

EMPREGO DE DICIONÁRIOS REDUNDANTES E DE MINIMIZAÇÃO DA INFORMAÇÃO MÚTUA PARA A SEPARAÇÃO CEGA DE FONTES EM AMBIENTES ANECOICOS COM ATRASOS FRACIONÁRIOS

Diego Barreto Haddad

Coordenação de Telecomunicações (CEFET-RJ) / PADS-PEE (UFRJ)

diego@pads.ufrj.br

Lisandro Lovisolo

PROSAICO - Departamento de Engenharia Eletrônica e de Telecomunicações (UERJ)

lisandro@uerj.br

Jorge Costa Pires Filho

Instituto de Pesquisas da Marinha (IPqM) / PADS-PEE (UFRJ)

jpires@ipqm.mar.mil.br

Mariane Rembold Petraglia

PADS-PEE (UFRJ)

mariane@pads.ufrj.br

Resumo – O presente artigo emprega dicionários redundantes multicanais e o princípio da minimização da informação mútua entre as estimativas para resolver o problema de separação cega de fontes, quando estas são misturadas num ambiente anecoico. Parte significativa das técnicas propostas na literatura para a resolução deste problema restringe-se a atrasos inteiros (múltiplos do período de amostragem) dos sinais. Nossa abordagem não compartilha desta restrição. Como inovação, ao contrário da quase totalidade dos métodos de separação de fontes que utilizam um princípio para a separação (independência, esparsidade ou diversidade espectral), nossa proposta conjuga a hipótese da independência entre as fontes (para a estimativa dos atrasos) com a da esparsidade entre as mesmas (para a estimativa dos coeficientes de atenuação). Dentre outras avaliações, verificamos que a técnica proposta pode atuar como uma etapa de pré-processamento para um sistema automático de reconhecimento de instrumentos musicais.

Palavras-chave – Matching pursuit, separação cega de fontes, atrasos fracionários, reconhecimento de padrões.

Abstract – This paper uses multichannel overcomplete dictionaries and the mutual information minimization principle between estimates to solve the blind source separation problem, in the anechoic mixture configuration. Most of the techniques proposed to solve this problem suppose that there exist only integer delays (multiples of the sampling period) among the signals. Our approach does not have this constraint. As novelty, differing from almost all source separation methods that use only one principle for separation (independence, sparsity or spectral diversity), our approach conjugates the independence between sources hypothesis (for delays estimates) with the sparsity hypothesis (for attenuation coefficients estimates). Among other evaluations, we verify that the proposed technique can be a pre-processing step for a musical instruments automatic recognizing system.

Keywords – Matching pursuit, blind source separation, fractional delays, pattern recognition.

1. INTRODUÇÃO¹

O problema da separação cega de fontes (BSS, do inglês *Blind Source Separation*) consiste em recuperar N sinais originais (também denominados de fontes) a partir do conhecimento de M misturas destes. Num ambiente onde coexistem diversas fontes sonoras, o ser humano se desincumbe facilmente desta tarefa, ao concentrar sua atenção em apenas uma das fontes, descartando as demais como interferências não passíveis de processamento de alto nível. O fato de que indivíduos com desordens no sistema auditivo central possam apresentar dificuldades de compreensão de uma fala num ambiente ruidoso ilustra a complexidade da resolução deste problema. Desta nossa habilidade advém o outro nome - *cocktail-party* [1] - atribuído ao problema de BSS. Embora as técnicas de separação cega de fontes em contextos genéricos, dada a pouca informação disponível, possam parecer à primeira vista um problema impossível para um sistema artificial [2], atualmente consistem num ramo maduro de pesquisa,

¹Este trabalho foi parcialmente financiado pelo CNPq.

sendo passíveis de inúmeras aplicações. Dentre estas, podemos elencar exemplos nas áreas de sensoriamento remoto [3], análise de sinais biomédicos [4] e sísmicos [5], bem como em comunicações digitais [6].

Para as configurações de misturas instantâneas, lineares, sem ruído e nas quais $M \geq N$, o problema de BSS pode ser considerado resolvido [1], porém em geral os atrasos entre sensores não podem ser negligenciados. Neste artigo, apresentamos uma técnica de BSS que contempla as configurações anecoicas com atrasos fracionários (em outras palavras, os atrasos entre sensores não são necessariamente múltiplos do período de amostragem).

O artigo está estruturado no seguinte modo: na Seção 2, modelamos o processo de mistura anecoica, bem como as simplificações decorrentes das ambiguidades inerentes ao processo de separação cega. Já a Seção 3 apresenta o algoritmo *Matching Pursuit* original, a nossa proposta que emprega dicionários redundantes multicanais e como podemos, a partir de nossa abordagem, estimar os coeficientes de atenuação do processo de mistura. A Seção 4 contempla o processo de estimativa dos atrasos via minimização da informação mútua, enquanto que as Seções 5 e 6 apresentam, respectivamente, as simulações e as conclusões do artigo.

2. O PROCESSO DE MISTURA ANECOICA

Seja $\tilde{s}_i(n)$ a n -ésima amostra da i -ésima fonte, para $1 \leq i \leq N$. As misturas x_j , $j \leq 1 \leq M$ podem ser expressas, num contexto anecoico, em função das fontes como $x_j(n) = \sum_{i=1}^N \tilde{h}_{ji} \tilde{s}_i(n - \tilde{\Delta}_{ji})$, onde \tilde{h}_{ji} e $\tilde{\Delta}_{ji}$ são, respectivamente, o coeficiente de atenuação e o atraso existentes no caminho da i -ésima fonte para o j -ésimo sensor. O processo de mistura pode ser resumido por uma matriz $\tilde{\mathbf{H}}$ de dimensões $M \times N$, cujos elementos são os filtros $\tilde{\mathbf{h}}_{ij} = \tilde{h}_{ij} z^{-\tilde{\Delta}_{ij}}$, de modo que podemos escrever $\mathbf{X} = \tilde{\mathbf{H}}\tilde{\mathbf{S}}$, com as fontes e as misturas, no domínio da transformada Z , dispostas nas N linhas da matriz $\tilde{\mathbf{S}}$ e nas M linhas da matriz \mathbf{X} , respectivamente.

