

Análise de Desempenho de Técnicas de Estimação de Canal na Desreverberação de Sinais de Fala

Gleyson A. da Silva, José A. Apolinário Jr. e Dirceu G. da Silva

Resumo—Este artigo apresenta uma comparação entre dois métodos diferentes de estimação de canal num contexto de desreverberação de sinais de fala captados por um único microfone: a estimação por média cepstral e a estimação baseada na identificação dos zeros associados ao sistema de degradação. Além disso, pretende-se determinar as condições de uso apropriadas para cada método. A motivação para esta análise de desempenho é a possibilidade de aplicação destas técnicas em um grande número de situações práticas que têm alguma dependência da melhoria de sinais de fala, incluindo reconhecimento automático de voz e, em um contexto mais geral, as telecomunicações.

Palavras-Chave—Desreverberação de fala, estimação de canal, melhoria de sinais de fala.

Abstract—This paper presents a comparison between two different channel estimation methods for single-microphone speech dereverberation: cepstral mean estimation and estimation based on the identification of the zeros associated with the degradation system. Besides this, we aim to establish the appropriate conditions of use for each method. The motivation for this performance analysis is the possibility of applications of these techniques in a number of practical situations which have some dependence on speech enhancement, including automatic speech recognition and, in a broader context, telecommunications.

Keywords—Speech dereverberation, channel estimation, speech enhancement.

I. INTRODUÇÃO

Os mais recentes avanços na área de telecomunicações, como teleconferência e aplicações multimídia, têm fomentado a necessidade de solução para diversos problemas que atingem o processo de comunicação da voz. Dentre estes, pode-se destacar a reverberação do sinal de fala, que é o efeito indesejável produzido pelas reflexões desse sinal no ambiente.

Numa situação em que o locutor se encontra distante do microfone, como, por exemplo, em aparelhos de viva-voz, o sinal adquirido pelo microfone difere do

original por ser afetado pelas distorções introduzidas no percurso entre o emissor sonoro e o receptor. Estas alterações podem, de um modo geral, ser elencadas em duas categorias principais: ruído aditivo e ruído convolucional [1], esta segunda possuindo íntima ligação com as propriedades acústicas do ambiente.

Uma vez que há uma relação direta entre a resposta ao impulso do ambiente e a reverberação produzida, geralmente a solução deste problema passa pela estimação da resposta inversa do meio, que é a base de ambos os métodos aqui expostos. Conseqüentemente, uma técnica capaz de realizar a normalização deste canal estará sendo responsável pela desreverberação do sinal.

Na literatura, as técnicas de desreverberação mais conhecidas podem ser agrupadas em duas categorias: métodos que consideram um único microfone [2], [3] e métodos que consideram um arranjo de microfones [4], [5]. Em particular, os aspectos teóricos de desreverberação de salas baseados na filtragem pela resposta inversa foram introduzidos há mais de duas décadas atrás para ambientes com resposta de fase mínima [6] e de fase mista [7]. Neste trabalho, para ambos os métodos comparados, será considerado um único microfone e resposta do ambiente de fase mista.

Neste trabalho, a primeira técnica de estimação de canal para fins de desreverberação é uma aplicação da normalização de canal fundamentada na teoria de filtragem homomórfica desenvolvida em [2]. Para sinais convoluídos num domínio no qual eles possam ser somados, uma separação por subtração é possível. O logaritmo do espectro, ou *cepstrum*, por exemplo, satisfaz essa condição. A partir deste procedimento, a deconvolução é perfeitamente possível quando se conhece algum dos sinais. Entretanto, quando não há informação alguma a respeito dos mesmos, pode-se utilizar os métodos de deconvolução cega para obter a separação.

O segundo método, proposto em [8], caracteriza-se por, além de ser uma técnica capaz de estimar respostas acústicas de fase mista considerando um único microfone, ser capaz de obter uma estimativa com alto grau de similaridade com a resposta original.

