

# Remoção do Efeito de Reverberação do Ambiente em Sinais de Fala Considerando um Único Microfone

Fernando S. Pacheco e Rui Seara

**Resumo**—Este artigo apresenta uma técnica de desreverberação de sinais de fala captados por um único microfone. A abordagem proposta baseia-se na estimação da resposta ao impulso do ambiente e posterior filtragem com a resposta inversa. A técnica de estimação consiste na segmentação do sinal degradado e na identificação dos zeros associados ao sistema de degradação (resposta ao impulso do ambiente). Resultados experimentais mostram a eficácia e a aplicabilidade da técnica proposta para respostas com característica de fase mista.

**Palavras-Chave**—Desreverberação de fala, melhoria de sinais de fala, reverberação.

**Abstract**—This paper presents an one-microphone speech dereverberation approach. The proposed technique consists of the room impulse response estimation followed by an inverse filtering. The estimation is carried out by identification of function zeros associated with the degradation system (room impulse response) in several speech segments. Experimental results confirm the effectiveness and applicability of the proposed approach for mixed-phase impulse responses.

**Keywords**—Speech dereverberation, speech enhancement, room reverberation.

## I. INTRODUÇÃO

No percurso entre o emissor e o receptor, o sinal de fala (como qualquer outro sinal acústico) está sujeito a sofrer alterações. Subjetivamente, essas alterações podem ser percebidas como agradáveis ou não. Uma sala de concertos, por exemplo, quando apropriadamente projetada, pode melhorar sensivelmente a percepção subjetiva de uma peça musical. Em outras aplicações, o efeito é oposto, degradando o sinal de fala. De modo geral, as distorções podem ser agrupadas em duas categorias: ruído aditivo e ruído convolucional [1]. Esse último está relacionado às propriedades acústicas do ambiente ou à resposta ao impulso do sistema de aquisição [1]. Em aplicações de telefonia, usando aparelhos convencionais, a distância entre o emissor sonoro (boca) e o receptor (microfone) é curta. Dessa forma, geralmente as distorções associadas ao ambiente são de pequeno nível. Já, em aparelhos telefônicos com viva-voz, o microfone é posicionado a uma distância maior do locutor. Assim, o sinal captado é composto pelo sinal original e por cópias atenuadas e atrasadas desse mesmo sinal geradas pelas reflexões nas paredes e outras superfícies do ambiente. Essa contaminação do sinal prejudica a inteligibilidade da

mensagem falada. Em aplicações de reconhecimento de fala, a contaminação do sinal pode comprometer severamente o desempenho do sistema de reconhecimento. Nesses casos, técnicas de desreverberação são fundamentais para reduzir tal degradação, aumentando a inteligibilidade do sinal de fala.

As abordagens para o problema de reverberação mais conhecidas na literatura têm considerado o processamento de sinais captados por dois ou mais microfones (arranjo de microfones). Uma técnica clássica que emprega arranjo de microfones é a de atraso e soma [2], [3]. Nessa abordagem, como os microfones são colocados em pontos distintos no ambiente, o sinal de fala percorre distâncias diferentes até atingir cada um deles. Através de atrasos apropriados, é possível compensar essas diferenças para que a soma dos sinais dos microfones reforce o sinal original e reduza o efeito das reflexões.

