

# Separação Cega de Fontes para Misturas Convolutivas Usando um Banco de Filtros Não-Uniforme

Paulo Bulkool Batalheiro, Mariane Rembold Petraglia e Diego Barreto Haddad.

**Resumo**—Estruturas adaptativas em subbandas têm sido propostas com o objetivo de acelerar a velocidade de convergência e/ou reduzir a complexidade computacional em aplicações que requerem um grande número de coeficientes adaptativos. Neste artigo propomos um método de separação cega de fontes em misturas convolutivas, que atua no domínio do tempo, sendo implementado por um banco de filtros não-uniforme com coeficientes reais e um novo esquema de normalização para o algoritmo de aprendizagem. Devido aos filtros de separação nas subbandas trabalharem em taxas de amostragem reduzidas, o método proposto apresenta menor complexidade computacional e maior taxa de convergência quando comparado ao algoritmo correspondente em banda cheia.

**Palavras-Chave**—Separação Cega de Fontes, Misturas Convolutivas, Processamento em Subbandas, Banco de Filtros Não-uniforme.

**Abstract**—Adaptive subband structures have been proposed with the objective of increasing the convergence speed and/or reducing the computational complexity of adaptation algorithms for applications which require a large number of adaptive coefficients. In this paper we propose a blind source separation method for convolutive mixtures which employs a real-coefficient non-uniform filter bank and a new normalization scheme for the adaptation algorithm. Since the separation filters in the subbands work at reduced sampling rates, the proposed method presents smaller computational complexity and faster convergence rate when compared to the corresponding fullband algorithm.

**Keywords**—Blind Source Separation, Convolutional Mixtures, Subband Processing, Non-uniform Filter Bank.

## I. INTRODUÇÃO

Separação cega de fontes (BSS, do inglês *Blind Source Separation*), técnica que tem sido bastante investigada nas últimas décadas, possibilita a extração do sinal de uma fonte desejada  $s_q(n)$  a partir dos sinais misturados de mais de uma fonte  $x_p(n)$  sem a necessidade de qualquer informação a respeito da posição das fontes originais, de suas estruturas espectrais ou de como foi realizada a mistura. São utilizadas hipóteses mínimas que, exatamente por serem fracas, aplicam-se a uma grande variedade de configurações. Entre estas hipóteses, destacamos as mais comuns: fontes independentes com distribuição não-gaussiana (frequentemente supergaussiana) e distribuição uniforme dos coeficientes da matriz de mistura.

Paulo Bulkool Batalheiro, Departamento de Eletrônica e Telecomunicações, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brasil, Mariane Rembold Petraglia, PEE/COPPE, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brasil, Diego Barreto Haddad, CEFET, Nova Iguaçu, Brasil, E-mails: bulkool@pads.ufrj.br, mariane@pads.ufrj.br, diego@pads.ufrj.br.

Estas misturas podem ser classificadas como não-lineares ou lineares, e estas divididas em instantâneas e convolutivas. Este artigo trata apenas o caso mais realístico de misturas lineares convolutivas envolvendo sinais de fala humana, permitindo assim modelar-se a reverberação sempre presente em ambientes ecóticos. Nestes casos, tipicamente são necessários filtros FIR (do inglês *Finite Impulse Response*) de separação da ordem de milhares de coeficientes, o que torna a tarefa de separação muito complexa. Para tentar resolver o problema, vários métodos no domínio do tempo e no domínio da frequência, baseados na análise de componentes independentes (ICA, do inglês *Independent Component Analysis*), têm sido propostos na literatura [1]–[6].

Algumas dessas soluções empregam filtros FIR de separação e tentam estimar via ICA seus coeficientes diretamente no domínio do tempo. Em situações reais, estes filtros têm milhares de coeficientes, e por isso, estas soluções demandam grande esforço computacional, apresentam tempo de convergência muito elevado e um branqueamento indesejado nas estimativas das fontes [1], [2]. Para contornar estas dificuldades, foram propostos métodos que transformam o problema de BSS para o domínio da frequência, em que as convoluções tornam-se produtos, e o problema de BSS pode ser tratado como um caso de mistura instantânea em cada raia de frequência [3], [4]. As desvantagens, nesse caso, são os problemas de escalamento e permutação entre as raias, além do uso de janelas grandes, necessárias para estimar filtros longos. Neste caso, devido à não-estacionalidade do ambiente acústico (sistema de misturas) e dos sinais de fala, não há amostras suficientes no domínio da frequência para uma correta estimativa das estatísticas em cada raia. Estas desvantagens podem comprometer severamente o desempenho destes algoritmos. Existem, ainda, soluções mistas que combinam as soluções no domínio da frequência e no domínio do tempo para melhorar o desempenho da BSS com menor complexidade computacional [5]. Neste cenário, algumas soluções em subbandas têm sido propostas, principalmente pela possibilidade de trabalhar com filtros de separação independentes e de ordem reduzida em cada canal, além da redução da taxa de amostragem. Estes métodos utilizam, normalmente, banco de filtros uniformes subamostrados com coeficientes complexos [6].

