individuais - Repique - músico 1.

Faixa no.	Instrumento	Tamanho	Baqueta	Revestimento	Ritmo
157	REPIQUE 1	12	MADEIRA	SINTÉTICA	SAMBA
158	REPIQUE 1	12	MADEIRA	SINTÉTICA	PARTIDO ALTO
159	REPIQUE 1	12	MADEIRA	SINTÉTICA	SAMBA DE ENREDO
160	REPIQUE 1	12	MADEIRA	SINTÉTICA	VIRADA
161	REPIQUE 2	10	MÃO	SINTÉTICA	SAMBA
162	REPIQUE 2	10	MÃO	SINTÉTICA	PARTIDO ALTO
163	REPIQUE 2	10	MÃO	SINTÉTICA	SAMBA DE ENREDO
164	REPIQUE 2	10	MÃO	SINTÉTICA	VIRADA
165	REPIQUE 3	12	ANEL	COURO	SAMBA
166	REPIQUE 3	12	ANEL	COURO	PARTIDO ALTO
167	REPIQUE 3	12	ANEL	COURO	SAMBA DE ENREDO
168	REPIQUE 3	12	ANEL	COURO	VIRADA

Tabela A.9: Gravações individuais - Cuíca - músico 1.

Faixa no.	Instrumento	Tamanho	Revestimento	Ritmo	Bpm
169	CUICA 1	8	COURO	SAMBA	80
170	CUICA 1	8	COURO	PARTIDO ALTO	100
171	CUICA 1	8	COURO	SAMBA DE ENREDO	130
172	CUICA 1	8	COURO	VIRADA	130

Tabela A.10: Gravações individuais - Agogô - músico 1.

Faixa no.	Instrumento	Ritmo	Bpm
173	AGOGO1 2 NOTAS	SAMBA	80
174	AGOGO1 2 NOTAS	PARTIDO ALTO	100
175	AGOGO1 2 NOTAS	SAMBA DE ENREDO	130
176	AGOGO1 2 NOTAS	VIRADA	100

Tabela A.11: Gravações individuais - Shake - músico 1.

Faixa no.	Instrumento	Material	Ritmo	Bpm
177	SHAKE 1	MADEIRA	SAMBA	80
178	SHAKE 1	MADEIRA	PARTIDO ALTO	100
179	SHAKE 1	MADEIRA	SAMBA DE ENREDO	130
180	SHAKE 2	METAL	SAMBA	80
181	SHAKE 2	METAL	PARTIDO ALTO	100
182	SHAKE 2	METAL	SAMBA DE ENREDO	130

Tabela A.12: Gravações individuais - Tantã - músico 1.

Faixa no.	Instrumento	Tamanho	Revestimento	Ritmo	Bpm
183	TANTÃ 1 CORTE	11	COURO	SAMBA	80
184	TANTÃ 1 CORTE	11	COURO	PARTIDO ALTO	100
185	TANTÃ 1 CORTE	11	COURO	SAMBA DE ENREDO	130
186	TANTÃ 1 CORTE	11	COURO	VIRADA	100

Tabela A.13: Gravações individuais - Surdo - músico 1.

Faixa no.	Instrumento	Tamanho	Afinação	Revestimento	Ritmo	Bpm
187	SURDO 1	18	PRIMEIRA (C)	COURO	SAMBA	80
188	SURDO 1	18	PRIMEIRA (C)	COURO	PARTIDO ALTO	100
189	SURDO 1	18	PRIMEIRA (C)	COURO	SAMBA DE ENREDO	130
190	SURDO 1	18	PRIMEIRA (C)	COURO	VIRADA	100
191	SURDO 2	18	SEGUNDA (E)	COURO	SAMBA DE ENREDO	130

Tabela A.14: Gravações individuais - Pandeiro - músico 2.

Faixa no.	Instrumento	Tamanho	Revestimento	Ritmo	Bpm
201	PANDEIRO 2	10	COURO	SAMBA	80
202	PANDEIRO 2	10	COURO	PARTIDO ALTO	100
203	PANDEIRO 2	10	COURO	SAMBA DE ENREDO	130
204	PANDEIRO 2	10	COURO	MARCHA	120
205	PANDEIRO 2	10	COURO	CAPOEIRA	65
206	PANDEIRO 2	10	COURO	VIRADA I	100
207	PANDEIRO 2	10	COURO	VIRADA II	120
208	PANDEIRO 3	11	SINTÉTICO	SAMBA	80
209	PANDEIRO 3	11	SINTÉTICO	PARTIDO ALTO	100
210	PANDEIRO 3	11	SINTÉTICO	SAMBA DE ENREDO	130
211	PANDEIRO 3	11	SINTÉTICO	MARCHA	120
212	PANDEIRO 3	11	SINTÉTICO	CAPOEIRA	65
213	PANDEIRO 3	11	SINTÉTICO	VIRADA I	100
214	PANDEIRO 3	11	SINTÉTICO	VIRADA II	120
215	PANDEIRO 3	—	—	—	—
	VIRADA III – 130				
216	PANDEIRO 5	12	SINTÉTICO	SAMBA	80
217	PANDEIRO 5	12	SINTÉTICO	PARTIDO ALTO	100
218	PANDEIRO 5	12	SINTÉTICO	SAMBA DE ENREDO	130
219	PANDEIRO 5	12	SINTÉTICO	MARCHA	120
220	PANDEIRO 5	12	SINTÉTICO	CAPOEIRA	65
221	PANDEIRO 5	12	SINTÉTICO	PARTIDO ALTO	100
222	PANDEIRO 5	12	SINTÉTICO	VIRADA	120
223	PANDEIRO 6	12	COURO	SAMBA	80
224	PANDEIRO 6	12	COURO	PARTIDO ALTO	100
225	PANDEIRO 6	12	COURO	SAMBA DE ENREDO	130
226	PANDEIRO 6	12	COURO	MARCHA	120
227	PANDEIRO 6	12	COURO	CAPOEIRA	65
228	PANDEIRO 6	12	COURO	VIRADA I	100
229	PANDEIRO 6	12	COURO	VIRADA II	120