É conhecido que as ambiguidades de filtragem (no contexto de misturas instantâneas, a ambiguidade é apenas de escalamento; quando os sinais são complexos, surge também uma ambiguidade de fase) e permutação, oriundas da ausência de informações concernentes às fontes [1], permitem-nos considerar bem-sucedida uma separação mesmo quando as estimativas são versões filtradas e possivelmente permutadas das fontes. Essa é a principal diferença entre as técnicas de BSS e de desconvolução, já que a última almeja preservar a estrutura temporal das fontes. Por isso, seja $s_i(n)$ a referência da i -ésima fonte a recuperar, a qual expressamos em função da fonte original $\tilde{s}_i(n)$, como $s_i(n) = \frac{\tilde{s}_i(n + \tilde{\Delta}_{1i})}{h_{1i}}$.

Podemos expressar então o processo de mistura como $\mathbf{X} = \mathbf{H}\mathbf{S}$, onde a matriz \mathbf{H} pode ser obtida da matriz $\tilde{\mathbf{H}}$, fazendo-se $\mathbf{h}_{ij} = \frac{\tilde{\mathbf{h}}_{ij}}{\tilde{\mathbf{h}}_{1j} z^{\tilde{\Delta}_{1j}}}$. Deste modo, doravante consideraremos, sem perda de generalidade para o problema de BSS, que a matriz \mathbf{H} de dimensões $M \times N$ que condensa o processo de mistura é:

$$\mathbf{H} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ h_{21} z^{-\Delta_{21}} & h_{22} z^{-\Delta_{22}} & \ddots & h_{2N} z^{-\Delta_{2N}} \\ \vdots & \ddots & \ddots & \vdots \\ h_{M1} z^{-\Delta_{M1}} & h_{M2} z^{-\Delta_{M2}} & \cdots & h_{MN} z^{-\Delta_{MN}} \end{bmatrix}, \quad (1)$$

com $\Delta_{ij} = \tilde{\Delta}_{ij} - \tilde{\Delta}_{1j}$. Importa ressaltar que, a partir de agora, $s_i(n)$ será a n -ésima amostra da i -ésima fonte. Um procedimento bastante semelhante de normalização da primeira linha da matriz de misturas costuma ser utilizado nas técnicas de análise de componentes esparsos. A principal diferença por nós adotada em relação a estes métodos reside no fato de contemplarmos o caso anecoico, o que implica a necessidade de incluir a ambiguidade de atraso em nossa análise.

As técnicas de BSS podem efetuar a busca de sua solução no espaço de mistura (almejando identificar o sistema de mistura) ou no espaço de separação (quando tentam otimizar parâmetros de um sistema de separação que resulte em estimativas fidedignas das fontes). Tipicamente, técnicas de Análise de Componentes Independentes [1] (ICA, do inglês *Independent Component Analysis*) atuam no espaço de separação, enquanto que técnicas de Análise de Componentes Esparsos [7] (SCA, do inglês *Sparse Component Analysis*) se articulam para identificar (a menos das ambiguidades supracitadas) o sistema de mistura.

O método proposto neste artigo reúne ambas as abordagens: na etapa I (estimativa dos coeficientes de atenuação), efetua-se uma busca no espaço de mistura, por meio de uma técnica de SCA fundamentada em dicionários redundantes multicanais, numa nova formulação do algoritmo MP [8] (do inglês *Matching Pursuit*). Já a etapa II (estimativa dos atrasos) efetua uma busca no espaço de separação, selecionando as estimativas dos atrasos Δ_{ij} por meio do recurso à informação mútua das estimativas das fontes como função custo. O processo de otimização da etapa II é dependente das estimativas obtidas na etapa I.

A relação entre o número de fontes e o de misturas tem grande efeito na escolha do algoritmo de BSS a empregar. As configurações $N > M$, $N = M$ e $N < M$ são denominadas, respectivamente, de configurações sobredeterminadas, determinadas e subdeterminadas (sendo estas últimas as mais desafiadoras de todas, por possuírem menor quantidade de informação disponível). O caso $M = 1$ (monocanal) deve ser destacado, dentre as configurações subdeterminadas, como extremamente difícil e por demandar técnicas diferenciadas. O foco deste artigo reside nas configurações determinadas.

As técnicas de SCA podem, em geral, atuar em configurações subdeterminadas, sendo por isso consideradas mais poderosas do que as técnicas de ICA [9]. Como a etapa I emprega um algoritmo de SCA, ela apenas exige que a condição $M \geq 2$ seja satisfeita (o caso monocanal não apresenta a diversidade espacial necessária para as técnicas de SCA nas quais estamos interessados). Porém a etapa II efetua uma otimização no espaço de separação à semelhança das técnicas de ICA e, a menos que nela efetue-

adaptações, compartilha com elas a restrição de que o número de fontes não pode ultrapassar o número de misturas.

