

Sistema de Localização de Fontes Sonoras Baseado em Algoritmo de Separação Cega

Luiz C. F. Nogueira^{1,2} e Mariane R. Petraglia¹

Resumo—Este artigo descreve um sistema de localização cega de fontes sonoras, que utiliza dois pares de microfones omnidirecionais. O método de localização desenvolvido se baseia em algoritmos de separação cega de fontes e de estimação do ângulo de chegada. A imprecisão na localização das fontes é analisada em função da distância entre os microfones em cada par, da reverberação do ambiente, da distância entre a fonte e o par de microfones, e da proximidade das fontes. Diretrizes para posicionar os microfones são derivadas dos resultados da análise, de forma a tornar o sistema de localização cega de fontes mais preciso.

Palavras-Chave—Localização cega de fontes, Separação cega de fontes, Direção de Chegada

Abstract—This article describes a system for blind sound source localization, which uses two pairs of omnidirectional microphones. The localization method is based on algorithms developed for blind source separation and estimation of the angle of arrival. The inaccuracy in the source localization is analyzed in terms of the distance between the microphones in each pair, the reverberation of the environment, the distance between the source and the pair of microphones, and the proximity of the sources. Guidelines for positioning the microphones are derived from the results of the analysis in order to render the blind source localization system more precise.

Keywords—Blind Source Localization, Blind Source Separation, Direction of Arrival

I. INTRODUÇÃO

O desenvolvimento de algoritmos que utilizam sinais capturados por múltiplos sensores para estimar as posições de fontes é uma linha de pesquisa de destaque na área de processamento de sinais, com diversas aplicações na área militar, em robótica e em telecomunicações.

Em particular, no rastreamento de fontes sonoras, cabe ao sistema de localização fornecer estimativas confiáveis das posições das fontes, sendo que a robustez das estimativas dependerá de uma série de fatores, como a distribuição dos microfones no ambiente, posições das fontes, presença de reverberação, entre outros.

O objetivo deste trabalho é desenvolver um sistema de localização cega de fontes sonoras em ambientes fechados (com reverberação), e indicar diretrizes para auxiliar na escolha do posicionamento dos microfones no ambiente, com a finalidade de tornar as estimativas das posições das fontes confiáveis. Para isso, foram empregadas técnicas de separação cega de fontes (BSS, do inglês Blind Source Separation) [1]

e de estimação das direções de chegadas (DoA, do inglês Direction of Arrival) dos sinais nos microfones [2].

O texto está organizado da seguinte forma: a Seção 2 aborda os fundamentos de BSS e de estimativa da DoA, os quais são necessários para a compreensão do sistema, baseando-se no modelo de campo distante; a Seção 3 analisa a precisão das estimativas da DoA, através de um conjunto de simulações, nas quais são variadas a distância entre os microfones de cada par, a intensidade da reverberação, T_{60} [5], a distância entre a fonte e o par de microfones, e as direções de chegada reais dos sinais das fontes em cada microfone; a Seção 4 discute a melhor forma de posicionar os pares de microfones na sala, de forma a obter estimativas das posições das fontes confiáveis.

Os experimentos foram efetuados com as fontes e os microfones posicionados a uma altura de aproximadamente 1,5 m do nível do solo. Portanto, as análises e conclusões apresentadas neste artigo são válidas para os casos em que as fontes e os microfones estão aproximadamente num mesmo plano.

II. FUNDAMENTOS TEÓRICOS

A. Canal de Áudio

Este trabalho aborda o problema de localização de fontes de áudio a partir de sinais adquiridos por microfones em uma sala. Portanto, os sinais disponíveis apresentarão o efeito de reverberação, fenômeno que ocorre devido às diversas reflexões da onda sonora no ambiente. Quanto maior for a intensidade da reverberação, maior será o grau de complexidade do problema da separação e da localização dos sinais das duas fontes.

Para a simulação do canal de áudio, foi utilizada neste trabalho uma versão do modelo de propagação de ondas acústicas numa sala conhecido como ISM (do inglês *image-source model*, o qual tem sido amplamente empregado em simulações para fins científicos. Este modelo foi originalmente apresentado em [4], sendo a versão utilizada nos nossos experimentos a desenvolvida em [5] e [6], por questões de eficiência no tempo de processamento e na modelagem da resposta do canal.