Tabela A.15: Gravações individuais - Tamborim - músico 2.

Faixa no.	Instrumento	Material	Baqueta	Ritmo	Bpm
230	TAMBORIM 3	SINTÉTICO	BAQUETA RÍGIDA	SAMBA	80
231	TAMBORIM 3	SINTÉTICO	BAQUETA RÍGIDA	PARTIDO ALTO	100
232	TAMBORIM 3	SINTÉTICO	BAQUETA RÍGIDA	SAMBA DE ENREDO	130
233	TAMBORIM 3	SINTÉTICO	BAQUETA RÍGIDA	VIRADA	100
234	TAMBORIM 3	SINTÉTICO	BAQUETA MALEÁVEL	SAMBA DE ENREDO	130
235	TAMBORIM 3	SINTÉTICO	BAQUETA MALEÁVEL	VIRADA	130

Tabela A.16: Gravações individuais - Reco-reco - músico 2.

Faixa no.	Instrumento	Material	Mola	Ritmo	Bpm
236	RECO 3	METAL	3	SAMBA	80
237	RECO 3	METAL	3	PARTIDO ALTO	100
238	RECO 3	METAL	3	SAMBA DE ENREDO	130
239	RECO 3	METAL	3	VIRADA	100
240	RECO 4	MADEIRA	3	SAMBA	80
241	RECO 4	MADEIRA	3	PARTIDO ALTO	100
242	RECO 4	MADEIRA	3	SAMBA DE ENREDO	130
243	RECO 4	MADEIRA	3	VIRADA	100

Tabela A.17: Gravações individuais - Caixa - músico 2.

Faixa no.	Instrumento	Tamanho	Bordões	Revestimento	Ritmo	Bpm
244	CAIXA 3	12	6	SINTÉTICA	SAMBA	80
245	CAIXA 3	12	6	SINTÉTICA	PARTIDO ALTO	100
246	CAIXA 3	12	6	SINTÉTICA	SAMBA DE ENREDO	130
247	CAIXA 3	12	6	SINTÉTICA	MARCHA	120
248	CAIXA 3	12	6	SINTÉTICA	VIRADA	130

Tabela A.18: Gravações individuais - Repique - músico 2.

Faixa no.	Instrumento	Tamanho	Baqueta	Revestimento	Ritmo	Bpm
249	REPIQUE 4	12	MADEIRA	SINTÉTICA	SAMBA	80
250	REPIQUE 4	12	MADEIRA	SINTÉTICA	PARTIDO ALTO	100
251	REPIQUE 4	12	MADEIRA	SINTÉTICA	SAMBA DE ENREDO	130
252	REPIQUE 4	12	MADEIRA	SINTÉTICA	VIRADA	130
253	REPIQUE 5	10	MÃO	SINTÉTICA	SAMBA	80
254	REPIQUE 5	10	MÃO	SINTÉTICA	PARTIDO ALTO	100
255	REPIQUE 5	10	MÃO	SINTÉTICA	SAMBA DE ENREDO	130
256	REPIQUE 5	10	MÃO	SINTÉTICA	VIRADA	100
257	REPIQUE 6	12	ANEL	COURO	SAMBA	80
258	REPIQUE 6	12	ANEL	COURO	PARTIDO ALTO	100
259	REPIQUE 6	12	ANEL	COURO	SAMBA DE ENREDO	130
260	REPIQUE 6	12	ANEL	COURO	VIRADA	100

Tabela A.19: Gravações individuais - Cuíca - músico 2.

Faixa no.	Instrumento	Revestimento	Ritmo	Bpm
261	CUICA 2	COURO	SAMBA	80
262	CUICA 2	COURO	PARTIDO ALTO	100
263	CUICA 2	COURO	SAMBA DE ENREDO	130
264	CUICA 2	COURO	VIRADA	100

Tabela A.20: Gravações individuais - Agogô - músico 2.

Faixa no.	Instrumento	Material	Ritmo	Bpm
265	AGOGO2 2 NOTAS	METAL	SAMBA	80
266	AGOGO2 2 NOTAS	METAL	PARTIDO ALTO	100
267	AGOGO2 2 NOTAS	METAL	SAMBA DE ENREDO	130
268	AGOGO2 2 NOTAS	METAL	VIRADA	100

Tabela A.21: Gravações individuais - Shake - músico 2.

