Estimação de Sinais de Voz Esparsificados em Misturas Subparametrizadas

Giulio Guiyti Rossignolo Suzumura, Ricardo Suyama
Centro de Engenharia, Modelagem e Ciências Sociais Aplicadas (CECS)
Universidade Federal do ABC (UFABC)
Santo André-SP, Brazil
{giulio.suzumura, ricardo.suyama}@ufabc.edu.br

***Resumo*— O problema de separação cega de fontes no contexto de misturas subparametrizadas tem sido investigado por meio de abordagens que exploram diferentes características dos sinais de interesse, dentre as quais se destaca a esparsidade. No presente trabalho, comparamos o desempenho de diferentes técnicas de estimação dos sinais de áudio, no contexto de misturas subparametrizadas, considerando que os mesmos são esparsos ou tenham sido esparsificados antes do processo de mistura. Os resultados obtidos estendem análises preliminares realizadas, e indicam que este pré-processamento traz ganhos relativos para o processo de estimação dos sinais.**

***Palavras-Chave*— Análise por Componentes Esparsas, Misturas subparametrizadas, Sinais Esparsificados.**

***Abstract*— The blind source separation problem have been investigated through approaches that explore specific characteristics of the signals of interest, among which stands out the sparsity. In this work we compare the performance of different estimation methods of audio signals, in the underdetermined context, considering that they are sparse or have been sparsified before the mixing process. The results extend preliminary studies and show that this process may relatively increase the performance of the estimation process.**

***Keywords*— Independent Component Analysis, Underdetermined Mixture, Sparsified Signals.**

I. Introdução

As técnicas de Separação Cega de Fontes (*Blind Source Separation* – BSS) têm como objetivo estimar um conjunto de sinais originais (fontes) a partir de suas versões misturadas (observações) [1], explorando características específicas dos sinais de interesse. Por exemplo, a Análise por Componentes Independentes (*Independent Component Analysis* – ICA), apoia-se na hipótese de que os sinais das fontes originais são mutuamente independentes, sendo capaz de recuperar as fontes em cenários nos quais a quantidade de observações é maior ou igual ao número de fontes [2].

Quando o número de fontes a ser estimado é maior do que o número de sensores, é necessário explorar outras informações a fim de estimar os sinais, como na abordagem da Análise por Componentes Esparsas (*Sparse Component Analysis* – SCA), que se utiliza da hipótese de que as fontes são esparsas em algum domínio [3].

Mesmo que os sinais não admitam, originalmente, uma representação esparsa, em algumas aplicações é possível ter acesso aos sinais das fontes antes do processo de mistura. Neste tipo de situação é possível manipular os sinais das fontes - e.g., em gravações de áudio em estúdio, nas quais os sinais das fontes são gravados separadamente e posteriormente misturados artificialmente -, tornando-os mais esparsos, uma abordagem usualmente associada à ideia de Separação Informada de Fontes (*Informed Source Separation* – ISS) [4].

Conforme observado em trabalhos anteriores [3][5] a utilização de sinais esparsos leva a uma melhora na estimação do modelo de mistura. Entretanto, ainda é necessário avaliar o impacto do grau de esparsidade das fontes na qualidade da estimação dos sinais. Dessa forma, o objetivo do presente trabalho é comparar diferentes métodos de estimação de fontes, em cenários de mistura subparametrizada, considerando que as fontes foram esparsificadas.

Assim, o artigo foi dividido da seguinte maneira: na Seção II é feita uma revisão do problema de Separação Cega de Fontes e procedimentos de esparsificação de sinais. As Seções III e IV apresentam métodos de recuperação em sistemas subparametrizados e métodos de avaliações utilizados, respectivamente. A Seção V apresenta diversas simulações realizadas, além da aplicação de uma proposta de modificação. Finalmente, na Seção VI são colocadas as conclusões finais e algumas perspectivas para trabalhos futuros.

