

Sensibilidade da Separação Cega de Fontes à Variação do Comprimento dos Filtros de Separação e da Ordem da Estatística Usada na Função Custo

Diego Barreto Haddad, Paulo Bulkool Batalheiro e Mariane Rembold Petraglia

Resumo—Métodos de separação cega de fontes têm sido propostos para melhoria de desempenho em sistemas automáticos de reconhecimento de fala, sistemas de auxílio à audição e análise de sinais biomédicos, entre outros. Neste artigo empregamos técnicas recentes, baseadas na análise de componentes independentes (ICA), para separação de sinais misturados de forma convolutiva e avaliamos experimentalmente a melhoria de desempenho destas estruturas quando os filtros de separação têm comprimento maior que os filtros de mistura e também quando a função custo contempla estatísticas de ordem superior.

Palavras-Chave—ICA, Misturas Convolutivas, Estatísticas de Segunda Ordem, Estatísticas de Ordem Superior.

Abstract—Methods of blind source separation have been proposed for performance improvement in systems of automatic speech recognition, hearing aids and biomedical signal analysis, among others. In this article, we employ recently proposed techniques, based on the independent components analysis (ICA), for separation of convolutive mixtures, and evaluate experimentally the performance improvements obtained when the separation filters have larger lengths than the mixture filters and when high-order statistics are employed in the cost function.

Keywords—ICA, Convolutional Mixtures, Second-Order Statistics, High-Order Statistics.

I. INTRODUÇÃO

Durante os últimos vinte anos houve intensa pesquisa em técnicas de separação cega de fontes, comumente denominadas BSS (do inglês *Blind Source Separation*), as quais possibilitam a extração do sinal de uma fonte desejada a partir de uma mistura linear de sinais sem a necessidade de qualquer informação da posição das fontes, de suas estruturas espectrais ou do conhecimento dos coeficientes do sistema que realiza a mistura. Para atingir este objetivo, comumente é utilizada uma técnica computacional e estatística baseada na análise de componentes independentes (ICA) [1], que possibilita a extração do sinal de cada uma das fontes a partir do sinal capturado pelos sensores, e que deve ser aplicada em sinais que possam ser modelados como misturas lineares de componentes estatisticamente independentes e não-gaussianas. Era comum a admissão do modelo de misturas lineares e instantâneas, porém mais recentemente surgiu um grande interesse em se contemplar modelos mais realistas. Estes modelos, embora

ainda lineares, passaram a ser convolutivos, permitindo assim modelar-se a reverberação sempre presente em ambientes ecóicos. O desafio no momento é obter uma boa solução para o problema de separação de misturas convolutivas, que tipicamente pode ser tratado no domínio do tempo ou no domínio da frequência. O problema no domínio do tempo é a demanda computacional para sistemas de ordem elevada. Tentando amenizar este problema, existem abordagens que transformam as misturas para o domínio da frequência, admitem (a partir do fato de que a convolução no tempo se torna produto na frequência) que em cada frequência a mistura é instantânea e resolvem o problema como se fosse um caso de mistura instantânea.

Essas abordagens possuem o mérito de reduzir enormemente a carga computacional, à custa de tornar não trivial o problema da permutação e do escalamento em cada frequência, criando a possibilidade de se trocarem as componentes na frequência entre as fontes na saída do sistema. O escalamento comumente é resolvido de forma razoável mediante o recurso ao Princípio da Mínima Distorção, apresentado em [3]. Já o problema da permutação, mais complexo, costuma ser abordado via correlação (ou uma versão modificada da mesma, inclusive utilizando critérios que contemplam simultaneamente diversas frequências) de envelopes [4],[5], direção de chegada de cada fonte [6],[7] ou uma conjunção de ambas as técnicas [8], opção esta que tenta atenuar as desvantagens de uma abordagem com as vantagens da outra. Seguiremos a nomenclatura adotada por Aichner et al. [9], denominando este tipo de método (transformada na frequência associada ao método ICA instantâneo) de banda estreita (*narrowband*), em contraposição aos métodos que não analisam cada frequência isoladamente, chamados banda larga (*broadband*).

Os métodos utilizados neste artigo são apresentados tanto no domínio do tempo quanto no domínio da frequência, usando estatísticas de segunda ordem (basicamente implementadas utilizando correlações) e também estatísticas de ordem superior (basicamente implementadas utilizando correlações não-lineares) na função custo utilizada na busca dos coeficientes do sistema que desfaz a mistura. Na Seção II a estrutura para separação cega de fontes é apresentada, focalizando a discussão em misturas instantâneas. O caso mais geral, envolvendo separação de sinais misturados de forma convolutiva, é abordado no domínio do tempo na Seção III, enquanto a separação no domínio da frequência é vista na Seção IV. Uma discussão entre a utilização de estatísticas de segunda ordem e de ordem superior é abordada na Seção V. Os resultados

Diego Barreto Haddad e Mariane Rembold Petraglia, Programa de Engenharia Elétrica, COPPE - DEL/Poli, Universidade Federal do Rio de Janeiro, CP 68504, CEP 21945-970, Rio de Janeiro, RJ, Brasil, E-mails: diego@pads.ufrj.br, mariane@pads.ufrj.br. Paulo Bulkool Batalheiro, Departamento de Engenharia Eletrônica e Telecomunicações, Universidade do Estado do Rio de Janeiro, CEP 20559-900, Rio de Janeiro, RJ, Brasil. Este trabalho foi parcialmente financiado pelo CNPq e pela FAPERJ.

experimentais são mostrados na Seção VI e em seguida são apresentadas, na Seção VII, as conclusões deste trabalho.

