

Desempenho do algoritmo beamformer MVDR através da variação de distância entre dois microfones

Roberto Cyrulnik e Ivandro Sanches

Resumo—O presente trabalho apresenta um estudo sobre as influências da variação da distância entre dois microfones em um algoritmo de filtragem espacial chamado *beamformer* MVDR (*minimum variance distortionless response*). Para obtenção dos resultados experimentais foram utilizados dois métodos: a ferramenta de simulação de arranjo de microfones e antenas do MATLAB, o *Phased Array* e através de dados obtidos na câmara varecólica do Bell Labs. Neste último, foram realizadas medições em um arranjo linear de 22 microfones, variando-se a posição de uma fonte sonora. Os detalhes de implementação do algoritmo e ensaios são demonstrados neste artigo.

Palavras-Chave—*beamformer*, arranjo de microfones, MVDR.

Abstract—This work presents a study about the influence of the distance variation between two microphones in a spacial filtering algorithm called MVDR (*minimum variance distortionless response*) beamformer. Experimental results were obtained in two procedures: via the MATLAB Phased Array toolbox that simulates microphone and antenna arrays and via Bell Labs varechoic chamber database. The latter contains measurements made on a 22 linear microphone array for many sound source localizations. Details of the algorithm implementation and experimental procedures were shown on this article.

Keywords—arranjo de microfones, *microphone array*, *beamforming*, MVDR

I. INTRODUÇÃO

Comunicação em ambientes ruidosos é um grande desafio para trabalhadores, militares e policiais pois a inteligibilidade fica comprometida. Os dispositivos de comunicação como rádios comunicadores ou telefones celulares utilizados nessas áreas captam a voz do locutor adicionada ao ruído, prejudicando a comunicação com o receptor.

São exemplos de ambientes extremamente ruidosos: pistas de aeroportos, estádios de futebol, shows, indústrias siderúrgicas e indústrias de mineração, ambientes onde o ruído pode alcançar níveis de pressão sonora muito elevados. Um policial durante sua guarda em um estádio de futebol recebe ordens enquanto a torcida gera ruídos extremos. Um operário em uma indústria siderúrgica deve comunicar-se com seu colega em um ambiente extremamente ruidoso. Um soldado deve transmitir informações sobre o inimigo dentro do campo de batalha imerso ao som de metralhadoras, explosões, etc. Em todos esses ambientes o dispositivo de comunicação, que pode ser um rádio comunicador ou um aparelho celular, deve

possuir um meio de suprimir o ruído ambiente para que o receptor consiga receber a informação com qualidade.

Existem algumas técnicas avançadas de supressão de ruído como CASA (Computational Auditory Scene Analysis) [13], BSS (Blind Source Separation) [13] e beamforming.

CASA utiliza o conceito sensorial da audição humana para separar diferentes fontes de sinais. Esse conceito utiliza um processo de dois estágios: a segmentação e o agrupamento. Durante a segmentação o sinal de entrada é decomposto em uma coleção de regiões tempo-frequência chamados segmentos. No segundo estágio, o agrupamento, os segmentos são agrupados de acordo com regras perceptuais para separar as diferentes fontes que geraram o sinal de entrada. Essa é uma técnica muito recente e ainda necessita de mais estudos para que possa ser implementada em dispositivos móveis.

Alguns algoritmos de BSS utilizam a técnica ICA (*independent component analysis*). Porém, esses algoritmos tendem a perder performance drasticamente em ambientes que possuem mais fontes sonoras do que microfones [13].

Beamforming é um método de filtragem espacial que diferencia os sinais desejados do ruído e interferência através de um arranjo de microfones. Essa técnica já está sendo implementada em DSP's (*Digital Signal Processor*) [6] e FPGA [12].

São alguns exemplos de *beamformers*: MVDR no domínio da frequência ou FMV (*frequency minimum variation*) [9], o algoritmo de Frost no domínio do tempo [7] e GSC (*generalized sidelobe canceler*) [8]. Testes com esses *beamformers* foram realizados utilizando dois microfones.

Como resultado dos testes entre os *beamformers*, o MVDR no domínio da frequência apresentou a melhor relação sinal ruído e a menor distorção do sinal desejado na saída do sistema [9].

