

Sobre o uso de GCC-PHAT na estimação de DoA para arranjos de dimensão reduzida

Felipe G. Serrenho, Luiz C. Sampaio Ramos e José A. Apolinário Jr.

Resumo— Este trabalho apresenta a importância da utilização de pesos denominados transformada de fase (PHAT) na correlação cruzada generalizada (GCC) para estimar corretamente a diferença de tempo de chegada entre microfones em arranjos de dimensões reduzidas. São apresentadas variações que visam aumentar a robustez quando o sinal de interesse é espectralmente esparsos e há presença de interferidores banda larga. Aborda-se a importância da utilização de algoritmos de interpolação para compensar os efeitos da discretização da função de correlação cruzada. A estimação da direção de chegada (DoA) utiliza um algoritmo de busca exaustiva de acordo com o critério de mínimos quadrados.

Palavras-Chave— GCC-PHAT, arranjo de microfones, diferença de tempo de chegada.

Abstract— This work presents the importance of the use of Phase Transform (PHAT) weights in generalized cross-correlation (GCC) for the correct estimation of time difference of arrival between microphones in small sized arrays. Are presented modified versions which aim to increase robustness when the signal of interest is spectrally sparse and there is broadband interferers. The importance of the use of interpolation algorithms as a way of compensating for the effects of the discretization of the cross-correlation function is also discussed. The direction of arrival (DoA) estimation is performed using an exhaustive search algorithm according to the least-squares criterion.

Keywords— GCC-PHAT, microphone array, time-delay estimation.

I. INTRODUÇÃO

A estimação da direção de chegada (DoA, do inglês *Direction of Arrival*) é um tópico de grande interesse acadêmico, possuindo aplicações em diversas áreas tais como radar [1], sonar [2], comunicações sem fio (*wireless*) [3], biomédica [4] e localização de alvos [5], [6].

Em geral divide-se os estimadores de DoA em dois grupos, de acordo com o sinal da fonte a ser localizada: os estimadores banda estreita e os banda larga. Em geral, os estimadores banda estreita utilizam arranjos de dimensões da ordem de metade do comprimento de onda do sinal da portadora [7]. Dentre os estimadores banda larga destacamos dois tipos: os baseados em estimadores banda estreita e os baseados em diferença de tempo de chegada (TDoA, do inglês *Time Difference of Arrival*) entre os elementos utilizando mínimos quadrados (LS, do inglês *Least-Squares*) [8]. Os estimadores do segundo tipo apresentam melhor desempenho com arranjos com sensores mais espaçados. Uma das principais vantagens dos estimadores de DoAs baseados em TDoA com minimização LS é uma

Felipe G. Serrenho, e-mail: felipe.serrenho@gmail.com; Luiz C. Sampaio Ramos, e-mail: sampaio@ime.eb.br; José A. Apolinário Jr, e-mail: apolin@ime.eb.br. Instituto Militar de Engenharia, Rio de Janeiro-RJ, Brasil. Este trabalho foi financiado parcialmente pela CAPES.

menor complexidade computacional quando comparados aos algoritmos baseados em busca em grade, típicos das técnicas SRP (do inglês, *steered response power*) [9], favorecendo aplicações em tempo real [10].

Neste trabalho procura-se conciliar o baixo custo computacional dos métodos baseados em TDoA, com a conveniência de utilizar arranjos de dimensões reduzidas, minimizando o impacto na precisão e robustez. É mostrado a vantagem da utilização do GCC-PHAT (do inglês, *generalized cross correlation phase transform*) sobre o GCC clássico em arranjos de dimensões reduzidas e, em seguida, propostas modificações no GCC-PHAT para aumentar sua robustez quando o sinal de interesse (SoI, do inglês *signal of interest*) é espectralmente esparsos e há a presença de interferidores banda-larga. Na sequência são aplicados algoritmos de interpolação para compensar a discretização da função de correlação-generalizada. Por fim, é utilizado um algoritmo de busca exaustiva para selecionar o subconjunto dos pares que fornece a melhor estimativa da DoA de acordo com um critério LS.

II. HARDWARE DE GRAVAÇÃO

Foi utilizado o arranjo circular UMA-8, da MiniDSP [11], composto por $N = 7$ microfones: um no centro da placa, e os outros seis dispostos nos vértices de um hexágono regular inscrito num círculo de diâmetro de 8,6cm, conforme a Fig. 1. A UMA-8 pode operar com frequências de amostragem de até 48kHz e 24 bits de resolução. Ela possui interface USB, através da qual é possível coletar os sinais de áudio dos microfones.