II. Análise por Componentes Esparsas

No problema geral de Separação Cega de Fontes buscam-se técnicas capazes de estimar sinais que foram misturados de uma maneira desconhecida. Para o caso em que o processo de mistura não apresenta memória, i.e., é *instantâneo*, e não apresenta ruído, o processo de mistura pode ser abordado por meio de um modelo onde as observações $\mathbf{x}(t)$ são compostas pela combinação linear - definidas por uma matriz de mistura $\mathbf{A}$ invariante no tempo - das fontes originais $\mathbf{s}(t)$ [6], ou seja,

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t), \tag{1}$$

onde $\mathbf{x}(t) = [x_1(t) ..., x_I(t)]^T$ é um vetor de tamanho I contendo todas observações e $\mathbf{s}(t) = [s_1(t) ..., s_J(t)]^T$ é um vetor de tamanho J contendo todas as fontes.

Uma estratégia para se obter a matriz de mistura inversa, de maneira não-supervisionada, consiste em tentar recuperar sinais que sejam mutuamente independentes [6], hipótese primária da Análise por Componentes Independentes. Esta abordagem funciona muito bem para cenários nos quais o número I de observações é igual, ou maior, ao número J

de fontes. Entretanto, em uma situação subparametrizada é necessário considerar outras informações a respeito dos sinais a serem estimados e/ou do processo de mistura.

A Análise por Componentes Esparsas se apoia na característica adicional de esparsidade das fontes, ou seja, é baseada na hipótese de que, em um domínio apropriado, as fontes assumem valores que, na maior parte do domínio, são próximos ou iguais a zero [2]. Mesmo havendo intersecção de fontes no domínio do tempo, pode ser possível realizar a separação no domínio tempo-frequência, pois os sinais de áudio são mais esparsos no domínio tempo-frequência do que no domínio do tempo [2, 7].

Considerando o modelo apresentado em (1), e observando que a Transformada de Fourier de Tempo Curto (*Short Time Fourier Transform* – STFT) é uma operação linear, obtemos

$$\mathbf{x}(t,f) = \mathbf{A}\mathbf{s}(t,f) \quad , \tag{2}$$

onde $\mathbf{x}(t,f)$ e $\mathbf{s}(t,f)$ são vetores de coeficientes complexos das observações e fontes, respectivamente. Assim, identificamos os coeficientes da STFT por meio de dois índices, um para referenciar a janela temporal e um para os coeficientes da transformada discreta de Fourier.

A Figura 1 mostra um exemplo de um cenário com misturas de 4 sinais de voz e apenas 2 sensores[1], ilustrando a distribuição dos dados tanto no domínio do tempo quanto no domínio tempo-frequência dado através da STFT. Pode-se perceber uma distinção entre a distribuição dos dados nos dois domínios, sendo que no domínio tempo-frequência nota-se certa estrutura subjacente aos dados - que está diretamente relacionada à matriz de mistura.

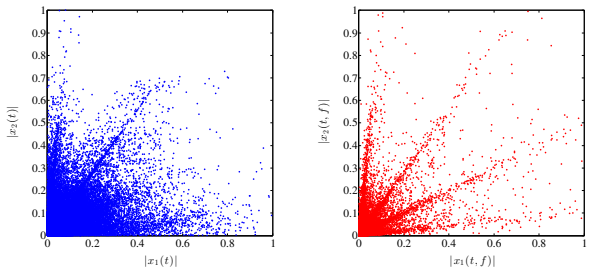

Fig. 1. Dispersão dos coeficientes de quatro sinais de voz em duas misturas no domínio do tempo (azul) e no domínio tempo-frequência (vermelho)

De forma mais objetiva é possível comparar a esparsidade das representações por meio de histogramas dos dados em cada domínio. Através da Figura 2 é possível verificar que a representação no domínio tempo-frequência é mais esparsa do que no domínio do tempo, pois apresenta uma maior concentração de valores próximos de zero e apenas uma pequena fração das amostras possui valores mais altos.

A. *Métodos de Esparsificação*

Em algumas aplicações, quando as fontes originais não são suficientemente esparsas para permitir boas estimativas na separação, é possível realizar um tipo de pré-processamento específico sobre os dados a fim de torná-los mais esparsos em algum domínio. A questão, no entanto, recai sobre o tipo de alteração que será efetuada nos sinais originais, a fim de não introduzir distorções adicionais às fontes e preservar as informações contidas nos sinais. Estas modificações nos sinais originais, por sua vez, se refletem nas observações (misturas das fontes), de maneira que os próprios sinais misturados apresentarão um maior grau de esparsidade.