Portanto, para esse trabalho, será utilizado o *beamformer* MVDR.

O objetivo desse artigo é apresentar a influência da distância entre dois microfones em um *beamformer* MVDR. Alguns artigos apresentam aplicações de *beamformer*, porém não apresentam um estudo sobre a influência da distância entre os microfones para a performance do sistema.

Para comprovar essa influência foram utilizados dois métodos experimentais: o *toolbox Phased Array System* do MATLAB [1] e dados gerados através de um arranjo de 22 microfones na câmara varecólica do laboratório Bell Labs [2]. O *toolbox Phased Array System* pode simular um arranjo de microfones de vários formatos e com diferentes números

de elementos além de fontes sonoras em diferentes azimutes e elevações. No caso da câmara varecólica, uma simulação bem mais próxima da realidade pode ser efetuada. Nesse ambiente foram instalados 22 microfones de forma linear e a posição de uma fonte sonora foi variada, gerando funções de transferência entre os microfones e a fonte sonora. Os experimentos utilizaram arranjos de dois microfones.

Esse artigo é estruturado da seguinte maneira: na Seção II será apresentada a implementação do algoritmo, e, a seguir, na Seção III as técnicas de ensaio. Por fim, na Seção IV serão apresentados os resultados desse trabalho e, logo após, as conclusões.

II. IMPLEMENTAÇÃO DO ALGORITMO

O sinal no domínio do tempo é transformado periodicamente (a cada $L = 16$ amostras) no domínio da frequência via DFT de comprimento N , utilizando a janela de Hamming. Em um sistema com dois microfones, os sinais no domínio da frequência são representados por componentes do vetor $\mathbf{X}_k = [X_{1k} \ X_{2k}]$ onde k é o índice da frequência. As F mais recentes DFT's são armazenadas em um *buffer* e uma matriz de correlação \mathbf{R}_k é calculada para cada índice de frequência k através de:

$$\mathbf{R}_k = \begin{bmatrix} \frac{M}{F} \sum_{i=1}^F X_{1k,i}^* X_{1k,i} & \frac{1}{F} \sum_{i=1}^F X_{1k,i}^H X_{2k,i} \\ \frac{1}{F} \sum_{i=1}^F X_{2k,i}^* X_{1k,i} & \frac{M}{F} \sum_{i=1}^F X_{2k,i}^H X_{2k,i} \end{bmatrix}, \quad (1)$$

onde H representa o conjugado transposto e M é um fator multiplicativo regulador, um pouco maior que 1.00 que ajuda a evitar a singularidade da matriz e melhora a robustez a descasamento de microfones. O valor de N utilizado foi de 1024, $M = 1.03$, 1.03 e $F = 32$. As matrizes de correlação foram atualizadas a cada $L = 16$ amostras, permitindo que elas fossem atualizadas rapidamente com variação do sinal de entrada. Para cada frequência k , a saída do *beamformer* é:

$$\mathbf{Y}_k = \mathbf{w}_k^H \mathbf{X}_k, \quad (2)$$

onde \mathbf{w}_k é um vetor de pesos no domínio da frequência. A otimização do algoritmo e restrição são expressas, para cada frequência, como:

$$\mathbf{w}_{opt,k} = \arg \min_{\mathbf{w}_k} E \{ |\mathbf{Y}_k|^2 \}, \quad (3)$$

restrito a

$$\mathbf{d}^H \mathbf{w}_k = 1, \quad (4)$$

onde $\arg \min$ representa a minimização da função em relação ao peso \mathbf{w}_k e E representa a operação esperança. O vetor \mathbf{d} indica a direção do sinal desejado. Para uma fonte sonora localizada no azimute 0° ambos os microfones devem receber os sinais ao mesmo tempo, com a mesma amplitude. Para isso, deve-se utilizar $\mathbf{d}^H = [1 \ 1]$. Como resultado da minimização representada pelas Equações 3 e 4, obtem-se o vetor peso ótimo dado por:

$$\mathbf{w}_{opt,k} = \frac{\mathbf{R}_k^{-1} \mathbf{d}}{\mathbf{d}^H \mathbf{R}_k^{-1} \mathbf{d} + \sigma}, \quad (5)$$