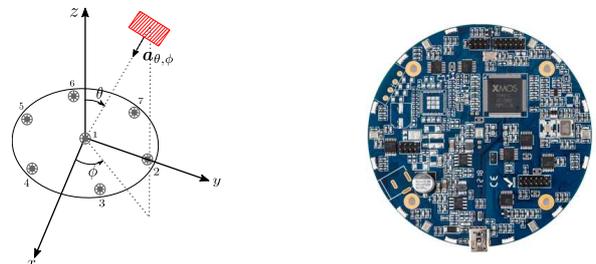


Fig. 1. Arranjo circular UMA-8, utilizado no experimento [11].

III. TÉCNICAS

Nesta seção são apresentadas as técnicas utilizadas na estimação da DoA, com destaque para a estimação da TDoA.

A. Correlação Cruzada Generalizada

Para estimar as TDoAs pode-se utilizar algoritmos baseados em GCC. A correlação-cruzada é definida como [12]:

$$r_{x_i x_j}(\tau) = E[x_i(k)x_j(k - \tau)]. \quad (1)$$

Como as estatísticas dos sinais no mundo real usualmente não são disponíveis para computar a Eq. (1), aproxima-se a correlação cruzada pela sua média temporal:

$$\hat{r}_{x_i x_j}(\tau) = \sum_{k=-\infty}^{\infty} x_i(k)x_j(k-\tau) = x_i(\tau) * x_j(-\tau), \quad (2)$$

sendo que $*$ representa a operação de convolução.

Definindo-se densidade espectral de potência como $R_x(e^{jw}) = \mathcal{F}\{r_x(\tau)\}$, pode-se estimar a densidade espectral cruzada de $x_i(k)$ e $x_j(k)$ como:

$$\begin{aligned} \hat{R}_{x_i x_j}(e^{jw}) &= \mathcal{F}\{\hat{r}_{x_i x_j}(\tau)\} = \mathcal{F}\{x_i(\tau) * x_j(-\tau)\} \\ &= X_i(e^{jw})X_j(e^{-jw}) = X_i(e^{jw})X_j^*(e^{jw}), \end{aligned} \quad (3)$$

para o caso de seqüências reais.

A partir desta equação chegamos ao método GCC clássico:

$$\hat{r}_{x_i x_j}^{\text{GCC}}(\tau) = \mathcal{F}^{-1}\{\hat{R}_{x_i x_j}(e^{jw})\}. \quad (4)$$

E o atraso é estimado por:

$$\hat{\tau}_{ij} = \max_{|\tau| \leq \tau_{\max}} |\hat{r}_{x_i x_j}(\tau)|, \quad (5)$$

onde τ_{\max} é o atraso máximo possível entre os microfones i e j , ou seja $|\tau| \leq \tau_{\max}$ é a região de atrasos válida para o referido par de sensores.

Contudo, devido ao arranjo ter dimensões reduzidas, a GCC clássica apresenta o lóbulo principal muito largo em relação a região de τ possíveis. Conforme pode ser visto na Fig. 2, na qual o GCC foi gerado a partir da simulação de recepção de um sinal de voz puro, sem ruídos ou reverberação, entre dois dos microfones o mais afastados possíveis no arranjo, sensores 2 e 5 da Fig. 1. Esse fenômeno faz com que o método GCC clássico seja vulnerável à reverberação, uma vez que uma reverberação irá gerar um novo lóbulo principal ligeiramente deslocado do original, dentro da região de atrasos possíveis, alterando o máximo da correlação.

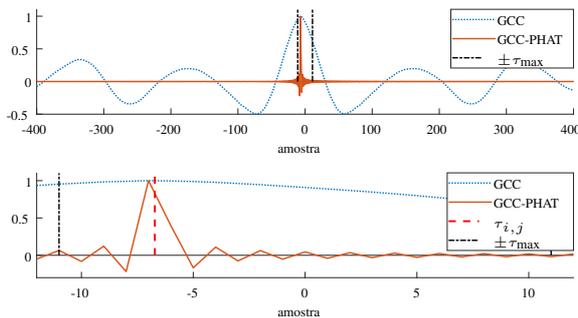


Fig. 2. GCC clássico e GCC-PHAT. Acima, visão panorâmica. Abaixo, zoom na região de τ possíveis.

Pode-se inserir pesos diferenciados em cada componente de frequência de $\hat{R}_{x_i x_j}(e^{jw})$, dando origem a diferentes funções GCC. Para identificar melhor os pesos utilizados, reescreve-se a Eq. (4) como:

$$\hat{r}_{x_i x_j}(\tau) = \mathcal{F}^{-1}\{\psi(w)\hat{R}_{x_i x_j}(e^{jw})\}, \quad (6)$$

onde, para o método GCC clássico, $\psi(w) = 1, \forall w$.

