

Localização de Fontes Acústicas por SRP-PHAT Volumétrico Robusto

Felipe Barboza da Silva e Wallace Alves Martins

Resumo— Este artigo propõe uma nova técnica para localização de fontes acústicas, denominada SRP-PHAT (do inglês *steered response power phase transform*) volumétrico com refinamento por LMedS (do inglês *least median of squares*), a qual congrega a precisão do LMedS na localização e a capacidade do SRP-PHAT volumétrico em definir, com acurácia, os locais mais prováveis de conter a posição da fonte. Apesar do LMedS localizar a fonte com baixo erro, sua complexidade computacional é demasiadamente elevada, uma vez que, para estimar a posição da fonte, realiza ao menos uma busca numa grade de pontos por todo ambiente. A técnica aqui proposta se mostrou eficaz na redução do espaço de busca da técnica LMedS e com um desempenho comparável. Para uma razão sinal-ruído de 20 dB e para uma sala com tempo de reverberação de 600 ms, a técnica LMedS obteve erros abaixo de 10 cm (mediana dos erros de estimação de localização), o mesmo ocorrendo para a técnica proposta, ainda que utilizando uma grade de busca com uma quantidade de pontos aproximadamente 60% menor.

Palavras-Chave— Localização de fontes acústicas, diferença de tempos de chegada, correlação cruzada, reverberação, SRP.

Abstract— This paper proposes a new technique for sound source localization, called volumetric steered response power phase transform (SRP-PHAT) with refinement via least median of squares (LMedS), which brings together the localization precision of the LMedS and the volumetric SRP-PHAT capability of determining, with accuracy, the most likely regions that contain the source position. Although the LMedS usually estimates the source position with low error, its computation burden is high for it performs at least a full grid search over the environment. The proposed technique is efficient on reducing the search space of the LMedS method, while keeping a comparable performance. Indeed, for a signal-to-noise ratio (SNR) of 20 dB and a room with reverberation time of 600 ms, the LMedS technique achieves median error under 10 cm, the same performance achieved by the proposed technique, even employing a search grid with nearly 60% less points.

Keywords— Sound source localization, time-difference of arrival, generalized cross-correlation, reverberation, SRP.

I. INTRODUÇÃO

A localização de fontes acústicas utilizando arranjo de microfones possui muitas aplicações na indústria, bem como na área militar. O vazamento de gás em um duto pode ter sua posição estimada através dessa técnica, assim como disparos de armas de fogo dentro de um quartel militar [1], [2]. Tal tecnologia também pode ser aplicada para realce de fala (do inglês *speech enhancement*) e audioconferências [3].

Este artigo lida com a localização de fontes em ambientes fechados, tais como salas, nos quais existe somente uma fonte de interesse. O esquema de localização de fontes baseia-se na captação dos sinais emitidos pela fonte por microfones e na posterior estimação da posição através de um processamento dos sinais adquiridos. O tempo que o sinal da fonte leva

para chegar em cada um dos microfones é diferente, sendo dependente das posições da fonte e dos microfones. Portanto, a estimação de tais tempos pode ser utilizada para aferir sua localização. No contexto deste artigo, as diferenças de tempos de chegada (TDOA, do inglês *time-difference of arrival*) [4] são estimadas a partir das correlações cruzadas entre os sinais dos pares de microfones.

Existem várias famílias de técnicas de localização de fontes utilizando arranjos de microfones, dentre as quais as mais utilizadas são as que explicitamente estimam os TDOAs e as que usam conformação de feixe (mais conhecida na literatura pelo termo *beamforming*) [5], [6]. Em relação aos métodos de localização baseados em TDOA, a forma mais utilizada de estimação desse parâmetro emprega a correlação cruzada entre os sinais captados pelos pares de microfones [4]. O tipo de correlação mais comum é a correlação cruzada generalizada (GCC, do inglês *generalized cross-correlation*) [4], [7].

O maior desafio para os sistemas de localização de fontes baseados em TDOA é a estimação correta desse parâmetro, a qual é substancialmente afetada por efeitos de reverberação e ruído. Desafio comparável é tornar o método de localização o mais computacionalmente eficiente o possível. Nesse sentido, este artigo descreve uma nova técnica de localização que visa a diminuir a complexidade computacional da técnica de mínima mediana dos quadrados (LMedS, do inglês *least median of squares*) [8], sendo esta originalmente concebida para melhorar a robustez a erros de estimação de TDOA sem se preocupar, *a priori*, com o custo computacional resultante.

