



PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS
Programa de Pós-graduação em Engenharia Elétrica

**ANÁLISE E APLICAÇÃO DAS TÉCNICAS DE FATORAÇÃO DE
MATRIZES E TENSORES NÃO NEGATIVOS PARA A
SEPARAÇÃO DE SINAIS SONOROS EM MISTURAS
CONTENDO ELEMENTOS HARMÔNICOS E PERCUSSIVOS**

Wellington Alexander Vieira da Fonseca

Dissertação de Mestrado PPGEE-176/2016

Orientadora: Prof.^a Dra. Zélia Myriam Assis Peixoto
Coorientadora: Prof.^a Dra. Flávia Magalhães Freitas Ferreira
Coorientador: Prof. Dr. Roberto de Maria Nunes Mendes

fevereiro de 2016



PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS
Programa de Pós-graduação em Engenharia Elétrica

WELLINGTON ALEXANDER VIEIRA DA FONSECA

**ANÁLISE E APLICAÇÃO DAS TÉCNICAS DE FATORAÇÃO DE
MATRIZES E TENSORES NÃO NEGATIVOS PARA A SEPARAÇÃO DE
SINAIS SONOROS EM MISTURAS CONTENDO ELEMENTOS
HARMÔNICOS E PERCUSSIVOS**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica, da Pontifícia Universidade Católica de Minas Gerais, como requisito parcial para obtenção do grau de Mestre em Engenharia Elétrica, elaborada sob a orientação da Prof.^a Dra. Zélia Myriam Assis Peixoto; Coorientadora: Prof.^a Dra. Flávia Magalhães Freitas Ferreira e Coorientador: Prof. Dr. Roberto de Maria Nunes Mendes.

Belo Horizonte
2016

FICHA CATALOGRÁFICA

Elaborada pela Biblioteca da Pontifícia Universidade Católica de Minas Gerais

F676a Fonseca, Wellington Alexander
Análise e aplicação das técnicas de fatoração de matrizes e tensores não negativos para a separação de sinais sonoros em misturas contendo elementos harmônicos e percussivos / Wellington Alexander Vieira da Fonseca. Belo Horizonte, 2016.
99 f. : il.

Orientadora: Zélia Myriam Assis Peixoto
Coorientadora: Flávia Magalhães Freitas Ferreira
Coorientador: Roberto de Maria Nunes Mendes
Dissertação (Mestrado) – Pontifícia Universidade Católica de Minas Gerais.
Programa de Pós-Graduação em Engenharia Elétrica.

1. Filtros elétricos. 2. Processamento de sinais - Técnicas digitais. 3. Separação Cega de Fontes. 4. MATLAB (Programa de computador). I. Peixoto, Zélia Myriam Assis. II. Ferreira, Flávia Magalhães Freitas. III. Mendes, Roberto de Maria Nunes. IV. Pontifícia Universidade Católica de Minas Gerais. Programa de Pós-Graduação em Engenharia Elétrica. V. Título.

CDU: 621.391



PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS

“ANÁLISE E APLICAÇÃO DAS TÉCNICAS DE FATORAÇÃO DE MATRIZES E TENSORES NÃO NEGATIVOS PARA A SEPARAÇÃO DE SINAIS SONOROS EM MISTURAS CONTENDO ELEMENTOS HARMÔNICOS E PERCUSSIVOS”

WELLINGTON ALEXANDER VIEIRA DA FONSECA

Dissertação de Mestrado submetida à banca examinadora designada pelo Colegiado do Programa de Pós-Graduação em Engenharia Elétrica da Pontifícia Universidade Católica de Minas Gerais, como parte dos requisitos necessários à obtenção do grau de Mestre em Engenharia Elétrica.

Aprovada em 01 de fevereiro de 2016.

Por:

Prof. Regis Rossi Alves Faria, Dr., USP

Profa. Zélia Myriam Assis Peixoto, Dra. - orientadora
PPGEE, PUC-Minas

Prof. Carlos Augusto Paiva da Silva Martins, Dr.,
PPGEE, PUC-Minas.

Profa. Flávia Magalhães Freitas Ferreira, Dra. – coorientadora,
Suplente- PPGEE, PUC Minas

A todos aqueles que contribuíram, direta ou indiretamente, para o desenvolvimento desta pesquisa

AGRADECIMENTOS

Em primeiro lugar agradeço a Deus por tornar possível todas as conquistas alcançadas.

Aos meus pais e minha irmã pelo apoio incondicional.

Aos meus padrinhos por confiarem em mim e estarem sempre dispostos a ajudar no possível e no impossível.

Às minhas avós pelo carinho e as incontáveis orações.

À minha orientadora, Dra. Zélia, pela paciência e apoio em todo esse período.

À minha coorientadora, Dra. Flávia, pelos ensinamentos técnicos e, sobretudo, os ensinamentos de vida. Ela, em conjunto com o Dr. Carlos, ajudaram-me a dar mais atenção ao mundo que existe além da engenharia.

Ao meu coorientador, Dr. Roberto, pela paciência e boa vontade em ajudar a tornar o bruto algo mais elegante.

Ao amigo Marcelo, precursor deste ramo de pesquisa no PPGEE que me passou muito do conhecimento técnico e artístico necessários para o desenvolvimento deste trabalho.

Aos demais amigos e colegas por estarem presentes quando necessário.

Por último e, não menos importante, agradeço à CAPES pela concessão da bolsa de estudos.

It ain't about how hard you hit. It's about how hard you can get hit
(Rocky Balboa)

RESUMO

Este trabalho trata da separação de sinais acústicos, em formatos mono e estéreo, compostos por instrumentos harmônicos e percussivos, especificamente, guitarra, baixo, prato de condução, bumbo e caixa. Para obter a separação dos diferentes sinais, utilizaram-se as técnicas de Fatoração de Matrizes Não Negativas (NMF – Non-negative Matrix Factorization) e Fatoração de Tensores Não Negativos, ambas associadas às divergências de Kullback-Leibler e de Itakura-Saito. A definição das técnicas aplicadas à separação dos sinais de áudio foi realizada a partir de uma ampla revisão bibliográfica, destacando-se o estado da arte sobre o tema. Em seguida, e após ter sido constatada a ausência de material bibliográfico dedicado ao desenvolvimento e aplicação dessas técnicas em engenharia de áudio, foram realizados estudos detalhados sobre as mesmas com a finalidade de se obter um maior entendimento das regras de atualização e aplicação dos algoritmos, além da delimitação das especificidades de cada uma das técnicas abordadas. Os algoritmos referentes às técnicas NMF e NTF foram implementados em ambiente MATLAB[®] e, a fim de se mensurar a qualidade da separação, foram realizadas análises comparativas entre os resultados obtidos com base nos critérios clássicos derivados da Relação Sinal-Ruído (SNR) e por meio de uma variação proposta para essas métricas, que inclui o janelamento com superposição de parcelas subsequentes dos sinais. Esse formato das métricas, denominado métrica janelada, permitiu avaliar a evolução das métricas ao longo do tempo e com base nesses resultados, foi proposto um novo método de separação, semi-supervisionado e recursivo que pressupõe uma etapa de treinamento, utilizando-se como informação *a priori* parcelas selecionadas do próprio sinal a ser submetido à separação. Os resultados foram avaliados, definindo-se os contextos de melhor desempenho das diversas técnicas e, posteriormente, realizada a restauração das misturas a partir dos sinais resultantes da separação. Pôde-se, com base nos estudos e resultados obtidos, viabilizar e validar a aplicação das técnicas mencionadas para a separação de sinais acústicos digitais.

Palavras-chave: Separação de sons. Elementos Harmônicos. Elementos Percussivos. Processamento de sinais de áudio. Fatoração de Matrizes Não Negativas. Fatoração de Tensores Não Negativos.

ABSTRACT

This work aims to study the techniques applied for sound source separation applied to harmonic and percussive elements (namely electric guitar, bass, ride, kick and snare). Both Mono and Stereo formats were considered, using the techniques of Non-negative Matrix Factorization (NMF) and Non-negative Tensor Factorization (NTF), respectively. Associated to these techniques, the Kullback-Leibler and Itakura-Saito Divergences were tested. A detailed study of the two techniques was conducted, both in conceptual and mathematical perspectives, aiming achieve a better understanding of the update rules deduction process and its application methods, and to better comprehend the peculiarities involved in each process. A literature review was conducted to analyze the state of the art in this field. The NMF and NTF algorithms were developed in Matlab, and a comparative analysis based on Signal-to-Noise Ratio (SNR) between the two methods was executed, to measure their performances. An alternative version for the measurement is proposed, based on its windowing to better adequate them to the separated signal. This approach also allowed the development of a method that is semi-supervised and recursive, since the training information is extracted from the original signal itself. Ultimately, aspects in which the techniques could be improved are highlighted, opening the field to future studies.

Keywords: Sound Separation. Harmonic Elements. Percussive Elements. Audio Signal Processing. Non-Negative Matrix Factorization. Non-Negative Tensor Factorization.

LISTA DE ILUSTRAÇÕES

Figura 1 – Distribuição ao longo dos anos das publicações na área de BASS . . .	32
Figura 2 – Indexação de Tensores	42
Figura 3 – Convexidade da função $\phi(p) = p \ln p$	45
Figura 4 – Exemplo: Variação da Divergência de Kullback-Leibler	47
Figura 5 – Convexidade da função $\phi = -\ln p$	47
Figura 6 – Exemplo: Variação da Divergência de Itakura-Saito	49
Figura 7 – Comparação entre as Variações das Divergências de Kullback-Leibler e Itakura-Saito	49
Figura 8 – Convergência da NMF para a Divergência de Kullback-Leibler	50
Figura 9 – Oscilação da NMF usando Divergência de Kullback-Leibler	51
Figura 10 – NMF Supervisionada	52
Figura 11 – Modelo de decomposição da NMF no contexto de áudio	55
Figura 12 – Bases do espaço tensorial	57
Figura 13 – Decomposição de Tensores	57
Figura 14 – Fluxograma da Simulação	66
Figura 15 – Espectrogramas da Mistura <i>Clean</i>	68
Figura 16 – Espectrogramas da Mistura <i>Delay</i>	69
Figura 17 – Espectrogramas da Mistura Distorção	69
Figura 18 – Espectrogramas da Guitarra <i>Clean</i>	70
Figura 19 – Espectrogramas da Guitarra <i>Delay</i>	71
Figura 20 – Espectrogramas da Guitarra Distorcida	71
Figura 21 – Espectrogramas do Baixo	72
Figura 22 – Espectrogramas do Bumbo	72
Figura 23 – Espectrogramas da Caixa	73
Figura 24 – Espectrogramas do Prato de Condução	73
Figura 25 – Espectrogramas da guitarra <i>clean</i> - (a) original; (b) separada via NMF-KL	77
Figura 26 – Espectrograma da guitarra <i>delay</i> - (a) original; (b) separada via NMF-KL	77
Figura 27 – Espectrograma da guitarra <i>dist</i> - (a) original; (b) separada via NMF-KL	77
Figura 28 – Comparação dos resultados da NTF usando as Divergências de Kullback- Leibler e de Itakura-Saito para <i>Clean.wav</i>	79
Figura 29 – Comparação dos resultados da NTF usando as Divergências de Kullback- Leibler e de Itakura-Saito para <i>Delay.wav</i>	79
Figura 30 – Comparação dos resultados da NTF usando as Divergências de Kullback- Leibler e de Itakura-Saito para <i>Dist.wav</i>	80
Figura 31 – Espectrogramas dos canais L e R da guitarra <i>clean</i> separada via NTF-KL	80
Figura 32 – Espectrogramas dos canais L e R da guitarra <i>delay</i> separada via NTF-IS	81

Figura 33 – Espectrogramas dos canais L e R da guitarra distorcida separada via NTF-KL	81
Figura 34 – Formas de onda dos sinais da guitarra <i>clean</i> , na versão original e após separação via NTF-KL.	82
Figura 35 – Formas de onda dos sinais do prato de condução <i>clean</i> , na versão original e após separação via NTF-KL.	82
Figura 36 – Aplicação da SDR Janelada à mistura <i>Dist.wav</i> , utilizando-se a técnica NMF-KL	86
Figura 37 – Aplicação da SIR Janelada à mistura <i>Dist.wav</i> , utilizando-se a técnica NMF-KL	86
Figura 38 – Aplicação da SDR Janelada à mistura <i>Dist.wav</i> , utilizando-se a técnica NMF-IS	87
Figura 39 – Aplicação da SIR Janelada à mistura <i>Dist.wav</i> , utilizando-se a técnica NMF-IS	87
Figura 40 – Comparação entre as métricas janeladas e clássicas: Som da guitarra na mistura <i>Dist.wav</i> via NMF-KL	88
Figura 41 – Comparação entre as métricas janeladas e clássicas: Som da guitarra na mistura <i>Dist.wav</i> via NMF-IS	88
Figura 42 – Comparação das métricas SIR Janeladas para o Baixo e Bumbo, na mistura <i>Dist.wav</i> via NMF-KL	88
Figura 43 – Comparação das métricas SIR Janeladas para o Baixo e Bumbo, na mistura <i>Dist.wav</i> via NTF-IS	89

LISTA DE TABELAS

Tabela 1	–	Especificações do microfone MX150	65
Tabela 2	–	Parâmetros de mixagem dos sinais de teste estéreo	66
Tabela 3	–	Parâmetros de Ajuste das técnicas NMF e NTF	67
Tabela 4	–	Número de Notas para a Simulação da NTF	68
Tabela 5	–	Resultado da Separação via NMF	76
Tabela 6	–	Resultado da Separação via NTF	78
Tabela 7	–	SAR dos sinais reconstruídos	83
Tabela 8	–	Comparação entre as técnicas NMF e NTF: $\Delta_S = S_{NTF} - S_{NMF}$. . .	83
Tabela 9	–	Parâmetros da métrica janelada	85
Tabela 10	–	Comparação NMF e NMF Semi-Supervisionada Recursiva, utilizando-se a mistura <i>Clean</i> e com treinamento via elementos percussivos.	91
Tabela 11	–	Comparação NMF e NMF Semi-Supervisionada Recursiva, com prato e caixa treinado - Clean	92

GLOSSÁRIO

Acrônimos

ALS ... Mínimos Quadrados Alternados (*Alternating Least Squares*)

BASS ... Separação cega de fontes de áudio (*Blind Audio Source Separation*)

BSS ... Separação cega de sinais (*Blind Source Separation*)

CANDECOMP ... Decomposição Canônica (*Canonical Decomposition*)

CASA ... Análise Computacional de Cenas Auditivas (*Computational Auditory Scene Analysis*)

CPD ... Decomposição Canônica Poliádica (*Canonical Polyadic Decomposition*)

dB ... Decibel

dB SPL ... Nível de Pressão Sonora em dB (*dB Sound Pressure Level*)

DFT ... Transformada Discreta de Fourier (*Discrete Fourier Transform*)

DTFT ... Transformada de Fourier de Tempo Discreto (*Discrete Time Fourier Transform*)

FFT ... Transformada Rápida de Fourier (*Fast Fourier Transform*)

FT ... Transformada de Fourier (*Fourier Transform*)

ICA ... Análise de Componentes Independentes (*Independent Component Analysis*)

IS ... Divergência de Itakura-Saito (*Itakura-Saito Divergence*)

ISS ... Separação não cega de sinais (*Informed Source Separation*)

KL ... Divergência de Kullback-Leibler (*Kullback-Leibler Divergence*)

MIDI ... Interface Digital para Instrumentos Digitais (*Musical Instrument Digital Interface*)

NeNMF ... Método Ótimo para Fatoração de Matrizes Não Negativas (*Optimal Gradient Method for Non-Negative Matrix Factorization*)

NMF ... Fatoração de Matrizes Não Negativas (*Nonnegative Matrix Factorization*)

NTF ... Fatoração de Tensores Não Negativos (*Nonnegative Tensor Factorization*)

PARAFAC ... Análise de Fatores Paralelos (*Parallel Factor Analysis*)

PCA ... Análise de Componente Principal (*Principal Component Analysis*)

PMF ... Fatoração de Matrizes Positivas (*Positive Matrix Factorization*)

PSNMF ... Fatoração Supervisionada Penalizada de Matrizes Não Negativas (*Penalized Supervised Non-Negative Matrix Factorization*)

relint ... Interior Relativo (*Relative Interior*)

SAR ... Relação Fonte-Artefato (*Sources to Artifacts Ratio*)

SDR ... Relação Fonte-Distorção (*Source to Distortion Ratio*)

SIR ... Relação Fonte-Interferência (*Source to Interference Ratio*)

SoNR ... Relação Fontes-Ruído (*Source to Noise Ratio*)

SNR ... Relação Sinal-Ruído (*Signal-to-Noise Ratio*)

STFT ... Transformada de Fourier de Curta Duração (*Short-Time Fourier Transform*)

SVD ... Decomposição em Valores Singulares (*Singular Value Decomposition*)

Símbolos

$\mathcal{A}$... Tensor dos padrões espectrais de cada fonte presente na mistura.

$\mathcal{G}$... Tensor dos ganhos de mixagem de cada fonte presente na mistura.

$\mathbf{H}$... Matriz dos ganhos de cada fonte, após separação.

$\mathcal{S}$... Tensor dos ganhos de cada fonte, após separação.

$\widehat{\mathbf{V}}$... Matriz que representa a aproximação do espectrograma original da mistura.

$\mathbf{V}$... Matriz que contém o espectrograma original da mistura.

$\widehat{\mathcal{X}}$... Tensor que representa a aproximação do espectrograma original da mistura.

$\mathcal{X}$... Tensor que contém o espectrograma original da mistura.

$\mathbf{W}$... Matriz dos padrões espectrais de cada fonte contida na mistura, após separação.

α ... Fator de minimização do Método Gradiente Descendente.

$\odot$... Operador que representa o Produto de Hadamard.

$\otimes$... Operador que representa o Produto Exterior.

∇f ... Gradiente de uma função.

SUMÁRIO

1	INTRODUÇÃO	31
1.1	Contextualização do Problema	31
1.2	Problema Motivador e Soluções Propostas	32
1.3	Objetivos e Metas	33
1.4	Justificativa	34
1.5	Escopo da Pesquisa	35
1.6	Contribuições da Pesquisa	35
1.7	Organização do Texto	36
2	FUNDAMENTOS TEÓRICOS E TRABALHOS RELACIONADOS	37
2.1	Fundamentos de Áudio	37
2.2	Tensores	38
2.2.1	<i>Produto Tensorial</i>	39
2.2.2	<i>Contração de Tensores</i>	40
2.2.3	<i>Produto Contraído</i>	40
2.3	Transformada de Fourier de Curta Duração	42
2.4	Divergências Utilizadas	43
2.4.1	<i>Funções Convexas</i>	44
2.4.2	<i>Divergência de Kullback-Leibler</i>	45
2.4.3	<i>Divergência de Itakura-Saito</i>	47
2.5	Métodos de otimização	49
2.6	Trabalhos Relacionados	51
2.7	Conclusão	54
3	DEDUÇÃO MATEMÁTICA DAS REGRAS DE ATUALIZAÇÃO DOS ALGORITMOS NMF E NTF	55
3.1	Non-Negative Matrix Factorization (NMF) e Non-Negative Tensor Factorization (NTF)	55
3.2	NMF usando Divergência de Kullback-Leibler	58
3.3	NMF usando Divergência de Itakura-Saito	60
3.4	NTF usando Divergência de Kullback-Leibler	61
3.5	NTF usando Divergência de Itakura-Saito	63
3.6	Conclusão	63
4	RESULTADOS COMPUTACIONAIS	65
4.1	Metodologia	65

4.2	Áudios de Teste	67
4.3	Métodos de Validação	70
<i>4.3.1</i>	<i>Avaliação Objetiva Baseada na SNR</i>	<i>74</i>
4.4	Análise dos Resultados	75
<i>4.4.1</i>	<i>Análise da separação dos sinais monoaurais</i>	<i>75</i>
<i>4.4.2</i>	<i>Análise da separação dos sinais estéreo</i>	<i>78</i>
4.5	Análise comparativa das técnicas NTF e NMF	80
<i>4.5.1</i>	<i>Particularidades da NTF</i>	<i>84</i>
4.6	Métricas Janeladas	85
4.7	Técnica NMF Semi-supervisionado Recursiva	89
4.8	Conclusão	90
5	CONSIDERAÇÕES FINAIS	93
	 REFERÊNCIAS	 95

1 INTRODUÇÃO

Este capítulo apresenta uma visão geral da pesquisa realizada, com destaque para os pontos-chave que definiram a escolha e o desenvolvimento do tema, como a contextualização do problema, as justificativas do trabalho, as metas e os objetivos, dentre outros. São apresentadas, também, as principais contribuições agregadas pela pesquisa e a organização o texto.

1.1 Contextualização do Problema

A concepção e evolução dos diversos tipos de separação de sinais de áudio baseiam-se em um fenômeno psicoacústico, conhecido por Problema do *Cocktail Party*, introduzido em (CHERRY; TAYLOR, 1954), que se refere à habilidade do ser humano em reconhecer e selecionar uma fonte de áudio em um ambiente contendo diversos tipos de ruídos, independentes, ou não uns dos outros (HAYKIN; CHEN, 2005). No entanto, caso uma das fontes sonoras presentes no cenário acústico seja caracterizada por um sinal de fala, frequentemente a atenção deste ser humano será desviada, tornando a tarefa de reconhecimento e seleção de uma fonte sonora específica mais complexa (LUND, 2002).

Em Cherry e Taylor (1954), esse fenômeno foi explorado experimentalmente. Os participantes, usando *headphones*, foram submetidos à leitura de textos e outros estímulos externos, explorando-se o uso de um ou ambos os lados auditivos. O objetivo dessa pesquisa era a identificação de sinais que permitissem que máquinas viessem a realizar o reconhecimento de voz. A partir daí, e mais recentemente, vários métodos têm sido propostos para solucionar o problema da separação de sinais de áudio, segundo (SANTOS, 2015) e (HAYKIN; CHEN, 2005), sendo os mais comuns a Análise Computacional de Cenas Auditivas e a Filtragem Espacial, destacando-se nessas áreas os trabalhos de (MITIANOUDIS, 2004) e (FARRELL; MAMMONE; FLANAGAN, 1992).

As técnicas de separação de sinais podem ser classificadas em função da ausência ou disponibilidade de informação *a priori* sobre as fontes contidas no sinal, respectivamente, a Separação Cega de Fontes (BSS - *Blind Source Separation*) ou Separação Supervisionada de Fontes (ISS - *Informed Source Separation*). Nesse contexto e, por vezes com adaptações para duas versões, destacam-se as técnicas de Fatoração de Matrizes Positivas (PAATERO; TAPPER, 1994), Fatoração de Matrizes Não Negativas (LEE; SEUNG, 1999), Análise de Componentes Principais (JOLLIFFE, 2002) e Análise de Componentes Independentes (MITIANOUDIS, 2004), dentre outras.

Embora essas técnicas não tenham sido originalmente desenvolvidas para o processamento de sinais áudio digital, encontram-se, na literatura, trabalhos que as aplicam para a separação de fontes sonoras. Dentre eles, observa-se mais frequentemente o uso das técnicas NMF, que se limitam apenas aos áudios em formato monoaural e, mais recentemente e introduzida por Fitzgerald, Cranitch e Coyle (2005), a técnica de Fatoração de Tensores

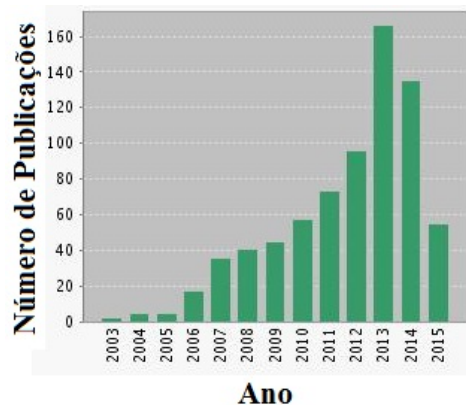
Não Negativos, que possibilita a separação de áudios em formato estéreo ou sua extensão para formatos multicanal.

No contexto musical, diversas aplicações das técnicas de separação de fontes sonoras podem ser encontradas, destacando-se redução de ruído *crosstalk*, restauração e remasterização de obras antigas, aumento na qualidade de reconhecimento de voz, transcrição automática de partituras, dentre outras (FITZGERALD, 2004)(MAHMOUD et al., 2009)(KITAMURA et al., 2013).

1.2 Problema Motivador e Soluções Propostas

O problema de separação cega de sinais de áudio (BASS - *Blind Audio Source Separation*) vem se tornando cada vez mais presente na literatura técnico-científica, conforme indica o levantamento bibliográfico mostrado na Figura 1, tendo sido até alvo de campanhas em prol do desenvolvimento da área como *SiSEC* (SIGNAL SEPARATION EVALUATION CAMPAIGN, 2013).

Figura 1 – Evolução, ao longo do tempo, das publicações relacionadas à BASS



Fonte: *Web of Science* (Web of Science, 2015)

Técnicas de separação de sinais como a NMF e, mais especificamente a NTF, envolvem complexas operações e ferramentas matemáticas específicas para o processamento dos sinais, onde pôde-se observar, claramente, a ausência de análises suficientemente elucidativas que facilitassem a aplicação dessas técnicas em processos de engenharia.

Em um primeiro momento concluiu-se, portanto, sobre importância de uma abordagem matemática formal sobre o tema, que pudesse contribuir para uma compreensão maior acerca das técnicas utilizadas, tanto do ponto de vista conceitual quanto do ponto de vista matemático e, a partir daí, para o desenvolvimento de novas pesquisas.

Além disso, durante a revisão bibliográfica, não foi encontrada nenhuma pesquisa nacional relativa à separação de sinais acústicos utilizando a técnica NTF, adequadas aos formatos estereofônicos ou multicanais. Devido ao fato de músicas em formato multicanal apresentar redundância de informação relativa aos instrumentos contidos na mistura,

simulando uma distribuição espacial dos instrumentos, uma comparação de desempenho das técnicas NMF e NTF pode indicar se a redundância de informação é um fator que contribui ou não para o incremento da qualidade da separação.

Optou-se, portanto, por investigar o uso da NTF na separação de fontes sonoras, levantando-se as suas particularidades em relação à NMF e explicando, todos os passos envolvidos na geração dos seus respectivos algoritmos.

Um outro desafio foi definido em relação às aplicações que deveriam envolver misturas de elementos percussivos (bumbo, caixa e prato de condução) e elementos harmônicos (baixo e guitarra), sinais cujas características espectrais poderiam implicar em um maior grau de dificuldade no processo de separação.

1.3 Objetivos e Metas

Esta pesquisa visa à análise conceitual e ao desenvolvimento matemático das técnicas NMF e NTF associadas às divergências de Kullback-Leibler e Itakura-Saito, bem como à sua aplicação para fins da separação de sinais de áudio gravados que incluem fontes percussivas e harmônicas, especificamente, bumbo, caixa, prato de condução, guitarra e baixo. Destacam-se como objetivos do trabalho:

- 1 Elaborar sínteses sobre os fundamentos conceituais e matemáticos necessários à compreensão e aplicação dos algoritmos referentes às técnicas NMF e NTF;
- 2 Selecionar as principais divergências utilizadas no contexto da separação de fontes sonoras e elaborar os desenvolvimentos matemáticos necessários à associação das técnicas de separação e divergências;
- 3 Implementar os algoritmos referente às técnicas NMF e NTF em ambiente computacional, associadas às Divergências de Kullback-Leibler e Itakura-Saito, para aplicação aos sinais de áudios gravados;
- 4 Adaptar as métricas baseadas na Relação Sinal-Ruído para a avaliação de desempenho das técnicas empregadas;
- 5 Comparar os métodos aplicados e avaliar as diversas combinações resultantes para a separação dos elementos percussivos e harmônicos;

Com base nos objetivos do trabalho, foram definidas as seguintes metas:

- 1 Desenvolvimento de *scripts* abertos, em ambiente MATLAB®, para a execução dos algoritmos referente às técnicas NMF e NTF;
- 2 Desenvolvimento de *scripts* abertos, em ambiente MATLAB®, para a associação das Divergências de Kullback-Leibler e Itakura-Saito às técnicas NMF e NTF;

- 3 Criação de um conjunto de sinais de testes compostos por elementos percussivos e harmônicos;
- 4 Modificação nas métricas clássicas, com base na relação SNR, incorporando métodos de janelamento e sobreposição de sinais;
- 5 Medição de desempenho das técnicas implementadas utilizando as métricas clássicas e janeladas. Comparação e validação dos resultados;
- 6 Proposição de uma versão supervisionada para os algoritmos com a finalidade de estipular novas formas de obtenção de informação de treino;
- 7 Elaboração e submissão de artigos em periódicos/congressos na área.

