

Nova Técnica de Separação Cega de Fontes em Subbandas

Jéssica Bartholdy Sanson, Mariane Rembold Petraglia

Resumo— Neste artigo é investigado o emprego de filtragem em subbandas em algoritmos de separação cega de fontes (BSS) para misturas convolutivas. O método proposto baseia-se na implementação no domínio da frequência do algoritmo TRINICON, que combina as vantagens de algoritmos de banda larga com a eficiência computacional de técnicas de banda estreita. São apresentadas modificações nas equações de atualização dos coeficientes do sistema de separação para realizá-las utilizando os sinais das misturas decompostos em subbandas, proporcionando uma melhora na taxa de convergência, maior razão sinal-interferência em regime estacionário e menor custo computacional em relação ao algoritmo em banda cheia. Resultados de simulações comprovam as vantagens do método proposto na separação de sinais de voz em ambientes com tempos de reverberação elevados.

Palavras-Chave— Separação Cega de Fontes, Misturas Convolutivas, Estruturas em Subbandas.

Abstract— In this paper we investigate the use of subband filtering in blind source separation (BSS) for algorithms convolutive mixtures. The proposed method is based on the frequency domain implementation of the TRINICON algorithm, which combines the advantages of wideband algorithms with the computational efficiency of narrow band techniques. The required modifications in the updating equations of the separation system coefficients are presented to perform them using the signals of the mixtures decomposed into subbands, providing an improvement in the convergence rate, higher signal-interference ratio at steady state and lower computational cost in relation to the fullband algorithm. Simulation results demonstrate the advantages of the proposed method in the separation of voice signals in environments with high reverberation times.

Keywords— Blind Source Separation, Convolutive mixtures, Subband structures.

I. INTRODUÇÃO

A maioria dos sinais de áudio capturados por microfones correspondem a misturas de sinais de várias fontes, como por exemplo, sinais de fala, música e ruídos ambientes e de equipamentos. Separação de fontes consiste na recuperação dos sinais originais das fontes de interesse a partir de um ou mais sinais de mistura [1]. Aplicações diretas incluem teleconferências, palestras em tempo real com tradução simultânea e amostragem de sons para composição de música eletrônica. Nessas situações, as fontes são os sinais desejados, mas apenas as gravações das fontes misturadas estão disponíveis e o processo de mistura é desconhecido.

As versões mais recentes de algoritmos de separação cega de fontes têm utilizado sinais capturados por arranjos de

microfones para a separação dos sons chegando de diferentes direções por meio da adaptação de filtros de separação. As razões sinal-interferência resultantes consideravelmente altas mostram a eficácia deste tipo de processamento. Porém, esses algoritmos possuem alguns desafios.

Muitas aplicações atuais exigem que os sensores que captam os sinais de voz estejam a uma distância considerável da fonte em ambientes com obstáculos e paredes que refletem as ondas sonoras, como, por exemplo, em telefones com vídeo, em que os microfones são geralmente colocados perto da câmera. Outro exemplo é no controle de voz em casas automatizadas ou em dispositivos móveis, cujo uso vem crescendo a cada dia e os usuários em geral não desejam estar com o aparelho muito próximo do emissor e/ou em ambientes silenciosos para o mesmo funcionar corretamente.

Um dos principais fatores para o mau desempenho de algoritmos de reconhecimento automático da fala em tais aplicações é a reverberação. Tempos de reverberação grandes provocam alterações temporais nas características da fala, que afetam a sua inteligibilidade [2]. Os efeitos da alta reverberação também prejudicam os desempenhos dos algoritmos de separação cega de fontes (BSS) e estudos para melhorar seus resultados nesses ambientes mostram-se ainda necessários.

Neste trabalho apresentamos um método de separação cega de fontes (BSS) para misturas convolutivas que combina as vantagens de algoritmos de banda larga com a eficiência computacional de técnicas de banda estreita. O método baseia-se no algoritmo de TRINICON (*Triple-N independent component analysis for convolutive mixtures*) proposto em [3]. Através da aplicação do teorema "Szegö", que explora as propriedades Toeplitz e circulantes de matrizes, uma implementação eficiente do algoritmo TRINICON no domínio da frequência foi derivada em [7]. O algoritmo resultante apresenta convergência rápida e reduzida complexidade computacionalmente, enquanto evita problemas típicos de algoritmos de banda estreita, tais como os efeitos de permutações e circularidade. É proposta a sua aplicação na separação cega de fontes em subbandas, resultando em um novo algoritmo com desempenho computacional e convergência melhores do que os do algoritmo em banda cheia. Os resultados das simulações em ambientes acústicos confirmam o melhor desempenho do novo algoritmo, principalmente em ambientes com altas reverberações.