onde \mathbf{R}_k^{-1} representa a matriz inversa de \mathbf{R}_k e σ é um valor muito pequeno que previne a divisão por zero. Para obter

a saída no domínio da frequência, os vetores peso $\mathbf{w}_{opt,k}$ para cada frequência k são aplicados ao *buffer* de DFT's (Equação 2). Os N valores no domínio da frequência são então transformados no domínio do tempo através de uma anti-transformada de comprimento N (IDFT). Isso ocorre a cada L amostras. Somente as L amostras centrais no domínio do tempo são utilizadas. Na Figura 1 é mostrado um diagrama de blocos do algoritmo *beamformer* MVDR.

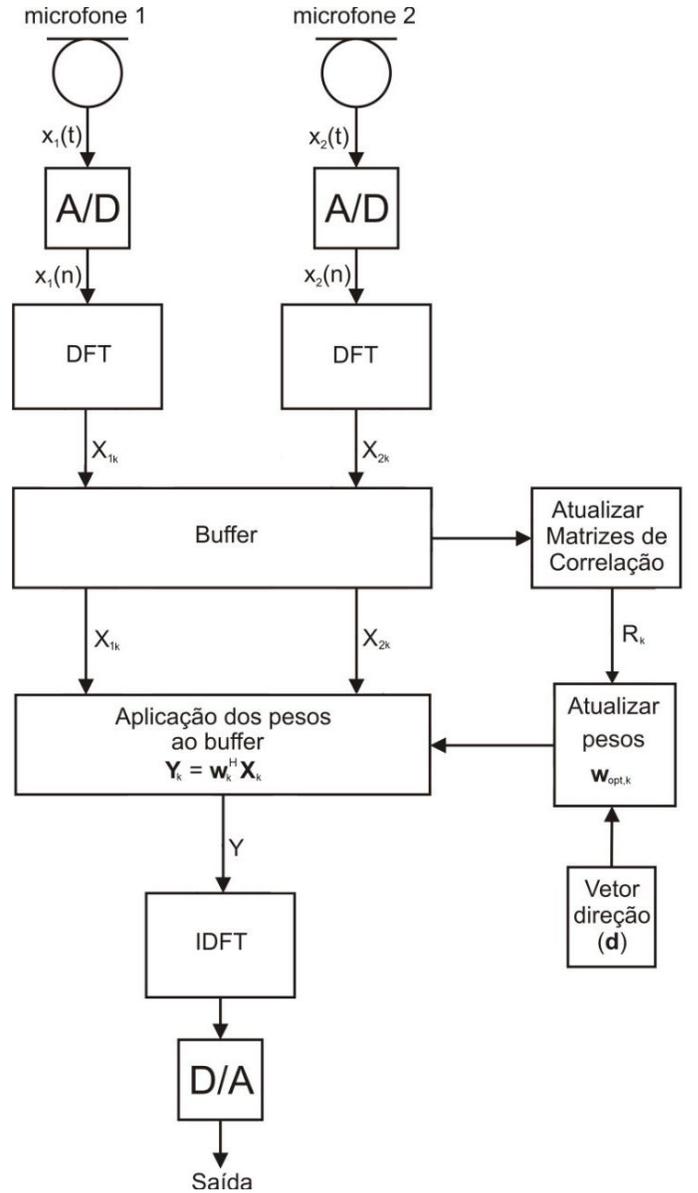


Fig. 1: Diagrama de blocos do algoritmo *beamformer* MVDR

III. MÉTODOS EXPERIMENTAIS

Os dados de entrada utilizados para obter os resultados experimentais foram gravados por um dispositivo M-Audio MicroTrack, a uma frequência de amostragem de 48kHz, 16 bits. Para obter o sinal de saída dos microfones através da variação de sua distância e a variação do azimute da fonte sonora, foram utilizados dois meios:

- MATLAB, utilizando a *toolbox Phased Array System* [1];

- respostas ao impulso obtidas no Bell Labs através de um arranjo de 22 microfones e fontes sonoras em várias posições [2].

Nas duas subseções seguintes serão detalhados esses dois meios.