3. ESTIMATIVA DOS COEFICIENTES DE ATENUAÇÃO VIA MP MULTICANAL

No contexto de decomposição de sinais utilizado, um dicionário redundante (cujo m -ésimo átomo é denotado por ϕ_m) é usado para a descrição esparsa de sinais [8]. Estando os átomos do dicionário bem localizados no espaço tempo-frequência, podemos empregar a diversidade temporal-espectral normalmente existente entre as diferentes fontes e supor que um átomo que descreva razoavelmente bem a mistura está associado, na maior parte das vezes, a apenas uma fonte. Na etapa de reconstrução, podemos determinar qual fonte está associada a um determinado átomo empregando a diversidade espacial, já que as fontes alcançam sensores distintos em instantes diferentes e com atenuações distintas. Porém, neste artigo a etapa de reconstrução não utiliza a decomposição em átomos.

O algoritmo MP padrão para a decomposição de um sinal monocanal \mathbf{x}_j como uma soma ponderada de J átomos de um dicionário redundante monocanal \mathcal{D} é apresentado a seguir [8]:

Algoritmo MP Padrão
Inicialização
 $\mathbf{r}^0 \leftarrow \mathbf{x}_j$
Laço Principal
Para $i = 1, 2, \dots, J$
 Escolha $\phi_{m_i} := \arg \max_{\phi \in \mathcal{D}} |\langle \mathbf{r}^{i-1}, \phi \rangle|$
 $\mathbf{r}^i \leftarrow \mathbf{r}^{i-1} - \langle \mathbf{r}^{i-1}, \phi_{m_i} \rangle \phi_{m_i}$
Fim

A notação $\langle \mathbf{x}, \mathbf{y} \rangle$ representa o produto interno entre os vetores \mathbf{x} e \mathbf{y} .

Para a resolução do problema de BSS, o emprego do algoritmo MP padrão para cada mistura de forma independente é proposto em [10], enquanto que em [11], para configurações instantâneas, criam-se átomos multicanais no estágio de reconstrução das fontes.

3.1. Nova Proposta de um Algoritmo MP Multicanal para BSS

Nossa proposta consiste em um algoritmo MP multicanal voltado para a etapa de estimativa dos coeficientes de atenuação. Seja ϕ um átomo multicanal de comprimento K e $\phi^{\{L\}}$ e $\phi_{\{L\}}$ o átomo monocanal formado, respectivamente, pelas primeiras e últimas L amostras de ϕ . Seja o operador \mathcal{Z}^d para um inteiro d definido como:

$$\mathcal{Z}^d\{\phi\} = \begin{cases} [\mathbf{0} \ \phi^{\{K+d\}}], & \text{para } d \leq 0 \\ [\phi_{\{K+d\}} \ \mathbf{0}], & \text{para } d > 0 \end{cases}, \quad (2)$$

onde $\mathbf{0}$ é um vetor-linha que contém $K - |d|$ zeros.

A partir de um átomo monocanal normalizado ϕ podemos formar átomos multicanais $\phi^{(d_1, d_2, \dots, d_{M-1})}$:

$$\phi^{(d_1, d_2, \dots, d_{M-1})} = (\phi, \mathcal{Z}^{d_1}\{\phi\}, \mathcal{Z}^{d_2}\{\phi\}, \dots, \mathcal{Z}^{d_{M-1}}\{\phi\}). \quad (3)$$

Seja \mathbf{x}_i^j o j -ésimo bloco, de comprimento K , da i -ésima mistura. Definamos o j -ésimo bloco multicanal das misturas como:

$$\mathcal{X}_j = (\mathbf{x}_1^j, \mathbf{x}_2^j, \dots, \mathbf{x}_M^j). \quad (4)$$

Após estas definições, podemos apresentar nossa proposta de decomposição MP multicanal para cada bloco multicanal \mathcal{X}_j :

Algoritmo MP Multicanal Proposto
Inicialização
 $\mathcal{R}^0 \leftarrow \mathcal{X}_j$
Laço Principal
Para $i = 1, 2, \dots, J$
 Escolha $\phi_{m_i}^{(d_1, d_2, \dots, d_{M-1})} := \arg \max_{\phi \in \mathcal{D}, d_1, d_2, \dots, d_{M-1}} |\langle \mathcal{R}^{i-1}, \phi^{(d_1, d_2, \dots, d_{M-1})} \rangle|$
 $\mathcal{R}^i \leftarrow \mathcal{R}^{i-1} - \langle \mathcal{R}^{i-1}, \phi_{m_i}^{(d_1, d_2, \dots, d_{M-1})} \rangle \phi_{m_i}^{(d_1, d_2, \dots, d_{M-1})}$
Fim

A implementação do algoritmo proposto pode recorrer a convoluções entre cada átomo $\phi \in \mathcal{D}$ e os blocos de cada mistura, pois os atrasos oriundos do operador \mathcal{Z} e o produto interno entre os átomos ϕ e um bloco monocanal \mathbf{x}_j podem ser utilizados para calcular $\langle \mathcal{R}^{i-1}, \phi_{m_i}^{(d_1, d_2, \dots, d_{M-1})} \rangle$. Estas convoluções podem ser aceleradas por meio de recurso a técnicas de transformadas rápidas de Fourier.

3.2. ESTIMATIVA DOS COEFICIENTES DE ATENUAÇÃO

O objetivo do uso do MP multicanal justifica-se por esperarmos que, caso o módulo do produto interno de um átomo multicanal $\phi^{(d_1, d_2, \dots, d_{M-1})}$ com \mathcal{X}_j seja elevado, este átomo esteja majoritariamente associado com uma das fontes.

