

CANCELAMENTO DE ECO ACÚSTICO ESTÉREO: ANÁLISE DE ALGORITMOS ADAPTATIVOS E UM NOVO MÉTODO DE REDUÇÃO DO DESALINHAMENTO

Roberto Roschel Belli¹, Phillip M. S. Burt

Laboratório de Comunicações e Sinais/Dept. de Eng. de Telecomunicações e Controle
Escola Politécnica da Universidade de São Paulo
<<http://www.lcs.poli.usp.br>>
{rbelli, philip}@lcs.poli.usp.br

RESUMO

Neste artigo, considera-se inicialmente o efeito de mudanças na posição do locutor em sistemas de cancelamento de eco acústico estéreo. Usando respostas impulsivas medidas em uma sala, verifica-se que mesmo pequenas mudanças na posição do locutor conduzem a reduções no desalinhamento do filtro adaptativo, aspecto pouco enfatizado até o presente. Este comportamento é, em seguida, explicado teoricamente, para os algoritmos RLS e NLMS. Com base nele é então proposto um novo método para reduzir o desalinhamento do filtro adaptativo: através do chaveamento de microfones adicionais, cria-se um efeito semelhante a pequenas trocas de posição do locutor. Mostra-se que este método tem o mesmo desempenho do que um método proposto por outros autores, com a vantagem de não introduzir degradações no sinal.

1. INTRODUÇÃO

O interesse atual por terminais de comunicação viva-voz é bastante grande. Exemplos da aplicação de equipamentos deste tipo são a vídeo conferência e a educação a distância. Sistemas de telecomunicações que implementam estes serviços, e outros semelhantes, podem ser genericamente denominados de *sistemas de telepresença*. De modo a tornar a telepresença mais semelhante à presença real do interlocutor, é preferível utilizar um sistema de voz estereofônico [1], [6], o que permite a um participante da sessão associar uma localização espacial distinta a cada locutor no outro extremo da conexão.

Em um terminal de comunicação viva-voz há, em geral, um acoplamento acústico grande entre o(s) seu(s) alto-falante(s) e microfone(s). Isto significa que o participante no outro extremo da conexão recebe de volta uma reprodução da própria fala, que é percebida como uma reverberação incômoda, ou mesmo, se o retardo da comunicação for maior que 30 milésimos de segundo, como um eco. Devido a estes efeitos de degradação de qualidade de áudio, denominados genericamente de *eco acústico*, um terminal de comunicação viva-voz deve incorporar um mecanismo de supressão de eco acústico [2].

A técnica preferível para a supressão do eco acústico é o cancelamento adaptativo [1], [3] e [4], sendo que o caso estéreo não é uma simples extensão do caso mono, uma vez que, como discutido adiante, apresenta problemas bastante específicos. Os métodos propostos para contornar estes problemas apresentam como contrapartida uma degradação da qualidade de voz. Neste artigo é

apresentada uma análise do problema de eco acústico onde são abordados aspectos até o presente pouco explorados. Além de permitir uma maior compreensão do comportamento de sistemas de cancelamento de eco acústico estéreo em condições reais, esta análise conduz a um método que não apresenta degradação da qualidade de voz.

2. O PROBLEMA BÁSICO NO CASO ESTÉREO

Na Figura 1 estão indicados os sinais de tempo discreto envolvidos no problema de cancelamento de eco acústico estéreo (o caso de um sistema multicanal com mais de dois canais corresponderia a uma generalização deste caso). Na figura, $g_1(n)$ e $g_2(n)$ são as respostas ao pulso unitário entre o locutor usando o terminal viva-voz em um extremo da conexão (denominado aqui *sala de transmissão*) e os microfones deste terminal; $x_1(n)$ e $x_2(n)$ são os sinais gerados por estes microfones e transmitidos para o terminal viva-voz no outro extremo da conexão (denominado aqui *sala de recepção*); $h_1(n)$ e $h_2(n)$ são as respostas ao pulso unitário entre os alto-falantes deste terminal viva-voz e um de seus microfones; $y(n)$ é o sinal produzido por este microfone; $w_{1,k}(n)$ e $w_{2,k}(n)$ são os coeficientes do filtro adaptativo no instante n , que produz a saída $\hat{y}(n)$; o sinal dito de erro $e(n) = y(n) - \hat{y}(n)$ é o sinal enviado a um dos alto-falantes do terminal viva-voz na sala de transmissão. Idealmente, quando o locutor na sala de transmissão está falando e na sala de recepção não há ninguém falando, deveríamos ter $e(n) \equiv 0$. Notar que os sinais indicados são suficientes para analisar o problema, dado que para os demais sinais (saída do outro microfone na sala de recepção e todos os sinais envolvidos no cancelamento de eco na sala de transmissão) o problema é exatamente equivalente.

Podemos determinar agora a condição a ser obedecida pelo filtro adaptativo, que corresponderá a uma generalização imediata da filtragem de Wiener [7] para o caso de um filtro de duas entradas. Temos, inicialmente, que o sinal de erro $e(n)$ pode ser escrito como (para maior simplicidade não é indicada a dependência em relação ao tempo dos coeficientes de w_1 e w_2):

$$e(n) = y(n) - \mathbf{x}_{1,L}^T(n)\mathbf{w}_{1,L} - \mathbf{x}_{2,L}^T(n)\mathbf{w}_{2,L}, \quad (1)$$

onde

$$y(n) = \mathbf{x}_{1,N}^T(n)\mathbf{h}_{1,L} + \mathbf{x}_{2,N}^T(n)\mathbf{h}_{2,N} \quad (2)$$

é a saída do microfone da sala de recepção e os vetores utilizados, com $i = 1, 2$ e $m = N, L$, são formados como:

¹-Este trabalho foi financiado pela FAPESP (proc. 98/13317-0).