O algoritmo PHAT é conhecido por apresentar bom desempenho contra reverberações [12]. A função de peso do PHAT é [13]:

$$\psi^{\text{PHAT}}(w) = \frac{1}{|X_i(e^{jw})X_j(e^{-jw})|}. \quad (7)$$

A GCC-PHAT apresenta um pico de correlação mais agudo do que a GCC tradicional, melhorando a precisão da TDE [14] (do inglês, *time delay estimation*), conforme pode ser visto na Fig. 2, onde as funções de correlação foram normalizadas para facilitar a comparação.

O GCC-PHAT dá o mesmo peso para todas as frequências; isso faz com que a estimativa do TDoA para sinais espectralmente esparsos seja vulnerável a interferências de maior banda, gerando erros mesmo com SNRs relativamente elevadas. Para aumentar a robustez do GCC-PHAT foi adicionado ao denominador um termo de normalização:

$$\psi^{\text{NPHAT}}(w) = \frac{1}{|X_i(e^{jw})X_j(e^{-jw})| + \frac{\sigma(X_i(e^{jw})X_j(e^{jw}))}{100}}, \quad (8)$$

onde σ é o operador desvio padrão. Esse termo é responsável por minimizar a influência das frequências com menor potência.

Outro método de aumentar a robustez da GCC-PHAT é elevar o denominador por um fator $\alpha \in]0, 1[$. Essa técnica foi proposta em [15] como uma forma de aumentar a robustez do GCC-PHAT em relação a ruídos. No artigo em questão o fator α é variado de acordo com a estimativa do SNR. Neste trabalho usaremos um valor fixo desse parâmetro e seu objetivo principal será aumentar a robustez dos sinais espectralmente esparsos. Assim a nova variante da GCC-PHAT, denominada neste trabalho de GCC- α PHAT, tem a função peso dada por:

$$\psi^{\alpha\text{PHAT}}(w) = \frac{1}{|X_i(e^{jw})X_j(e^{-jw})|^\alpha}. \quad (9)$$

Por último, fazemos uma composição das duas modificações; dessa forma dá-se origem ao GCC- α NPHAT:

$$\psi^{\alpha\text{NPHAT}}(w) = \frac{1}{|X_i(e^{jw})X_j(e^{-jw})|^\alpha + \frac{\sigma(X_i(e^{jw})X_j(e^{jw}))}{100}}. \quad (10)$$

A Fig. 3 mostra as diferenças entre as funções GCC-PHAT propostas em um cenário no qual o sinal de interesse é voz e está presente um interferidor banda-larga (ruído-branco). O sinal possui SNR de 20dB, mas, mesmo assim, o pico do GCC-PHAT corresponde ao τ do interferidor.

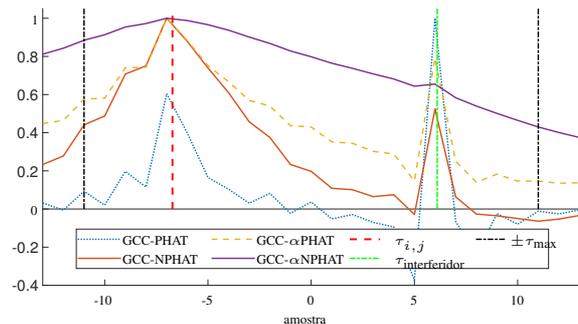


Fig. 3. Desempenho de diferentes funções GCC-PHAT de um sinal de voz na presença de um interferidor banda-larga com SNR de 20dB. $\alpha = 0,6$.

B. Interpolação

Uma limitação dos métodos baseados em GCC é que o atraso estimado entre dois microfones é um número inteiro de amostras, ou seja, um múltiplo do período de amostragem. Isto pode ocasionar estimativas imprecisas mesmo considerando uma captação perfeita (sem ruídos, interferências ou distorções), conforme visto na Fig. 4. Este fenômeno tem maior impacto no desempenho da TDE quanto menor o espaçamento entre os sensores (redução dos atrasos) e menor a frequência de amostragem (período de amostragem grande).

Para minimizar a influência dessa fonte de erros poderíamos propor dois tipos de técnicas: a primeira é aumentar a taxa de amostragem (*upsampling*) dos sinais captados antes do cálculo da GCC; e a segunda é utilizar um algoritmo de interpolação em cada τ_{ij} .

O *upsampling* teria o impacto positivo de diminuir o período de amostragem, aumentando a resolução da correlação cruzada. Contudo, conforme visto na Seção III-A, os métodos baseados em GCC-PHAT perdem de desempenho em sinais espectralmente esparsos (a operação de *upsampling* gera uma vazia na parte superior do espectro). Adicionalmente, aumentar a frequência de amostragem impacta de maneira relevante o custo computacional, uma vez que aumenta o tamanho dos vetores sobre os quais são efetuadas as operações.