II. O PROBLEMA DE SEPARAÇÃO DE FONTES

Misturas lineares convolutivas de sinais podem ser definidas de acordo com a equação:

$$\mathbf{x}(n) = \mathbf{H}(n) * \mathbf{s}(n), \quad (1)$$

Jéssica Bartholdy Sanson pertence ao Instituto de Telecomunicações, Universidade de Aveiro, Portugal e Mariane Rembold Petraglia pertencem ao Programa de Engenharia Elétrica da COPPE/UFRJ, Rio de Janeiro-RJ, Brasil. E-mails: brunalrpv@gmail.com, jessica.sanson@ua.pt, mariane@pads.ufrj.br. Este trabalho foi parcialmente financiado pelo CAPES.

em que $*$ é o operador convolução, $\mathbf{H}(n)$ é a matriz de mistura de dimensões $P \times Q$, sendo Q o número de fontes e P o número de sensores, contendo as respostas impulsivas dos filtros de mistura $h_{ij}(n)$, i.e.,

$$\mathbf{H}(n) = \begin{bmatrix} h_{11}(n) & h_{12}(n) & \dots & h_{1Q}(n) \\ h_{21}(n) & h_{22}(n) & \dots & h_{2Q}(n) \\ \vdots & \vdots & \ddots & \vdots \\ h_{P1}(n) & h_{P2}(n) & \dots & h_{PQ}(n) \end{bmatrix}, \quad (2)$$

$\mathbf{s}(n)$ é o vetor composto pelos sinais das fontes, i.e.,

$$\mathbf{s}(n) = [s_1(n) \quad s_2(n) \quad \dots \quad s_Q(n)]^T \quad (3)$$

e $\mathbf{x}(n)$ é o vetor formado pelos sinais que chegam aos sensores, dado por

$$\mathbf{x}(n) = [x_1(n) \quad x_2(n) \quad \dots \quad x_P(n)]^T. \quad (4)$$

O problema de separação de fontes convolutivas lineares consiste em encontrar a matriz de separação $\mathbf{W}(n)$ capaz de obter estimativas dos sinais das fontes a partir dos sinais de mistura, ou seja,

$$\mathbf{y}(n) = \mathbf{W}(n) * \mathbf{x}(n) \quad (5)$$

sendo

$$\mathbf{y}(n) = [y_1(n) \quad y_2(n) \quad \dots \quad y_Q(n)]^T \quad (6)$$

e a matriz de separação $\mathbf{W}(n)$, de dimensões $Q \times P$, definida por

$$\mathbf{W}(n) = \begin{bmatrix} w_{11}(n) & w_{12}(n) & \dots & w_{1P}(n) \\ w_{21}(n) & w_{22}(n) & \dots & w_{2P}(n) \\ \vdots & \vdots & \ddots & \vdots \\ w_{Q1}(n) & w_{Q2}(n) & \dots & w_{QP}(n) \end{bmatrix}. \quad (7)$$

Assume-se filtros de separação $w_{ij}(n)$ de comprimento N . Quando o número de sensores P é igual ao número de fontes Q , o problema de separação é denominado determinado, sendo esse o caso considerado neste trabalho. É necessário introduzir dois parâmetros adicionais para a formulação geral do algoritmo BSS com processamento de blocos de misturas convolutivas:

- número de atrasos no tempo D das correlações consideradas na função custo ($1 \leq D \leq N$);
- comprimento dos blocos dos sinais de saída L , como base para as estimativas das correlações.

Neste trabalho, foi escolhido $L = 2N$, seguindo a recomendação de [3].

A matriz \mathbf{W}_{pq} , de dimensões $2N \times D$, contém em cada coluna os coeficientes do filtro de separação do p -ésimo sensor à q -ésima saída que geram as D amostras de y_q .