[1]Nos gráficos de dispersão utilizados, cada eixo do plano bi-dimensional representa um sensor i, e os pontos o módulo dos coeficientes normalizados das fontes misturadas.

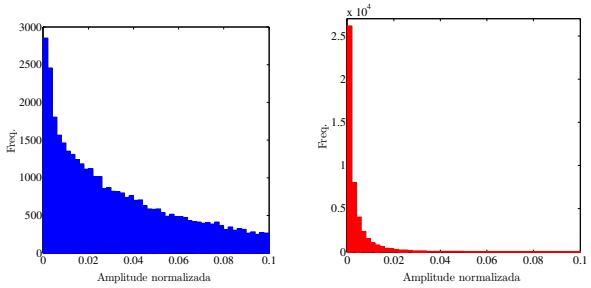

Fig. 2. Histograma da média dos coeficientes da Figura 1 no domínio do tempo (azul) e no domínio tempo-frequência (vermelho)

Algumas abordagens de Separação Informada de Fontes realizam a manipulação das fontes com intuito de torná-las mais esparsas para obter maior facilidade de separação, onde três técnicas se destacam: *Filtro de Irrelevância*, *Doping Watermarking* e *Perceptual Doping Watermarking*.

1) Filtro de Irrelevância: Desenvolvida por Pinel [7], esta técnica de modificação tenta preservar a qualidade do sinal original utilizando o Modelo de Assimilação Perceptiva *(Perceptual Assimilation Model* – PAM) que realiza o mascaramento simultâneo e, a partir de um valor limiar, filtra componentes irrelevantes. Além disso, este filtro aplica a Transformada Discreta de Cosseno Modificada (*Modified Discrete Cosine Transform* – MDCT), a qual é robusta em relação aos efeitos de borda na etapa de reconstrução do sinal, pela sua inversa.

2) Doping Watermarking: Proposto por Mahé [5], este método é baseado no conhecimento da Função Densidade de Probabilidade (*Probability Density Function* – PDF) dos coeficientes tempo-frequência das fontes de áudio, e que pode ser remodelada, aproximadamente, por uma Distribuição da Gaussiana Generalizada (*Generalized Gaussian Distribution* – GGD) com fatores de forma específicos. Portanto, os sinais originais de áudio, que tem pouca característica esparsa no domínio tempo-frequência, podem ser processados de forma a atribuir um fator de forma mais esparso em relação ao original. Dessa forma, o método de esparsificação por *Doping Watermarking* procura "dopar" o fator de forma original modificando-o para um fator de forma alvo.

3) Perceptual Doping Watermarking: A fim de se conseguir um compromisso entre desempenho na separação e degradação da qualidade perceptiva dos sinais de áudio no processo de dopagem, esta nova proposta de Mahé [4] apresenta um método no qual a qualidade sonora é avaliada por meio da Distorção Espectral de Bark (*Bark Spectral Distortion* – BSD), controlando assim o nível de modificação realizada nos sinais sobre o método anterior de *Doping Watermarking*.

III. Recuperação das Fontes

Foram avaliados três diferentes algoritmos: o primeiro, denominado *FASST* (`FST`), proposto por Ozerov [6] e utilizado por Mahé [4] na obtenção dos sinais recuperados; o segundo método avaliado é a *Minimização da Norma* ℓ_p (`NLP`), proposta por Vincent [8]; o último algoritmo avaliado é a *Máscara Binária Ideal* (`BIN`), descrita por Vincent [9], que tem como intuito principal servir como parâmetro de comparação para os demais métodos.

1) Algoritmo FASST: Este método explora a hipótese básica de que as distribuições dos coeficientes tempo-frequência dos sinais podem ser bem modelados por meio de um modelo "localmente Gaussiano", i.e., ao menos em uma certa região do domínio tempo-frequência, os coeficientes podem ser descritos com relativa acuidade por meio de uma distribuição Gaussiana. É proposta uma modificação sobre a transformação padrão STFT para a MDCT na Seção V-D. – por possuir propriedades excelentes de esparsificação para sinais de áudio e de voz [7].