A. Dados gerados pelo MATLAB

O MATLAB possui um *toolbox* chamado *Phased Array System* [1] que pode simular os sinais obtidos nas saídas de um arranjo de microfones através de sinais gerados por fontes sonoras em diferentes azimutes e elevações. Para obter as saídas dos microfones devem ser utilizados os seguintes parâmetros de entrada:

- resposta em magnitude do microfone que pode ser omnidirecional ou customizada;
- posicionamento do arranjo de microfones que pode ser ULA (*uniform linear array*), URA (*uniform rectangular array*) ou *conformal* onde pode-se customizar o posicionamento do arranjo de microfones;
- número de microfones do arranjo;
- distância entre os microfones do arranjo;
- velocidade do som em metros por segundo;
- vetores contendo os arquivos de áudio que irão simular as fontes sonoras;
- vetores contendo as informações de azimute e elevação de cada fonte sonora;
- frequência de amostragem.

Um exemplo para o *Phased Array* é ilustrado nas Figuras 2 e 3. Na Figura 2 são mostrados os parâmetros de entrada no *Phased Array* para um arranjo linear de dois microfones, composto por 3 fontes sonoras representadas pelos vetores contendo **X1**, **X2** e **X3**, posicionadas nos respectivos azimutes θ_1 , θ_2 e θ_3 . Na Fig 3 é ilustrado o comportamento físico dos parâmetros de entrada da Fig 2.

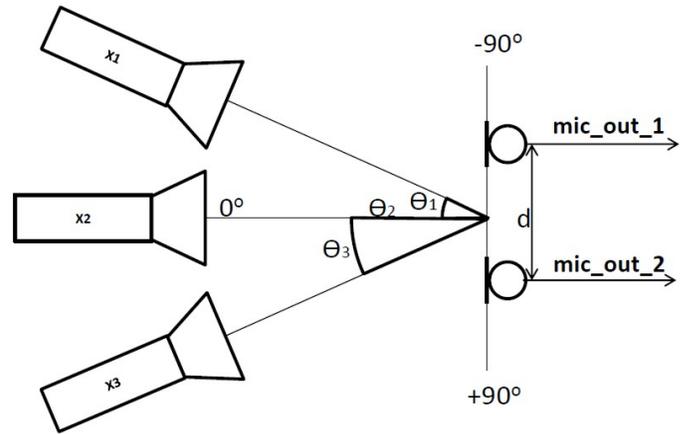


Fig. 3: Modelo físico do *toolbox Phased Array* segundo os parâmetros indicados na Figura 2.

o som. Em função da colocação relativa das lâminas pode-se controlar as propriedades acústicas da sala (reverberação). Nessa câmara realizou-se uma série de medições através de um arranjo linear de microfones [2]. Um arranjo de 22 microfones é posicionado na parede norte de câmara varecólica, em frente a uma placa de fibra de vidro. Uma caixa de som foi posicionada em 31 locais distintos para obter a resposta ao impulso na saída dos 22 microfones. A câmara varecólica foi configurada para uma abertura de 89%, que corresponde a um tempo T60 de reverberação de 0.28 segundos. As respostas ao impulso foram adquiridas a uma frequência de amostragem de 48 kHz e foram gravadas em arquivos WAV de 22 canais. Assim, foram fornecidos 31 arquivos, com 22 canais cada um. As posições dos microfones e da fonte sonora podem ser observadas na Figura 4.

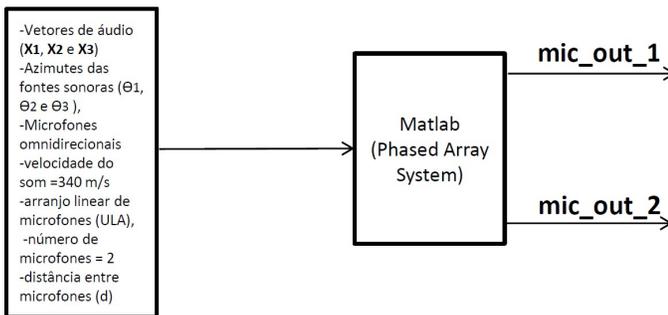


Fig. 2: Dados de entrada para o *toolbox Phased Array* do MATLAB e vetores de saída **mic_out_1** e **mic_out_2** para a configuração de 2 microfones.