II. SEPARAÇÃO CEGA DE FONTES

O problema de separação cega de fontes (BSS - *Blind Source Separation*) consiste em estimar fontes de sinais usando somente a informação dos sinais misturados observados em cada canal de entrada, ou seja, sem nenhuma informação das fontes ou do sistema de mistura destes sinais. Alguns exemplos típicos incluem misturas simultâneas de sinais de fala ou áudio que são capturados por alguns microfones, ondas cerebrais gravadas a partir de múltiplos sensores e interferências entre sinais de rádio que chegam em uma estação móvel.

No caso de separação de fontes de áudio, alguns microfones são colocados em diferentes posições de forma que cada um capture uma mistura diferente das fontes originais, ainda que com pequenas diferenças de atraso e de potência, de forma a inserir uma diversidade espacial que forneça mais informações para o método de separação. Considerando situações do mundo real, podemos considerar sinais de fala como sendo as fontes originais e uma sala e/ou um auditório como o sistema que realiza a mistura. Neste caso, os sinais capturados pelo microfones são afetados pela reverberação do ambiente. A Fig. 1 mostra o esquema geral de um sistema de separação cega de fontes. Os sinais das Q fontes capturados pelos P microfones

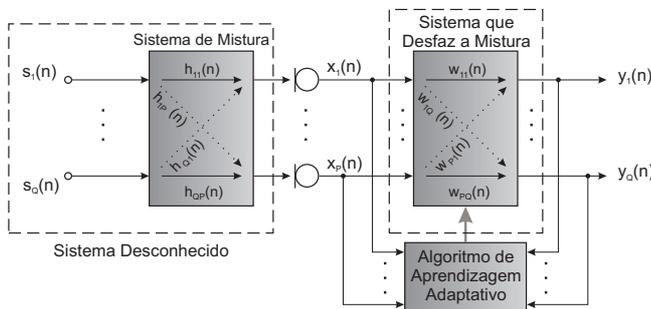


Fig. 1. Esquema geral para separação cega de fontes.

podem ser modelados como

$$x_p(n) = \sum_{q=1}^Q \sum_{k=1}^M h_{qp}(k) s_q(n-k+1) \text{ para } p = 1, \dots, P, \quad (1)$$

sendo s_q o sinal da q -ésima fonte, x_p o sinal captado pelo p -ésimo microfone e h_{qp} a resposta ao impulso, de comprimento M , do caminho acústico entre a fonte q e o microfone p .

Para desfazer a mistura a partir dos sinais capturados pelos microfones são estimados, através de um algoritmo de aprendizagem adaptativo, os filtros w_{pq} de comprimento L de forma que os sinais y_q tornem-se mutuamente independentes. Portanto, os sinais não mais misturados são do tipo $y_p = F(s_q)$ e podem ser escritos da seguinte forma:

$$y_q(n) = \sum_{p=1}^P \sum_{k=1}^L w_{pq}(k) x_p(n-k+1) \text{ para } q = 1, \dots, P. \quad (2)$$

A. Separação Cega de Fontes para Misturas Instantâneas

No caso dos coeficientes h_{qp} e w_{pq} serem escalares ($L = M = 1$, inexistindo atraso e reverberação, como seria o caso de um misturador de áudio), nosso problema é reduzido a um caso de mistura instantânea que já foi muito estudada e onde se consegue resultados muito bons. Neste caso, o sistema de mistura pode ser descrito por

$$\mathbf{x} = \mathbf{H}\mathbf{s}, \quad (3)$$

sendo $\mathbf{x} = [x_1(n) \ x_2(n) \ \dots \ x_P(n)]^T$ o vetor que contém os sinais observados, $\mathbf{s} = [s_1(n) \ s_2(n) \ \dots \ s_Q(n)]^T$ o vetor que contém os sinais das fontes independentes, \mathbf{H} a matriz de mistura formada pelos coeficientes h_{qp} , e o sistema que desfaz a mistura é dado por

$$\mathbf{y} = \mathbf{W}\mathbf{x}, \quad (4)$$

sendo $\mathbf{y} = [y_1(n) \ y_2(n) \ \dots \ y_Q(n)]^T$ o vetor contendo as estimativas das fontes $s_i(n)$ e \mathbf{W} a matriz formada pelos os coeficientes w_{pq} que desfazem a mistura. Substituindo a Eq. (3) na Eq. (4), temos que