Seja $\phi_m^{(d_1, d_2, \dots, d_{M-1})} = (\phi, \mathcal{Z}^{d_1} \{\phi\}, \mathcal{Z}^{d_2} \{\phi\}, \dots, \mathcal{Z}^{d_{M-1}} \{\phi\})$ o átomo multidimensional selecionado numa iteração MP. Por meio deste átomo, podemos aproximar o k -ésimo bloco da primeira mistura e o k -ésimo bloco da j -ésima mistura por meio de:

$$\begin{aligned} \mathbf{x}_1^k &\approx \langle \mathbf{x}_1^k, \phi_m \rangle \phi_m \\ \mathbf{x}_j^k &\approx \langle \mathbf{x}_j^k, \mathcal{Z}^{d_j} \{\phi_m\} \rangle \mathcal{Z}^{d_j} \{\phi_m\} \end{aligned} \quad (5)$$

Se o átomo predominantemente descreve a i -ésima fonte, temos que $d_j \approx \Delta_{ij}$ e, portanto:

$$\gamma = \frac{\langle \mathbf{x}_j^k, \mathcal{Z}^{d_j} \{\phi_m\} \rangle}{\langle \mathbf{x}_1^k, \phi_m \rangle} \approx h_{ij}. \quad (6)$$

Note que γ é calculado a partir de dados do problema (as misturas e o dicionário), sem qualquer apelo às fontes, que são desconhecidas, e que um histograma de γ conterà picos localizados nos valores de h_{ij} . Importa ressaltar que, caso os atrasos Δ_{ij} sejam fracionários, a aproximação acima pode jamais se tornar verdadeira, dado que os atrasos implementados nos átomos são inteiros. Poder-se-ia questionar se o uso de atrasos inteiros para formar os átomos $\phi_m^{(d_1, d_2, \dots, d_{M-1})}$ não introduziria um viés nas estimativas dos coeficientes de atenuação (obtidas a partir da aproximação acima), no caso de os atrasos serem fracionários. Porém, as simulações (cujos resultados são apresentados na Subseção 5.2) demonstram que este viés é estatisticamente pouco relevante. As naturais oscilações estatísticas de γ apresentam uma simetria em torno dos valores reais de h_{ij} , o que nos permite aproximá-los com razoável precisão, por meio do procedimento descrito a seguir.

O janelamento das misturas (que as segmenta em blocos) e o número de iterações MP determinam a quantidade P de estimativas γ . A n -ésima estimativa será doravante denotada por $\gamma(n)$, para $1 \leq n \leq P$. Por simplicidade, suporemos daqui em diante que o processo de mistura ocorre numa configuração do tipo 2×2 (2 fontes e 2 misturas); logo, cabe estimar h_{21} e h_{22} .

Nossa abordagem, inspirada em [7], consiste em 1) gerar um histograma das frequências de $\gamma(n)$ e 2) obter os picos do histograma. A densidade do histograma para cada valor candidato \hat{h}_{2i}^c será aproximada pela função potencial $\Phi(\hat{h}_{2i}^c)$, definida como:

$$\Phi(\hat{h}_{2i}^c) = \sum_{n=1}^P \Theta(\gamma(n), \hat{h}_{2i}^c), \quad (7)$$

onde $\Theta(\gamma(n), \hat{h}_{2i}^c)$ é calculado da forma a seguir:

$$\Theta(\gamma(n), \hat{h}_{2i}^c) = \begin{cases} 1 - \frac{|\gamma(n) - \hat{h}_{2i}^c|}{\lambda}, & \text{para } |\gamma(n) - \hat{h}_{2i}^c| \leq \lambda \\ 0, & \text{para } |\gamma(n) - \hat{h}_{2i}^c| > \lambda \end{cases}, \quad (8)$$

Em nossa implementação, λ assume o valor 0,1. As localizações dos picos do histograma são razoavelmente insensíveis à escolha de λ , o qual pode ser compreendido como um fator de suavidade do histograma (quanto maior o valor de λ , mais suave é a curva do histograma).

4. ESTIMATIVA DOS ATRASOS VIA MINIMIZAÇÃO DA INFORMAÇÃO MÚTUA

Em contextos determinados, lineares e instantâneos, conhecer a matriz de mistura (supostamente não singular) equivale a recuperar as fontes, já que, para tanto, basta efetuar a transformação linear nas misturas inversa à implementada pela matriz de mistura. Dadas as ambiguidades de escalamento e permutação, inerentes à separação cega de fontes, as estimativas acabam por ser versões escaladas e possivelmente permutadas das fontes originais. No caso de misturas anecoicas, há ainda a ambiguidade de atraso, o que implica estimativas que, mesmo nos casos bem-sucedidos, são versões filtradas [12] (e possivelmente permutadas) das fontes.

Neste caso ($N = M = 2$), se o filtro estimado na posição (i, j) da matriz de mistura for $\hat{\mathbf{h}}_{ij}$, podemos utilizar a seguinte matriz de separação [12]:

$$\hat{\mathbf{W}} = \text{inv}(\mathbf{H}) \cdot \det(\mathbf{H}) = \begin{bmatrix} \hat{\mathbf{h}}_{22} & -\hat{\mathbf{h}}_{12} \\ -\hat{\mathbf{h}}_{21} & \hat{\mathbf{h}}_{11} \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{h}}_{22} & -1 \\ -\hat{\mathbf{h}}_{21} & 1 \end{bmatrix}. \quad (9)$$