A finalidade deste artigo é, portanto, a análise comparativa entre os métodos citados à luz dos parâmetros

Gleyson A. da Silva, José A. Apolinário Jr. e Dirceu G. da Silva: Departamento de Engenharia Elétrica, IME, Rio de Janeiro, Brasil, E-mails: gleyson131@gmail.com, apolin@ime.eb.br e dirceu@ime.eb.br.

eficácia e abrangência da técnica. A eficácia deve ser entendida, neste contexto, como a precisão na estimativa da resposta do canal. Já a abrangência da técnica será testada mediante a manutenção da validade do método sobre condições mais adversas como, por exemplo, a inserção de ruído aditivo.

Este artigo está dividido como segue: uma descrição dos métodos de estimação de resposta ao impulso utilizados é apresentada na Seção II; os resultados experimentais, mostrando o desempenho de cada método, são exibidos na Seção III e, por fim, as conclusões obtidas são apresentadas na Seção IV.

II. MÉTODOS DE ESTIMAÇÃO DE RESPOSTA DO AMBIENTE

O processo de desreverberação, que pode ser enquadrado como um procedimento de normalização de canal, possui basicamente três etapas que devem ser atingidas na seguinte seqüência: estimação da resposta ao ambiente, obtenção da resposta inversa e, por último, filtragem com a resposta inversa. Neste trabalho, serão abordadas apenas técnicas para a realização da primeira etapa, haja vista que a solução para as demais é consagrada.

A. Estimação por Média Cepstral

O trabalho pioneiro em deconvolução cega foi apresentado em [9], inspirando algumas das mais comumente utilizadas técnicas de normalização de canal. Naquele trabalho, o objetivo era restaurar uma gravação musical removendo as distorções convolucionais introduzidas por um aparelho de gravação antigo (ano de 1907). Deste modo, nenhum conhecimento explícito existia do sinal ou do aparelho de gravação. Para obter algum conhecimento do sinal original, foi realizada uma gravação mais recente da música que eles desejavam restaurar (*Vesti la Guibba* por Enrico Caruso). A gravação recente foi realizada com equipamentos modernos e foi cantada por um tenor com voz de características similares à de Enrico Caruso. O espectro médio desta nova versão foi utilizado para estimar as características originais da música de Caruso. Dividindo o espectro médio da antiga gravação por esta nova versão, eles obtiveram uma estimativa da função de transferência do equipamento original. Um dos pressupostos do método era que a resposta em frequência do equipamento utilizado recentemente fosse plana. Após obter a estimativa da resposta do gravador, foi projetado um filtro inverso para compensar as suas distorções, reconstituindo-se, assim, a gravação original.

A proposta do método de estimação por média cepstral é seguir passos similares aos adotados em [9] e [10].

Deste modo, seja $y(t)$ o resultado da convolução entre o sinal de voz $s(t)$ e a resposta ao impulso do canal $h(t)$. Então:

$$y(t) = s(t) * h(t). \quad (1)$$

Como no caso da recuperação da gravação antiga, onde não era conhecida a resposta ao impulso do aparelho de gravação, aqui se desconhece a resposta ao impulso do ambiente que irá contaminar o sinal original com ruído convolucional, ou seja, o objetivo é, dado $y(t)$, estimar um sinal tão próximo quanto possível do original.

Após a digitalização de $y(t)$ e a aplicação da DFT para cada quadro de voz, a operação acima transforma-se numa multiplicação no domínio da frequência:

$$|Y_{k,i}| \approx |S_{k,i}| |H_k|, \quad (2)$$

onde k é o índice da DFT e i é o índice do quadro. A aproximação decorre do fato de que

$$\begin{aligned} y_i(t) &= w_i(t) \cdot y(t) = w_i(t) \cdot [s(t) * h(t)] \\ &\approx [w_i(t) \cdot s(t)] * h(t) = s_i(t) * h(t), \end{aligned} \quad (3)$$

onde $w_i(t)$ é o janelamento aplicado em $y(t)$ para obter-se $y_i(t)$. A validade dessa aproximação depende de que o tamanho da janela seja maior que a resposta ao impulso do canal [9].

Para simplificar a notação, serão utilizados apenas \mathbf{Y}_i , \mathbf{S}_i e \mathbf{H} para indicar os vetores de pontos da DFT para cada quadro.