O aumento da eficiência computacional da técnica baseada em mediana foi obtido através de um pré-processamento com base em uma adaptação da técnica SRP-PHAT (do inglês *steered response power phase transform*) volumétrica [5], [9], [10], a qual pertence à classe de técnicas que usam *beamforming*. Esse pré-processamento tem como meta reduzir o espaço de busca utilizado pelo LMedS, diminuindo portanto a complexidade computacional do sistema de localização, sem deteriorar de forma significativa o desempenho do sistema.

II. ESTIMAÇÃO DE TDOA

A. Definição de TDOA

O TDOA entre os microfones de um dado par quantifica a diferença temporal que a onda sonora leva para viajar da fonte até tais microfones através da menor distância possível (caminho direto ou visada direta). Matematicamente, definindo $M \in \mathbb{N}$ como o número de microfones, $p \in \{1, \dots, P\}$ como o índice de um dado par de microfones, onde $P = \frac{M(M-1)}{2}$, e $c \in \mathbb{R}_+$ como a velocidade de propagação da onda, assumida conhecida, é possível definir o p -ésimo TDOA τ_p como

$$\tau_p(\mathbf{s}) \triangleq \frac{\|\mathbf{m}_p^{(1)} - \mathbf{s}\| - \|\mathbf{m}_p^{(2)} - \mathbf{s}\|}{c}, \quad (1)$$

F. B. da Silva e W. A. Martins são da Universidade Federal do Rio de Janeiro (UFRJ), RJ, Brasil (emails: {felipe.silva, wallace.martins}@smt.ufrj.br).

onde $\mathbf{s} \in \mathbb{R}^3$ representa a posição da fonte, $\mathbf{m}_p^{(1)}, \mathbf{m}_p^{(2)} \in \mathbb{R}^3$ denotam as posições do p -ésimo par de microfones e $\|\cdot\|$ denota a norma euclidiana de um vetor.

B. Estimação por Correlação Cruzada

Existem diversos métodos para se estimar TDOAs, em sua maioria baseados em correlações cruzadas (CC), que permitem avaliar qual é a melhor correspondência entre um sinal capturado por um microfone e as diferentes versões temporalmente atrasadas do sinal capturado pelo outro microfone do par. Assim, o p -ésimo TDOA é estimado ao se avaliar o atraso temporal $\hat{\tau}_p$ que maximiza a função de correlação-cruzada (CCF, do inglês *cross-correlation function*).

Assumindo um modelo discreto no tempo onde os sinais captados pelos microfones são amostrados a uma taxa $F \in \mathbb{R}_+$ (em Hz), pode-se calcular a CCF $R_p^{CC}[\cdot] : \mathbb{Z} \rightarrow \mathbb{R}$ como

$$R_p^{CC}[l] \triangleq \sum_{n \in \mathbb{Z}} y_p^{(1)}[n] y_p^{(2)}[n+l], \quad (2)$$

onde $y_p^{(1)}[n]$ representa o sinal real capturado pelo primeiro microfone do p -ésimo par e $y_p^{(2)}[n]$ representa o sinal capturado pelo segundo.

Então, para se estimar o atraso temporal τ_p em segundos a partir da CCF entre os sinais dos microfones do p -ésimo par, basta calcular:

$$\hat{\tau}_p \triangleq \left(\operatorname{argmax}_{l \in \mathbb{Z}} \{R_p^{CC}[l]\} \right) / F, \quad (3)$$

estimativa cuja resolução é limitada, dentre outros fatores, pela taxa de amostragem.

C. Correlação Cruzada Generalizada

Os algoritmos de correlação cruzada generalizada, GCC, são os mais populares para a estimação de TDOA. Um desses algoritmos, proposto em [4], e que é largamente utilizado para o cálculo de correlação, é o da transformada de fase, PHAT. Matematicamente, tal técnica consiste em modificar a definição da correlação cruzada calculado-a no domínio da frequência e normalizando-a adequadamente:

$$R_p^{\text{PHAT}}[l] = \frac{1}{L} \sum_{k=0}^{L-1} \frac{Y_p^{(1)}[k] \cdot \left(Y_p^{(2)}[k]\right)^*}{|Y_p^{(1)}[k] \cdot Y_p^{(2)}[k]|} e^{j\left(\frac{2\pi}{L}\right)lk}, \quad (4)$$

onde $Y_p^{(i)}[k]$ representa a transformada discreta de Fourier de $y_p^{(i)}[n]$ com comprimento $L \in \mathbb{N}$, em que $i \in \{1,2\}$.