Neste artigo propomos a BSS em subbandas empregando um banco de filtros subamostrado não-uniforme com coeficientes reais e filtros FIR de separação de ordem reduzida. Os coeficientes dos filtros de separação nas diversas subbandas são ajustados independentemente por um algoritmo adaptativo

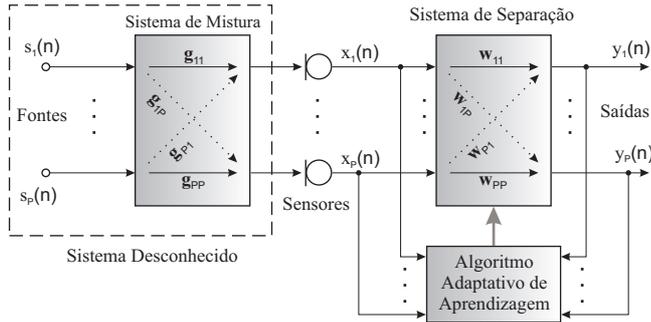


Fig. 1. Configuração MIMO linear para BSS em banda cheia.

no domínio do tempo, baseado em estatísticas de segunda ordem [1], com um novo esquema de normalização, o qual proporciona aceleração da taxa de convergência e redução da complexidade computacional quando comparado com o esquema de normalização original. A estrutura proposta emprega processamento em múltiplas taxas de amostragem, com taxas menores nas subbandas mais baixas onde, tipicamente em sinais de fala, há maior concentração de energia. Outra vantagem do algoritmo proposto é o uso de filtros com coeficientes reais, que são atrativos em implementações em DSPs.

Na Seção II descreveremos o problema de separação cega de fontes para o caso de misturas convolutivas. Na Seção III apresentaremos o algoritmo em banda cheia proposto em [1] e o novo esquema de normalização proposto. A estrutura em subbandas não-uniforme será descrita na Seção IV. Na Seção V apresentaremos os resultados experimentais, comparando o desempenho dos algoritmos BSS em banda cheia e em subbandas. Na Seção VI serão apresentadas as conclusões.

## II. BSS PARA MISTURAS CONVOLUTIVAS

Em sistemas de teleconferência, as fontes originais são sinais de fala e a mistura convolutiva das fontes é provocada pela reverberação do auditório. A Figura 1 ilustra um sistema de separação cega de fontes determinado, ou seja, com número de fontes igual ao número de sensores. Considerando que o sistema desconhecido de mistura pode ser modelado por um conjunto de filtros FIR de comprimento  $U$  (mistura linear convolutiva), os sinais capturados pelos microfones  $x_p(n)$  podem ser escritos como

$$x_p(n) = \sum_{q=1}^P \sum_{k=0}^{U-1} g_{qp}(k) s_q(n-k) \quad (1)$$

sendo  $g_{qp}(n)$  a resposta ao impulso do filtro que modela o caminho acústico (eco) da  $q$ -ésima fonte até o  $p$ -ésimo sensor, e  $P$  o número de fontes e de microfones.

No problema de BSS, para desfazer a mistura a partir dos sinais capturados pelos microfones, são estimados, através de um algoritmo de aprendizagem adaptativo baseado na análise de componentes independentes, os coeficientes dos filtros FIR de separação  $w_{pq}(n)$  de forma a maximizar a independência dos sinais de saída  $y_q(n)$ , os quais podem ser descritos da

seguinte forma:

$$y_q(n) = \sum_{p=1}^P \sum_{k=0}^{S-1} w_{pq}(k) x_p(n-k) \quad \text{para } q = 1, \dots, P \quad (2)$$

sendo  $S$  o comprimento dos filtros de separação.

## III. ALGORITMO BSS EM BLOCOS NO DOMÍNIO DO TEMPO

Para sinais coloridos e não-estacionários, como sinais de fala, o problema de BSS pode ser resolvido diagonalizando a matriz de correlação de saída considerando múltiplos blocos em diferentes instantes de tempo (TDD - *Time-Delayed Decorrelation*) [7]. Nesta seção vamos revisar a solução tipo banda larga baseada em estatística de segunda ordem proposta em [1].

Estendendo a formulação dos sinais de saída  $y_q$  no instante  $n$  (Eq. (2)) para uma forma matricial, podemos descrevê-los como

$$y_q(n) = \sum_{p=1}^P \mathbf{x}_p^T(n) \mathbf{w}_{pq}, \quad (3)$$

sendo  $\mathbf{x}_p(n) = [x_p(n), x_p(n-1), \dots, x_p(n-S+1)]^T$  o vetor que contém as últimas  $S$  amostras dos sinais capturados pelo  $p$ -ésimo microfone e  $\mathbf{w}_{pq}(n) = [w_{pq,0}, w_{pq,1}, \dots, w_{pq,S-1}]^T$  o vetor formado pelos  $S$  coeficientes do filtro FIR que modela o caminho do eco entre o  $p$ -ésimo sensor e a estimativa da  $q$ -ésima fonte.