B. Dados gerados pela câmara varecólica

A câmara varecólica localizada no Bell Labs [3] consiste em uma sala onde o piso, as paredes e o teto estão recobertos por um conjunto de painéis controlados digitalmente. Os painéis são compostos por lâminas de aço perfuradas e móveis, contendo em sua parte traseira fibra de vidro para absorver

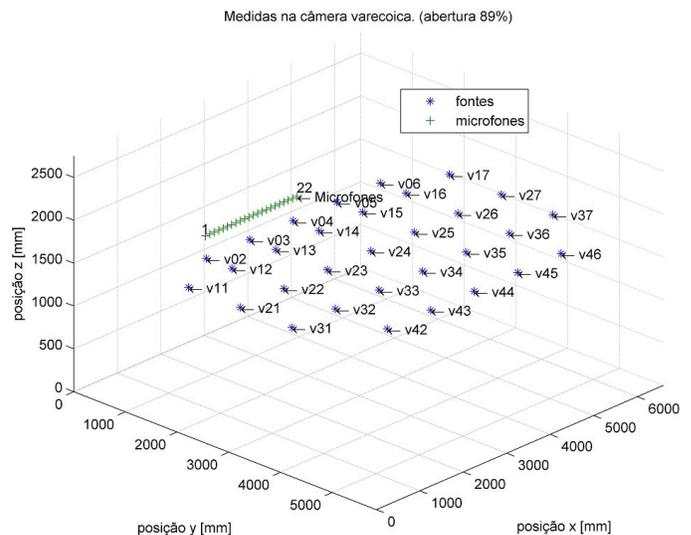


Fig. 4: Posições da fonte sonora e do arranjo de microfones na câmara varecólica durante o experimento de Aki Härrmä [2].

Para obtenção dos resultados desse artigo, foram utilizados os pares de microfones 9 - 11 e 8 - 12, correspondendo as respectivas distâncias entre microfones de 200 mm e 400 mm. As posições da fonte sonora utilizadas foram: v17, v26, v35,

v34, v33, v22 e v11, correspondendo respectivamente aos azimutes: -64.39° , -39.36° , -16.22° , 0° , $+16.22^\circ$, $+39.36^\circ$ e $+64.39^\circ$.

Na Figura 4 pode-se verificar as dimensões da sala assim como a disposição dos microfones e as posições que o altofalante assumiu para obter as repostas ao impulso nas saídas dos microfones.

IV. RESULTADOS

A. Resultados dos dados gerados pelo MATLAB

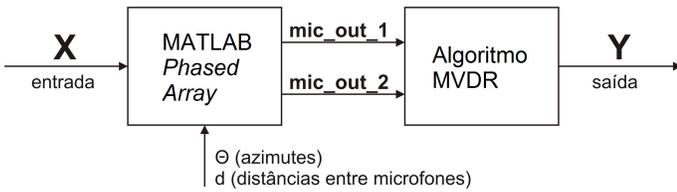


Fig. 5: Setup para obter os resultados da variação de distância entre os microfones através do *toolbox Phased Array* do MATLAB.

Para esse experimento, foram utilizados dois microfones omnidirecionais, em um arranjo uniforme linear (ULA). O vetor (\mathbf{X}) contendo o arquivo de áudio foi obtido na frequência de amostragem de 48 kHz, 16 bits, utilizando o dispositivo de gravação M-Audio Microtrack II onde o narrador é do sexo masculino com idade de 35 anos, com a emissão de um sinal acústico totalmente sonoro constituído pela pronúncia de 5 vogais. Esse mesmo vetor \mathbf{X} foi utilizado como fonte sonora para todos os azimutes. Após passar pelo *Phased Array*, são obtidos dois vetores: o **mic_out_1** e o **mic_out_2**. Esses vetores são submetidos ao algoritmo MVDR e a saída \mathbf{Y} é obtida. O vetor de entrada \mathbf{X} é então correlacionado com o vetor de saída \mathbf{Y} como mostra a Equação 6:

$$\mathbf{R}_{XY} = R_{XY}(m) = E[X(n+m) * Y^H(n)], \quad (6)$$

onde $R_{XY}(m)$ é a correlação cruzada entre os vetores \mathbf{X} e \mathbf{Y} para um determinado azimute θ , E é a esperança, $*$ é a convolução entre os dois vetores e $Y^H(n)$ é o conjugado transposto de $Y(n)$. Considerando que o maior vetor possui comprimento M , m pode variar entre $-M+1$ e $M-1$. A maior energia para um determinado azimute θ é calculada através do maior valor de \mathbf{R}_{XY} , isto é, $\max(\mathbf{R}_{XY})$. As energias são normalizadas para cada uma das distâncias entre microfones (d), obtendo-se o gráfico da Figura 6. As distâncias entre microfones consideradas foram 20 mm, 50 mm e 100 mm. O gráfico da Figura 6 mostra que quanto maior for a distância entre os microfones, mais seletivo é o algoritmo.

B. Resultados dos dados gerados pela câmara varecólica

Para cada uma das posições da fonte sonora v17, v26, v35, v34, v33, v22 e v11 escolhidas na Seção III-B e para cada par de microfones (9, 11) e (8, 12) o mesmo sinal de entrada, \mathbf{X} , utilizado na Seção IV-A foi utilizado como entrada para essa configuração. O sinal de entrada \mathbf{X} foi convoluído com as repostas ao impulso relativas a posição da fonte sonora

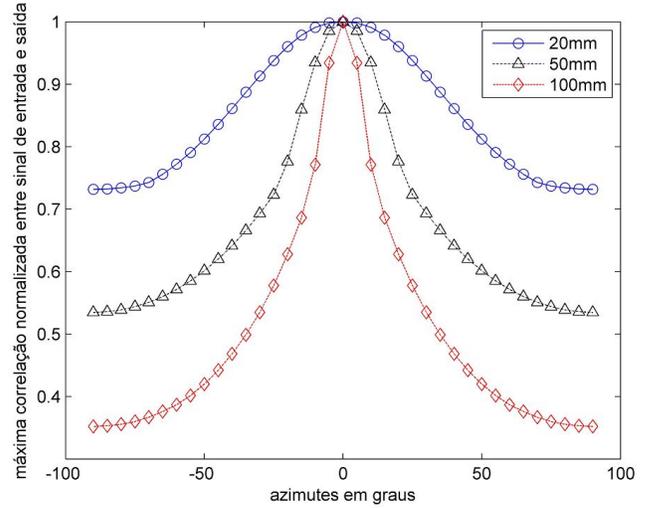


Fig. 6: Máxima correlação normalizada (energia) entre o sinal de entrada \mathbf{X} e o sinal de saída \mathbf{Y} .

e os microfones escolhidos. Os sinais obtidos (**mic_out_1** e **mic_out_2**) foram normalizados em energia como mostra a Equação 7:

$$\mathbf{mic_out_1}' = \frac{\mathbf{mic_out_1}}{\sqrt{\mathbf{mic_out_1}^H \mathbf{mic_out_1}}}, \quad (7)$$

onde o símbolo H representa o transposto conjugado do vetor **mic_out_1**. O mesmo procedimento é realizado para o vetor **mic_out_2**. Essa normalização, sem alterar o alvo da investigação, visa reduzir efeitos indesejados causados por eventuais diferenças nas repostas de magnitude entre os 4 microfones utilizados e nas várias distâncias resultantes entre fonte sonora e microfones. O vetor é submetido então ao algoritmo MVDR e o sinal de saída \mathbf{Y} é obtido.

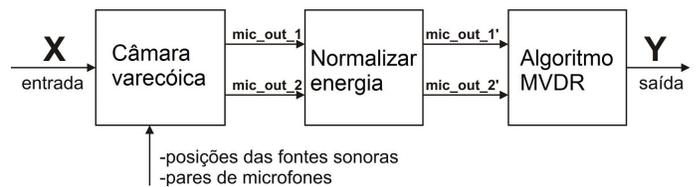


Fig. 7: Setup para obter os resultados da variação de distância entre os microfones através da câmara varecólica.