2) Minimização da Norma ℓ_p*:* O segundo método de separação estudado se baseia na hipótese de que os coeficientes associados às fontes são independentes e esparsamente distribuídos. Além disso, utiliza a distribuição Gaussiana Generalizada para modelar a característica esparsa dos coeficientes. Diferentemente do *FASST*, este método não estima parâmetros das fontes e da matriz de mistura, levando em consideração que esta última já é conhecida ou pode ser estimada. Com estes dados as fontes são estimadas a partir do critério de *Maximum a posteriori* – MAP [8].

3) Máscara Binária Ideal: Estudada por Vincent [9], este método foi escolhido como base de comparação. Cabe ressaltar, no entanto, que a utilização deste estimador "oráculo" é restrita a uma avaliação em que os sinais originais são disponibilizados. Resumidamente, o algoritmo cria uma máscara no domínio tempo-frequência, com intervalos definidos no tempo e nas faixa de frequência, e aplica esta máscara nas misturas x_i a partir da comparação com as fontes s_j.

IV. Critério de Avaliação dos Métodos

Uma família de critérios utilizada em pesquisas relacionadas à separação de fontes foi proposta por Vincent [10] - posteriormente corrigida por Emiya [11] -, consiste em medir as razões de energia dos sinais em distorção.

Embora as métricas objetivas forneçam uma base de comparação para os diferentes métodos, as avaliações de sinais de voz dadas por razões de energia não estão diretamente ligadas à interpretação ao estímulo sensorial humano. Nesse sentido, critérios perceptuais foram promovidos por Emiya [11], através da Medida de Similaridade Perceptual (*Perceptual Similarity Measure* – PSM), modelo auditivo PEMO-Q e um mapeamento não linear.

Inicialmente foram utilizadas duas métricas para avaliação dos sinais, uma objetiva - a SDR (*Signal to Distortion Ratio*) em decibel (dB) - e uma perceptual - a OPS (*Overall Perceptual Score*) que atribui uma nota entre 0 e 100 (pior para melhor).

V. Simulações

No conjunto de simulações realizadas foram utilizados sinais do banco *TIMIT Corpus Sample* [12], que compreende diversos bancos de dados em seu conteúdo. O banco específico utilizado – denominado no presente trabalho como banco `original`, é composto por 96 sinais de voz com duração de 5 segundos cada e amostrados a 16kHz.

Além do banco de dados com fontes sem modificações, foram utilizados os outros três bancos de dados modificados pelas abordagens de esparsificação mencionadas na seção II-A. Os bancos processados foram nomeados da seguinte forma:

- banco `pinel`: sinais modificados pelo *Filtro de Irrelevância* e fornecidos por Pinel [7];
- banco `gael`: sinais modificados pelo *Doping Watermarking* e fornecidos por Mahé [5];
- banco `gael2`: sinais modificados pelo *Perceptual Doping Watermarking* e fornecidos por Mahé [4].

A fim de avaliar a recuperação de fontes dos algoritmos estudados (i.e., `FST`, `NLP` e `BIN`), sobre misturas de fontes de todos os bancos (i.e., `original`, `pinel`, `gael` e `gael2`), foram realizadas simulações considerando diferentes cenários sintéticos, descritos na sequência.

A. *Cenários com misturas* ideais

No primeiro conjunto de simulações considerou-se um número variado de fontes do banco `original`, emulando uma situação na qual as fontes apresentam-se igualmente espaçadas. Dessa forma, pode-se considerar que se trata de um cenário favorável para a estimação das fontes, mesmo em uma situação de mistura subparametrizada.

Para avaliar os métodos de separação estudados serão utilizadas as métricas de estudo a SDR e a OPS. Através da Figura 3 é possível observar que a $\overline{\text{SDR}}$ e $\overline{\text{OPS}}$[2,3] atribuídas aos métodos `FST` são menores do que as atribuídas ao método `NLP` para todos os sistemas subparametrizados.