Realizar a desreverberação com o sinal captado por um único microfone e sem conhecimento prévio da resposta ao impulso do ambiente tem sido uma tarefa desafiadora e de difícil solução. Em [4], é discutida uma abordagem para desreverberação usando um único microfone, que emprega análise cepstral para o processamento do sinal degradado e redução do efeito de reverberação. Em tal abordagem, é assumido que a resposta ao impulso do ambiente e o sinal de fala original ocupem regiões distintas no domínio cepstral. Nesse caso, se os componentes do cepstro complexo do sinal degradado, vinculados à reverberação, apresentarem uma estrutura na forma de impulsos, um procedimento de filtragem cepstral (usando um filtro tipo pente) pode ser considerado para a redução (ou mesmo eliminação) do efeito de reverberação [4]. Uma outra alternativa [4] consiste na ponderação do cepstro complexo do sinal contaminado por uma janela que preserve os componentes associados à fala (baixos tempos) e remova aqueles relacionados à reverberação (altos tempos). Análise cepstral também pode ser empregada com o objetivo de estimar a resposta ao impulso do ambiente, como discutido em [5]. Bons resultados em relação à estimação são obtidos para respostas com característica de fase mínima ou de fase mista com poucos zeros fora do círculo de raio unitário [5]. Nesses trabalhos, a grande dificuldade ao empregar análise cepstral é a validade da suposição de que o sinal original e a resposta do ambiente ocupam distintas regiões no domínio cepstral. Essa consideração é válida para respostas ao impulso de fase mínima, sendo possível distinguir a contribuição do sinal de fala nos baixos tempos e da resposta ao impulso do ambiente nos altos tempos. Entretanto, para respostas de fase mista, há contribuição da resposta do ambiente na região

Fernando S. Pacheco e Rui Seara, LINSE – Laboratório de Circuitos e Processamento de Sinais, Departamento de Engenharia Elétrica, Universidade Federal de Santa Catarina, Florianópolis, SC, E-mails: {fernando, seara}@linse.ufsc.br.

Este trabalho foi parcialmente financiado pelo CNPq.

de baixos tempos, dificultando a separação. Em condições reais, as respostas acústicas de ambientes tem característica de fase mista [1], não favorecendo o uso de técnicas de desreverberação baseadas em análise cepstral.

Em [6], [7], um método baseado na estrutura harmônica do sinal de fala é empregado para desreverberação considerando um único microfone. Apesar dos resultados mostrarem-se aceitáveis, uma das maiores dificuldades encontradas nesse método é a necessidade de uma longa etapa de treinamento.

Este artigo apresenta uma nova abordagem para a estimação de respostas ao impulso de ambientes. A principal contribuição da técnica proposta encontra-se na estimação de respostas acústicas com característica de fase mista. A partir da obtenção de uma estimativa da resposta do ambiente, a desreverberação obtida através da filtragem do sinal degradado pela resposta inversa revela-se uma solução trivial.

Este trabalho está organizado como segue. A Seção II apresenta a abordagem de desreverberação de fala proposta, destacando a técnica de estimação da resposta ao impulso do ambiente. Resultados experimentais, levando em consideração a característica de fase mista da resposta do ambiente, são mostrados na Seção III. Por fim, as conclusões do presente artigo são apresentadas na Seção IV.

## II. DESREVERBERAÇÃO DE FALA

O procedimento de desreverberação, adotado neste trabalho, é o de filtragem do sinal degradado pela resposta inversa do sistema de degradação (resposta acústica do ambiente). Esse processo é constituído por três etapas: estimação da resposta do ambiente, determinação da resposta inversa e filtragem do sinal pela resposta inversa obtida. A Figura 1 mostra um diagrama em blocos desse procedimento. A seguir, são discutidas em detalhe cada uma dessas etapas.

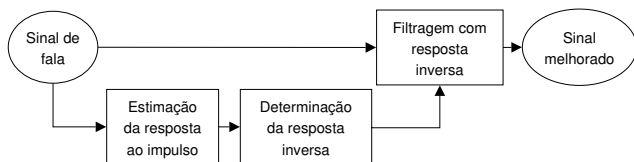


Fig. 1. Diagrama em blocos do procedimento de desreverberação de fala.

### A. Estimação da Resposta do Ambiente

A distorção de um sinal de fala por reverberação pode ser modelada por uma operação de convolução linear no domínio da seqüência. Assim, o sinal de fala modificado  $y(n)$ , captado por um microfone, pode ser expresso como

$$y(n) = s(n) * h(n), \quad (1)$$

onde  $s(n)$  representa o sinal de fala original;  $h(n)$ , a resposta ao impulso do ambiente; e “\*”, a operação de convolução linear.