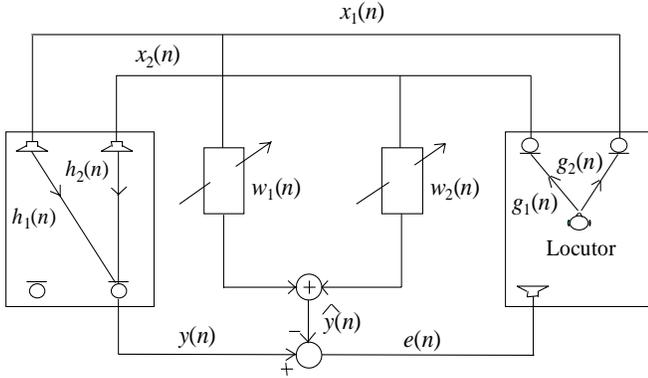


Fig. 1. Definição dos sinais e filtros utilizados.

$$\mathbf{h}_{i,N} = [h_i(0) \quad h_i(1) \quad \dots \quad h_i(N-1)]^T,$$

$$\mathbf{x}_{i,m}(n) = [x_i(n) \quad x_i(n-1) \quad \dots \quad x_i(n-m+1)]^T,$$

$$\mathbf{w}_{i,L} = [w_{i,0} \quad w_{i,1} \quad \dots \quad w_{i,L-1}]^T.$$

Nestas expressões admitiu-se que as respostas ao pulso unitário da sala de recepção, $h_1(n)$ e $h_2(n)$, têm uma duração finita de N amostras. Na realidade, estas respostas têm duração infinita, com um decaimento gradual dos termos a zero. O modelamento destas respostas por respostas de duração finita é, no entanto, mais conveniente para a análise que segue, sendo que N pode ser tomado tão grande quanto se queira para atingir uma determinada precisão no modelamento. Notar também que o filtro adaptativo tem um total de $2L$ coeficientes.

Os coeficientes $\mathbf{w}_{1,L}$ e $\mathbf{w}_{2,L}$ podem ser agrupados no vetor $\mathbf{w} = [\mathbf{w}_{1,L}^T \quad \mathbf{w}_{2,L}^T]^T$, assim como as amostras $\mathbf{x}_{1,L}(n)$ e $\mathbf{x}_{2,L}(n)$, no vetor $\mathbf{x}(n) = [\mathbf{x}_{1,L}^T(n) \quad \mathbf{x}_{2,L}^T(n)]^T$. A minimização do erro quadrático médio σ_e^2 leva à condição de optimalidade $\mathbf{R}\mathbf{w}_o = \mathbf{p}$ para o filtro adaptativo, onde $\mathbf{p} = \mathbf{E}[y(n)\mathbf{x}(n)]$ é o vetor da correlação cruzada entre os sinais de entrada e a saída $y(n)$ e $\mathbf{R} = \mathbf{E}[\mathbf{x}(n)\mathbf{x}^T(n)]$ é a matriz de autocorrelação da entrada. Se a matriz \mathbf{R} for inversível os coeficientes ótimos serão dados por $\mathbf{w}_o = \mathbf{R}^{-1}\mathbf{p}$. Além de determinar esta solução ótima é de interesse calcular a redução de eco acústico proporcionada por um filtro qualquer \mathbf{w} . A redução de eco acústico *ERLE* ("Echo Return Loss Enhancement") utilizada aqui é definida por $ERLE = \sigma_e^2 / \sigma_y^2$, sendo que σ_e^2 é dado por $\sigma_e^2 = \sigma_y^2 - \mathbf{w}^T \mathbf{R} \mathbf{w} + 2\mathbf{p}^T \mathbf{w}$.

Como poderá ser visto a seguir, um outro conceito importante é o de *solução ideal*, dada pelos primeiros L termos das respostas ao pulso unitário da sala de recepção:

$$\mathbf{w}_{id} \triangleq [\mathbf{w}_{1,L}^T \quad \mathbf{w}_{2,L}^T]^T. \quad (3)$$

Com esse conceito pode-se definir o *desalinhamento* dos coeficientes do filtro adaptativo, dado por $\|\mathbf{w} - \mathbf{w}_{id}\|$, onde $\|\cdot\|$ denota

norma euclidiana. Pode-se mostrar facilmente a seguinte relação entre as soluções ótima e ideal [5]:

$$\mathbf{w}_o = \mathbf{w}_{id} + \mathbf{R}^{-1} \overline{\mathbf{R}} \overline{\mathbf{h}} \quad (4)$$

onde $\overline{\mathbf{R}} \triangleq \mathbf{E}[\mathbf{x}(n)\overline{\mathbf{x}}^T(n)]$ com $\overline{\mathbf{x}}(n) = [\overline{\mathbf{x}}_{1,L}^T(n) \quad \overline{\mathbf{x}}_{2,L}^T(n)]^T$,

$$\overline{\mathbf{x}}_{i,L}(n) = [x_i(n-L) \quad \dots \quad x_i(n-N+1)]^T,$$

$$\overline{\mathbf{h}} = [\overline{\mathbf{h}}_{1,L}^T \quad \overline{\mathbf{h}}_{2,L}^T]^T \text{ e}$$

$$\overline{\mathbf{h}}_{i,L} = [h_i(L) \quad h_i(L+1) \quad \dots \quad h_i(N-1)]^T.$$