1.4 Justificativa

A utilização das técnicas NMF e NTF é cada vez mais frequente na literatura. Por se tratar de técnicas multidisciplinares, suas aplicações podem ser encontradas em áreas diversas, tais como a mineração de dados (KROMER; PLATOS; SNÄ;SEL, 2010), análises de texto (BERRY et al., 2006), detecção de intrusos em redes de computadores (SNASEL et al., 2008), dentre outras .

No contexto do áudio digital, suas aplicações mais comuns abrangem áreas que vão desde filtragem de ruídos até criação musical. A seguir são destacadas pesquisas que utilizam essas técnicas e suas respectivas áreas de aplicação:

- 1 Filtragem de ruídos em sinais de fala (JODER et al., 2012);
- 2 Separação de fontes sonoras percussivas (SANTOS, 2015);
- 3 Transcrição automática de partitura (FITZGERALD, 2004) (PHON-AMNUAISUK, 2010);
- 4 Classificação de instrumentos (BENETOS; KOTTI; KOTROPOULOS, 2006);
- 5 Supressão de ruídos mecânicos em câmeras de vídeo (JEON et al., 2013).

Além da crescente utilização na área de áudio digital, a escolha das técnicas de fatoração de matrizes e tensores não negativos para a separação dos sinais musicais, nesta pesquisa, foi feita devido a algumas características intrínsecas dessas técnicas, destacadas, a seguir:

- 1 Facilidade de parametrização;
- 2 Flexibilidade quanto à associação de diferentes divergências;
- 3 Facilidade de implementação computacional;

- 4 Menor tempo de processamento em relação aos outros métodos disponíveis, como o PCA, por exemplo.

1.5 Escopo da Pesquisa

A aplicação das técnicas NMF e NTF às misturas de sons digitais, descrita nesta pesquisa, foi realizada em ambiente MATLAB® R2012b, onde foram desenvolvidos os *scripts* responsáveis pela implementação das respectivas técnicas NMF e NTF, em suas versões não supervisionada e semi-supervisionada. Utilizam-se para testes, sinais de áudio gravados compostos por guitarra, baixo, prato de condução, bumbo e caixa, de forma a comparar os resultados obtidos para um dos casos analisados.

Cada uma das versões das técnicas NMF e NTF foram associados a dois tipos de divergências, Kullback-Leibler e Itakura-Saito. Dessa forma, pôde-se comparar os resultados, buscando-se os melhores índices de desempenho quanto à separação dos sinais de teste, já conhecidos *a priori*.

A avaliação de desempenhos das técnicas foram realizadas de forma objetiva, usando-se métricas baseadas na Relação Sinal-Ruído. Além disso, a partir de uma modificação dessas métricas, proposta nesta pesquisa, buscou-se uma avaliação mais precisa da qualidade de separação dos sinais, incluindo a evolução temporal dos critérios aplicados. Essa abordagem permitiu também identificar os intervalos de melhor desempenho dos critérios avaliados, viabilizando uma nova forma de obtenção de informação *a priori* para fins de treinamento e supervisão.

1.6 Contribuições da Pesquisa

Como principal contribuição desta pesquisa destacam-se as análises conceitual e matemática para as técnicas NMF e NTF, voltadas às suas aplicações no contexto de separação de áudio digital. Além disso, destaca-se a continuidade da pesquisa realizada por Santos (2015), agregando, principalmente estudos acerca da extensão multicanal e testes em misturas contendo elementos harmônicos e percussivos.

A partir desta pesquisa, espera-se ter contribuído também em outros sentidos, dentre eles:

- 1 Análise conceitual e desenvolvimento matemático das técnicas envolvidas, incluindo-se a dedução das regras de atualização;
- 2 Desenvolvimento de *scripts* em ambiente MATLAB® para as técnicas NMF e NTF, ambas associadas às divergências Kullback-Leibler e Itakura-Saito;
- 3 Proposição de nova métrica janelada, mais adequada à análise de desempenho em situações onde a presença de fortes variações nos sinais impeçam que as métricas clássicas, baseadas em SNR, não possam atender aos requisitos usuais de medição;

- 4 Proposição de novo método para a obtenção de informação *a priori* voltada às versões supervisionadas;
- 5 Proposição de uma versão NMF semi-supervisionada recursiva, utilizando-se informação *a priori*;
- 6 Publicação do artigo "Aplicação da Técnica de Fatoração de Matrizes Não-Negativas à Separação de Fontes Sonoras em Misturas contendo Elementos Harmônicos e Percussivos" ao EATI 2015 - 6º Encontro Anual de TI, Universidade Federal de Santa Maria - Rio Grande do Sul - Brasil (FONSECA et al., 2015);
- 7 Submissão do artigo "*Analysis and Application of Non-negative Matrix and Tensor Factorization Techniques for the Sound Signal Separation in Mixtures Containing Harmonic and Percussive Elements*" no periódico da AES.

1.7 Organização do Texto

O Capítulo 2 visa à apresentação dos fundamentos teóricos necessários para o entendimento da pesquisa realizada. Assim, são apresentadas as ferramentas matemáticas envolvidas no processo de geração das regras de atualização bem como no algoritmo, como um todo. É introduzido, também, o conceito de tensor, levando-se em conta o desenvolvimento da NTF.

Trata também das características do áudio, em especial sua representação no formato estéreo. Um levantamento dos trabalhos correlatos e uma breve discussão sobre as particularidades dos mesmos também são apresentados.

No Capítulo 3 são demonstradas as deduções das regras de atualização das técnicas NMF e NTF, considerando as Divergências de Kullback-Leibler e Itakura-Saito, com base nos fundamentos matemáticos descritos no Capítulo 2.

O Capítulo 4 inicia-se com a descrição da metodologia de testes a ser utilizada na pesquisa. Descreve-se, dentre outros, a diferença entre os sinais de áudio utilizados para teste e a parametrização dos algoritmos. É descrita a modificação proposta na métrica e a motivação da mesma, bem como a sua implicação em uma nova forma de se obter sinais de treinamento para as técnicas NMF e NTF. São apresentados, também, as comparações entre as diversas combinações de técnica e divergência.

Por fim, o Capítulo 5 apresenta as considerações finais e as propostas de continuidade desta pesquisa.

2 FUNDAMENTOS TEÓRICOS E TRABALHOS RELACIONADOS

Este capítulo apresenta os conceitos e fundamentos matemáticos necessários ao desenvolvimento e aplicação das técnicas NMF e NTF à separação de sinais no contexto do áudio digital. Os aspectos abordados foram selecionados a partir da revisão bibliográfica realizada, utilizando-se as bases de dados que compõem a *Web of Science*, especialmente o *IEEEExplore*, disponíveis no Portal de Periódicos da CAPES e os estudos desenvolvidos posteriormente. Os trabalhos correlatos e o estado da arte sobre o tema, resultantes desse levantamento bibliográfico, são apresentados ao final do capítulo.

2.1 Fundamentos de Áudio

Os sistemas de áudio destinam-se ao processamento de sons, em geral. Os sons possuem uma natureza dual, podendo ser considerados como um distúrbio em algum meio físico, como o ar, bem como uma percepção psicofísica resultante de estímulos no córtex acústico do cérebro (BALLOU, 2005).

Do ponto de vista da Física, os sons constituem um tipo de energia cinética, descrita como energia acústica. A energia acústica é, basicamente, a variação de pressão em um meio físico, comumente produzida pela vibração de diafragmas, cordas vocais, cordas de instrumentos ou algum outro material sólido (DAVIS et al., 1989) (BALLOU, 2005).

A conversão dos sons em sinais elétricos é realizada pelo uso de transdutores, mais comumente, por meio de microfones que fornecem variações de tensão equivalentes ao perfil de variação da pressão provocada pela propagação do som no ar. A unidade de medida da pressão sonora associada à variação de pressão, decorrente de uma onda sonora, é o *dB SPL* (*dB* - decibel, *SPL* - *Sound Pressure Level*), diretamente determinado pela amplitude das variações de tensão do sinal elétrico correspondente (DAVIS et al., 1989).

Da mesma forma que o sistema visual humano permite perceber a profundidade de campo em uma cena observada, o sistema auditivo permite localizar as fontes sonoras a partir da diferença entre os sinais recebidos. Músicas em formato estéreo utilizam essa propriedade para a realização do *panning*¹. Em resumo, a diferença entre o som monoaural e estéreo é que, enquanto no primeiro tem-se apenas uma representação do áudio, no segundo, utilizam-se duas representações para possibilitar a localização das fontes sonoras (FARNELL, 2010).

¹A expressão *pan* é a abreviação para *panorama*, no sentido de percepção ampla, em todas as direções. A ação de *panning* refere-se à distribuição espacial das fontes sonoras, presentes em determinado áudio estéreo, ao longo de uma imagem ou campo do panorama. Tipicamente, adotam-se imagens com abertura de 90° ou seja, o *panning* realiza a localização das fontes variando-se a posição das mesmas, para a esquerda ou para a direita, considerando um arco descrito por um ângulo variável de 0 a 90°. A relação graus/ganho ou de *panning* varia de acordo com a regra (linear, da raiz quadrada ou do cosseno) e o *software* utilizado. (FARNELL, 2010).

2.2 Tensores

Sejam $\xi = (\mathbf{v}_1, \dots, \mathbf{v}_n)$ e $\mathcal{F} = (\mathbf{w}_1, \dots, \mathbf{w}_m)$, respectivamente, as bases ordenadas dos espaços vetoriais reais $\mathbf{V}$ e $\mathbf{W}$. Seja $\mathbf{T}$ o espaço vetorial real cujos elementos são as combinações lineares formais, $\sum_{i,j} a_{ij} t_{ij}$ dos símbolos t_{ij} , onde a_{ij} e $t_{ij} \in \mathbb{R}$, $1 \leq i \leq n$ e $1 \leq j \leq m$.

Tomando os vetores resultantes das combinações lineares dadas por $\mathbf{v} = \sum_{i=1}^n x_i \mathbf{v}_i$ e $\mathbf{w} = \sum_{j=1}^m y_j \mathbf{w}_j$, $x_i \in \mathbb{R}$, $y_j \in \mathbb{R}$, pode-se definir o produto tensorial de vetores $\mathbf{v} \otimes \mathbf{w} = \sum_{i,j} x_i y_j t_{ij}$, em particular $\mathbf{v}_i \otimes \mathbf{w}_j = t_{ij}$. Os símbolos a_{ij} indicam os coeficientes do produto tensorial. Verifica-se que $(\mathbf{v}, \mathbf{w}) \mapsto \mathbf{v} \otimes \mathbf{w}$ é bilinear, ou seja, linear nos dois argumentos e que $\{\mathbf{v}_i \otimes \mathbf{w}_j; 1 \leq i \leq n, 1 \leq j \leq m\}$ é base de $\mathbf{T}$. Diz-se, então, que o espaço vetorial $\mathbf{T}$ é o produto tensorial de $\mathbf{V}$ por $\mathbf{W}$, ou seja, $\mathbf{T} = \mathbf{V} \otimes \mathbf{W}$, cuja dimensão é dada por $\dim(\mathbf{V} \otimes \mathbf{W}) = \dim(\mathbf{V}) \times \dim(\mathbf{W}) = n \times m$. Os elementos de $\mathbf{T} = \mathbf{V} \otimes \mathbf{W}$ são chamados tensores de ordem 2.

Sejam $\mathcal{G} = (\mathbf{u}_1, \dots, \mathbf{u}_p)$, $\xi = (\mathbf{v}_1, \dots, \mathbf{v}_n)$ e $\mathcal{F} = (\mathbf{w}_1, \dots, \mathbf{w}_m)$, respectivamente, as bases ordenadas dos espaços vetoriais reais $\mathbf{U}$, $\mathbf{V}$ e $\mathbf{W}$. A aplicação $\mathbf{u} \otimes (\mathbf{v} \otimes \mathbf{w}) \in \mathbf{U} \otimes (\mathbf{V} \otimes \mathbf{W}) \mapsto \mathbf{u} \otimes (\mathbf{v} \otimes \mathbf{w}) \in \mathbf{U} \otimes (\mathbf{V} \otimes \mathbf{W})$ é um isomorfismo, sendo possível, então, a omissão dos parêntesis e a escrita da forma $\mathbf{U} \otimes \mathbf{V} \otimes \mathbf{W}$. Os elementos de $\mathbf{U} \otimes \mathbf{V} \otimes \mathbf{W}$ são chamados tensores de ordem 3.

O espaço dual do espaço vetorial $\mathbf{V}$ é definido como $\mathbf{V}^* = \{\mathbf{f} : \mathbf{V} \rightarrow \mathbb{R} \text{ linear}\}$. Em outras palavras, o espaço dual $\mathbf{V}^*$ é o espaço vetorial das formas lineares sobre o espaço vetorial $\mathbf{V}$. Se $\xi = (\mathbf{v}_1, \dots, \mathbf{v}_n)$ é a base de $\mathbf{V}$, então $\xi^* = (\mathbf{f}_1, \dots, \mathbf{f}_n)$ é a base de $\mathbf{V}^*$ ou, a base dual de ξ , onde:

$$\mathbf{f}_i(\mathbf{v}_j) = \delta_{ij} = \begin{cases} 1, & \text{se } i = j \\ 0, & \text{se } i \neq j \end{cases}$$

A aplicação:

$$\mathbf{V}^* \rightarrow \mathcal{L}(\mathbf{V}, \mathbf{W}) = \{\mathbf{A} : \mathbf{V} \rightarrow \mathbf{W} \text{ linear}\}$$

$$\mathbf{f} \otimes \mathbf{w} \mapsto \mathbf{A}, \mathbf{A}(\mathbf{v}) = \mathbf{f}(\mathbf{v})\mathbf{w}$$

é um isomorfismo, tal que as transformações lineares $\mathbf{A} : \mathbf{V} \rightarrow \mathbf{W}$ são tensores mistos (ARFKEN; WEBER, 2008) (RILEY; HOBSON; BENCE, 2006) (BORISENKO; TARAPOV, 1979). Se $\mathbf{V} = \mathbb{R}^N$ e $\mathbf{W} = \mathbb{R}^M$, é comum a utilização da notação $t \in \mathbb{R}^{N \times M}$ para indicar que $t \in \mathbb{R}^N \otimes \mathbb{R}^M$, embora seja um abuso de notação.

Seja um tensor de ordem 3, isto é, $t \in \mathbf{U}^P \otimes \mathbf{V}^N \otimes \mathbf{W}^M$. Se $\mathcal{G} = (\mathbf{u}_1, \dots, \mathbf{u}_p)$, $\xi = (\mathbf{v}_1, \dots, \mathbf{v}_n)$ e $\mathcal{F} = (\mathbf{w}_1, \dots, \mathbf{w}_m)$ são as bases dos espaços vetoriais $\mathbf{U}$, $\mathbf{V}$ e $\mathbf{W}$, respectivamente, então $t = \sum_{i,j,k} a_{ijk} \mathbf{u}_i \otimes \mathbf{v}_j \otimes \mathbf{w}_k$ e t será, então, caracterizado por suas componentes $a_{ijk} \in \mathbb{R}$.

Por exemplo, se $P = N = 2$ e $M = 3$, as componentes de t são dadas na forma:

$$t(, , 1) = \begin{bmatrix} a_{111} & a_{121} \\ a_{211} & a_{221} \end{bmatrix}$$

$$\begin{aligned}
t(, , 2) &= \begin{bmatrix} a_{112} & a_{122} \\ a_{212} & a_{222} \end{bmatrix} \\
t(, , 3) &= \begin{bmatrix} a_{113} & a_{123} \\ a_{213} & a_{223} \end{bmatrix} \\
t(1, ,) &= \begin{bmatrix} a_{111} & a_{112} & a_{113} \\ a_{121} & a_{122} & a_{123} \end{bmatrix} \\
t(2, ,) &= \begin{bmatrix} a_{211} & a_{212} & a_{213} \\ a_{221} & a_{222} & a_{223} \end{bmatrix} \\
t(, 1,) &= \begin{bmatrix} a_{111} & a_{112} & a_{113} \\ a_{211} & a_{212} & a_{213} \end{bmatrix} \\
t(, 2,) &= \begin{bmatrix} a_{121} & a_{122} & a_{123} \\ a_{221} & a_{222} & a_{223} \end{bmatrix}
\end{aligned}$$

2.2.1 Produto Tensorial

Sejam as aplicações lineares $\mathbf{T} : \mathbf{U} \rightarrow \mathbf{W}$, $\mathbf{Y} : \mathbf{V} \rightarrow \mathbf{Z}$. O produto tensorial de $\mathbf{T}$ por $\mathbf{Y}$ é a aplicação linear $\mathbf{T} \otimes \mathbf{Y} : \mathbf{U} \otimes \mathbf{V} \rightarrow \mathbf{W} \otimes \mathbf{Z}$ tal que $(\mathbf{T} \otimes \mathbf{Y})(\mathbf{u} \otimes \mathbf{v}) = \mathbf{T}(\mathbf{u}) \otimes \mathbf{Y}(\mathbf{v})$, $\mathbf{u} \in \mathbf{U}$, $\mathbf{v} \in \mathbf{V}$.

Sejam $\mathbf{T}, \mathbf{Y} : \mathbf{V} \rightarrow \mathbf{V}$ aplicações lineares, $\xi = (\mathbf{v}_1, \dots, \mathbf{v}_n)$ a base de $\mathbf{V}$ e $\mathbf{A} = (a_{ij}) = [\mathbf{T}]_\xi^\xi$, $\mathbf{B} = (b_{ij}) = [\mathbf{Y}]_\xi^\xi$ as matrizes de $\mathbf{T}$ e $\mathbf{Y}$, na base ξ . A matriz de $\mathbf{T} \otimes \mathbf{Y}$ na base $\xi \otimes \xi = (\mathbf{v}_i \otimes \mathbf{v}_j; 1 \leq i, j \leq n)$ de $\mathbf{V} \otimes \mathbf{V}$, é denominada produto de Kronecker ou produto tensorial das matrizes $\mathbf{A}$ e $\mathbf{B}$, sendo indicada como $\mathbf{A} \otimes \mathbf{B}$. Tem-se:

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \dots & a_{1n}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \dots & a_{2n}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{i1}\mathbf{B} & a_{i2}\mathbf{B} & \dots & a_{in}\mathbf{B} \end{bmatrix} \quad (1)$$

Se $\mathbf{A}$ e $\mathbf{B}$ são invertíveis, então o produto tensorial $\mathbf{A} \otimes \mathbf{B}$ também será invertível e dado por $(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$. De maneira mais geral, as matrizes $\mathbf{A}$ e $\mathbf{B}$ não precisam ser quadradas e nem apresentar as mesmas dimensões. Quando os argumentos do produto de Kronecker são vetores, $\mathbf{u} \otimes \mathbf{v}$, $\mathbf{u} \in \mathbb{R}^N$ e $\mathbf{v} \in \mathbb{R}^M$, também chamado de produto exterior (KOLDA; BADER, 2009), tem-se:

$$\mathbf{u} \otimes \mathbf{v} = \mathbf{u}\mathbf{v}^T = \begin{bmatrix} u_1v_1 & u_1v_2 & \dots & u_1v_m \\ u_2v_1 & u_2v_2 & \dots & u_2v_m \\ \vdots & \vdots & \ddots & \vdots \\ u_nv_1 & u_nv_2 & \dots & u_nv_m \end{bmatrix} \quad (2)$$

2.2.2 Contração de Tensores

Seja $\mathbf{T}_s^r$ o espaço dos tensores mistos do tipo (r, s) sobre o espaço vetorial $\mathbf{V}$, isto é:

$$\mathbf{T}_s^r = V \otimes \overset{r}{\underbrace{\dots}} \otimes V \otimes V^* \otimes \overset{s}{\underbrace{\dots}} \otimes V^*$$

Sejam $i, j \in \mathbb{Z}$, $1 \leq i \leq r$ e $1 \leq j \leq s$. A contração do índice i (superior) com o índice j (inferior) é a aplicação linear:

$$C_j^i : \mathbf{T}_s^r \rightarrow \mathbf{T}_{s-1}^{r-1}$$

tal que $C_j^i(\mathbf{v}_1 \otimes \dots \otimes \mathbf{v}_r \otimes \mathbf{f}_1 \otimes \dots \otimes \mathbf{f}_s) = \mathbf{f}_j(\mathbf{v}_i)(\mathbf{v}_1 \otimes \dots \hat{\mathbf{v}}_i \otimes \dots \otimes \mathbf{v}_r \otimes \mathbf{f}_1 \otimes \dots \otimes \hat{\mathbf{f}}_j \otimes \dots \otimes \mathbf{f}_s)$, onde o acento circunflexo, $\hat{\cdot}$, indica o vetor a ser omitido. Por exemplo, se $a_{m_1 \dots m_s}^{k_1 \dots k_r}$ são as coordenadas de $t \in T_s^r$ na base ξ , então as coordenadas de $C_j^i \in T_{s-1}^{r-1}$ na base ξ são os números $b_{m_1 \dots \hat{m}_j \dots m_s}^{k_1 \dots \hat{k}_i \dots k_r}$ dados por:

$$b_{m_1 \dots \hat{m}_j \dots m_s}^{k_1 \dots \hat{k}_i \dots k_r} = \sum_{q=1}^N a_{m_1 \dots \hat{m}_j \dots m_s}^{k_1 \dots \hat{k}_i \dots q \dots \hat{k}_{i+1} \dots k_r}$$

2.2.3 Produto Contraído

Considerando dois tensores descritos conforme as expressões dadas em (3):

$$\begin{aligned} \mathcal{X} &\in \mathbb{R}^{I_1 \times \dots \times I_N \times K_1 \times \dots \times K_M} \\ \mathcal{Y} &\in \mathbb{R}^{I_1 \times \dots \times I_N \times J_1 \times \dots \times J_P} \end{aligned} \quad (3)$$

onde $I_{(\cdot)}$, $J_{(\cdot)}$ e $K_{(\cdot)} \in \mathbb{Z}_+^*$, o produto tensorial contraído $\langle \mathcal{X}, \mathcal{Y} \rangle_{1:N, 1:N}$ é definido como o tensor $\mathcal{Z} \in \mathbb{R}^{K_1 \times \dots \times K_M \times J_1 \times \dots \times J_P}$, que é calculado segundo a expressão (5) (KOLDA; BADER, 2009)(FITZGERALD; CRANITCH; COYLE, 2005)(FEVOTTE; OZEROV, 2010):

$$\mathcal{Z} = \langle \mathcal{X}, \mathcal{Y} \rangle_{1:N, 1:N} \quad (4)$$

$$\langle \mathcal{X}, \mathcal{Y} \rangle_{1:N, 1:N} = \sum_{i_1=1}^{I_1} \dots \sum_{i_N=1}^{I_N} x_{i_1, \dots, i_N, k_1, \dots, k_M} y_{i_1, \dots, i_N, j_1, \dots, j_P} \quad (5)$$

Alguns aspectos devem ser ressaltados em relação à Equação (4). O primeiro é que os índices da notação indicam as dimensões a serem consideradas na operação. Dessa forma, reescrevem-se os tensores $\mathcal{X}$ e $\mathcal{Y}$, na forma de tensores de ordem N , como $\mathcal{X} \in \mathbb{R}^{I_1 \times \dots \times I_N}$ e $\mathcal{Y} \in \mathbb{R}^{I_1 \times \dots \times I_N}$, evidenciando-se que ambos os tensores resultantes apresentam a mesma ordem. Ainda, que o produto tensorial contraído é uma generalização do produto interno de tensores. Assim, $\mathcal{Z}$ é um tensor de ordem $M + P$, em $\mathbb{R}^{K_1 \times \dots \times K_M \times J_1 \times \dots \times J_P}$, com dimensão definida pelas dimensões não consideradas no produto e cujas entradas são os produtos internos das matrizes resultantes do reordenamento.

Para fins de ilustração, vamos considerar dois tensores de ordem 3 especificados como $\mathcal{X} \in \mathbb{R}^{2 \times 3 \times 4}$ e $\mathcal{Y} \in \mathbb{R}^{3 \times 4 \times 3}$, tal que:

$$\mathcal{X}(:, :, 1) = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}$$

$$\mathcal{X}(:, :, 2) = \begin{bmatrix} 7 & 9 & 11 \\ 8 & 10 & 12 \end{bmatrix}$$

$$\mathcal{X}(:, :, 3) = \begin{bmatrix} 13 & 15 & 17 \\ 14 & 16 & 18 \end{bmatrix}$$

$$\mathcal{X}(:, :, 4) = \begin{bmatrix} 19 & 21 & 23 \\ 20 & 22 & 24 \end{bmatrix}$$

$$\mathcal{Y}(:, :, 1) = \begin{bmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & 9 & 12 \end{bmatrix}$$

$$\mathcal{Y}(:, :, 2) = \begin{bmatrix} 13 & 16 & 19 & 22 \\ 14 & 17 & 20 & 23 \\ 15 & 18 & 21 & 24 \end{bmatrix}$$

$$\mathcal{Y}(:, :, 3) = \begin{bmatrix} 25 & 28 & 31 & 34 \\ 26 & 29 & 32 & 35 \\ 27 & 30 & 33 & 36 \end{bmatrix}$$

Reescrevendo $\mathcal{X}$ na forma de tensores em $\mathbb{R}^{3 \times 4}$, tem-se:

$$\mathcal{X}(1, :, :) = \begin{bmatrix} 1 & 7 & 13 & 19 \\ 3 & 9 & 15 & 21 \\ 5 & 11 & 17 & 23 \end{bmatrix}$$

$$\mathcal{X}(2, :, :) = \begin{bmatrix} 2 & 8 & 14 & 20 \\ 4 & 10 & 16 & 22 \\ 6 & 12 & 18 & 24 \end{bmatrix}$$

Definindo $\mathcal{Z} = \langle \mathcal{X}, \mathcal{Y} \rangle_{2:3,1:2}$, tem-se:

$$\mathcal{Z}(1, 1) = \langle \mathcal{X}(1, :, :), \mathcal{Y}(:, :, 1) \rangle$$

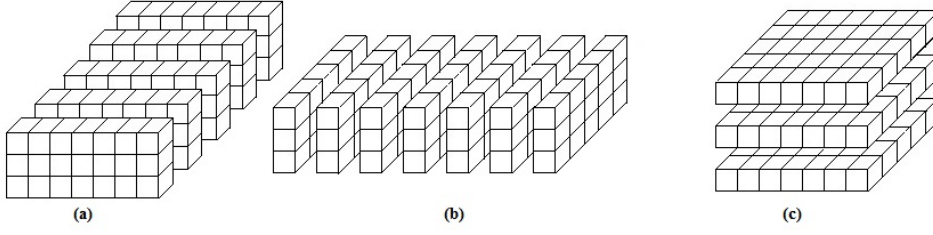
$$\mathcal{Z}(1, 2) = \langle \mathcal{X}(1, :, :), \mathcal{Y}(:, :, 2) \rangle$$

$$\mathcal{Z}(1, 3) = \langle \mathcal{X}(1, :, :), \mathcal{Y}(:, :, 3) \rangle$$

$$\mathcal{Z}(2, 1) = \langle \mathcal{X}(2, :, :), \mathcal{Y}(:, :, 1) \rangle$$

$$\mathcal{Z}(2, 2) = \langle \mathcal{X}(2, :, :), \mathcal{Y}(:, :, 2) \rangle$$

Figura 2 – Indexação de Tensores: (a) - (:,:,k); (b) - (:,k,:); (c) - (k,:,:)



Fonte: Elaborado pelo autor

$$\mathcal{Z}(2,3) = \langle \mathcal{X}(2, :, :), \mathcal{Y}(:, :, 3) \rangle$$

Assim:

$$\mathcal{Z} = \begin{bmatrix} 1222 & 2950 & 4678 \\ 1300 & 3172 & 5044 \end{bmatrix}$$

A Figura 2 ilustra o processo de indexação dos tensores, considerado no exemplo anterior.

2.3 Transformada de Fourier de Curta Duração

Usualmente, em processamento digital de sinais, utilizam-se ferramentas matemáticas capazes de representar os sistemas e/ou sinais discretos no tempo em outros domínios que sejam mais adequados ao tipo de aplicação, tais como o domínio da frequência, domínio z ou domínio tempo-frequência, obtendo-se uma melhor descrição do problema, simplificação dos modelos envolvidos, redução do número de operações aritméticas, dentre outras (OPPENHEIM; SCHAFER, 2010) (SANTOS, 2015).