Para a recuperação de um dos sinais, aplica-se o logaritmo a ambos os lados da equação, fazendo com que a multiplicação transforme-se em uma soma.

$$\log |\mathbf{Y}_i| \approx \log |\mathbf{S}_i| + \log |\mathbf{H}|. \quad (4)$$

Dessa forma, é possível recuperar um dos dois sinais, s ou h , conhecendo-se um deles. Aplicando-se a transformada inversa em (4), passamos para o domínio denominado *quefrequency* [11], ou

$$\hat{\mathbf{y}}_i = \hat{\mathbf{s}}_i + \hat{\mathbf{h}}, \quad (5)$$

onde $\hat{\mathbf{y}}_i$, $\hat{\mathbf{s}}_i$ e $\hat{\mathbf{h}}$ são os vetores de coeficientes *cepstrum* para cada quadro do sinal de voz distorcido do sinal limpo e do canal, respectivamente.

Para que (5) seja verdadeira, como já mencionado, é necessário que o tamanho da janela seja maior que a resposta ao impulso do canal, conforme [9].

Tomando a média do sinal no domínio *quefrequency*, temos:

$$\bar{\mathbf{y}} = \frac{1}{N} \sum_{i=1}^N (\hat{\mathbf{s}}_i + \hat{\mathbf{h}}), \quad (6)$$

onde N é o número total de quadros do sinal.

Considerando que o canal seja invariante no tempo e desconsiderando quaisquer efeitos de não-linearidade que possam ser introduzidos pelos circuitos ou pelo canal de transmissão, podemos expressar (6) como:

$$\bar{\mathbf{y}} = \hat{\mathbf{h}} + \frac{1}{N} \sum_{i=1}^N \hat{\mathbf{s}}_i, \quad (7)$$

ou

$$\bar{\mathbf{y}} = \hat{\mathbf{h}} + \bar{\mathbf{s}}. \quad (8)$$

De acordo com [12], caso o sinal de voz seja balanceado¹, a média *cepstral* do sinal de voz tende para zero, ou $\bar{\mathbf{s}} \rightarrow 0$, e então $\bar{\mathbf{y}} \approx \hat{\mathbf{h}}$.

Esta técnica apresenta uma polarização devido ao fato que o balanceamento não ocorre, fazendo com que $\bar{\mathbf{s}}$ não tenda a zero.

A modificação proposta neste artigo visa à compensação da polarização na estimação cega da resposta ao ambiente, introduzida pelo termo $\bar{\mathbf{s}}$ de (8). Para isso, foi verificado que a média da evolução dos coeficientes *cepstrum* tende para uma constante e que ela pode refletir a tendência do peso fonético de uma locução. Considerando \mathbf{m} como a média dos coeficientes *cepstrum* obtidos do sinal limpo, de acordo com [11] ela pode ser encontrada utilizando-se uma segunda gravação limpa com características estatísticas similares às do sinal que foi convoluído com o canal, ou seja, de mesmo texto. Neste artigo, ela foi obtida a partir de sinais de voz de contextos fonéticos distintos gravados por vários locutores.

Assumindo que o termo $\bar{\mathbf{s}}$ seja aproximadamente igual à média \mathbf{m} , pode-se obter uma estimativa mais fiel do canal, conforme a equação abaixo:

$$\bar{\mathbf{y}} - \mathbf{m} = \hat{\mathbf{h}} + \bar{\mathbf{s}} - \mathbf{m}. \quad (9)$$

Admitindo que $\bar{\mathbf{s}} \approx \mathbf{m}$, tem-se:

$$\bar{\mathbf{y}} - \mathbf{m} \approx \hat{\mathbf{h}}, \quad (10)$$

de onde pode-se tirar que a estimativa do canal é dada por $\hat{\mathbf{y}} - \mathbf{m}$.

B. Estimação com Base na Identificação dos Zeros

A idéia básica deste método consiste na obtenção do padrão de resposta do ambiente mediante a observação de valores repetidos de raízes (zeros) extraídas de polinômios representados por segmentos do sinal $y(n)$. Considerando que a resposta $h(n)$ é estacionária, e uma vez que o sinal de fala é variante, a repetição de valores

de zeros em número igual ou superior à quantidade de segmentos evidencia a existência de um padrão que pode ser atribuído ao ambiente.