No domínio da transformada  $Z$ , (1) é representada por

$$Y(z) = S(z)H(z), \quad (2)$$

onde  $Y(z)$ ,  $S(z)$  e  $H(z)$  denotam as transformadas  $Z$  de  $y(n)$ ,  $s(n)$  e  $h(n)$ , respectivamente.

Pode-se, então, representar  $S(z)$  e  $H(z)$ , em suas formas fatoradas, como

$$S(z) = A \prod_{k=1}^{M_1} (1 - r_k z^{-1}) \prod_{k=1}^{M_2} (1 - s_k z^{-1})(1 - s_k^* z^{-1}) \quad (3)$$

$$H(z) = B \prod_{k=1}^{M_3} (1 - g_k z^{-1}) \prod_{k=1}^{M_4} (1 - h_k z^{-1})(1 - h_k^* z^{-1}) \quad (4)$$

onde  $A$  caracteriza um valor de ganho;  $r_k$  e  $\{s_k, s_k^*\}$  representam, respectivamente, os  $M_1$  zeros reais e  $M_2$  pares complexos conjugados associados à função  $S(z)$ . Na função  $H(z)$ ,  $B$  denota um ganho;  $g_k$  e  $\{h_k, h_k^*\}$  representam, respectivamente, os  $M_3$  zeros reais e  $M_4$  pares de zeros complexos conjugados.

Considere-se agora a partição do sinal de fala em  $N$  segmentos e que o comprimento de cada segmento seja, no mínimo, igual ao comprimento da resposta ao impulso do ambiente. Avaliando  $Y(z)$  para cada um dos segmentos, nota-se que, devido à natureza variante do sinal de fala, para a função  $S(z)$ , é pouco provável a ocorrência de zeros comuns a esses segmentos. Por outro lado, fazendo a consideração de que a resposta  $h(n)$  é estacionária, para o conjunto de segmentos analisados, os zeros associados a  $H(z)$  permanecem em localizações fixas no plano  $Z$ . Dessa forma, a partir da constelação de zeros de  $Y(z)$ , avaliada para um certo número de segmentos, é possível identificar um padrão fixo (ou de pequena variação), relacionado a  $H(z)$ , recuperá-lo e determinar uma estimativa de  $h(n)$ .

Assim, com base neste princípio, que é conceitualmente simples, a técnica de estimação proposta consiste em:

- i) segmentação do sinal contaminado (pelo efeito de reverberação);
- ii) obtenção das raízes (zeros) associados a cada um dos segmentos;
- iii) identificação do padrão da resposta do ambiente entre os segmentos;
- iv) desfatoração para obtenção da resposta estimada.

A Figura 2 apresenta o diagrama em blocos do procedimento considerado.

No processo de segmentação, as considerações estabelecidas aos zeros de  $Y(z)$  permanecem válidas se algumas condições forem atendidas. Como mencionado anteriormente, cada segmento analisado deve conter completamente a resposta  $h(n)$ . Além disso, para reduzir a contaminação do segmento corrente com a “cauda” da resposta do segmento anterior, deve-se empregar a mesma estratégia adotada em [5]. No citado trabalho, com o objetivo de reduzir erros de segmentação, é usada uma janela (de ponderação) exponencial decrescente, iniciando no instante subsequente a um certo período de silêncio. Tal janela é dada por  $w(n) = \gamma^n$ , para  $|\gamma| < 1$ .

