Redução de Ruído em Sinais de Voz Usando Função de Limiar *SoftSoft* para Minimizar o Erro Quadrático Médio

Irineu Antunes Júnior

***Resumo*— A redução de ruído em sinais corrompidos por AWGN pode ser feita procurando-se uma versão suave para o sinal ruidoso. Num domínio transformado, a regra Soft é aplicada com sucesso para a classe de sinais suaves por trechos, a qual, infelizmente, não inclui sinais de voz. Neste trabalho, inova-se ao empregar uma abordagem variacional para obter regras especializadas (SoftSoft e LogSoft) aplicadas aos coeficientes de voz com o objetivo de minimizar o erro médio quadrático. Experimentos indicam que, ao se minimizar este erro, a regra SoftSoft proporciona menor distorção do que o método Soft convencional.**

***Palavras-Chave*— Soft, SoftSoft, redução de ruído em voz.**

***Abstract*— The problem of denoising signals corrupted by AWGN can be solved by finding a smoother version of the noisy signal. In a transformed domain, Soft-thresholding is used successfully to denoise smooth signals. Unfortunately, speech signals are not smooth, and non-vocalic sounds are similar to white noise contamination. For such situations, we consider the use of new specialized thresholding functions (LogSoft and SoftSot) obtained by a variational approach which assumes a compromise between perceptual related measures. Simulations show that, when the objective is to minimize the mean-square error, the SoftSoft method provides smaller distortion than conventional Soft-thresholding.**

***Keywords*— Soft-thresholding, SoftSoft, Speech Denoising.**

I. Introdução

Muitas aplicações envolvendo sinais de voz podem se beneficiar com o emprego de um método de redução de ruído. Dentre as possíveis aplicações, pode-se citar a redução de ruído para melhoria de inteligibilidade de fala ou para aprimoramento do desempenho de sistemas de reconhecimento de voz, proporcionando um pré-processamento da entrada e melhorando a sua relação sinal-ruído.

Uma forma de contaminação quase sempre encontrada na maioria das aplicações é o ruído de banda larga que, comumente, pode ser modelado como *AWGN* (*additive white Gaussian noise*). Este trabalho considera a redução de ruído em sinal de voz com *AWGN* por meio de técnicas de modificação de coeficientes transformados, fazendo alterações no espectro do sinal de voz. Nesta categoria de métodos, dentre os quais se podem citar [1], [2] e [3], uma das primeiras técnicas experimentadas foi a subtração espectral [4, p.333] que consiste em subtrair do espectro da entrada uma estimativa da densidade espectral de potência do ruído. No caso de ruído branco, o processamento é simplificado, bastando subtrair um valor constante (a potência do ruído) de todos os coeficientes, já que, neste caso, o espectro do ruído é plano.

Incluído na mesma categoria de métodos, pode-se considerar o uso de funções de limiar (*thresholding functions*) concebidas para produzir aproximação de sinais pela modificação de seus coeficientes *wavelet*. Embora, a aplicação inicial destas funções não tenha sido para sinais de voz, elas também podem ser empregadas com sucesso para redução de ruído em voz por meio da modificação dos coeficientes *wavelets* ou de outras tranformadas como a *FFT* ou a *DCT* (*discrete cosine transform*), sendo normalmente empregada a função *Soft* ou funções semelhantes a ela.

Neste trabalho, consideram-se propostas de modificação dos coeficientes usando funções cujo formato é bem diferente das usualmente empregadas. Em [5] empregou-se uma abordagem estatística para justificar o uso de uma delas (a função *SoftSoft*) e, também, procedeu-se minimização do erro médio quadrático (MSE). No presente trabalho, inova-se ao se considerar uma abordagem variacional para obter uma nova função (a *LogSoft*) e justificar o uso da função especializada *SoftSoft* com parâmetros otimizados para minimizar o erro. O objetivo aqui considerado consiste em determinar o melhor resultado que se pode alcançar com *SoftSoft* em comparação com o *Soft* quando o objetivo é minimizar este erro. Para isto, emprega-se um estimador "clarividente", isto é, que supõe conhecido o sinal de voz livre de ruído, possibilitando o cálculo do MSE em função dos parâmetros de ajuste das funções *SoftSoft* e *Soft*, sendo os valores ótimos desses parâmetros encontrados por experimentos Monte Carlo.