B. Misturas Convolutivas

Quando há mais de uma fonte sonora ativa no ambiente, os sinais obtidos nos microfones são misturas dos sinais das fontes modificados pelo canal acústico entre cada fonte e cada microfone. A Fig. 1 ilustra o processo de mistura para duas fontes e dois sensores, onde h_{ji} representa a resposta

¹PADS, Programa de Engenharia Elétrica, COPPE, UFRJ, Rio de Janeiro - RJ, Brasil. ²Coordenação de telecomunicações, CEFET/RJ Uned NI, Nova Iguaçu - RJ, Brasil. E-mails: lctelecom@pads.urfj.br, mariane@pads.urfj.br.

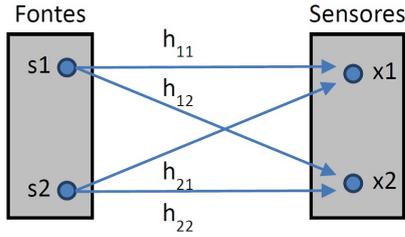


Fig. 1. Representação do sistema de mistura com duas fontes e dois sensores.

impulsiva do canal entre a fonte i e o sensor j . Para N fontes e M microfones, o sinal no sensor j pode ser escrito como

$$x_j(n) = \sum_{i=1}^N \sum_{l=0}^{L-1} h_{ij}(l) s_i(n-l), \quad j = 1, \dots, M \quad (1)$$

onde s_i representa o sinal da fonte i e L o comprimento das respostas impulsivas dos canais $h_{ij}(\cdot)$. As misturas para as quais $L > 1$ são chamadas de misturas convolutivas. É importante destacar que quanto maior for a quantidade de fontes (maior N) e/ou maior for a duração da reverberação (maior L), maior será a complexidade do sistema de separação. Neste trabalho, consideraremos o caso de duas fontes ($N = 2$) e dois sensores ($M = 2$).

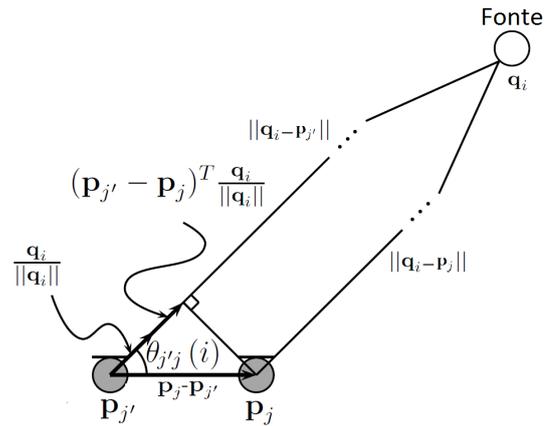
C. Separação Cega de Fontes

A separação cega de fontes tem como base a análise de componentes independentes (ICA, do inglês Independent Component Analysis). O ICA possui uma série de versões [7],[8], sendo elas sozinhas adequadas apenas para a separação de misturas instantâneas. No caso das salas acústicas, onde as misturas são convolutivas, há a necessidade de uma transformação para o domínio tempo-frequência (de forma a tornar as misturas instantâneas em cada raia) e a utilização de uma versão do ICA adequada a sinais de valor complexo. Partindo dos sinais $x_j(n)$ nos sensores, para separá-los realizam-se as seguintes etapas:

- 1) Aplica-se a transformada de Fourier de curta duração (STFT, do inglês Short-Time Fourier Transform) às misturas $x_j(n)$ gerando $X(f_{bin}, j, n)$, sendo f_{bin} , j e n os índices correspondentes à raia de frequência, à mistura, e ao bloco de dados, respectivamente;
- 2) Aplica-se uma versão do algoritmo ICA adequada para sinais de valor complexos [2] em cada raia de frequência para obter a matriz de separação, $W(f_{bin})$, e a estimativa do sinal separado por faixa de frequência $Y(f_{bin}, i, n)$ para cada fonte i ;
- 3) Usa-se o princípio da correlação espectral [9] [10] e/ou cálculos das DOAs [2] para corrigir permutações das estimativas das fontes entre as raias de frequência;
- 4) Usa-se o princípio da distorção mínima [11] para corrigir diferentes escalamentos nas diferentes raias de frequência;
- 5) Aplica-se a STFT inversa para obter a estimativa dos sinais das fontes no domínio do tempo.