III. ALGORITMO TRINICON

O algoritmo TRINICON explora as seguintes características dos sinais de voz: não-gaussianidade, através de estatísticas de ordens elevadas utilizadas para formular sua função-custo; espectro colorido, através da diagonalização simultânea de matrizes de correlação das saídas para múltiplos atrasos; não-estacionariedade, através da diagonalização simultânea das

matrizes de correlação de saída para diferentes instantes de tempo.

Para apresentação do algoritmo TRINICON numa forma compacta, definem-se os vetores contendo os dados dos sinais de entrada $\mathbf{x}_p(b, j)$ e de saída $\mathbf{y}_q(b, j)$ do b -ésimo bloco, de tamanho $L = 2N$, por:

$$\mathbf{x}_p(b, j) = [x_p(bN + j) \quad \dots \quad x_p(bN - 2N + 1 + j)] \quad (8)$$

$$\mathbf{y}_q(b, j) = \sum_{p=1}^P \mathbf{x}_p(b, j) \mathbf{W}_{pq}(b) \quad (9)$$

sendo $j = 0, \dots, L - 1$ o deslocamento no tempo dentro do bloco. Combinando todos os canais, podemos escrever:

$$\mathbf{Y}(b) = \mathbf{X}(b) \mathbf{W}(b) \quad (10)$$

onde

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

e

$$\mathbf{X}(b) = [\mathbf{X}_1(b) \quad \dots \quad \mathbf{X}_P(b)] \quad (12)$$

Pode-se agora definir a seguinte função de custo, com base na sequência de matrizes de correlação reduzidas e atrasadas dos sinais de saída, como:

$$\mathcal{J} = - \sum_{i=0}^b \beta(i, b) \{ \log[\det(\text{bdiag}(\mathbf{Y}^H(i) \mathbf{Y}(i)))] \dots - \log[\det(\mathbf{Y}^H(i) \mathbf{Y}(i))] \} \quad (13)$$

cujo ponto de equilíbrio corresponde exatamente à solução da BSS desejada. O operador bdiag . seleciona as submatrizes da diagonal. Empregando-se o algoritmo baseado no método do gradiente natural [5] para minimização da função custo \mathcal{J} , cujo gradiente natural em relação à matriz de separação \mathbf{W} é:

$$\Delta \mathbf{W} \propto \mathbf{W} \mathbf{W}^H \frac{\partial \mathcal{J}}{\partial \mathbf{W}^*} \quad (14)$$

obtem-se a seguinte regra de atualização:

$$\mathbf{W}(b) = \mathbf{W}(b - 1) - \mu \Delta \mathbf{W}(b) \quad (15)$$

onde

$$\Delta \mathbf{W}(b) = 2 \sum_{i=0}^{\infty} \beta(i, b) \mathbf{W}(b) \{ \mathbf{R}_{yy}(i) \dots - \text{bdiag}(\mathbf{R}_{yy}(i)) \} \text{bdiag}^{-1}(\mathbf{R}_{yy}(i)) \quad (16)$$

e

$$\mathbf{R}_{yy}(b) = \mathbf{Y}^H(b) \mathbf{Y}(b). \quad (17)$$

Observa-se que a matriz (17) é posto completo desde que $L \geq PD$.

Considerando o caso $P = Q = 2$ por simplicidade, a equação de atualização do algoritmo do modo *off-line*, em que todo o sinal é processado para gerar uma iteração, é dada por:

$$\mathbf{W}(b) = \mathbf{W}(b - 1) - \frac{2\mu}{B} \sum_{b=0}^B \begin{bmatrix} \mathbf{W}_{12} \mathbf{R}_{y_2 y_1} \mathbf{R}_{y_1 y_1}^{-1} & \mathbf{W}_{11} \mathbf{R}_{y_1 y_2} \mathbf{R}_{y_2 y_2}^{-1} \\ \mathbf{W}_{22} \mathbf{R}_{y_2 y_1} \mathbf{R}_{y_1 y_1}^{-1} & \mathbf{W}_{21} \mathbf{R}_{y_1 y_2} \mathbf{R}_{y_2 y_2}^{-1} \end{bmatrix} \quad (18)$$

onde $\mathbf{R}_{y_p y_q}$, $p, q \in \{1, 2\}$, são as submatrizes de dimensão $D \times D$ correspondentes de \mathbf{R}_{yy} , μ é o passo de adaptação do algoritmo.