Conforme exposto em [8], esta técnica de estimação consiste no seguinte procedimento:

- 1) segmentação do sinal contaminado pelo efeito de reverberação;
- 2) obtenção dos zeros associados a cada um dos segmentos;
- 3) identificação do padrão de resposta do ambiente entre os segmentos;
- 4) desfatoração para obtenção da resposta estimada.

Como condições de contorno do método, cada segmento deve possuir completamente a resposta $h(n)$, o que implica em que cada segmento a ser analisado deve ter, no mínimo, tamanho igual ao da resposta ao impulso do ambiente. Outra restrição a ser considerada é a de que a resposta $h(n)$ deve ser estacionária. Esta condição é que permite que se considere que os zeros associados a $H(z)$, $z_i = \alpha_i e^{j\theta_i}$, permanecem em local fixo.

A fim de assegurar uma maior precisão nos resultados, a obtenção das raízes dos polinômios representativos dos segmentos, durante o processo de fatoração, é realizada através do algoritmo Lindsey-Fox [13], uma vez que os segmentos possuem tamanho razoável e este algoritmo é mais adequado à determinação de raízes de polinômios de ordem mais elevada, que apresentem uma constelação de zeros muito próxima à circunferência de raio unitário [13].

A estratégia de identificação do padrão de resposta do ambiente é a representação dos zeros através de uma matriz de singularidades quantizadas. Dada a simetria do problema, uma vez que polinômios de coeficientes reais (caso aqui estudado) geram raízes complexas conjugadas, pode-se mapear parte do plano Z em uma grade de localização de zeros. Inicializando a matriz com o valor nulo, cada zero obtido pela fatoração dos polinômios incrementa em 1 a sua correspondente célula de localização bidimensional. Após realizar a análise dos N segmentos, as células da matriz de quantização que tiverem armazenado o valor N representam as posições dos zeros relativos à resposta ao impulso do ambiente.

A quantização adotada para a localização das raízes no plano Z é realizada da seguinte forma: utilizando coordenadas polares, a magnitude é representada por 2^{B_m} posições, com B_m representando o número de *bits* para a designação da magnitude, enquanto a fase é composta por 2^{B_f} posições, com B_f representando o número de *bits* representativos da fase. Para a fase, é estabelecida uma quantização linear, uma vez que a distribuição dos zeros de acordo com esse parâmetro é uniforme. Já para a magnitude, utiliza-se uma quantização não-linear, com

¹Equilíbrio entre a ocorrência de sons sonoros, fricativos e oclusivos.

uma concentração maior de posições próximas ao círculo unitário, e que permita o estabelecimento de valores máximos e mínimos, conforme (11) [8].

$$\hat{\alpha}_{iq} = D_0 \left[D_1 + tg \left(\frac{\pi}{D_2} \alpha_i q - \frac{\pi}{D_2} \right) \right], \quad (11)$$

onde D_0 , D_1 e D_2 são parâmetros de controle da função e α_{iq} é o valor da magnitude α_i quantizado linearmente.

Por fim, através do emprego de um algoritmo de desfatoração adequado, chega-se à estimativa da resposta ao impulso do ambiente.

III. RESULTADOS EXPERIMENTAIS

Para ambos os métodos de desreverberação, foi adotado o seguinte modelo de resposta ao impulso do ambiente, capaz de simular as reflexões geradas pelo ambiente no sinal de voz:

$$h(n) = \begin{cases} 0, & n < 0 \\ 1, & n = 0 \\ \sum_{i=1}^p \sum_{j=0}^{N_i} q_i^j \delta(n - jn_0), & n > 0, \end{cases} \quad (12)$$