A seguir abordamos os métodos baseados em interpolação. Serão abordadas a interpolação cúbica [16] e a interpolação através de sinais *sinc* [15], também conhecida por interpolação de *Whittaker–Shannon* [17].

Na interpolação cúbica foram utilizados todos os pontos compreendidos entre $-\tau_{\max}$ e τ_{\max} ; desta forma cobre-se todos os valores de atraso possíveis sem haver perda de informação.

Na interpolação utilizando *sinc* é utilizada a seguinte expressão [18]:

$$\tau_{x_i x_j}^{\text{sinc}}(\tau) = \sum_{t=-\tau_{\max}-\Delta}^{\tau_{\max}+\Delta} \hat{\tau}_{x_i x_j}(t) \frac{\text{sen}(\tau-t)}{\tau-t}, \quad (11)$$

onde Δ é um fator que permite que amostras vizinhas ao intervalo de $-\tau_{\max}$ a τ_{\max} influenciem o valor da função dentro do intervalo. Porém, a busca do máximo de $\tau_{x_i x_j}^{\text{sinc}}$ deve se manter restrita ao intervalo em que $|\tau| < \tau_{\max}$. Aumentar o valor de Δ aumenta a precisão da interpolação *sinc*, porém aumenta sua complexidade computacional.

Para evidenciar o impacto de dimensões reduzidas do arranjo e de uma frequência de amostragem baixa é mostrado na Fig. 4 o efeito das diferentes interpolações na estimação da TDoA entre um par de microfones.

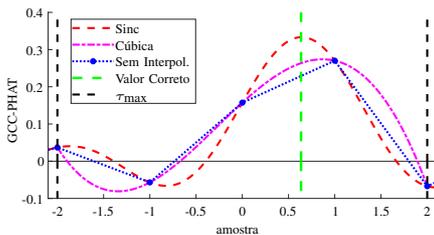


Fig. 4. GCC-PHAT entre um par de microfones e suas interpolações. Frequência de amostragem de 8KHz

C. Solução de mínimos quadrados (LS)

Uma DoA é caracterizada por dois ângulos [19]: o azimute (ϕ) e o zênite (θ), conforme a Fig. 1. Explicitando o vetor unitário $\mathbf{a}_{\theta,\phi}$ na direção de propagação da onda, escreve-se:

$$\mathbf{a}_{\theta,\phi} = - \begin{bmatrix} \text{sen}(\theta)\cos(\phi) \\ \text{sen}(\theta)\text{sen}(\phi) \\ \cos(\theta) \end{bmatrix}. \quad (12)$$

Obtidos os TDoAs τ_{ij} para todos os pares possíveis dos N microfones do arranjo, o atraso (em segundos) entre os microfones i e j é expresso por $\tau_{ij} = \frac{\tau_{ij}}{f_s}$. Dessa forma, tem-se que:

$$\bar{\tau}_{ij} = \frac{d_{ij}}{v_{\text{som}}} = \frac{\mathbf{a}_{\theta,\phi}^T \mathbf{p}_i - \mathbf{a}_{\theta,\phi}^T \mathbf{p}_j}{v_{\text{som}}} = \mathbf{a}_{\theta,\phi}^T \Delta \bar{\mathbf{p}}_{ij}, \quad (13)$$

onde \mathbf{p}_i representam as coordenadas do i -ésimo microfone e v_{som} representa a velocidade do som no ar.

A solução de mínimos quadrados (LS) [8] considera todos os pares de microfones disponíveis para encontrar uma estimativa da DoA [20] minimizando, com respeito a $\mathbf{a}_{\theta,\phi}$, a função de custo mínimos quadrados dada por:

$$\xi(\theta, \phi) = (\bar{\tau}_{12} - \Delta \bar{\mathbf{p}}_{12}^T \mathbf{a}_{\theta,\phi})^2 + (\bar{\tau}_{13} - \Delta \bar{\mathbf{p}}_{13}^T \mathbf{a}_{\theta,\phi})^2 + \dots + (\bar{\tau}_{(N-1)N} - \Delta \bar{\mathbf{p}}_{(N-1)N}^T \mathbf{a}_{\theta,\phi})^2, \quad (14)$$

um total de $\frac{N(N-1)}{2}$ termos. Minimizando Eq. (14) em relação a $\mathbf{a}_{\theta,\psi}$ obtém-se:

$$\hat{\mathbf{a}}_{\text{DoA}} = \mathbf{R}^{-1} \mathbf{p}, \quad (15)$$

onde:

$$\begin{aligned} \mathbf{R} &= \Delta \bar{\mathbf{p}}_{12} \Delta \bar{\mathbf{p}}_{12}^T + \dots + \Delta \bar{\mathbf{p}}_{(N-1)N} \Delta \bar{\mathbf{p}}_{(N-1)N}^T; \\ \mathbf{p} &= \bar{\tau}_{12} \Delta \bar{\mathbf{p}}_{12} + \dots + \bar{\tau}_{(N-1)N} \Delta \bar{\mathbf{p}}_{(N-1)N}. \end{aligned} \quad (16)$$