D. Estimação da Direção de Chegada

A determinação da direção de chegada é baseada no modelo de campo distante [2], com a finalidade de resolver o problema da permutação nas diferentes raias de frequência. O princípio deste modelo é apresentado na Fig. 2, o qual considera a distância entre os microfones, $d_{jj'} = \|\mathbf{p}_{j'} - \mathbf{p}_j\|$, muito pequena quando comparada às distâncias entre a fonte e os sensores. Assim, admite-se que os percursos diretos pelos quais os sinais passam até chegar aos sensores tendem a ser paralelos, e que as atenuações sofridas pelos sinais são praticamente iguais.


 Fig. 2. Ilustração do modelo de campo distante, onde \mathbf{q}_i é o vetor com as coordenadas da posição da fonte, e \mathbf{p}_j e $\mathbf{p}_{j'}$ são os vetores com as coordenadas das posições dos sensores.

Neste modelo, a única diferença significava entre os sinais nos sensores está na variação de tempo em que os sinais atingem os sensores, influenciando na diferença de fase entre os sinais, parâmetro dependente do ângulo de incidência do sinal, $\theta_{j'j}(i)$, da distância entre os sensores e da frequência do sinal. Portanto, considerando as aproximações acima, a relação entre as respostas ao impulso dos filtros correspondentes aos dois percursos é dada por

$$h_{j'i}(t) = h_{ji}(t + \delta t) \quad (2)$$

Da equação 2, pode-se estimar a DoA do sinal da fonte i nos microfones por [2]

$$\theta(f, k) = \arccos \left(\frac{\arg \left(\frac{H_{ji}(f)}{H_{j'i}(f)} \right)}{2\pi f c^{-1} d_{jj'}} \right) \quad (3)$$

onde $H_{ji}(f)$ e $H_{j'i}(f)$ são as transformadas de Fourier de $h_{ji}(t)$ e de $h_{j'i}(t)$, as quais para cada raia podem ser estimadas através da matriz inversa de $W(f_{bin})$ obtida pelo algoritmo ICA no domínio da frequência, e c é a velocidade do som. Uma restrição deste modelo é que ele só pode ser utilizado para frequências

$$f < c/2d_{jj'} \quad (4)$$

devido ao aliasing espacial.

III. ANÁLISE DO ERRO DE ESTIMAÇÃO DO DOA

A acurácia do método de estimação da DoA utilizado neste trabalho foi analisada através de simulações com o modelo de propagação acústica ISM, com sinais de voz amostrados em 8 KHz, em uma sala de comprimento e largura iguais a 6 m.

As posições das fontes foram escolhidas aleatoriamente, obedecendo a uma distribuição uniforme, conforme ilustrado na Fig. 3, onde as posições das fontes utilizadas nas simulações estão indicadas pelo símbolo +. As fontes estão sempre a pelo menos um metro de distância das paredes, condição necessária para preservar os parâmetros de qualidade do modelo. Na parede localizada na parte inferior da figura, as fontes estão a uma distância de 1,5 m para não prejudicar a validade do modelo de campo distante, pois assim as fontes estão ao menos a 0,5 m de distância dos microfones, valor muito maior que a distância entre os microfones de cada par, a qual nas simulações foi sempre menor que 10 cm.

O centro do par de microfones foi posicionado no ponto de coordenadas (3;1), conforme ilustrado na Fig. 3, sendo utilizadas as distâncias entre microfones de 2, 6 e 10 cm. Para cada uma dessas distâncias foi feito um conjunto de simulações, com T_{60} igual a 100, 250 e 400 ms, estimando-se a DoA.