A atualização dessa função resulta numa complexidade alta para o cálculo de $\mathbf{R}_{qq}^{-1}(b)$ em uma aplicação simples. No entanto, para reduzir a complexidade computacional do algoritmo, os fatores de normalização $\mathbf{R}_{qq}^{-1}(b)$ podem ser simplificados considerando $\mathbf{R}_{qq}(b)$ uma matriz diagonal; desta maneira, $\mathbf{R}_{qq}^{-1}(b)$ pode ser obtida invertendo os elementos da sua diagonal, que correspondem aos inversos das potências dos blocos atrasados no tempo do sinal da q -ésima saída, ou seja:

$$\mathbf{R}_{qq}(b) \approx \text{diag}\{\mathbf{R}_{qq}(b)\} \quad (19)$$

Para reduzir ainda mais o custo computacional no cálculo de $\mathbf{R}_{qq}^{-1}(b)$, foi proposto em [4] se reduzir o fator de normalização a um escalar, ou seja:

$$\mathbf{R}_{qq}(b) \approx \mathbf{y}_q^T(b) \mathbf{y}_q(b) \mathbf{I} \quad (20)$$

IV. ALGORITMO TRINICON NO DOMÍNIO DFT

Em [6], foi demonstrado que qualquer matriz Toeplitz pode ser expressa de forma equivalente no domínio da frequência através de uma matriz circulante gerada pela extensão apropriada da matriz Toeplitz. Em seguida, a matriz circulante é diagonalizada aplicando-se a matriz DFT \mathbf{F}_R , de tamanho $R \times R$, onde $R \geq L + N$ representa o comprimento da transformada. Esses dois passos são aplicados em [7] à matriz Toeplitz do sinal de saída $\mathbf{Y}_q(b)$, conforme as equações:

$$\mathbf{Y}_q(b) = \mathbf{W}_{(L+N) \times R}^{01_{L+N}} \mathbf{F}_R^{-1} \underline{\mathbf{Y}}_q(b) \mathbf{F}_R \mathbf{W}_{R \times N}^{1_{N^0}} \quad (21)$$

ou

$$\mathbf{Y}_q(b) = \mathbf{W}_{(L+N) \times R}^{01_{L+N}} \mathbf{C}_{\mathbf{Y}_q}(b) \mathbf{W}_{R \times N}^{1_{N^0}} \quad (22)$$

onde $\mathbf{C}_{\mathbf{Y}_q}(b)$ é uma matriz circulante de tamanho $R \times R$, e as matrizes correspondentes ao "janelamento" dos sinais são dadas por:

$$\mathbf{W}_{(L+N) \times R}^{01_{L+N}} = [\mathbf{0}_{(L+N) \times (R-L-N)}, \mathbf{I}_{(L+N) \times (L+N)}] \quad (23)$$

e

$$\mathbf{W}_{R \times N}^{1_{N^0}} = [\mathbf{I}_{N \times N}, \mathbf{0}_{(R-N) \times N}] \quad (24)$$

Neste trabalho, a convenção utilizada é que o índice inferior da matriz de "janelamento" denota as suas dimensões e o índice superior descreve as posições dos uns e zeros. O tamanho das submatrizes é indicado em subscrito (por exemplo, "01_L"). A matriz de saída no domínio da frequência, $\underline{\mathbf{Y}}_q(b)$, é diagonal e dada por:

$$\underline{\mathbf{Y}}_q(b) = \text{Diag}\{\mathbf{F}_R[0, \dots, 0, y_q(bN), \dots, y_q(bN + L - 1), 0, \dots, 0]^T\} \quad (25)$$

onde $\text{Diag}\{\mathbf{x}\}$ denota a matriz com os elementos do vetor " \mathbf{x} " na sua diagonal principal. Por fim, temos que

$$\mathbf{R}_{y_p y_q}(b) = \mathbf{W}_{N \times R}^{1_{N^0}} \mathbf{F}_R^{-1} \underline{\mathbf{Y}}_q^H(b) \underline{\mathbf{Y}}_q(b) \mathbf{F}_R \mathbf{W}_{R \times N}^{1_{N^0}} \quad (26)$$