Para generalizar a formulação com múltiplos blocos, são definidos o parâmetro  $D$  como o número de blocos atrasados no tempo levados em conta na estimativa da correlação ( $1 \leq D \leq S$ ) e  $N$  como o tamanho de um bloco de sinal de saída ( $N \geq D$ ), sendo o vetor que contém um bloco de sinal de saída dado por

$$\mathbf{y}_q(m) = [y_q(mS), y_q(mS+1), \dots, y_q(mS+N-1)]^T. \quad (4)$$

Desta forma, os  $m$ -ésimos blocos atrasados no tempo correspondentes à  $q$ -ésima saída podem ser descritos por:

$$\mathbf{Y}_q(m) = \sum_{p=1}^P \mathbf{X}_p(m) \mathbf{W}_{pq}, \quad (5)$$

com

$$\mathbf{X}_p(m) = [\mathbf{U}_p^T(m), \mathbf{U}_p^T(m-1)], \quad (6)$$

$$\mathbf{U}_p^T(m) = \begin{bmatrix} x_p(mS) & \dots & x_p(mS-S+1) \\ x_p(mS+1) & \dots & x_p(mS-S+2) \\ \vdots & \ddots & \vdots \\ x_p(mS+N-1) & \dots & x_p((m-1)S+N) \end{bmatrix}. \quad (7)$$

As dimensões das matrizes descritas acima são  $N \times D$ ,  $N \times 2S$  e  $N \times S$ , respectivamente. A matriz  $\mathbf{W}_{pq}$ , uma matriz do tipo

Sylvester de dimensões  $2S \times D$ , é definida como:

$$\mathbf{W}_{pq} = \begin{bmatrix} w_{pq,0} & 0 & \cdots & 0 \\ w_{pq,1} & w_{pq,0} & \ddots & \vdots \\ \vdots & w_{pq,1} & \ddots & 0 \\ w_{pq,S-1} & \vdots & \ddots & w_{pq,0} \\ 0 & w_{pq,S-1} & \ddots & w_{pq,1} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & w_{pq,S-1} \\ 0 & \cdots & 0 & 0 \\ \vdots & \cdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 \end{bmatrix}. \quad (8)$$

Combinando todos os canais, a Eq. (5) pode ser escrita concisamente como:

$$\mathbf{Y}(m) = \mathbf{X}(m)\mathbf{W}, \quad (9)$$

sendo

$$\mathbf{Y}(m) = [\mathbf{Y}_1(m), \dots, \mathbf{Y}_P(m)], \quad (10)$$

$$\mathbf{X}(m) = [\mathbf{X}_1(m), \dots, \mathbf{X}_P(m)], \quad (11)$$

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_{11} & \cdots & \mathbf{W}_{1P} \\ \vdots & \ddots & \vdots \\ \mathbf{W}_{P1} & \cdots & \mathbf{W}_{PP} \end{bmatrix}. \quad (12)$$

As matrizes acima têm dimensões  $N \times PD$ ,  $N \times 2SP$  e  $2SP \times PD$ , respectivamente.

Definida a formulação matricial, a função custo para a BSS baseada na TDD é dada por:

$$\mathfrak{S} = \sum_{i=1}^b \frac{1}{b} \{ \log(\det(\text{bdiag}(\mathbf{Y}^T(i)\mathbf{Y}(i)))) - \log(\det(\mathbf{Y}^T(i)\mathbf{Y}(i))) \}, \quad (13)$$

sendo  $b$  o número de blocos considerados na otimização e  $\text{bdiag}(\cdot)$  o operador que atua sobre uma matriz de submatrizes, anulando todas as submatrizes que não estejam situadas na diagonal principal da matriz  $\mathbf{A}$ .

Aplicando o método do gradiente natural à função custo (Eq. (13)), temos

$$\nabla_{\mathbf{W}}^{GN} \mathfrak{S}(m) = \frac{2}{b} \sum_{i=1}^b \mathbf{W} \{ \mathbf{R}_{yy} - \text{bdiag}(\mathbf{R}_{yy}) \} \{ \text{bdiag}(\mathbf{R}_{yy}) \}^{-1}, \quad (14)$$

sendo  $\mathbf{R}_{yy}$  a matriz de ordem  $PD \times PD$  dada por:

$$\mathbf{R}_{yy}(m) = \mathbf{Y}^H(m)\mathbf{Y}(m). \quad (15)$$

O algoritmo *off-line* tipo batelada (*batch*) para ajustar os coeficientes dos filtros de separação, considerando um sistema com duas fontes e dois sensores (TITO - *Two Input Two Output*), é dado por:

$$\mathbf{W}(i) = \mathbf{W}(i-1) - \frac{2\mu}{b} \sum_{m=1}^b \begin{bmatrix} \mathbf{W}_{12}\mathbf{R}_{y2y1}\mathbf{R}_{y1y1}^{-1} & \mathbf{W}_{11}\mathbf{R}_{y1y2}\mathbf{R}_{y2y2}^{-1} \\ \mathbf{W}_{22}\mathbf{R}_{y2y1}\mathbf{R}_{y1y1}^{-1} & \mathbf{W}_{21}\mathbf{R}_{y1y2}\mathbf{R}_{y2y2}^{-1} \end{bmatrix} \quad (16)$$

sendo  $\mathbf{R}_{pq}$ , de dimensão  $D \times D$ , uma submatriz de  $\mathbf{R}_{yy}$  (Eq. (15)),  $i$  o número de iterações (*off-line*) e  $\mu$  o passo de adaptação do algoritmo.