A motivação para esse método é realizar um branqueamento no espectro cruzado dos sinais, fazendo com que a correlação seja dependente somente da defasagem entre tais sinais, possibilitando o trabalho com sinais de áudio. Outra característica interessante do PHAT é tornar a estimação de TDOA mais robusta aos efeitos de reverberação [11].

III. TÉCNICAS DE LOCALIZAÇÃO DE FONTES ACÚSTICAS

A. Mínimos Quadrados sem Restrição – LS

A técnica de mínimos quadrados sem restrição (LS, do inglês *least-squares*) [12] estima a posição da fonte por

$$\hat{\mathbf{s}} = (\hat{\Phi}^T \hat{\Phi})^{-1} \hat{\Phi}^T \hat{\mathbf{b}}, \quad (5)$$

onde

$$\hat{\Phi} = \begin{bmatrix} c\hat{\tau}_2 & \mathbf{m}_2^T \\ \vdots & \vdots \\ c\hat{\tau}_M & \mathbf{m}_M^T \end{bmatrix} \quad \text{e} \quad \hat{\mathbf{b}} = \begin{bmatrix} \frac{\|\mathbf{m}_2\| - c\hat{\tau}_2}{2} \\ \vdots \\ \frac{\|\mathbf{m}_M\| - c\hat{\tau}_M}{2} \end{bmatrix}, \quad (6)$$

em que é assumido que um dos microfones do arranjo, numerado como 1, seja utilizado como referência para cálculo dos TDOAs, e que sua posição, denotada por \mathbf{m}_1 , seja tomada como a origem do sistema de coordenadas cartesianas.

Utilizando (5), pode-se estimar a posição da fonte com

$$\hat{\mathbf{s}} = [\mathbf{0}_{3 \times 1} \quad \mathbf{I}_3] \hat{\mathbf{s}}. \quad (7)$$

O método de mínimos quadrados apresenta baixa complexidade computacional, podendo ser utilizado em aplicações em tempo real. No entanto, ele é muito sensível a possíveis erros de estimação de TDOA, sendo aplicado da maneira que é descrito neste artigo somente em situações onde a SNR é alta e o tempo de reverberação é baixo.

B. Máxima Verossimilhança – ML

Os autores em [5] descrevem um método de máxima verossimilhança (ML, do inglês *maximum likelihood*) baseado em TDOA, que mostra uma maneira de determinar o ponto no espaço que possui os TDOAs que mais se assemelham com os TDOAs estimados. A ideia básica do método consiste em dividir a sala em uma grade de pontos $\mathcal{G} \subset \mathbb{R}^3$, onde cada ponto $\mathbf{g} \in \mathcal{G}$ da grade é considerado pelo algoritmo como um candidato a posição de fonte. A partir disso, são avaliados para cada ponto candidato seus TDOAs correspondentes a cada par de microfones. Definindo o erro de TDOA do p -ésimo par

$$e_p(\mathbf{g}) = \tau_p(\mathbf{g}) - \hat{\tau}_p, \quad (8)$$

realiza-se uma busca para avaliar qual ponto da grade possui a menor média de e_p^2 . Matematicamente, a estimação da posição da fonte é definida como

$$\hat{\mathbf{s}} = \operatorname{argmin}_{\mathbf{g} \in \mathcal{G}} \{ \mathbb{E}_P [e_p^2(\mathbf{g})] \}, \quad (9)$$

onde

$$\mathbb{E}_P [e_p^2] = \frac{e_1^2 + e_2^2 + \dots + e_P^2}{P}. \quad (10)$$

C. Mínima Mediana dos Quadrados – LMedS

O método de mínima mediana dos quadrados, LMedS, proposto em [8], emprega a mediana dos erros quadrados de TDOA (e_p^2) como função-custo, em vez da média. Com isso, obtém-se um aumento na robustez do método frente a possíveis TDOAs mal estimados. O LMedS dá menos foco aos grandes erros, suavizando seus efeitos na localização da fonte. O método é dividido em dois passos, denotados como passo A e passo B. O primeiro calcula $\hat{\mathbf{s}}^{(A)} \in \mathcal{G}$ a partir de

$$\hat{\mathbf{s}}^{(A)} = \operatorname{argmin}_{\mathbf{g} \in \mathcal{G}} \{ \mathbb{M}_P [e_p^2(\mathbf{g})] \}, \quad (11)$$

onde $\mathbb{M}_P [e_p^2]$ computa a mediana das amostras $e_1^2, e_2^2, \dots, e_P^2$. A seguir, recalculam-se os TDOAs referentes a todos os pares de microfones assumindo que a fonte realmente esteja em $\hat{\mathbf{s}}^{(A)}$. A partir disso, uma verificação é feita para se descartar possíveis TDOAs mal estimados dado um limiar $\gamma \in \mathbb{R}_+$: se