Como os segmentos são relativamente longos, no processo de fatoração, polinômios de ordem elevada devem ser avaliados. Para essa condição, métodos clássicos de determinação de raízes como o método de Newton ou o baseado nos autovalores da matriz associada ao polinômio apresentam sérios problemas de convergência e um considerável custo computacional. Nesse caso, uma alternativa sedutora é o uso do algoritmo de fatoração Lindsey-Fox [8], apropriado à determinação de

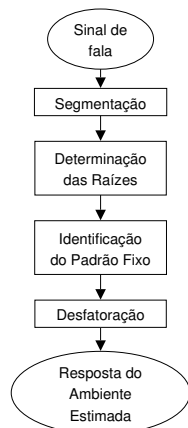


Fig. 2. Diagrama em blocos do procedimento de estimação da resposta ao impulso do ambiente.

raízes de polinômios de ordem elevada, que apresentem uma constelação de zeros muito próxima à circunferência de raio unitário [8]. A idéia central desse algoritmo é o emprego da transformada rápida de Fourier (FFT) para a avaliação do polinômio em anéis concêntricos centrados na origem e de raios próximos à unidade.

Para a identificação do padrão fixo de zeros, associado à função  $H(z)$ , a estratégia proposta é representar os zeros de  $Y(z)$  ( $z_i = \alpha_i e^{j\theta_i}$ ) através de uma matriz de singularidades quantizadas. O plano  $Z$  (parte dele) fica, assim, mapeado em uma grade de localização de zeros. Essa matriz é inicializada com o valor 0. Para cada segmento analisado, a célula de localização bidimensional de cada zero é incrementada em 1. Assim, ao final do processamento dos  $N$  segmentos, as células que representam os zeros (fixos) de  $H(z)$  devem armazenar o valor  $N$ . Dessa forma, a identificação dos zeros de  $H(z)$  torna-se simples e muito rápida. Alguns aspectos do processo de geração da grade de localização de zeros são importantes para melhorar a resolução do procedimento de identificação. Essa grade pode ser representada em coordenadas retangulares ou polares. Como os zeros de  $Y(z)$  estão mais concentrados em torno da circunferência de raio unitário, a grade magnitude-fase (em coordenadas polares) mostra-se como a melhor alternativa. A quantização de fase é composta por  $2^{B_f}$  posições, onde  $B_f$  denota o número de *bits* alocados para a fase. De modo semelhante, a quantização de magnitude apresenta  $2^{B_m}$  posições, onde  $B_m$  denota o número de *bits* alocados para a magnitude. Para respostas contendo apenas coeficientes reais (nosso caso), é suficiente avaliar somente o semi-plano  $Z$  superior (fase entre 0 e  $\pi$ ). Para o problema em questão, a distribuição da fase dos zeros nesse intervalo é considerada uniforme e, desse modo, usa-se uma quantização linear para a fase. Já, a magnitude dos zeros de  $Y(z)$  tem uma distribuição com concentração em torno de 1. Assim, para a magnitude, uma quantização não-linear mostra-se mais eficiente. Neste trabalho, adota-se uma função tangente deslocada para tal quantização. O valor de magnitude quantizado de forma não-linear  $\hat{\alpha}_{iq}$  é obtido pela seguinte relação de mapeamento:

$$\hat{\alpha}_{iq} = D_0 \left[ D_1 + \text{tg} \left( \frac{\pi}{D_2} \alpha_{iq} - \frac{\pi}{D_2} \right) \right], \quad (5)$$

onde  $D_0$ ,  $D_1$  e  $D_2$  são parâmetros de controle da função; e  $\alpha_{iq}$  é o valor de magnitude  $\alpha_i$  quantizado linearmente. Os parâmetros são ajustados de modo a concentrar a grade em torno de 1 e admitir valores entre um mínimo  $\hat{\alpha}_q^{\min}$  e um máximo  $\hat{\alpha}_q^{\max}$ .

Para ilustrar a distribuição das células na grade, a Figura 3 mostra um detalhe da grade obtida com  $B_f = B_m = 5$  bits,  $D_0 = 1/7$ ,  $D_1 = 6,9552$ ,  $D_2 = 2,2$ ,  $\hat{\alpha}_q^{\min} = 0$  e  $\hat{\alpha}_q^{\max} = 1,6$ .