Na Eq. (15), encontra-se a DoA em coordenadas retangulares, ou seja, $\mathbf{a}_{\text{DoA}} = [a_x \ a_y \ a_z]^T$. Contudo não há garantias de que esta equação forneça um vetor de norma unitária, o que deve ser assegurado através de uma normalização. Os ângulos azimute e zênite podem ser encontrados por simples operações trigonométricas, de acordo com a Eq. (12).

É importante ressaltar que a Eq. (15) só pode fornecer as três coordenadas se o arranjo for espacial; caso, por exemplo, o arranjo seja planar, há ambiguidade, e a matriz \mathbf{R} não é inversível. No caso do arranjo estar contido em um dos planos dos eixos, como o utilizado nesse trabalho que está contido no plano xy , podemos adaptar as variáveis que guardam as posições dos sensores (\mathbf{p}_i), de forma a suprimir a coordenada do eixo perpendicular, no nosso caso a coordenada z . Dessa forma a matriz \mathbf{R} é inversível e a Eq. (15) fornece o resultado $\hat{\mathbf{a}}_{\text{DoA}}^{\text{incompleto}} = [a_x \ a_y]^T$. Sabendo que \mathbf{a}_{DoA} , quando com as três coordenadas, deve ser um vetor unitário e assumindo que a fonte está acima ou abaixo do arranjo, pode-se calcular o vetor $\hat{\mathbf{a}}_{\text{DoA}}$ completo.

A captação de sinais está sujeita a fatores externos como ruído, multipercursos, interferidores, entre outros. Esses fatores podem fazer com que o atraso relativo entre alguns pares de microfones apresentem erros consideravelmente maiores que os demais pares, os chamados *outliers*. Para atenuar

TABELA I
SIMULAÇÃO COM FS=44,1kHz - VOZ

	Sem interpolação		Interpolação cúbica		Interpolação <i>sinc</i>	
	Erro Médio	Desvio Padrão	Erro Médio	Desvio Padrão	Erro Médio	Desvio Padrão
GCC-PHAT	9,5178°	11,0484°	8,4069°	13,0308°	7,5843°	13,2529°
GCC-NPHAT	7,9557°	7,9614°	5,9015°	8,0753°	4,9494°	8,7307°
GCC- α PHAT°	7,9688°	7,7662°	5,8509°	8,0564°	4,9058°	8,6335°
GCC- α NPHAT°	7,8247°	7,5626°	5,5294°	7,4034°	5,3042°	7,1348°

o efeito desses erros espúrios utiliza-se algoritmos capazes de selecionar quais pares devem ser utilizados, dentre os quais, podemos destacar o ILS (do inglês, *iterative least-squares*) [21] e o de Busca Exaustiva (ES, do inglês *exhaustive search*) [22]. A diferença entre eles é, basicamente, que o ILS utiliza iterações para remover um par de microfones por vez, o que resulta no maior decréscimo da função custo ξ , até restar uma quantidade de n pares pré-definida. Por outro lado, o ES testa todas as combinações de n pares e seleciona a que resulta na menor função custo (ξ). Claramente o ILS é um algoritmo guloso (sub-ótimo) e de menor complexidade computacional, enquanto que o ES obtém melhores resultados ao buscar a solução ótima do ponto de vista da função custo.

IV. SIMULAÇÃO

A fim de estudar o comportamento estatístico dos métodos apresentados, foi utilizado um sinal de voz de 5 segundos de duração como sinal de interesse e permitiu-se que o azimute variasse aleatoriamente entre 0° e 360° e o zênite, entre 0° e 90°. Foram realizadas 100 iterações independentes com frequência de amostragem de 44,1kHz.

Para simular o ambiente foi utilizado o pacote *MCRoomSim* para *MATLAB* [23] que fornece a resposta impulsiva esperada do ambiente. Desta forma, as simulações consideraram a reverberação do ambiente. O ambiente simulado foi uma sala de 6,4m por 6,4m com 3 m de altura, com RT60 de aproximadamente 200ms. Foi considerada uma SNR de 10 dB, sendo utilizado ruído espacialmente branco. Ainda foi considerada um interferidor banda-larga (ruído-branco) chegando no arranjo de uma direção desconhecida com SNI (relação entre o sinal de interesse e o interferidor) de 25dB.