Seja $\mathbf{C} = \hat{\mathbf{W}}\mathbf{H}$ a matriz que condensa os processos de mistura e de separação. Não é difícil concluir que a escolha de $\hat{\mathbf{W}}$ conforme a Equação (9) resultará numa separação bem-sucedida caso os filtros \mathbf{h}_{2i} sejam bem-estimados. Para tanto, avaliemos \mathbf{C} quando $\hat{\mathbf{h}}_{2i} = \mathbf{h}_{2i}$:

$$\mathbf{C} = \begin{bmatrix} h_{22}z^{-\Delta_{22}} & -1 \\ -h_{21}z^{-\Delta_{21}} & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ h_{21}z^{-\Delta_{21}} & h_{22}z^{-\Delta_{22}} \end{bmatrix} = \begin{bmatrix} h_{22}z^{-\Delta_{22}} - h_{21}z^{-\Delta_{21}} & 0 \\ 0 & h_{22}z^{-\Delta_{22}} - h_{21}z^{-\Delta_{21}} \end{bmatrix}, \quad (10)$$

de modo que, supondo que não houve permutação nas estimativas, teríamos $y_i(n) = h_{22}s_i(n - \Delta_{22}) - h_{21}s_i(n - \Delta_{21})$, onde $y_i(n)$ é a n -ésima amostra da i -ésima estimativa.

Como o algoritmo MP multicanal proposto estima os coeficientes, para determinar $\hat{\mathbf{H}}$, cumpre apenas estimar Δ_{2i} , para $i = 1, 2$. Parametrizando $\hat{\mathbf{H}}$ em função de $\hat{\Delta}_{21}$ e $\hat{\Delta}_{22}$, temos:

$$\hat{\mathbf{H}}(\hat{\Delta}_{21}, \hat{\Delta}_{22}) = \begin{bmatrix} h_{22}z^{\hat{\Delta}_{22}} & -1 \\ -h_{21}z^{\hat{\Delta}_{21}} & 1 \end{bmatrix}. \quad (11)$$

Com esta parametrização de $\hat{\mathbf{H}}$, pode-se calcular as estimativas das fontes para cada combinação possível de atrasos $\hat{\Delta}_{2i}$. Variando-se as combinações, temos uma série de estimativas possíveis. Para selecionar as estimativas ótimas (indiretamente selecionando $\hat{\Delta}_{2i}$), partimos da hipótese estatisticamente forte, mas em geral fisicamente plausível de que as fontes são independentes [13]. Logo, podemos concluir que os atrasos $\hat{\Delta}_{2i}$ serão os mais próximos possíveis aos que desejamos estimar quando a informação mútua das saídas $y_i(n)$ for mínima (porque neste caso recuperamos versões filtradas das fontes nas estimativas).

Portanto, para determinar os atrasos testamos várias combinações ($\hat{\Delta}_{21}, \hat{\Delta}_{22}$), calculamos as estimativas das fontes, estimamos a informação mútua entre elas e escolhemos, dentre as combinações testadas, a que apresenta menor informação mútua.

Como medida de informação mútua, empregamos a proposta por [14], onde se demonstra que é possível aproximar tanto quanto se queira (em probabilidade) a estimativa da informação mútua por meio de cálculos das frequências relativas dos sinais em partições apropriadas e angariando a independência condicional nos retângulos nos quais as partições são efetuadas. O artigo [15] propôs o uso desta medida de informação mútua para o problema de separação cega de fontes em configurações instantâneas.

5. RESULTADOS

5.1. ESTIMATIVA DOS COEFICIENTES DE ATENUAÇÃO VIA MP MULTICANAL

Neste experimento, foram testadas quatro combinações de sinais de voz gravados em 16 kHz, entre vozes masculinas e femininas, reamostrados para uma taxa de amostragem de 8 kHz. Para a decomposição MP, foram empregados 1500 blocos e 10 iterações por bloco, sem sobreposição entre blocos adjacentes. Os trechos dos sinais empregados têm duração de 7,5s. Cada átomo do dicionário, bem como os blocos das misturas, tem comprimento igual a 40. Os átomos são reais e construídos a partir de amostras de versões moduladas, defasadas e comprimidas/expandidas da função sinc:

$$f(t, \alpha_1, \alpha_2, \alpha_3) = \text{sinc}(0, 2\alpha_1 t) \cos(\alpha_2 t + \alpha_3), \quad (12)$$

onde $\alpha_1 \in [0, 39]$, $\alpha_2 \in [0, 01, 0, 05]$ e $\alpha_3 \in [0, \frac{9\pi}{5}]$ correspondem, respectivamente, ao fator de compressão, à frequência da portadora modulante e à fase da portadora. Cada átomo (de um total de 32000 átomos monocanais) foi escalado de modo a apresentar norma unitária.

Na Figura 1, apresentamos os histogramas, bem como os valores reais dos coeficientes (escolhidos aleatoriamente). Claramente, os dois maiores picos dos histogramas estão associados aos valores reais de h_{2i} , mostrando a robustez do método proposto.

A Tabela 1 compara os valores reais dos coeficientes de atenuação com os estimados, revelando a precisão do método.

Tabela 1: Coeficientes de atenuação reais e estimados.