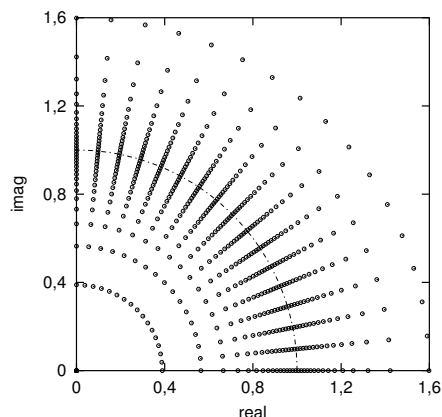


Fig. 3. Grade de localização dos zeros de  $Y(z)$  para o processo de identificação do padrão fixo. Grade obtida com  $B_f = B_m = 5$  bits,  $D_0 = 1/7$ ,  $D_1 = 6,9552$ ,  $D_2 = 2,2$ ,  $\hat{\alpha}_q^{\min} = 0$  e  $\hat{\alpha}_q^{\max} = 1,6$ .

Após a identificação dos zeros associados a  $H(z)$ , obtém-se uma resposta estimada  $h_{\text{est}}(n)$  por meio de um processo de desfatoração.

### B. Determinação da Resposta Inversa e Filtragem

Uma vez estimada a resposta ao impulso do ambiente, realiza-se a determinação da resposta inversa para posterior desconvolução com  $y(n)$ . Como, no caso geral,  $h_{\text{est}}(n)$  é de fase mista, sua inversa não pode ser obtida de forma direta, pois isso levaria a um sistema instável ou não-causal [9]. Nesse caso, uma alternativa é a decomposição de  $h_{\text{est}}(n)$  em componentes de fase mínima e passa-tudo usando análise cepstral [9], [10]. Do componente de fase mínima,  $h_{\text{est}}^{\min}(n)$ , com transformada de Fourier  $H_{\text{est}}^{\min}(e^{j\omega})$ , obtém-se o componente inverso  $g_{\text{est}}^{\min}(n)$ , usando

$$G_{\text{est}}^{\min}(e^{j\omega}) = \frac{1}{H_{\text{est}}^{\min}(e^{j\omega})}, \quad (6)$$

onde  $G_{\text{est}}^{\min}(e^{j\omega})$  é a transformada de Fourier de  $g_{\text{est}}^{\min}(n)$ .

Do componente passa-tudo  $h_{\text{est}}^{\text{pt}}(n)$  não é possível obter diretamente o componente inverso, recorrendo-se então a  $g_{\text{est}}^{\text{pt}}(n) = h_{\text{est}}^{\text{pt}}(M-n)$ , onde  $M$  é o comprimento de  $h_{\text{est}}^{\text{pt}}(n)$ . A resposta inversa é dada então por  $g_{\text{est}}(n) = g_{\text{est}}^{\min}(n) * g_{\text{est}}^{\text{pt}}(n)$ .

Por fim, a obtenção da estimativa do sinal original,  $s_{\text{est}}(n)$ , pode ser realizada através de convolução no domínio da seqüência de  $y(n)$  com  $g_{\text{est}}(n)$  ou por um processo de filtragem no domínio da freqüência, dado por

$$S_{\text{est}}(e^{j\omega}) = Y(e^{j\omega})G_{\text{est}}(e^{j\omega}), \quad (7)$$

onde  $S_{\text{est}}(e^{j\omega})$ ,  $Y(e^{j\omega})$  e  $G_{\text{est}}(e^{j\omega})$  são as transformadas de Fourier de  $s_{\text{est}}(n)$ ,  $y(n)$  e  $g_{\text{est}}(n)$ , respectivamente.