A relação entre $\mathbf{W}_{pq}(b)$ no domínio de tempo e $\underline{\mathbf{W}}_{pq}(b)$ no domínio DFT é dada por:

$$\mathbf{W}_{pq}(b) = \mathbf{W}_{2N \times R}^{1_{2N^0}} \mathbf{F}_R^{-1} \underline{\mathbf{W}}_{pq}(b) \mathbf{F}_R \mathbf{W}_{R \times D}^{1_{D^0}} \quad (27)$$

sendo

$$\underline{\mathbf{W}}_{pq}(b) = \text{Diag}\{\mathbf{F}_R[\mathbf{W}_{pq}^T(b), \mathbf{0}_{R-N \times 1}^T]^T\} \quad (28)$$

A. Cálculo da matriz $\mathbf{R}_{y_p y_q}^{-1}(b)$

Utilizando o teorema *Szego*, podemos aproximar o inverso da matriz $\mathbf{R}_{y_p y_q}(b)$ pelo inverso da matriz circulante, como feito em [7], ou seja,

$$\mathbf{R}_{y_p y_q}^{-1}(b) = \mathbf{W}_{N \times R}^{1_{N^0}} \mathbf{F}_R^{-1} (\underline{\mathbf{Y}}_q^H(b) \underline{\mathbf{Y}}_q(b))^{-1} \mathbf{F}_R \mathbf{W}_{R \times N}^{1_{N^0}} \quad (29)$$

Este procedimento também pode ser interpretado como uma aproximação de banda estreita, uma vez que os autovalores $\underline{\mathbf{Y}}_q^H(b) \underline{\mathbf{Y}}_q(b)$ podem ser facilmente determinados pela DFT da primeira coluna da matriz circulante $\mathbf{C}_{\mathbf{Y}_p \mathbf{Y}_q}(b) = \mathbf{F}_R^{-1} (\underline{\mathbf{Y}}_q^H(b) \underline{\mathbf{Y}}_q(b))^{-1} \mathbf{F}_R$. Sua inversa pode agora ser obtida de forma eficiente através da inversão de escalares, uma vez que $\underline{\mathbf{Y}}_q^H(b) \underline{\mathbf{Y}}_q(b)$ é uma matriz diagonal. Além disso, é importante notar que a inversa de uma matriz circulante é também circulante. Assim, após o janelamento a matriz $\mathbf{R}_{y_p y_q}^{-1}(b)$ resultante apresenta novamente uma estrutura Toeplitz. Em resumo, essa função pode ser eficazmente implementada pela DFT da primeira coluna de $\mathbf{C}_{\mathbf{Y}_p \mathbf{Y}_q}(b)$ seguida por uma inversão escalar dos valores no domínio da frequência e pela DFT inversa. Após a operação de janelamento, estes valores são então replicados para gerar a estrutura Toeplitz de $\mathbf{R}_{y_p y_q}^{-1}(b)$.

B. Regra de atualização

Como apresentado em [8] utilizando-se o gradiente natural, temos que a equação de atualização dos coeficientes da matriz de separação no domínio da frequência é dada por:

$$\underline{\mathbf{W}}(b) = \underline{\mathbf{W}}(b-1) - \mu \Delta \underline{\mathbf{W}}(b) \quad (30)$$

onde

$$\Delta \underline{\mathbf{W}}(b) = \sum_{i=0}^{\infty} \beta(i, b) \mathbf{G}_{PR \times PR}^{1_{N^0}} \Delta \tilde{\underline{\mathbf{W}}}(i) \mathbf{L}_I \quad (31)$$

sendo

$$\Delta \tilde{\underline{\mathbf{W}}}(i) \propto \underline{\mathbf{W}}(i) \text{boff}\{\underline{\mathbf{Y}}(i) \underline{\mathbf{Y}}^H(i)\} \text{bdiag}^{-1}\{\underline{\mathbf{Y}}^H(i) \underline{\mathbf{Y}}(i)\} \quad (32)$$

$$\mathbf{G}_{PR \times PR}^{1_{N^0}} = \text{bdiag}\{\mathbf{F}_R \mathbf{W}_{R \times R}^{1_{N^0}} \mathbf{F}_R^{-1}, \dots, \mathbf{F}_R \mathbf{W}_{R \times R}^{1_{N^0}} \mathbf{F}_R^{-1}\} \quad (33)$$