Faixa no.	Instrumento	Material	Ritmo	Bpm
269	SHAKE 3	PLÁSTICO	SAMBA	80
270	SHAKE 3	PLÁSTICO	PARTIDO ALTO	100
271	SHAKE 3	PLÁSTICO	SAMBA DE ENREDO	130
272	SHAKE 3	PLÁSTICO	MARCHA	120
273	SHAKE 2	METAL	SAMBA	80
274	SHAKE 2	METAL	PARTIDO ALTO	100
275	SHAKE 2	METAL	SAMBA DE ENREDO	130
276	SHAKE 2	METAL	MARCHA	120

Tabela A.22: Gravações individuais - Tantã - músico 2.

Faixa no.	Instrumento	Tamanho	Revestimento	Ritmo	Bpm
277	TANTÃ 3	12	COURO	SAMBA	80
278	TANTÃ 3	12	COURO	PARTIDO ALTO	100
279	TANTÃ 3	12	COURO	SAMBA DE ENREDO	130
280	TANTÃ 3	12	COURO	VIRADA	120
281	TANTÃ 4	14	NAPA/NÁILON	SAMBA	80
282	TANTÃ 4	14	NAPA/NÁILON	PARTIDO ALTO	100
283	TANTÃ 4	14	NAPA/NÁILON	SAMBA DE ENREDO	130
284	TANTÃ 4	14	NAPA/NÁILON	MARCHA	120
285	TANTÃ 4	14	NAPA/NÁILON	VIRADA	120

Tabela A.23: Gravações individuais - Surdo - músico 2.

Faixa no.	Instrumento	Tamanho	Afinação	Revestimento	Ritmo	Bpm
286	SURDO 4	18	PRIMEIRA (C#)	COURO	SAMBA	80
287	SURDO 4	18	PRIMEIRA (C#)	COURO	PARTIDO ALTO	100
288	SURDO 4	18	PRIMEIRA (C#)	COURO	MARCHA	120
289	SURDO 4	18	PRIMEIRA (C#)	COURO	VIRADA	100
290	SURDO 4	20	PRIMEIRA (C#)	COURO	SAMBA	80
291	SURDO 4	20	PRIMEIRA (C#)	COURO	PARTIDO ALTO	100
292	SURDO 4	20	PRIMEIRA (C#)	COURO	SAMBA DE ENREDO	130
293	SURDO 4	20	PRIMEIRA (C#)	COURO	MARCHA	120
294	SURDO 4	20	PRIMEIRA (C#)	COURO	VIRADA	100
295	SURDO 5	20	SEGUNDA (E)	COURO	SAMBA DE ENREDO	130
296	SURDO 6	18	TERCEIRA (F#)	COURO	SAMBA DE ENREDO	130

Tabela A.24: Gravações individuais - Pandeiro - músico 3.

Faixa no.	Instrumento	Tamanho	Revestimento	Ritmo	Bpm
301	PANDEIRO 2	10	COURO	SAMBA	80
302	PANDEIRO 2	10	COURO	PARTIDO ALTO	100
303	PANDEIRO 2	10	COURO	SAMBA DE ENREDO	130
304	PANDEIRO 2	10	COURO	MARCHA	120
305	PANDEIRO 2	10	COURO	CAPOEIRA	65
306	PANDEIRO 2	10	COURO	VIR1	100
307	PANDEIRO 2	10	COURO	VIR2	130
308	PANDEIRO 3	11	SINTÉTICO	SAMBA	80
309	PANDEIRO 3	11	SINTÉTICO	PARTIDO ALTO	100
310	PANDEIRO 3	11	SINTÉTICO	SAMBA DE ENREDO	130
311	PANDEIRO 3	11	SINTÉTICO	MARCHA	120
312	PANDEIRO 3	11	SINTÉTICO	CAPOEIRA	65
313	PANDEIRO 3	11	SINTÉTICO	VIR1	100
314	PANDEIRO 3	11	SINTÉTICO	VIR2	130
315	PANDEIRO 5	12	SINTÉTICO	SAMBA	80
316	PANDEIRO 5	12	SINTÉTICO	PARTIDO ALTO	100
317	PANDEIRO 5	12	SINTÉTICO	SAMBA DE ENREDO	130
318	PANDEIRO 5	12	SINTÉTICO	MARCHA	120
319	PANDEIRO 5	12	SINTÉTICO	CAPOEIRA	65
320	PANDEIRO 5	12	SINTÉTICO	vir1	100
321	PANDEIRO 5	12	SINTÉTICO	VIRADA2	130

Tabela A.25: Gravações individuais - Tamborim - músico 3.

Faixa no.	Instrumento	Tamanho	Revestimento	Ritmo	Bpm
322	TAMBORIM 5	SINTÉTICO	BAQUETA RÍGIDA	SAMBA	80
323	TAMBORIM 5	SINTÉTICO	BAQUETA RÍGIDA	PARTIDO ALTO	100
324	TAMBORIM 5	SINTÉTICO	BAQUETA RÍGIDA	SAMBA DE ENREDO	130
325	TAMBORIM 5	SINTÉTICO	BAQUETA RÍGIDA	VIRADA	100
326	TAMBORIM 5	SINTÉTICO	BAQUETA MALEÁVEL	SAMBA DE ENREDO	130
327	TAMBORIM 5	SINTÉTICO	BAQUETA MALEÁVEL	VIRADA	130

Tabela A.26: Gravações individuais - Reco-reco - músico 3.