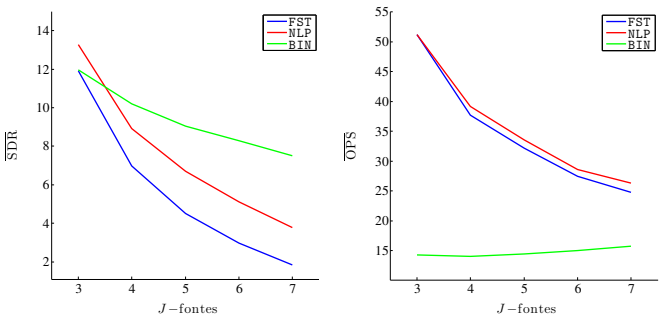

Fig. 3. $\overline{\text{SDR}}$ e $\overline{\text{OPS}}$ das fontes estimadas, quando utilizadas fontes do banco `original`, em cenários com o número de fontes variando entre 3 e 7.

Também se pode verificar que, na média, o melhor resultado objetivo é apresentado pelo método `BIN`, o que pode ser explicado por este levar em consideração um fator limiar

[2] $\overline{\text{SDR}}$ e $\overline{\text{OPS}}$ indicam as médias de 50 simulações realizadas

[3] A análise subjetiva da percepção auditiva dos autores considerou que experimentos com $J > 7$ fontes não trariam informações relevantes aos resultados obtidos.

como critério de SNR local, e inversamente o pior resultado perceptual, causada por dar ao som um aspecto "metálico".

Embora os resultados providos pela métrica objetiva sejam válidos para uma comparação de similaridade em termos da forma de onda dos sinais recuperados, elas podem não avaliar a capacidade de interpretação do sinal para o usuário. Nesse sentido, a pontuação perceptual tende a ser mais adequada em avaliações com sinais de voz.

Para se avaliar a hipótese estudada por Mahe [4], referente ao benefício da esparsificação, os mesmos experimentos foram realizados para os outros bancos. Além disso, simulações realizadas com muitas fontes ativas podem ser encaradas como limites inferiores de desempenho, fazendo-se necessário realizar uma análise criteriosa em cenários mais relacionados com o mundo real, como a simultaneidade de 3 e 4 fontes.

Através de uma análise perceptual, duas observações pertinentes podem ser realizadas sobre a Tabela I: a primeira se refere a superioridade do algoritmo `NLP` frente ao `FST`; a segunda se dá pela melhor estimativa dada por `gael2` para 3 fontes, e `original` para 4 fontes.

TABELA I
RESULTADOS DE $\overline{\text{OPS}}$ PARA CENÁRIOS IDEAIS DE 3 E 4 FONTES

Cenário	Método	original	pinel	gael	gael2
Ideal J=3	FST	51.22	51.06	48.08	50.43
	NLP	51.24	51.84	50.41	52.38
Ideal J=4	FST	37.72	37.09	35.59	36.62
	NLP	39.19	38.85	37.91	38.42

B. Cenários com misturas reais

A fim de realizar uma avaliação em uma situação mais próxima da encontrada em aplicações reais, foram considerados dois cenários propostos nas campanhas de avaliação SiSEC[4], na qual se propoem cenários com 3 e 4 fontes em 2 sensores.

É possível verificar na Tabela II que o melhor algoritmo separador é o `FST`, situação inversa à observada nos resultados dos cenários ideais. Além disso, a utilização de sinais esparsificados obteve relativamente melhores resultados, comparado as fontes originais.

TABELA II
$\overline{\text{OPS}}$ DOS SINAIS ESTIMADOS PELOS MÉTODOS `FST` E `NLP`, PARA UM CENÁRIO REAL DE 3 FONTES COM O USO DE DIFERENTES BANCOS

Cenário	Método	original	pinel	gael	gael2
Real J=3	FST	47.75	48.41	47.84	49.19
	NLP	44.85	46.84	47.32	48.50
Real J=4	FST	37.20	37.40	36.86	37.44
	NLP	34.50	35.03	36.59	36.92

Para os dois cenários reais estudados, o método de estimação `FST` obteve resultado superior, por outro lado, os cenários ideais mostraram-se melhor estimados pelo algoritmo `NLP`. Logo, a estimação ótima das fontes é dependente do processo de mistura.

[4]Para mais detalhes sobre os sistemas de misturas utilizados, os leitores são convidados a acessarem o domínio do SiSEC2016 em `https://sisec.inria.fr/home/2016-underdetermined-speech-and-music-mixtures/`

C. Estimação dos coeficientes da mistura

Todos os resultados prévios foram obtidos a partir da hipótese da estimação perfeita do processo de mistura, porém se a estimação do processo não for precisa, é necessário descobrir se há discrepância entre os resultados dos métodos `FST` e `NLP`.