No processamento descrito a seguir, a redução de ruído num sinal de voz corrompido por *AWGN* é obtida aplicando-se funções de limiar aos coeficientes *DCT*. Inicialmente, na Seção II, é detalhado o processamento em blocos aqui empregado. Na Seção III, apresenta-se uma abordagem variacional para se obter uma dada função de limiar, que pode ser considerada como o funcional que minimiza uma relação de compromisso entre a fidelidade (preservação de coeficientes) e a suavidade (redução de coeficientes) do resultado. Por fim, na Seção IV, confirma-se por meio de simulações computacionais que o emprego da função de limiar especializada *SoftSoft* fornece um resultado com menor erro quadrático médio do que a função *Soft* convencional.

II. Descrição do Processamento Empregado

O sinal de voz pode ser considerado aproximadamente estacionário em intervalos da ordem de 30 ms de duração ([4,

Irineu Antunes Júnior, CECS, Engenharia de Informação, Universidade Federal do ABC, São Paulo, Brasil, e-mail: irineu.antunes@ufabc.edu.br.

p.13]), por isto, considera-se a transformada em blocos de curta duração. Inicialmente, dado um trecho de voz ruidoso $\{y(n)\}_{n=0}^{N-1}$, com N amostras, este é dividido em M blocos, com índices $m = 0, 1, \ldots, M-1$, sendo cada bloco de comprimento $B = N/M$, suposto inteiro. No m-ésimo bloco, o trecho ruidoso, $\mathbf{y}_m = \{y(n+mB)\}_{n=0}^{B-1}$, é dado pela a soma do trecho de voz ($\mathbf{s}_m$) com uma realização ($\mathbf{w}_m$) do *AWGN* de média nula e densidade espectral de potência constante. Como a transformada é um operador linear: $\mathrm{DCT}(\mathbf{y}_m) = \mathrm{DCT}(\mathbf{s}_m + \mathbf{w}_m)$, ou seja, os coeficientes $\mathbf{Y}_m = \mathbf{S}_m + \mathbf{W}_m$. (Note-se que a DCT tem comprimento $K = B$, o número de amostras no bloco.)

Em seguida, como é usual nos métodos de *thresholding*, consideram-se apenas funções que modificam individualmente cada um dos coeficientes, sendo uma estimativa do coeficiente do sinal de voz

$$\hat{S}_m[k] = g_{\mathbf{t}}(Y_m[k]), \tag{1}$$

onde $Y_m[k]$ é o coeficiente ruidoso de índice k do m-ésimo bloco.

Finalmente, a transformada inversa de cada bloco permite reconstruir um sinal com menor nível de ruído ($\hat{\mathbf{s}}$). Neste trabalho, a função de modificação dos coeficientes, $g_{\mathbf{t}}(\cdot)$, tem o conjunto de parâmetros ($\mathbf{t}$) ajustados para minimizar o MSE do resultado.

III. Funções para Modificação de Coeficientes

A. A Função de Soft-thresholding

Uma função frequentemente usada para redução de ruído em sinais de voz é a função *Soft*

$$g_t^{(S)}(X) = \mathrm{sgn}(X) \cdot \max\{\, (|X| - t)\ ;\ 0\,\} \tag{2}$$

que depende de um único parâmetro t. Nesta expressão, X é empregado para representar um dado coeficiente a ser modificado e $\mathrm{sgn}(.)$ é a função *signum*.

Essa forma de modificar os coeficientes foi concebida em [6] empregando argumentos estatísticos. Contudo, também pode-se chegar à mesma função usando uma abordagem variacional. Mais especificamente, a regra *Soft* é o funcional que minimiza o erro quadrático penalizado pela norma L_1 do resultado [7, p.451], isto é,

$$\sum_{k=0}^{K-1} \left\{ \left(S_m[k] - \hat{S}_m[k]\right)^2 + 2t\left|\hat{S}_m[k]\right| \right\}. \tag{3}$$

Ou melhor, escolhida uma dada relação de compromisso igual a $2t$, em cada bloco, os coeficientes $\hat{S}_m[k]$, $k = 0, 1, \ldots, K-1$, que minimizam (3) são fornecidos pela função Soft $\hat{S}_m[k] = g_t^{(S)}(Y_m[k])$.

Como as funções $g(.)$ aqui consideradas modificam individualmente os coeficientes, a minimização de (3) pode ser feita de maneira unidimensional. Ou seja, para um dado valor de coeficiente X e relação de compromisso $2t$ fixados, o valor G que minimiza

$$Q_1(G) = (X-G)^2 + 2t|G|, \tag{4}$$

é a conhecida curva *Soft* com limiar t, dada por (2). Pode-se ver o formato desta curva na Fig. 1.a, obtida por minimização numérica de (4) para limiar $t = 0.2$ e que, igualmente, poderia ter sido obtida diretamente de (2).