O problema básico do cancelamento de eco acústico estéreo (e multicanal) resulta do condicionamento muito ruim de \mathbf{R} , que por sua vez decorre da alta correlação existente entre os sinais $x_1(n)$ e $x_2(n)$ (ver Figura 1), gerados pela convolução do mesmo sinal (do locutor) e as respostas \mathbf{g}_1 e \mathbf{g}_2 , respectivamente. Sendo \mathbf{R} mal-condicionada, \mathbf{R}^{-1} terá elementos de alto valor absoluto, e de (4) vemos que a solução ótima será fortemente dependente da posição do locutor, uma vez que alterações em \mathbf{g}_1 e \mathbf{g}_2 levarão a grandes variações dos elementos de \mathbf{R}^{-1} . Mesmo sendo os elementos de $\overline{\mathbf{R}}\overline{\mathbf{h}}$ pequenos (em módulo), devido a um valor de L elevado, adotado para obter uma boa redução do eco, esta dependência, tipicamente, ainda será elevada. Consequentemente, ao ocorrer uma mudança de locutor ou uma mudança na posição de um mesmo locutor na sala de *transmissão*, o filtro adaptativo terá, a princípio, que convergir para uma nova solução ótima muito distante da solução anterior. Além disso, o filtro adaptativo terá que desempenhar seu papel usual de acompanhar variações na sala de recepção.

Uma técnica que tem sido considerada para contornar este problema é diminuir a correlação dos sinais de entrada do filtro adaptativo $x_1(n)$ e $x_2(n)$, melhorando assim o condicionamento da matriz \mathbf{R} . Dado que estes sinais vão também para os alto-falantes da sala de recepção (ver Figura 1), resulta que a contrapartida do melhor condicionamento de \mathbf{R} é, em geral, uma redução da qualidade de áudio (com possível redução do efeito estereofônico) devido à manipulação dos sinais. Deve ser buscada, neste caso, portanto, uma solução de compromisso. Em [6], [8] e [9] algumas técnicas deste tipo são comentadas e utilizadas. Como exemplo, podemos citar a adição de ruído não correlacionado aos dois canais fazendo com que o nível de correlação entre os mesmos diminua. O nível de ruído acrescido, porém, não pode ser alto para não degradar a qualidade de áudio, o que limita a decorrelação obtida.

3. MEDIDAS ACÚSTICAS E MÉTODOS COMPUTACIONAIS

Para obtenção das respostas $h_i(n)$ e $g_i(n)$ (ver Figura 1) em uma determinada sala, foi utilizado o método descrito em [10]. Na Figura 2 tem-se um esquema das posições dos microfones e alto-falantes nas medidas das respostas ao pulso unitário. Na sala de transmissão foram feitas 3 medidas de resposta ao pulso unitário e na sala de recepção, 1 medida. A pequena variação de posição entre as medidas 1 e 2 da sala de transmissão teve como objetivo simular uma pequena mudança de posição de um mesmo locutor, o que, tipicamente, ocorre diversas vezes ao longo do tempo em uma conversação. A maior variação entre as medidas 1 e 3, por outro lado, teve como objetivo representar uma mudança de locutor. Pode-se ver um par de respostas medidas, $h_1(n)$ e $h_2(n)$, na Figura 3. A taxa de amostragem nas medidas efetuadas foi de

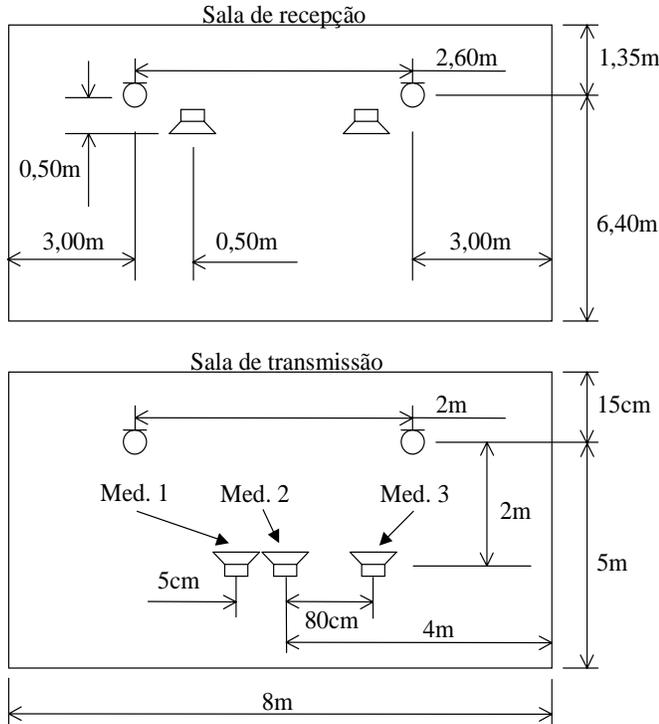


Fig. 2. Posição dos microfones e alto-falantes na sala de transmissão e recepção.

16 KHz. Observa-se que as primeiras 4000 amostras representam a parte mais significativa desta resposta. Devido a este fato, utilizou-se o comprimento de 4000 coeficientes para o filtro adaptativo. Um comprimento menor implicaria numa redução considerável da redução máxima de eco e um aumento do comprimento não traria ganhos consideráveis.