Caso	Coef. Real 1	Coef. Estimado 1	Coef. Real 2	Coef. Estimado 2
1	-2,0123	-2,007	2,0669	2,064
2	-1,6718	-1,671	-0,9439	-0,947
3	-2,3214	-2,319	2,1700	2,182
4	-0,27207	-0,275	1,0468	1,038

5.2. APLICAÇÃO NUM SISTEMA DE RECONHECIMENTO AUTOMÁTICO DE INSTRUMENTOS MUSICAIS (SRAIM)

Após a separação de sinais efetuada por algum algoritmo, existem diversas métricas capazes de avaliar a qualidade de separação, como por exemplo a SIR (razão sinal-ruído, do inglês *signal-to-interference ratio*) e a SAR (razão sinal-artefato, do inglês *signal-to-artifact ratio*). Optamos por verificar o efeito da separação num sistema de reconhecimento automático de instrumentos musicais (SRAIM). A partir da seleção (aleatória) de dois instrumentos, as fontes são geradas pela construção de sequências monofônicas de notas oriundas de cada um dos instrumentos pré-selecionados. Estas sequências são então misturadas de forma anecoica e separadas pelo algoritmo de separação proposto, supondo que os coeficientes de atenuação foram corretamente estimados (ou seja, apenas a estimativa dos atrasos é realmente efetuada). Após a separação, um extrator de notas é empregado para extrair as notas de cada estimativa [16]. Estas notas são então classificadas pelo sistema SRAIM [16]. Após todas as notas de cada estimativa serem estimadas, a decisão a respeito de qual instrumento será associado a cada estimativa decorre do instrumento mais frequentemente estimado ao longo da mesma. O fato de existirem diversas notas a classificar em

cada estimativa significa que mesmo uma taxa de acerto relativamente baixa (no tocante às notas de uma sequência) não resulta necessariamente num erro de estimativa de instrumento da sequência como um todo. Cumpre notar que o processo de separação insere algumas distorções nas estimativas, mesmo que estas sejam bem-sucedidas, e que estas distorções podem impactar o desempenho dos classificadores empregados.

O SRAIM empregado recorre a um banco de classificadores [16], conforme ilustrado pela Figura 2. O desempenho do banco

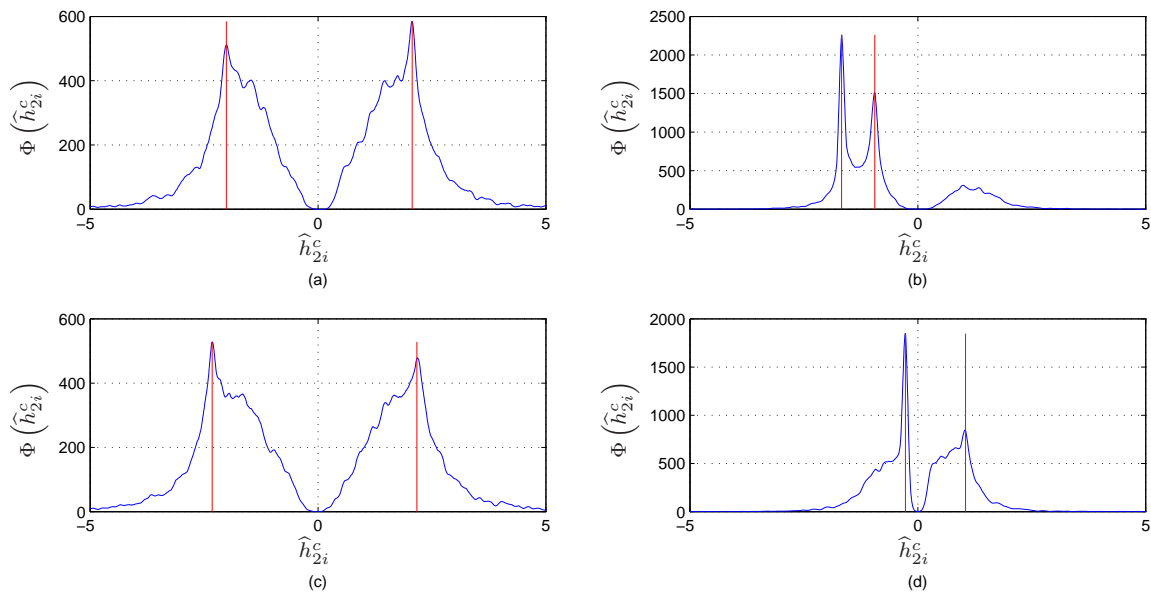


Figura 1: Histogramas construídos a partir do cálculo de $\Phi(\hat{h}_{2i}^c)$ (em azul), para quatro casos distintos. Os valores reais estão assinalados em vermelho.

de classificadores supera o de qualquer um dos classificadores, caso empregados isoladamente. Este SRAIM foi projetado para o reconhecimento de notas isoladas, a partir de um segmento das mesmas.

Os três primeiros blocos que constam na Figura 2 (Pot. Inst., IMF e ED) condensam parte do pré-processamento aplicado ao