A Transformada de Fourier de Tempo Discreto (DTFT - *Discrete Time Fourier Transform*) estabelece a relação entre as representações dos sistemas/sinais discretos nos domínios do tempo-frequência. Especificamente nas aplicações de áudio, trata-se de uma transformação linear cujo domínio é o dos números reais e a imagem é o conjunto dos números complexos ou seja, $DTFT(x[n]) : \mathbb{R} \rightarrow \mathbb{C}$, $n \in \mathbb{Z}$, definida como (OPPENHEIM; SCHAFER, 2010):

$$X(e^{j\omega}) = \sum_{n=-\infty}^{\infty} x[n]e^{-j\omega n} \quad (6)$$

onde $X(e^{j\omega})$ é a Transformada de Fourier de Tempo Discreto, $x[n]$ é uma sequência numérica composta pelas amplitudes do sinal no domínio do tempo discreto e ω ($rad/amostra$) é a frequência digital. A representação $X(e^{j\omega})$ é uma função complexa, periódica com período 2π , que poderá ser expressa em termos de suas parcelas real e imaginária, segundo a equação (7) (OPPENHEIM; SCHAFER, 2010)(MITRA, 2011).

$$X(e^{j\omega}) = X_{re}(e^{j\omega}) + jX_{im}(e^{j\omega}) \quad (7)$$

ou, na forma polar,

$$X(e^{j\omega}) = |X(e^{j\omega})|e^{j\theta(\omega)} \quad (8)$$

sendo $|X(e^{j\omega})|$ e $\theta(\omega) = \arg\{X(e^{j\omega})\}$, respectivamente, os espectros de amplitude e de fase do sinal.

Para sinais periódicos, representados por $\tilde{x}[n]$, com período $T = N$ amostras, faz-se uso da chamada Transformada de Fourier Discreta (DFT - *Discrete Fourier Transform*), $\tilde{X}[k]$, tal que (OPPENHEIM; SCHAFER, 2010)(MITRA, 2011):

$$\tilde{X}[k] = X \left(e^{j \left(\frac{2k\pi}{N} \right)} \right) = \sum_{n=0}^{N-1} \tilde{x}[n] e^{-j \left(\frac{2k\pi}{N} \right) n} \quad (9)$$

Um algoritmo computacional eficiente para o cálculo da DFT é a Transformada Rápida de Fourier (FFT - *Fast Fourier Transform*), proposto em 1965 por Cooley e Tukey (COOLEY; TUKEY, 1965).

Para sinais tipicamente não estacionários, como no caso dos sinais de áudio, a representação no domínio da frequência através da DFT não é adequada (OPPENHEIM; SCHAFER, 2010)(GOSWAMI; CHAN, 1999). Mais comumente, utiliza-se a Transformada de Fourier de Tempo Curto (STFT - *Short-Time Fourier Transform*), na qual o sinal de interesse é dividido em janelas de menor comprimento e, em seguida, é calculada a FFT para cada janela. A STFT é expressa matematicamente como (MITRA, 2011)(DINIZ; SILVA; NETTO, 2014)(OPPENHEIM; SCHAFER, 2010):

$$X[n, \lambda] = \sum_{m=-\infty}^{\infty} x[n+m]w[m]e^{-j\lambda m} \quad (10)$$

onde $w[m]$ é a função de janelamento e λ é uma frequência variável. Dessa forma, a STFT representa um sinal unidimensional $x[n]$ na forma de uma função bidimensional tempo-frequência (OPPENHEIM; SCHAFER, 2010). Em geral, um espectrograma é representado por uma matriz $X \in \mathbb{C}^{I \times J}$ onde a dimensão N equivale à excursão da frequência em relação a uma janela específica, enquanto M equivale ao número de janelas presentes no sinal. Os tipos de janelas mais comuns são as de Hanning e Hamming (TYGEL, 2009).

2.4 Divergências Utilizadas

No contexto matemático, uma função de distância ou métrica permite avaliar a proximidade entre dois elementos de um determinado conjunto, vetor ou tensor (VITANYI, 2011).

Definição 1: Uma métrica ou distância em um conjunto F é uma função $d : F \times F \rightarrow \mathbb{R}_+$ se, para todos P, Q e $U \in F$, tem-se que:

Axioma 1 : $d(\mathbf{P}, \mathbf{Q}) \geq 0$ e $d(\mathbf{P}, \mathbf{Q}) = 0$ se $\mathbf{P} = \mathbf{Q}$.

Axioma 2 : $d(\mathbf{P}, \mathbf{Q}) = d(\mathbf{Q}, \mathbf{P})$

Axioma 3 : $d(\mathbf{P}, \mathbf{Q}) \leq d(\mathbf{P}, \mathbf{U}) + d(\mathbf{U}, \mathbf{Q})$

sendo o Axioma 3 conhecido como Condição da Desigualdade Triangular.

Quando se deseja encontrar a distância entre dois vetores de probabilidade n -dimensionais, ação requerida nas aplicações das técnicas NMF e NTF e em diversas outras técnicas aplicáveis aos problemas de otimização, são utilizadas, em geral, as chamadas famílias de divergência (CICHOCKI et al., 2009).

Dentre essas, os exemplos mais conhecidos são as famílias de divergências baseadas nas Funções de Csiszár e na Divergência de Bregman, das quais são derivados os subconjuntos denominados divergências Alfa e Beta, respectivamente (CICHOCKI et al., 2009).

A Divergência de Bregman constitui a base matemática das métricas mais usualmente aplicadas, como a Distância Euclidiana, a Divergência de Kullback-Leibler e a Divergência de Itakura-Saito, sendo as duas últimas representantes da subclasse beta. A exemplo das métricas, a Divergência de Bregman é também definida positiva, propriedade fundamental no contexto das técnicas NMF e NTF. A seguir, a Divergência de Bregman é apresentada mais detalhadamente, com destaque para os aspectos de interesse neste trabalho (BANERJEE et al., 2005).

Definição 2: Seja $\phi : S \mapsto \mathbb{R}$ uma função estritamente convexa, definida em um conjunto convexo $S \subseteq \mathbb{R}^d$ tal que ϕ é diferenciável em $\mathbf{relint}(S) \neq \emptyset$. A Divergência de Bregman, no domínio $d_\phi : S \times \mathbf{relint}(S) \mapsto [0, \infty)$, é definida como:

$$d_\phi(p, q) = \phi(p) - \phi(q) - \langle p - q, \nabla \phi(q) \rangle \quad (11)$$

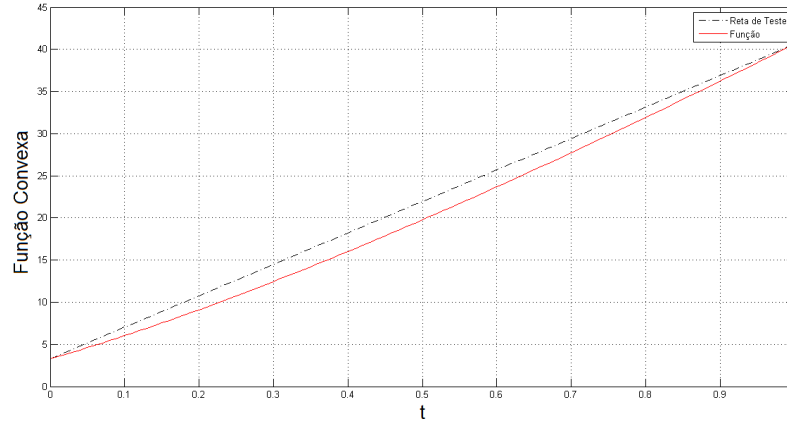
onde $\mathbf{relint}(S)$ refere-se ao conjunto relativo interior de S que contém todos os elementos de S com exceção daqueles pertencentes aos limites de S , $\langle \cdot, \cdot \rangle$ indica o produto interno dos argumentos e $\nabla(f)$ indica o gradiente da função f .

A escolha de uma função ϕ que obedeça à restrição de convexidade, quando aplicada à divergência de Bregman, gera as medidas de divergência. Por exemplo, definindo-se as funções $\phi(p)$ e $\phi(q)$ como os produtos internos $\langle p, p \rangle$ e $\langle q, q \rangle$, respectivamente, tem-se como consequência a Distância Euclidiana. É fácil notar que a Distância Euclidiana é uma métrica, uma vez que atende aos três axiomas.

2.4.1 Funções Convexas

Situações nas quais se deseja maximizar ou minimizar uma determinada função exigem que a mesma apresente a propriedade da concavidade ou convexidade, respectivamente, de tal forma que seja possível a obtenção de um máximo ou mínimo local para a referida função. Para que se entenda uma função convexa, relacionada aos problemas de minimização, deve-se primeiro entender o conceito de convexidade de uma maneira geral.

Figura 3 – Convexidade da função $\phi(p) = p \ln p$



Fonte: Elaborado pelo autor

Definição 3: Seja V um espaço vetorial e seja $X \subset V$. Diz-se que X é convexo se (GALLAGER, 1968):

$$[t\alpha + (1 - t)\beta] \in V \quad \forall t \in [0, 1] \quad (12)$$

onde $\alpha, \beta \in X$.

Avaliando-se a definição dada pela expressão (12) em $V = \mathbb{R}$, α e β serão unidimensionais e a variação de t descreverá uma linha reta conectando α a β . Ou seja, qualquer linha reta traçada entre dois vetores pertencentes a V não possuirá pontos externos a V .

Define-se, então, uma função convexa tal que sua imagem seja convexa. Em outras palavras, o segmento de reta que liga os valores da função em dois vetores específicos, pertencentes a esse espaço vetorial, é sempre maior ou igual aos valores da função para os vetores intermediários. Considerando uma desigualdade estrita, diz-se que a função f é estritamente convexa, se é expressa matematicamente conforme a inequação (13):

$$tf(\alpha) + (1 - t)f(\beta) > f(t\alpha + (1 - t)\beta) \quad (13)$$

No contexto das técnicas NMF e NTF, funções convexas são usadas para a obtenção das divergências utilizadas pelo método de otimização. Os vetores considerados para a avaliação de proximidade são vetores de probabilidade, ou seja, todas as entradas são não negativas e possuem soma unitária.

2.4.2 Divergência de Kullback-Leibler

Para a função convexa $\phi(p) = p \ln p$, a entropia negativa definida por (SCHRODINGER, 1944), ou entropia de Shannon (SHANNON, 1948), pode ser calculada como (CICHOCKI et al., 2009):

$$H_\phi(p) = - \sum_{j=1}^d p_j \ln p_j \quad (14)$$

onde p é um vetor de probabilidade e d , a dimensão desse vetor. A Figura 3 ilustra a convexidade da função $\phi(p) = p \ln p$.

A partir da equação (11) e da função convexa $\phi(p) = p \ln p$, a Divergência de Kullback-Leibler é calculada como:

$$\begin{aligned}
 d_\phi(p, q) &= \sum_{j=1}^d p_j \ln p_j - \sum_{j=1}^d q_j \ln q_j - \langle p - q, \sum_{j=1}^d 1 + \ln q_j \rangle \\
 d_\phi(p, q) &= \sum_{j=1}^d (p_j \ln p_j - q_j \ln q_j - (p_j - q_j)(1 + \ln q_j)) \\
 d_\phi(p, q) &= \sum_{j=1}^d (p_j \ln p_j - q_j \ln q_j - p_j - \\
 &\quad p_j \ln q_j + q_j + q_j \ln q_j) \\
 \therefore D_{KL}(\mathbf{p} \parallel \mathbf{q}) &= \mathbf{p} \ln \frac{\mathbf{p}}{\mathbf{q}} - \mathbf{p} + \mathbf{q}
 \end{aligned} \tag{15}$$

Obtendo-se, assim, a formulação para a divergência generalizada de Kullback-Leibler (BANERJEE et al., 2005)(CICHOCKI et al., 2009).

As divergências, em geral, não podem ser consideradas métricas uma vez que não atendem aos axiomas da simetria e da desigualdade triangular. Porém, uma vez que atende ao Axioma 1, é possível sua utilização como medida de distância nas aplicações de separação de áudio.

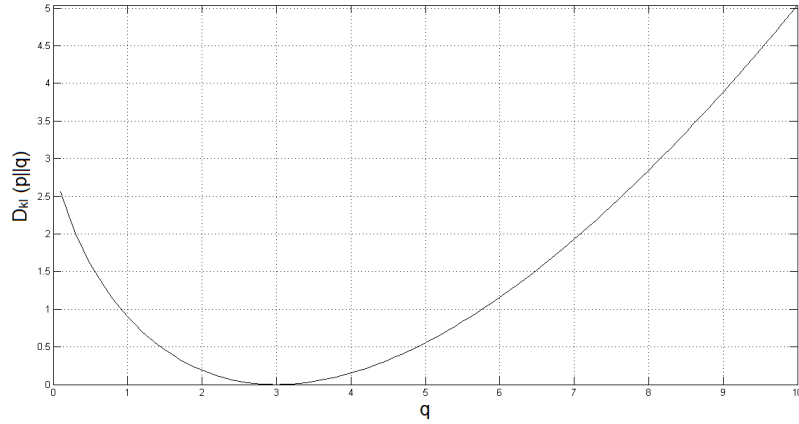
Análise do Axioma 1:

Para provar que a Divergência de Kullback-Leibler apresenta valores sempre positivos, usa-se a desigualdade $e^{x-1} \geq x$, tomando-se $x = \frac{q}{p}$. Assim:

$$\begin{aligned}
 e^{\frac{q}{p}-1} &\geq \frac{q}{p} \\
 \frac{q}{p} - 1 &\geq \ln \frac{q}{p} \\
 \frac{q}{p} - \ln \frac{q}{p} - 1 &\geq 0 \\
 q - p \ln q + p \ln p - p &\geq 0 \\
 p \ln \frac{p}{q} - p + q &\geq 0
 \end{aligned}$$

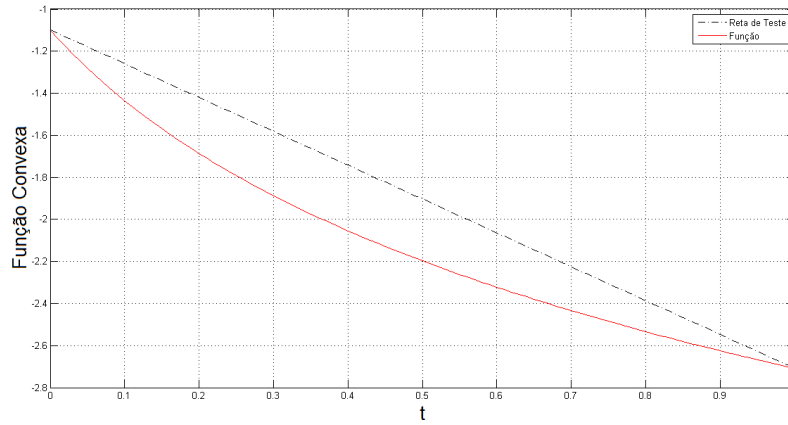
A Figura 4 mostra a variação da Divergência de Kullback-Leibler para $p = [0, 10]$ em relação ao valor constante $q = 3$. Vale ressaltar que tais valores foram escolhidos visando apenas uma melhor visualização. É possível confirmar que o Axioma 1 foi completamente atendido, ou seja, a Divergência de Kullback-Leibler sempre retorna valores positivos, sendo 0 apenas quando $p = q = 3$.

Figura 4 – Exemplo: Variação da Divergência de Kullback-Leibler



Fonte: Elaborado pelo autor

Figura 5 – Convexidade da função $\phi = -\ln p$



Fonte: Elaborado pelo autor

2.4.3 Divergência de Itakura-Saito

Para a função convexa $\phi = -\ln p$, ilustrada na Figura 5, obtém-se a entropia de Burg, dada por (CICHOCKI et al., 2009):

$$H_{\phi}(p) = \sum_i \ln p_i$$

Similarmente, com base na equação (11) e na função convexa $\phi = -\ln p$, a Divergência de Itakura-Saito é calculada como:

$$d_{\phi}(p, q) = -\sum_{j=1}^d \ln p_j + \sum_{j=1}^d \ln q_j + \langle p - q, \sum_{j=1}^d \frac{1}{q_j} \rangle$$

$$d_{\phi}(p, q) = \sum_{j=1}^d \left(-\ln p_j + \ln q_j + (p_j - q_j) \left(\frac{1}{q_j} \right) \right)$$

$$d_\phi(p, q) = \sum_{j=1}^d \left(-\ln p_j + \ln q_j + \frac{p_j}{q_j} - 1 \right)$$

$$\therefore D_{IS}(\mathbf{p} \parallel \mathbf{q}) = \frac{\mathbf{p}}{\mathbf{q}} - \ln \frac{\mathbf{p}}{\mathbf{q}} - 1 \quad (16)$$

Obtendo-se, portanto, a formulação geral para a Divergência de Itakura-Saito. Assim como para a Divergência de Kullback-Leibler, a Divergência de Itakura-Saito não atende aos axiomas de simetria e da desigualdade triangular, não sendo consideradas métricas conforme a definição matemática. Em ambos os casos, porém, não há restrições quanto à aplicação dessas divergências para mensurar a aproximação dos resultados estimados pelas técnicas NMF e NTF, uma vez que, além de serem definidas positivas (característica fundamental para a utilização nas técnicas NMF e NTF), as divergências são usadas apenas na fase de geração das regras de atualização.

Análise do Axioma 1:

Para provar que a Divergência de Itakura-Saito apresenta valores sempre positivos usa-se a desigualdade $e^{x-1} \geq x$ para $x = \frac{p}{q}$. Assim:

$$e^{\frac{p}{q}-1} \geq \frac{p}{q}$$

$$\frac{p}{q} - 1 \geq \ln \frac{p}{q}$$

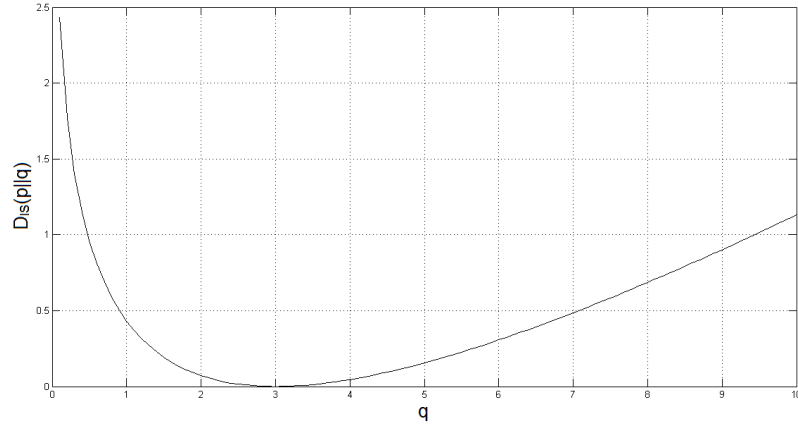
$$\frac{p}{q} - \ln \frac{p}{q} - 1 \geq 0$$

Interpretações alternativas para a Divergência de Itakura-Saito ocorrem quando é possível aproximar o sinal de interesse a um processo gaussiano estacionário. Definem-se os vetores de probabilidade, consideradas no cálculo do erro, como sendo os espectros de potência ($F(e^{j\theta})$) dos sinais envolvidos. Utilizando como função convexa $\phi(F) = -\ln F(e^{j\theta})$, obtém-se a Divergência de Itakura-Saito (BANERJEE et al., 2005)(PALUS, 1997)(COVER; THOMAS, 1991).

A Figura 6 mostra a variação da Divergência de Itakura-Saito para $p = [0, 10]$ em relação ao valor constante $q = 3$. É possível confirmar que o Axioma 1 foi completamente atendido, ou seja, a Divergência de Itakura-Saito sempre retorna valores positivos, sendo 0 apenas quando $p = q = 3$.

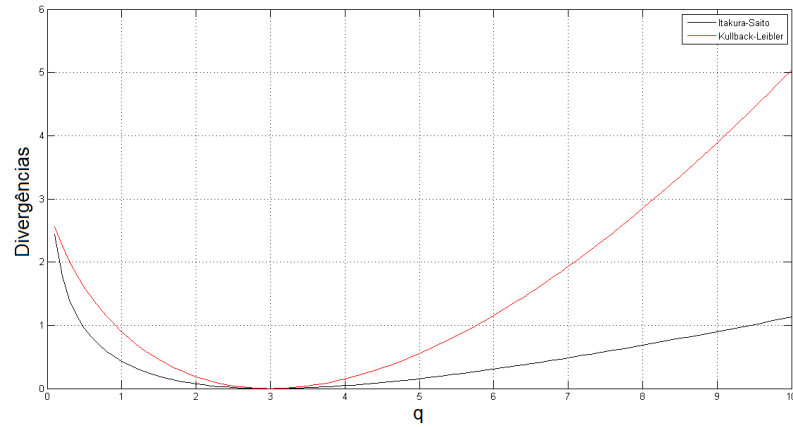
A Figura 7 mostra os gráficos das variações das divergências de Kullback-Leibler e de Itakura-Saito, tomando-se $p = 3$ e $q = [0, 10]$. Novamente, os valores são escolhidos apenas com a finalidade de se obter uma melhor visualização. Pode-se observar que a Divergência de Kullback-Leibler apresenta uma inclinação mais acentuada em relação à Divergência Itakura-Saito, notadamente mais suave. Como os parâmetros considerados em ambas são os mesmos, Itakura-Saito e Kullback-Leibler convergem para 0 quando $p = q = 3$.

Figura 6 – Exemplo: Variação da Divergência de Itakura-Saito



Fonte: Elaborado pelo autor

Figura 7 – Comparação entre as Variações das Divergência de Kullback-Leibler e Itakura-Saito



Fonte: Elaborado pelo autor

A seguir, tem-se um resumo das divergências beta de Bregman, D_B^β (CICHOCKI et al., 2009):

$$D_B^\beta(\mathbf{p} \parallel \mathbf{q}) = \begin{cases} \frac{1}{\beta(\beta+1)} \sum_i (p_i^{\beta+1} - q_i^{\beta+1} - (\beta+1)q_i^\beta(p_i - q_i)), & \beta > -1 \\ \sum_i \left(p_i \ln \frac{p_i}{q_i} - p_i + q_i \right), & \beta = 0 \\ \sum_i \left(\frac{p_i}{q_i} - \ln \frac{p_i}{q_i} - 1 \right), & \beta = -1 \end{cases} \quad (17)$$

2.5 Métodos de otimização

A partir da diferença ou erro calculado entre os sinais real e estimado, deve-se aplicar algum método para a minimização desse erro.

Historicamente, os modelos CANDECOMP (*Canonical Decomposition*) utilizam,

como algoritmo de otimização, o método dos mínimos quadrados alternados (ALS - *Alternating Least Squares*), no qual fixam-se todos, exceto um, termos do modelo matemático utilizado e atualiza o que não foi fixado. Dessa forma, o método de otimização ALS é aproximado ao método clássico dos mínimos quadrados (LS - *Least Squares*). O processo deve ser repetido para os outros termos, até que o erro total atinja o valor mínimo (KOLDA; BADER, 2009). Diversos métodos de otimização podem ser associadas à NMF e à NTF, sendo o mais usual nas aplicações de áudio o Método do Gradiente Descendente (*Gradient Descent ou Steepest Descent*).

O Método do Gradiente Descendente é um dos mais antigos e conhecidos métodos de minimização de funções de várias variáveis e consiste em um algoritmo de otimização de primeira ordem (LUENBERGER; YE, 2008).

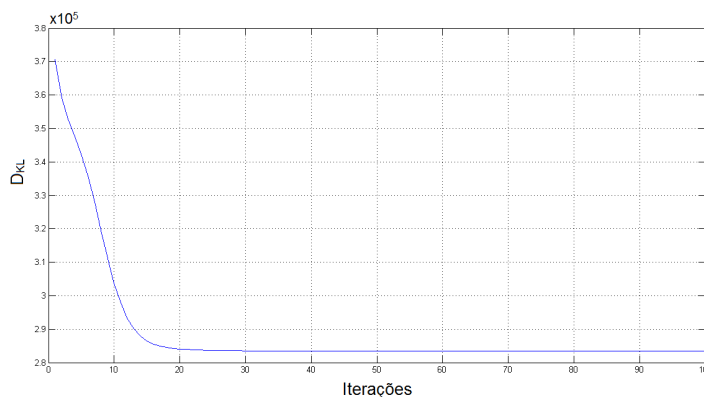
Definição 4: Considerando f uma função continuamente derivável, define-se o Método do Gradiente Descendente como um algoritmo iterativo tal que (LUENBERGER; YE, 2008):

$$x_{k+1} = x_k - \alpha \times \nabla f(x_k) \quad (18)$$

onde $\alpha \in \mathbb{R}$, x representa a variável que se deseja otimizar, k e $k + 1$ referem-se, respectivamente, aos instantes de tempo atual e futuro e ∇f é o gradiente da função f .

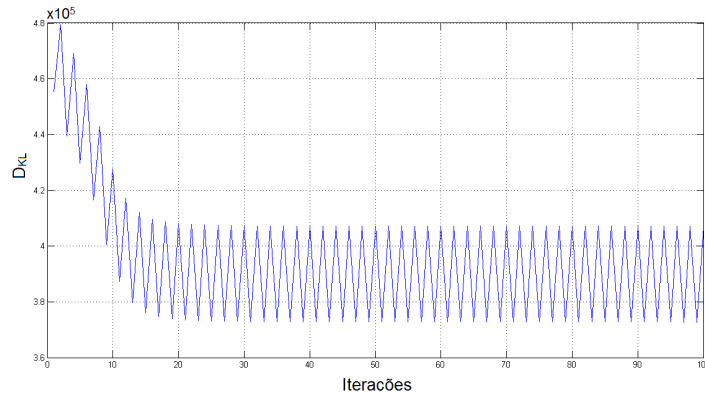
Em relação às técnicas NMF e NTF, uma propriedade das divergências beta que vale ressaltar, em relação à equação (18), refere-se à convexidade. As divergências beta são convexas em seu primeiro argumento, mas não necessariamente no segundo (BANERJEE et al., 2005), espectrograma observado e estimado, respectivamente. Por consequência, o problema de otimização da NMF e NTF não é convexo como um todo, mas será convexo para o espectrograma observado e/ou espectrograma estimado. Dessa forma, é possível obter um valor ótimo para o espectrograma estimado (mínimo local), tal que o erro convirja para valores muito pequenos ou nulos, transformando a NMF e a NTF em problemas de otimização convexas. Essa característica é perdida quando se atualizam todos os termos de uma vez (CICHOCKI; CRUCES; AMARI, 2011).

Figura 8 – Convergência da NMF para a Divergência de Kullback-Leibler



Fonte: Elaborado pelo autor

Figura 9 – Oscilação da NMF usando Divergência de Kullback-Leibler



Fonte: Elaborado pelo autor

As Figuras 8 e 9 exemplificam, respectivamente, a convergência da técnica NMF usando a divergência de Kullback-Leibler e a não convergência, para o caso no qual as matrizes W e H , que serão introduzidas no Capítulo 3, são atualizadas ao mesmo tempo.

2.6 Trabalhos Relacionados

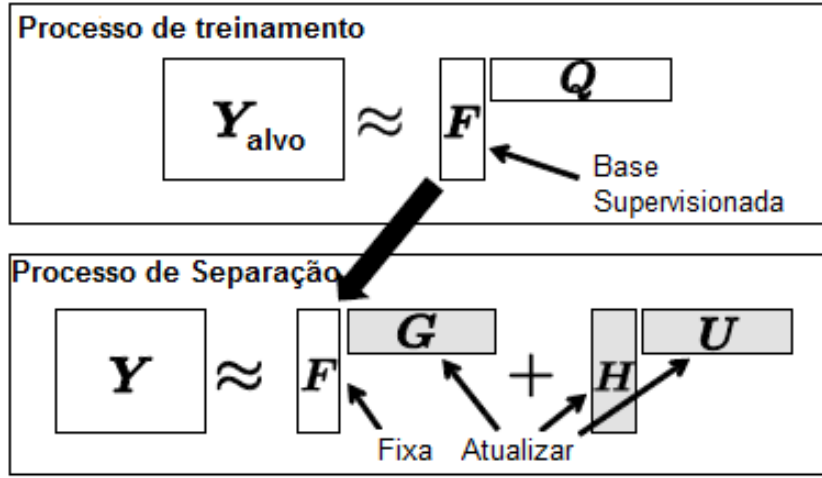
Em Kitamura et al. (2013) é apresentada uma breve revisão sobre a técnica NMF básica, utilizando as divergências beta como função de custo. Após destacar as dificuldades da NMF básica quanto à clusterização dos padrões espectrais de um instrumento específico, os autores investigam uma versão supervisionada da NMF (SNMF - *Supervised Non-Negative Matrix Factorization*), que utiliza amostras do sinal de interesse para o treinamento *a priori*. O processo de treinamento consiste em obter a matriz de base (que contem o padrão espectral de uma fonte) do sinal de treino através da NMF básica. Posteriormente, no processo de separação, é composto um espectrograma observado que consiste do sinal desejado e fontes de interferências utilizando a base supervisionada, obtida na etapa anterior, e outras.