Para determinar o filtro ótimo a partir de $\mathbf{w}_o = \mathbf{R}^{-1}\mathbf{p}$ é necessário inverter a matriz \mathbf{R} que, neste caso, possui dimensão 8000. A inversão dessa matriz se torna inviável utilizando algoritmos normais de inversão de matriz (os quais têm complexidade computacional proporcionais a N^3 , N sendo a ordem da matriz). No entanto, utilizando o fato da matriz \mathbf{R} ser bloco Toeplitz, implementou-se um algoritmo eficiente [11] para a solução de $\mathbf{R}\mathbf{w}_o = \mathbf{p}$, que tem uma complexidade computacional proporcional a N^2 , apenas. Para o cálculo do erro quadrático médio $\sigma_e^2 = \sigma_y^2 - \mathbf{w}^T\mathbf{R}\mathbf{w} + 2\mathbf{p}^T\mathbf{w}$, ainda, a multiplicação $\mathbf{w}^T\mathbf{R}$ pode ser realizada eficientemente por meio de uma FFT, novamente explorando a estrutura bloco Toeplitz de \mathbf{R} .

Com as respostas medidas, foi possível então analisar o efeito de mudanças na posição do locutor na sala de transmissão.

4. EFEITOS DE MUDANÇAS NA POSIÇÃO DO LOCUTOR

Mudanças nas respostas ao pulso unitário da sala de transmissão ocorrem em todos os instantes em que o locutor troca de posição, ou um locutor pára de falar e outro começa. Percebe-se, portanto, que estas mudanças ocorrem com frequência maior do que na sala de recepção, e de forma mais abrupta. Para caracterizar os

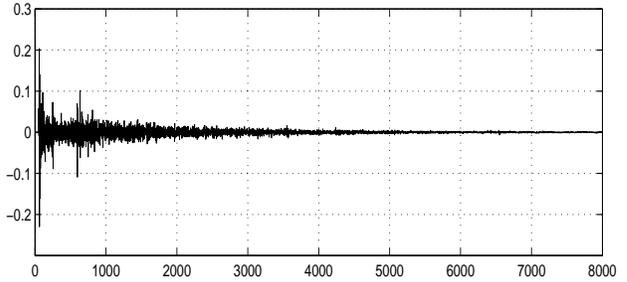


Fig. 3. Resposta ao pulso unitário da sala de recepção.

seus efeitos no cancelamento de eco, calculou-se, inicialmente, o valor da redução de eco para todas as combinações dos pares de respostas de transmissão medidos \mathbf{g}_1 , \mathbf{g}_2 e \mathbf{g}_3 (ver item anterior) com as respectivas soluções ótimas $\mathbf{w}_{o,1}$, $\mathbf{w}_{o,2}$ e $\mathbf{w}_{o,3}$. Os resultados estão na Tabela I. Estes valores fornecem uma estimativa dos transitórios na redução de eco que ocorrem devido a mudanças na posição do locutor.

Pode-se notar que quando a resposta ao pulso unitário da sala de transmissão deixa de ser aquela para a qual foi calculado o filtro ótimo, a redução de eco cai drasticamente. Ou seja, o filtro ótimo utiliza a alta correlação entre os sinais $x_1(n)$ e $x_2(n)$ produzidos na sala de transmissão para alcançar a maior redução de eco possível. Por outro lado, se o filtro adaptativo estivesse não na resposta ótima mas na ideal (desalinhamento nulo, portanto), a redução de eco não seria prejudicada com mudanças na sala de transmissão (ver última coluna da Tabela I). Pode-se ver que com a solução ideal a redução de eco é significativamente menor do que com a solução ótima. Isto ocorre pois a alta correlação entre os sinais $x_1(n)$ e $x_2(n)$ deixa de ser aproveitada. Por outro lado, obtém-se uma menor variação da redução de eco com a posição do locutor.

	$\mathbf{w}_{o,1}$	$\mathbf{w}_{o,2}$	$\mathbf{w}_{o,3}$	\mathbf{w}_{id}
\mathbf{g}_1	26,6dB	9,87dB	5,88dB	19,12dB
\mathbf{g}_2	9,02dB	26,6dB	5,04dB	19,11dB
\mathbf{g}_3	5,47dB	5,41dB	26,8dB	19,25dB

Tabela I

REDUÇÃO DO ECO ACÚSTICO. INFLUÊNCIA DE TROCAS NA POSIÇÃO DO LOCUTOR NA SALA DE TRANSMISSÃO.

Na Tabela II tem-se o mesmo resultado para o caso mono-canal. Nota-se que o problema que ocorre no caso estéreo praticamente não existe no caso mono, não havendo grandes variações da redução de eco com trocas de locutor.

Com base nos resultados acima, uma abordagem para resolver o problema básico de cancelamento de eco acústico estéreo é agir sobre os sinais $x_1(n)$ e $x_2(n)$ de modo a que a solução ótima do cancelamento de eco se aproxime da solução ideal, como em [6], [8] e [9]. Como já comentado, esta abordagem tem o inconveniente de introduzir uma degradação no sinal de áudio. Além disso, como será visto a seguir, a análise apenas com base no comportamento da solução ótima não é suficiente para caracterizar o comporta-

	$w_{o,1}$	$w_{o,2}$	$w_{o,3}$
g_1	21,17dB	20,00dB	18,89dB
g_2	20,11dB	21,13dB	18,86dB
g_3	18,73dB	18,68dB	21,31dB

TabelaII

REDUÇÃO DO ECO ACÚSTICO, CASO MONO. INFLUÊNCIA DE TROCAS NA POSIÇÃO DO LOCUTOR NA SALA DE TRANSMISSÃO.