Foi utilizado $\Delta = 10$ na interpolação *sinc*, Eq. (11), e passo de interpolação de 0,001° para as duas técnicas. No algoritmo ES foi utilizado $n = 13$ de um total de 21 possíveis TDoAs.

O erro na estimação da DoA foi calculado como o ângulo entre a direção de chegada correta e a estimada; como ambos os vetores são unitários esse cálculo pode ser realizado da seguinte forma:

$$\text{Erro Angular} = |\cos^{-1}(\mathbf{a}_{\text{DoA}}^T \hat{\mathbf{a}}_{\text{DoA}})|. \quad (17)$$

Essa forma de calcular o erro permite comparar objetivamente duas estimações, ao mesmo tempo que evita levar em consideração as distorções na estimação do azimute quando o zênite possui valor próximo de zero.

De acordo com os resultados exposto na Tabela I, relativo ao sinal de voz, espectralmente esparsos, percebe-se ganho de desempenho ao se utilizar as técnicas GCC- α PHAT e GCC- α NPHAT, principalmente quando aplicadas em conjunto com a interpolação *sinc*, a qual apresentou o melhor desempenho

TABELA II
SIMULAÇÃO COM FS=44,1kHz - RUÍDO BRANCO

	Sem interpolação		Interpolação cúbica		Interpolação <i>sinc</i>	
	Erro Médio	Desvio Padrão	Erro Médio	Desvio Padrão	Erro Médio	Desvio Padrão
GCC-PHAT	6,5396°	4,3447°	3,5422°	3,0195°	0,1585°	0,1221°
GCC-NPHAT	6,5396°	4,3447°	3,5332°	3,0106°	0,1580°	0,1184°
GCC- α PHAT	6,5396°	4,3447°	3,4553°	2,9466°	0,1646°	0,1278°
GCC- α NPHAT	6,5396°	4,3447°	3,3859°	2,9312°	0,1830°	0,1613°

em todos os cenários simulados. Já quando o SoI é banda-larga, Tabela II, os algoritmos GCC-PHAT e GCC-NPHAT, obtiveram resultados superiores.

V. EXPERIMENTOS PRÁTICOS

Para estudar o método proposto com dados reais, coletou-se sinais em uma sala de aula de 6,4m por 6,4m e 3 m de altura. Foram feitas gravações de seis posições diferentes, nas quais foram variadas as posições do arranjo e da caixa de som, a caixa ficando em uma posição mais elevada do que o arranjo. Desta forma, foram testadas diversas DoAs e condições de reverberação. De cada posição, foram gravados dois sinais de 2 minutos cada, um constituído de ruído branco e outro de voz. Durante todas as gravações, o ambiente encontrava-se em silêncio, exceto pelo ar-condicionado mantido ligado durante todo o experimento. Os dados foram coletados em uma taxa de amostragem de 44,1kHz.

Cada gravação foi processada utilizando as técnicas descritas na Seção III. Os sinais foram divididos em segmentos de 5 segundos, totalizando 24 realizações por gravação. Para tornar as gravações mais robustas aos erros provocados pela reverberação, foi utilizado o algoritmo de seleção de pares Busca Exaustiva com $n = 13$ [22].

Os resultados quando o SoI é ruído branco está exposto nas Tabelas III, em relação ao erro angular, e IV, em relação ao azimute. O baixo desvio padrão do erro associado com o erro médio relativamente alto ao obtido nas simulações indica que possivelmente houveram erros de viés associados ao experimento, como, por exemplo, na medição das posições da caixa e do arranjo. Em especial, percebe-se na Tabela IV que o azimute apresentou um erro médio de -3° de forma consistente, sugerindo que a posição dos microfones no arranjo esteja ligeiramente rotacionada em relação ao presumido.

Os resultados das gravações de sinal de voz, Tabela III e VI, evidenciam a perda de desempenho ao se utilizar sinais espectralmente esparsos, que pode ser minimizada com o uso das técnicas GCC- α PHAT e GCC- α NPHAT, como pode ser visto na redução do desvio padrão do erro.

Como os erros angulares, Tabelas III e V, são maiores que os erros no azimute, Tabelas IV e VI, fica evidenciado que existe uma dificuldade maior na estimação do zênite. Possivelmente isso decorre da utilização de um arranjo em que todos os microfones estão no mesmo plano, que como visto na Seção III-C, precisa utilizar o fato de que \mathbf{a}_{DoA} é unitária para estimar o zênite.