Devido às redundâncias na matriz  $\mathbf{W}_{pq}$  (Eq. (8)) e por razões de convergência [1], atualizamos a cada iteração somente os  $S$  primeiros elementos de sua primeira coluna, que são suficientes para formar toda a matriz do tipo Sylvester. Para reduzir a complexidade computacional do algoritmo, o cálculo do fator de normalização  $\mathbf{R}_{qq}^{-1}(m)$  pode ser simplificado desprezando-se as submatrizes fora da diagonal principal da matriz de autocorrelação [9], ou seja,

$$\mathbf{R}_{qq}(m) \approx \text{diag}\{\mathbf{R}_{qq}(m)\} = \sigma_{y_q}^2(m)\mathbf{I}. \quad (17)$$

Assim, a inversão desta matriz fica reduzida ao cálculo do inverso de cada elemento de sua diagonal.

Com o objetivo de reduzir ainda mais o custo computacional, propomos um novo esquema que reduz o fator de normalização a um escalar. Neste caso,

$$\mathbf{R}_{qq}(m) \approx \mathbf{y}_q^T(m)\mathbf{y}_q(m)\mathbf{I}, \quad (18)$$

com  $\mathbf{y}_q(m)$  dado pela Eq. (4), sendo sua inversa reduzida ao inverso da potência de um único bloco do sinal de saída  $y_q(n)$ .

#### IV. MÉTODO BSS EM SUBBANDAS

Nesta seção vamos investigar o uso de uma estrutura em subbandas em conjunto com o algoritmo BSS em blocos no domínio do tempo apresentado na seção anterior. A idéia é explorar as características de melhoria da taxa de convergência e de redução da complexidade computacional inerentes às estruturas em subbandas. Uma estrutura que emprega banco de filtros uniforme com coeficientes complexos aplicada à BSS foi usada em [6]. Neste artigo, propomos a utilização de uma estrutura em subbandas não-uniforme, baseada em [10], que emprega decomposição em frequência por oitavas com diferentes taxas de amostragem nas diferentes subbandas, ou seja, bandas mais estreitas e taxas menores durante a adaptação dos coeficientes em faixas de frequência onde o sinal de fala concentra maior energia.

A Figura 2 mostra um esquema TITO para BSS em subbandas considerando um banco de filtros não-uniforme de  $M$  canais. Esta estrutura é uma versão modificada, em relação à estrutura apresentada em [10], onde os sinais presentes nas entradas dos filtros de separação de cada subbanda  $w_{pq}^k(n)$  são decimados pela metade do fator de decimação crítico com o objetivo de reduzir o efeito do *aliasing* durante o processo de adaptação dos coeficientes. As decimações dos sinais nas saídas dos filtros de separação de cada subbanda por um fator igual a 2 restauram a taxa de amostragem crítica da estrutura antes da etapa de reconstrução do sinal de saída. Os filtros de análise ( $H_i(z)$ ) e síntese ( $F_i(z)$ ) equivalentes de um banco de filtros em árvore não-uniforme de  $M$ -canais são dados por [11]

$$\begin{aligned} H_0(z) &= \prod_{j=0}^{M-2} H^{0,j}(z^{2^j}), & F_0(z) &= \prod_{j=0}^{M-2} F^{0,j}(z^{2^j}), \\ H_k(z) &= z^{-\Delta_k} H^{1,M-1-k}(z^{2^{M-1-k}}) \prod_{j=0}^{M-k-2} H^{0,j}(z^{2^j}), \\ F_k(z) &= z^{-\Delta_k} F^{1,M-1-k}(z^{2^{M-1-k}}) \prod_{j=0}^{M-k-2} F^{0,j}(z^{2^j}), \end{aligned} \quad (19)$$

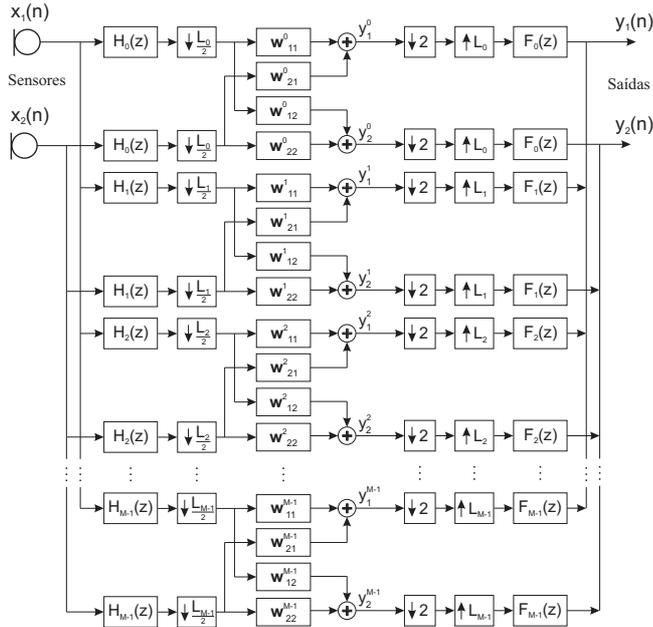


Fig. 2. Configuração TITO linear para BSS em subbandas.