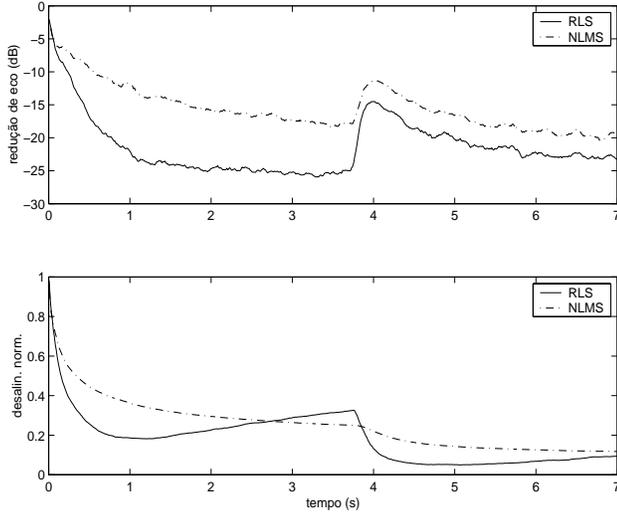


Fig. 4. Simulação com uma grande troca de posição do locutor.

mento de algoritmos adaptativos quando aplicados ao cancelamento de eco acústico estéreo.

Isto pode ser constatado na Figura 4, onde temos o comportamento da redução de eco e do desalinhamento normalizado $\|\mathbf{w} - \mathbf{w}_{id}\|/\|\mathbf{w}_{id}\|$ para os algoritmos *RLS* (implementado em uma versão rápida) e *NLMS* [7] quando há uma grande mudança na posição do locutor (troca no par de respostas ao pulso unitário da sala de transmissão, de g_1 para g_3) em $t = 3,75s$ e na Figura 5, onde há uma pequena mudança na posição do locutor (de g_1 para g_2). Pode-se notar que imediatamente após as mudanças há uma grande variação na redução de eco, como era de se esperar pelos resultados obtidos na Tabela I. Entretanto, nota-se também que após as mudanças de posição do locutor o desalinhamento diminui, comportamento bastante interessante pois implica em que nas mudanças seguintes de posição do locutor a variação na redução de eco será menor. Além disso, ocorrendo mais mudanças de posição do locutor o desalinhamento continuará baixo.

Este comportamento dos algoritmos adaptativos sugere um procedimento alternativo para o cancelamento de eco adaptativo estéreo, como será visto no Item 5. Antes porém, será feita uma análise para buscar explicá-lo.

4.1. Algoritmo *RLS*

O algoritmo *RLS* [7] calcula a matriz \mathbf{R} segundo a expressão

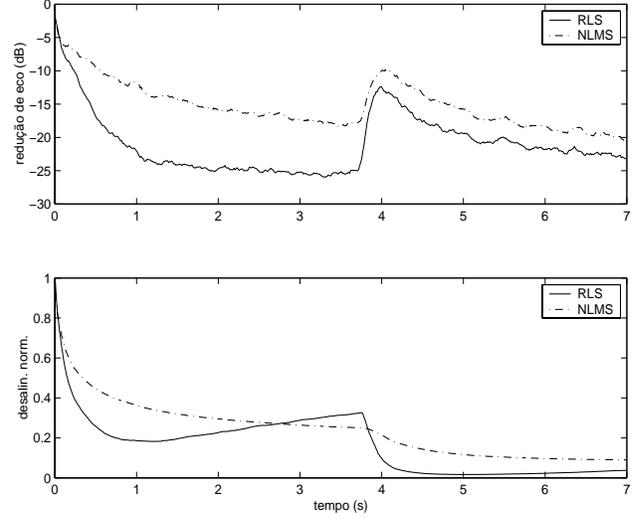


Fig. 5. Simulação com uma pequena troca de posição do locutor.

$$\mathbf{R}_{RLS}(n) = \alpha \mathbf{I} \lambda^n + \sum_{i=1}^n \mathbf{x}(i) \mathbf{x}^T(i) \lambda^{n-i}, \quad (5)$$

onde $\alpha \cdot \mathbf{I}$ é a inicialização para $n = 0$ e λ é o fator de esquecimento. Se considerarmos que os sinais $x_1(n)$ e $x_2(n)$ são estacionários podemos escrever que $\mathbf{E}[\mathbf{x}(n) \cdot \mathbf{x}^T(n)] = \mathbf{R}$ e assim:

$$\begin{aligned} \mathbf{E}[\mathbf{R}_{RLS}(n)] &= \alpha \cdot \mathbf{I} \cdot \lambda^n + \sum_{i=1}^n \mathbf{R} \cdot \lambda^{n-i} \\ &= \alpha \cdot \mathbf{I} \cdot \lambda^n + \mathbf{R} \cdot \frac{1 - \lambda^{n+1}}{1 - \lambda}. \end{aligned} \quad (6)$$

Este é o caso em que não há trocas de respostas ao pulso unitário na sala de transmissão. Caso haja uma mudança na posição do locutor, podemos, desprezando os instantes próximos à troca, considerar que temos um sinal estacionário antes da mudança e outro sinal estacionário diferente depois. Considerando que a troca ocorreu em $n = N_1$, temos que, depois da mudança

$$\begin{aligned} \mathbf{E}[\mathbf{R}_{RLS}(n)] &\cong \alpha \cdot \mathbf{I} \cdot \lambda^n + \mathbf{R}_1 \cdot \frac{1 - \lambda^{N_1+1}}{1 - \lambda} \cdot \lambda^{n-N_1} \\ &\quad + \mathbf{R}_2 \cdot \frac{1 - \lambda^{n-N_1}}{1 - \lambda}, \end{aligned} \quad (7)$$

onde $\mathbf{E}[\mathbf{x}(n) \cdot \mathbf{x}^T(n)] = \mathbf{R}_1$, para $n < N_1$, e $\mathbf{E}[\mathbf{x}(n) \cdot \mathbf{x}^T(n)] = \mathbf{R}_2$, para $n > N_1 + L$, onde L é o comprimento do filtro adaptativo. A imposição $n > N_1 + L$ é feita para que vetor $\mathbf{x}(n)$ só tenha amostras obtidos após a mudança. Por causa disto, a igualdade na Equação (6) se transforma em aproximação em (7).