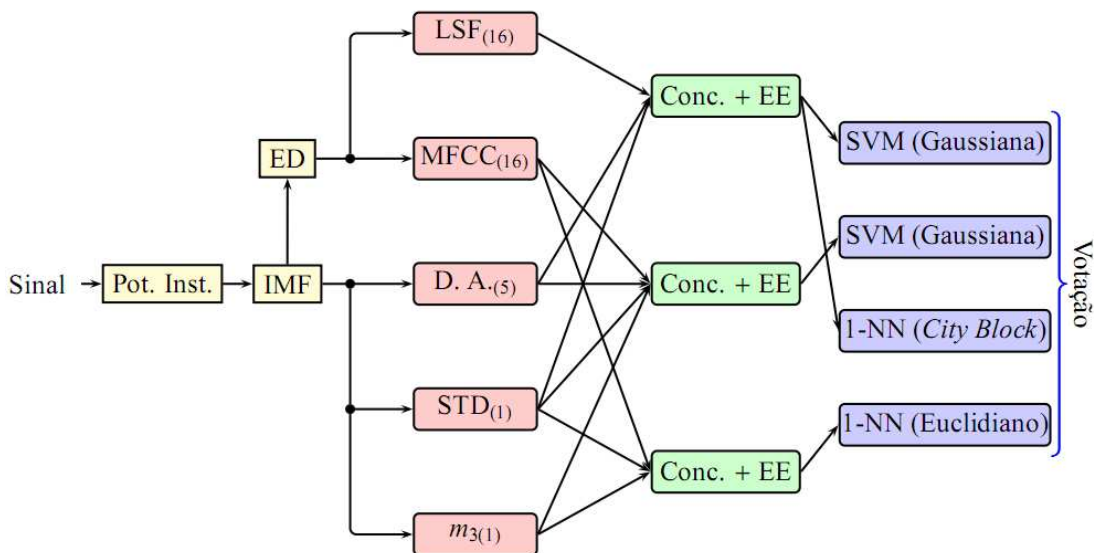


Figura 2: Sistema de Reconhecimento Automático de Instrumentos Musicais Empregado.

sinal (nota de um instrumento musical). O bloco Pot. Inst. obtém a potência instantânea do sinal, através de um limiar definido como 90% da média da potência instantânea da nota musical. A segmentação da nota é efetuada pelo bloco IMF (Início, Meio e Fim), proposto em [16], o qual estabelece limiares dependentes da potência da nota de modo a estimar seus instantes de início e de fim. Após a detecção destes instantes, as medidas de desvio padrão (STD) e momento central de terceira ordem (m_3) são extraídas, bem como cinco descritores de áudio (DA): taxa de cruzamento por zero, fluxo espectral, valor RMS do *frame*, centroide

espectral e largura do centroide espectral. Por fim, o segmento extraído passa por um processo de escalamento dinâmico (bloco ED), visando sua normalização. A partir desse momento são extraídos os vetores de características da nota que são formados basicamente por dois tipos de codificadores: coeficientes LSF (do inglês *Line Spectral Frequencies*) [17] e MFCC (do inglês *Mel Frequencies Cepstral Coefficients*) [18]. Ambos os codificadores utilizam 16 coeficientes representativos. Essas medidas são concatenadas (daí o termo “Conc.” na Figura 2 de 3 diferentes formas, resultando em 3 vetores de características distintos, os quais, após passar por um processo de escalamento estatístico (EE), servem como entrada para 4 classificadores. Foram utilizados como classificadores 2 SVMs (do inglês *Support Vector Machines*) [19] e 2 K-NN (do inglês *K-nearest neighbor* [20]). Ao final a classe predita representa a classe mais votada entre os quatro classificadores (caso haja empate, há um sorteio entre as predições).

O SRAIM foi treinado com um subconjunto de notas (90% das notas disponíveis) oriundo de três bases de dados distintas: RWC [21], MIS (do inglês *Musical Instruments Samples* [22], e MUMS (*McGill University Master Samples* [23]). O sistema adotado é capaz de classificar 20 instrumentos distintos. O conjunto usado na fase de treinamento não possui notas em comum com o conjunto de notas usado para formar as sequências monofônicas de teste (10% das notas oriundas da base de dados RWC). Para testar a viabilidade do algoritmo de BSS proposto como uma etapa de pré-processamento de um SRAIM, efetuamos 6 misturas de instrumentos. Para cada uma das misturas foram escolhidos aleatoriamente os coeficientes de atenuação, sendo testados 6 casos distintos. A diferença entre os casos reside na escolha (também aleatória) dos atrasos Δ_{ij} . No total, temos portanto 6 (misturas) \times 2 (fontes) \times 6 (casos) = 72 sequências cuja informação de alto nível (instrumento gerador) deve ser recuperada. Do total de 72 casos testados, o SRAIM acertou 70. Ou seja, a taxa de acerto ficou em torno de 97%. Caso não empreguemos o método de separação, aplicando diretamente às misturas o SRAIM, temos uma taxa de acerto de 58%, aproximadamente. A significativa redução da taxa de acerto (quando o método de separação não é utilizado) é um bom indicativo da eficiência da separação obtida pelo algoritmo proposto.

A Tabela 2 apresenta a SDR (razão sinal-distorção, a qual contempla degradações nas estimativas oriundas tanto de interferências quanto de artefatos) obtida pelo sistema separador nos casos simulados. Observa-se que as estimativas obtidas com o método de BSS proposto obtém uma alta SDR.

Tabela 2: SDR (razão sinal-distorção) das estimativas para as misturas e os casos simulados. Misturas diferentes apresentam coeficientes de atenuação distintos. Casos distintos apresentam atrasos diferentes.

Mistura	Estimativa	Caso 1	Caso 2	Caso 3	Caso 4	Caso 5	Caso 6
1	1	31.6	25.3	21.9	28.3	35.2	34.7
1	2	21.2	19.5	57.1	62.8	41.9	23.9
2	1	30.1	24.2	37.3	29.2	24.9	34.0
2	2	25.9	24.3	55.0	54.8	45.3	28.6
3	1	37.9	30.0	33.5	21.1	23.9	24.3
3	2	33.4	32.2	48.5	48.6	44.1	35.1
4	1	24.5	22.6	25.7	42.2	34.5	18.5
4	2	24.5	24.1	38.3	38.4	33.7	25.9
5	1	25.9	24.7	36.1	56.6	28.9	31.8
5	2	22.0	27.1	30.3	30.3	26.1	21.2
6	1	17.8	25.0	31.6	44.1	17.8	29.1
6	2	29.5	31.2	18.1	22.2	34.9	29.4

6. CONCLUSÃO

Neste artigo, o problema de separação de fontes em contextos anecoicos com atrasos possivelmente fracionários foi contemplado. O método de separação proposto estima primeiramente os coeficientes de atenuação do sistema de mistura, por meio de uma decomposição via MP multicanal das misturas. Esta etapa atua no espaço de mistura e emprega princípios de SCA, podendo ser facilmente estendida para o caso subdeterminado, desde que tenhamos $M \geq 2$. Na segunda etapa, o princípio da informação mútua é empregado para a seleção dos atrasos presentes nas misturas. O método proposto apresentou elevado grau de acurácia na estimativa dos coeficientes de atenuação, bem como se mostrou viável como uma etapa de pré-processamento para um sistema de reconhecimento automático de instrumentos musicais.