Faixa no.	Instrumento	Material	Ritmo	Bpm
328	RECO 5	METAL	SAMBA	80
329	RECO 5	METAL	PARTIDO ALTO	100
330	RECO 5	METAL	SAMBA DE ENREDO	130
331	RECO 5	METAL	MARCHA	120
332	RECO 5	METAL	VIRADA	100
333	RECO 5	METAL	BAIÃO	96
334	RECO 5	METAL	MAXIXE	106
335	RECO 5	METAL	VIRADA II	130

Tabela A.27: Gravações individuais - Caixa - músico 3.

Faixa no.	Instrumento	Tamanho	Bordões	Revestimento	Ritmo	Bpm
336	CAIXA 5	12	6	SINTÉTICA	SAMBA	80
337	CAIXA 5	12	6	SINTÉTICA	PARTIDO ALTO	100
338	CAIXA 5	12	6	SINTÉTICA	SAMBA DE ENREDO	130
339	CAIXA 5	12	6	SINTÉTICA	MARCHA 1	120
340	CAIXA 5	12	6	SINTÉTICA	MARCHA 2	120
341	CAIXA 5	12	6	SINTÉTICA	MARCHA 3	120
342	CAIXA 5	12	6	SINTÉTICA	VIRADA	130

Tabela A.28: Gravações individuais - Repique - músico 3.

Faixa no.	Instrumento	Tamanho	Bordões	Revestimento	Ritmo	Bpm
343	REPIQUE 4	12	MADEIRA	SINTÉTICA	SAMBA	80
344	REPIQUE 4	12	MADEIRA	SINTÉTICA	PARTIDO ALTO	100
345	REPIQUE 4	12	MADEIRA	SINTÉTICA	SAMBA DE ENREDO	130
346	REPIQUE 4	12	MADEIRA	SINTÉTICA	VIRADA 1	100
347	REPIQUE 4	12	MADEIRA	SINTÉTICA	VIRADA 2	130
348	REPIQUE 5	10	MÃO	SINTÉTICA	SAMBA	80
349	REPIQUE 5	10	MÃO	SINTÉTICA	PARTIDO ALTO	100
350	REPIQUE 5	10	MÃO	SINTÉTICA	SAMBA DE ENREDO	130
351	REPIQUE 5	10	MÃO	SINTÉTICA	VIRADA	100
352	REPIQUE 6	12	ANEL	COURO	SAMBA	80
353	REPIQUE 6	12	ANEL	COURO	PARTIDO ALTO	100
354	REPIQUE 6	12	ANEL	COURO	SAMBA DE ENREDO	130
355	REPIQUE 6	12	ANEL	COURO	VIRADA	100

Tabela A.29: Gravações individuais - Agogô - músico 3.

Faixa no.	Instrumento	Material	Ritmo	Bpm
356	AGOGO2 2 NOTAS	METAL	SAMBA	80
357	AGOGO2 2 NOTAS	METAL	PARTIDO ALTO	100
358	AGOGO2 2 NOTAS	METAL	SAMBA DE ENREDO	130
359	AGOGO2 2 NOTAS	METAL	VIRADA	100

Tabela A.30: Gravações individuais - Cuíca - músico 3.

Faixa no.	Instrumento	Material	Ritmo	Bpm
360	SHAKE 4	MADEIRA	SAMBA	80
361	SHAKE 4	MADEIRA	PARTIDO ALTO	100
362	SHAKE 4	MADEIRA	SAMBA DE ENREDO	130
363	SHAKE 4	MADEIRA	MARCHA	120
364	SHAKE 2	METAL	SAMBA	80
365	SHAKE 2	METAL	PARTIDO ALTO	100
366	SHAKE 2	METAL	SAMBA DE ENREDO	130
367	SHAKE 2	METAL	MARCHA	120

Tabela A.31: Gravações individuais - Tantã - músico 3.

Faixa no.	Instrumento	Tamanho	Revestimento	Ritmo	Bpm
368	TANTÃ 5	10	COURO	SAMBA	80
369	TANTÃ 5	10	COURO	PARTIDO ALTO	100
370	TANTÃ 5	10	COURO	SAMBA DE ENREDO	130
371	TANTÃ 5	10	COURO	VIRADA I	100
372	TANTÃ 5	10	COURO	VIRADA II	130
373	TANTÃ 6	14	NAPA/náilon	SAMBA	80
374	TANTÃ 6	14	NAPA/náilon	PARTIDO ALTO	100
375	TANTÃ 6	14	NAPA/náilon	SAMBA DE ENREDO	130
376	TANTÃ 6	14	NAPA/náilon	MARCHA	120
377	TANTÃ 6	14	NAPA/náilon	VIRADA	130

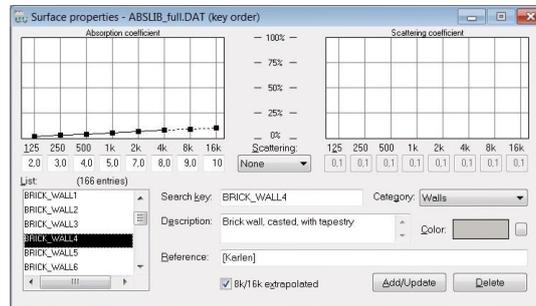
Tabela A.32: Gravações individuais - Surdo - músico 3.