A Tabela III apresenta a degradação do valor da OPS obtida no cenário real com mistura estimada - com erro de estimação de $\pm 5\%$ no cenário de mistura com 3 fontes - em relação aos valores obtidos com o conhecimento informado (Tabela II).

TABELA III
VARIAÇÃO DE $\overline{\text{OPS}}$ ENTRE O PROCESSO DE SEPARAÇÃO COM MISTURA REAL CONHECIDA E ESTIMADA (QUANTO MENOR MELHOR)

Cenário	Método	original	pinel	gael	gael2
J=3	FST	9.06	4.85	4.12	5.52
	NLP	3.32	4.47	3.87	4.76

É possível verificar que a abordagem de sinais esparsos no método de estimação, promovido pelo `NLP`, proporciona mais robustez ao sistema quando este atua em um contexto no qual a matriz de separação não foi acuradamente estimada. Além disso, a superioridade dos bancos esparsificados corrobora a ideia de que sinais mais esparsos facilitam a separação em situações não ideais.

D. Aplicação da proposta de modificação

Com o intuito de se obter melhores avaliações perceptuais, foi proposto alterar o domínio transformado da STFT para a MDCT, o que pode trazer benefícios em relação à estimação dos sinais esparsificados, pelo fato de que esta é uma transformada de decomposição/reconstrução mais eficiente [7], assegurando maior qualidade dos sinais esparsificados reconstruídos.

Após todas as devidas alterações, as simulações realizadas com os cenários reais, que resultaram dados da Tabela II, foram refeitas, fornecendo os resultados da Tabela IV.

TABELA IV
$\overline{\text{OPS}}$ DOS SINAIS ESTIMADOS PELOS MÉTODOS `FST` E `NLP`, PARA OS CENÁRIOS REAIS COM O USO DA MDCT

Cenário	Método	original	pinel	gael	gael2
Real J=3	FST	45.66	47.25	45.54	47.57
	NLP	49.18	49.70	47.48	49.94
Real J=4	FST	35.82	35.85	34.81	35.29
	NLP	37.18	37.12	36.36	37.75

É possível discorrer que a mudança da transformada atua positivamente para o método `NLP`, pois a MDCT leva a um grau de esparsidade maior do que a STFT. No entanto, o mesmo não ocorre para o método `FST`, e provavelmente isso se

deva à hipótese básica explorada neste algoritmo que considera um modelo localmente gaussiano para os coeficientes das fontes.

E. Análise de Eficiência

O tempo de processamento de cada simulação[5] de estimação dos métodos `FST` e `NLP`, realizadas a partir dos cenários reais, são apresentados na Tabela V[6].

TABELA V
MÉDIA DO TEMPO DE PROCESSAMENTO GASTO PELOS MÉTODOS DE SEPARAÇÃO PARA OS CENÁRIOS REAIS AVALIADOS

Cenário	Transf.	Método	Tempo (s)
J=3	STFT	`FST`	56.18
		`NLP`	0.14
	MDCT	`FST`	52.58
		`NLP`	0.12
J=4	STFT	`FST`	70.95
		`NLP`	0.20
	MDCT	`FST`	59.75
		`NLP`	0.17

Os resultados apresentados pela TabelaV, mostram que as estimações realizadas pelo algoritmo `NLP` são, no mínimo, 400 vezes mais rápidas comparadas ao `FST`, para o cenário de 3 fontes, e 350 vezes mais rápido para o cenário de 4 fontes, algo que torna o algoritmo `NLP` interessante para trabalhos em tempo real. Outro resultado, não tão expressivo, se dá pela diferença entre as transformadas utilizadas, com vantagem da MDCT frente a STFT.

VI. CONCLUSÕES E PERSPECTIVAS

Partindo do princípio que o interesse em aplicações de áudio é obter sinais com boa qualidade sonora - algo, portanto, relacionado a uma avaliação subjetiva do sinal - é importante perceber que, embora as métricas objetivas tenham sido utilizadas para comparação dos métodos de separação em geral, a análise perceptiva dos sinais pode trazer informações bastante ricas a respeito da qualidade dos sinais recuperados.