$$\mathbf{y} = \mathbf{W}\mathbf{H}\mathbf{s}, \quad (5)$$

ou seja, se a matriz de mistura for não-singular, existe uma matriz $\mathbf{W} = \mathbf{H}^{-1}$ que desfaz a mistura. Porém, mesmo sendo bem sucedido o método de separação, a matriz \mathbf{W} pode não ser exatamente a inversa da matriz \mathbf{H} por duas razões. Primeiramente, há uma ambigüidade de escalamento; ou seja, caso multiplicássemos uma fonte $s_k(n)$ por uma constante a , isso seria o equivalente a multiplicar uma coluna da matriz \mathbf{H} pelo mesmo valor, e, desconhecida qualquer informação tanto sobre a fonte quanto sobre a matriz de mistura, $s_k(n)$ e $as_k(n)$ são, ambas, soluções igualmente aceitáveis. Assim, podemos restringir o objetivo a obter uma matriz $\mathbf{W}\mathbf{H}$ diagonal não singular. Porém, além dessa ambigüidade, há o problema da permutação, que implica não podermos sempre recuperar (a menos da ambigüidade de escalamento já abordada) a k -ésima linha de \mathbf{x} na k -ésima linha de \mathbf{y} . Em última análise, este problema também ocorre devido à nossa falta de informação sobre as fontes e sobre a matriz de mistura, pois inverter as fontes é uma operação que possui os mesmos efeitos do que trocar as colunas da matriz \mathbf{H} . Logo, somente devemos esperar que o produto $\mathbf{W}\mathbf{H}$, ao invés de ser a matriz identidade, seja uma matriz diagonal ou anti-diagonal (no caso de duas fontes e duas misturas).

B. Gradiente Natural da Informação Mútua

Almejando tornar nula a informação mútua entre as saídas y_i , os métodos de separação cega de fontes não raro têm por objetivo anular as diagonais secundárias da matriz \mathbf{R}_Y definida como:

$$\mathbf{R}_Y = \begin{bmatrix} E[\Phi(y_1)y_1] & E[\Phi(y_1)y_2] & \dots & E[\Phi(y_1)y_Q] \\ E[\Phi(y_2)y_1] & E[\Phi(y_2)y_2] & \dots & E[\Phi(y_2)y_Q] \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ E[\Phi(y_Q)y_1] & E[\Phi(y_Q)y_2] & \dots & E[\Phi(y_Q)y_Q] \end{bmatrix}, \quad (6)$$

sendo

$$\Phi(y_i) = -\frac{\partial}{\partial |y_i|} \{\ln[p(y_i)]\}, \quad (7)$$

onde $p(y_i)$ é a função densidade de probabilidade do sinal y_i . Observe que $\Phi(y_i) = y_i$ implica um método que emprega apenas estatísticas de segunda ordem, não recorrendo às mais altas. O operador $E[\cdot]$ é implementado apenas sobre blocos dos sinais de saída, gerando uma componente estocástica que torna o algoritmo menos sensível a mínimos locais (uma opção muito utilizada seria branquear o sinal, antes de se aplicar o ICA, por meio de uma matriz de branqueamento). Por meio do gradiente natural (mais rápido que o gradiente padrão normalmente utilizado, além de mais robusto ao mau condicionamento da matriz de misturas) proposto em [2], a matriz de separação é obtida iterativamente da seguinte forma:

$$\mathbf{W}_{i+1} = \mathbf{W}_i + \eta \{ \mathbf{I}_{Q \times Q} - E[\Phi(\mathbf{Y})\mathbf{Y}^T] \} \mathbf{W}_i, \quad (8)$$

onde η é um coeficiente de aprendizagem, $\mathbf{I}_{Q \times Q}$ é a matriz identidade de ordem Q e $(\cdot)^T$ indica transposição.

III. MISTURAS CONVOLUTIVAS SEPARADAS NO DOMÍNIO DO TEMPO

No mundo real, devido à reverberação do ambiente, os sinais originais das fontes são filtrados por um sistema MIMO linear antes de serem capturados pelos sensores (ver Fig. 1). A partir deste momento, vamos assumir que o número de fontes é igual ao número de microfones ($P = Q$) e apresentar a solução baseada em estatística de segunda ordem proposta em [10].

Os filtros de separação são do tipo FIR, de comprimento L , gerando a q -ésima saída $y_q(n)$ dada por:

$$y_q(n) = \sum_{p=1}^P \mathbf{x}_p^T(n) \mathbf{w}_{pq}, \quad (9)$$

onde

$$\mathbf{x}_p = [x_p(n), x_p(n-1), \dots, x_p(n-L+1)]^T, \quad (10)$$

e

$$\mathbf{w}_{pq} = [w_{pq,0}, w_{pq,1}, \dots, w_{pq,L-1}]^T. \quad (11)$$

Dois novos parâmetros devem ser introduzidos para generalizar a formulação, a saber: o número D de intervalos no tempo considerados para as correlações na função custo ($1 \leq D \leq L$) e o comprimento N dos blocos de sinais de saída utilizados para estimar as correlações de intervalos de curta duração ($N \geq PD$). O m -ésimo bloco correspondente à q -ésima saída é dado por:

$$\mathbf{Y}_q(m) = \sum_{p=1}^P \mathbf{X}_p(m) \mathbf{W}_{pq}, \quad (12)$$

com

$$\mathbf{X}_p(m) = [\mathbf{U}_p^T(m), \mathbf{U}_p^T(m-1)] \quad (13)$$

e

$$\mathbf{U}_p(m) = \begin{bmatrix} x_p(mL) & \cdots & x_p(mL-L+1) \\ x_p(mL+1) & \cdots & x_p(mL-L+2) \\ \vdots & \ddots & \vdots \\ x_p(mL+N-1) & \cdots & x_p((m-1)L+N) \end{bmatrix}. \quad (14)$$