Foram utilizadas 100 posições diferentes das duas fontes, sendo realizados testes com as 3 distâncias diferentes entre os microfones, e com os 3 valores de T_{60} , totalizando 1800 estimativas das DoAs, as quais foram comparadas aos seus valores reais.

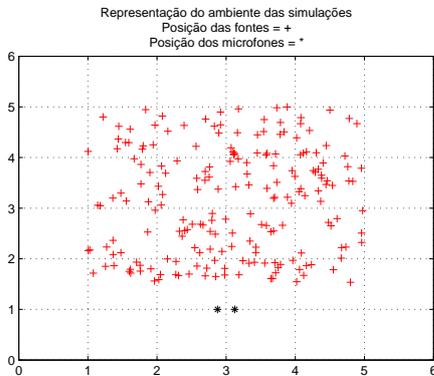


Fig. 3. Dimensões da sala, posições das fontes e microfones nas simulações.

A. Resultados das Estimativas da DoA

Os resultados das 1800 simulações são apresentados na Fig. 4, através de gráficos que relacionam cada faixa de valores da DoA real (no eixo horizontal) ao valor médio do erro de suas estimativas (no eixo vertical).

Os gráficos para os diferentes valores de T_{60} e da distância entre microfones são apresentados em um grid. O T_{60} varia no grid horizontalmente, assim cada coluna corresponde a uma intensidade de reverberação, enquanto a distância entre os microfones varia verticalmente, com cada linha correspondendo a uma distância.

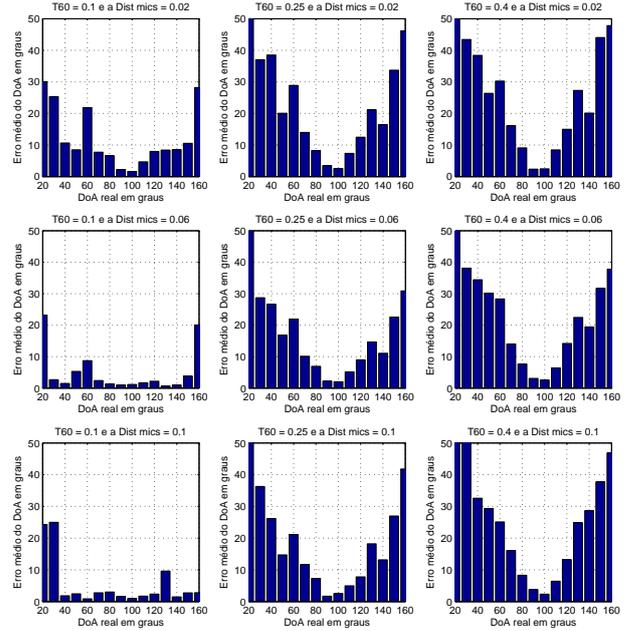


Fig. 4. Relação entre o erro médio da estimativa e a DoA real, variando-se o T_{60} e a distância entre os microfones.

B. Análise dos resultados

A análise dos resultados é apresentada inicialmente considerando-se a variação de cada um dos parâmetros.

1) *Distância entre os microfones num par:* As estimativas da DoA obtidas para a distância de 2 cm, com erros no gráfico da primeira linha do grid da Fig. 4, apresentam baixa qualidade quando comparadas às obtidas com as distâncias maiores, mesmo em situações em que há pouca reverberação. A distância de 6 cm, com erros das estimativas representados na segunda linha, apresentou resultados significativamente melhores quando comparados com os da distância de 2 cm. Na comparação com os obtidos com a distância de 10 cm, os resultados são levemente melhores. Então, é possível concluir que a distância de 2 cm não é uma boa opção, sendo as distâncias próximas a 6 cm boas escolhas.

2) *Reverberação T_{60} :* Analisando a Fig. 4, verificamos que, com tempo de reverberação de 100 ms, há boa acurácia nas estimativas da DoA para praticamente todos os ângulos e distâncias entre microfones; porém esses casos não são os predominantes em um ambiente reais. Para valores maiores de T_{60} , os erros aumentam significativamente, exceto para DoAs reais em torno de 90° .