$|e_p(\hat{\mathbf{s}}^{(A)})| \geq \gamma$, então $\hat{\tau}_p$ é descartado. Então, executa-se o passo B, onde mais uma vez uma busca exaustiva é realizada, assim como ocorre em (11), porém dessa vez utilizando um conjunto reduzido contendo $\mathbb{N} \ni \bar{P} \leq P$ amostras de erro correspondentes às estimativas remanescentes de TDOA:

$$\hat{\mathbf{s}}^{(B)} = \underset{\mathbf{g} \in \mathcal{G}}{\operatorname{argmin}} \left\{ \mathbb{M}_{\bar{P}} [e_p^2(\mathbf{g})] \right\}. \quad (12)$$

Este método de localização tem como ponto forte a robustez perante erros de estimação de TDOA ocorridos por efeito de ruído e de reverberação. No entanto, para se estimar a posição da fonte faz-se necessária pelo menos uma busca em toda a grade de pontos da sala, o que pode demorar, a depender dos recursos computacionais disponíveis. Com base nisso será sugerida a aplicação do método SRP-PHAT juntamente com o LMedS, como será visto na Seção IV.

D. SRP-PHAT Volumétrico

Uma sala com grande tempo de reverberação e altos níveis de ruído pode tornar a estimação dos TDOAs imprecisa, o que acarreta erros de estimação da posição da fonte. O SRP-PHAT combina diversos trechos de correlações cruzadas dos vários pares de microfones, amenizando tal problema. Ele assim o faz por meio de *beamforming* [5], [6], o que permite que o arranjo de microfones privilegie sinais provenientes de uma determinada região de interesse e atenuar sinais oriundos de outras sem a necessidade de um direcionamento físico. Quando o foco do arranjo corresponde à posição da fonte, ou seja, quando o lobo principal do arranjo estiver direcionado para a região da fonte, a função-custo do SRP-PHAT alcança, em teoria, seu máximo.

Em [9], foi proposta uma adaptação do SRP-PHAT para lidar com regiões espaciais (*e.g.*, paralelepípedos retangulares) em vez de pontos. A técnica resultante, chamada SRP-PHAT volumétrico, utiliza também uma busca hierárquica sofisticada com base em algoritmos *branch-and-bound*. Será descrita a seguir uma simplificação da técnica que implementa uma busca exaustiva nas regiões, nas mesmas linhas de [9], [10].

De início, divide-se o espaço de busca $\mathcal{F} \subset \mathbb{R}^3$ em pequenos volumes $\mathcal{V} \in \mathbb{V} \subset \mathcal{P}(\mathcal{F})$ que aqui serão denominados cuboides (paralelepípedos retangulares), onde \mathbb{V} é uma família de conjuntos que exaurem o espaço de busca e $\mathcal{P}(\mathcal{F})$ denota o conjunto das partes de \mathcal{F} . Então, para cada cuboide são computados os TDOAs admissíveis associados àquele volume no espaço referentes a todos os pares de microfones. Pode-se definir o conjunto de TDOAs admissíveis como

$$\Lambda_p(\mathcal{V}) \triangleq \left\{ \tau_p^{(a)} \triangleq \lfloor F \cdot \tau_p(\mathbf{x}) \rfloor \mid \mathbf{x} \in \mathcal{V} \right\} \subset \mathbb{Z}, \quad (13)$$

onde $\lfloor \cdot \rfloor$ representa o operador arredondamento para o inteiro mais próximo e $\tau_p^{(a)}(\mathbf{x})$ denota o TDOA em amostras do p -ésimo par de microfones considerando a posição candidata \mathbf{x} .

Em vez de calcular os muitos TDOAs referentes a cada cuboide, definem-se quais são seus limites mínimos e máximos. O mínimo valor $\tau_{p,\min}^{(a)}(\mathcal{V}) \in \mathbb{N}$ do conjunto $\left\{ \tau_p^{(a)}(\mathbf{x}) \mid \mathbf{x} \in \mathcal{V} \right\}$ pode ser aproximado por [9]

$$\tau_{p,\min}^{(a)}(\mathcal{V}) \approx \min_{\mathbf{x} \in \mathcal{S}_p^v(\mathcal{V})} \left\{ \tau_p^{(a)}(\mathbf{x}) \right\}, \quad (14)$$

onde $\mathcal{S}_p^v(\mathcal{V})$ representa um conjunto que contém as posições (x, y, z) dos vértices de \mathcal{V} . Com isso, conclui-se que dado um cuboide, os TDOAs mínimos daquela região para todos os pares de microfones podem ser calculados considerando cada vértice desse cuboide como candidato a posição de fonte. Então, calcula-se o TDOA mínimo global referente aos 8 possíveis candidatos. Note que para calcular $\tau_{p,\max}^{(a)}(\mathcal{V})$, o raciocínio é o mesmo utilizado para calcular o mínimo.