sendo  $H^{0,j}(z)$  e  $F^{0,j}(z)$  os filtros passa-baixas,  $H^{1,j}(z)$  e  $F^{1,j}(z)$  os filtros passa-altas, utilizados no  $j$ -ésimo estágio de uma estrutura binária, e projetados para obter reconstrução perfeita [12]. Os filtros de análise têm ordens dadas por

$$N_{H_0} = \sum_{j=0}^{M-2} 2^j N_{H^{0,j}},$$

$$N_{H_k} = \sum_{j=0}^{M-k-2} 2^j N_{H^{0,j}} + 2^{M-k-1} N_{H^{1,j}}, \quad (20)$$

sendo  $N_H^{0,j}$  e  $N_H^{1,j}$  as ordens de  $H^{0,j}(z)$  e  $H^{1,j}(z)$ , respectivamente. O número de coeficientes de cada subfiltro de separação  $w_{pq}^k(n)$  deve ser no mínimo [10]

$$S_k = 2 \left\lceil \frac{S-1+N_{F_k}}{L_k} \right\rceil + 1, \quad (21)$$

sendo  $N_{F_k}$  a ordem do  $k$ -ésimo filtro do banco de síntese correspondente, e  $L_k$  os fatores de decimação, que são dados por  $L_0 = 2^{M-1}$  e  $L_k = 2^{M-k}$  para  $k = 1, \dots, M-1$ . Os atrasos (Eq. (19)) necessários para compensar a introdução de atrasos diferentes nas subbandas são  $\Delta_k = N_{H_0} - N_{H_k}$ , e o atraso entrada-saída total introduzido pela estrutura é  $\Delta_D = N_{H_0}$ .

Para ajustar os coeficientes de cada subfiltro de separação usamos o algoritmo em banda cheia (Eq. (16)) adaptado para subbandas, sendo a equação de atualização dos coeficientes do filtro da  $k$ -ésima banda dada por:

$$\mathbf{W}^k(i) = \mathbf{W}^k(i-1) - \frac{2}{b_k} \sum_{m=1}^{b_k} \begin{bmatrix} \mathbf{W}_{12}^k \mathbf{R}_{21}^k \mathbf{R}_{11}^{k-1} & \mathbf{W}_{11}^k \mathbf{R}_{12}^k \mathbf{R}_{22}^{k-1} \\ \mathbf{W}_{22}^k \mathbf{R}_{21}^k \mathbf{R}_{11}^{k-1} & \mathbf{W}_{21}^k \mathbf{R}_{12}^k \mathbf{R}_{22}^{k-1} \end{bmatrix} \cdot \begin{bmatrix} \mu_1^k \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mu_2^k \mathbf{I} \end{bmatrix} \quad (22)$$

sendo

$$\mathbf{R}_{pq}^k(m) = \mathbf{Y}_p^{kH}(m) \mathbf{Y}_q^k(m) \quad (23)$$

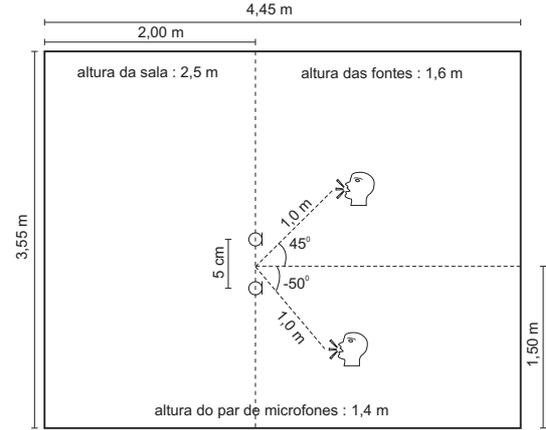


Fig. 3. Cenário virtual usado nos experimentos.

e

$$\mathbf{Y}_q^k(m) = \begin{bmatrix} y_q^k(mS_k) & \dots & y_q^k(mS_k - D_k + 1) \\ y_q^k(mS_k + 1) & \ddots & y_q^k(mS_k - D_k + 2) \\ \vdots & \ddots & \vdots \\ y_q^k(mS_k + N_k - 1) & \dots & y_q^k(mS_k - D_k + N_k) \end{bmatrix}. \quad (24)$$

As matrizes acima têm ordens  $D_k \times D_k$  (com  $1 \leq D_k \leq S_k$ ) e  $N_k \times D_k$  (com  $N_k \geq D_k$ ), respectivamente,  $b_k$  é o número de blocos,  $N_k$  é o tamanho de cada bloco,  $\mu_q^k$  é o  $q$ -ésimo passo de adaptação da  $k$ -ésima subbanda,  $i$  é o número de iterações (*off-line*) para todas as subbandas e  $y_q^k$  é o sinal da  $q$ -ésima saída na  $k$ -ésima subbanda (ver Figura 2).