onde p é o número de obstáculos, $|q_i| < 1$ é uma constante que determina a taxa de decaimento da resposta; $\delta(n)$ denota a função impulso unitário; N_i representa o número de reflexões e n_0 , o intervalo de tempo entre cada reflexão (retardo). Neste trabalho foi considerada uma situação de reverberação em três obstáculos ($p = 3$), sendo adotados os seguintes parâmetros para cada situação: na primeira, $N_1 = 20$, $n_0 = 5$ e $q_1 = 0,9$; na segunda, $N_2 = 8$, $n_0 = 11$ e $q_2 = 0,6$, e na terceira, $N_3 = 12$, $n_0 = 8$ e $q_3 = 0,8$. Com este modelo, é obtido um efeito de reverberação com 36 reflexões e a resposta de fase mista $h(n)$, obtida a partir de (12), possui comprimento de 101 amostras. Os resultados práticos apresentados neste artigo levaram em consideração uma base contendo várias frases proferidas por vários locutores que foram apresentadas em [14]. Foi utilizado um processo aleatório de escolha do(s) locutor(es) e da(s) locução(ões). Para ambos os métodos, as locuções, que foram gravadas a uma taxa de $22.050Hz$ e reamostradas a $8.000Hz$. Inicialmente não foi acrescentado ruído.

A. Estimação por Média Cepstral

Foram estabelecidas três estimativas para $|\mathbf{H}(z)|$ com a intenção de extrair informações relevantes sobre o processo. A estimativa A foi obtida através do cálculo da média de um mesmo locutor falando dez frases em sequência (sem a subtração da média de um sinal limpo). A estimativa B foi conseguida a partir da subtração no domínio cepstral entre uma locução corrompida pela

resposta do ambiente e outra locução de mesmo texto, porém proferida por um locutor diferente. Já a estimativa C, foi levantada a partir do processo similar à estimativa B, porém com texto e locutor diferentes.

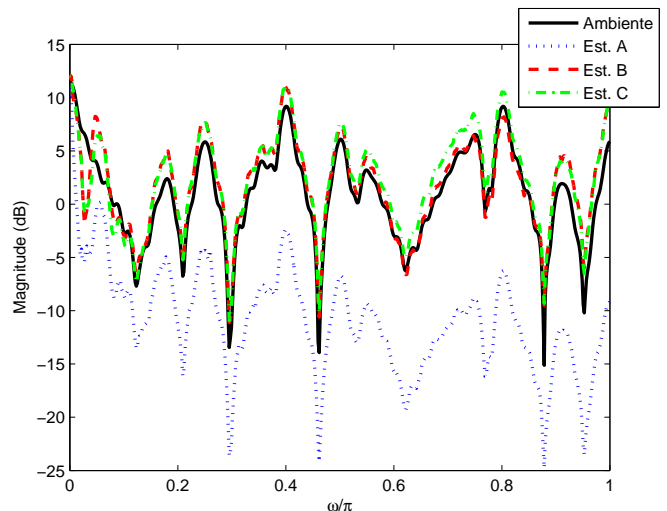


Fig. 1. Resposta em Frequência do Ambiente e Estimativas A, B e C

Pela observação da Figura 1, pode-se chegar às seguintes conclusões:

- pelo comportamento em baixas frequências e a inclinação descendente da curva, pode-se deduzir que na estimativa A há grande influência do contexto fonético;
- as estimativas B e C são mais independentes do contexto fonético, pois não apresentam comportamento similar ao da estimativa A;
- a estimativa B reproduz com razoável fidelidade o comportamento do modelo de reverberação;
- apesar de ser a estimativa B mais precisa que a C, percebe-se claramente que essas duas curvas se sobrepõem na maior parte do espectro; isto traduz a independência de texto e locutor que o procedimento possui.

B. Estimação pela Identificação dos Zeros

Utilizando uma locução e um locutor aleatoriamente escolhidos a partir da mesma base, obteve-se o padrão de identificação de zeros apresentado na Figura 2. Os zeros de cada um dos segmentos são mapeados na grade de quantização. No exemplo apresentado, foram adotados os seguintes valores: $B_f = 12$, $B_m = 8$, $D_0 = 1/7$, $D_1 = 6,9552$, $D_2 = 2,2$, $\alpha_q^{\hat{min}} = 0$ e $\alpha_q^{\hat{max}} = 2,2$.

Após a desfatoração do polinômio de zeros obtido, encontrou-se a resposta estimada do ambiente conforme pode ser visto na Figura 3.