A Figura 10 mostra um diagrama simplificado do processo, com destaque para a base F , construída na fase de treinamento e aplicada na fase de separação.

Os autores indicam alguns aspectos que degradam a qualidade da separação, como por exemplo, o compartilhamento da base F durante o treinamento e a separação. Propõem uma nova técnica para a separação de sinais de áudio, denominada NMF Supervisionada Penalizada (PSNMF - *Penalized Supervised Nonnegative Matrix Factorization*) que, além de melhorar a qualidade da separação, ainda responde bem quando utilizada para a separação de instrumentos com característica de sustentação harmônica.

Na versão PSNMF trabalha-se com termos de penalidade nas funções de custo, buscando operar a ortogonalidade e maximização das funções de divergência, de forma a se obter fontes independentes na composição das bases. A PSNMF obteve melhores resultados se comparada às versões NMF e SNMF básicas, chegando a apresentar uma

Figura 10 – NMF Supervisionada



Fonte: Adaptado de (KITAMURA et al., 2013)

diferença de cerca de 8dB, em média, na relação sinal distorção (SDR) para o caso de separação em uma mistura com 2 sinais. Para esses experimentos foi utilizado sintetizador MIDI (*Musical Digital Instrument Interface*), em uma mistura monocal.

Sawada et al. (2012) trata do problema de separação de voz e violão em uma música gravada comercialmente, enquanto a grande maioria dos trabalhos encontrados na literatura utiliza sintetizador MIDI, o que configura um ambiente mais controlado. Além disso, trata de sistemas de áudio multicanais, no caso específico, sinais do tipo estéreo descritos por:

$$\hat{\mathbf{X}}_{ij} = \sum_{k=1}^K \mathbf{H}_{ik} t_{ik} v_{kj}$$

onde H é uma matriz Hermitiana definida positiva e $\|H\| = 1$ que descreve a propriedade espacial do sinal de áudio analisado.

Além disso, propõe uma nova função de custo dada pela equação 19. Dessa forma são gerados algoritmos de atualização mais eficientes e que expandem a NMF para o uso em problemas multicanais sem, portanto, utilizar a NTF (*Nonnegative Tensor Factorization*).

$$\mathcal{L}(\mathbf{T}, \mathbf{V}, \mathbf{H}, \mathbf{R}, \mathbf{U}) = \sum_{i,j} \left[\frac{\text{tr}(X_{ij} R_{ijk} H_{ik}^{-1} R_{ijk})}{t_{ik} v_{kj}} + \log \det U_{ij} + \frac{\det \hat{X}_{ij} - \det U_{ij}}{\det U_{ij}} \right] \quad (19)$$

onde tr indica o traço da matriz, R_{ijk} e U_{ij} são variáveis auxiliares introduzidas tal que a matriz Hermitiana definida positiva, H , seja satisfeita e tal que $\sum_k R_{ijk} = I_M$, onde I é a matriz identidade.

Os resultados apresentam tempo de convergência relativamente baixo e uma relação sinal distorção (SDR) compatível com os métodos monocanais, o que mostra a aplicabilidade do método proposto.

Santos (2015) utiliza a NMF para a separação dos elementos de uma bateria composta por Bumbo, chimbau e caixa, utilizando como funções de custo a Distância Euclidiana e a Divergência de Kullback-Leibler. Apresenta prova matemática das regras

de atualização utilizadas, bem como um estudo sobre o histórico da separação de sinais de áudio. Foi testada uma forma supervisionada do método, onde os sinais de treino são descritos a seguir:

- 1 Chimal: Foram realizados diversos ataques variando-se intensidade, velocidade e localização na peça da bateria, considerando a mesma aberta, semi-aberta ou fechada;
- 2 Bumbo: Variam-se a intensidade e a velocidade dos ataques;
- 3 Caixa: Variam-se intensidade e a velocidade dos ataques, bem como a localização na qual a peça pode ser tocada.

O conjunto de testes utilizados é composto por 10 áudios distintos, sendo 5 compostos por elementos reais e 5 por elementos sintetizados. Para a validação foram utilizadas as métricas objetivas SDR (Relação Fonte-Distorção), SIR (Relação Fonte-Interferência) e SAR (Relação Fonte-Artefato), baseadas na SNR, bem como a métrica subjetiva PEAQ (*Perceptual Evaluation of Quality Audio* - Avaliação Perceptual da Qualidade do Áudio). Esta última apresenta como resultado uma nota de saída chamada ODG (*Objective Difference Grade* - Grau de Diferença Objetiva) que varia de 0 a -4, sendo 0 referente a uma diferença imperceptível e -4 muito incômoda.

Os resultados foram considerados satisfatórios e promissores. Para o método não supervisionado as divergências utilizadas não apresentaram diferenças significativas, enquanto que para o método supervisionado a Distância Euclidiana se mostrou mais adequada.

Fitzgerald, Cranitch e Coyle (2005) trabalha com a separação de áudio estéreo fazendo uso da NTF, sendo o primeiro trabalho a usar essa abordagem para o áudio. É dada uma breve descrição matemática da técnica e apresentadas as regras de atualização para a divergência de Kullback-Leibler, bem como uma visão geral do algoritmo. A mistura considerada é composta pelos instrumentos: baixo, saxofone e violino. Os resultados apresentados são todos baseados em formas de onda, o que dá apenas um indicativo visual da qualidade da separação. Diferente da NMF, a NTF separa o áudio em função de seus componentes (acordes), em vez de suas fontes. Dessa forma faz-se necessário o uso de formas de agrupar essas notas nas fontes, efetivamente. É utilizado pelo autor, então, o algoritmo knn (*k-nearest neighbour*) para essa finalidade.

Fevotte e Ozerov (2010) também trata do problema de separação de áudio multicanal, usando tanto a divergência de Kullback-Leibler, quanto a de Itakura-Saito. Apresenta uma parte teórica bem consistente com deduções e interpretações do ponto de vista estatístico das técnicas utilizadas. Além disso, propõe um método diferente de agrupamento das notas, realizado pelo próprio algoritmo referente à técnica NTF.

Guan et al. (2012) tratou da NMF de maneira mais teórica focando bastante em algoritmos de atualização (chamados, pelo autor, como solucionadores) que apresentam melhores resultados. Um amplo estudo foi desenvolvido, comparando-se os diversos métodos encontrados na literatura quanto à convergência e velocidade. Propõe um método (NeNMF

– Optimal Gradient Method for Nonnegative Matrix Factorization) que busca uma solução ótima em relação ao compromisso entre velocidade e convergência. Embora seus resultados sejam todos baseados em testes na área de clusterização de documentos, as técnicas são bem descritas matematicamente, tornando possível o teste das mesmas para sistemas de áudio. Provas matemáticas da técnica são apresentadas e também disponibilizados alguns códigos em MatLab no site.

Lee e Seung (1999) foi o primeiro a tratar do tema como Fatoração de Matrizes Não Negativas. Trabalho semelhante já havia sido proposto anteriormente por Pateiro, sob o nome de Fatoração de Matrizes Positivas; porém, não foi um trabalho muito conhecido na época. Lee e Seung propuseram o primeiro algoritmo da técnica NMF que realmente apresentava resultados satisfatórios e, com uma velocidade relativamente alta. O artigo, porém, é bastante teórico e não apresenta testes das técnicas propostas. Entretanto, a descrição matemática, provas de convergência da técnica, dentre outros temas importantes para a implementação, são tratados.

Wang e Zhang (2013), Lin (2007) e Berry et al. (2006) trazem estudos detalhados da NMF e suas variações. Tratam de problemas quanto à convergência e formas de melhoria dos mesmos, descrição detalhada das técnicas e seus algoritmos, comparações entre as funções de custos e os métodos de otimização, áreas de aplicação, dentre outros. São, portanto, artigos de caráter mais teórico; entretanto, por se tratarem de Surveys na área, são fundamentais para o entendimento da técnica.

2.7 Conclusão

Este capítulo apresentou os fundamentos matemáticos necessários à análise das técnicas NMF e NTF. Foram apresentados, de maneira sucinta, os principais aspectos dessas técnicas, bem como os requisitos matemáticos para a obtenção das mesmas. Foram consideradas como medidas de erro, as divergências de Kullback-Leibler e Itakura-Saito.

Foi realizada também uma revisão bibliográfica sobre o tema com destaque para o estado da arte, além de um levantamento dos pontos-chave, sob o ponto de vista do áudio, necessários ao entendimento e desenvolvimento da pesquisa.

3 DEDUÇÃO MATEMÁTICA DAS REGRAS DE ATUALIZAÇÃO DOS ALGORITMOS NMF E NTF

Nesta seção serão apresentadas as deduções das regras de atualização referentes às técnicas NMF e NTF para as divergências de Kullback-Leibler e Itakura-Saito, com base nas aplicações para a separação de sinais de áudio digital.

3.1 Non-Negative Matrix Factorization (NMF) e Non-Negative Tensor Factorization (NTF)

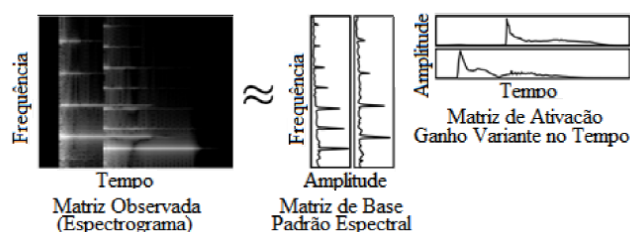
Conforme apresenta a eq. (20), a técnica NMF busca aproximar uma matriz não-negativa $\mathbf{V} \in \mathbb{R}_+^{N \times M}$ à matriz $\mathbf{A}$, que é o produto de duas outras matrizes não-negativas $\mathbf{W} \in \mathbb{R}_+^{N \times R}$ e $\mathbf{H} \in \mathbb{R}_+^{R \times M}$, de tal forma que $R < \min(N, M)$ (CICHOCKI et al., 2009).

$$\mathbf{V} \approx \mathbf{A} = \mathbf{WH} \quad (20)$$

Nas aplicações voltadas à separação de fontes sonoras em sinais de áudio digital, a matriz de entrada $\mathbf{V} \in \mathbb{R}_+^{N \times M}$ é composta pelas N faixas de frequência consideradas no espectrograma do sinal e os M intervalos de tempo onde ocorreram as observações. $\mathbf{W} \in \mathbb{R}_+^{N \times R}$ é chamada matriz de base, cujas colunas contêm o padrão espectral de cada fonte e $\mathbf{H} \in \mathbb{R}_+^{R \times M}$, denominada matriz de ativação, fornece os coeficientes da combinação linear para a composição de cada fonte, conforme ilustrado na Figura 11 (KITAMURA et al., 2013). O posto R refere-se ao número de fontes contidas na mistura, ou seja, o número de instrumentos a serem separados no áudio original (LIN, 2007).

Assim, em aplicações cujos sinais de entrada são dados em formato estéreo ou multicanais, faz-se necessário o uso de técnicas capazes de lidar com estruturas tensoriais. No contexto da separação de fontes sonoras, considera-se um tensor como uma extensão das matrizes, isto é, um arranjo multidimensional de dados (KOLDA; BADER, 2009). Em outras palavras, assim como uma matriz pode ser entendida como um vetor bidimensional, um tensor pode ser entendido como uma matriz n-dimensional. Nos últimos anos, a decomposição de tensores tem sido crescente alvo de pesquisa no contexto de separação cega de fontes, fazendo com que o estudo dos tensores assuma, cada vez mais, uma

Figura 11 – Modelo de decomposição da NMF no contexto de áudio - Kitamura, 2013



Fonte: Adaptado de Kitamura, 2013

importância maior para a comunidade de processamento de sinais (COMON, 2014).

A Decomposição Canônica Poliádica (CPD - canonical polyadic decomposition), também conhecida por PARAFAC (Parallel Factor Analysis) (HARSHMAN, 1970) ou CANDECOMP (Cannonical Decomposition) (CARROLL; CHANG, 1970), foi introduzida em (HITCHCOCK, 1927) . A equação (21) mostra a expressão geral da CPD.

$$\mathcal{T} \approx \hat{\mathcal{T}} = \sum_{k=1}^K T_{:,k}^1 \otimes \cdots \otimes T_{:,k}^Z \quad (21)$$

sendo $\mathcal{T}$ e $\hat{\mathcal{T}}$ tensores de ordem Z , onde $\hat{\mathcal{T}}$ é dado pela soma de K tensores de ordem Z , obtido através do produto exterior de Z tensores de ordem 1 ($T_{:,k}^1 \cdots T_{:,k}^Z$).

Tendo sido inicialmente aplicada às áreas de psicometria e quimiometria, mais recentemente, o método vem sendo utilizado em aplicações nas áreas de processamento digital de sinais, visão computacional, mineração de dados, dentre outros. Porém, da forma como foi inicialmente proposta, a Decomposição Canônica Poliádica tem, por base, a Decomposição em Valores Singulares (SVD - Singular Value Decomposition) (KOLDA; BADER, 2009) e, devido à um espalhamento da energia ao longo de todo o espectro, não se mostra adequada para o processamento de sinais de áudio digital, especificamente, à separação de fontes sonoras (FITZGERALD; CRANITCH; COYLE, 2005). Nesse caso, a técnica de fatoração de tensores não-negativos (NTF - Non-Negative Tensor Factorization) mostra-se mais atraente na medida em que utiliza algoritmos de otimização como, por exemplo, o método gradiente descendente, para a aproximação entre os dados de entrada e o modelo a ser estimado.

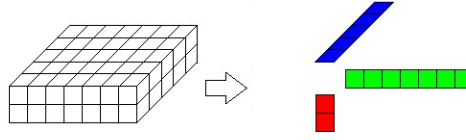
Visando, especificamente, à separação de sinais de áudio digital, a técnica de fatoração de tensores não-negativos é expressa pela equação (22), onde um tensor de entrada $\mathcal{X}$ é aproximado a um tensor estimado $\hat{\mathcal{X}}$ da forma,

$$\mathcal{X} \approx \hat{\mathcal{X}} = \sum_{k=1}^K G_{:,k} \otimes A_{:,k} \otimes S_{:,k} \quad (22)$$

sendo G , A e S tensores de ordem 2 referentes aos ganhos de mixagem, padrão espectral e ativação no tempo, respectivamente. O subscrito " $:k$ " indica uma coluna específica dos tensores, o que permite escrevê-los como tensores de ordem 1. Assim, o espectrograma estimado, de ordem 3, é a soma de K tensores também de ordem 3, obtido através do produto exterior dos tensores equivalentes G , A e S , reescritos como tensores de ordem 1.

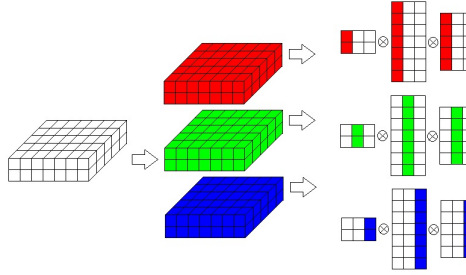
Tipicamente, para aplicações em áudio estéreo, $\mathcal{X}$ e $\hat{\mathcal{X}}$ pertencem a $\mathbb{R}^{2 \times N \times M}$, $G \in \mathbb{R}^{2 \times K}$, $A \in \mathbb{R}^{N \times K}$ e $S \in \mathbb{R}^{M \times K}$, onde $K \in \mathbb{Z}_+^*$ e indica o número de tensores que compõem o tensor estimado. No contexto de separação de áudio, K é equivalente ao número de notas musicais contidas no áudio. Além disso, como um espectrograma é expresso por uma matriz (tensor de ordem 2), a consideração de mais canais de áudio conferem a esse espectrograma uma terceira dimensão, cujo tamanho depende do número de canais do

Figura 12 – Bases do espaço tensorial



Fonte: Elaborado pelo autor

Figura 13 – Decomposição de Tensores



Fonte: Elaborado pelo autor

áudio. Assim, no contexto de separação de sinais de áudio, estéreo ou multicanais, os tensores considerados são de ordem 3.

O espaço tensorial no qual o espectrograma estimado $\widehat{\mathcal{X}}$ está contido é gerado pelo produto tensorial de 3 tensores de ordem 1, pertencentes a $\mathbb{R}^2$, $\mathbb{R}^N$ e $\mathbb{R}^M$, conforme Figura 12, sendo $\widehat{\mathcal{X}}$, então, uma combinação linear de tensores de ordem 1 (CICHOCKI et al., 2009).

O espectrograma estimado depende então de $k \times 3$ tensores de ordem 1, onde k é o número total de notas contidas no sinal de áudio. As colunas dos tensores G , A e S referem-se, respectivamente, ao ganho, ao padrão espectral e ao tempo de ativação de cada nota específica. A Figura 13 exemplifica a decomposição de um tensor $\mathcal{X} \in \mathbb{R}^{2 \times 7 \times 5}$.

Nas aplicações para a separação de fontes sonoras, as técnicas NMF e NTF apresentam grande semelhança conceitual, exceto pelo formato dos seus dados de entrada. Enquanto a NMF utiliza, como dados de entrada, o espectrograma de um áudio monoaural em uma matriz $N \times M$ e o número de fontes R , a NTF recebe os espectrogramas dos canais esquerdo e direito de um áudio estéreo, ou os espectrogramas dos diversos canais para o caso de um formato de áudio multicanal, em forma de um tensor $2 \times N \times M$ e o número de notas presentes na mistura. No caso específico de sinais estéreos, 2 indica o número existente de canais do áudio (FITZGERALD; CRANITCH; COYLE, 2005).

Ambas as técnicas NMF e NTF são implementados segundo uma mesma sequência, descrita a seguir:

1. Definição do modelo de aproximação, sendo a matriz $\mathbf{A}$ ou o tensor $\widehat{\mathcal{X}}$, para a NMF e NTF, respectivamente;
2. Escolha da métrica ou medidas de divergência a serem utilizadas para mensurar a aproximação entre os dados de entrada e os modelos estimados;

3. Escolha do método de otimização para a atualização e convergência do modelo.

Em geral, utilizam-se regras de atualização multiplicativa, conforme as equações (20) ou (22), com o intuito de garantir a não negatividade do algoritmo em todas as iterações, desde que os tensores/matrizas tenham sido devidamente inicializados com valores positivos (FITZGERALD; CRANITCH; COYLE, 2005).

3.2 NMF usando Divergência de Kullback-Leibler

Seja uma matriz $V \in \mathbb{R}_+^{N \times M}$ cujas entradas correspondem ao espectrograma de um áudio sob análise. O método de fatoração de matrizes não negativas busca aproximar V de $A \in \mathbb{R}_+^{N \times M}$, obtida através do produto de $W \in \mathbb{R}_+^{N \times K}$ por $H \in \mathbb{R}_+^{K \times M}$, conforme (23):

$$V \approx A = WH \quad (23)$$

onde W é a matriz contendo o padrão espectral das fontes e H contém os coeficientes das combinações lineares das fontes.

Usando a divergência de Kullback-Leibler entre V e A tem-se:

$$D_{KL}(V \parallel A) = X \ln \frac{X}{A} - X + A$$

expressa em (24) na forma matricial:

$$D_{KL}(V \parallel A) = E = \sum_{n=1}^N \sum_{m=1}^M \left(v_{nm} \ln \frac{v_{nm}}{a_{nm}} - v_{nm} + a_{nm} \right) \quad (24)$$

Aplicando-se o método do gradiente descendente às matrizes W e H tem-se:

$$W \leftarrow W - \alpha \times \nabla f = W - \alpha \frac{\partial E_{nm}}{\partial W_{nd}} \quad (25)$$

$$H \leftarrow H - \alpha \times \nabla f = H - \alpha \frac{\partial E_{nm}}{\partial H_{dm}} \quad (26)$$

onde ∇f indica o gradiente da função de custo ou divergência utilizada.

A derivada parcial $\frac{\partial E_{nm}}{\partial W_{nd}}$ pode ser reescrita da forma:

$$\frac{\partial E_{nm}}{\partial W_{nd}} = \frac{\partial E_{nm}}{\partial A_{nm}} \frac{\partial A_{nm}}{\partial W_{nd}} \quad (27)$$

Desenvolvendo o primeiro termo de (27), tem-se:

$$\begin{aligned} \frac{\partial E_{nm}}{\partial A} &= \frac{\partial E}{\partial a_{nm}} = \frac{\partial}{\partial a_{nm}} \sum_{n=1}^N \sum_{m=1}^M (v_{nm} \ln v_{nm} - \ln a_{nm} - v_{nm} + a_{nm}) \\ \frac{\partial E_{nm}}{\partial a_{nm}} &= \sum_{n=1}^N \sum_{m=1}^M \frac{\partial}{\partial a_{nm}} (v_{nm} \ln v_{nm} - \ln a_{nm} - v_{nm} + a_{nm}) \end{aligned}$$

$$\therefore \frac{\partial E_{nm}}{\partial a_{nm}} = - \left(\frac{v_{nm}}{a_{nm}} - 1 \right) \quad (28)$$

Enquanto para o segundo termo de (27), tem-se:

$$\frac{\partial A_{nm}}{\partial w_{nd}} = \begin{bmatrix} \frac{\partial a_{11}}{\partial w_{nd}} & \frac{\partial a_{12}}{\partial w_{nd}} & \dots & \frac{\partial a_{1m}}{\partial w_{nd}} \\ \frac{\partial a_{21}}{\partial w_{nd}} & \frac{\partial a_{22}}{\partial w_{nd}} & \dots & \frac{\partial a_{2m}}{\partial w_{nd}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial a_{n1}}{\partial w_{nd}} & \frac{\partial a_{n2}}{\partial w_{nd}} & \dots & \frac{\partial a_{nm}}{\partial w_{nd}} \end{bmatrix}$$

Que também pode ser escrito como:

$$\frac{\partial a_{n'm}}{\partial w_{nd}} = \frac{\partial}{\partial w_{nd}} \sum_{k=1}^D w_{n'k} h_{km}$$

Deve-se observar que $\frac{\partial a_{n'm}}{\partial w_{nd}}$ só assume valores não nulos para o caso onde $n = n'$ e $k = d$. Assim:

$$\begin{aligned} \frac{\partial a_{n'm}}{\partial w_{nd}} &= \sum_{d=1}^D \frac{\partial}{\partial w_{nd}} w_{nd} h_{dm} \\ \therefore \frac{\partial a_{n'm}}{\partial w_{nd}} &= h_{dm} \end{aligned} \quad (29)$$

Substituindo (28) e (29) em (27):

$$\begin{aligned} \frac{\partial E_{nm}}{\partial W_{nd}} &= - \sum_{n=1}^N \sum_{m=1}^M \left(\frac{v_{nm}}{a_{nm}} - 1 \right) h_{dm} \\ \frac{\partial E_{nm}}{\partial W_{nd}} &= - \sum_{n=1}^N \sum_{m=1}^M \left(\frac{v_{nm}}{a_{nm}} - 1 \right) \left(\frac{\partial}{\partial w_{nd}} \sum_{k=1}^D w_{n'k} h_{km} \right) \end{aligned}$$

Vale lembrar que valor de $n' \neq n$ zera $\frac{\partial E_{nm}}{\partial W_{nd}}$. Assim:

$$\frac{\partial E_{nm}}{\partial W_{nd}} = - \sum_{m=1}^M \left(\frac{v_{nm}}{a_{nm}} - 1 \right) h_{dm}$$

Que pode ser reescrita como:

$$\frac{\partial E_{nm}}{\partial W_{nd}} = - \left(\frac{V}{A} - U \right) H^T \quad (30)$$

onde U é uma matriz cujos elementos são todos iguais a um e de dimensão $N \times M$.

Substituindo (30) em (25):

$$W \leftarrow W + \alpha \left(\frac{V}{WH} H^T - U H^T \right)$$

e fazendo $\alpha = \frac{W}{UH^T}$ para que seja gerada uma regra de atualização na forma multiplicativa. Tem-se:

$$W \leftarrow W + \frac{W}{UH^T} \odot \left(\frac{V}{WH} H^T - UH^T \right)$$

$$W \leftarrow W \odot \frac{\left(\frac{V}{WH} H^T \right)}{UH^T} \quad (31)$$

Repetindo o processo para H , tem-se:

$$H \leftarrow H \odot \frac{\left(W^T \frac{V}{WH} \right)}{W^T U} \quad (32)$$

onde $\odot$ representa a multiplicação elemento a elemento ou Produto de Hadamard.

Quando não utilizado o símbolo $\odot$ considera-se o produto matricial; além disso, todas as divisões envolvidas são realizadas elemento a elemento. Vale notar que as dimensões das matrizes são todas compatíveis, possibilitando assim o uso das regras geradas.

3.3 NMF usando Divergência de Itakura-Saito

O mesmo processo pode ser repetido para a obtenção das regras de atualização usando a divergência de Itakura-Saito, expressa em (33):

$$D(V \parallel A) = \frac{V}{A} - \ln \frac{V}{A} - 1 \quad (33)$$

Tomando-se o gradiente de (33),

$$\frac{\partial E}{\partial A} = - \left(\frac{V}{A^2} - \frac{U}{A} \right) H^T \quad (34)$$

E substituindo (34) em (25), obtém-se:

$$W \leftarrow W + \alpha \left(\frac{V}{A^2} H^T - \frac{U}{A} H^T \right) \quad (35)$$

Para $\alpha = \frac{W}{\frac{U}{A} H^T}$, a regra de atualização para W pode ser escrita como:

$$W \leftarrow W \odot \frac{\frac{V}{(WH)^2} H^T}{\frac{U}{WH} H^T} \quad (36)$$

Similarmente, repetindo-se o processo para a matriz H , tem-se:

$$H \leftarrow H \odot \frac{W^T \frac{V}{(WH)^2}}{W^T \frac{U}{WH}} \quad (37)$$

3.4 NTF usando Divergência de Kullback-Leibler

Seja $\mathcal{X}$ um tensor de ordem 3 e de dimensões $2 \times N \times M$, composto pelas matrizes referentes aos espectrogramas de cada um dos 2 canais de um áudio digital estéreo. De uma maneira geral, tensores de quaisquer ordens podem ser decompostos em uma combinação linear de tensores de ordem 1, com base na decomposição canônica poliádica (Canonical Polyadic).

Conforme mencionado anteriormente, o método utilizado neste trabalho baseia-se na CPD, também conhecida por PARAFAC (Parallel Factor Analysis) ou CANDECOMP (Canonical Decomposition) onde o tensor $\mathcal{X}$ pode ser expresso conforme (38):

$$\mathcal{X} \approx \widehat{\mathcal{X}} = \sum_{k=1}^K G_{:k} \otimes A_{:k} \otimes S_{:k} \quad (38)$$

onde $G_{:k}$, $A_{:k}$ e $S_{:k}$ são as k -ésimas colunas de G , A e S , respectivamente. O índice k é o número de fatores paralelos do modelo, ou seja, representa o número de tensores que somados geram o tensor $\widehat{\mathcal{X}}$. No caso específico, k representa o número total de notas que compõem o áudio. O símbolo $\otimes$ representa o produto exterior. As dimensões de G , A e S são, respectivamente $2 \times K$, $N \times K$ e $M \times K$.