## V. RESULTADOS EXPERIMENTAIS

Todos os experimentos foram realizados usando dois sinais de fala com 10 segundos de duração e frequência de amostragem  $F_s = 8\text{kHz}$ : uma voz feminina em inglês e uma voz masculina em português. Estas sentenças foram convoluídas com as respostas impulsivas sintéticas obtidas para a sala mostrada na Figura 3, com tempo de reverberação de 250 ms. Em nossos experimentos as misturas foram realizadas considerando diferentes condições de reverberação, reamostrando as respostas impulsivas obtidas em [13] para  $F_s = 8\text{kHz}$  e truncando-as com o tamanho desejado (ou seja, considerando somente suas  $S$  primeiras amostras). O comprimento dos filtros de separação foram fixados, em todas as simulações, igual ao dos filtros de mistura ( $U = S$ ). Os algoritmos foram executados em um processador Intel Core 2 Duo 2GHz. Adotamos, para avaliação de desempenho, a relação sinal interferência (SIR), definida como

$$\text{SIR} = 10 \log_{10} \left( \frac{\text{SIR}_1 + \text{SIR}_2}{2} \right) \quad (25)$$

sendo  $\text{SIR}_i$  a relação entre a potência do sinal presente na saída  $y_i$  quando somente a fonte  $s_i$  está ativa e a potência do sinal presente na saída  $y_i$  quando somente a fonte  $s_i$  está inativa.

### A. Experimento 1

Neste experimento comparamos o desempenho do algoritmo em banda cheia usando dois diferentes esquemas de normalização: (a) normalização antiga (NA - Eq. (17)) e (b)

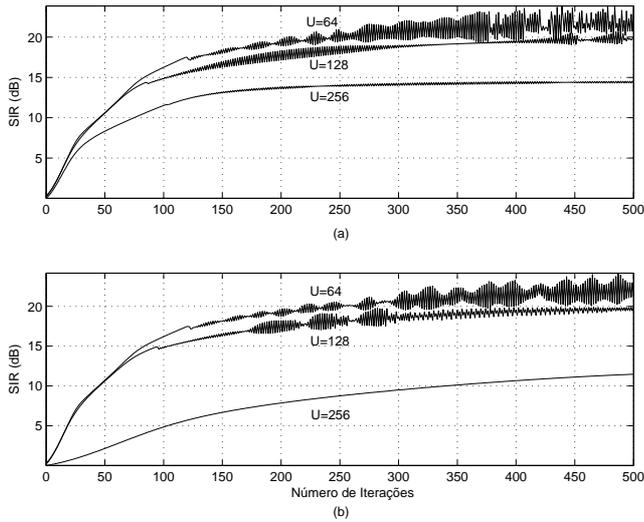


Fig. 4. Evolução do SIR para dois esquemas de normalização: (a) NN e (b) NA.

TABELA I  
TEMPO DE PROCESSAMENTO EM MINUTOS.

S=U	NA	NN
64	22	13
128	55	23
256	181	57

normalização nova (NN - Eq. (18)). A Figura 4 mostra a evolução do SIR considerando filtros de mistura de diferentes comprimentos  $U = 64, 128$  e  $256$ . O passo de adaptação em todos os casos foi  $\mu = 5 \times 10^{-3}$ , exceto para a normalização proposta com  $U = 256$ , onde foi usado  $\mu = 1 \times 10^{-3}$  por razões de convergência. A Tabela I mostra o tempo de processamento para as simulações exibidas na Figura 4. Analisando estes resultados fica evidente que em ambientes mais reverberantes (ou seja, para filtros de mistura de comprimentos maiores) o novo esquema de normalização reduz significativamente o tempo de processamento e melhora sensivelmente a velocidade de convergência do algoritmo em banda cheia.

### B. Experimento 2

Neste experimento confrontamos as estruturas em banda cheia e em subbandas, ambas com o esquema de normalização proposto na Seção III (Eq. (18)). A estrutura em subbandas não-uniforme foi implementada utilizando um banco de filtros com bandas em oitavas e com reconstrução perfeita para  $M = 4$  subbandas. A Tabela II apresenta os fatores de decimação  $L_k$ , as ordens dos filtros de análise  $N_{H_k}$  (que são iguais às ordens dos filtros de síntese  $N_{F_k}$ ), os atrasos  $\Delta_k$ , as ordens dos filtros de separação  $S_k$  e os passos de adaptação,  $\mu_1^k$  e  $\mu_2^k$ , usados nas simulações em subbandas com  $U = S = 1024$ . O passo de adaptação para o algoritmo em banda cheia foi o mesmo usado no Experimento 1, exceto para  $U = 1024$  onde foi usado  $\mu = 3 \times 10^{-3}$ . Estes passos resultaram na melhor convergência e foram obtidos experimentalmente por tentativa e erro. A Figura 5 mostra as respostas em frequência dos filtros de análise utilizados.