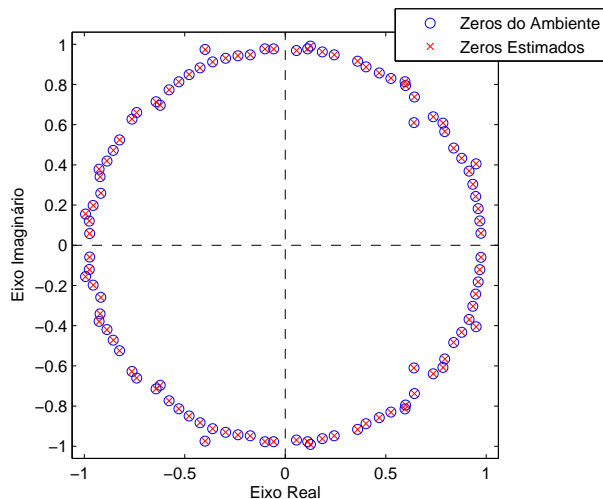


Fig. 2. Constelação de zeros de $H(z)$ e de sua estimativa.

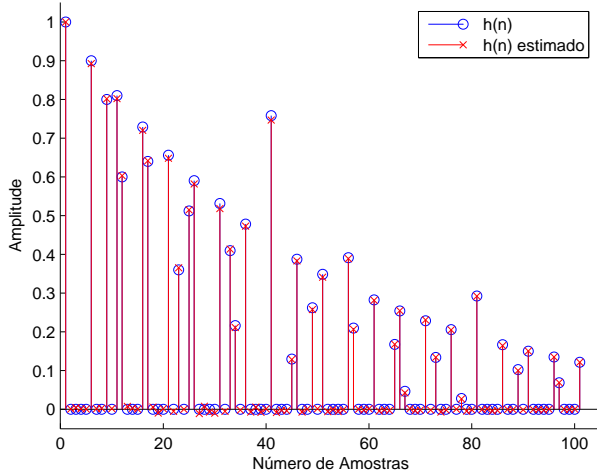


Fig. 3. Comparação entre a resposta do ambiente e a estimada.

O erro médio quadrático entre a resposta estimada e a do ambiente com este método chegou a um valor de aproximadamente $-35,87$ dB, apontando para uma grande similaridade entre as respostas. A curva de resposta em frequência deste resultado seria basicamente idêntica à do ambiente.

C. Robustez a Ruído

Com vistas a estabelecer uma comparação mais efetiva entre as técnicas, escolheu-se verificar a sensibilidade das mesmas a um ruído de fundo de valor muito pequeno. Foram escolhidos dois valores de relação sinal-ruído (RSR) muito altos: 100 e 50 dB.

A técnica de estimação pela identificação dos zeros mostrou-se bastante sensível a esse tipo de excitação e não foi capaz de estimar adequadamente os zeros para

ambos os valores. As Figuras 4 e 5 apresentam, respectivamente, a comparação entre a constelação estimada e a correspondente à resposta do ambiente para uma relação sinal-ruído de 100 dB e o espectro original e o estimado para o ambiente. Pode-se notar uma profunda disparidade, principalmente no espectro. Outra deficiência desta técnica fica exposta aqui: a similaridade da estimativa com a resposta original é altamente dependente do acerto dos zeros, pois tanto o processo numérico quanto matemático de obtenção do polinômio representativo da estimação contribuem para a instabilidade do método.

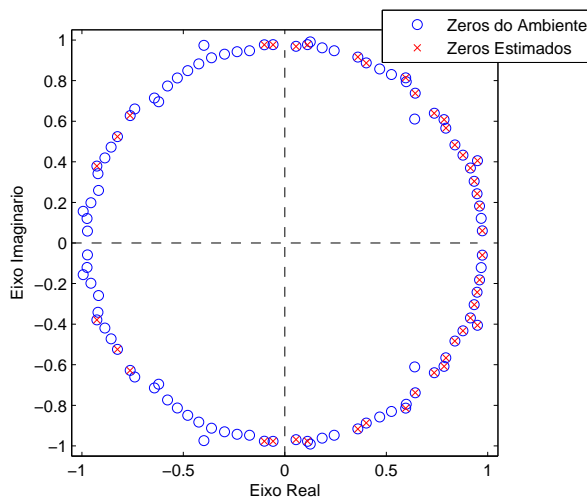


Fig. 4. Constelação de zeros obtida com RSR = 100dB.