Com um maior número de trocas basta acrescentar mais termos à Equação (7). Para o cálculo do vetor de correlação cruzada \mathbf{p} o procedimento é análogo, e se chega à seguinte relação, semelhante à do cálculo de \mathbf{R} :

$$\mathbf{E}[\mathbf{p}_{RLS}(n)] \cong \mathbf{p}_0 \cdot \lambda^n + \mathbf{p}_1 \cdot \frac{1 - \lambda^{N_1+1}}{1 - \lambda} \cdot \lambda^{n-N_1} + \mathbf{p}_2 \cdot \frac{1 - \lambda^{n-N_1}}{1 - \lambda}. \quad (8)$$

Vemos, portanto, que durante um certo período de tempo após uma mudança de posição do locutor na sala de transmissão, a matriz \mathbf{R} passa a ser a combinação de duas matrizes de autocorrelação distintas, o que melhora bastante o seu condicionamento. Consequentemente, o filtro adaptativo obtido pela solução recursiva de $\mathbf{w}(n) = \mathbf{R}_{RLS}^{-1}(n)\mathbf{p}_{RLS}(n)$ fica mais próximo da solução ideal (menor desalinhamento).

4.2. Algoritmo NLMS

Com (4) o desalinhamento dos coeficientes do filtro adaptativo pode ser escrito como

$$\mathbf{w} - \mathbf{w}_{id} = \mathbf{w} - \mathbf{w}_o - \delta, \quad (9)$$

onde $\delta = \mathbf{R}^{-1}\overline{\mathbf{R}}\overline{\mathbf{h}}$ (para maior simplicidade, foi omitida a dependência em relação ao tempo de \mathbf{w}). Após um número suficientemente grande de iterações do algoritmo adaptativo, temos $\mathbf{w} \simeq \mathbf{w}_o$ e $\mathbf{w} - \mathbf{w}_{id} \simeq -\delta$. No entanto, em função do mau condicionamento de \mathbf{R} , $\mathbf{w} - \mathbf{w}_o$ tende a zero bem mais rapidamente em algumas direções do que em outras, o que, como veremos a seguir, pode fazer com que o desalinhamento, de início, diminua. Ocorrendo então uma mudança em \mathbf{w}_o e \mathbf{R} (devido a uma mudança na posição do locutor), pode iniciar-se um novo período de redução do desalinhamento, e assim sucessivamente a cada mudança de posição do locutor, até que o desalinhamento atinja um certo valor mínimo (não necessariamente nulo, porém).

Para analisar em maior detalhe este comportamento, considere inicialmente os autovalores λ_i da matriz de autocorrelação \mathbf{R} (de dimensão $2L$), para os quais sempre vale $\lambda_i \geq 0$ e $\sum_i \lambda_i = L\sigma^2$, onde $\sigma^2 = (\sigma_{x,1}^2 + \sigma_{x,2}^2)$ e $\sigma_{x,i}^2$ é a potência de $x_i(n)$. Devido ao mau condicionamento de \mathbf{R} temos $\min\{\lambda_i\} \ll \sigma^2$ e $\max\{\lambda_i\} > \sigma^2$. Estabelecendo um certo limiar $m\sigma^2$, $m > 1$, podemos então dividir o conjunto dos autovalores de \mathbf{R} nos conjuntos

$$\mathcal{A} : \lambda_i \mid \lambda_i \geq m\sigma^2 \quad \text{e} \quad \mathcal{B} : \lambda_i \mid \lambda_i < m\sigma^2, \quad (10)$$

e reescrever (9) como

$$\mathbf{w} - \mathbf{w}_{id} = [\mathbf{w} - \mathbf{w}_o - \delta]_{\mathcal{A}} + [\mathbf{w} - \mathbf{w}_o - \delta]_{\mathcal{B}}, \quad (11)$$

onde $[\cdot]_{\mathcal{A}}$ e $[\cdot]_{\mathcal{B}}$ indicam projeções do argumento nos sub-espacos dos autovetores de \mathbf{R} associados aos autovetores dos conjuntos \mathcal{A} e \mathcal{B} , respectivamente. Uma vez que os autovetores de uma matriz de autocorrelação são ortogonais vale ainda

$$\|\mathbf{w} - \mathbf{w}_{id}\|^2 = \|[\mathbf{w} - \mathbf{w}_o - \delta]_{\mathcal{A}}\|^2 + \|[\mathbf{w} - \mathbf{w}_o - \delta]_{\mathcal{B}}\|^2. \quad (12)$$