A divergência de Kullback-Leibler pode ser expressa conforme a (39),

$$D(\mathcal{X} \parallel \widehat{\mathcal{X}}) = \mathcal{X} \ln \frac{\mathcal{X}}{\widehat{\mathcal{X}}} - \mathcal{X} + \widehat{\mathcal{X}} \quad (39)$$

como uma métrica de avaliação do erro entre $\mathcal{X}$ e $\widehat{\mathcal{X}}$. Aplicando-se o método do gradiente descendente em (39), tem-se:

$$G \leftarrow G - \alpha \times \frac{\partial E_{ijl}}{\partial G_{ik}} \quad (40)$$

$$A \leftarrow A - \alpha \times \frac{\partial E_{ijl}}{\partial A_{jk}} \quad (41)$$

$$S \leftarrow S - \alpha \times \frac{\partial E_{ijl}}{\partial S_{lk}} \quad (42)$$

Em (40), tem-se $\frac{\partial E_{ijl}}{\partial G_{ik}}$, expressa como:

$$\frac{\partial E_{ijl}}{\partial G_{ik}} = \frac{\partial}{\partial G} \sum_{l=1}^m \sum_{j=1}^n \sum_{i=1}^2 \left(x_{ijl} \ln x_{ijl} - x_{ijl} \sum_{k=1}^K G_{:k} \otimes A_{:k} \otimes S_{:k} - x_{ijl} + \sum_{k=1}^K G_{:k} \otimes A_{:k} \otimes S_{:k} \right) \quad (43)$$

A derivada parcial $\frac{\partial E_{ijl}}{\partial G_{ik}}$ em (40), pode ser expressa como:

$$\frac{\partial E_{ijl}}{\partial G_{ik}} = \frac{\partial E_{ijl}}{\partial \widehat{\mathcal{X}}_{ijl}} \frac{\partial \widehat{\mathcal{X}}_{ijl}}{\partial G_{ik}} \quad (44)$$

O primeiro termo de (44) pode ser reescrito como:

$$\begin{aligned}\frac{\partial E_{ijl}}{\partial \widehat{\mathcal{X}}_{ijl}} &= \frac{\partial}{\partial \widehat{\mathcal{X}}_{ijl}} (x_{ijl} \ln x_{ijl} - x_{ijl} \ln \widehat{x_{ijl}} - x_{ijl} + \widehat{x_{ijl}}) \\ \frac{\partial E}{\partial \widehat{\mathcal{X}}} &= -\frac{\mathcal{X}}{\widehat{\mathcal{X}}} + \mathcal{U} \\ \therefore \frac{\partial E}{\partial \widehat{\mathcal{X}}} &= -(\mathcal{D} - \mathcal{U})\end{aligned}\tag{45}$$

onde $\mathcal{D}$ e $\mathcal{U}$ são tensores de ordem 3 e dimensão $2 \times N \times M$. $\mathcal{D}$ é o tensor resultante da divisão elemento a elemento de $\mathcal{X}$ e $\widehat{\mathcal{X}}$, e $\mathcal{U}$ é um tensor composto de 1's.

O segundo termo de (44) pode ser reescrito como:

$$\frac{\partial \widehat{\mathcal{X}}_{ijl}}{\partial G_{ik}} = \frac{\partial}{\partial G_{ik}} \sum_{k=1}^K G_{:k} \otimes A_{:K} \otimes S_{:k}$$

Como cada termo desse somatório só será não nulo quando considera-se a k-ésima coluna da matriz G , pode-se reescrever $\frac{\partial \widehat{\mathcal{X}}_{ijl}}{\partial G_{ik}}$ como:

$$\frac{\partial \widehat{\mathcal{X}}_{ijl}}{\partial G_{ik}} = \frac{\partial \widehat{\mathcal{X}}_{ijl}}{\partial G_{:k}} = A_{:k} \otimes S_{:k}$$

Pode-se notar que, para cada valor de k , $\frac{\partial \widehat{\mathcal{X}}_{ijl}}{\partial G_{ik}}$ é um tensor de ordem 2 e dimensão $N \times M$. Assim, pode-se expressar $\frac{\partial \widehat{\mathcal{X}}_{ijl}}{\partial G_{ik}}$ como um tensor de ordem 3 e dimensão $N \times M \times K$ como em (46):

$$\frac{\partial \widehat{\mathcal{X}}_{ijl}}{\partial G_{ik}} = \mathcal{P}(:, :, k) = A_{:k} \otimes S_{:k}\tag{46}$$

Substituindo (45) e (46) em (44):

$$\frac{\partial E}{\partial G} = -(\mathcal{D} - \mathcal{U}) \mathcal{P}\tag{47}$$

Fazendo $\alpha = \frac{\mathcal{G}}{\mathcal{UP}}$ e substituindo (47) em (40):

$$G \leftarrow G \odot \left(\frac{\mathcal{DP}}{\mathcal{UP}} \right)$$

Por fim, deve-se atentar às dimensões das matrizes e tensores envolvidas. Como G é uma matriz de dimensão $2 \times K$, o tensor multiplicado elemento a elemento por G também deve ter dimensão $2 \times K$. Da mesma forma, $\mathcal{DP}$ e $\mathcal{UP}$. Assim:

$$G \leftarrow G \odot \left[\frac{\langle \mathcal{D}, \mathcal{P} \rangle_{2:3,1:2}}{\langle \mathcal{U}, \mathcal{P} \rangle_{2:3,1:2}} \right]\tag{48}$$

Desenvolvendo para (41) e (42), tem-se:

$$A \leftarrow A \odot \left[\frac{\langle \mathcal{D}, \mathcal{Q} \rangle_{1:3,1:2}}{\langle \mathcal{U}, \mathcal{Q} \rangle_{1:3,1:2}} \right]\tag{49}$$

$$S \leftarrow S \odot \left[\frac{\langle \mathcal{D}, \mathcal{R} \rangle_{1:2,1:2}}{\langle \mathcal{U}, \mathcal{R} \rangle_{1:2,1:2}} \right] \quad (50)$$

com:

$$\mathcal{P}(:, :, k) = A_{:k} \otimes S_{:k}$$

$$\mathcal{Q}(:, :, k) = G_{:k} \otimes S_{:k}$$

$$\mathcal{R}(:, :, k) = G_{:k} \otimes A_{:k}$$

onde $\langle \mathcal{D}, \mathcal{P} \rangle_{2:3,1:2}$, $\langle \mathcal{U}, \mathcal{P} \rangle_{2:3,1:2}$, $\langle \mathcal{D}, \mathcal{Q} \rangle_{1:3,1:2}$, $\langle \mathcal{U}, \mathcal{Q} \rangle_{1:3,1:2}$, $\langle \mathcal{D}, \mathcal{R} \rangle_{1:2,1:2}$ e $\langle \mathcal{U}, \mathcal{R} \rangle_{1:2,1:2}$ são produtos tensoriais, definidos na Seção 2.2.3, de dimensões $2 \times K$, $2 \times K$, $N \times K$, $N \times K$, $M \times K$ e $M \times K$, respectivamente.

3.5 NTF usando Divergência de Itakura-Saito

O mesmo desenvolvimento pode ser estendido para a Divergência de Itakura-Saito, para $\alpha = \frac{\mathcal{U}}{\widehat{\mathcal{X}}}\mathcal{P}$, que levará às Eq. (51), (52) e (53) como regras de atualização para os tensores G , A e S , respectivamente:

$$G \leftarrow G \odot \left[\frac{\langle \mathcal{D}, \mathcal{P} \rangle_{2:3,1:2}}{\left\langle \frac{\mathcal{U}}{\widehat{\mathcal{X}}}, \mathcal{P} \right\rangle_{2:3,1:2}} \right] \quad (51)$$

$$A \leftarrow A \odot \left[\frac{\langle \mathcal{D}, \mathcal{Q} \rangle_{1:3,1:2}}{\left\langle \frac{\mathcal{U}}{\widehat{\mathcal{X}}}, \mathcal{Q} \right\rangle_{1:3,1:2}} \right] \quad (52)$$

$$S \leftarrow S \odot \left[\frac{\langle \mathcal{D}, \mathcal{R} \rangle_{1:2,1:2}}{\left\langle \frac{\mathcal{U}}{\widehat{\mathcal{X}}}, \mathcal{R} \right\rangle_{1:2,1:2}} \right] \quad (53)$$

O tensor $\mathcal{D}$ é definido segundo (54):

$$\mathcal{D} = \left(\frac{\mathcal{X}}{\widehat{\mathcal{X}}^2} \right) \quad (54)$$

Os tensores $\mathcal{P}$, $\mathcal{Q}$ e $\mathcal{R}$ são definidos da mesma forma para as divergências de Kullback-Leibler e de Itakura-Saito.

3.6 Conclusão

Este capítulo apresentou, de forma detalhada, as deduções matemáticas das regras de atualização das técnicas NMF e NTF, ressaltando os pontos mais importantes dos mesmos. As regras de atualização foram deduzidas levando em consideração as Divergências utilizadas no presente trabalho, Kullback-Leibler e Itakura-Saito.

4 RESULTADOS COMPUTACIONAIS

Neste capítulo são apresentadas as análises e os resultados obtidos sobre a separação de fontes sonoras em sinais de áudio compostos por sons de guitarra, baixo, prato de condução, bumbo e caixa. A separação é realizada aplicando-se as técnicas NMF e NTF, respectivamente, em sinais monoaurais e estéreos sendo, em ambos os casos, utilizadas as divergências de Kullback-Leibler e de Itakura-Saito. Além disso, como uma contribuição deste trabalho, é apresentado a técnica NMF Semi-Supervisionada Recursivo, cuja informação de treinamento é extraída de parcelas pré-selecionadas dos sinais de áudio a serem separados.

Todo o desenvolvimento é realizado em ambiente MATLAB[®]. São apresentados os parâmetros de configuração dos algoritmos, espectrogramas dos sinais, os resultados das métricas clássicas e janeladas e a restauração dos sinais obtida a partir dos sinais separados.

4.1 Metodologia

Para a realização dos testes foram utilizadas músicas com características próximas, tanto quanto possível, de músicas gravadas profissionalmente. Nesse sentido, a mixagem dessas músicas foi realizada com frequência de amostragem igual a 48kHz e codificação 16 bits. Os sinais acústicos foram gravados utilizando um microfone monoaural, cujas especificações são mostradas na Tabela 1.

Tabela 1 – Especificações do microfone MX150

Característica	Especificação Técnica
Tipo	Dinâmico
Polaridade	Unidirecional
Faixa de frequência	56Hz a 14kHz
Sensibilidade	56dB $\pm$ 3dB (0dB = 1V/PA a 1kHz)
Impedância	250 Ω $\pm$ 30% (em 1kHz)

Fonte: Adaptado de (Yoga Electronics Co., Ltd., 2014)

A Tabela 2 mostra as informações de mixagem utilizadas nos áudios de teste considerados no presente trabalho, onde R e L referem-se às parcelas nos canais R e L , respectivamente e C , indica uma mesma intensidade nos dois canais. O *software* utilizado para a mixagem foi o *Pro Tools*.

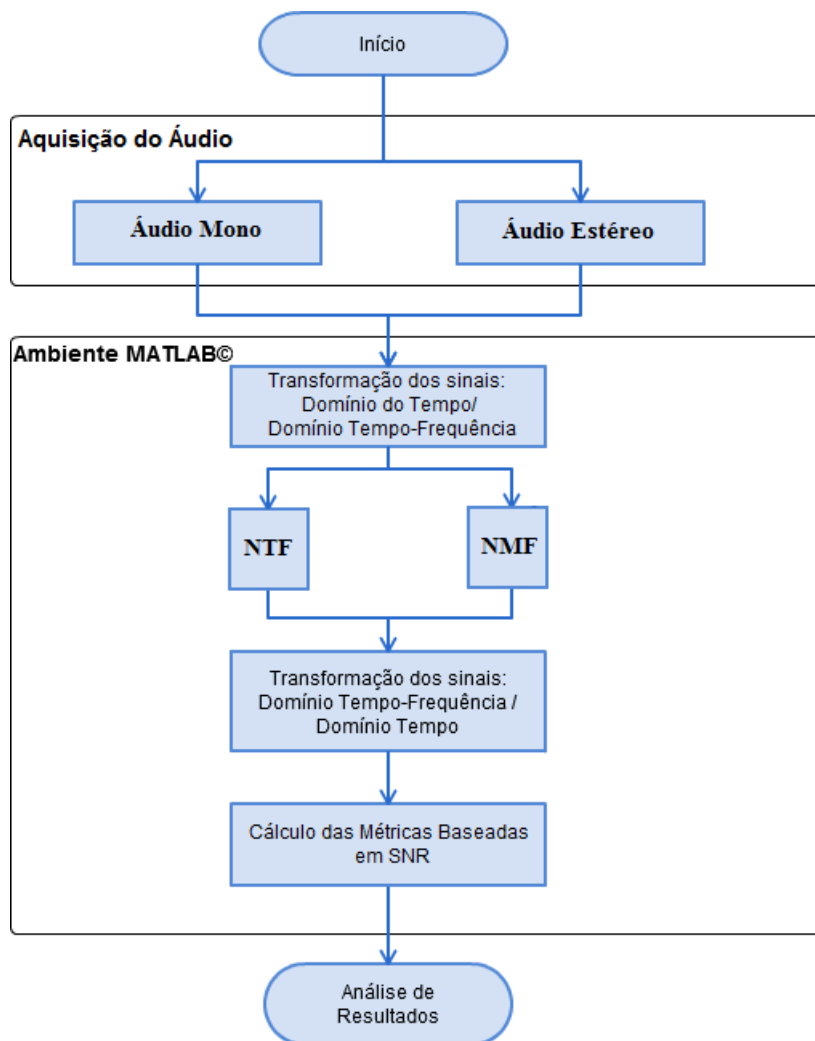
A Figura 14 mostra o fluxograma referente às etapas de simulação e validação. A aquisição dos sinais dos instrumentos é realizada utilizando o microfone descrito na Tabela 1. A mixagem dos áudios, em formato mono e estéreo, é feita utilizando o *Pro Tools*. A transformação entre os domínios do tempo e tempo-frequência é realizada por meio da Transformada de Fourier de Curta Duração (STFT - *Short-Time Fourier Transform*) e, o processo inverso, utilizando-se a Transformada Inversa de Fourier de Curta Duração

Tabela 2 – Parâmetros de mixagem dos sinais de teste estéreo

Fonte	<i>Panning</i>	(L - R)(dB)
Guitarra Clean	R83	-17,3
Guitarra Delay	L39	5,71
Guitarra Distorção	R33	-4,77
Baixo	C	0
Prato de Condução	R36	-5,23
Bumbo	C	0
Caixa	L22	3,02

(ISTFT - *Inverse Short-Time Fourier Transform*). A aplicação das técnicas NMF e NTF é realizada no domínio tempo-frequência, ou seja, nos espectrogramas referentes aos áudios a serem separados. O processo de obtenção do desempenho, através das métricas baseadas na SNR, é realizado utilizando como informação os sinais no domínio do tempo.

Figura 14 – Fluxograma da Simulação



Fonte: Adaptado de [Santos, 2015]

A fim de se estabelecer uma comparação entre os resultados obtidos a partir da

NMF e NTF, os parâmetros de simulação utilizados foram os mesmos para ambos os métodos. A Tabela 3 resume as principais configurações dos algoritmos, onde se destaca a diferença entre o número de iterações, devido à convergência mais lenta do algoritmo da NTF. Vale ressaltar que para a implementação da NTF, foi utilizado um *toolbox* específico para operações com tensores disponível em (BADER; KOLDA et al., 2015), uma vez que o MATLAB® não suporta esse tipo de dado.

Tabela 3 – Parâmetros de Ajuste das técnicas NMF e NTF

	NMF	NTF
Comprimento da Janela (Amostras)	1024	1024
Sobreposição (Amostras)	512	512
Comprimento da FFT (Amostras)	1024	1024
Número de Iterações	100	500

Fonte: Elaborado pelo autor

4.2 Áudios de Teste

Para que os testes dos algoritmos pudessem ser realizados, tanto quanto possível, próximos das condições práticas usuais, foram utilizadas músicas contendo 5 fontes distintas: guitarra, baixo e bateria, sendo essa última composta por bumbo, caixa e prato de condução. Além disso, todos os áudios apresentam um mesmo arranjo para o baixo e bateria, com variações apenas a guitarra. Com a escolha desse tipo de composição buscou-se investigar o desempenho dos algoritmos de separação de sinais quando aplicados a sons gerados pelos instrumentos harmônicos, estabelecer análises comparativas e dar continuidade aos estudos realizados em relação à bateria acústica, apresentados por (SANTOS, 2015). Os sinais de teste utilizados são descritos a seguir:

- 1 *Clean.wav* - Mistura contendo um arranjo de guitarra sem efeitos sonoros e de ritmo lento, apresentando 10 notas² e 2 acordes³.
- 2 *Delay.wav* - Mistura contendo um arranjo de guitarra incluindo-se o efeito sonoro *delay* e de ritmo lento, apresentando 10 notas.
- 3 *Dist.wav* - Mistura contendo um arranjo de guitarra incluindo-se o efeito sonoro de distorção, apresentando 6 notas, 2 *powerchord*⁴ e um harmônico artificial⁵.

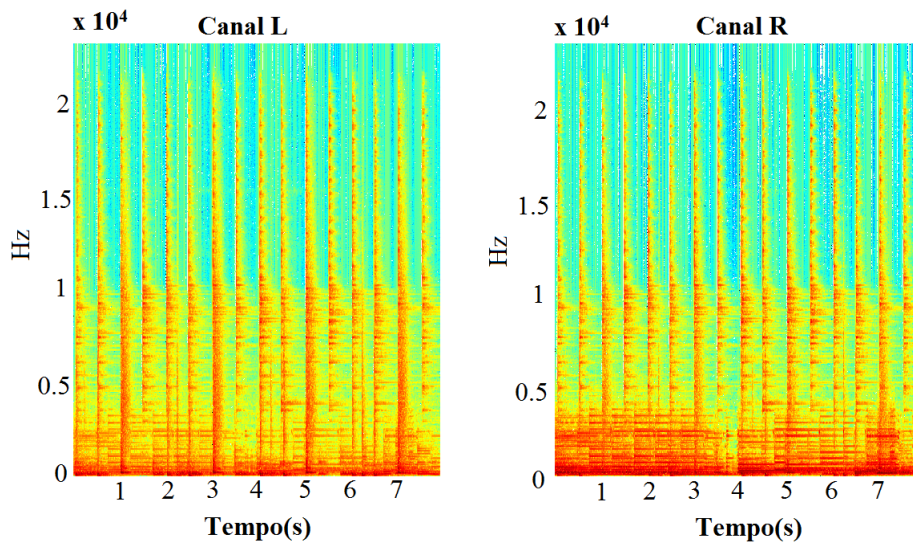
A Tabela 4 mostra a distribuição do número de notas em relação às fontes contidas nos sinais de teste indicados, respectivamente, nas 1^{as} linha e coluna da referida tabela.

²Nota: denominação utilizada quando apenas uma corda é tocada, solta ou em uma determinada casa

³O acorde refere-se a um conjunto de notas específicas, tocadas simultaneamente

⁴*Powerchord*: Também chamado de diáde, o *powerchord* é composto por duas notas quaisquer que formam um intervalo de tônica e quinta

⁵O harmônico artificial ocorre quando a nota, logo após ser tocada, é levemente abafada, produzindo-se um som mais próximo da componente de frequência fundamental da própria nota

Figura 15 – Espectrogramas da Mistura *Clean*

Fonte: Elaborado pelo autor

Conforme citado anteriormente na Seção 3, enquanto a NMF recebe o número de fontes contidas no áudio como um dos parâmetros de entrada, a NTF recebe o número total de notas. Considerou-se como número total de notas na guitarra, a soma do número de acordes e notas. O tempo total de duração dos áudios considerados foi o mesmo para todas as misturas, igual a 7,921s.

Tabela 4 – Número de Notas para a Simulação da NTF

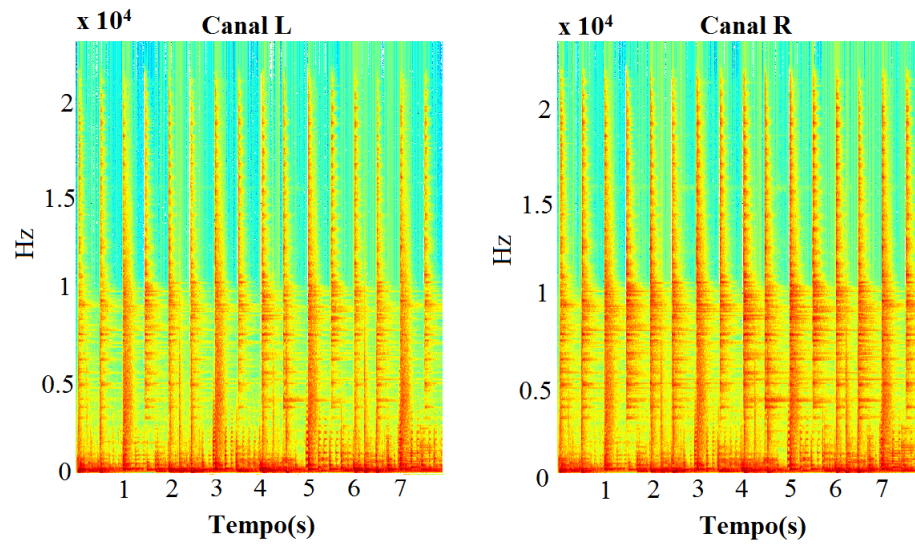
	Gui- tarra Clean	Gui- tarra Delay	Gui- tarra Distor- ção	Baixo	Bumbo	Caixa	Condu- ção	Total
Mix Clean	12			8	1	1	1	23
Mix Delay		10		8	1	1	1	21
Mix Distor- ção			9	8	1	1	1	20

Fonte: Elaborado pelo autor

As Figuras 15 a 17 apresetam os espectrogramas dos canais *L* e *R* para as misturas *Clean*, *Delay* e *Distorção*, obtidos a partir dos sinais de teste *Clean.wav*, *Delay.wav* e *Dist.wav*, respectivamente. Pode-se observar variações mais significativas nos arquivos *Clean* e *Distorção*, na faixa de frequência abaixo de $10kHz$, com maior espalhamento e intensidade até $5kHz$.

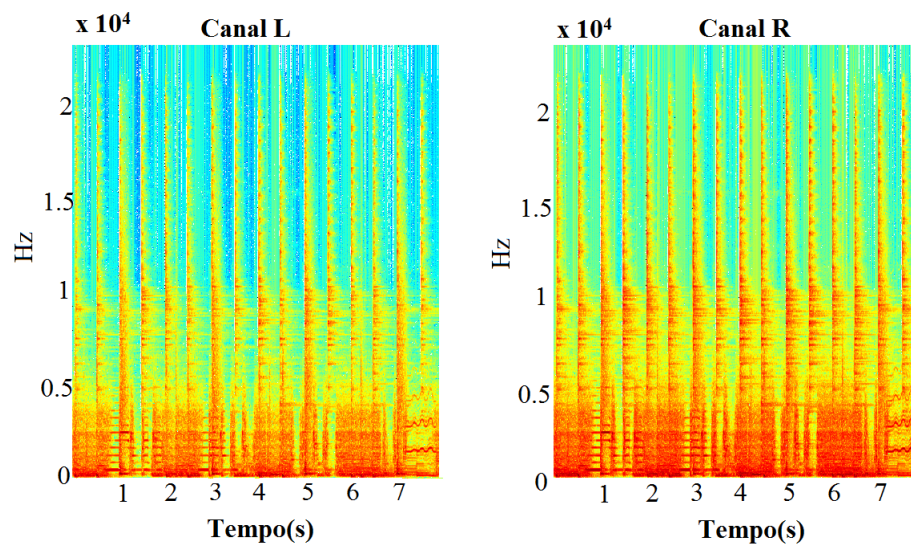
As Figuras 18 a 24 mostram os espectrogramas de referência dos canais *L* e *R* da Guitarra *Clean*, Guitarra *Delay*, Guitarra Distorcida, Baixo, Bumbo e caixa, respectiva-

Figura 16 – Espectrogramas da Mistura *Delay*



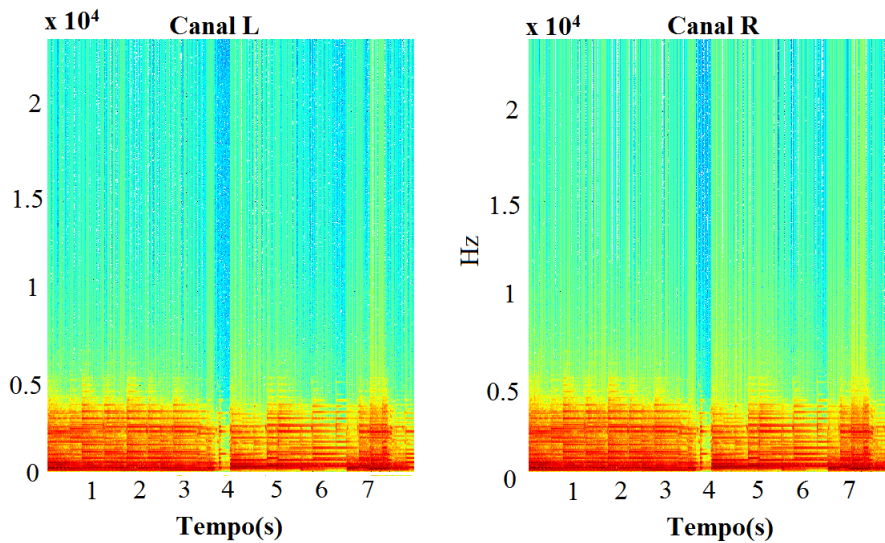
Fonte: Elaborado pelo autor

Figura 17 – Espectrogramas da Mistura Distorção



Fonte: Elaborado pelo autor

Figura 18 – Espectrogramas da Guitarra *Clean*



Fonte: Elaborado pelo autor

mente.

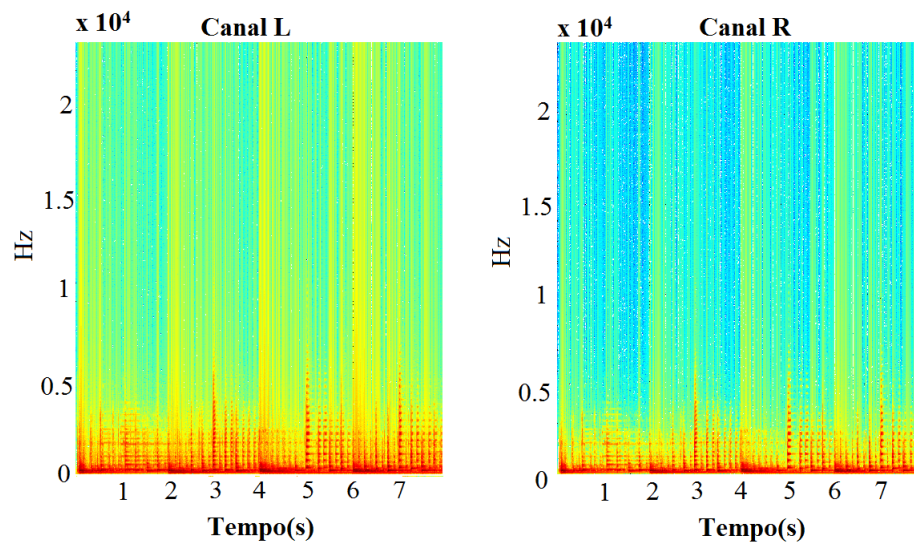
Observando-se a Figura 18, é possível notar que as componentes de frequência estão concentradas na faixa de 0 a 5kHz. Para a guitarra *Delay*, representada na Figura 19, nota-se como principal diferença uma menor homogeneidade de espectro se comparada com a guitarra *Clean*, devido principalmente ao fato dessa fonte ser composta basicamente de notas e apresentar menor sustentação harmônica. Já para a guitarra *Dist*, representada na Figura 20, é possível notar um maior espalhamento espectral em relação às demais, uma vez que a mesma foi submetida a um efeito de distorção. Além disso, nota-se a inserção do harmônico forçado no período de 7 a 8s, caracterizado oscilações bem definidas no espectrograma.

Em relação aos sons percussivos, os espectrogramas do bumbo e a caixa apresentam componentes de frequência em momentos definidos dos ataques, Figuras 22 e 23 respectivamente, ao contrário do prato de condução, cujo espectrograma na Figura 24 mostra espalhamento das frequências e maiores intensidades em torno de 5kHz a 10kHz.