Assim como SRP-PHAT, sua versão volumétrica privilegia determinadas regiões a partir do direcionamento dos lobos do arranjo de microfones de forma a maximizar a função-custo

$$W_{\text{média}}(\mathcal{V}) = \mathbb{E}_{\mathcal{P}} \left[\sum_{l \in \Lambda_p(\mathcal{V})} R_p^{\text{PHAT}}[l] \right]. \quad (15)$$

IV. SRP-PHAT VOLUMÉTRICO REFINADO POR LMedS

É possível realizar um pré-processamento de forma que o LMedS faça uma busca apenas nos lugares que têm maiores chances de conter a posição da fonte. Tais chances podem ser determinadas através do SRP-PHAT volumétrico. Os cuboides que obtiverem os maiores valores de $W_{\text{média}}(\mathcal{V})$ em (15) seriam os candidatos potenciais a conter a posição da fonte. O uso acoplado do método LMedS como etapa de refinamento do SRP-PHAT volumétrico permite a melhoria da eficiência computacional da técnica de localização, além de conferir robustez e acurácia que não estão presentes na mesma medida no SRP-PHAT volumétrico atuando isoladamente.

A técnica proposta se divide em três passos, denotados como passos A, B e C. O passo A consiste em dividir cada dimensão da região de busca \mathcal{F} em 2^I partes, de forma a se obter uma divisão total da região em 8^I cuboides, $I \in \mathbb{N}$. Então, para todos os cuboides calcula-se a função-custo descrita em (15). Entretanto, não há garantias de que o cuboide que contém a posição da fonte vai ser o que irá obter o maior valor da função-custo. Então, selecionam-se os $V \in \mathbb{N}$ cuboides com maiores valores da função-custo, em vez de somente o cuboide associado ao maior valor. Isso produz maior segurança de que a posição da fonte esteja contida no espaço de busca resultante.

Definindo o conjunto composto pela união dos V cuboides selecionados como $\bar{\mathcal{F}}$, avança-se para o passo B da técnica, o qual consiste em definir uma grade (*e.g.*, uma grade uniforme com uma resolução pré-fixada) de pontos $\mathcal{G}_{\bar{\mathcal{F}}} \subset \bar{\mathcal{F}}$, onde cada ponto $\mathbf{g} \in \mathcal{G}_{\bar{\mathcal{F}}}$ é considerado como candidato a posição de fonte. Utilizando a mesma métrica de avaliação de erro de TDOA do método LMedS, descrita em (8), realiza-se uma busca para se avaliar qual ponto da grade possui a menor mediana de e_p^2 . Este passo é descrito em (11), trocando-se \mathcal{G} por $\mathcal{G}_{\bar{\mathcal{F}}}$.

A partir do cálculo da posição de fonte $\hat{\mathbf{s}}^{(B)} \in \mathbb{R}^3$, utiliza-se a mesma ideia de um limiar para descarte de possíveis TDOAs mal estimados utilizada no LMedS. Se houver TDOAs mal estimados, executa-se o passo C realizando outra busca nas grades de pontos pertencentes ao cuboides, nas mesmas linhas de (12), novamente trocando-se \mathcal{G} por $\mathcal{G}_{\bar{\mathcal{F}}}$, o que resultará na estimativa $\hat{\mathbf{s}}^{(C)} \in \mathbb{R}^3$.

V. AVALIAÇÃO DE DESEMPENHO

Nesta seção será comparado o desempenho das técnicas de localização apresentadas na Seção III (LS, ML e LMedS) frente à técnica proposta na Seção IV (SRP+LMedS).

A. Metodologia

Foram realizadas diversas simulações com o objetivo de comparar o desempenho dos métodos de localização de fontes abordados neste artigo. Visando a simular diversas condições reais de um experimento prático, foram aplicados diversos níveis de ruído nos sinais captados pelos microfones. Além disso, foram empregados diferentes níveis de RT60 (tempo de reverberação [4]) para a sala utilizada nas simulações.