### III. RESULTADOS EXPERIMENTAIS

Para avaliação da abordagem proposta, utilizou-se uma resposta ao impulso simulada, representando as reflexões em um ambiente.

Um modelo simples de resposta ao impulso que simula o efeito de reverberação é obtido por

$$p(n) = \sum_{k=0}^{N_r} q^k \delta(n - kn_0), \quad (8)$$

onde  $|q| < 1$  é uma constante que determina a taxa de decaimento da resposta;  $\delta(n)$  denota a função impulso unitário;  $N_r$  representa o número de reflexões e  $n_0$ , o intervalo de tempo (atraso) entre cada reflexão. A resposta obtida através desse modelo é de fase mínima. Como neste trabalho propomos considerar o problema de desreverberação para respostas de fase mista (caso real), usamos para a resposta ao impulso  $h(n)$  a seguinte expressão

$$h(n) = \begin{cases} 0, & n < 0, \\ 1, & n = 0, \\ p_1(n) + p_2(n), & n > 0, \end{cases} \quad (9)$$

onde  $p_1(n)$  e  $p_2(n)$  são respostas obtidas através de (8). Para  $p_1(n)$ , são empregados os parâmetros  $N = 20$ ,  $n_0 = 5$  e  $q = 0,9$ , enquanto que, para  $p_2(n)$ , os parâmetros considerados são  $N = 8$ ,  $n_0 = 11$  e  $q = 0,8$ . A resposta de fase mista  $h(n)$ , obtida através de (9), modela o efeito de 27 reflexões e tem comprimento de 101 amostras. A Figura 4 apresenta a constelação de zeros associada à resposta  $h(n)$ .

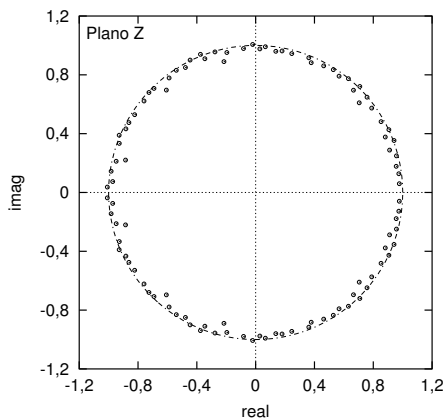


Fig. 4. Constelação de zeros de  $H(z)$ .

A partir de uma gravação realizada com taxa de amostragem de 16 kHz e reamostrada a 4 kHz, são extraídos segmentos de fala que compõem o sinal  $s(n)$ . Cinco segmentos, correspondentes a trechos das sílabas “ca”, “ze”, “mi”, “co”, “zu”, de aproximadamente 110 ms cada, são considerados em nosso experimento. A Figura 5 mostra o sinal original no domínio da seqüência (tempo). O sinal de fala contaminado por reverberação  $y(n)$  é obtido através da operação de convolução linear entre  $s(n)$  e  $h(n)$ , sendo mostrado na Figura 6.

Neste trabalho, os pontos de segmentação são selecionados de forma manual, no entanto, um procedimento de segmentação automática poderia ser facilmente implementado. Os

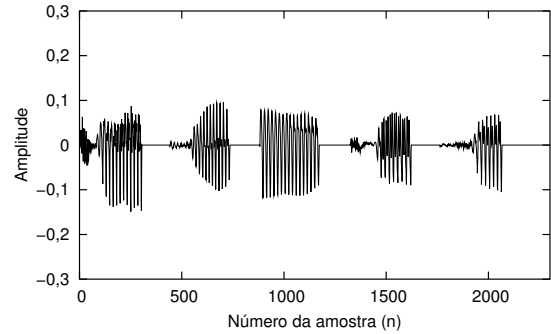


Fig. 5. Sinal original.