Por fim, a matriz Sylvester \mathbf{W}_{pq} de dimensões $2L \times D$ é definida como:

$$\mathbf{W}_{pq} = \begin{bmatrix} w_{pq,0} & 0 & \cdots & 0 \\ w_{pq,1} & w_{pq,0} & \ddots & \vdots \\ \vdots & w_{pq,1} & \ddots & 0 \\ w_{pq,L-1} & \vdots & \ddots & w_{pq,0} \\ 0 & w_{pq,L-1} & \ddots & w_{pq,1} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & w_{pq,L-1} \\ 0 & \cdots & 0 & 0 \\ \vdots & \cdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 \end{bmatrix}. \quad (15)$$

A redundância desta matriz significa que conhecer uma de suas colunas implica conhecê-la completamente. Há razões de convergência [10] que requerem o procedimento de, a cada atualização de \mathbf{W}_{pq} , utilizar somente a primeira coluna para preencher a matriz, desprezando todos os valores contidos nas outras colunas. Esse procedimento será adotado em nossa implementação. Combinando todos os canais, podemos escrever concisamente:

$$\mathbf{Y}(m) = \mathbf{X}(m)\mathbf{W}, \quad (16)$$

com as matrizes:

$$\mathbf{Y}(m) = [\mathbf{Y}_1(m), \dots, \mathbf{Y}_P(m)], \quad (17)$$

$$\mathbf{X}(m) = [\mathbf{X}_1(m), \dots, \mathbf{X}_P(m)], \quad (18)$$

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_{11} & \cdots & \mathbf{W}_{1P} \\ \vdots & \ddots & \vdots \\ \mathbf{W}_{P1} & \cdots & \mathbf{W}_{PP} \end{bmatrix}. \quad (19)$$

Definida a formulação, sendo b o número de blocos, a implementação *off-line* do algoritmo tem por função custo:

$$\mathfrak{S} = \frac{1}{b} \sum_{i=1}^b \log \{ \det [\text{bdiag}(\mathbf{Y}^T(i)\mathbf{Y}(i))] \} - \log[\det(\mathbf{Y}^T(i)\mathbf{Y}(i))], \quad (20)$$

onde o operador $\text{bdiag}(\cdot)$ opera sobre uma matriz de submatrizes, anulando todas as submatrizes que não estejam situadas na diagonal principal da matriz (neste artigo, todas as matrizes sobre as quais atua este operador são constituídas por 4 submatrizes de idênticas dimensões). O gradiente natural da função custo acima é fornecido por:

$$\nabla_{\mathbf{W}}^{GN} \mathfrak{S}(m) = \frac{2}{b} \sum_{i=1}^b \mathbf{W} [\mathbf{R}_{yy} - \text{bdiag}(\mathbf{R}_{yy})] [\text{bdiag}(\mathbf{R}_{yy})]^{-1}. \quad (21)$$

A separação das fontes é efetuada por filtros que idealmente geram saídas independentes. Porém, se um sinal s_i é independente de s_j , uma versão filtrada de s_i também é independente de uma versão filtrada de s_j . Portanto, mesmo sendo bem sucedido o processo de separação, as saídas do sistema não são, em geral, versões meramente escaladas e permutadas das

fontes, mas versões *filtradas* das mesmas. Vale notar que essa filtragem costuma não ser perceptível auditivamente.

IV. MISTURAS CONVOLUTIVAS SEPARADAS NO DOMÍNIO DA FREQUÊNCIA

Seja $\mathbf{F}_{4L \times 4L}$ a matriz DFT de dimensão $4L \times 4L$. Assumiremos, sem perda de generalidade, $N = L$, exigindo que os blocos possuam comprimento não mais igual a $2N$, mas a $4N$, pois o tamanho do bloco deve dobrar após a convolução e novamente duplicar após a estimativa de correlação. Definimos:

$$\underline{\mathbf{X}}_p(m) = \text{diag} \{ \mathbf{F}_{4L \times 4L} [x_p(mL-3L), \dots, x_p(mL-1), x_p(mL), x_p(mL+1), \dots, x_p(mL+L-1)]^T \}, \quad (22)$$

$$\underline{\mathbf{X}}(m) = [\underline{\mathbf{X}}_1(m), \dots, \underline{\mathbf{X}}_p(m)], \quad (23)$$

$$\mathbf{W}_{pq} = \text{diag} \{ \mathbf{F}_{4L \times 4L} [w_{pq,0}, \dots, w_{pq,L-1}, 0, \dots, 0]^T \}. \quad (24)$$

Deve-se ressaltar que, a cada iteração na frequência, deve-se voltar com os filtros na frequência para o domínio do tempo de forma a zerar todos os componentes que ultrapassam o tamanho do filtro, restringindo o comprimento deste para L . Esta restrição pode ser representada como um produto pela inversa de uma matriz DFT, em seguida por uma matriz de restrição (que zera os componentes indesejáveis) e então pela matriz DFT. Definimos ainda:

$$\underline{\mathbf{W}} = \begin{bmatrix} \mathbf{W}_{11} & \cdots & \mathbf{W}_{1P} \\ \vdots & \ddots & \vdots \\ \mathbf{W}_{P1} & \cdots & \mathbf{W}_{PP} \end{bmatrix}, \quad (25)$$

$$\mathbf{S}_{xx} = \underline{\mathbf{X}}^H \mathbf{F}_{4L \times 4L} \begin{bmatrix} \mathbf{0}_{3L \times 3L} & \mathbf{0}_{3L \times 3L} \\ \mathbf{0}_{L \times 3L} & \mathbf{I}_{L \times L} \end{bmatrix} \mathbf{F}_{4L \times 4L}^{-1} \underline{\mathbf{X}}, \quad (26)$$