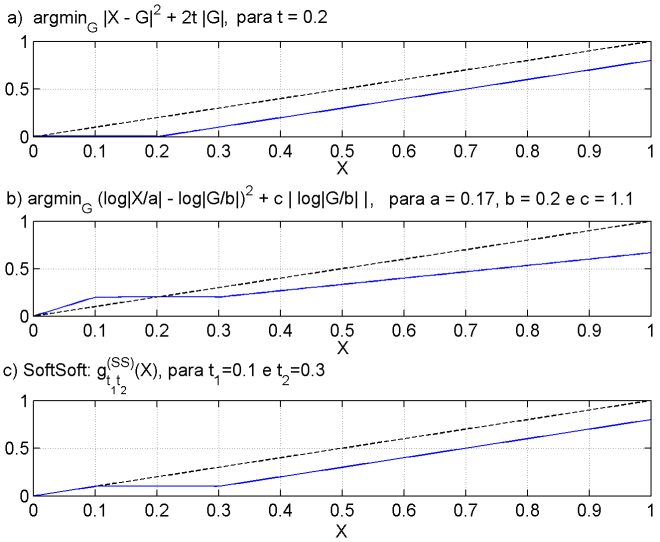

Fig. 1. a) Função *Soft* para $t = 0.2$; b) Função *LogSoft* para $a = 0.13$, $b = 0.2$ e $c = 1.1$; c) Função *SoftSoft* para $t_1 = 0.1$ e $t_2 = 0.3$.

B. Uma Função Especializada para Voz

Tendo como motivação o fato de que a percepção de intensidade sonora proporcionada pelo sistema auditivo humano é mais bem descrita pelo logaritmo da potência do sinal[1], propõe-se investigar o valor $\mathcal{G}$ que minimiza a seguinte relação de compromisso

$$Q_2(\mathcal{G}) = \left(\log\left|\frac{\mathcal{X}}{a}\right| - \log\left|\frac{\mathcal{G}}{b}\right|\right)^2 + c\left|\log\left|\frac{\mathcal{G}}{b}\right|\right| \tag{5}$$

que depende do conjunto de parâmetros (a, b, c). Nesta expressão, substituindo-se $\log\left|\frac{\mathcal{X}}{a}\right| \triangleq X$, $\log\left|\frac{\mathcal{G}}{b}\right| \triangleq G$ e $2t \triangleq c$, obtém-se justamente (4), cuja solução é dada pela função *Soft*, ou seja,

$$\log\left|\frac{\mathcal{G}}{b}\right| = g_t^{(S)}\left(\log\left|\frac{\mathcal{X}}{a}\right|\right),$$

que, usando a definição da curva *Soft*, (2), fornece

$$\log\left|\frac{\mathcal{G}}{b}\right| = \begin{cases} \log\left|\frac{\mathcal{X}}{a}\right| - t, \text{ para } \log\left|\frac{\mathcal{X}}{a}\right| > t \\ 0, \text{ para } -t \le \log\left|\frac{\mathcal{X}}{a}\right| \le t \\ \log\left|\frac{\mathcal{X}}{a}\right| + t, \text{ para } \log\left|\frac{\mathcal{X}}{a}\right| < -t \end{cases}.$$

Esta, por sua vez, pode ser reescrita, fornecendo uma regra especializada para modificação de coeficientes de voz:

$$g_{t,a,b}^{(LS)}(\mathcal{X}) = \begin{cases} b\frac{|\mathcal{X}|}{a}e^{-t}\,\mathrm{sgn}(\mathcal{X}), \text{ para } |\mathcal{X}| > ae^{t} \\ b\,\mathrm{sgn}(\mathcal{X}), \text{ para } ae^{-t} \le |\mathcal{X}| \le ae^{t} \\ b\frac{|\mathcal{X}|}{a}e^{t}\,\mathrm{sgn}(\mathcal{X}), \text{ para } |\mathcal{X}| < ae^{-t} \end{cases}. \tag{6}$$

Chamou-se a função $g_{t,a,b}^{(LS)}(\cdot)$ de *LogSoft* pois ela foi obtida da solução *Soft* usando logaritmos naturais. Apenas a título de verificação, a Fig. 1.b apresenta a curva obtida por

[1]Deve-se comentar que, talvez, um possível aperfeiçoamento possa ser conseguido considerando modelos mais elaborados para a percepção auditiva.

minimização numérica de (5) com parâmetros $a = 0.13$, $b = 0.2$ e $c = 1.1$. Conforme esperado, esta curva é a prevista por (6).