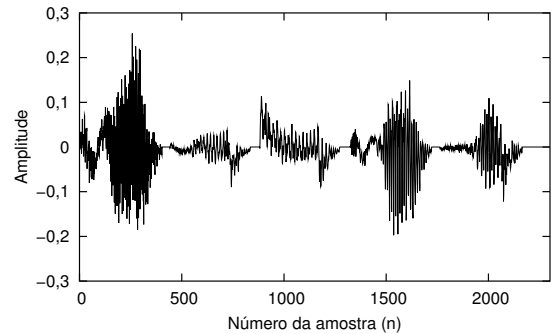


Fig. 6. Sinal degradado por reverberação.

segmentos são ponderados por uma janela exponencial com  $\gamma = 0,996$ . Para cada um dos segmentos, são determinados os zeros associados aos sinais usando o algoritmo de fatoração Lindsey-Fox [8]. A Figura 7 ilustra a constelação de zeros do primeiro segmento do sinal com reverberação  $y(n)$ .

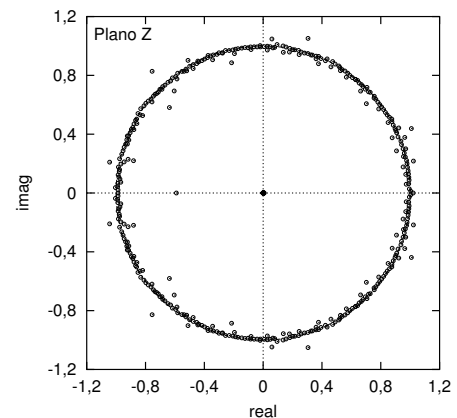


Fig. 7. Constelação de zeros associada ao primeiro segmento do sinal com reverberação.

Os zeros de cada um dos segmentos do sinal são mapeados para a grade de quantização. Para o exemplo, os parâmetros usados são  $B_f = 12$  bits,  $B_m = 8$  bits,  $D_0 = 1/7$ ,  $D_1 = 6,9552$ ,  $D_2 = 2,2$ ,  $\hat{\alpha}_q^{\min} = 0$  e  $\hat{\alpha}_q^{\max} = 2$ . Selecionando as células da grade com valor igual a 5 ( $N$  segmentos utilizados) são obtidos os zeros associados a  $h_{est}(n)$ . Após a desfatoração, obtém-se  $h_{est}(n)$ . A Figura 8 mostra a resposta original  $h(n)$  e a resposta estimada  $h_{est}(n)$ . A partir da resposta estimada,

é obtida a inversa usando o procedimento descrito em [9] e o sinal desreverberado é determinado pela convolução de  $y(n)$  com  $g_{est}(n)$ . A Figura 9 mostra o sinal desreverberado resultante do processamento.

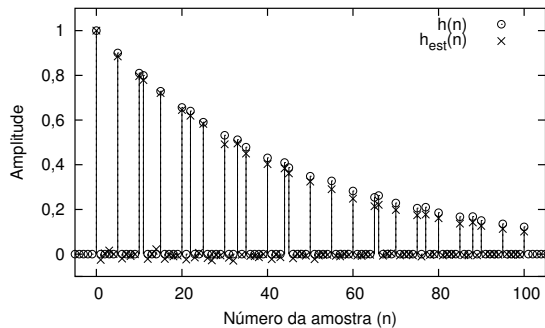


Fig. 8. Resposta original  $h(n)$  (o) e estimada  $h_{est}(n)$  (x).

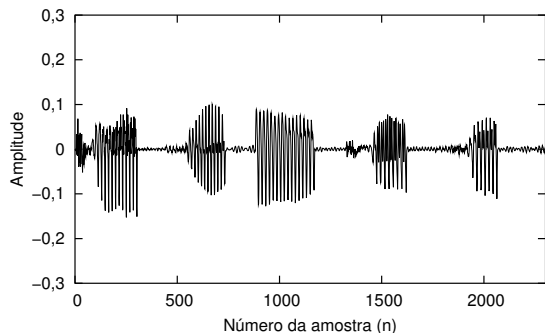


Fig. 9. Sinal desreverberado.