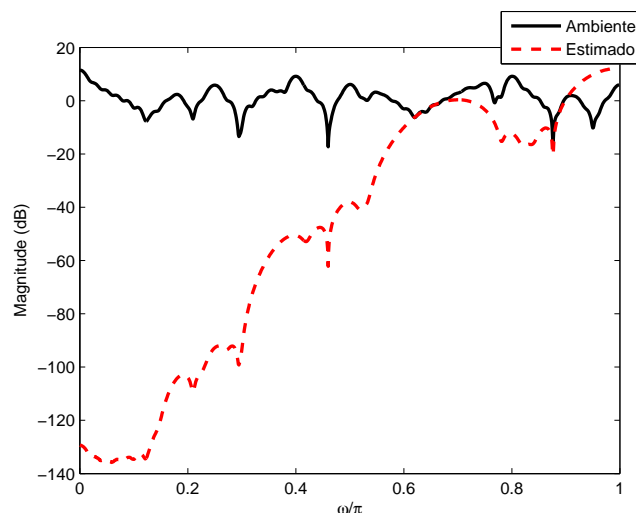


Fig. 5. Resposta em Frequência do Ambiente e Estimativa por Zeros para RSR = 100dB.

A técnica de estimação pela média cepstral apresentou comportamento bem mais robusto à inserção do ruído. Para uma RSR igual a 100 dB, a resposta em frequência fica basicamente idêntica ao montado na Figura 1. Para

uma RSR de 50dB, verifica-se um desempenho muito melhor que o segundo método, além disso, pode-se constatar que a relação de similaridade entre as estimativas B e C se mantém. A menos de alguns desajustes nos vales, pode-se perceber que as estimativas acompanham o comportamento da resposta original na maior parte do espectro. O segundo método mostrou-se inadequado mesmo para valores de RSR muito além dos encontrados na prática. Em tais casos (RSR menor que 30dB), o primeiro método apresenta resultados razoáveis, mas novos gráficos não foram mostrados por limitações de espaço.

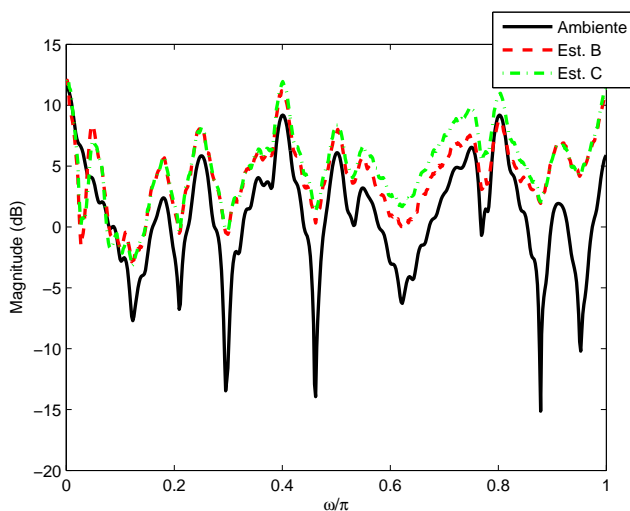


Fig. 6. Comportamento das estimativas A e B com RSR = 50dB.

IV. CONCLUSÕES

Neste trabalho, foi realizada uma análise comparativa entre a utilização de técnicas de estimação de canal para fins de desreverberação de sinais de fala. A técnica de estimação baseada na identificação do padrão de zeros originado pela resposta ao impulso de um ambiente apresentou como vantagem a precisão de acerto sob determinadas condições, contudo, apesar de a estimação baseada na média cepstral não assegurar a mesma precisão, esta última destacou-se no que diz respeito a implementações práticas, demonstrando maior robustez a variações ruidosas pouco significativas, mas que podem interferir no processo de transmissão da voz, e uma velocidade de processamento consideravelmente superior.