Deve-se destacar que o logaritmo enfatiza coeficientes de pequena magnitude, incluindo-os na composição da saída. Isto não ocorre na curva *Soft*, já que coeficientes abaixo do limiar são eliminados (cf. Fig. 1.a). Já a curva *LogSoft* permite que estes coeficientes passem para a saída (cf. Fig. 1.b). Como a distribuição dos coeficientes de voz tem uma grande concentração próximo da origem, isto é, pode ser aproximada por uma distribuição laplaciana [8], é possível melhorar o resultado se esses coeficientes forem preservados. Tal característica também é observada em outra regra de *thresholding* (a *SoftSoft*) que pode ser usada como aproximação da *LogSoft* com a vantagem de necessitar o ajuste de apenas dois limiares.

C. A Função SoftSoft Especializada para Voz

Esta função, introduzida em [8], recebe o nome *SoftSoft* pois possui duas transições suaves e, ainda, pode ser escrita combinando duas funções *Soft*:

$$g_{t_1,t_2}^{(SS)}(X) = g_{t_2}^{(S)}(X) + \left(X - g_{t_1}^{(S)}(X)\right), \qquad (7)$$

sendo os parâmetros: um limiar inferior (t_1) e um superior (t_2). Observe-se que a *SoftSoft* depende de apenas dois parâmetros, ao invés de três como no caso da *LogSoft*.

Normalmente, os valores de limiar que minimizam o erro são da mesma ordem de grandeza que os coeficientes de pequena magnitude a serem preservados. Ou melhor, os limiares possuem pequena magnitude, por isto, pode-se supor que as curvas *LogSoft* e *SoftSoft* otimizadas para mínimo erro tenham formatos semelhantes. Por exemplo, a Fig. 1.c apresenta uma curva *SoftSoft* e sugere que é possível ajustar os parâmetros da curva desta figura para que se torne semelhante à curva *LogSoft* da Fig. 1.b.

A seguir, passa-se a considerar o uso da proposta *SoftSoft*, em comparação com o método *Soft*, para a redução de ruído em sinais de voz. Como a *SoftSoft* se assemelha à curva que minimiza de (5), mesmo que o objetivo seja minimizar o erro médio quadrático, espera-se que a ela também consiga reduzir a distância quadrática entre os logaritmos, proporcionando menor distorção (distância) log-espectral (LSD)

A seguir, são apresentados resultados de simulação computacional para ambos os métodos, sendo os parâmetros das funções ajustados empiricamente de maneira a minimizar o erro (MSE). O cálculo deste erro e dos limiares ótimos é possível pois se supõe disponível o sinal sem ruído.

IV. Simulações Computacionais

A. Arquivos de Voz e Medidas Usadas para Comparar Desempenho

Foram escolhidas duas gravações [9] de frases foneticamente balanceadas, isto é, cujos fonemas aparecem com a mesma proporção de uma conversação usual, ambas com duração de cerca de 2,3 segundos:

- **Frase 1:** *"clear pronunciation is appreciated"*; voz feminina; com $M = 142$ blocos de $K = 128$ amostras;
- **Frase 2:** *"oak is strong and also gives shade"*; voz masculina; com $M = 144$ blocos de $K = 128$ amostras.

Em seguida, os sinais tiveram a taxa de amostragem convertida para 8000 amostras/s que é a taxa empregada nas simulações.[2]

Como o nível de ruído das gravações é muito baixo, pode-se considerar estes sinais "limpos" ou "originais", sendo ambos denotados por $s(n)$. Já os respectivos sinais ruidosos são obtidos acrescentando-se *AWGN* com nível σ_w para produzir uma relação sinal-ruído calculada a partir de

$$\mathrm{SNR} = 10\log_{10}\left(\frac{\frac{1}{N}\sum_{n=0}^{N-1}|s(n)|^2}{\mathrm{MSE}}\right), \qquad (8)$$

sendo o $\mathrm{MSE} \triangleq \frac{1}{N}\sum_{n=0}^{N-1}|s(n) - y(n)|^2$. O sinal ruidoso $y(n)$ é empregado como entrada para o método de redução de ruído, sendo a saída $\hat{s}(n)$ o resultado do processamento. A SNR da saída também é calculada pela mesma expressão, empregando-se $\hat{s}(n)$ no lugar de $y(n)$. Deve-se observar que, fixado o sinal $s(n)$, uma redução de 1 dB no MSE corresponde a um acréscimo de mesmo valor na SNR, podendo-se ter como objetivo maximizar esta ou minimizar aquele.