Faixa no.	Instrumento	Tamanho	Afinação	Revestimento	Ritmo	Bpm
378	SURDO 7	20	PRIMEIRA (C)	COURO	SAMBA	80
379	SURDO 7	20	PRIMEIRA (C)	COURO	PARTIDO ALTO	100
380	SURDO 7	20	PRIMEIRA (C)	COURO	SAMBA DE ENREDO	130
381	SURDO 7	20	PRIMEIRA (C)	COURO	MARCHA	120
382	SURDO 7	20	PRIMEIRA (C)	COURO	VIRADA	100
383	SURDO 7	20	PRIMEIRA (C)	COURO	SAMBA	80
384	SURDO 7	20	PRIMEIRA (C)	COURO	PARTIDO ALTO	100
385	SURDO 7	20	SEGUNDA (E)	COURO	SAMBA DE ENREDO	130
386	SURDO 7	20	PRIMEIRA (C)	COURO	VIRADA	100
387	SURDO 9	20	TERCEIRA (G)	COURO	SAMBA DE ENREDO	130

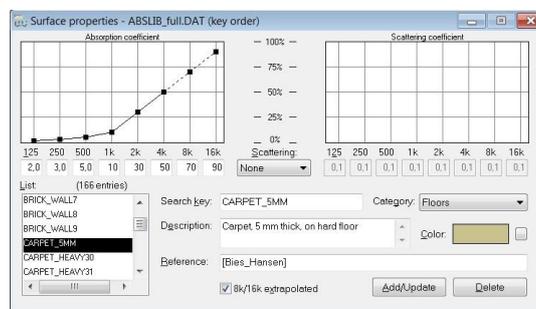
Apêndice B

Ambiente simulado: materiais e características

O CATT permite a seleção de vários tipos de materiais diferentes para a composição do ambiente. Além disso, é possível inserir novos materiais, ajustar curvas de reflexão e absorção acústicas e alterar a cor do material na representação 3D. As curvas e características dos materiais selecionados são apresentadas nas Figuras B.1 a B.9.

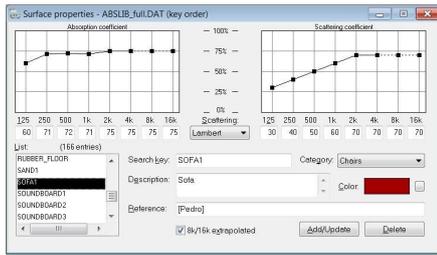


(a) Paredes e teto.

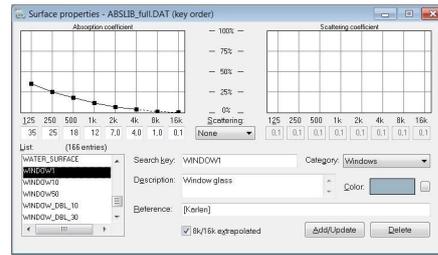


(b) Piso.

Figura B.1: Sala residencial.

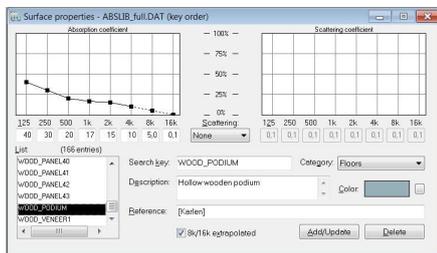


(a) Sofá.

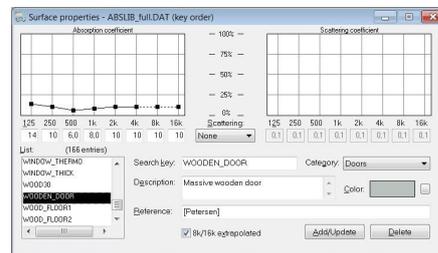


(b) Janela.

Figura B.2: Sala residencial.

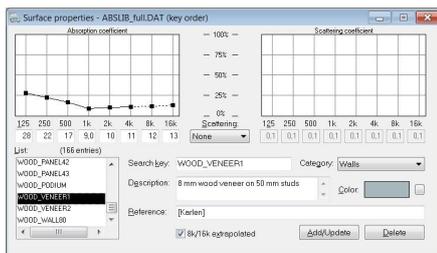


(a) Armário.

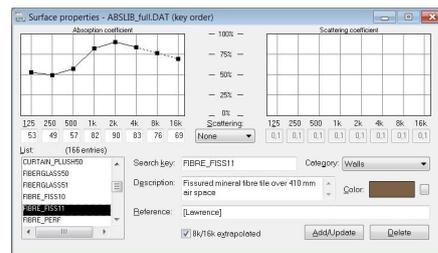


(b) Mesa de centro.

Figura B.3: Sala residencial.

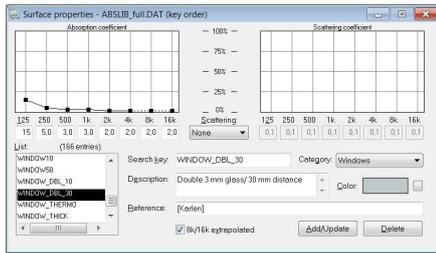


(a) Paredes.

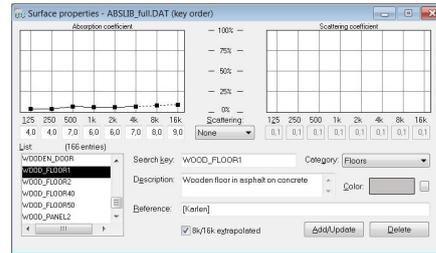


(b) Teto.