A correlação entre o sinal de entrada \mathbf{X} e o sinal de saída \mathbf{Y} é realizada através da Equação 6. Assim como na seção anterior, são utilizados os máximos valores de \mathbf{R}_{XY} e as energias são normalizadas para cada par de microfones selecionados.

O resultado pode ser observado na Figura 8.

Assim como na Seção IV-A, distanciando-se os microfones, o algoritmo torna-se mais seletivo.

Como sugerido por um dos revisores do artigo, compararam-se os resultados simulados pelo phased array nas mesmas condições experimentais da Figura 8. Essa comparação é apresentada na Figura 9. O caso simulado com phased array ficou muito semelhante ao experimental. Esse fato surpreende pois condições reais introduzem efeitos

que produzem resultados bem diferentes da situação ideal, como por exemplo: descasamento entre microfones (tanto em amplitude como em fase), reverberação do ambiente, o fato da frente de onda não poder ser considerada plana em condições em que a fonte sonora está muito próxima dos microfones.

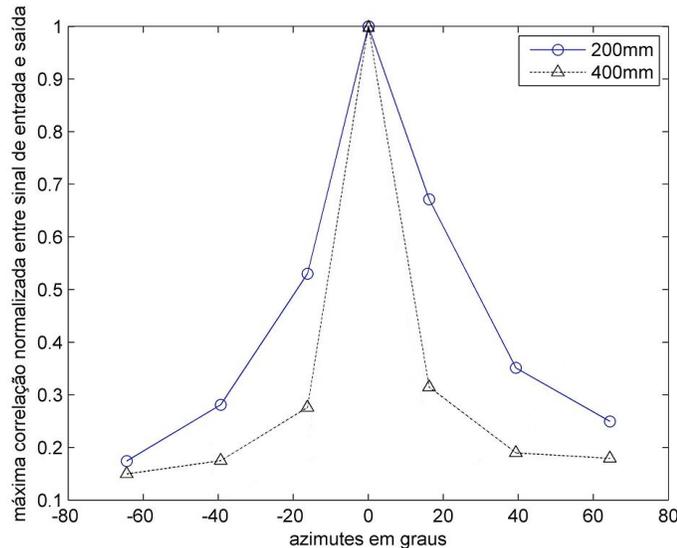


Fig. 8: Máxima correlação normalizada (energia) entre o sinal de entrada X e o sinal de saída Y para a câmara varecólica.

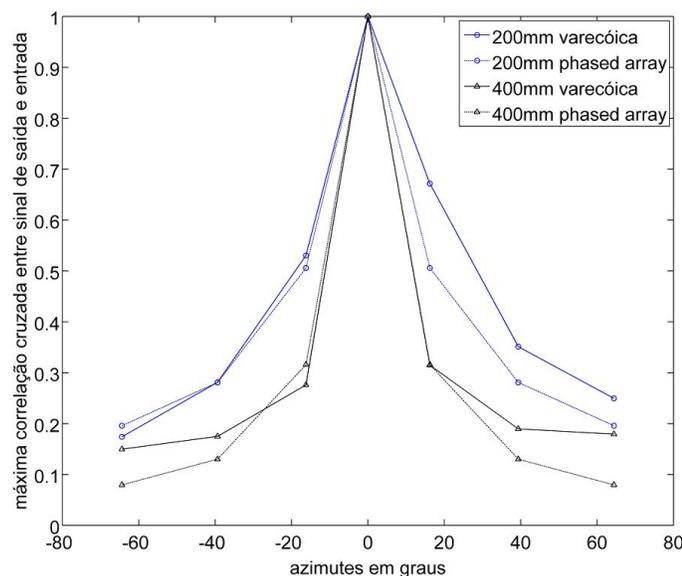


Fig. 9: Comparação entre os resultados simulados pelo phased array com os resultados experimentais obtidos na Figura 8.