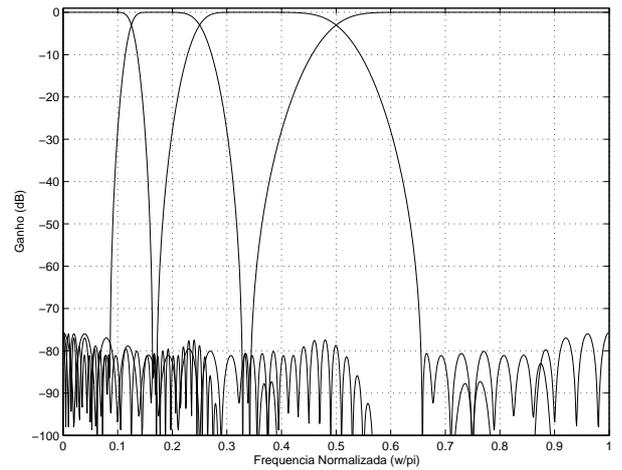


Fig. 5. Resposta em frequência dos filtros de análise  $H_k(z)$ .

TABELA II  
PARÂMETROS DA ESTRUTURA NÃO-UNIFORME PARA  $M = 4$ .

k	$L_k$	$N_{H_k}$	$\Delta_k$	$S_k$	$\mu_1^k$	$\mu_2^k$
0	8	441	0	366	$8.8 \times 10^{-3}$	$8 \times 10^{-3}$
1	8	441	0	366	$17.6 \times 10^{-3}$	$16 \times 10^{-3}$
2	4	189	252	606	$35.2 \times 10^{-3}$	$32 \times 10^{-3}$
3	2	63	378	575	$70.4 \times 10^{-3}$	$64 \times 10^{-3}$

Para reduzir o custo computacional sem degradação significativa no processo de separação, a ordem dos filtros de separação na banda de maior frequência ( $k = 3$ ), que são adaptados na mesma taxa dos sinais capturados pelos microfones, foi reduzida em relação a da Eq. (21) para  $S_3 = 2 \left\lfloor \frac{S}{2} - 1 + N_{F_k} \right\rfloor + 1$ . Esta redução foi possível devido às características de reverberação em altas frequências e também à concentração da energia dos sinais de fala em baixas frequências.

A Figura 6 mostra a evolução do SIR para os algoritmos em banda cheia (Eq. (16)) e em subbandas (Eq. (24)), considerando filtros de mistura de diferentes comprimentos ( $U = 256, 512$  e  $1024$ ). A Tabela III exhibe o SIR máximo em banda cheia e em cada uma das 4 bandas da estrutura não-uniforme. A Tabela IV mostra o tempo de processamento para as simulações da Figura 6. Analisando estas tabelas e os resultados da Figura 6, percebe-se que, à medida que aumenta a ordem do sistema de mistura (ambientes mais reverberantes), as vantagens da estrutura em subbandas sobre a estrutura em banda cheia tornam-se mais evidentes, resultando numa taxa de convergência mais rápida e num menor tempo de processamento, que pode ser ainda reduzido usando processamento paralelo para implementações em subbandas.

O desempenho da BSS na banda de frequência mais alta ( $k = 3$ ) foi sempre pior que nas demais subbandas (devido ao uso de filtros de separação de comprimento reduzido nesta banda), causando uma redução no SIR final da estrutura em subbandas. Porém, nesta faixa de frequência a potência dos sinais de fala é reduzida e auditivamente os resultados ficaram significativamente melhores que em banda cheia.

A Figura 7 mostra o espectro de potência da fonte original

TABELA III  
SIR MÁXIMO (EM dB).

Filtros de Mistura	SIR em subbandas				SIR
	$k = 0$	$k = 1$	$k = 2$	$k = 3$	
$U=S$					--
256	21.72	11.06	12.57	6.63	14.56
512	13.81	9.41	10.37	6.31	10.70
1024	12.40	8.73	9.39	6.44	9.92

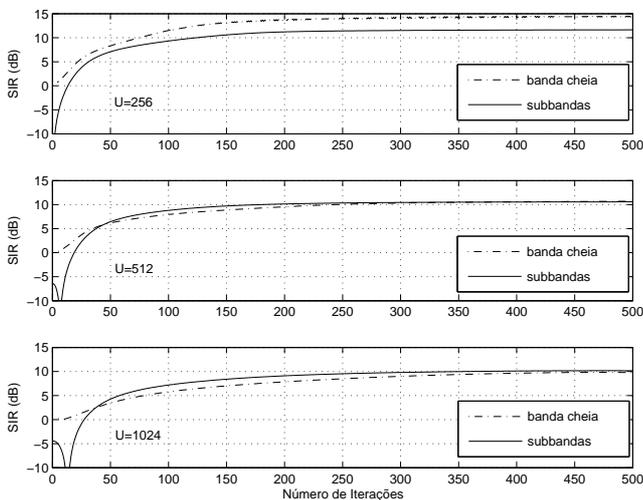


Fig. 6. Evolução do SIR para os algoritmos em banda cheia e em subbandas com filtros de misturas de diferentes comprimentos:  $U=256$ ,  $512$  e  $1024$  (de cima para baixo, respectivamente).

e de suas estimativas em banda cheia e em subbandas, para filtros de mistura de diferentes comprimentos. Estes resultados mostram a robustez dos algoritmos para o problema de branqueamento das fontes, e também para o escalamento dos sinais de saída.