e $\mathbf{L}_I = \text{bdiag}\{\mathbf{1}_{R \times 1}, \dots, \mathbf{1}_{R \times 1}\}$. A operação boff configura todas as submatrizes na diagonal para zero, a multiplicação no lado direito pela matriz \mathbf{L}_I converte a matriz $\Delta \tilde{\underline{\mathbf{W}}}$ no domínio DFT, de dimensões $R \times R$, em vetores colunas contendo os R valores da DFT. A multiplicação no lado esquerdo pela matriz de restrições $\mathbf{G}_{PR \times PR}^{1_{N^0}}$ transforma a matriz do domínio DFT para o domínio do tempo, restringindo todos os coeficientes de índices maiores do que o comprimento do filtro N a zero, e depois volta para o domínio DFT.

V. BSS EM SUBBANDAS NO DOMÍNIO DA FREQUÊNCIA

Nesta seção, algumas modificações foram feitas no algoritmo TRINICON no domínio da frequência para a sua execução em subbandas. Derivaremos assim um novo método para separação cega de misturas convolutivas em subbandas no domínio da frequência, que apresenta melhor desempenho que os algoritmos em banda cheia, sendo o ganho de desempenho maior em ambientes com altos tempos de reverberação. Esses cenários serão os que focaremos neste artigo, devido ao bom desempenho do algoritmo proposto em relação ao no domínio do tempo compensar seu gasto computacional superior.

Comparado ao algoritmo TRINICON no domínio DFT apresentado na seção anterior, o algoritmo proposto à primeira vista apresenta uma complexidade computacional maior, devido a realizar M vezes o cálculo de $\Delta\mathbf{W}_m$, sendo M o número de subbandas. Porém, para ambientes com tempos de reverberação elevados, ele apresenta resultado superior com filtros de separação de tamanho consideravelmente menor do que o usado em banda completa, tendo assim uma complexidade computacional comparável à deste, além de apresentar maiores taxa de convergência e razão sinal-interferência em regime estacionário.

Cada sinal observado da mistura, $x_p(n)$, é decomposto pelos filtros de análise $\mathbf{F}_m(z)$, sendo os sinais resultantes processados pelos subfiltros de separação de banda completa, dados por $\mathbf{W}_{pq}(n)$, gerando sinais de saída $y_p^m(n)$ para a m -ésima subbanda, os quais serão passados para o domínio DFT, como mostrado nas seções anteriores, para o cálculo de atualização da nova matriz \mathbf{W} .

A atualização da matriz \mathbf{W} se dá em banda completa, utilizando um somatório das funções $\Delta\tilde{\mathbf{W}}_m$ das M subbandas, como realizado no algoritmo apresentado no domínio do tempo.

Por fim para obter a equação de atualização da matriz de filtros de separação \mathbf{W} , usamos a mesma recursão apresentada na seção anterior, com a adição dos cálculos de $\Delta\tilde{\mathbf{W}}_m$ em cada subbanda, como mostrado a seguir:

$$\Delta\mathbf{W}(b) = \sum_{i=0}^{\infty} \beta(i, b) \mathbf{G}_{PR \times PR}^{1N^0} \Delta\tilde{\mathbf{W}}(i) \mathbf{L}_I \quad (34)$$

sendo

$$\Delta\tilde{\mathbf{W}}(i) = \sum_{m=0}^M \mathbf{W}(i) \text{boff} \{ \mathbf{Y}^m(i) \mathbf{Y}^{mH}(i) \} \dots \text{bdiag}^{-1} \{ \mathbf{Y}^{mH}(i) \mathbf{Y}^m(i) \} \quad (35)$$

VI. RESULTADOS DE SIMULAÇÕES

Os algoritmos foram avaliados empregando-se duas misturas convolutivas de dois sinais de voz, supondo $Q = P = 2$, compostas de um sinal de voz masculina e outro de voz feminina, disponibilizados em [9], com duração de 10 s e frequência de amostragem 16 kHz.

As simulações da propagação acústica de salas foram desenvolvidas usando o modelo "Image-Source" descrito em [11], sendo utilizados os parâmetros da Tabela I.