No algoritmo NLMS, as componentes de $\mathbf{w} - \mathbf{w}_o$ nas direções dos autovetores associados aos maiores autovalores de \mathbf{R} tendem a zero mais rapidamente do que nas direções associadas aos menores autovalores [7]. Consideremos então que em (10) escolhe-se um valor suficientemente alto para m tal que após um certo número

de iterações do algoritmo adaptativo os coeficientes \mathbf{w}' sejam tais que, em termos práticos, $\mathbf{w}' - \mathbf{w}_o = 0$. O quadrado da norma do desalinhamento valerá então

$$\|\mathbf{w}' - \mathbf{w}_{id}\|^2 = \|[\delta]_{\mathcal{A}}\|^2 + \zeta + \|[\mathbf{w} - \mathbf{w}_o - \delta]_{\mathcal{B}}\|^2, \quad (13)$$

onde $\zeta = \|[\mathbf{w}' - \mathbf{w}_o - \delta]_{\mathcal{B}}\|^2 - \|[\mathbf{w} - \mathbf{w}_o - \delta]_{\mathcal{B}}\|^2$ terá um valor absoluto relativamente pequeno pois a variação de $[\mathbf{w}' - \mathbf{w}_o]_{\mathcal{B}}$ será relativamente pequena. Portanto, de (12) e (13), desde que

$$\|[\mathbf{w} - \mathbf{w}_o - \delta]_{\mathcal{A}}\|^2 > \|[\delta]_{\mathcal{A}}\|^2 + \zeta \quad (14)$$

teremos, após o número de iterações em questão, uma redução no desalinhamento, isto é, $\|\mathbf{w}' - \mathbf{w}_{id}\|^2 < \|\mathbf{w} - \mathbf{w}_{id}\|^2$. Adiante analisaremos a validade de (14). Antes porém, considere o efeito de uma mudança na posição do locutor: com \mathbf{w}'_o e δ' associados a esta nova posição podemos escrever, analogamente a (9), $\mathbf{w}' - \mathbf{w}_{id} = \mathbf{w}' - \mathbf{w}'_o - \delta'$, e analogamente a (12), $\|\mathbf{w}' - \mathbf{w}_{id}\|^2 = \|[\mathbf{w}' - \mathbf{w}'_o - \delta']_{\mathcal{A}'}\|^2 + \|[\mathbf{w}' - \mathbf{w}'_o - \delta']_{\mathcal{B}'}\|^2$, de onde resulta que

$$\|\mathbf{w}' - \mathbf{w}_{id}\|^2 \geq \|[\mathbf{w}' - \mathbf{w}'_o - \delta']_{\mathcal{A}'}\|^2. \quad (15)$$

Para que novas iterações do algoritmo levem a uma nova redução do desalinhamento, temos, analogamente a (14), a condição

$$\|[\mathbf{w}' - \mathbf{w}'_o - \delta']_{\mathcal{A}'}\|^2 > \|[\delta']_{\mathcal{A}'}\|^2 + \zeta'. \quad (16)$$

Podemos ver que à medida que o desalinhamento diminui, expressões análogas a (15) e (16) referentes às novas posições do locutor podem se tornar mutuamente exclusivas, de onde resulta que o desalinhamento pode parar de diminuir com sucessivas trocas na posição do locutor, antes de chegar a zero. Este comportamento é, de fato, observado na prática.

Consideremos agora a validade de (14). Da desigualdade triangular $\|\mathbf{a} + \mathbf{b}\| \geq \|\mathbf{a}\| - \|\mathbf{b}\|$ resulta que

$$\|[\mathbf{w} - \mathbf{w}_o]_{\mathcal{A}}\| > 2\|[\delta]_{\mathcal{A}}\| + \sqrt{\zeta} \quad (17)$$

com $\zeta = \max(\zeta, 0)$ é uma condição suficiente para (14). Por outro lado, uma vez que para uma matriz de autocorrelação sempre vale $\|\mathbf{R}\mathbf{v}\| \leq \max\{\lambda_i\}\|\mathbf{v}\|$ e $\|\mathbf{R}^{-1}\mathbf{v}\| \leq \|\mathbf{v}\|/\min\{\lambda_i\}$, temos, com (10), que

$$\|[\delta]_{\mathcal{A}}\| = \|[\mathbf{R}^{-1}\overline{\mathbf{R}}\overline{\mathbf{h}}]_{\mathcal{A}}\| \leq \frac{\|\overline{\mathbf{R}}\overline{\mathbf{h}}\|}{\min(\mathcal{A})} \leq \frac{\|\overline{\mathbf{R}}\overline{\mathbf{h}}\|}{m\sigma^2}. \quad (18)$$

Portanto, com (17), uma condição suficiente para (14) é

$$\|[\mathbf{w} - \mathbf{w}_o]_{\mathcal{A}}\| > 2\frac{\|\overline{\mathbf{R}}\overline{\mathbf{h}}\|}{m\sigma^2} + \sqrt{\zeta}. \quad (19)$$

Para um \mathbf{w} sem uma direção preferencial, isto é, para o qual tanto $\|[\mathbf{w} - \mathbf{w}_o]_{\mathcal{A}}\|$ quanto $\|[\mathbf{w} - \mathbf{w}_o]_{\mathcal{B}}\|$ têm valores apreciáveis, a condição (19) será atendida, uma vez que ζ é relativamente pequeno e pela construção de $\overline{\mathbf{R}}$ e $\overline{\mathbf{h}}$, $\|\overline{\mathbf{R}}\overline{\mathbf{h}}\| \ll \sigma^2$. Além disso, ainda que (19) não seja atendida para o valor inicial de \mathbf{w} (por exemplo, se $\mathbf{w} = \mathbf{w}_o$) e que, portanto, o desalinhamento permaneça elevado, ao ocorrer uma mudança na posição do locutor, tipicamente $\|[\mathbf{w} - \mathbf{w}'_o]_{\mathcal{A}}\|$ terá um valor apreciável, e o desalinhamento, durante um certo tempo, apresentará então uma redução. O prosseguimento das mudanças de posição do locutor levará então, como visto, o desalinhamento a permanecer em um valor reduzido.