Por ser uma medida global do nível de ruído, a SNR não reflete bem a qualidade do sinal de voz, sendo esta mais bem representada pelo valor médio dado pela relação sinal-ruído segmentada

$$\mathrm{SegSNR} = \frac{1}{M}\sum_{m=0}^{M-1}\mathrm{SNR}_m\,, \qquad (9)$$

em que SNR_m é a relação sinal-ruído do m-ésimo bloco. Esta medida pode ser calculada para a entrada e para a saída proporcionando uma melhor comparação da redução média de ruído nos blocos. Caso, para algum bloco, ocorra logaritmo de zero, tal problema pode ser contornado limitando-se os valores de SNR_m à faixa dinâmica de 40 a -40 dB.

Considerando-se aspectos perceptuais do sistema auditivo humano, o nível de ruído face ao sinal é mais bem avaliado pelo logaritmo das potências, como na estimativa da distorção log-espectral do sinal de entrada:

$$\mathrm{LSD} = \frac{1}{M}\sum_{m=0}^{M-1} D_m, \qquad (10)$$

$$D_m = \sqrt{\frac{1}{\frac{K}{2}+1}\sum_{k=0}^{K/2}\left(10\log_{10}\frac{\mathsf{S}_m[k]}{\mathsf{Y}_m[k]}\right)^2},$$

em que $\mathsf{S}_m[k] = \max\left\{|\{\mathrm{FFT}(s_m)\}_k|^2,\, 10^{-10}\mathsf{S}_{\max}^2\right\}$ é a potência dentro de uma faixa dinâmica de 100 dB e $\mathsf{S}_{\max} = \max_{k,m}\{|\{\mathrm{FFT}(s_m)\}_k|\}$ é o coeficiente FFT de maior magnitude, tendo-se empregado FFT de comprimento $K = 128$ pontos. De maneira análoga, $\mathsf{Y}_m[k] = \max\left\{|\{\mathrm{FFT}(y_m)\}_k|^2,\, 10^{-10}\mathsf{Y}_{\max}^2\right\}$, com FFT de mesmo comprimento e $\mathsf{Y}_{\max} = \max_{k,m}\{|\{\mathrm{FFT}(y_m)\}_k|\}$.

Observe-se que, de igual maneira, essa distorção também pode ser calculada para a saída, bastando empregar $\hat{\mathsf{S}}_m[k]$ no lugar de $\mathsf{Y}_m[k]$.

[2] Isto foi feito para manter as mesmas condições experimentais de outros trabalhos, [5] [8] [10], possibilitando a comparação de resultados.

B. Determinação dos Valores dos Limiares

O MSE do resultado $\hat{s}(n)$ pode ser calculado a partir do sinal limpo $s(n)$ e do sinal ruidoso $y(n) = s(n) + w(n)$, que é obtido pela adição de uma realização $w(n)$ de *AWGN*, sendo o nível de ruído σ_w ajustado para proporcionar uma SNR de entrada desejada. Por exemplo, para sinal ruidoso de entrada com SNR $= 3$ dB, os valores de MSE podem ser calculados em função dos limiares t_1 e t_2 da função *SoftSoft*. A Fig. 2 exibe as curvas de nível assim obtidas. Nesta figura, deve-se observar que, para tornar os valores dos limiares independentes da amplificação do sinal, empregam-se limiar inferior e superior normalizados, respectivamente, $\bar{t}_1 \triangleq t_1/t_2$ e $\bar{t}_2 \triangleq t_2/Y_{\max}$, sendo $Y_{\max}$ a maior magnitude de coeficiente, $Y_{\max} = \max_{k,m} |Y_m[k]|$.

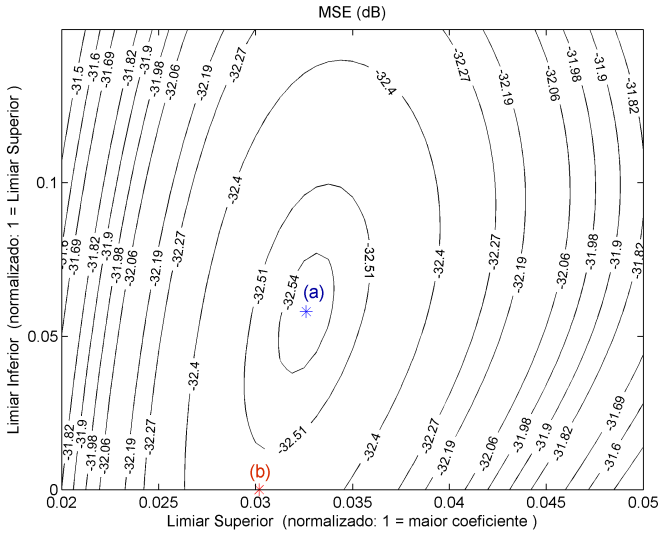

Fig. 2. Curvas de nível da superfície do MSE em função do limiar inferior ($\bar{t}_1$) e superior ($\bar{t}_2$), para Frase 1 com $SNR = 3$dB.