4.3 Métodos de Validação

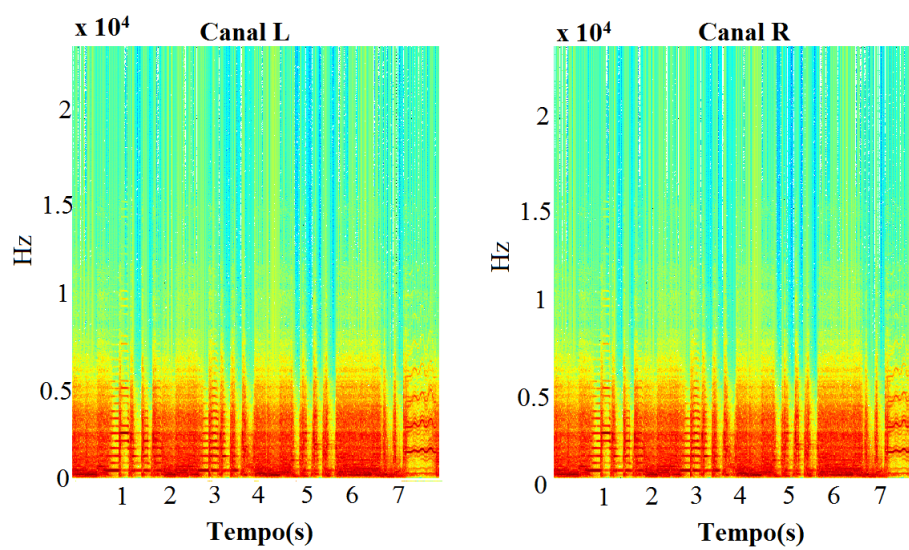
Os métodos de avaliação da qualidade da separação de fontes sonoras podem ser agrupados como métodos de validação subjetiva ou objetiva (BECH; ZACHAROV, 2006).

Para a validação subjetiva, usualmente, disponibiliza-se os sinais de áudio de cada uma das fontes, originais e obtidos pelo algoritmo de separação, a um grupo pré-selecionado de ouvintes para a avaliação cega dos diversos arquivos musicais. Esse grupo de avaliadores deverá ouvir, comparar os áudios e então, responder a questionários elaborados com o intuito de qualificar a separação. Esse tipo de abordagem apresenta grandes dificuldades, principalmente devido aos aspectos subjetivos dos indivíduos envolvidos na formulação

Figura 19 – Espectrogramas da Guitarra *Delay*

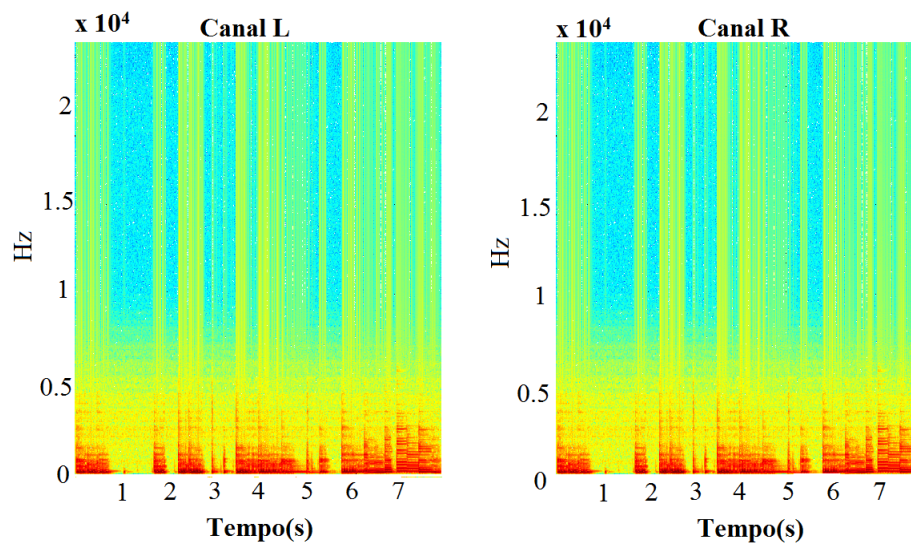
Fonte: Elaborado pelo autor

Figura 20 – Espectrogramas da Guitarra Distorcida



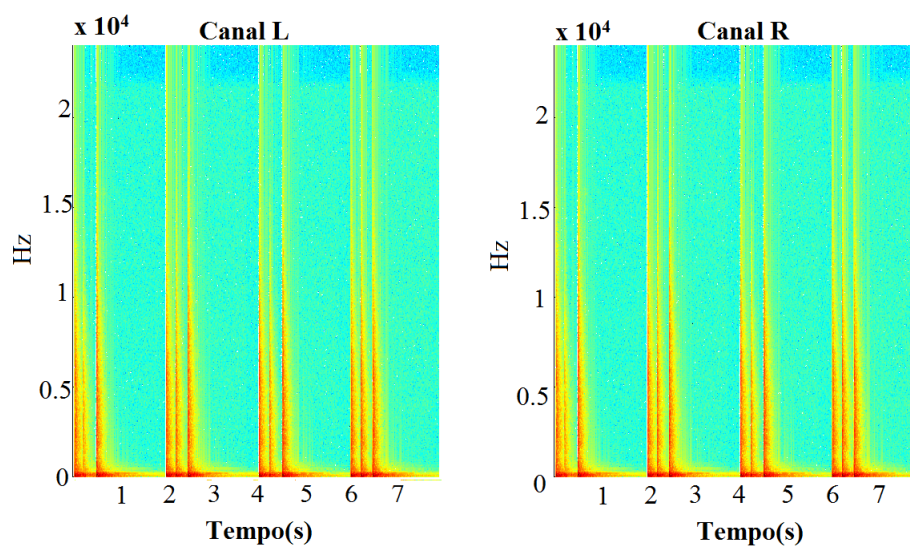
Fonte: Elaborado pelo autor

Figura 21 – Espectrogramas do Baixo



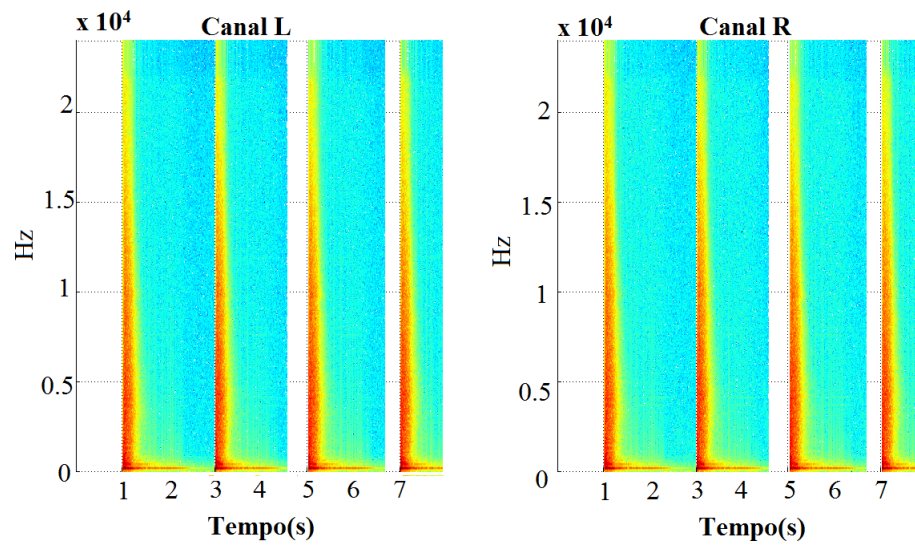
Fonte: Elaborado pelo autor

Figura 22 – Espectrogramas do Bumbo



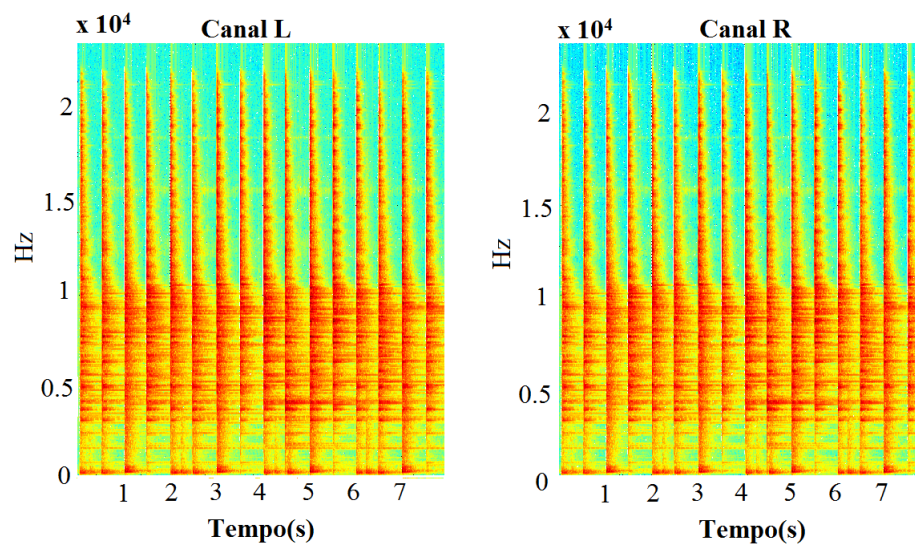
Fonte: Elaborado pelo autor

Figura 23 – Espectrogramas da Caixa



Fonte: Elaborado pelo autor

Figura 24 – Espectrogramas do Prato de Condução



Fonte: Elaborado pelo autor

e respostas dos testes, que estabelecem uma forte dependência quanto à formulação dos quesitos a serem avaliados, à escolha e tipo de avaliadores, às condições de execução dos testes, dentre outros. Em todos os casos, o aumento da confiabilidade sempre implicará fortemente em maior complexidade da avaliação. Esse tipo de avaliação é investigado para os casos de áudio digital em (BECH; ZACHAROV, 2006).

A validação objetiva baseia-se, usualmente, no espectro de frequência dos sinais ou em modelos matemáticos que simulam a percepção humana de um determinado fenômeno, no caso a sensibilidade auditiva e, dessa forma, possibilitam uma comparação quantitativa dos resultados (TYGEL, 2009).

A validação objetiva da separação de fontes acústicas é, comumente, realizada por meio de critérios derivados a partir da Relação Sinal-Ruído (SNR - *Signal-to-Noise Ratio*) (VINCENT; GRIBONVAL; FEVOTTE, 2006) ou por meio de algoritmos baseados em modelos psicoacústicos (FASTL, 2006).

Neste trabalho serão utilizados os critérios baseados na relação sinal-ruído introduzidos para sinais de áudio, mono e estéreo, em (VINCENT; GRIBONVAL; FEVOTTE, 2006), cuja implementação em ambiente MATLAB[®] encontra-se disponível para *download* em (VINCENT, 2015).

4.3.1 Avaliação Objetiva Baseada na SNR

De acordo com (VINCENT; GRIBONVAL; FEVOTTE, 2006), as métricas baseadas em SNR consideram que o sinal da fonte estimada da mistura (s_j) é igual à soma do sinal da fonte original (s_{alvo}) às interferências (e_{inter}) causadas por outras fontes, aos erros de quantização ou artefatos (e_{artef}) do processo de aquisição/separação e a uma parcela e_{noise} referente ao ruído contido na mistura original. A representação matemática do sinal estimado da mistura é, então, dada pela equação (55):

$$s_j = s_{alvo} + e_{inter} + e_{artef} + e_{noise} \quad (55)$$

As métricas a serem utilizadas são apresentadas a seguir (VINCENT; GRIBONVAL; FEVOTTE, 2006):

- 1 Razão Fonte-Distorção (SDR - *Source to Distortion Ratio*): Visa quantificar a qualidade da separação de uma forma ampla, a partir da razão entre o valor quadrático do sinal da fonte original em relação à soma dos quadrados das parcelas referentes às interferências, erros de quantização e ruídos. A equação (56) apresenta a expressão utilizada para o cálculo da SDR.

$$SDR_{dB} = 10 \log \left(\frac{\|s_{alvo}\|^2}{\|e_{inter} + e_{artef} + e_{noise}\|^2} \right) \quad (56)$$

- 2 Razão Fonte-Interferência (SIR - *Source-to-Interference Ratio*): Mensura a interferência das demais fontes na mistura estimada, sendo calculada pela expressão (57),

$$SIR_{dB} = 10 \log \left(\frac{\|s_{alvo}\|^2}{\|e_{inter}\|^2} \right). \quad (57)$$

- 3 Razão Fonte-Artefatos (SAR - *Sources-to-Artifacts Ratio*): Essa razão visa medir a quantidade de imperfeições geradas nas fases de aquisição/processamento e pode ser avaliada por meio da Equação (58).

$$SAR_{dB} = 10 \log \left(\frac{\|s_{alvo} + e_{inter} + e_{noise}\|^2}{\|e_{artef}\|^2} \right) \quad (58)$$

Para todas as métricas baseadas em SNR, valores mais altos indicam uma presença mais significativa do sinal original no sinal estimado ou recuperado, de tal forma que quanto maior o valor dessas métricas, melhor será a qualidade da separação.

4.4 Análise dos Resultados

Os experimentos envolveram a aplicação das técnicas de separação NMF e NTF, respectivamente, aos sinais de áudio monoaural e estéreo/multicanal descritos na Seção 4.2. Em ambos os casos, os algoritmos foram associados às divergências de Kullback-Leibler e de Itakura-Saito e, como regra de atualização, utilizou-se o Método Gradiente Descendente. A seguir, são apresentados os resultados obtidos.

4.4.1 Análise da separação dos sinais monoaurais

A Tabela 5 mostra os valores das métricas SIR, SDR e SAR calculados após a separação dos sons dos instrumentos por meio da técnica NMF, destacados em vermelho ou azul, os maiores valores, em dB, relativos às divergências Kullback-Leibler ou Itakura-Saito, respectivamente.

Com base nas métricas SIR, SDR e SAR, os resultados indicam, em sua maioria, melhor desempenho na separação se aplicada a Divergência de Kullback-Leibler, alcançando maiores valores, respectivamente, 10, 11 e 12 ocorrências para as misturas *Clean*, *Delay* e *Distorção*. Observa-se que a Divergência de Itakura-Saito apresentou melhor desempenho em relação à Kullback-Leibler apenas na separação do baixo, para todos os sinais de testes.

Os resultados relativos à métrica SAR indicam a presença menos significativa dos efeitos de truncamento e precisão numérica finita quando utilizada a Divergência de Kullback-Leibler. Porém as dificuldades para se identificar esses efeitos na análise espectral dos sinais impedem que conclusões mais objetivas possam ser elaboradas, até o momento.

Os resultados em relação à guitarra apresentam altos valores SIR para a mistura *Dist.wav*, se comparados aos demais. Embora esses resultados sejam indicativos de um bom desempenho na separação dos sons desse instrumento, essa condição não foi comprovada nas audições realizadas. Pelo contrário, as suas notas foram praticamente suprimidas, com ocorrências fracas apenas em instantes onde o instrumento na mistura original era mais significativo. Evidencia-se, portanto, limitações das métricas para a análise dos resultados.

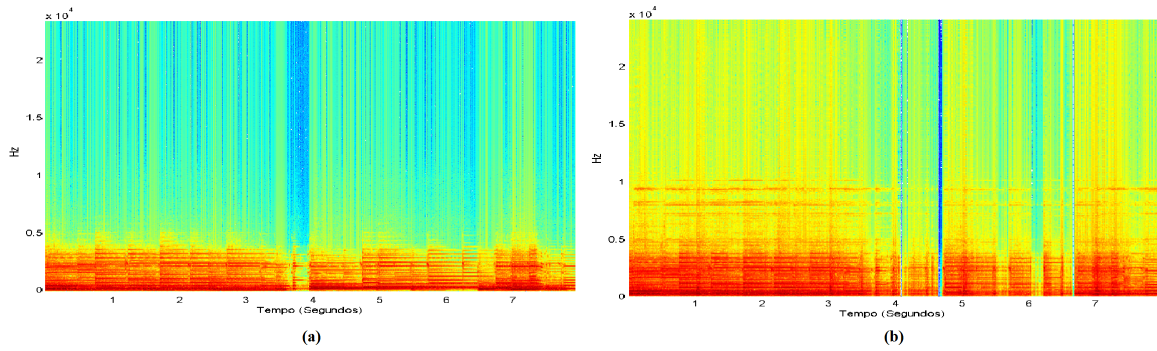
Tabela 5 – Resultado da Separação via NMF

		SDR(dB)		SIR(dB)		SAR(dB)	
	<i>Fonte</i>	<i>KL</i>	<i>IS</i>	<i>KL</i>	<i>IS</i>	<i>KL</i>	<i>IS</i>
<i>Clean</i>	Guitarra	4,076	2,139	11,731	5,680	5,176	5,716
	Baixo	-12,604	-7,008	-11,350	-4,959	5,063	3,401
	Condução	7,410	7,165	20,518	24,126	7,666	7,270
	Bumbo	-0,204	-3,294	1,081	0,014	8,215	2,441
	Caixa	-6,232	-14,825	-1,344	-11,229	-0,795	-0,785
<i>Delay</i>	Guitarra	-2,647	-8,409	2,143	-4,823	1,175	0,152
	Baixo	-8,303	-3,652	-7,429	-1,833	7,244	5,030
	Condução	8,638	7,358	22,618	21,604	8,839	7,554
	Bumbo	-0,019	-3,193	1,141	-0,699	8,756	3,777
	Caixa	-3,996	-10,236	0,635	-8,597	0,542	3,948
<i>Dist</i>	Guitarra	5,918	4,999	19,994	18,393	6,134	5,265
	Baixo	-19,186	-6,543	-18,043	-4,421	5,280	3,346
	Condução	5,944	-0,068	12,898	5,995	7,139	2,142
	Bumbo	-0,008	-7,031	1,419	-6,092	7,878	7,131
	Caixa	-8,839	-5,013	-7,454	-1,084	4,970	0,825

As Figuras 25, 26 e 27 mostram os espectrogramas dos sinais monoaurais das guitarras, originais e após a separação pela técnica NMF, para as misturas *Clean.wav*, *Delay.wav* e *Dist.wav*, respectivamente. Foram considerados apenas os melhores resultados, obtidos por meio da Divergência de Kullback-Leibler.

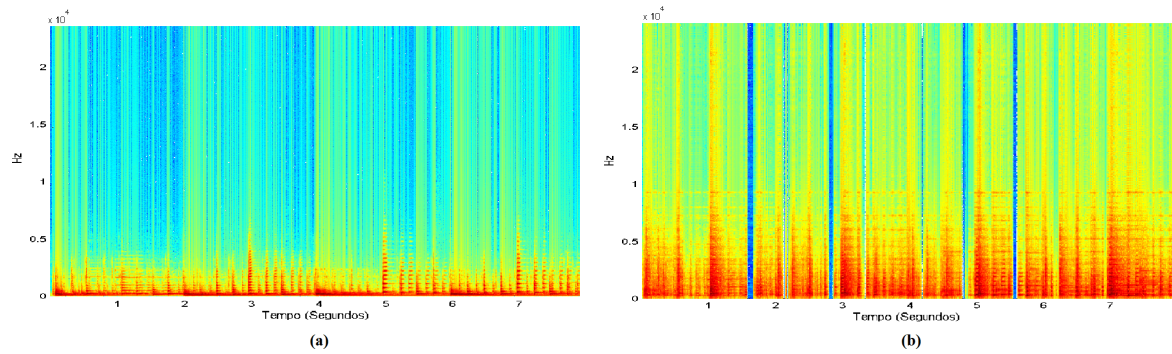
Pode-se observar a interferência das demais fontes, em especial do prato de condução, comparando-se os espectrogramas apresentados nas Figuras 24, 26 e 27. As faixas em cor azul representam períodos de tempo onde todas as componentes de frequência, na faixa indicada, têm pouco ou nenhuma contribuição no sinal. Em termos práticos, um breve silêncio foi adicionado nesse período.

Figura 25 – Espectrogramas da guitarra *clean* - (a) original; (b) separada via NMF-KL



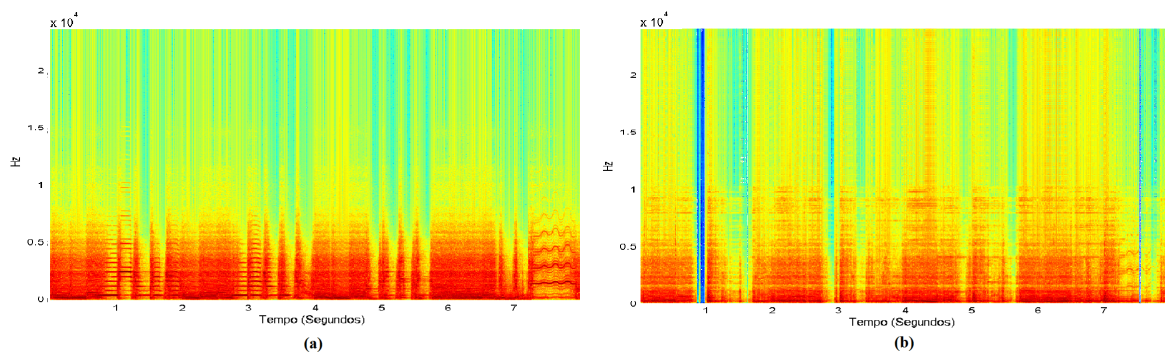
Fonte: Elaborado pelo autor

Figura 26 – Espectrograma da guitarra *delay* - (a) original; (b) separada via NMF-KL



Fonte: Elaborado pelo autor

Figura 27 – Espectrograma da guitarra *dist* - (a) original; (b) separada via NMF-KL



Fonte: Elaborado pelo autor

4.4.2 Análise da separação dos sinais estéreo

A Tabela 6 mostra os valores das métricas SIR, SDR e SAR calculados após a separação dos sons dos instrumentos por meio da técnica NTF, destacados em vermelho ou azul, os maiores valores, em dB, relativos às divergências Kullback-Leibler ou Itakura-Saito, respectivamente.

Novamente, os resultados relativos à métrica SAR indicam a presença menos significativa dos efeitos de truncamento e precisão numérica finita quando aplicada a Divergência Kullback-Leibler.

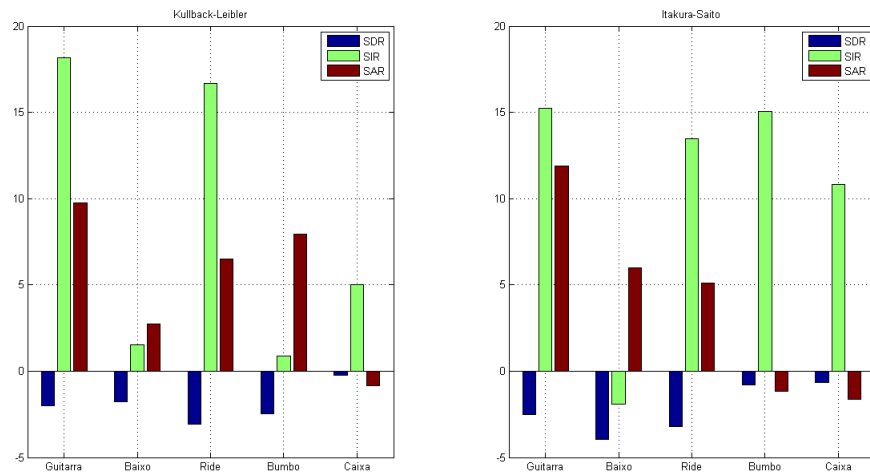
Tabela 6 – Resultado da Separação via NTF

		SDR(dB)		SIR(dB)		SAR(dB)	
	<i>Fonte</i>	<i>KL</i>	<i>IS</i>	<i>KL</i>	<i>IS</i>	<i>KL</i>	<i>IS</i>
<i>Clean</i>	Guitarra	-2,010	-2,509	18,162	15,220	9,735	11,900
	Baixo	-1,763	-3,951	1,542	-1,889	2,725	5,989
	Condução	-3,068	-3,200	16,673	13,461	6,517	5,116
	Bumbo	-2,489	-0,782	0,865	15,037	7,952	-1,186
	Caixa	-0,232	-0,662	5,034	10,803	-0,851	-1,611
<i>Delay</i>	Guitarra	-2,797	-2,438	5,168	5,455	9,308	5,833
	Baixo	-2,864	-0,346	-1,676	2,111	8,048	-4,453
	Condução	-2,801	-3,394	21,492	11,876	8,834	5,161
	Bumbo	-0,854	-3,333	11,043	2,128	1,293	6,159
	Caixa	0,012	0,044	11,441	17,020	-1,371	1,124
<i>Dist</i>	Guitarra	-3,125	-3,468	11,432	10,207	10,508	9,959
	Baixo	-2,116	0,016	-3,494	8,637	5,283	0,225
	Condução	-3,586	-3,226	8,623	10,349	3,523	2,181
	Bumbo	-1,480	-2,562	1,628	5,583	6,104	5,311
	Caixa	0,062	-0,692	10,338	14,734	-0,808	-1,028

Em geral, os dois tipos de divergência apresentaram níveis de desempenho equivalentes, para todos os tipos de misturas. Embora os resultados apresentados em (FEVOTTE; OZEROV, 2010) indiquem uma melhor separação via Divergência de Kullback-Leibler para elementos percussivos, os resultados obtidos no presente trabalho mostram que a Divergência de Itakura-Saito apresenta melhores resultados, se considerados os elementos menos evidentes na mistura. Em outras palavras, com exceção do baixo para a mistura *Clean* e do bumbo para a mistura *Delay*, a Divergência de Itakura-Saito mostrou-se mais adequada para a separação do baixo, bumbo e caixa, misturas de características de frequência próximas.

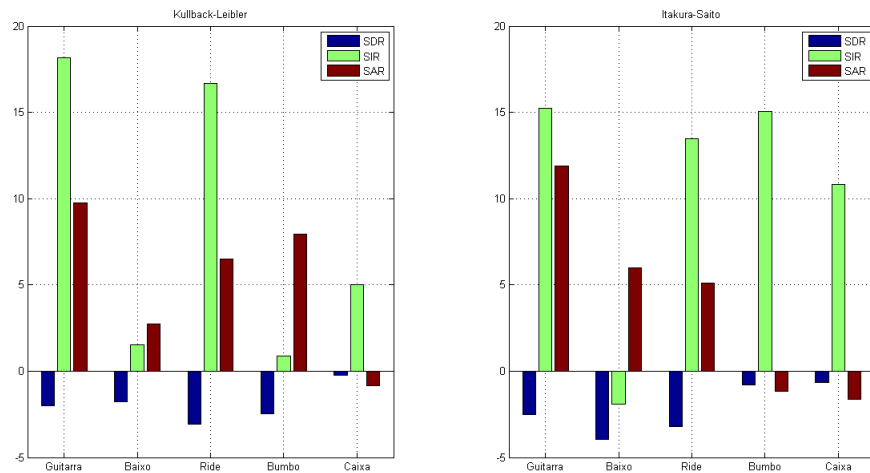
Isso se deve à própria natureza das divergências consideradas, ou seja, devido à Divergência de Kullback-Leibler ser basicamente uma medida da entropia, ela pode apresentar melhores resultados quando se considera fontes bem destacadas na mistura. A Divergência de Itakura-Saito, por ser invariante à escala, comporta-se de maneira

Figura 28 – Comparação dos resultados da NTF usando as Divergências de Kullback-Leibler e de Itakura-Saito para *Clean.wav*



Fonte: Elaborado pelo autor

Figura 29 – Comparação dos resultados da NTF usando as Divergências de Kullback-Leibler e de Itakura-Saito para *Delay.wav*



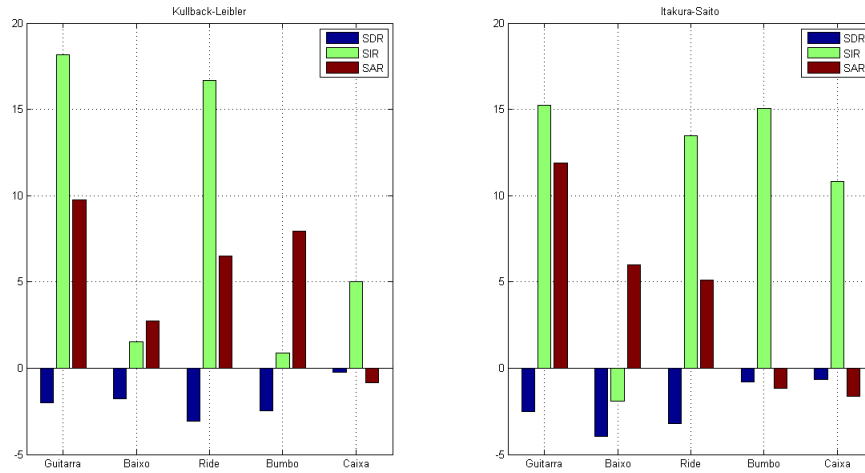
Fonte: Elaborado pelo autor

equivalente para sinais de maior ou menor contribuição na mistura.

As Figuras 28, 29 e 30 mostram, graficamente, as diferenças dos resultados da NTF usando as Divergências de Kullback-Leibler e Itakura-Saito para os 3 áudios considerados. Evidencia-se uma maior uniformidade das métricas obtidas usando Itakura-Saito.