$$\mathbf{G}_{4L} = \mathbf{F}_{4L \times 4L} \begin{bmatrix} \mathbf{I}_{2L \times 2L} & \mathbf{0}_{2L \times 2L} \\ \mathbf{0}_{2L \times 2L} & \mathbf{0}_{2L \times 2L} \end{bmatrix} \mathbf{F}_{4L \times 4L}^{-1}, \quad (27)$$

$$\mathbf{G}_{8L} = \begin{bmatrix} \mathbf{G}_{4L} & \mathbf{0}_{4L \times 4L} \\ \mathbf{0}_{4L \times 4L} & \mathbf{G}_{4L} \end{bmatrix}, \quad (28)$$

$$\mathbf{S}_{yy} = \underline{\mathbf{W}}^H \mathbf{G}_{8L} \mathbf{S}_{xx} \mathbf{G}_{8L} \underline{\mathbf{W}}, \quad (29)$$

$$\mathbf{L} = \begin{bmatrix} \mathbf{F}_{4L \times 4L} \begin{bmatrix} \mathbf{I}_{D \times D} \\ \mathbf{0}_{4L-D \times D} \end{bmatrix} & \mathbf{0}_{4L \times D} \\ \mathbf{0}_{4L \times D} & \mathbf{F}_{4L \times 4L} \begin{bmatrix} \mathbf{I}_{D \times D} \\ \mathbf{0}_{4L-D \times D} \end{bmatrix} \end{bmatrix}, \quad (30)$$

$$\mathbf{L}_I = \begin{bmatrix} \mathbf{1}_{4L \times 1} & \mathbf{0}_{4L \times 1} \\ \mathbf{0}_{4L \times 1} & \mathbf{1}_{4L \times 1} \end{bmatrix}, \quad (31)$$

onde $(\cdot)^H$ significa hermitiano e $\mathbf{0}_{m \times n}$ é a matriz nula de dimensões m e n . O gradiente natural da função custo nesta abordagem é fornecido por:

$$\nabla_{\underline{\mathbf{W}}}^{GN} \mathfrak{S} = \frac{2}{b} \sum_{i=1}^b \mathbf{G}_{8L} \underline{\mathbf{W}} \mathbf{L} \mathbf{L}^H [\mathbf{S}_{yy} - b \text{diag}(\mathbf{S}_{yy})] \mathbf{L} [\text{bdiag}(\mathbf{L}^H \mathbf{S}_{yy} \mathbf{L})]^{-1} \mathbf{L}^H. \quad (32)$$

A matriz $\underline{\mathbf{W}}$, à maneira de \mathbf{W} , também possui redundâncias, que são retiradas após cada iteração, por meio da transformação:

$$\mathbf{W}' = \underline{\mathbf{W}} \mathbf{L}_I, \quad (33)$$

com a matriz \mathbf{W}' possuindo duas colunas, cada qual com $2L$ linhas (as primeiras L linhas da q -ésima coluna contendo as informações das primeiras L linhas da primeira coluna de \mathbf{W}_{1q} e as restantes armazenando as primeiras L linhas da primeira coluna de \mathbf{W}_{2q}). O gradiente natural de \mathbf{W}' é:

$$\nabla \mathbf{W}' = \mathbf{G}'_{8L} \nabla_{\underline{\mathbf{W}}}^{GN} \mathfrak{S} \mathbf{L}_I, \quad (34)$$

com

$$\mathbf{G}'_{4L} = \mathbf{F}_{4L \times 4L} \begin{bmatrix} \mathbf{I}_{L \times L} & \mathbf{0}_{L \times L} \\ \mathbf{0}_{3L \times L} & \mathbf{0}_{3L \times 3L} \end{bmatrix} \cdot \mathbf{F}_{4L \times 4L}^{-1}, \quad (35)$$

e

$$\mathbf{G}'_{8L} = \begin{bmatrix} \mathbf{G}'_{4L} & \mathbf{0}_{4L \times 4L} \\ \mathbf{0}_{4L \times 4L} & \mathbf{G}'_{4L} \end{bmatrix}. \quad (36)$$

V. ESTATÍSTICA DE SEGUNDA ORDEM \times ESTATÍSTICA DE ORDEM SUPERIOR

Os procedimentos descritos anteriormente envolvem estatísticas de segunda ordem (SOS - *Second Order Statistics*), as quais basicamente implementam simples desconexões das saídas. Como é sabido, estatísticas de segunda ordem não são suficientes para separação de fontes estacionárias porque fornecem menos restrições do que necessário, tornando o problema indeterminado. Porém, para fontes não-estacionárias (como é o caso de sinais de fala), quando contemplamos vários blocos das saídas com diferentes atrasos (TDD - *Time-Delayed Decorrelation*) [11], ou seja,

$$\langle \Phi(\mathbf{Y}_i) \mathbf{Y}_j \rangle = \langle \mathbf{Y}_i(m) \mathbf{Y}_j(m + \tau_i) \rangle = 0, \quad (37)$$

sendo $\langle \Phi(\mathbf{Y}_i) \rangle = \mathbf{Y}_i$, a desconexão consegue gerar restrições suficientes e resolver razoavelmente o problema.