Ainda na Fig. 2, percebe-se que há um ponto de mínimo (a) que é alcançado empregando a função *SoftSoft* com $\bar{t}_1$ =0.059, $\bar{t}_2$ =0.033. Já a função *Soft*, que corresponde à *SoftSoft* com limiar inferior nulo, atinge o mínimo MSE em (b), ou seja, $\bar{t}_1 = 0$ e $\bar{t}_2$=0.031 que é igual ao parâmetro $\bar{t}$ da *Soft*.

Utilizando o método *SoftSoft,* a determinação dos pontos de mínimo MSE foi refeita para diversos níveis de ruído, SNR entre -3 e 20 dB, sendo as posições destes pontos expostas na Fig. 3. Novamente, deve-se alertar quanto à normalização dos limiares que, na Fig. 3.a, foram ambos normalizados em relação ao coeficiente de maior magnitude e, na Fig. 3.b, segundo a normalização anteriormente introduzida e denotada por $\bar{t}_1$ e $\bar{t}_2$. De maneira semelhante, também foram encontrados os pontos de mínimo MSE para o caso *Soft* e determinados os valores ótimos de limiar $\bar{t}$, normalizados em relação ao coeficiente de maior magnitude. Os resultados experimentais são apresentados e discutidos a seguir.

C. Comparação Experimental: Soft × SoftSoft

As Tabelas I e II apresentam os valores de SNR, SegSNR e LSD para o sinal ruidoso de entrada e, em comparação,

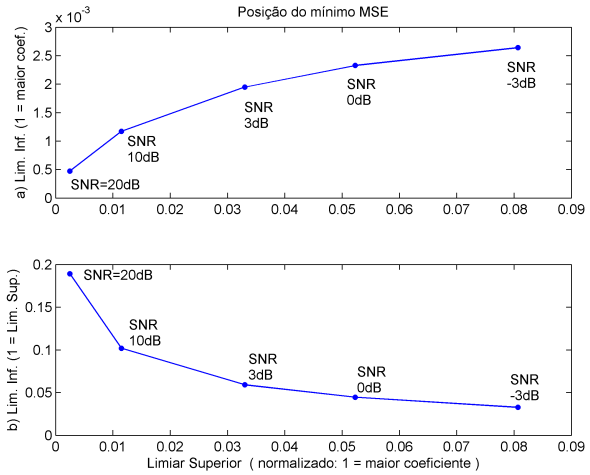

Fig. 3. Valores de limiar da função *SoftSoft* que minimizam o MSE para $SNR = 20$, 10, 3, 0 e -3dB, com **a)** os dois limiares normalizados em relação ao coeficiente de maior magnitude e **b)** limiares $\bar{t}_1$ e $\bar{t}_2$.

os respectivos resultados proporcionados pelos métodos *Soft* e *SoftSoft* empregando limiares ótimos, isto é, que produzem mínimo erro (MSE). Na primeira tabela, empregou-se a Frase 1 e, na segunda, a Frase 2.

Como se pode constatar nessas tabelas, o método *SoftSoft* tem uma vantagem da ordem de 0.05 dB evidenciada pelo maior valor de SNR. Ou seja, o erro (MSE) realmente é menor do que aquele proporcionado pelo método *Soft* convencional. Conforme comentado em [8] isto ocorre porque a introdução de coeficientes de pequena magnitude proporcionada pelo limiar inferior possibilita um ganho de sinal maior do que a quantidade de ruído que é acrescentada, resultando maior SNR e, consequentemente, menor MSE.

É importante comentar que, na prática, um ganho de SNR desta ordem não justificaria o emprego de dois limiares, como indicado pela SegSNR, medida mais adequada da potência de sinal face ao ruído, que se mostrou equivalente nos dois métodos. Contudo, a LSD que é uma medida mais apropriada da distorção espectral sempre é melhor no caso *SoftSoft*, justificando o esforço de se utilizar dois limiares.

Como os limiares foram ajustados para minimizar o erro (MSE) de todo o sinal, pode-se perguntar se o ajuste de

TABELA I

COMPARAÇÃO ENTRE OS MÉTODOS *Soft* E *SoftSoft* PARA "Sinal Ruidoso de entrada" COM DIVERSAS SNR. FRASE 1 COM COMPRIMENTO: $N = M \times K = 142 \times 128$. MÉDIA DE 1000 REALIZAÇÕES DO RUÍDO.