B. Procedimentos das Simulações

As dimensões da sala utilizada para as simulações foram de 5,2 m × 7,5 m × 2,6 m. Cinco diferentes posições para a fonte foram utilizadas, sendo o arranjo composto por 16 microfones. As posições de fonte e microfones foram baseadas nos procedimentos de simulações descritos em Fig. 4 e TABLE I de [9].¹ Todos os sinais captados pelos microfones foram corrompidos por ruído branco aditivo gaussiano (AWGN, do inglês *additive white Gaussian noise*). Durante as simulações, um detector de atividade de voz [14], [15] (VAD, do inglês *voice-activity detector*) foi aplicado antes da execução do sinal da fonte, o qual era um sinal de voz feminina de 4,5 s de duração dividido em 60 blocos de 100 ms com 25 ms de sobreposição, a uma taxa de amostragem de 48 kHz. A velocidade de propagação c da onda foi considerada 340 m/s, referente a uma temperatura de 20 °C.

Para as técnicas ML e LMedS, foi utilizada uma grade de pontos uniforme em cada direção cartesiana com a menor distância entre pontos vizinhos de 10 cm, totalizando 67932 pontos. Foi descartado qualquer ponto da grade que estivesse a uma distância inferior a 10 cm das paredes laterais, assim como qualquer ponto que estivesse a menos de 30 cm do chão da sala ou a 60 cm do teto. Tais premissas se justificam pois é improvável que a fonte que se deseja localizar esteja nos lugares em que foram descartados os pontos da grade, especialmente se a fonte for uma pessoa falando.

Para a técnica proposta SRP+LMedS a sala foi dividida em $8^4 = 4096$ cuboides. Foram selecionados os 300 cuboides que possuíam os menores valores da função-custo. Com isso, torna-se bastante provável que a posição da fonte estaria contida no espaço de busca da técnica. É importante mencionar que cada cuboide, de dimensões 32,5 cm × 46,9 cm × 16,3 cm, possuía uma grade de pontos espaçados de 10 cm, de forma que o total de pontos em cada volume era 90.

Os métodos que utilizam descarte através dos erros e_p (LMedS e SRP+LMedS) utilizaram $\gamma = 1$ ms como valor máximo de $|e_p|$.

As TABELAS I-IV mostram os resultados das simulações para SNRs de 10 dB, 20 dB, 30 dB, sem ruído e para cinco diferentes valores de RT60. As linhas das tabelas mostram estatísticas do erro de localização considerando todos os

blocos de sinal emitidos pela fonte e suas cinco diferentes posições ($60 \times 5 = 300$ erros de localização no total). Foi assumido que, quando qualquer erro fosse superior a 1 metro, o respectivo método falhou, o que é indicado nas tabelas por um traço.

C. Resultados

As TABELAS I-IV mostram que as técnicas clássicas da literatura apresentam pequenos erros de localização somente para RT60 abaixo de 200 ms. Isso é válido mesmo quando não há a presença de ruído, como mostra a TABELA I, o que indica a baixa robustez dos métodos à reverberação. É importante lembrar que tais técnicas são sensíveis a grandes erros de estimação de TDOA, cuja ocorrência pode ser observada ao se analisar o desvio padrão para os métodos LS e ML de todas as tabelas de resultados aqui mostradas, em especial a TABELA IV, a qual só apresentou valores de desvio padrão abaixo de 1 metro no ambiente anecoico. Note que o método LS apresentou melhores resultados para RT60 de 200 ms do que para 0 ms na TABELA I. Na verdade, os resultados foram bem próximos, possivelmente dentro da incerteza do método LS, cuja acurácia é sensível a pequenas diferenças nas estimativas de TDOAs.

O método LMedS e o método SRP+LMedS são os que, em geral, apresentaram os menores erros de localização dentre os métodos abordados neste artigo. Observando a mediana do erro para as duas técnicas, pode-se afirmar que somente para SNR de 10 dB e RT60 a partir de 400 ms (ambiente extremamente hostil à localização) é que o método SRP+LMedS apresenta erros de localização muito maiores que o método LMedS. Do ponto de vista computacional, esses resultados são promissores, pois se conseguiu manter o desempenho da técnica LMedS nos casos práticos de interesse mesmo utilizando apenas uma parcela muito menor do espaço de busca, correspondente a 40% do espaço original. É válido ressaltar que a técnica LMedS avaliou sua função custo ao menos 67932 vezes (tal número de avaliações pode crescer se o passo B do método LMedS for executado), enquanto que na execução da técnica SRP+LMedS foram feitas 4096 chamadas da função custo (15) e, no mínimo, 27000 avaliações de função custo do LMedS (que também pode se elevar, a depender da execução do passo C do método SRP+LMedS). Também é importante mencionar que o ganho computacional proporcionado pela diminuição da região de busca no método SRP+LMedS é dependente da forma como é implementada a seleção dos cuboides, o que significa que o percentual de diminuição da região de busca não necessariamente se traduz no mesmo percentual de ganho computacional.