Através dos resultados obtidos no exemplo considerado, pode-se constatar o bom desempenho da técnica proposta. Determinando-se o erro quadrático médio (EQM), definido por (10), entre a resposta original e a resposta estimada, obtém-se um valor de -27,7 dB, o que caracteriza um grau de similaridade elevado entre as duas respostas.

$$\text{EQM} = 10 \log_{10} \frac{\sum_{i=0}^{L_h-1} [h(i) - h_{est}(i)]^2}{\sum_{i=0}^{L_h-1} [h(i)]^2}, \quad (10)$$

onde  $L_h$  representa o comprimento das respostas consideradas e  $i$  caracteriza o índice da amostra. Para o exemplo em questão,  $L_h = 101$  amostras.

#### IV. CONCLUSÕES

Neste trabalho, uma nova abordagem para estimação de resposta ao impulso de ambientes tem sido apresentada. Resultados experimentais corroboram a estratégia proposta para estimar respostas de fase mista em aplicações de desreverberação de sinais de fala contaminados pelo efeito acústico do ambiente, considerando apenas um microfone. O método de desreverberação proposto tem um apelo importante para aplicação em reconhecimento automático de fala. Em trabalhos

futuros, pretende-se avaliar seu desempenho considerando respostas de ambientes reais e usando uma estratégia de segmentação automática do sinal de fala. Além disso, um procedimento de avaliação perceptual (como, por exemplo, o descrito em [11]) poderia nos dar melhores indicativos da qualidade de desreverberação alcançada pela técnica proposta.

#### AGRADECIMENTOS

Os autores agradecem James W. Fox por ter gentilmente disponibilizado o código-fonte do algoritmo Lindsey-Fox.

#### 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] J. B. Allen, D. A. Berkley, and J. Blauert, "A multicrophone signal-processing technique to remove room reverberation from speech signals," *J. Acoust. Soc. Amer.*, vol. 62, pp. 912–915, Oct. 1977.
- [3] K. Eneman, J. Duchateau, M. Moonen, D. V. Comperolle, and H. V. Hamme, "Assessment of dereverberation algorithms for large vocabulary speech recognition systems," in *Proc. Europ. Conf. Speech Commun. Technol. (EUROSPEECH'03)*, Geneva, Sept. 2003, pp. 1–4.
- [4] A. V. Oppenheim, R. W. Schafer, and T. G. Stockham Jr., "Nonlinear filtering of multiplied and convolved signals," *IEEE Trans. Audio Electroacoust.*, vol. AU-16, no. 3, pp. 437–466, Sept. 1968.
- [5] D. Bees, M. Blostein, and P. Kabal, "Reverberant speech enhancement using cepstral processing," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP'91)*, vol. II, Toronto, May 14–17, 1991, pp. 977–980.
- [6] T. Nakatani and M. Miyoshi, "Blind dereverberation of single channel speech signal based on harmonic structure," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP'03)*, Hong Kong, Apr. 2003, pp. 1–4.
- [7] T. Nakatani, M. Miyoshi, and K. Kinoshita, "Implementation and effects of single channel dereverberation based on the harmonic structure of speech," in *Proc. Int. Workshop Acoust. Echo Noise Control (IWAENC'03)*, Kyoto, Sept. 2003, pp. 1–4.
- [8] G. A. Sittou, 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.
- [9] B. D. Radlović and R. A. Kennedy, "Nonminimum-phase equalization and its subjective importance in room acoustics," *IEEE Trans. Speech Audio Processing*, vol. 8, no. 6, pp. 728–737, Nov. 2000.
- [10] A. V. Oppenheim and R. W. Schafer, *Discrete-time signal processing*. Englewood Cliffs: Prentice Hall, 1989.
- [11] *Methods for Subjective Determination of Transmission Quality*, ITU-T Recommendation P.800, Aug. 1996.