	SNR	SegSNR	LSD
Sinal Ruidoso de entrada	10.00	-0.33	14.5
Soft, $\bar{t}$=0.0101	13.10	3.62	12.7
SoftSoft, $\bar{t}_1$=0.1018, $\bar{t}_2$=0.0116	13.15	3.68	10.5
Sinal Ruidoso de entrada	3.00	-7.30	19.5
Soft, $\bar{t}$=0.0308	8.16	-0.53	15.4
SoftSoft, $\bar{t}_1$=0.0591, $\bar{t}_2$=0.0333	8.23	-0.50	12.2
Sinal Ruidoso de entrada	0.00	-10.22	21.9
Soft, $\bar{t}$=0.0493	6.32	-1.91	16.5
SoftSoft, $\bar{t}_1$=0.0445 , $\bar{t}_2$=0.0523	6.39	-1.90	12.8

TABELA II

COMPARAÇÃO ENTRE OS MÉTODOS *Soft* E *SoftSoft* PARA "Sinal Ruidoso de Entrada" COM DIVERSAS SNR. FRASE 2 COM COMPRIMENTO: $N = M \times K = 144 \times 128$. MÉDIA DE 100 REALIZAÇÕES DO RUÍDO (EXCETO EM c E d).

	SNR	SegSNR	LSD
Sinal Ruidoso de entrada	9.99	2.39	14.5
a) Soft, $\bar{t}=0.0192$	13.05	6.01	12.6
b) SoftSoft, $\bar{t}_1=0.0950$, $\bar{t}_2=0.0216$	13.09	6.05	10.7
c) Soft, $\bar{t}$ ótimos nos segmentos	13.5	7.5	12.5
d) SoftSoft, $\bar{t}_1$ e $\bar{t}_2$ ótimos nos segm.	13.6	7.5	9.4
Sinal Ruidoso de entrada	2.99	-4.61	19.5
e) Soft, $\bar{t}=0.0553$	7.81	1.22	15.4
f) SoftSoft, $\bar{t}_1=0.0550$, $\bar{t}_2=0.0592$	7.86	1.26	12.6
Sinal Ruidoso de entrada	-0.01	-7.56	21.9
g) Soft, $\bar{t}=0.0866$	5.77	-0.51	16.5
h) SoftSoft, $\bar{t}_1=0.0450$, $\bar{t}_2=0.0918$	5.82	-0.48	13.2

limiares em segmentos menores poderia aperfeiçoar o resultado, fornecendo uma melhor aproximação deste com o sinal original. Na Fig. 4.a, dividiu-se o sinal em segmentos de $128 \times 6 = 768$ amostras e, na Tabela II, em (c) e (d), são comparados os dois métodos, desta vez, usando limiares ajustados para mínimo MSE em cada um dos segmentos. Como se pode constar nessa tabela, ambos têm igual melhora em termos de relação sinal-ruído, no entanto, o *SoftSoft* consegue reduzir ainda mais a distorção. Provavelmente, isto ocorre pois os segmentos acompanham de maneira aproximada os fonemas da frase (cf. Fig. 4.b), possibilitando uma melhor adaptação às suas características espectrais.

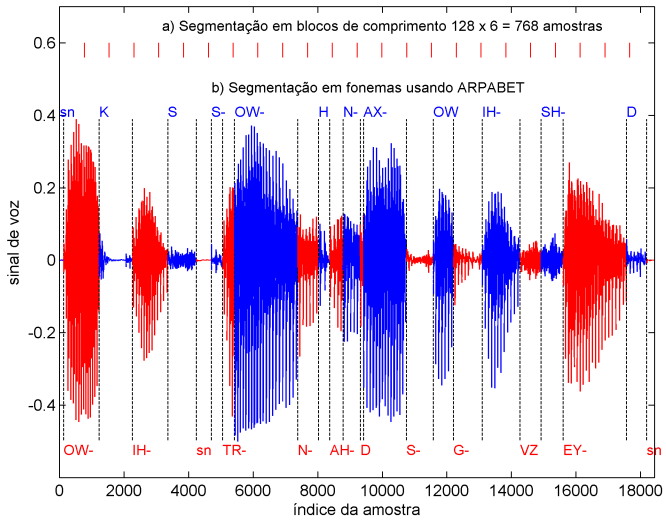

Fig. 4. **a)** Segmentação em trechos de comprimento constante; **b)** Sinal de voz da Frase 2 e segmentação em fonemas usando ARPABET [9]: "sn OW-K IH-S sn S-TR-OW-N-H AH–N-D AX-S-OW G-IH-VZ SH-EY-D sn", sendo denotado por "sn", um segmento de silêncio ou com baixo ruído de fundo.