Figura B.4: Estúdio.

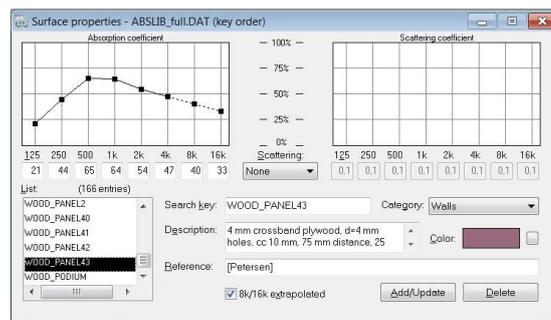


(a) Janela.



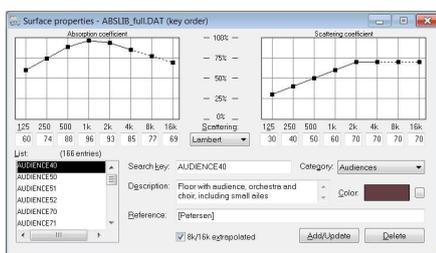
(b) Piso.

Figura B.5: Estúdio.

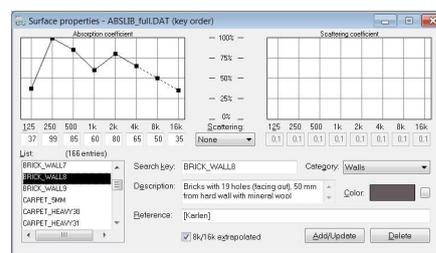


(a) Parede frontal (do piso até 1 m de altura).

Figura B.6: Estúdio.

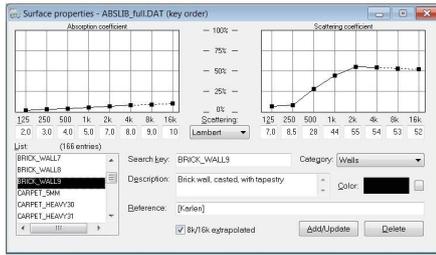


(a) Plateia.

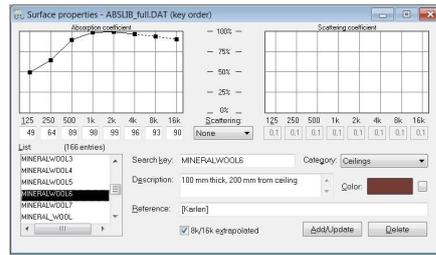


(b) Parede traseira.

Figura B.7: Quadra.

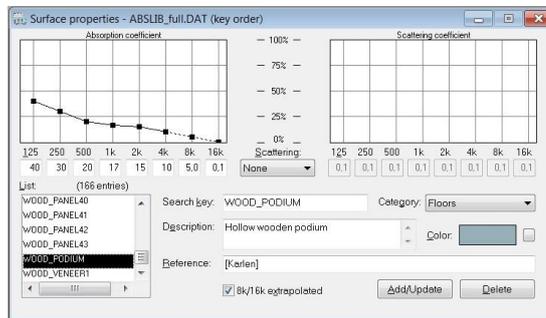


(a) Paredes laterais e frontal.



(b) Teto.

Figura B.8: Quadra.



(a) Palco.

Figura B.9: Quadra.

Apêndice C

Microfones

Tabela C.1: Microfones utilizados na captação dos instrumentos da base, a 20 centímetros de distância.

	Instrumento	Microfone	
		Marca	Modelo
1	Pandeiro náilon	Shure	SM57
2	Pandeiro Couro	Shure	Beta A57
3	Tamborim	Shure	SM81
4	Reco-reco	Shure	SM81
5	Caixa	Shure	Beta A57
6	Repique (mão e baqueta)	Shure	Beta A57
7	Repique de anel ¹	Shure	SM57
		Shure	Beta A57
8	Cuíca	Shure	SM81
9	Agogô	Shure	SM81
10	Shake	Shure	SM81
11	Tantã corte	Shure	Beta A57
12	Tantã marcação	AKG	D112
13	Surdo	AKG	D112

¹No caso do repique de anel, o microfone Beta A57 captou o som mais agudo do repique, produzido pelo toque do anel na lateral do instrumento, enquanto o microfone SM57 captou o som do couro (mais grave).

Apêndice D

Medidas objetivas

As medidas objetivas para avaliação de separação de fontes descritas em [25] e obtidas para os casos discutidos no Capítulo 6 são mostradas nas Tabelas D.1 a D.4.

Um sinal de áudio, modelado por

$$\hat{s} = s + e_i + e_a + e_r, \quad (\text{D.1})$$

onde s é o sinal original, e_a o artefato gerado pela separação, e_i a interferência causada por outras fontes e e_r o ruído (caso haja) na mistura, tem suas medidas de razão fonte-distorção SDR (*source-to-distortion ratio*), razão fonte-interferência SIR (*source-to-interference ratio*) e razão fonte-artefato SAR (*source-to-artifact ratio*) calculadas como

$$\text{SDR} = 10 \log \frac{\|s\|^2}{\|e_i + e_a + e_r\|^2} \quad (\text{D.2})$$

$$\text{SIR} = 10 \log \frac{\|s\|^2}{\|e_i\|^2} \quad (\text{D.3})$$

$$\text{SAR} = 10 \log \frac{\|s\|^2}{\|e_a\|^2}. \quad (\text{D.4})$$

As siglas nas tabelas representam os métodos MF=*Median filter*, SB NMFD = NMFD aplicado nas sub-bandas, MF NMFD = *Median filter*, seguido de NMFD e SB MF NMFD = *Median filter* aplicado nas sub-bandas, seguido de NMFD.