VI. CONCLUSÃO

Neste artigo propomos um novo algoritmo em subbandas para separação cega de fontes que emprega um banco de filtros não-uniforme para decomposição dos sinais observados pelos sensores. Os filtros de separação têm diferentes comprimentos e são adaptados em diferentes taxas, ou seja, as bandas mais baixas são mais estreitas e operam em menores taxas de amostragem. A adaptação é realizada por um algoritmo baseado no gradiente natural com um novo esquema de normalização, que resulta na redução do custo computacional. Simulações computacionais foram realizadas com sinais de fala, mostrando as vantagens do novo esquema de normalização e da estrutura em subbandas em relação ao tempo de processamento, taxa de convergência e razão sinal interferência.

AGRADECIMENTOS

Os autores gostariam de agradecer o parcial apoio financeiro do CNPq.

REFERÊNCIAS

[1] H. Buchner and R. Aichner and W. Kellermann, "A Generalization of Blind Source separation Algorithms for Convolutional Mixtures Based on Second-Order Statistics," *IEEE Trans. on Speech and Audio Processing*, vol. 13, pp. 120-134, Jan. 2005.

TABELA IV  
TEMPO DE PROCESSAMENTO EM MINUTOS.

S=U	Banda Cheia	Subbandas
256	57	45
512	127	83
1024	377	167

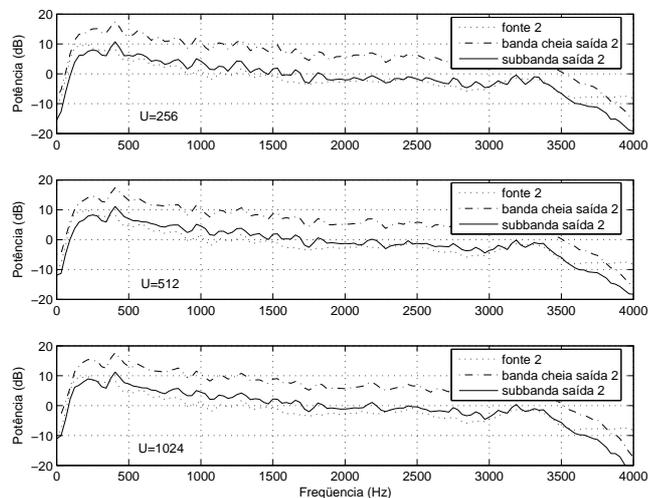


Fig. 7. Espectro da fonte original masculina, do sinal separado em banda cheia e em subbandas, com filtros de mistura de diferentes comprimentos:  $U=256$ ,  $512$  e  $1024$  (de cima para baixo, respectivamente).

[2] A. Hyvärinen and J. Karhunen and E. Oja, *Independent Component Analysis*. John Wiley & Sons, 2001.

[3] H. saruwatari and T. Kawamura and T. Nishikawa and A. Lee and K. Shikano, "Blind Source separation Based on a Fast-Convergence Algorithm Combining ICA and Beamforming," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 14, no. 2, pp. 666-678, Mar. 2006.

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

[5] T. Nishikawa and H. Saruwatari and K. Shikano, "Comparison of Time-Domain ICA, Frequency-Domain ICA and Multistage ICA for Blind Source Separation," in *Proc. Eur. Signal Processing Conf.*, 2006, pp. 15-18.

[6] S. Araki and S. Makino and R. Aichner and T. Nishikawa and H. Saruwatari, "Subband-Based Blind Separation for Convolutional Mixtures of speech," *IEICE Trans. Fundamentals*, vol. E88-A, pp. 3593-3603, Dec. 2005.

[7] H. Buchner and R. Aichner and W. Kellermann, "A Generalization of a Class of Blind Source Separation Algorithms for convolutional mixtures," in *Proc. Int. Symp. Independent Component Analysis Blind Signal Separation*, pp. 945-950, Apr. 2003.

[8] S. Amari and A. Cichocki and H. Yang, "A New Learning Algorithm for Blind Signal Separation," in *Advances in Neural Information Processing Systems*, vol. 8, pp. 757-763, 1996.

[9] R. Aichner and H. Buchner and W. Kellermann, "Exploiting Narrow-band Efficiency for Broadband Convolutional Blind Source Separation," *EURASIP Journal on Applied Signal*, vol. 2007, pp. 1-9, Sep. 2006.

[10] M. R. Petraglia and P. B. Batalheiro, "Nonuniform Subband Adaptive Filtering With Critical Sampling," *IEEE Trans. on Signal Processing*, vol. 56, no. 2, pp. 565-575, Feb. 2008.

[11] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*. Prentice-Hall, Englewood Cliffs, New Jersey, 1993.

[12] T. Q. Nguyen, "Digital filter banks design - quadratic constrained formulation," *IEEE Transactions on Signal Processing*, vol. 43, pp. 2103-2108, Sep. 1995.

[13] <http://sassec.gforge.inria.fr/>