Constatamos, portanto, nos nossos experimentos, que quanto menor o tempo de reverberação, melhores são as estimativas da DoA. É importante destacar que a intensidade da reverberação depende do ambiente, cabendo ao projetista elaborar soluções para diminuir os efeitos da tal fenômeno.

3) *DoA real:* Os resultados dos testes com direções de chegadas reais entre 70° e 110° possuem as melhores estimativas, com erro médio máximo em torno de 10° , característica que pode ser explorada para posicionar os microfones e aumentar a robustez em suas aplicações. Apenas nos casos onde há pouca reverberação, os resultados com DoAs reais fora da faixa de 70° a 110° foram satisfatórios.

Analisando a figura e os resultados de todas as simulações, é possível perceber que a DoA real é um parâmetro de peso extremamente significativo.

4) *Distância entre a fonte e o par de microfones:* A distância entre a fonte e o par de microfones influencia a intensidade com que um sinal chega a um sensor, podendo atrapalhar a estimação da DoA em função da razão sinal-ruído ficar muito baixa; porém, se o ruído não for significativo em relação ao sinal principal, não haverá grande aumento nos erros de estimação. De acordo com o ambiente, a distância entre a fonte e o par de microfones pode influenciar a intensidade da reverberação em relação ao caminho direto entre a fonte e o microfone, contribuindo para aumentar os erros na localização.

A conclusão geral dessa análise experimental é que há boa acurácia na estimativa da DoA para os casos em que o seu valor real está entre 70° e 110° , a distância entre microfones está em torno de 6 cm, e o tempo de reverberação é relativamente baixo (em torno de 100 ms). Portanto, essas são boas características para operação do sistema de localização.

IV. POSICIONAMENTO DE MICROFONES NO SISTEMA DE LOCALIZAÇÃO DE FONTES

Considerando a utilização de dois pares de microfones, estimativas da localização de fontes sonoras podem ser obtidas calculando-se os pontos de encontro das retas definidas pelas estimativas das DoAs de cada sinal em relação a cada par de sensores posicionados adequadamente no ambiente de operação do sistema.

Este trabalho compara três diferentes configurações de posicionamento dos pares de microfones numa sala (de largura e comprimento iguais a 6 m) para estimar a localização das fontes, conforme ilustrado na Fig. 5. A primeira configuração utiliza dois pares de microfones paralelos a uma das paredes, representados na Fig. 5 como Mics1. A segunda configuração posiciona cada par de microfones paralelo a uma das paredes, representados na Fig. 5 como Mics2. Na terceira configuração, os pares de microfones estão posicionados nas quinas da área em que as fontes podem estar localizadas, estando o par da esquerda inclinado de -45° em relação ao eixo horizontal, e o da direita de 45° , representados na Fig. 5 como Mics3.

As retas mostradas na Fig. 6 identificam a região em que cada par de microfones estimará a DoA com confiabilidade. Estas regiões correspondem a valores de DoA reais entre 70° e 110° em relação à cada par de microfones, sendo o erro esperado na estimativa da DoA inferior a 10° .

Comparando as três configurações, observa-se que o primeiro caso (*Mics1*) apresenta boas características de operação do método de estimação da DoA apenas numa pequena região da sala (o triângulo central superior entre as duas retas mais internas). Já na segunda configuração (*Mics2*), a região na qual as fontes seriam localizadas com menores erros é o quadrilátero central. A terceira configuração (*Mics3*) apresenta a maior região na qual o sistema de localização apresenta baixos erros.

A comparação das três formas de posicionamento dos pares de microfones permite concluir que a configuração *Mics3* é a

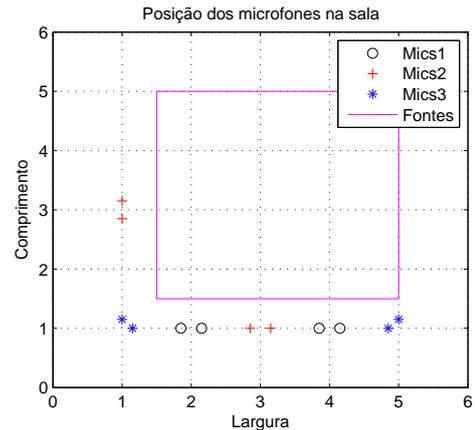


Fig. 5. Posições dos microfones nas três diferentes configurações.