REFERÊNCIAS

- [1] A. Hyvärinen, J. Karhunen and E. Oja. *Independent Component Analysis*. John Wiley & Sons, New York, 2001.
- [2] X. Cao and R. Liu. “General Approach for Blind Source Separation”. *IEEE Transactions on Signal Processing*, vol. 44, no. 3, pp. 562–571, Mar. 1996.

- [3] M. Naceur, M. A. Loghmari and M. R. Boussema. “The Contribution of the Sources Separation Method in the Decomposition of Mixed Pixels”. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 11, pp. 2642–2653, Nov. 2004.
- [4] J. Anemuller, T. Sejnowski and S. Makeig. “Complex Independent Component Analysis of Frequency-domain Electroencephalographic Data”. *Neural Networks*, vol. 16, no. 9, pp. 1311–1323, Nov. 2003.
- [5] F. Acernese, A. Ciaramella, S. D. Martino, R. D. Rosa, M. Falanga and R. Tagliaferri. “Neural Networks for Blind-Source Separation of Stromboli Explosion Quakes”. *IEEE Transactions on Neural Networks*, vol. 14, no. 1, pp. 167–175, Jan. 2003.
- [6] A. T. Erdogan. “Globally Convergent Deflationary Instantaneous Blind Source Separation Algorithm for Digital Communication Signals”. *IEEE Transactions on Signal Processing*, vol. 55, no. 5, pp. 2182–2192, May. 2007.
- [7] P. Bofill and M. Zibulevsky. “Underdetermined Blind Source Separation Using Sparse Representations”. *Signal Processing*, vol. 81, pp. 2353–2362, 2001.
- [8] S. G. Mallat and Z. Zhang. “Matching Pursuits with Time-Frequency Dictionaries”. *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397 – 3415, Dec. 1993.
- [9] P. Kisilev, M. Zibulevsky and Y. Y. Zeevi. “Multiscale Framework for Blind Separation of Linearly Mixed Signals”. *Journal of Machine Learning Research*, pp. 1–26, 2003.
- [10] P. Sugden and N. Canagarajah. “Underdetermined Noisy Blind Separation Using Dual Matching Pursuits”. *ICASSP*, vol. V, pp. 557–560, May. 2004.
- [11] S. Lesage, S. Krstulovic and R. Gribonval. “Underdetermined Source Separation: Comparison of Two Approaches Based on Sparse Decompositions”. *ICA*, pp. 633–640, Mar. 2006.
- [12] H. Buchner and W. Kellerman. “A Fundamental Relation Between Blind and Supervised Adaptive Filtering Illustrated for Blind Source Separation and Acoustic Echo Cancellation”. *HSCMA*, pp. 17–20, May. 2008.
- [13] J.-F. Cardoso. “Blind Signal Separation: Statistical Principles”. *Proceedings of the IEEE*, vol. 86, no. 10, pp. 2009 – 2025, Oct. 1998.
- [14] G. A. Darbellay and I. Vajda. “Estimation of the Information by an Adaptive Partitioning of the Observation Space”. *IEEE Transactions on Information Theory*, vol. 45, no. 4, pp. 1315–1321, May. 1999.
- [15] G. A. Darbellay and P. Tichavsky. “Independent Component Analysis Through Direct Estimation of the Mutual Information”. *ICA'2000 Proceedings*, pp. 69–75, Jun. 2000.
- [16] J. C. P. Filho. “Classificação de Instrumentos Musicais em Configurações Monofônicas e Polifônicas”. Master’s thesis, COPPE - Universidade Federal do Rio de Janeiro, Rio de Janeiro, Sep. 2009.
- [17] P. Kabal and R. P. Ramachandran. “The Computation of Line Spectral Frequencies using Chebyshev Polynomials”. *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 34, no. 6, pp. 1419–1426, Dec. 1986.
- [18] S. B. Davis and P. Mermelstein. “Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences”. *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 28, no. 4, pp. 357–366, Aug. 1980.
- [19] C. Cortes and V. Vapnik. “Support-Vector Networks”. *Machine Learning*, vol. 5, pp. 273–297, 1995.
- [20] A. Aizerman, E. Braverman and L. Rozonoer. “Theoretical Foundations of the Potential Function Method in Pattern Recognition Learning”. *Automation and Remote Control*, vol. 25, pp. 821–837, 1964.
- [21] M. Goto. “RWC Music Database”. In <http://sta.aist.go.jp/m.goto/RWC-MDB/>. National Institute of Advanced Industrial Science and Technology (AIST), Japan.
- [22] L. Fritts. “Musical Instrument Samples”. In <http://theremin.music.uiowa.edu/MIS.html>. University of Iowa Electronic Music Studios, University of Iowa, USA.
- [23] F. Opolko and J. Wapnick. “McGill University’s Master Samples”. In <http://www.music.mcgill.ca/resources/mums/html/mums.html>. McGill University, USA.