Outra forma possível para separar misturas convolutivas é baseada em estatísticas de mais alta ordem (HOS - *Higher Order Statistics*). De acordo com [2], para sinais de fala podemos aproximar a Eq. 7 para $\langle \Phi(\mathbf{Y}_i) \rangle = \tanh(\mathbf{Y}_i)$ e aplicar na Eq. (37). Este procedimento pode ser visto como um caso de desconexão não-linear [12], como mostra a expansão de Taylor da função $\tanh(\mathbf{Y}_i)$ dada por:

$$\left\langle \left(\mathbf{Y}_i - \frac{\mathbf{Y}_i^3}{3} + \frac{2\mathbf{Y}_i^5}{15} - \frac{17\mathbf{Y}_i^7}{315} + \dots \right) \mathbf{Y}_j \right\rangle = 0. \quad (38)$$

Desta forma, a utilização de HOS para separação de misturas no domínio do tempo (ver Seção III) pode ser facilmente implementada através da Eq. (21) considerando [13]:

$$\mathbf{R}_{yy} = \langle \Phi(\mathbf{Y}) \mathbf{Y} \rangle. \quad (39)$$

VI. RESULTADOS EXPERIMENTAIS

Para efeitos de simulação, utilizamos dois sinais de fala, amostrados em 8kHz, um de voz feminina e outro de voz masculina, e dois microfones ($P = Q = 2$). A duração dos sinais é de mais ou menos 10 segundos. Os coeficientes dos filtros de mistura (de comprimento $M = 8$), que foram arbitrariamente escolhidos e que serão utilizados em todos os experimentos, são mostrados na Fig. 2. O desempenho do sistema é avaliado a partir da razão média Sinal/Interferência (sigla em inglês, SIR) [8], dada em dB.

A. Primeiro Experimento

Neste experimento, verificaremos a influência do aumento do tamanho dos filtros de separação no desempenho do sistema, usando apenas estatísticas de segunda ordem. Utilizando o algoritmo no domínio do tempo (Eq. (21)), a Fig. 3 mostra a evolução do SIR para filtros de separação de comprimentos $L = 4, 8, 32$ e 64 , com número de intervalos $D = L$, comprimento dos blocos $N = 2D$ e passo de adaptação $\mu = 4 \times 10^{-3}$. A Fig. 4 evidencia o SIR máximo em relação a diversos comprimentos de filtros. Analisando estes resultados, observamos que aumentar o comprimento dos filtros de separação implica uma clara melhora do SIR. Com um filtro de comprimento 64 obtemos um SIR acima de 20dB, onde a tarefa de separação se revela auditivamente quase perfeita. Devido ao maior custo computacional do algoritmo de separação no domínio da frequência do tipo banda larga (este fato é o inverso do que ocorre quando o algoritmo é do tipo banda estreita, de acordo com a Seção I) utilizamos comprimentos menores, a saber: $L = 2, 4, 6, 8$ e 16 (com $\mu = 1,25 \times 10^{-4}$), cujos resultados, apresentados na Fig. 5, confirmam as observações anteriores.

B. Segundo Experimento

Nesta subseção, analisaremos o impacto que o uso de estatísticas de ordem mais elevadas provoca no desempenho do sistema de separação. A estatística de ordem mais alta foi testada utilizando-se $\Phi(x) = \tanh(\alpha x)$, função clássica por ter sido utilizada em muitos artigos e que não aumenta em demasia o custo computacional. Nesta subseção, sempre utilizamos filtros de comprimento 8 e algoritmo de separação no domínio do tempo, por ser mais rápido. Para diferentes valores de α , o SIR máximo encontrado é mostrado na Fig. 6, para $\mu = 5 \times 10^{-3}$. Nota-se a diminuta influência da estatística de mais alta ordem no SIR máximo obtido (observe a escala do gráfico). Porém esse resultado não persiste quando utilizamos sinais de comprimento menor (definido em porcentagem do comprimento do sinal total), quando a estatística de mais alta ordem aumenta o SIR em até de 1,4 dB. À medida que possuímos mais informação (comprimento do sinal maior), a vantagem da estatística de mais alta ordem possui uma clara tendência a se reduzir, como se vê na Fig. 7 (escolhido um $\alpha = 0.45$).

VII. CONCLUSÕES

No problema de separação de fontes misturadas de forma convolutiva, é comum utilizar-se filtros de separação de comprimento menor ou igual aos comprimentos dos filtros de mistura. Desconhecemos artigos que tratam como interessante a possibilidade de filtros de separação maiores, sendo que em [14] há uma análise que argumenta o contrário; ou seja, que aumentar o comprimento dos filtros de separação piora o SIR na saída do sistema. Esse resultado é muitas vezes válido para algoritmos no domínio da frequência do tipo banda estreita, que não são os algoritmos investigados neste trabalho. Os resultados mostrados neste artigo indicam que a utilização de filtros de separação de comprimentos maiores que os comprimentos dos filtros de mistura introduzem, ao