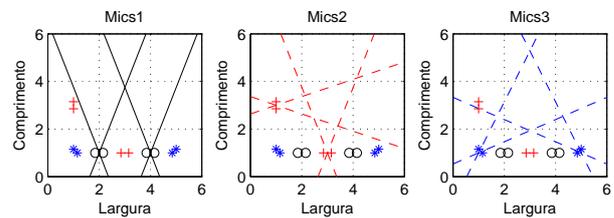


Fig. 6. Região de atuação eficiente para cada configuração dos pares de microfones.

mais eficiente para a sala considerada. Para outros ambientes, uma análise semelhante pode ser feita. No caso geral, podem-se adotar as seguintes diretrizes para posicionar os microfones:

- Aumentando-se a distância entre os pares de microfones, reduz-se a tendência das direções de chegadas nos dois pares de microfones a serem próximas, facilitando a busca do ponto de encontro das retas referentes às DoAs estimadas;
- Inclinando-se os pares de microfones em relação ao eixo horizontal, obtém-se uma maior região na qual as estimativas das DoAs apresentam baixos erros, levando a uma localização mais precisa das fontes.

V. CONCLUSÃO

Neste artigo descrevemos um sistema de localização cega de fontes sonoras, que utiliza dois pares de microfones omnidirecionais. Esse sistema utiliza um método de separação cega de fontes no domínio da frequência, seguido de um método de estimação da direção de chegada. A partir das estimativas das DoAs de uma fonte em relação aos dois pares de microfones, pode-se determinar a posição da mesma. Três diferentes configurações de posicionamento dos pares de microfones foram analisadas. A partir dos resultados obtidos, elaboramos diretrizes para tornar a operação do sistema de localização mais eficiente.

AGRADECIMENTOS

Este trabalho foi parcialmente financiado pelo CNPq e pela FAPERJ.

REFERÊNCIAS

- [1] S. Makino, T.-W. Lee e H. Sawada, *Blind Speech Separation*. Springer, 2007.
- [2] S. Makino, H. Sawada, R. Mukai and S. Araki, "Blind source separation of convolutive mixtures of speech in frequency domain", *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, v. E88-A, Issue 7, pp. 1640-1655, 2005.
- [3] M. Omologo and P. Savier, "Use of the crosspower-spectrum phase in acoustic event location", *IEEE Trans. on Speech and Audio Processing*, pp. 288-292, 1997.
- [4] J. Allen and D. Berkeley, "Image method for efficiently simulating," *Journal of the Acoustical Society of America*, vol. 65(4), pp. 943-950, 1979.
- [5] E. Lehmann, A. Johansson and S. Nordholm, "Reverberation-time prediction method for room impulse responses simulated with the image-source model," *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 159-162, Oct. 2007.
- [6] E. Lehmann and A. Johansson, "Prediction of energy decay in room impulse responses simulated with an image-source model," *Journal of the Acoustical Society of America*, vol. 124, pp. 269-277, 2008.
- [7] J. Romano, R. Attux, C. Cavalcante and R. Suyama, *Unsupervised signal processing: channel equalization and source separation*. CRC Press, 2011.
- [8] A. Hyvarinen and E. Oja, "A fast fixed-point algorithm for Independent Component Analysis," *Neural Computation*, vol. 9(7), pp. 1483-1492, 1997.
- [9] J. Anemuller and B. Kollmeier, "Amplitude modulation decorrelation for convolutive blind source separation," *Proceedings of ICA*, pp. 215-220, Jun. 2000.
- [10] H. Sawada and S. Makino, "Measuring dependence of bin-wise separated signals for permutation Alignment in Frequency-domain BSS," *IEEE International Symposium on Circuits and Systems*, pp. 3247-3250, May 2007.
- [11] K. Matsuoka and S. Nakashima, "Minimal distortion principle for blind source separation," *Proceedings of ICA*, pp. 722-727, Dec. 2001.