VI. CONCLUSÃO

Neste trabalho mostrou-se como utilizar a modificação da técnica GCC-PHAT, a depender das características do SoI, e

TABELA III

 GRAVAÇÕES - ERRO MÉDIO E DESVIO PADRÃO DO ERRO ANGULAR,
 VIDE EQ. (17) - RUÍDO BRANCO

	Sem interpolação		Interpolação cúbica		Interpolação <i>sinc</i>	
	Erro Médio	Desvio Padrão	Erro Médio	Desvio Padrão	Erro Médio	Desvio Padrão
GCC-PHAT	9,4816 °	0,7335 °	5,9510 °	1,0686 °	9,2005 °	0,4961 °
GCC-NPHAT	15,6524 °	1,7678 °	7,8600 °	1,3742 °	10,1612 °	0,3091 °
GCC- α PHAT	15,9863 °	0,3972 °	8,2038 °	0,7931 °	10,4397 °	0,4151 °
GCC- α NPHAT	16,1636 °	0,5086 °	8,8366 °	2,2263 °	10,6344 °	0,4500 °

TABELA IV

 GRAVAÇÕES - ERRO MÉDIO E DESVIO PADRÃO DO AZIMUTE - RUÍDO
 BRANCO

	Sem interpolação		Interpolação cúbica		Interpolação <i>sinc</i>	
	Erro Médio	Desvio Padrão	Erro Médio	Desvio Padrão	Erro Médio	Desvio Padrão
GCC-PHAT	-3,3153 °	0,5265 °	-3,3624 °	0,2693 °	-2,6953 °	0,1575 °
GCC-NPHAT	-2,8303 °	0,4824 °	-3,4227 °	0,4441 °	-2,8097 °	0,1512 °
GCC- α PHAT	-3,5548 °	0,3260 °	-3,3718 °	0,2628 °	-2,9034 °	0,1931 °
GCC- α NPHAT	-3,3498 °	0,3014 °	-3,3498 °	0,3014 °	-2,9272 °	0,2156 °

o uso da interpolação contribuem para refinar a estimação das TDoA, melhorando a estimação da DoA através da minimização da função de custo de mínimos quadrados.

Com base nos resultados obtidos, recomenda-se a utilização do GCC-NPHAT quando o SoI é banda larga, como o ruído branco. E, caso o sinal de interesse seja espectralmente esparsos, a exemplo da voz, o GCC- α PHAT e o GCC- α NPHAT tiveram desempenhos parecidos. A interpolação *sinc* apresentou o melhor resultado em praticamente todos os cenários, sendo a recomendada para atingir a TDE mais precisa, para todos os algoritmos de correlação testados.

Para sinais temporalmente esparsos vale a pena investigar a utilização de detectores, por exemplo para a voz a utilização de um VAD (do inglês, *voice activity detector*).

REFERÊNCIAS

- [1] S. Qin, Y. D. Zhang, and M. G. Amin, "DoA estimation of mixed coherent and uncorrelated targets exploiting coprime MIMO radar," *Digital Signal Processing*, vol. 61, pp. 26 – 34, 2017, special Issue on Coprime Sampling and Arrays. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1051200416300690>
- [2] J. Huang, L. Zhang, Q. Zhang, Y. Jin, and M. Jiang, "Performance analysis of DoA estimation for MIMO sonar based on experiments," in *2009 IEEE/SP 15th Workshop on Statistical Signal Processing*, August 2009, pp. 269–272.
- [3] S.-S. Jeon, Y. Wang, Y. Qian, and T. Itoh, "A novel smart antenna system implementation for broad-band wireless communications," *IEEE Transactions on Antennas and Propagation*, vol. 50, no. 5, pp. 600–606, May 2002.
- [4] N. Dey, A. S. Ashour, F. Shi, and R. S. Sherratt, "Wireless capsule gastrointestinal endoscopy: Direction-of-arrival estimation based localization survey," *IEEE Reviews in Biomedical Engineering*, vol. 10, pp. 2–11, 2017.
- [5] E. C. dos Santos, "Avaliação de desempenho de algoritmos de localização de emissores de RF baseados em medidas de marcação em um cenário naval," *Dissertação de Mestrado, Instituto Militar de Engenharia*, 2017.
- [6] I. L. Freire, J. A. Apolinário Jr., "GCC-based DoA Estimation of Overlapping Muzzleblast and Shockwave Components of Gunshot Signals," *IEEE Second Latin American Symposium on Circuits and Systems (LASCAS)*, 2011.
- [7] H. L. Van Trees, "Detection, estimation, and modulation theory, Part IV, optimum array processing," vol. 1, Jan. 2002.
- [8] J. Smith and J. Abel, "Closed-form least-squares source location estimation from range-difference measurements," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 35, no. 12, pp. 1661–1669, Dec. 1987.