Para o caso 1, é possível observar, principalmente para a medida SIR, que o método de SB MF NMFD (escolhido como o melhor resultado subjetivo) é o que atinge o maior grau de separação para o surdo. Em relação às outras medidas é importante destacar que a característica de multiplicar o sinal original por máscaras complementares é o que possivelmente confere à MF, na maioria dos casos, os melhores resultados. Vale ressaltar que isso não necessariamente significa, em termos

Tabela D.1: Caso 1: SDR, SIR e SAR.

	SAR		SIR		SDR	
	Reco-reco	Surdo	Reco-reco	Surdo	Reco-reco	Surdo
NMFD	-13,979	-13,603	11,088	25,894	-14,317	-13,614
MF	3,435	18,618	11,043	45,277	2,458	18,608
SB NMFD	-13,342	-6,414	11,914	27,988	-13,625	-6,422
MF NMFD	-16,082	-8,820	6,290	36,853	-17,019	-8,821
SB MF NMFD	-13,944	18,520	11,350	46,629	-14,263	18,513

de percepção, uma boa separação. No caso 2, a SIR mostra o melhor resultado para o reco-reco com SB MF NMFD. No caso do surdo, MF NMFD obteve o melhor resultado dentre os demais métodos. No caso 3, a NMFD (como a avaliação subjetiva) obteve os melhores resultados. No entanto, as medidas objetivas de SIR, apesar de menores, mostram os valores de SIR para NMFD bem próximos dos melhores, principalmente para o pandeiro. No caso 4, os resultados objetivos ficam distantes do que foi avaliado subjetivamente. Por exemplo, a SIR da estratégia MF NMFD aparece com o melhor resultado. E, de fato, pouca interferência é percebida na separação do pandeiro. Contudo, o timbre do instrumento deixa a desejar e, por isso, como conjunto subjetivo de avaliação, a SB NMFD atinge o desempenho satisfatório.

Tabela D.2: Caso 2: SDR, SIR e SAR.

	SAR		SIR		SDR	
	Reco-reco	Surdo	Reco-reco	Surdo	Reco-reco	Surdo
NMFD	-15,977	-10,940	15,158	22,205	-16,110	-10,968
MF	6,086	27,553	13,473	30,667	5,196	25,824
SB NMFD	-19,487	-3,120	3,582	38,812	-21,080	-3,121
MF NMFD	-20,510	-7,041	3,912	32,343	-22,002	-7,044
SB MF NMFD	-11,353	-16,262	22,579	26,361	-11,378	-16,272

Tabela D.3: Caso 3: SDR, SIR e SAR.

	SAR		SIR		SDR	
	Pandeiro	Tamborim	Pandeiro	Tamborim	Pandeiro	Tamborim
NMFD	-5,082	-15,821	22,897	7,990	-5,111	-16,476
MF	13,791	15,877	24,034	5,822	13,383	5,312
SB MF	14,962	15,364	16,409	11,007	12,557	9,559
SB NMFD	-16,637	-7,785	12,772	7,858	-16,865	-8,544
MF NMFD	-6,354	-11,545	22,160	-3,082	-6,386	-16,564

Tabela D.4: Caso 4: SDR, SIR e SAR.

	SAR		SIR		SDR	
	Pandeiro	Tamborim	Pandeiro	Tamborim	Pandeiro	Tamborim
NMFD	-13,718	-12,583	11,862	7,587	-14,003	-13,316
MF	13,382	2,064	6,789	3,037	5,768	-1,545
SB MF	12,607	2,015	6,700	2,944	5,523	-1,630
SB NMFD	-19,041	-15,093	13,847	1,205	-19,218	-17,600
MF NMFD	-13,762	-17,186	23,061	2,914	-13,784	-19,007

Referências Bibliográficas

- [1] LINDSAY-SMITH, H., MCDONALD, S., SANDLER, M. “Real-time transcription and separation of drum recordings based on NMF decomposition”. In: *Proceedings of the 17th International Conference on Digital Audio Effects (DAFx-14)*, pp. 187–194, Erlangen, Setembro 2014.
- [2] SPICH, A., ZANONI, M., SARTI, A., et al. “Drum music transcription using prior subspace analysis and pattern recognition”. In: *Proceedings of the 13th International Conference on Digital Audio Effects (DAFx-10)*, pp. 1–5, Graz, Setembro 2010.
- [3] GILLET, O., RICHARD, G. “Transcription and separation of drum signals from polyphonic music”, *IEEE Transactions on Audio, Speech, and Language Processing*, v. 16, n. 3, pp. 529–540, Março 2008.
- [4] ÇATALTEPE, Z., ALTINEL, B. “Music recommendation based on adaptive feature and user grouping”. In: *Proceedings of the 22nd International Symposium on Computer and Information Sciences (ISCIS 2007)*, pp. 1–6, Ankara, Junho 2007. IEEE.
- [5] AZZERBONI, B., LA FORESTA, F., MAMMONE, N., et al. “A new approach based on wavelet-ICA algorithms for fetal electrocardiogram extraction”. In: *Proceedings of the 13th European Symposium on Artificial Neural Networks (ESANN2005)*, pp. 193–198, Bruges, Abril 2005.
- [6] NOSE FILHO, K. *Desconvolução e Separação Cega de Sinais Esparsos e Aplicações em Sísmica de Reflexão*. Tese de doutorado, Programa de Pós-Graduação em Engenharia Elétrica, Universidade Estadual de Campinas, Campinas, 2015.
- [7] KLAPURI, A., DAVY, M. *Signal Processing Methods for Music Transcription*. Nova York, Springer, 2006.
- [8] DINIZ, P. S. R., DA SILVA, E. A. B., NETTO, S. L. *Processamento Digital de Sinais – Projeto e Análise de Sistemas*. 2 ed. Porto Alegre, Bookman, 2014.