Como última avaliação dos métodos, um dado ouvinte, geralmente, percebe menor distorção e menos presença de ruídos espúrios (como "ruído musical") no *SoftSoft*. Além do mais, o ruído de fundo apresenta maior semelhança com ruído branco, possivelmente, por causa da manutenção de um grande número de coeficientes de pequena magnitude. Cabe observar que a introdução de mais coeficientes não é inconveniente uma vez que a finalidade dos métodos aqui considerados é realizar redução de ruído.

V. CONCLUSÕES E PERSPECTIVAS

Neste trabalho, estudou-se uma abordagem variacional para obter funções especializadas na modificação de coeficientes de voz com o objetivo de minimizar o erro (MSE). Em [10], fez-se uma investigação semelhante, exceto que foi considerado o objetivo de minimizar a distorção (LSD). É importante comentar que, mesmo quando o objetivo não é minimizar a distorção, em geral, obtém-se um resultado com LSD menor quando se emprega uma função especializada, a exemplo da *SoftSoft* aqui avaliada.

Na prática, os valores ótimos de limiar não são conhecidos e devem ser estimados, sendo disponíveis métodos para estimar os limiares que minimizam o MSE, como os estudados em [8]. Os resultados aqui obtidos sugerem que mesmo que não se disponha de valores muito precisos para os limiares, ainda assim o método *SoftSoft* deve proporcionar menor distorção e, portanto, melhor qualidade avaliada por um ouvinte, sendo, portanto, um método mais indicado do que o uso de *Soft-thresholding* convencional.

Finalmente, deve-se acrescentar que métodos espectrais mais elaborados podem se beneficiar com o emprego de *SoftSoft*. Por exemplo, o uso de *SoftSoft* no método que faz ajuste dos limiares em segmentos (Tabela I.d) possibilitou aprimorar um método que pode ser usado para compensar variações lentas na potência do ruído, caso o mesmo não seja estacionário. Além do mais, acredita-se que outros trabalhos que lançam mão da regra *Soft* possam ser aperfeiçoados pelo uso da *SoftSoft* ou de uma outra função baseada na *LogSoft*.

REFERÊNCIAS

[1] M. A. Q. Duarte, J. V. Filho e F. Villarreal, "Um Novo Método de Redução de Ruído em Sinais de Voz Baseado em Wavelets," XXI Simp. Bras. de Telecom. - SBT'04, 6pp., Set., 2004, Belém - PA.

[2] C. A. Medina S., J. A. Apolinário Jr. e A. Alcaim, "Modern Speech Enhancement Techniques in Text-Independent Speaker Verification", XX Simp. Bras. de Tel. - SBT'03, Out., 2003, Rio de Janeiro, RJ.

[3] L. A. da Silva e M. B. Joaquim, "Redução de Ruído em Sinais de Voz Usando Filtros de Kalman de Tempo e Freqüência Discretos Combinados com Subtração Espectral de Potência e/ou Wavelets", XXV Simp. Bras. de Tel. - SBrT 2007, Set., 2007, Recife, PE.

[4] S. V. Vasegui, "Advanced Signal Processing and Noise Reduction", *John Wiley & Sons*, 2nd edition, 2000.

[5] I. Antunes Jr., and P. M. S. Burt, "Speech Denoising by SoftSoft Thresholding", *Proc. ISIE, IEEE International Symposium on Industrial Electronics*, v.1., pp. 532-536, Montreal, 2006.

[6] D. L. Donoho, and I. M. Johnstone, "Ideal spatial adaptation via wavelet shrinkage", *Biometrika*, vol. 1, pp. 425-455, 1992.

[7] S. Mallat, "A Wavelet Tour of Signal Processing", *Academic Press*, 2nd edition, 1999.

[8] I. Antunes Jr., "Redução de ruído em sinais de voz usando curvas especializadas de modificação dos coeficientes da transformada em co-seno.", *Tese (Doutorado)*, Escola Politécnica da Universidade de São Paulo, Departamento de Telecomunicações e Controle, 2006.

[9] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallet, and N. L. Dahlgren, "The darpa timit acoustic-phonetic continuous speech corpus cdrom", Cdrom, NIST, Gaithersburg, MD, 1996.

[10] I. Antunes Jr., "Redução de Ruído em Sinais de Voz Usando Funções de Limiar SoftSoft para Minimizar a Distorção Log-Espectral", Congresso de Engenharia de Áudio da AES-Brasil, 2011, São Paulo - SP. *(Submetido em março de 2011.)*