TABELA I
VALORES UTILIZADOS NAS SIMULAÇÕES

Parâmetro	Valor
Ângulo da fonte 1	-45°
Ângulo da fonte 2	45°
Distância entre os microfones	5 cm
Coefficiente de absorção das paredes	0.8
Distância das fontes até o centro do arranjo de microfones	1 m

Os tempos de reverberação avaliados (T_{60}) foram: 0.3 s e 0.5 s. A medida de avaliação utilizada foi a SIR (*Signal-to-Interference Ratio*). Para a decomposição em subbandas, empregou-se um banco de filtros modulado por cosseno [10] com $M = 4$ subbandas e filtro protótipo de comprimento 32. O tamanho dos blocos foi fixado em $L = 2D$, sendo o valor do atraso D utilizado igual ao comprimento do filtro de separação, ou seja, $D = N$.

Para tornar a adaptação mais robusta, um controle no tamanho do fator de aprendizagem do algoritmo é desejável. Usamos uma estratégia simples, mas eficaz, para atualizar o tamanho do passo, baseado no método apresentado em [8]. Os limites $[\mu_{min}, \mu_{max}]$ variaram para cada simulação, com menor limite inferior $\mu_{min} = 0.00001$ e maior limite superior $\mu_{max} = 0.005$.

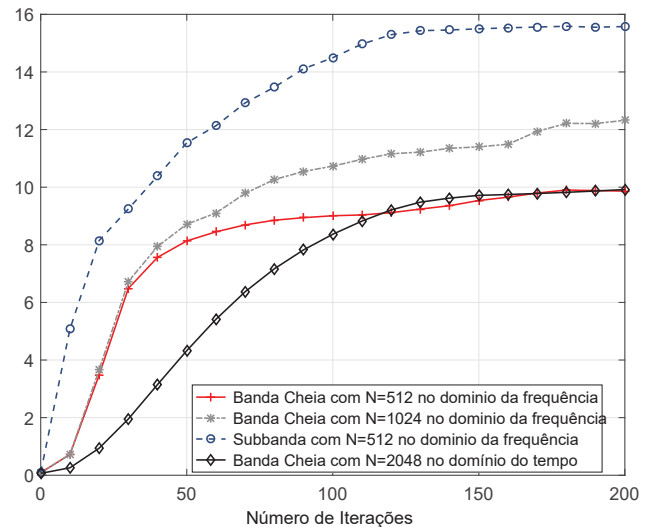


Fig. 1. Evolução da SIR (em dB) para $T_{60} = 0.3$ s.

A Figura 1 apresenta as evoluções da SIR ao longo das iterações obtidas com os algoritmos em banda cheia e em subbandas para $T_{60} = 0.3$ s, com filtros de separação de comprimentos $N = 2048$ no domínio do tempo em banda cheia, $N = 1024$ e $N = 512$ para o algoritmo em banda cheia e $N = 512$ para o algoritmo em subbandas no domínio da frequência. Pode-se verificar nesta figura que o algoritmo em subbandas proposto neste trabalho apresenta desempenho superior, obtendo SIR de 15.6 dB, enquanto que o algoritmo em banda cheia alcançou SIR de 9.89 dB com o mesmo comprimento de filtro $N = 512$ e 12.32 dB com $N = 1024$ e 9.92 dB no domínio do tempo, novamente inferiores à SIR obtida com o algoritmo em subbandas.

Na Figura 2 são mostrados os resultados das simulações para $T_{60} = 0.5$ s com filtros de separação de comprimentos $N = 4096$ no domínio do tempo em banda cheia, $N = 1024$ e $N = 512$ para o algoritmo de banda cheia e $N = 512$ para o algoritmo em subbandas no domínio da frequência. Novamente pode-se verificar a superioridade do algoritmo

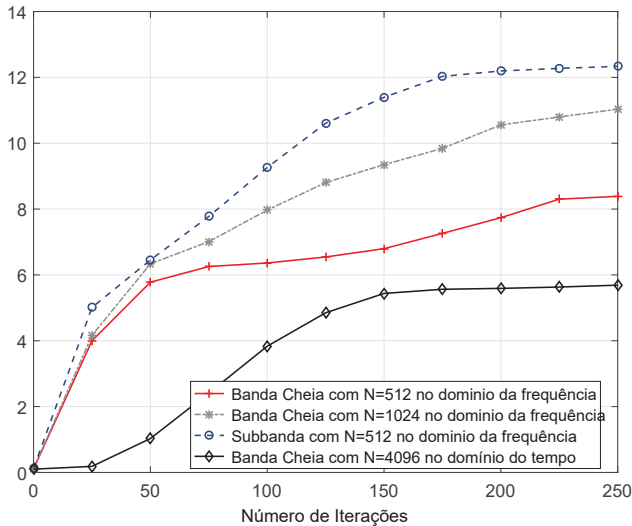


Fig. 2. Evolução da SIR (em dB) para $T_{60} = 0.5$ s.