- [9] CAVACO, S., ALMEIDA, H. “Automatic cymbal classification using non-negative matrix factorization”. In: *Proceedings of the 19th International Conference on Systems, Signals and Image Processing (IWSSIP 2012)*, pp. 468–471, Viena, Setembro 2012. IEEE.
- [10] JANER, J., MARXER, R., ARIMOTO, K. “Combining a harmonic-based NMF decomposition with transient analysis for instantaneous percussion separation”. In: *Proceedings of the 2012 International Conference on Acoustics, Speech and Signal Processing (ICASSP 2012)*, pp. 281–284, Kyoto, Março 2012. IEEE.
- [11] YAZAWA, S., HAMANAKA, M., UTSURO, T. “Novel approach to separation of musical signal sources by NMF”. In: *Proceedings of the 12th International Conference on Signal Processing (ICSP 2014)*, pp. 610–615, HangZhou, Outubro 2014. IEEE.
- [12] VIRTANEN, T. “Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria”, *IEEE Transactions on Audio, Speech, and Language Processing*, v. 15, n. 3, pp. 1066–1074, Março 2007.
- [13] BECKER, J. M., SOHN, C., ROHLFING, C. “NMF with spectral and temporal continuity criteria for monaural sound source separation”. In: *Proceedings of the 22nd European Signal Processing Conference (EUSIPCO 2014)*, pp. 316–320, Lisboa, Setembro 2014. EURASIP.
- [14] LEE, D. D., SEUNG, H. S. “Algorithms for non-negative matrix factorization”. In: *Proceedings of the 13th Neural Information Processing Systems Conference (NIPS 2000)*, pp. 556–562, Vancouver, Junho 2000. NIPS.
- [15] ANTONIOU, A., LU, W. S. *Practical Optimization: Algorithms and Engineering Applications*. Nova Iorque, Springer, 2007.
- [16] SMARAGDIS, P. “Non-negative matrix factor deconvolution; Extraction of Multiple Sound Sources from Monophonic Inputs”. In: *Proceedings of the 5th International Conference on Independent Component Analysis and Blind Signal Separation (ICA 2004)*, pp. 494–499, Grenada, Setembro 2004.
- [17] SCHMIDT, M. N., MØRUP, M. “Nonnegative matrix factor 2-D deconvolution for blind single channel source separation”. In: *Proceedings of the 6th International Conference on Independent Component Analysis and Blind Signal Separation (ICA 2006)*, pp. 700–707, Charleston, Março 2006.

- [18] LINDSAY-SMITH, H., MCDONALD, S., SANDLER, M. “Drumkit transcription via convolutive NMF”. In: *Proceedings of the 15th International Conference on Digital Audio Effects (DAFx-12)*, pp. 1–4, York, Setembro 2012.
- [19] ALMEIDA, R. M. *Separação de Fontes Sonoras por Fatoração Duplamente Deconvolutiva de Matrizes Não-Negativas com Uso de Restrições*. Dissertação de mestrado, Programa de Engenharia Elétrica, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2014.
- [20] FITZGERALD, D. “Harmonic/percussive separation using median filtering”. In: *Proceedings of the 13th International Conference on Digital Audio Effects (DAFx-10)*, pp. 1–4, Graz, Setembro 2010.
- [21] BAYARRES, I. I. *Transient and steady-state component separation for audio signals*. Dissertação de mestrado, Facultad de Ingeniería, Universidad de la República, Montevidéo, 2014.
- [22] GRIFFIN, D. W., LIM, J. S. “Signal estimation from modified short-time Fourier transform”, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, v. 32, n. 3, pp. 236–243, Abril 1984.
- [23] ZHU, X., BEAUREGARD, G. T., WYSE, L. L. “Real-time signal estimation from modified short-time Fourier transform magnitude spectra”, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, v. 15, n. 5, pp. 1645–1653, Julho 2007.
- [24] ZHU, X., BEAUREGARD, G. T., WYSE, L. L. “An efficient algorithm for real-time spectrogram inversion”. In: *Proceedings of the 8th Conference on Digital Audio Effects (DAFX-05)*, pp. 116–121, Espanha, Setembro 2007.
- [25] VINCENT, E., GRIBONVAL, R., FÉVOTTE, C. “Performance measurement in blind audio source separation”, *IEEE Transactions on Audio, Speech, and Language Processing*, v. 14, n. 4, pp. 1462–1469, Julho 2006.