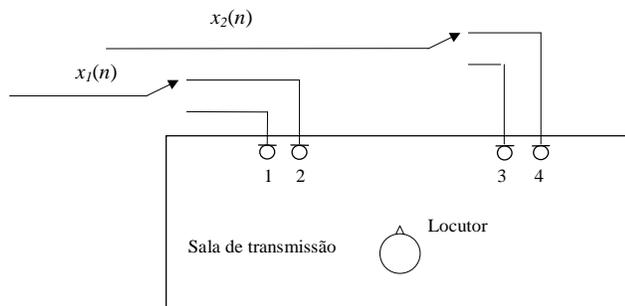


Fig. 6. Esquema do método de trocas de microfones.

5. MÉTODO DE CHAVEAMENTO DE MICROFONES

Uma maneira de provocar um efeito semelhante a mudanças na posição do locutor, tentando assim diminuir o desalinhamento, é utilizar microfones adicionais na sala de transmissão, como representado na Figura 6. Assim, tanto o sinal $x_1(n)$ quanto o sinal $x_2(n)$ passam a ser gerados ora por um e ora por outro microfone. Este chaveamento de microfones equivale a pequenas trocas na posição do locutor, levando, portanto, a uma redução no desalinhamento e não degradando a qualidade de voz.

A melhora obtida na variação da redução de eco e no desalinhamento com o método descrito acima, foi verificada com simulações empregando respostas medidas. Foi utilizada uma separação de 5 cm entre os microfones 1 e 2 assim como entre os microfones 3 e 4. O chaveamento de microfones foi realizado a cada 2000 iterações, ou seja, 8 trocas por segundo. Na Figura 7 estão os resultados obtidos. Para efeitos de comparação considerou-se também o método de acrescentar uma não-linearidade em cada sinal $x_i(n)$ segundo $x'_i(n) = x_i(n) + 0.5|x_i(n)|$ [8]. Observa-se que a melhoria introduzida pelos dois métodos é bastante semelhante, devendo-se repetir que com o método proposto nenhuma deterioração é introduzida na voz.

6. CONCLUSÃO

Usando respostas impulsivas medidas em uma sala, verificou-se que mesmo pequenas mudanças na posição do locutor conduzem a reduções no desalinhamento do filtro adaptativo em sistemas de cancelamento de eco acústico estéreo. Este comportamento foi explicado teoricamente, para os algoritmos RLS e NLMS. Um novo método para reduzir o desalinhamento do filtro adaptativo foi proposto, baseado no chaveamento de microfones adicionais. Este método tem o mesmo desempenho do que um método proposto por outros autores, com a vantagem de não introduzir degradações no sinal.

7. REFERÊNCIAS

- [1] S. Makino, "Acoustic Echo Cancellation", *IEEE Signal Processing Magazine*, Setembro 1997, pp. 39-41.
- [2] E. Hänsler, "The hands-free telephone problem - An annotated bibliography", *Signal Processing*, vol. 27, no. 3, Junho 1992, pp.259-271.
- [3] E. Hänsler, ed., *Signal Processing- Special Issue Acoustic Echo and Noise Control*, vol. 64, no. 1, Janeiro 1998.

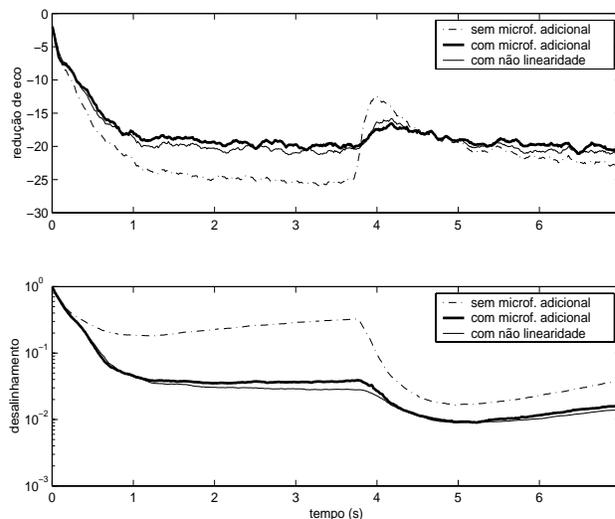


Fig. 7. Simulação com e sem troca de microfones e com uso de não linearidade, troca da posição do locutor em $t = 3, 75$ s.

- [4] A. Gilloire et al., "State of the Art in Acoustic Echo Cancellation", em A.R. Figueiras-Vidal, Ed., *Digital Signal Processing in Telecommunications*, Springer, 1996.
- [5] A. Gilloire, "Current Issues in Stereophonic and Multi-Channel Acoustic Echo Cancellation", *IWAENC97*
- [6] M.M. Sondhi, D. R. Morgan e J. L. Hall, "Stereophonic Acoustic Echo Cancellation-An Overview of the Fundamental Problem", *IEEE Signal Processing Letters*, Vol. 2, NO. 8, Agosto de 1995, pp. 148-151
- [7] S. Haykin, *Adaptive Filter Theory*, 3a. ed., Prentice Hall, 1996.
- [8] J. Benesty, D. R. Morgan e M.M. Sondhi, "A Better Understanding and an Improved Solution to the Problems of Stereophonic Acoustic Echo Cancellation", *ICASSP 97*, pp.303-306.
- [9] S. Shimauchi, Y. Haneda, S. Makino e Y. Kaneda, "New Configuration for a Stereo Echo Canceller with Nonlinear Pre-Processing", *IWAENC97*.
- [10] P.M.S. Burt, "Measuring Acoustic Responses with Maximum-length Sequences", *International Telecommunications Symposium 1998 (ITS'98)*, pp.284-289.
- [11] H. Akaike, "Block Toeplitz Matrix Inversion", *SIAM J. Appl. Math.*, Vol.24, NO.2, Março de 1973.