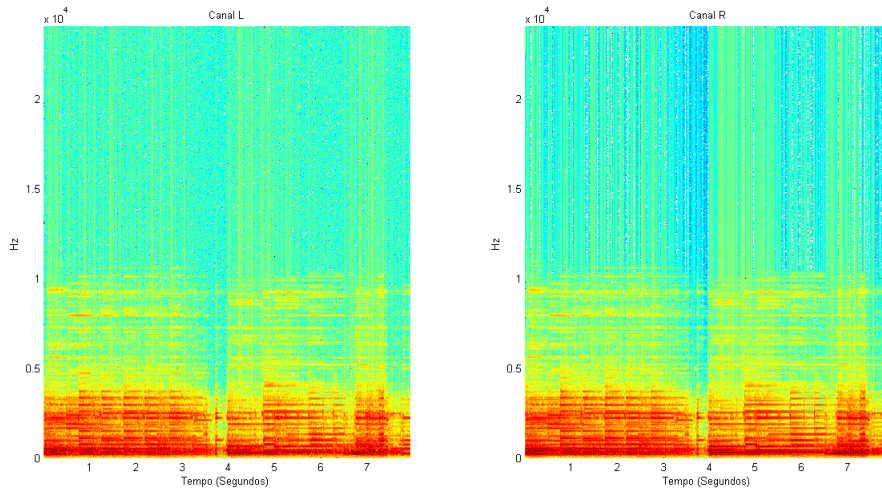
As Figuras 31, 32 e 33 mostram os espectrogramas dos canais L e R referente à separação da guitarra nas misturas *Clean.wav*, *Delay.wav* e *Dist.wav*, respectivamente, considerando-se a divergência que apresentou os melhores resultados SIR ou seja, Itakura-Saito para *Delay.wav* e Kullback-Leibler para *Clean.wav* e *Dist.wav*. É possível notar uma melhor separação por meio da NTF se comparados com os espectrogramas obtidos

Figura 30 – Comparação dos resultados da NTF usando as Divergências de Kullback-Leibler e de Itakura-Saito para *Dist.wav*



Fonte: Elaborado pelo autor

Figura 31 – Espectrogramas dos canais L e R da guitarra *clean* separada via NTF-KL



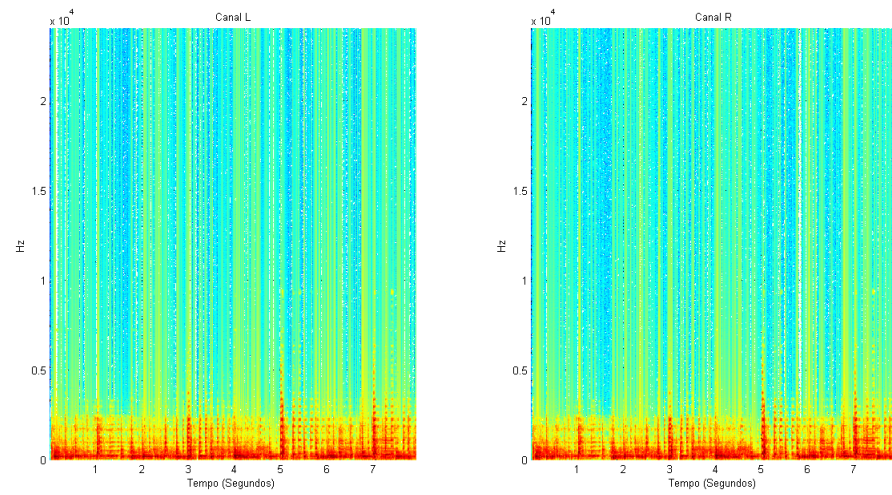
Fonte: Elaborado pelo autor

via NMF, Figuras 25 a 27. Além da redução das interferências entre as fontes, a Figura 33 evidencia que a NTF é capaz de representar algumas particularidades dos sinais de áudio tais como o harmônico artificial incluído na mistura *Dist.wav* indicado por três oscilações que se destacam no intervalo de 7 a 8s.

4.5 Análise comparativa das técnicas NTF e NMF

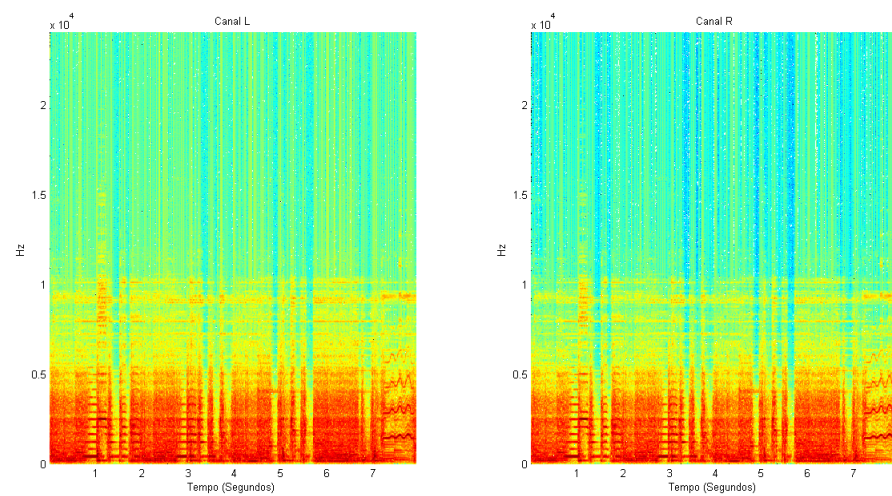
No contexto da separação de fontes sonoras, outra importante tarefa refere-se à reconstrução da mistura original através das fontes separadas, buscando-se avaliar a inserção de ruídos, interferências e/ou artefatos pelos algoritmos utilizados.

Figura 32 – Espectrogramas dos canais L e R da guitarra *delay* separada via NTF-IS



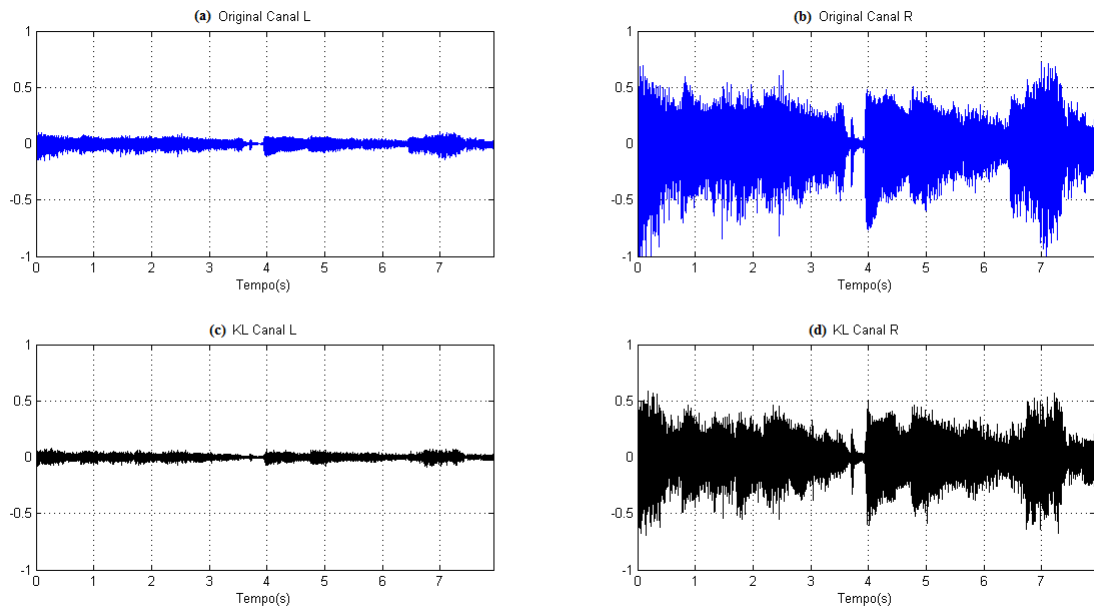
Fonte: Elaborado pelo autor

Figura 33 – Espectrogramas dos canais L e R da guitarra distorcida separada via NTF-KL



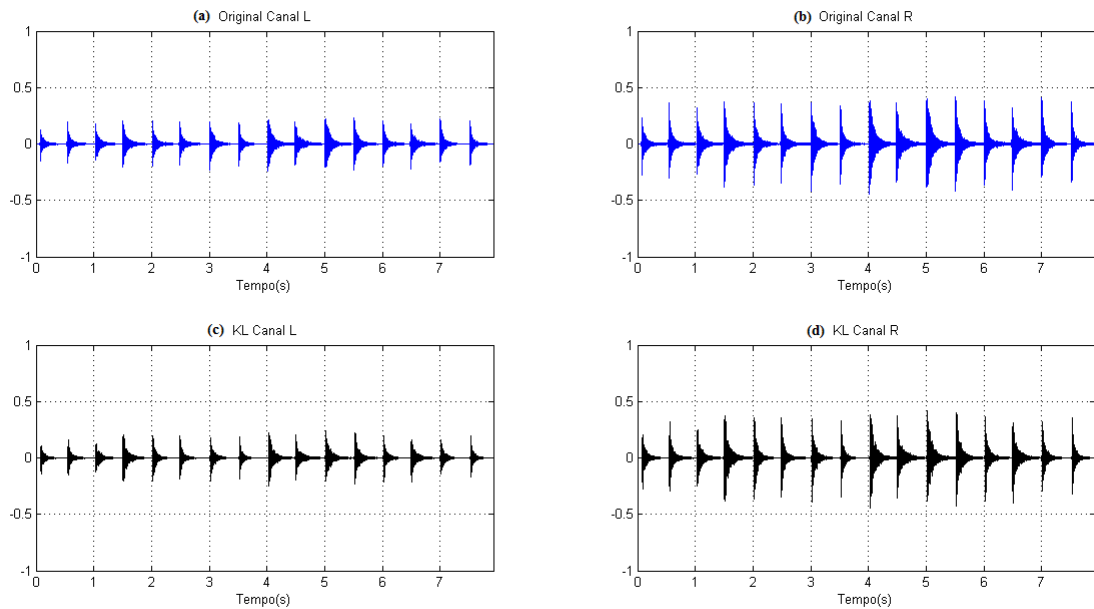
Fonte: Elaborado pelo autor

Figura 34 – Formas de onda dos sinais da guitarra *clean*, na versão original e após separação via NTF-KL.



Fonte: Elaborado pelo autor

Figura 35 – Formas de onda dos sinais do prato de condução *clean*, na versão original e após separação via NTF-KL.



Fonte: Elaborado pelo autor

As Figuras 34 e 35 mostram as formas de onda dos sinais da guitarra e do prato de condução originais da mistura *clean* e obtidos por meio das técnicas NTF-KL para os canais L e R, respectivamente. Pode-se observar uma boa aproximação entre as fontes originais e separadas exceto, principalmente, pelas amplitudes.

Ainda nesse sentido, a Tabela 7 apresenta um comparativo da métrica SAR referente às técnicas NTF, utilizando as Divergências de Kullback-Leibler e Itakura-Saito. Nota-se o melhor desempenho da Divergência de Itakura-Saito, principalmente nas misturas *clean* e *dist*, o que fornece um indicativo para essa divergência ser mais usual nas aplicações de áudio.

Tabela 7 – SAR dos sinais reconstruídos

Mistura	KL	IS
<i>Clean</i>	13,049	20,286
<i>Delay</i>	20,278	20,278
<i>Dist</i>	17,688	18,381

A Tabela 8 apresenta as diferenças entre os resultados das métricas conforme as Tabelas 6 e 5, onde os valores (dB) favoráveis à técnica NTF são positivos e foram destacados em vermelho.

Tabela 8 – Comparação entre as técnicas NMF e NTF: $\Delta_S = S_{NTF} - S_{NMF}$

	<i>Fonte</i>	$\Delta_{SDR}(\text{dB})$		$\Delta_{SIR}(\text{dB})$		$\Delta_{SAR}(\text{dB})$	
		<i>KL</i>	<i>IS</i>	<i>KL</i>	<i>IS</i>	<i>KL</i>	<i>IS</i>
<i>Clean</i>	Guitarra	-6,086	-4,648	6,431	9,540	4,559	6,184
	Baixo	10,841	3,057	12,892	3,070	-2,338	2,588
	Condução	-10,478	-10,365	-3,845	-10,665	-1,149	-2,154
	Bumbo	-2,285	2,512	-0,216	15,023	-0,263	-3,627
	Caixa	6,000	14,163	6,378	22,032	-0,056	-0,826
<i>Delay</i>	Guitarra	-0,150	5,971	3,025	10,278	8,133	5,681
	Baixo	5,439	3,306	5,573	3,944	0,804	-9,483
	Condução	-11,439	-10,752	-1,126	-9,728	-0,005	-2,393
	Bumbo	-0,835	-0,140	9,902	2,827	-7,463	2,382
	Caixa	4,008	10,280	10,806	25,617	-1,913	-2,824
<i>Dist</i>	Guitarra	-9,043	-8,467	-8,562	-8,186	4,374	4,694
	Baixo	17,070	6,559	14,549	13,058	0,003	-3,121
	Condução	-9,530	-3,158	-4,275	4,354	-3,616	0,039
	Bumbo	-1,472	4,469	0,209	11,675	-1,774	-1,820
	Caixa	8,901	4,321	17,792	15,818	-5,778	-1,853

Considerando os resultados obtidos, pode-se concluir que:

- 1 Levando em consideração a métrica SIR, com exceção do prato de condução e da guitarra para a mistura *Dist.wav*, a técnica NTF apresentou desempenho superior à NMF. Isso deve-se aos fatos de a NTF dividir o áudio em seus componentes, enquanto a NMF divide nas fontes e, principalmente, por uma redundância espacial conferida ao áudio pela característica multicanal (estéreo, no presente caso). Levando-se em consideração a SDR e SAR não foi possível indicar qual das técnicas apresentou

melhor desempenho.

- 2 Embora a NTF apresente resultados superiores em relação a NMF, destaca-se alguns pontos negativos, como por exemplo a complexidade do algoritmo e menor velocidade de convergência;
- 3 As Divergências de Kullback-Leibler e Itakura-Saito apresentaram desempenhos equivalentes, quando comparadas entre si. No entanto, a Divergência de Itakura-Saito apresenta uma característica mais constante para todas as fontes, além de sua reconstrução apresentar melhor desempenho;
- 4 As métricas com base em SNR, apesar de serem muito utilizadas na área de processamento de áudio, não conseguem representar de forma conclusiva os resultados. Nesse sentido, a avaliação subjetiva, apesar de todas as restrições e dificuldades indicadas anteriormente, ainda constitui uma importante ferramenta auxiliar no processo de validação. Devido à limitação de tempo inerente ao desenvolvimento das dissertações de mestrado, esse tipo de avaliação não foi incluído no escopo do trabalho.

4.5.1 Particularidades da NTF

A técnica NTF, apesar de bastante semelhante à NMF, apresenta algumas diferenças significativas no contexto do áudio digital. Dentre essas, as matrizes e tensores utilizados na decomposição e o significado dos seus respectivos postos em cada uma dessas versões.

Enquanto para a NMF, os postos das matrizes W e H constituem o número de fontes, para a NTF os postos dos tensores G , A e S referem-se ao número de notas contidas na mistura (FITZGERALD; CRANITCH; COYLE, 2005). Embora as notas sejam mais facilmente identificáveis em misturas mais simples, em misturas mais complexas, os postos referem-se ao número de faixas de frequência do sinal de áudio de entrada. Neste trabalho, nos áudios utilizados para teste, o prato de condução apresentou um amplo espalhamento espectral, elevando a contribuição das notas contidas nessa fonte. Esse comportamento não era previsto, uma vez que o prato de condução deveria apresentar uma nota única.

Além disso, como o áudio é decomposto em diversas faixas de frequência, torna-se necessária a clusterização dessas faixas nas fontes componentes do áudio. Embora sejam encontrados trabalhos que utilizam algoritmos para essa finalidade como, por exemplo, o *k-nearest Neighbour* (FITZGERALD; CRANITCH; COYLE, 2005), de acordo com Ozerov et al. (2011), a clusterização manual ainda fornece melhores resultados. No presente trabalho, como o baixo e o bumbo foram mixados com a mesma relação entre canais, o algoritmo não foi capaz de formar a fonte correspondente ao baixo e, portanto, os resultados foram obtidos por meio da clusterização manual

Tabela 9 – Parâmetros da métrica janelada

Parâmetro	Configuração
Tipo de Janela	Hamming
Comprimento da Janela	$\approx 500\text{ms}$
Sobreposição	50%

4.6 Métricas Janeladas

Conforme citado anteriormente, apesar das métricas baseadas em SNR serem comumente utilizadas para a avaliação objetiva em um grande número de trabalhos sobre áudios digitais, disponíveis na literatura, nem sempre essas métricas apresentam resultados conclusivos. Essa limitação pode ser evidenciada na separação da guitarra através da técnica NMF e considerando a mistura *Dist.wav*, cujos valores das métricas indicaram uma qualidade de separação elevada, muito diferentemente dos resultados audíveis.

Para contornar essa limitação, foi proposta uma modificação quanto a aplicação das métricas clássicas neste trabalho. Optou-se por aplicar um janelamento aos sinais de áudio a serem analisados e, assim, observar a evolução dessas métricas ao longo do tempo. Em consequência do janelamento, substituiu-se a FFT para a estimação dos parâmetros apresentados na subseção (4.3.1) pela STFT, mais adequada para a análise de sinais não estacionários. A Tabela 9 resume os parâmetros utilizados para o cálculo das métricas janeladas.

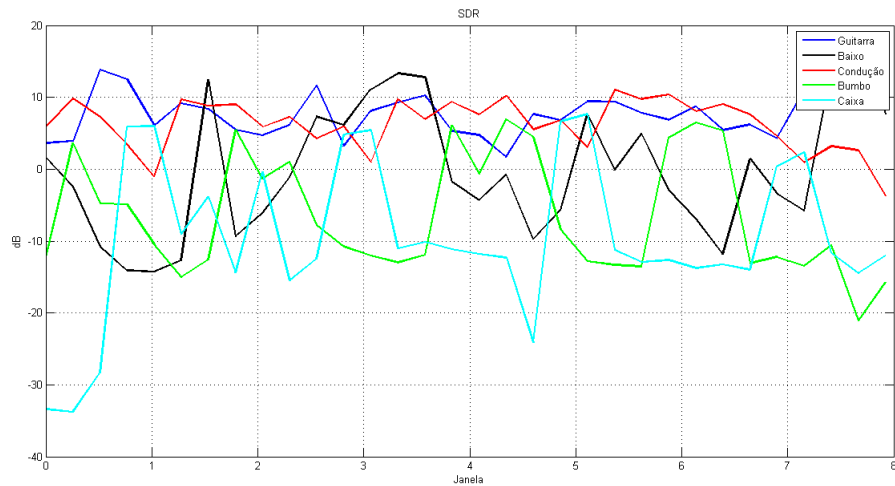
Alguns dos resultados são apresentados nas Figuras 36 a 39. Vale destacar os comportamentos semelhantes de todas as métricas analisadas (SDR, SIR e SAR), com bruscas variações entre quadros e, na maioria da vezes, contidas em faixas de amplitude aproximadas (dB), salvo alguns picos não-repetitivos.

As Figuras 36 e 37 mostram os resultados das métricas SDR e SIR janeladas, considerando a mistura *Dist.wav* e a técnica NMF-KL. Sob condições idênticas, as Figuras 38 e 39 apresentam os resultados obtidos pela técnica NMF-IS.

As Figuras 40 e 41 mostram as variações no tempo e as médias calculadas para as métricas janeladas e clássicas, para a guitarra na mistura *Dist.wav*, considerando-se a NMF e as divergências de Kullback-Leibler e Itakura-Saito, respectivamente. Apesar de as médias ainda apresentarem valores relativamente elevados, nota-se uma redução de cerca de 6dB e 8dB, no valor de SIR, para as divergências Kullback-Leibler e Itakura-Saito, respectivamente. Pôde-se constatar que os resultados se mostraram mais condizentes com as observações audíveis.

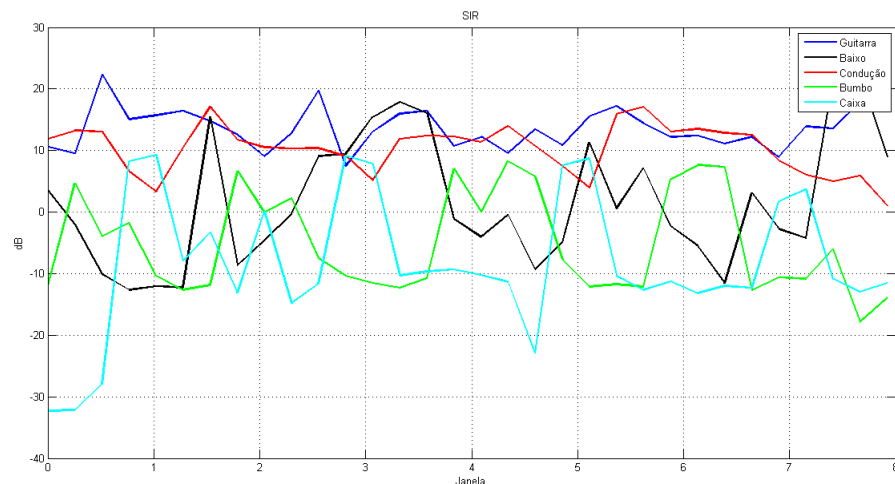
Por apresentarem faixas de frequência muito próximas, avaliou-se também a variação da métrica SIR janelada para o baixo e bumbo. As Figuras 42 e 43 apresentam os resultados obtidos, considerando-se a mistura *Dist.wav* para as divergências de Kullback-Leibler e Itakura-Saito, bem como os espectrogramas do baixo e bumbo, respectivamente.

Figura 36 – Aplicação da SDR Janelada à mistura *Dist.wav*, utilizando-se a técnica NMF-KL



Fonte: Elaborado pelo autor

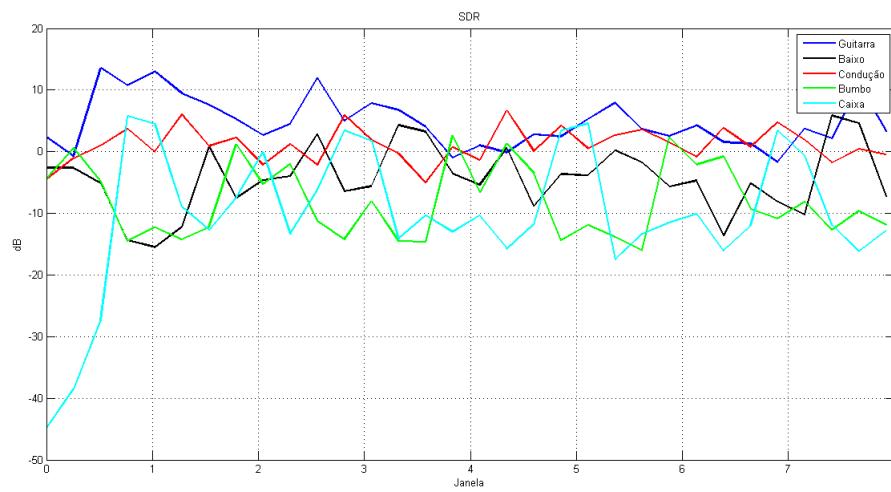
Figura 37 – Aplicação da SIR Janelada à mistura *Dist.wav*, utilizando-se a técnica NMF-KL



Fonte: Elaborado pelo autor

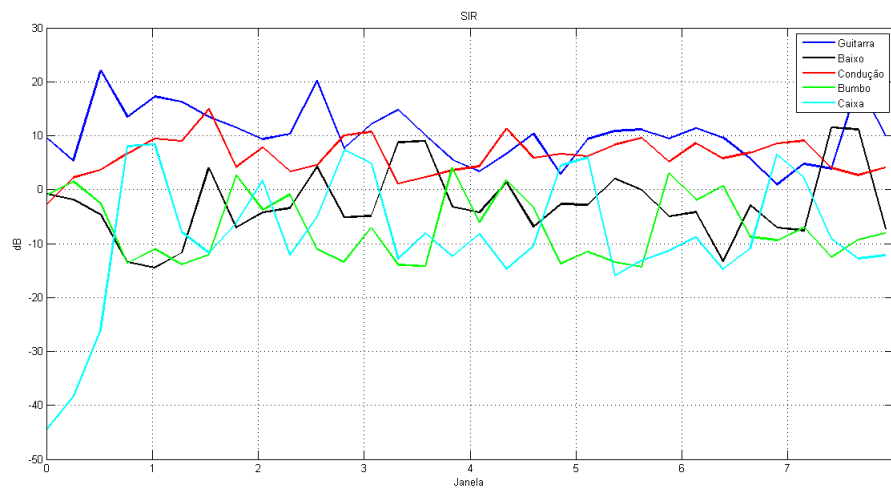
Como o bumbo é um instrumento que não apresenta sustentação harmônica e, pela composição dos áudios de teste, sua presença na mistura ocorre em momentos bem definidos, permitindo avaliar a influência na SIR do ataque ao bumbo. Como esperado, observa-se com base nas Figuras 42 e 43, que nos instantes de tempo em que o baixo e bumbo estão presentes, ocorre uma alternância nos sentidos das inclinações da SIR para cada um dos instrumentos.

Figura 38 – Aplicação da SDR Janelada à mistura *Dist.wav*, utilizando-se a técnica NMF-IS



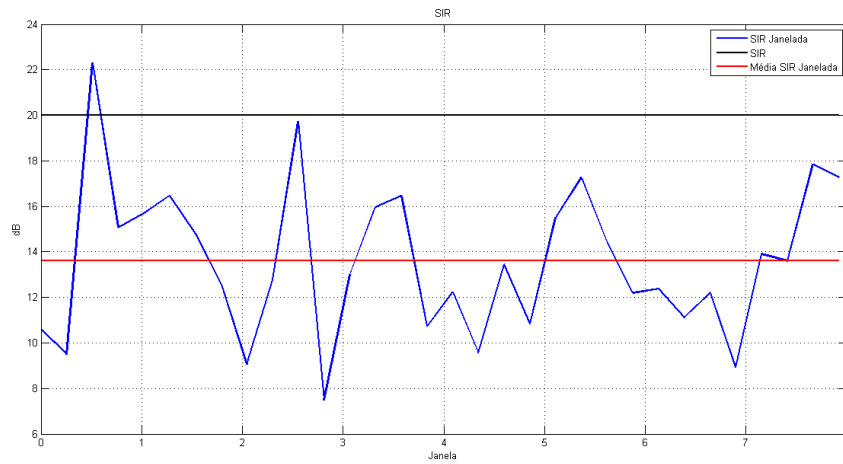
Fonte: Elaborado pelo autor

Figura 39 – Aplicação da SIR Janelada à mistura *Dist.wav*, utilizando-se a técnica NMF-IS



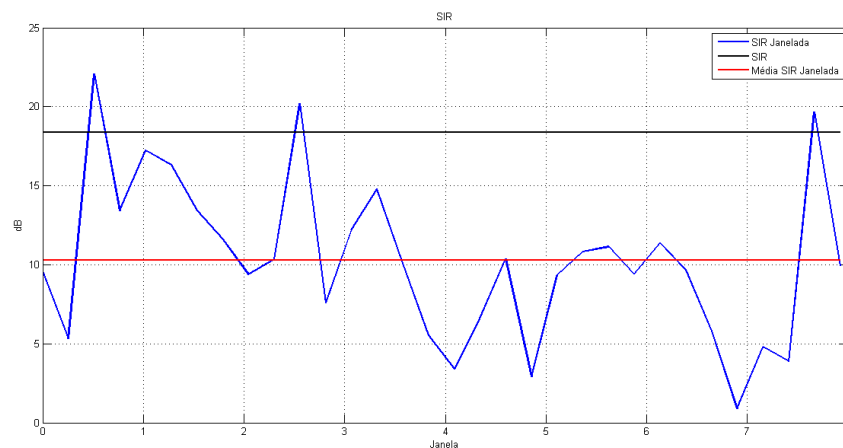
Fonte: Elaborado pelo autor

Figura 40 – Comparação entre as métricas janeladas e clássicas: Som da guitarra na mistura *Dist.wav* via NMF-KL



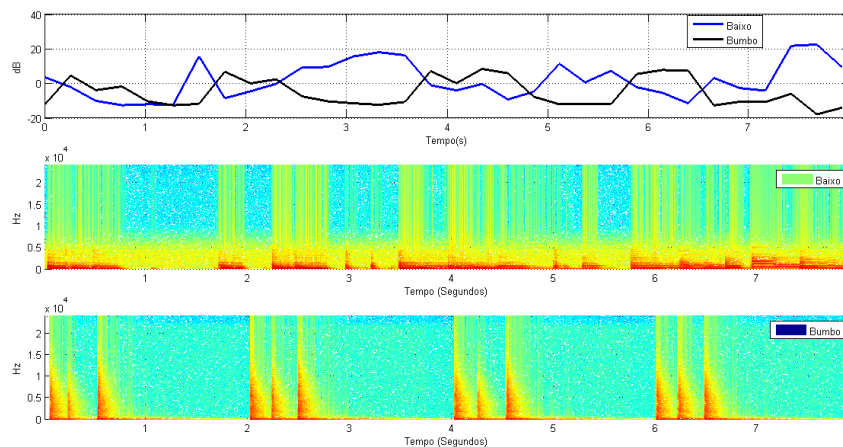
Fonte: Elaborado pelo autor

Figura 41 – Comparação entre as métricas janeladas e clássicas: Som da guitarra na mistura *Dist.wav* via NMF-IS



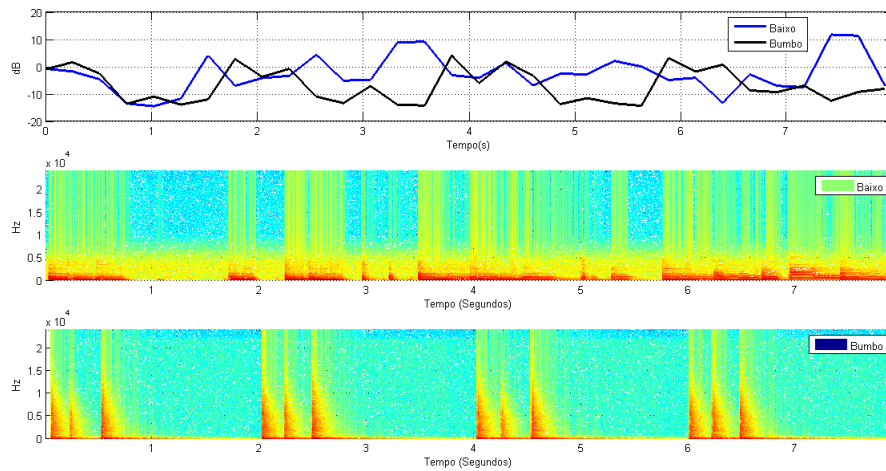
Fonte: Elaborado pelo autor

Figura 42 – Comparação das métricas SIR Janeladas para o Baixo e Bumbo, na mistura *Dist.wav* via NMF-KL



Fonte: Elaborado pelo autor

Figura 43 – Comparação das métricas SIR Janeladas para o Baixo e Bumbo, na mistura *Dist.wav* via NTF-IS



Fonte: Elaborado pelo autor

4.7 Técnica NMF Semi-supervisionado Recursiva

As métricas janeladas permitem a identificação dos intervalos de ocorrência dos valores máximos das respectivas métricas para cada fonte contida na mistura.