A separação realizada pela *Máscara Binária Ideal* retorna o melhor resultado objetivo - como esperado, pois visa maximizar uma medida objetiva -, porém uma diferença positiva na qualidade dos sinais recuperados, detectada na análise feita com o índice OPS, é dada pelos outros dois métodos.

Inicialmente o método *FASST* obteve melhores resultados em cenários com menor espaçamento angular entre as fontes (misturas reais), porém, o efeito observado com a utilização da nova representação MDCT, fez com que a *Minimização da Norma* ℓ_p obtivesse um desempenho ligeiramente superior frente ao *FASST* (0,5 ponto no OPS). Além disso, em termos do custo computacional, verificou-se que a abordagem baseada na norma ℓ_p se destaca, tornando o algoritmo bastante atrativo para aplicação em sinais de áudio esparsificados e misturas convolutivas.

Estudos relacionados ao ganho obtido com a MDCT no algoritmo de estimação baseada na norma ℓ_p e novas propostas de métodos para esparsificação de sinais são alvos de pesquisas futuras, para fornecer uma solução completa para o problema de separação de sinais.

[5]O processamento foi realizado em uma máquina HP Z210 Workstation Intel® Xeon® CPU E3-1270 @ 3.40GHz com 16GB de memória RAM e Sistema Operacional Windows®.

[6]O custo computacional da etapa de mistura, suposta estimação da mistura e qualificação, não estão inclusos nos tempos apresentados.

REFERÊNCIAS

[1]Comon, Pierre e Jutten, Christian. *Handbook of Blind Source Separation: Independent component analysis and applications*. Academic press, 2010.

[2]Bofill, Pau e Zibulevsky, Michael. "Underdetermined blind source separation using sparse representations". Em: *Signal processing* 81.11 (2001), pp. 2353–2362.

[3]Nadalin, Everton Z, Suyama, Ricardo e Faissol Attux, Romis de. "An ICA-Based Method for Blind Source Separation in Sparse Domains." Em: *ICA*. Springer. 2009, pp. 597–604.

[4]Mahé, Gaël et al. "Perceptually controlled doping for audio source separation". Em: *EURASIP Journal on Advances in Signal Processing* 2014.1 (2014), pp. 1–14.

[5]Mahé, Gaël, Nadalin, Everton Z e Romano, Joao-Marcos T. "Doping audio signals for source separation". Em: *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European*. IEEE. 2012, pp. 2402–2406.

[6]Ozerov, Alexey, Vincent, Emmanuel e Bimbot, Frédéric. "A general flexible framework for the handling of prior information in audio source separation". Em: *Audio, Speech, and Language Processing, IEEE Transactions on* 20.4 (2012), pp. 1118–1133.

[7]Pinel, Jonathan e Girin, Laurent. ""Sparsification"of Audio Signals Using the MDCT/IntMDCT and a Psychoacoustic Model—Application to Informed Audio Source Separation". Em: *Audio Engineering Society Conference: 42nd International Conference: Semantic Audio*. Audio Engineering Society. 2011.

[8]Vincent, Emmanuel. "Complex nonconvex l p norm minimization for underdetermined source separation". Em: *Independent Component Analysis and Signal Separation*. Springer, 2007, pp. 430–437.

[9]Vincent, Emmanuel, Gribonval, Rémi e Plumbley, Mark D. "Oracle estimators for the benchmarking of source separation algorithms". Em: *Signal Processing* 87.8 (2007), pp. 1933–1950.

[10]Vincent, Emmanuel, Gribonval, Rémi e Févotte, Cédric. "Performance measurement in blind audio source separation". Em: *Audio, Speech, and Language Processing, IEEE Transactions on* 14.4 (2006), pp. 1462–1469.

[11]Emiya, Valentin et al. "Subjective and objective quality assessment of audio source separation". Em: *Audio, Speech, and Language Processing, IEEE Transactions on* 19.7 (2011), pp. 2046–2057.

[12]Garofolo, John S. "DARPA TIMIT acoustic-phonetic speech database". Em: *National Institute of Standards and Technology (NIST)* 15 (1988), pp. 29–50.