VI. CONCLUSÕES E TRABALHOS FUTUROS

Este artigo propôs uma nova técnica de localização de fontes acústicas que acopla dois métodos recentemente propostos pelo grupo de trabalho dos autores, a saber: o emprego do método SRP-PHAT volumétrico com refinamento através do método LMedS. Os resultados obtidos mostram que tal técnica apresentou baixos erros de localização mesmo para elevados níveis de ruído e em ambientes muito reverberados. Tal desempenho foi obtido com uma redução significativa do

¹Cabe destacar aqui três pontos: (i) a técnica proposta não se restringe a geometrias específicas de arranjos de microfones, sendo possível utilizá-la em cenários não-estruturados como o utilizado aqui; (ii) devido à relativamente grande abertura (*aperture*) do arranjo e ao uso de uma taxa de amostragem elevada, não se faz necessário o emprego de técnicas de atraso fracionário (*fractional delay*); (iii) embora a abertura do arranjo seja um fator determinante para se evitar a ocorrência de *aliasing* espacial no contexto de sinais harmônicos de banda estreita (*narrowband*), o presente trabalho lida com sinais de banda larga (*wideband*) e, portanto, o *aliasing* espacial não é um fator predominante, conforme indicado em [13]. De fato, a acurácia dos resultados obtidos aqui corrobora com tal constatação.

TABELA I
ERRO DE LOCALIZAÇÃO (CM) PARA SNR INFINITA.

	RT60 (ms)	0	200	400	600	800
Mediana	LS	5,08	3,32	—	76,73	—
	ML	5,03	5,03	30,79	64,42	56,38
	LMedS	5,03	5,03	5,03	5,03	5,03
	SRP+LMedS	5,76	5,76	5,76	6,13	6,13
Média	LS	5,28	3,79	—	—	—
	ML	6,94	6,94	33,99	—	52,07
	LMedS	6,94	6,94	6,94	6,94	6,94
	SRP+LMedS	9,29	9,29	7,49	6,43	7,57
Desvio Padrão	LS	3,27	1,88	—	—	—
	ML	4,46	4,46	26,56	—	38,46
	LMedS	4,49	4,49	4,49	4,49	4,49
	SRP+LMedS	7,97	7,97	4,51	2,65	4,51

TABELA II
ERRO DE LOCALIZAÇÃO (CM) PARA SNR DE 30 dB

	RT60 (ms)	0	200	400	600	800
Mediana	LS	5,08	5,08	—	—	—
	ML	5,03	5,03	36,85	—	—
	LMedS	5,03	5,03	5,03	5,03	5,03
	SRP+LMedS	5,76	6,13	6,13	6,13	6,13
Média	LS	5,14	6,53	—	—	—
	ML	6,94	7,93	99,11	—	—
	LMedS	6,94	7,26	6,94	8,21	7,89
	SRP+LMedS	9,32	9,05	10,08	10,16	10,32
Desvio Padrão	LS	3,61	6,00	—	—	—
	ML	4,46	12,50	—	—	—
	LMedS	4,49	4,99	4,49	6,17	5,81
	SRP+LMedS	7,96	7,51	8,29	8,05	12,13

TABELA III
ERRO DE LOCALIZAÇÃO (CM) PARA SNR DE 20 dB

	RT60 (ms)	0	200	400	600	800
Mediana	LS	6,58	13,03	—	—	—
	ML	5,03	9,41	—	—	—
	LMedS	5,03	5,03	5,03	7,85	9,41
	SRP+LMedS	5,94	6,13	6,13	7,80	18,57
Média	LS	9,21	43,06	—	—	—
	ML	7,13	32,69	—	—	—
	LMedS	7,57	7,57	7,89	9,51	11,35
	SRP+LMedS	9,04	9,15	10,34	16,60	31,08
Desvio Padrão	LS	11,23	—	—	—	—
	ML	4,77	86,00	—	—	—
	LMedS	5,43	5,43	5,82	8,03	10,62
	SRP+LMedS	7,52	7,10	8,49	19,72	39,63

TABELA IV
ERRO DE LOCALIZAÇÃO (CM) PARA SNR DE 10 dB.