V. CONCLUSÕES

O *beamformer* MVDR no domínio da frequência ou FMV (*frequency minimum variation*) é um algoritmo que possui boa relação sinal ruído e menor distorção do sinal quando comparado a outros *beamformers* [9]. Existem muitas aplicações para o *beamformer*, dentre elas, os dispositivos portáteis como

notebooks, *tablets* [14], sistemas de viva-voz para veículos [12] e sistemas de comunicação ao trabalhador [15]. Este artigo apresentou um estudo sobre a influência da distância entre dois microfones na performance de um *beamformer* MVDR. Para comprovar essa influência, foram utilizados dois métodos experimentais. O primeiro com a utilização do *toolbox Phased Array System* do MATLAB [1] que simula arranjo de microfones. O segundo com dados gerados através de um arranjo de 22 microfones na câmara varecólica do laboratório Bell Labs [2]. Os experimentos contemplaram arranjos de dois microfones. Ao se analisar os resultados desses experimentos, nota-se que tanto na Figura 6 como na Figura 8 nota-se que a seletividade do conjunto aumenta com o aumento da distância entre os microfones. Esse resultado confirmou a expectativa inicial, pois sinais vindos de direções mais afastadas da direção de azimute zero chegarão mais descorrelacionados entre os microfones.

REFERÊNCIAS

- [1] The MathWorks Inc., *Phased Array Toolbox*, Natick, MA. Acesso em 30/01/2012. Disponível em: <http://www.mathworks.com/products/phased-array/>
- [2] A. Härmä, *Microphone array measurements from the varechoic chamber*. Bell Laboratories, Murray Hill, NJ, EUA, outubro 2001. Acesso em 15/02/2012. Disponível em: http://www.acoustics.hut.fi/aqi/vardata/Varechoic_array_data.html
- [3] Bell Labs, *Varechoic Chamber*. Acesso em 15/02/2012. Disponível em: <http://www.bell-labs.com/org/1133/Research/Acoustics/VarechoicChamber.html>
- [4] R. Adve, *Optimal beamforming*. University of Toronto, 2011. Acesso em 21/08/2011. Disponível em: <http://www.comm.toronto.edu/~rsadve/Notes/BeamForming.pdf>
- [5] M. D. Burkhard; R. M. Sachs, *Anthropometric manikin for acoustic research*. Illinois, EUA, 1978. Acesso em 21/08/2011. Disponível em: <http://www.gras.dk/documents/00316.pdf>
- [6] M. E. Elledge et al., "Real-time implementation of a frequency-domain beamformer on the TI C62X EVM." *10th Annual DSP Technology Education and Research Conference*, Texas Instrument, Houston, TX, 2000.
- [7] O. L. Frost, "An algorithm for linearly constrained adaptive array processing", *Proceeding of the IEEE*, IEEE, v. 60, n. 8, p. 926–935, 1972.
- [8] J. E. Greenberg, "Modified LMS algorithms for speech processing with an adaptive noise canceller", *IEEE Transactions on speech and audio processing*, IEEE, v. 6, p. 338–351, 1998.
- [9] M. E. Lockwood et al., "Performance of time- and frequency-domain binaural beamformers based on recorded signals from real rooms", *J. Acoustic. Soc. Am.*, Acoustical Society of America, v. 115, p. 379–391, 2004.
- [10] J. Bitzer, K.U. Simmer, "Superdirective microphone arrays", in *Microphone Arrays*, M. S. Brandstein and D. B. Ward, Eds., chapter 2, pp. 19–38. Springer, 2001.
- [11] B. D. V. Veen, K. M. Buckley, "Beamforming: A versatile approach to spatial filtering", *IEEE ASSP Magazine*, IEEE, p. 338–351, abr 1998.
- [12] N. V. Vu et al., "Small footprint implementation of dual-microphone delay-and-sum beamforming for in-car speech enhancement," *ICASSP 2010*, pp. 1482–1485, 2010.
- [13] D. Wang; G. J. Brown, *Computational Auditory Scene Analysis. Principles, Algorithms and Applications*. Hoboken, New Jersey: Wiley-Interscience, pp. 1–37 and 187–205, 2005
- [14] MICROSOFT. *Microphone Array Support in Windows Vista*. Acesso em 02/02/2012. Disponível em: <http://download.microsoft.com/download/9/c/5/9c5b2167-8017-4bae-9fde-d599bac8184a/MicArrays.doc>
- [15] SENSEAR, P. *Protetores auditivos com comunicação Sensear*. 2011. Acesso em 31/08/2011. Disponível em: <http://www.sensear.com>