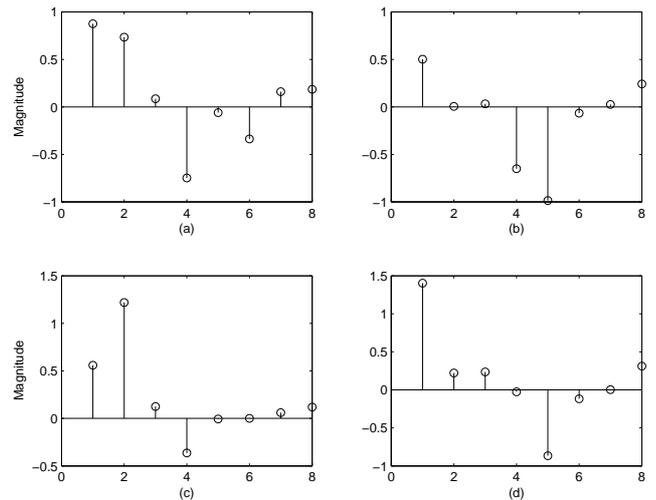


Fig. 2. Coeficientes dos filtros de mistura: (a) h_{11} , (b) h_{12} , (c) h_{21} e (d) h_{22} .

menos em alguns casos, melhoras não desprezíveis no SIR que, dependendo da aplicação, podem se mostrar cruciais. O fato de a complexidade aumentar com filtros de separação maiores pode tornar este resultado em geral pouco prático. Porém, como é sabido, a complexidade dessa implementação em subbandas pode ser muito menor. Eis, então, um forte argumento a favor da separação em subbandas: a potencialidade de se implementar filtros de separação de tamanho maior que o dos filtros de mistura com um custo computacional relativamente baixo, obtendo assim um desempenho melhor. Planejamos implementar esta abordagem no futuro. Quanto ao uso de estatísticas de ordem mais elevada, Kellerman et al. [13] comparou o uso de um modelo SIRP (do inglês *Spherically Invariant Random Processes*, o qual gera uma função $\Phi(\cdot)$ bem mais complexa e pesada computacionalmente que a \tanh aqui utilizada) com estatísticas de segunda ordem, obtendo valores máximos de SIR equivalentes. Neste artigo verificamos que, quando não há muita informação disponível (o que é crítico na implementação *on-line*, por exemplo), mesmo um modelo simples de estatísticas de ordem mais elevada pode obter um ganho de até 1,4 dB no SIR médio do sistema, à custa de um pequeno aumento de complexidade.

REFERÊNCIAS

- [1] A. Hyvärinen, J. Karhunen, e E. Oja, "Independent Component Analysis", John Wiley & Sons, 2001.
- [2] S. Amari, "Natural gradient works efficiently in learning". Proc. ICASSP, 2002.
- [3] K. Matusoka, "Minimal Distortion Principle for Blind Source Separation". SICE, pp. 2138-2143, 2002.
- [4] J. Anemüller e B. Kollmeier, "Amplitude Modulation decorrelation for convolutive blind source separation". Proc. ICA, pp. 215-220, 2000.
- [5] N. Murata, S. Ikeda e A. Ziehe, "An approach to blind source separation based on temporal structure of speech signals". Neurocomputing, vol. 41, pp. 1-24, 2001.
- [6] S. Kurita, H. Saruwatari, S. Kajita, K. Takeda e F. Itakura, "Evaluation of blind signal separation method using directivity pattern under reverberant conditions". Proc. ICASSP, pp. 3140-3143, 2002.
- [7] M. Z. Ikram e D. R. Morgan, "A beamforming approach to permutation alignment for multichannel frequency-domain blind speech separation". Proc. ICASSP, pp. 881-884, 2002.

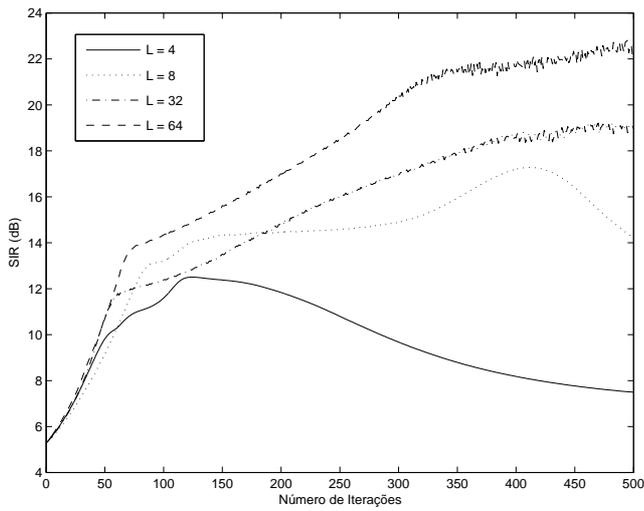


Fig. 3. Evolução do SIR para diferentes tamanhos dos filtros de separação.

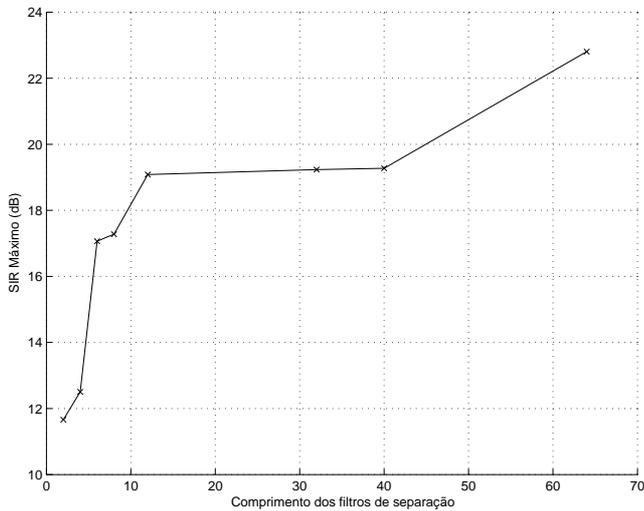


Fig. 4. SIR máximo para diferentes tamanhos dos filtros de separação.