Como uma das contribuições deste trabalho, foi proposta uma versão semi-supervisionada da técnica NMF, realizada a partir da seleção de trechos das fontes que apresentaram melhores resultados em relação às métricas janeladas, mais especificamente, a SIR, durante a primeira fase de aplicação da versão original da NMF. Posteriormente, os trechos selecionados foram utilizados como base de treinamento para o algoritmo, implementando-se a versão que foi denominada NMF Semi-Supervisionada Recursiva. A característica recursiva foi atribuída na medida em que o sinal de treinamento é obtido da própria mistura que se deseja separar.

Vale ressaltar que, como são utilizadas informações contidas em intervalos equivalentes às janelas aplicadas às métricas, nesse caso $\approx 500\text{mS}$, esse tipo de abordagem mostra-se adequado apenas para elementos percussivos, uma vez que o perfil dos mesmos são repetitivos, enquanto os elementos harmônicos apresentam grande variação no tempo.

Foram realizados experimentos, considerando a mistura *Clean.wav* e técnica NMF para as divergências de Kullback-Leibler e Itakura-Saito, visando determinar o melhor treinamento que poderia ser obtido, para o presente caso. A Tabela 10 mostra os resultados anteriores, na versão cega indicada apenas como *NMF*, e os demais obtidos utilizando-se para como sinais de treinamento, os sons indicados como prato de condução, bumbo e caixa, além das combinações prato de condução e bumbo (*PeB*), prato de condução e caixa (*PeC*), bumbo e caixa (*BeC*), prato de condução, bumbo e caixa (*P, BeC*).

De forma geral, o treinamento das fontes garante, no mínimo uma melhoria na fonte treinada porém, tipicamente, outras fontes são melhoradas juntas. Esse fato pode

ser observado para o caso da NMF usando Divergência de Kullback-Leibler e treinamento do prato de condução, no qual observa-se aumento na SIR do baixo, prato de condução e bumbo.

Para o caso da NMF utilizando Divergência de Itakura-Saito, o treinamento do prato de condução piorou a SIR do mesmo, sendo o único caso em que esse comportamento foi observado. Os melhores resultados de separação foram obtidos usando como fontes de treinamento o prato de condução e a caixa, tanto para Kullback-Leibler quanto para Itakura-Saito; nesses casos nota-se incremento na SIR de todas as fontes. Vale mencionar que, para Itakura-Saito, o conjunto de treino prato de condução, baixo e bumbo apresentou resultados absolutos maiores que o conjunto de treino prato de condução e caixa; porém na separação do baixo se mostrou pior que o obtido sem treinamento e, por isso, não foi considerado o melhor caso.

A Tabela 11 resume os melhores resultados a partir da comparação entre a NMF não supervisionada e a Semi-Supervisionada Recursiva (utilizando-se os sinais do prato de condução e caixa para o treinamento), destacando-se em vermelho os valores onde a versão Semi-Supervisionada Recursiva se mostrou superior.

4.8 Conclusão

A aplicação da NTF, embora represente uma complexidade teórica mais elevada que a NMF, tempo de convergência mais elevado e exija do usuário um conhecimento musical mínimo mostrou melhor desempenho para a separação/restauração dos sinais. Além disso, o formato mono já não é mais o padrão e existe uma tendência crescente à utilização de formatos multicanais, o que torna a NMF uma técnica limitada no contexto de separação de sinais de áudio.

Quanto às funções de custo utilizadas, a Divergência de Itakura-Saito apresentou, no geral, valores maiores para as métricas em relação à Kullback-Leibler, confirmando o descrito na literatura. Por apresentar menor inserção de artefatos e característica de invariância à escala, conforme observado nos resultados, a Divergência de Itakura-Saito é comumente preferida nas aplicações de áudio.

A aplicação das métricas janeladas possibilitou a construção da versão semi-supervisionada recursiva, utilizando-se a informação de treinamento da própria música a ser separada. Vale observar que a utilização de parcelas do sinal a ser separado como treinamento, nesse caso, não compromete os resultados visto que não há a intenção de se inferir sobre um conjunto diferente de informações, muito pelo contrário. Sua aplicação pode ser uma alternativa em casos onde a obtenção de informação *a priori* nos métodos convencionais tornem-se indisponíveis. A técnica semi-supervisionada recursiva mostrou-se promissora com base nos resultados relativamente melhores.

Tabela 10 – Comparação NMF e NMF Semi-Supervisionada Recursiva, utilizando-se a mistura *Clean* e com treinamento via elementos percussivos.

		SDR(dB)		SIR(dB)		SAR(dB)	
	<i>Fonte</i>	<i>KL</i>	<i>IS</i>	<i>KL</i>	<i>IS</i>	<i>KL</i>	<i>IS</i>
<i>NMF</i>	Guitarra	4,076	2,139	11,731	5,680	5,176	5,716
	Baixo	-12,604	-7,008	-11,350	-4,959	5,063	3,401
	Condução	7,410	7,165	20,518	24,126	7,666	7,270
	Bumbo	-0,204	-3,294	1,081	0,014	8,215	2,441
	Caixa	-6,232	-14,825	-1,344	-11,229	-0,795	-0,785
<i>Prato</i>	Guitarra	3,737	1,477	11,379	4,180	4,863	6,224
	Baixo	-11,732	-6,772	-10,457	-4,510	5,043	2,968
	Condução	6,729	-1,212	21,274	15,021	6,917	-0,973
	Bumbo	0,135	-3,986	1,678	-0,599	7,629	1,997
	Caixa	-6,804	-19,956	-2,666	-18,563	-0,143	4,285
<i>Bumbo</i>	Guitarra	3,302	1,864	12,228	5,002	4,150	5,944
	Baixo	-12,931	-6,950	-11,690	-4,441	5,087	2,402
	Condução	6,941	4,953	16,467	15,078	7,552	5,530
	Bumbo	0,547	-0,033	2,368	7,131	7,186	1,662
	Caixa	-10,040	-13,816	-8,681	-10,010	4,903	-1,053
<i>Caixa</i>	Guitarra	4,113	2,367	11,952	5,669	5,161	6,145
	Baixo	-12,157	-6,823	-10,930	-4,539	5,197	2,908
	Condução	7,529	4,779	18,498	14,659	7,953	5,396
	Bumbo	-0,183	-2,470	1,204	0,529	7,899	3,305
	Caixa	-5,768	-12,808	-0,039	-7,514	-1,387	-3,064
<i>PeB</i>	Guitarra	3,875	1,217	10,908	3,973	5,150	5,960
	Baixo	-11,471	-7,697	-10,236	-4,520	5,221	0,986
	Condução	6,735	5,691	21,242	20,783	6,924	5,863
	Bumbo	0,521	-1,637	2,307	5,310	7,249	0,463
	Caixa	-7,056	-12,524	-3,085	-9,087	-0,011	-0,310
<i>PeC</i>	Guitarra	4,135	1,911	12,109	5,600	5,149	5,391
	Baixo	-11,789	-7,391	-10,561	-3,999	5,222	0,723
	Condução	7,823	6,470	21,580	21,271	8,041	6,649
	Bumbo	-0,037	-1,615	1,418	1,358	7,775	3,816
	Caixa	-6,040	-11,770	-0,963	-7,090	-0,905	-2,098
<i>BeC</i>	Guitarra	3,978	1,549	11,148	3,964	5,224	6,715
	Baixo	-11,553	-7,745	-10,309	-4,900	5,174	1,555
	Condução	7,877	2,186	19,090	9,170	8,272	3,652
	Bumbo	0,402	-0,698	2,125	4,992	7,325	1,823
	Caixa	-5,891	-11,751	-0,295	-6,257	-1,330	-3,132
<i>P, BeC</i>	Guitarra	3,865	-0,942	11,599	7,433	4,956	0,462
	Baixo	-11,404	-7,851	-10,158	-6,217	5,186	4,332
	Condução	7,906	6,387	21,607	21,163	8,125	6,567
	Bumbo	0,403	-1,781	2,114	5,517	7,357	0,189
	Caixa	-6,2186	-10,970	-1,342	-6,528	-0,776	-1,634

Tabela 11 – Comparação NMF e NMF Semi-Supervisionada Recursiva, com prato e caixa treinado - Clean

		SDR(dB)		SIR(dB)		SAR(dB)	
	<i>Fonte</i>	<i>KL</i>	<i>IS</i>	<i>KL</i>	<i>IS</i>	<i>KL</i>	<i>IS</i>
<i>NMF</i>	Guitarra	4,076	2,139	11,731	5,680	5,176	5,716
	Baixo	-12,604	-7,008	-11,350	-4,959	5,063	3,401
	Condução	7,410	7,165	20,518	24,126	7,666	7,270
	Bumbo	-0,204	-3,294	1,081	0,014	8,215	2,441
	Caixa	-6,232	-14,825	-1,344	-11,229	-0,795	-0,785
<i>PeC</i>	Guitarra	4,135	1,911	12,109	5,600	5,149	5,391
	Baixo	-11,789	-7,391	-10,561	-3,999	5,222	0,723
	Condução	7,823	6,470	21,580	21,271	8,041	6,649
	Bumbo	-0,037	-1,615	1,418	1,358	7,775	3,816
	Caixa	-6,040	-11,770	-0,963	-7,090	-0,905	-2,098

5 CONSIDERAÇÕES FINAIS

Com base nos resultados obtidos é possível confirmar os ganhos obtidos quando do uso da técnica NTF, em relação ao uso da NMF. Em primeiro lugar isso se deve ao fato de o formato estéreo, utilizado pela NTF, apresentar uma espécie de redundância de observações, uma vez que, a menos de amplitude das fontes, os canais L e R apresentam informações iguais das fontes. Além disso, a NTF reduz o áudio considerado em suas notas componentes, enquanto a NMF o faz para as fontes.

Em relação às divergências, avaliando-se o quesito separação, nota-se em geral, uma tendência de melhores da Itakura-Saito para elementos percussivos, enquanto Kullback-Leibler leva vantagem nos elementos harmônicos. De maneira geral é possível afirmar que Itakura-Saito leva vantagem quando as fontes possui menor potência. Avaliando-se a reconstrução dos sinais nota-se uma ligeira vantagem da Divergência de Itakura-Saito. Os resultados obtidos comprovam a melhor adequação desta, em relação a Kullback-Leibler, quando em aplicações de áudio.

As métricas utilizadas para a validação, apesar de serem clássicas na área, nem sempre representam a realidade. Nesse sentido, foi proposta uma modificação das métricas, com o intuito de avaliar a evolução temporal dos resultados. Para o caso específico considerado, o valor apresentado pela métrica janelada se mostrou mais condizente com a realidade, embora ainda se mostre elevado.

Um ponto bastante importante de se ressaltar, em relação à métrica janelada, é a possibilidade da proposição de uma nova abordagem de obtenção de informação de treino para os métodos de separação. Torna-se possível a identificação de trechos que apresentem melhores resultados e, portanto, a extração de informação de treino a partir do próprio áudio a ser separado. Essa abordagem permite a obtenção de informação de treino especificamente para os elementos percussivos.

De maneira geral, apesar das limitações de tempo de convergência da NTF, considera-se o mesmo bem sucedido. Além disso, a abordagem de treinamento proposta para a NMF pode representar uma boa alternativa também para o NTF, quando a obtenção de informação *a priori* mostrar-se complexa. Assim, considera-se que essa nova forma de construção mostrou-se bastante promissora.

Destacam-se como principais contribuições desta pesquisa:

- 1 Descrição matemática detalhada das técnicas NMF e NTF, bem como suas particularidades teóricas e práticas;
- 2 Proposição de uma modificação nas métricas, permitindo avaliação temporal da evolução dos resultados;
- 3 Proposição de uma nova abordagem para a obtenção de informação *a priori* para as versões semi-supervisionadas dos métodos de separação;

- 4 Publicação do artigo "Aplicação da Técnica de Fatoração de Matrizes Não-Negativas à Separação de Fontes Sonoras em Misturas contendo Elementos Harmônicos e Percussivos" ao EATI 2015 - 6º Encontro Anual de TI, Universidade de Santa Maria - Rio Grande do Sul, Brasil - 2015 (FONSECA et al., 2015).
- 5 Submissão do artigo "*Analysis and Application of Non-negative Matrix and Tensor Factorization Techniques for the Sound Signal Separation in Mixtures Containing Harmonic and Percussive Elements*" no periódico da AES.

Como trabalhos futuros, podem ser destacados:

- 1 Extensão da versão semi-supervisionada recursiva à técnica de separação NTF;
- 2 Avaliação do uso de ferramentas alternativas para a representação espectral, tal como a Transformada Q-constante;
- 3 Avaliação do uso de restrições aos métodos, tais como esparsidade, *smoothness*, ortogonalidade, dentre outras, visando ao incremento da qualidade;
- 4 Pesquisa das formas de implementação dos métodos em tempo real e seu impacto na qualidade da separação;
- 5 Aplicação de outros métodos de validação, nas versões objetiva e/ou subjetiva, que possibilitem maior precisão e melhor avaliação do resultados.

REFERÊNCIAS

- ARFKEN, G.; WEBER, H. **Mathematical methods for physicists**. [S.l.]: Elsevier Acad. Press, 2008. ISBN 9780120598762.
- BADER, B. W.; KOLDA, T. G. et al. **MATLAB Tensor Toolbox Version 2.6**. February 2015. Available online. Disponível em: <<http://www.sandia.gov/~tgkolda/TensorToolbox/>>.
- BALLOU, G. **Handbook for Sound Engineers**. [S.l.]: Focal, 2005. (Focal Press). ISBN 9780240807584.
- BANERJEE, A. et al. Clustering with bregman divergences. **J. Mach. Learn. Res.**, JMLR.org, v. 6, p. 1705–1749, dez. 2005. ISSN 1532-4435.
- BECH, S.; ZACHAROV, N. **Perceptual Audio Evaluation - Theory, Method and Application**. 1. ed. Chichester, Sussex do Oeste, Inglaterra: John Wiley & Sons, Ltd., 2006.
- BENETOS, E.; KOTTI, M.; KOTROPOULOS, C. Musical instrument classification using non-negative matrix factorization algorithms. In: **Circuits and Systems, 2006. ISCAS 2006. Proceedings. 2006 IEEE International Symposium on**. [S.l.: s.n.], 2006. p. 4 pp.–.
- BERRY, M. W. et al. Algorithms and applications for approximate nonnegative matrix factorization. In: **Computational Statistics and Data Analysis**. [S.l.: s.n.], 2006. p. 155–173.
- BORISENKO, A.; TARAPOV, I. **Vector and Tensor Analysis with Applications**. [S.l.]: Dover Publications, 1979. (Dover Books on Mathematics). ISBN 9780486638331.
- CARROLL, J.; CHANG, J. J. Analysis of individual differences in multidimensional scaling via an n-way generalization of ‘Eckart-Young’ decomposition. **Psychometrika**, v. 35, n. 3, p. 283–319, set. 1970.
- CHERRY, E. C.; TAYLOR, W. K. Some further experiments upon the recognition of speech, with one and, with two ears. **The Journal of the Acoustic Society of America**, v. 26, n. 5, p. 554 – 559, Julho 1954.
- CICHOCKI, A.; CRUCES, S.; AMARI, S.-i. Generalized alpha-beta divergences and their application to robust nonnegative matrix factorization. **Entropy**, v. 13, n. 1, p. 134, 2011. ISSN 1099-4300.
- CICHOCKI, A. et al. **Nonnegative matrix and tensor factorizations: applications to exploratory multiway data analysis and blind source separation**. 1. ed. Chichester, Sussex do Oeste, Inglaterra: John Wiley & Sons, Ltd, 2009. 477 p.
- COMON, P. Tensors : A brief introduction. **Signal Processing Magazine, IEEE**, v. 31, n. 3, p. 44–53, May 2014. ISSN 1053-5888.
- COOLEY, J. W.; TUKEY, J. W. An algorithm for the machine calculation of complex fourier series. **Mathematics of computation**, JSTOR, v. 19, n. 90, p. 297–301, 1965.

COVER, T. M.; THOMAS, J. A. **Elements of Information Theory**. New York, NY, USA: Wiley-Interscience, 1991. ISBN 0-471-06259-6.

DAVIS, G. et al. **The Sound Reinforcement Handbook**. [S.l.]: Hal Leonard, 1989. (Recording and Audio Technology Series). ISBN 9780881889000.

DINIZ, P.; SILVA, E. da; NETTO, S. **Processamento Digital de Sinais - 2.ed.: Projeto e Análise de Sistemas**. [S.l.]: Bookman Editora, 2014. ISBN 9788582601242.

FARNELL, A. **Designing Sound**. 1. ed. Cambridge, Massachusetts, Estados Unidos: MIT Press, 2010. ISBN 9780262014410.

FARRELL, K.; MAMMONE, R.; FLANAGAN, J. Beamforming microphone arrays for speech enhancement. In: **17th IEEE International Conference on Acoustics, Speech, and Signal Processing (XVII ICASSP'1992)**. [S.l.: s.n.], 1992. v. 1, p. 285–288.

FASTL, H. Psychoacoustic basis of sound quality evaluation and sound engineering. In: **13th International Congress on Sound and Vibration (XIII ICSV'2006)**. Viena, Austria: [s.n.], 2006.

FEVOTTE, C.; OZEROV, A. Notes on nonnegative tensor factorization of the spectrogram for audio source separation: statistical insights and towards self-clustering of the spatial cues. In: YSTAD, S. et al. (Ed.). **Proc. 7th International Symposium on Computer Music Modeling and Retrieval (CMMR)**. Malaga, Spain, 2010.: Springer, 2010. v. 6684, p. 102–115.

FITZGERALD, D. **Automatic Drum Transcription and Source Separation**. 197 p. Tese (Doutorado) — Dublin Institute of Technology, 2004.

FITZGERALD, D.; CRANITCH, M.; COYLE, E. Non-negative tensor factorisation for sound source separation. In: **IN: PROCEEDINGS OF IRISH SIGNALS AND SYSTEMS CONFERENCE**. [S.l.: s.n.], 2005.

FONSECA, W. et al. Aplicação da técnica de fatoração de matrizes não-negativas à separação de fontes sonoras em misturas contendo elementos harmônicos e percussivos. In: **EATI 2015**. [S.l.: s.n.], 2015.

GALLAGER, R. G. **Information Theory and Reliable Communication**. New York, NY, USA: John Wiley & Sons, Inc., 1968. ISBN 0471290483.

GOSWAMI, J. C.; CHAN, A. K. **Fundamentals of Wavelets: theory, algorithms, and applications**. Nova York, NY, Estados Unidos: John Wiley & Sons, Inc., 1999. ISBN 0471197483.

GUAN, N. et al. Nnmf: An optimal gradient method for nonnegative matrix factorization. **Signal Processing, IEEE Transactions on**, v. 60, n. 6, p. 2882–2898, June 2012. ISSN 1053-587X.

HARSHMAN, R. A. Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-modal factor analysis. **UCLA Working Papers in Phonetics**, v. 16, p. 84, 1970.

HAYKIN, S.; CHEN, Z. The cocktail party problem. **Neural Computation**, v. 17, n. 9, p. 1875–1902, 2005.

HITCHCOCK, F. L. The expression of a tensor or a polyadic as a sum of products. **Journal of Mathematics and Physics**, v. 6, n. 1, p. 164–189, 1927. ISSN 1467-9590. Disponível em: <<http://dx.doi.org/10.1002/sapm192761164>>.

JEON, K. M. et al. Mechanical noise suppression based on non-negative matrix factorization and multi-band spectral subtraction for digital cameras. **Consumer Electronics, IEEE Transactions on**, v. 59, n. 2, p. 296–302, May 2013.

JODER, C. et al. Real-time speech separation by semi-supervised nonnegative matrix factorization. In: **Latent Variable Analysis and Signal Separation**. [S.l.]: Springer Berlin Heidelberg, 2012, (Lecture Notes in Computer Science, v. 7191). p. 322–329.

JOLLIFFE, I. T. **Principal Component Analysis**. 2. ed. Nova York, NY, Estados Unidos: Springer-Verlag New York, Inc., 2002. 519 p.

KITAMURA, D. et al. Robust music signal separation based on supervised nonnegative matrix factorization with prevention of basis sharing. In: **Signal Processing and Information Technology (ISSPIT), 2013 IEEE International Symposium on**. [S.l.: s.n.], 2013. p. 000392–000397.

KOLDA, T. G.; BADER, B. W. Tensor decompositions and applications. **SIAM Review**, v. 51, n. 3, p. 455–500, September 2009.

KROMER, P.; PLATOS, J.; SNÁĎSEL, V. Data mining using nmf and generalized matrix inverse. In: **Intelligent Systems Design and Applications (ISDA), 2010 10th International Conference on**. [S.l.: s.n.], 2010. p. 409–414.

LEE, D. D.; SEUNG, H. S. Learning the parts of objects by non-negative matrix factorization. **NATURE - International Weekly Journal of Science**, v. 401, p. 788–791, Outubro 1999.

LIN, C.-J. On the convergence of multiplicative update algorithms for nonnegative matrix factorization. **Neural Networks, IEEE Transactions on**, v. 18, n. 6, p. 1589–1596, Nov 2007. ISSN 1045-9227.

LUENBERGER, D.; YE, Y. **Linear and Nonlinear Programming**. [S.l.]: Springer US, 2008. (International Series in Operations Research & Management Science). ISBN 9780387745022.

LUND, N. **Attention and Pattern Recognition**. [S.l.]: Taylor & Francis, 2002. (Routledge Modular Psychology). ISBN 9781134570263.

MAHMOUD, A. et al. Improving the performance of the instantaneous blind audio source separation algorithms. In: **Signal Processing and Information Technology (ISSPIT), 2009 IEEE International Symposium on**. [S.l.: s.n.], 2009. p. 519–526.

MITIANOUDIS, N. **Audio Source Separation using Independent Component Analysis**. Tese (Doutorado) — Queen Mary, University of London, Abril 2004.

MITRA, S. **Digital Signal Processing: A Computer-based Approach**. [S.l.]: McGraw-Hill, 2011. ISBN 9780073380490.

OPPENHEIM, A.; SCHAFER, R. **Discrete-time Signal Processing**. Upper Saddle River, Nova Jersey, Estados Unidos: Prentice-Hall, Inc., 2010. (Prentice-Hall signal processing series). ISBN 9780131988422.

OZEROV, A. et al. Multichannel nonnegative tensor factorization with structured constraints for user-guided audio source separation. In: **Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)**. Prague, Czech Republic: [s.n.], 2011.

PAATERO, P.; TAPPER, U. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. **Environmetrics**, John Wiley & Sons, Ltd., v. 5, n. 2, p. 111–126, 1994.

PALUS, M. On entropy rates of dynamical systems and gaussian processes. **Physics Letters A**, v. 227, p. 301–308, 1997.

PHON-AMNUAISUK, S. Transcribing bach chorales using non-negative matrix factorisation. In: **2nd International Conference on Audio, Language and Image Processing (II ICALIP'2010)**. Xangai, China: [s.n.], 2010. p. 688–693.

RILEY, K.; HOBSON, M.; BENICE, S. **Mathematical Methods for Physics and Engineering: A Comprehensive Guide**. [S.l.]: Cambridge University Press, 2006. ISBN 9781139450997.

SANTOS, M. B. **Aplicação e análise de desempenho da técnica de fatoração de matrizes não-negativas para a separação de fontes acústicas percussivas**. 74 p. Dissertação (Mestrado) — PPGEE/PUCMG, Minas Gerais, Brasil, Julho 2015.

SAWADA, H. et al. Efficient algorithms for multichannel extensions of itakura-saito nonnegative matrix factorization. In: **Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on**. [S.l.: s.n.], 2012. p. 261–264. ISSN 1520-6149.

SCHRODINGER, E. **What is Life?: The Physical Aspect of the Living Cell**. [S.l.]: The University Press, 1944. (What is Life?: The Physical Aspect of the Living Cell).

SHANNON, C. A mathematical theory of communication. **Bell System Technical Journal**, The, v. 27, n. 4, p. 623–656, Oct 1948. ISSN 0005-8580.

SIGNAL SEPARATION EVALUATION CAMPAIGN. **Signal Separation Evaluation Campaign**. 2013. <https://sisec.wiki.irisa.fr/tiki-index.php>.

SNASEL, V. et al. Matrix factorization approach for feature deduction and design of intrusion detection systems. In: **Information Assurance and Security, 2008. ISIAS 08. Fourth International Conference on**. [S.l.: s.n.], 2008. p. 172–179.

TYGEL, A. F. **Métodos de fatoração de matrizes não-negativas para separação de sinais musicais**. Dissertação (Mestrado) — COPPE/UFRJ, Rio de Janeiro, Brasil, Dezembro 2009.

VINCENT, E.; GRIBONVAL, R.; FEVOTTE, C. Performance measurement in blind audio source separation. **IEEE Transactions on Audio, Speech, and Language Processing**, v. 14, n. 4, p. 1462–1469, Julho 2006.

VINCENT, F. G. **BSS Eval - A toolbox for performance measurement in blind source separation**. 2015. <https://www.irisa.fr/metiss/bsseval>.

VITANYI, P. Information distance in multiples. **Information Theory, IEEE Transactions on**, v. 57, n. 4, p. 2451–2456, April 2011. ISSN 0018-9448.

WANG, Y.-X.; ZHANG, Y.-J. Nonnegative matrix factorization: A comprehensive review. **Knowledge and Data Engineering, IEEE Transactions on**, v. 25, n. 6, p. 1336–1353, June 2013. ISSN 1041-4347.

Web of Science. **Web of Science**. 2015. <https://webofknowledge.com>.

Yoga Electronics Co., Ltd. **Headphones, Microphones Manufacturer - Yoga Electronics Co., Ltd.** 2014. <http://www.yoga.com.tw/>.