	RT60 (ms)	0	200	400	600	800
Mediana	LS	14,60	—	—	—	—
	ML	5,03	68,25	—	—	—
	LMedS	5,03	5,03	8,96	9,41	9,41
	SRP+LMedS	5,94	6,24	18,11	22,66	53,56
Média	LS	91,22	—	—	—	—
	ML	15,49	—	—	—	—
	LMedS	6,94	7,88	12,22	16,13	28,04
	SRP+LMedS	8,81	10,41	23,58	65,16	78,15
Desvio Padrão	LS	—	—	—	—	—
	ML	48,54	—	—	—	—
	LMedS	4,49	6,08	13,14	24,22	47,89
	SRP+LMedS	6,84	8,60	30,32	78,87	69,05

espaço de busca originalmente utilizado pela técnica LMedS,

tornando o sistema ao mesmo tempo eficiente e robusto.

Os trabalhos futuros devem focar em técnicas mais sofisticadas de filtragem de possíveis TDOAs mal estimados e em algoritmos que refinem ainda mais o volume da busca do método LMedS, o que possibilitaria seu uso em aplicações em tempo real. Além disso, experimentos práticos podem ser realizados visando a atestar o desempenho dos métodos de localização de fontes perante condições reais.

AGRADECIMENTOS

Os autores são gratos pelo suporte financeiro de CNPq, Capes, e Faperj.

REFERÊNCIAS

- [1] Borzino, A. M. C. R., Apolinario, J. A. e de Campos, M. L. R., "Estimating direction of arrival of long range gunshot signals," *Anais do International Telecommunications Symposium (ITS 2014)*, pp. 1–5, São Paulo, Brasil, Agosto 2014.
- [2] Lv, X., Zhang, M., Yuan, G., Chen, Q. e Zhao, H., "Robot sound source search strategy based on multi-blackboard model," *Anais da IEEE International Conference on Robotics and Biomimetics (ROBIO 2010)*, pp. 633–638, Tianjin, China, Dezembro 2010.
- [3] Maganti, H. K., Gatica-Perez, D. e McCowan, I., "Speech enhancement and recognition in meetings with an audio-visual sensor array," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 8, pp. 2257–2269, Novembro 2007.
- [4] Benesty, J., Chen, J. e Huang, Y., *Microphone Array Signal Processing*, Berlim, Alemanha, Springer 2008.
- [5] Brandstein, M. e Ward, D., *Microphone Arrays: Signal Processing Techniques and Applications*, Berlim, Alemanha, Springer, 2001.
- [6] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, Março 1986.
- [7] Knapp, C. e Carter, G., "The generalized correlation method for estimation of time delay," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, no. 4, pp. 320–327, Agosto 1976.
- [8] da Silva, F. B. e Martins, W. A., "Robust TDOA-based sound source localization," *Anais do XXXIII Simpósio Brasileiro de Telecomunicações (SBt 2015)*, pp. 1–2, Juiz de Fora, Brasil, Setembro 2015.
- [9] Nunes, L. O., Martins, W. A., Lima, M. V. S., Biscainho, L. W. P., Costa, M. V. M., Goncalves, F. M., Said, A. e Lee, B., "A steered-response power algorithm employing hierarchical search for acoustic source localization using microphone arrays," *IEEE Transactions on Signal Processing*, vol. 62, no. 19, pp. 5171–5183, Outubro 2014.
- [10] Lima, M. V. S., Martins, W. A., Nunes, L. O., Biscainho, L. W. P., Ferreira, T. N., Costa, M. V. M. e Lee, B., "A volumetric SRP with refinement step for sound source localization," *IEEE Signal Processing Letters*, vol. 22, no. 8, pp. 1098–1102, Agosto 2015.
- [11] Gustafsson, T., Rao, B. D. e Trivedi, M., "Source localization in reverberant environments: modeling and statistical analysis," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 6, pp. 791–803, Novembro 2003.
- [12] Stoica, P. e Li, J., "Source localization from range-difference measurements," *IEEE Signal Processing Magazine*, vol. 23, no. 6, pp. 63–66, Novembro 2006.
- [13] Dmochowski, J., Benesty, J., e Affes, S., "On spatial aliasing in microphone arrays," *IEEE Transactions on Signal Processing*, vol. 57, no. 4, pp. 1383–1395, Abril 2009.
- [14] Gazor, S. e Zhang, W., "A soft voice activity detector based on a Laplacian-Gaussian model," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 5, pp. 498–505, Setembro 2003.
- [15] Sohn, J., Kim, N. S., Sung, W., "A statistical model-based voice activity detection," *IEEE Signal Processing Letters*, vol. 6, no. 1, pp. 1–3, Janeiro 1999.