em subbandas, apresentando convergência mais rápida e razão sinal-interferência maior quando comparado ao de banda cheia, mesmo com comprimento de filtro menor. Para o tempo de reverberação de $T_{60} = 0.5$ s o algoritmo em banda cheia alcançou SIR de 6.15 dB no domínio do tempo, 11.03 dB para $N = 1024$ e de 8.38 dB para $N = 512$ no domínio da frequência, enquanto que o algoritmo em subbandas alcançou SIR de 12.34 dB, demonstrando o ganho no processo de separação em subbandas tanto em relação à SIR quanto em relação ao custo computacional, uma vez que o tamanho dos filtros de separação é de apenas $N = 512$.

Analisando os resultados obtidos, verificou-se a superioridade do método em subbandas proposto, tendo eles uma complexidade computacional menor devido ao reduzido tamanho de filtro utilizado no sistema de separação, um menor tempo de convergência e maior razão sinal-interferência em regime estacionário quando comparados aos algoritmos correspondentes em banda cheia, sendo, portanto, a melhor opção em ambientes com alta reverberação (acima de 0.5 s).

VII. CONCLUSÃO

Neste trabalho foi derivado um novo algoritmo de separação cega de fontes, baseado no algoritmo TRINICON no domínio DFT, que utiliza os sinais de misturas em subbandas. O algoritmo proposto apresenta complexidade computacional menor do que o algoritmo em banda cheia para separação de sinais de voz em ambientes com alta reverberação, devido ao reduzido tamanho dos filtros necessários no sistema de separação, além de um menor tempo de convergência e maior razão sinal-interferência em regime estacionário.

REFERÊNCIAS

- [1] S. Haykin, and Z. Chen, "The Cocktail Party Problem," *Neural Comput.*, pp.1875–1902, MAR. 2005.
- [2] A. Schwarz, K. Reindl and W. Kellermann, "A two-channel reverberation suppression scheme based on blind signal separation and wiener filtering," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp.113–116, MAR. 2012.
- [3] H. Buchner, R. Aichner and W. Kellermann, "Blind Source Separation for Convulsive Mixtures Exploiting Nongaussianity, Nonwhiteness, and Nonstationarity," *International Workshop on Acoustic Echo and Noise Control*, pp.275–278, Sep. 2003.
- [4] R. Aichner, H. Buchner and S. Araki, "On-Line Time-Domain Blind Source Separation of Nonstationary Convolved Signals," *European Signal Processing Conference*, v. 9, no. 10, pp. 987–992, 2003.
- [5] S. Amari, A. Cichocki and H. H. Yang, "A New Learning Algorithm for Blind Signal Separation," *Advances in Neural Information Processing Systems*, pp. 757–763, 1996.
- [6] R. Gray, "On the asymptotic eigenvalue distribution of Toeplitz matrices," *IEEE Transactions on Information Theory*, pp. 725–730, 1972.
- [7] R. Aichner, and H. Buchner, and W. Kellermann, "A Novel Normalization and Regularization Scheme for Broadband Convulsive Blind Source Separation," *Independent Component Analysis and Blind Signal Separation: 6th International Conference, ICA*, pp. 527–535, 2006.
- [8] R. Aichner, *Acoustic Blind Source Separation in Reverberant and Noisy Environments*. Tese de doutorado, Technischen Fakultät der Friedrich-Alexander-Universität, Erlangen, Nurnberg, 2007.
- [9] E. Vincent and H. Sawada, *Código-fonte MATLAB para simulação de sala reverberante*, howpublished = <http://www.irisa.fr/metiss/SASSECO7>. Jun., 2015.
- [10] S. K. Mitra. *Digital Signal Processing: A Computer-Based Approach*. McGraw-Hill Higher Education, 2000.
- [11] E. A. Lehmann and A. M. Johansson and S. Nordholm, "Reverberation-Time Prediction Method for Room Impulse Responses Simulated with the Image-Source Model," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 159–162, 2007.