TABELA V

 GRAVAÇÕES - ERRO MÉDIO E DESVIO PADRÃO DO ERRO ANGULAR,
 VIDE EQ. (17) - VOZ

	Sem interpolação		Interpolação cúbica		Interpolação <i>sinc</i>	
	Erro Médio	Desvio Padrão	Erro Médio	Desvio Padrão	Erro Médio	Desvio Padrão
GCC-PHAT	30,6388 °	8,9809 °	28,3814 °	8,9003 °	27,6020 °	9,1282 °
GCC-NPHAT	36,3016 °	7,4952 °	35,9277 °	7,2973 °	35,8087 °	7,3107 °
GCC- α PHAT	36,9098 °	5,3945 °	36,0954 °	6,2232 °	36,3652 °	6,1086 °
GCC- α NPHAT	38,7751 °	7,1560 °	38,3510 °	7,1703 °	38,2166 °	7,0809 °

TABELA VI

GRAVAÇÕES - ERRO MÉDIO E DESVIO PADRÃO DO AZIMUTE - VOZ

	Sem interpolação		Interpolação cúbica		Interpolação <i>sinc</i>	
	Erro Médio	Desvio Padrão	Erro Médio	Desvio Padrão	Erro Médio	Desvio Padrão
GCC-PHAT	-5,2047 °	9,9994 °	-3,6291 °	4,1160 °	-5,6980 °	15,3483 °
GCC-NPHAT	0,4065 °	5,0695 °	0,3453 °	3,1669 °	0,2740 °	3,1472 °
GCC- α PHAT	-2,9749 °	5,0074 °	-0,6044 °	3,2661 °	-0,4573 °	3,2641 °
GCC- α NPHAT	0,5988 °	5,6207 °	0,6382 °	2,9926 °	0,6023 °	3,0059 °

- [9] J. Dmochowski, J. Benesty, and S. Affes, "Direction of arrival estimation using the parameterized spatial correlation matrix," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 4, pp. 1327–1339, May 2007.
- [10] M. Strauss, P. Mordel, V. Miguet, and A. Deleforge, "DREGON: Dataset and Methods for UAV-Embedded Sound Source Localization," Oct. 2018, pp. 1–8.
- [11] "MiniDSP UMA-8 USB MicArray," <https://www.minidsp.com/products/usb-audio-interface/uma-8-microphone-array>.
- [12] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, no. 4, pp. 320–327, Aug. 1976.
- [13] M. Brandstein and H. Silverman, "A robust method for speech signal time-delay estimation in reverberant rooms," vol. 1, May 1997, pp. 375 – 378 vol.1.
- [14] B. Van Den Broeck, A. Bertrand, P. Karsmakers, B. Vanrumste, H. Van hamme, and M. Moonen, "Time-domain generalized cross correlation phase transform sound source localization for small microphone arrays," in *EDERC 2012 - Proceedings of the 5th European DSP in Education and Research Conference*, 2012.
- [15] Bo Qin, Heng Zhang, Qiang Fu, and Yonghong Yan, "Subsample time delay estimation via improved gcc phat algorithm," in *2008 9th International Conference on Signal Processing*, Oct. 2008, pp. 2579–2582.
- [16] I. L. Freire, J. A. Apolinário Jr., "DoA of gunshot signals in a spatial microphone array: performance of the interpolated generalized cross-correlation method," in *6th Argentine School of Micro-Nanoelectronics, Technology and Applications (EAMTA)*, 2011.
- [17] A. J. Jerri, "The Shannon sampling theorem—its various extensions and applications: A tutorial review," *Proceedings of the IEEE*, vol. 65, no. 11, pp. 1565–1596, Nov. 1977.
- [18] F. Gensun, "Whittaker–Kotelnikov–Shannon sampling theorem and aliasing error," *Journal of Approximation Theory*, vol. 85, no. 2, pp. 115 – 131, 1996. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0021904596900337>
- [19] I. L. Freire, J. A. Apolinário Jr., "Localização de atirador por arranjo de microfones," in *Simpósio Brasileiro de Automação Inteligente*, 2011.
- [20] M. S. Brandstein, J. E. Adcock, and H. F. Silverman, "A localization-error-based method for microphone-array design," in *1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings*, vol. 2, May 1996, pp. 901–904 vol. 2.
- [21] I. L. Freire, "Robust direction-of-arrival by matched-lags, applied to gunshots," *The Journal of the Acoustical Society of America*, vol. 135, no. 6, pp. EL246–EL251, 2014.
- [22] A. M. C. R. Borzino, J. A. Apolinário Jr., M. L. R. de Campos, "Robust DoA estimation of heavily noisy gunshot signals," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Brisbane: IEEE, April 2015, pp. 449–453.
- [23] A. Wabnitz, N. Epain, C. Jin, and A. van Schaik, "Room acoustics simulation for multichannel microphone arrays," in *International Symposium on Room Acoustics*, Jan. 2010.