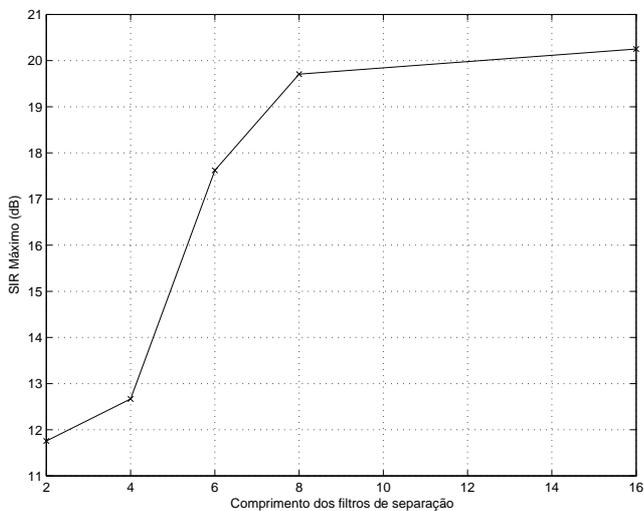


Fig. 5. SIR máximo para diferentes tamanhos dos filtros de separação (algoritmo no domínio da frequência).

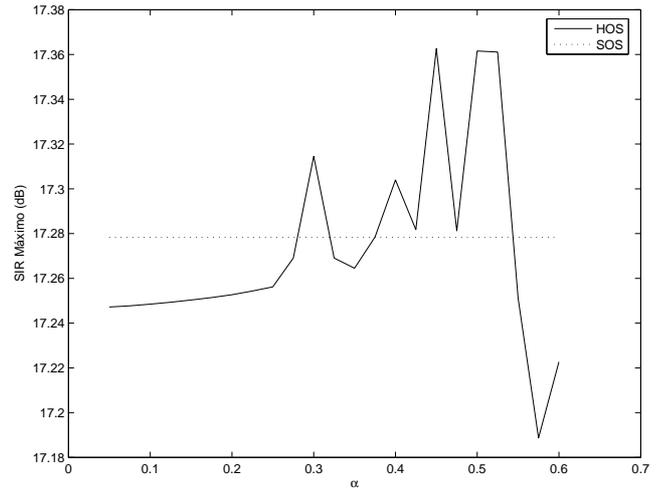


Fig. 6. SIR máximo para diferentes valores do parâmetro α .

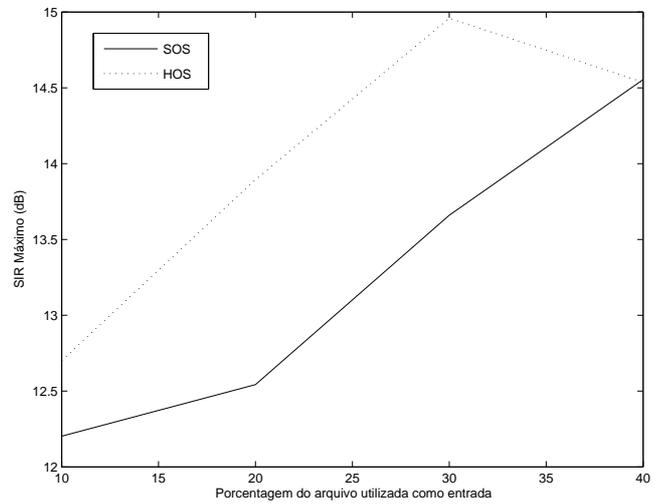


Fig. 7. SIR máximo para diferentes comprimentos de sinal.

[8] H. Sawada, R. Mukai, S. Araki e S. Makino, "A Robust and Precise Method for Solving the Permutation Problem of Frequency-Domain Blind Source Separation", IEEE Transactions on Speech And Audio Processing, vol. 12, pp. 530-538, 2004.

[9] R. Aichner, H. Buchner, e W. Kellerman, "Exploiting Narrowband Efficiency for Broadband Convolutional Blind Source Separation", EURASIP Journal on Advances in Signal Processing, 2007.

[10] H. Buchner, R. Aichner, e W. Kellerman, "A Generalization of Blind Source Separation Algorithms for Convolutional Mixtures Based on Second-Order Statistics". IEEE Transactions on Speech and Audio Processing, vol. 13, no. 1, 2005.

[11] K. Matsuoka and M. Ohya and M. Kawamoto, "A Neural Net for Blind Separation of Nonstationary Signals". Neural Networks, vol. 08, no. 3, p. 411-419, 1995.

[12] P. Smaragdis, "Blind Separation of convolved mixtures in the Frequency Domain". Neurocomputing, vol. 22, p. 21-34, 1998.

[13] H. Buchner, R. Aichner, e W. Kellerman, "Audio Signal Processing-Chapter 10, Kluwer Academic Publishers,2004.

[14] S. Araki, R. Mukai, S. Makino, T. Nishikawa e H. Saruwatari, "The Fundamental Limitation of Frequency Domain Blind Source Separation for Convolutional Mixtures of Speech", IEEE Transactions on Speech and Audio Processing, vol. 11, no. 2, 2003.