Além disso, o método de estimação pela identificação dos zeros demonstrou ser bastante dependente de uma quantização adequada, principalmente em relação à amplitude. Além das conclusões comparativas entre os métodos, pôde-se constatar que a estimação baseada

na média cepstral pode fornecer um mecanismo de identificação cega de canal independente do texto.

AGRADECIMENTOS

Os autores agradecem a Fernando S. Pacheco por ter gentilmente cedido o código-fonte do algoritmo Lindsey-Fox. Os mesmos também são gratos à CAPES, à FAPERJ e ao CNPQ por viabilizarem esta pesquisa.

REFERÊNCIAS

- [1] S. Subramanian, A. P. Petropulu and C. Wendt, "Cepstrum-based deconvolution for speech dereverberation," *IEEE Trans. Speech Audio Processing*, vol. 4, no. 5, pp. 392-396, Sept. 1996.
- [2] A. V. Oppenheim, R. W. Shafer and T. G. Stockham Jr, "Nonlinear filtering of multiple and convolved signals," in *Proc. IEEE*, vol. 56, pp. 1264-1291, Aug. 1968.
- [3] D. Dees, M. Blostein, and P. Cabal, "Reverberant speech enhancement using cepstral processing," in *Proc. IEEE Int. Conf. Acoust. Speech and Signal Processing (ICASSP'91)*, vol. II, pp. 977-980, Toronto, May 14-17, 1991.
- [4] J. B. Allen, D. A. Berkley and J. Blauert, "A multi-microfone signal processing technique to remove room reverberation from speech signals," *J. Acoust. Soc. Amer.*, vol. 62, pp. 912-915, Oct. 1977.
- [5] K. Eneman, J. Duchateau, M. Moonen, D.V. Compernelle and H. V. Hamme, "Assessment of dereverberation algorithms for large vocabulary speech recognition systems," in *Proc. Europ. Conf. Speech Commun. Technol. (EUROSPEECH'03)*, pp. 1-4, Geneva, Sept. 2003.
- [6] S. T. Neely and J. B. Allen, "Invertibility of a Room Impulse Response," *J. Acoust. Soc. Am.*, vol. 66, pp.165-169, 1979.
- [7] J. Mourjopoulos, P. M. Clarkson, J. K. Hammond, "A comparative study of least-squares and homomorphic techniques for the inversion of mixed-phase signals," *Proc. IEEE ICASSP'82*, pp.1858-1861, 1982.
- [8] F. Pacheco e R. Seara, "Remoção do efeito de reverberação do ambiente em sinais de fala considerando um único microfone," *XXI Simpósio Brasileiro de Telecomunicações-SBT'04*, Belém, setembro 2004.
- [9] T. G. Stockham Jr., T. M. Cannon, and R. B. Ingebretsen, "Blind deconvolution through digital signal processing," *International Telecommunications Symposium-ITS2002*, vol. 63, no. 4, pp. 678-692, April, 1975.
- [10] D. G. da Silva, J. A. Apolinário Jr., and C. B. de Lima, "On the effect of the language in CMS channel normalization," in *Proc. IEEE*, Natal, setembro 2002.
- [11] D. Childers, D. Skinner and R. Kemerait, "The cepstrum: a guide to processing," *Proc. IEEE*, Oct. 1977.
- [12] R. J. Mammone, X. Zhang, R. P. Ramachandran, "Robust speaker recognition - a feature-based approach," *IEEE Signal Processing Mag.*, pp. 58-71, Sept. 1996.
- [13] G. A. Sitton, C. S. Burrus, J. W. Fox and S. Treitel, "Factoring very-high-degree polynomials," *IEEE Signal Processing Mag.*, vol. 20, no. 6, pp. 27-42, Nov. 2003.
- [14] A. Alcaim, J. Solewicz, e J. Moraes, "Frequência de ocorrência dos fonos e listas de frases foneticamente balanceadas no português falado no Rio de Janeiro," *Revista da Sociedade Brasileira de Telecomunicações*, vol. 7, no. 1, p. 23-41, dezembro 1992.