REDUÇÃO DE RUÍDO EM SINAIS DE VOZ USANDO CRITÉRIOS PSICOACÚSTICOS

JOZUÉ VIEIRA FILHO

Universidade Estadual Paulista
(DEE/FEIS/UNESP)[1]
Av. Brasil Centro, 56, Ilha Solteira/SP

RESUMO

As diferentes aplicações de redução de ruído em telecomunicações exigem que qualquer processamento introduza o mínimo de distorção no sinal de voz original. Técnicas clássicas, como as baseadas na subtração espectral, apresentam bom desempenho na redução do ruído, mas introduzem distorções relevantes no sinal de processado. Visando minimizar essas distorções, existem diferentes propostas para técnicas de redução de ruído que usam critérios psicoacústicos. Neste trabalho, propõe-se um novo filtro baseado na subtração espectral, com modificações que consideram critérios psicoacústicos na estimação dos parâmetros do filtro. Avaliações objetivas mostram que o novo método apresenta ótimo desempenho na redução de ruído e introduz baixas distorções no sinal de voz.

1 - INTRODUÇÃO

A psicoacústica é o estudo da percepção dos sons pelo ser humano, visando buscar soluções para os problemas mais diversos relacionados à audição humana. Do ponto de vista de processamento de sinais de voz, os estudos da psicoacústica têm sido usado no melhoramento de técnicas de codificação, na avaliação objetiva de qualidade e na redução de ruído [1],[2],[3]. A maioria das aplicações explora dois fenômenos importantes: *bandas críticas e mascaramento*. Um dos trabalhos pioneiros nessa área foi publicado por Schroeder [1] em 1979. Nesse artigo, foi apresentado um método para medição de degradação em sinais codificados com base no princípio do mascaramento. Uma das principais contribuições do trabalho foi o cálculo da sonoridade, onde foi explorado um modelo de espalhamento do sinal acústico que chega à membrana basilar. Outros trabalhos importantes foram publicados [4], [5], [6]. Entretanto, nem todos apresentam bons resultados quando aplicados à redução do ruído. Por exemplo, nos trabalhos apresentados em [4] e [6], obtêm-se curvas de mascaramento através de métodos que usam uma redução de ruído auxiliar (normalmente a subtração espectral clássica) e na seqüência aplica-se a redução de ruído nas faixas onde a potência do ruído está acima do limiar de mascaramento. O objetivo, nesses casos, é minimizar as distorções no sinal processado. Porém, o uso de métodos auxiliares de redução de ruído significa a introdução de critérios não lineares para a determinação das freqüências do ruído com potência abaixo ou acima da potência do sinal de voz, provocando outras distorções no sinal processado.

Neste estudo, o objetivo foi obter um método de redução de ruído que permitisse um mínimo de degradação ao sinal processado. Usando como base os princípios da psicoacústica, particularmente o de banda crítica, foram realizadas modificações no método proposto em [7] que permitem uma nova proposta para redução de ruído em sinais de voz. Os resultados obtidos mostram que o novo método mantém a boa redução de ruído do método original e apresenta uma menor degradação do sinal processado.

2-ALGUNS CONCEITOS DE PSICOACÚSTICA

2.1 - Bandas Críticas

Um dos pontos mais importantes no modelamento do sistema auditivo é o estudo de como é feita a discriminação de freqüências, que é a base da psicoacústica. A membrana basilar desloca-se para cada componente de freqüência do som que chega ao ouvido, com espalhamento em torno de um ponto central. Isto faz com que duas componentes próximas e com amplitudes semelhantes possam ou não ser separadas. Por exemplo: sejam dois tons com freqüências F1 e F2 e amplitudes A1 e A2, respectivamente, que são disparados simultaneamente. Fixando-se F1 e variando-se F2 (abaixo ou acima de F1), ouvem-se “batidas” ou tons com freqüência F2-F1 ou F1-F2 e amplitudes variadas. De acordo com testes subjetivos [8], as “batidas” são percebidas quando F1 e F2 apresentam uma diferença máxima de 12,5 Hz. Quando essa diferença passa de 15 Hz, um som áspero é ouvido. Essa sensação de som áspero continua até que os dois tons possam ser discriminados

[1] *Trabalho desenvolvido com o apoio financeiro da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), processo 98/04257-4.*

corretamente pelo sistema auditivo. A diferença entre esses dois pontos é denominada de BANDA CRÍTICA. A largura da banda crítica aumenta com a freqüência e existem algumas fórmulas aproximadas. Por exemplo, de acordo com [1], a banda crítica para as freqüências de 200 Hz e 2000Hz são, respectivamente, 47,7 Hz e 240 Hz. Cada faixa de freqüência correspondente a uma banda crítica pode ser modelada como um filtro passa-faixa.

2.2 - Mascaramento

A definição de mascaramento pode ser relacionada ao conceito de bandas críticas, principalmente se for considerada uma análise a partir de tons puros. Como visto anteriormente, dois tons puros de mesma intensidade só são percebidos quando a diferença de freqüência entre os mesmos estiver acima do valor da banda crítica na faixa analisada. Essa percepção será tão mais difícil quanto maior for a diferença de intensidade desses tons. Quando o tom de menor intensidade não pode mais ser percebido, diz-se então que ele foi "mascarado". Monitorando-se todo o espectro auditivo para um determinado tom ou som, obtém-se então a curva ou limiar de mascaramento, que indicará que todo som abaixo desse limiar não será percebido. Considerando um tom puro, sabe-se que o limiar de mascaramento é mais significativo para freqüências acima do tom mascarador, como indicado na figura 1. Sabe-se também que a curva de mascaramento depende da intensidade do som mascarador. Nesse caso, quanto maior a intensidade, maior o efeito do mascaramento nas freqüências superiores. A figura 1 ilustra também o que ocorre para diferentes limiares de mascaramento. Observa-se que há um limiar onde o efeito do mascaramento é praticamente o mesmo para as freqüências abaixo e acima da freqüência do tom mascarador.

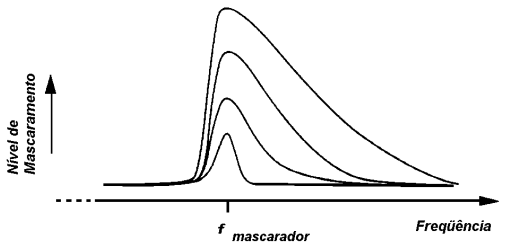

Figura 1 – Variação do nível de mascaramento

3 - FILTRO DE REDUÇÃO DE RUÍDO BASEADO NA SUBTRAÇÃO ESPECTRAL MODIFICADA

A subtração espectral original proposta por Boll [9] não apresenta bons resultados [10], [7] (redução de ruído média e baixa qualidade auditiva), mas tem como grande vantagem a simplicidade (com relação a carga computacional). Os estudos realizados por [7] resultaram em um método mais eficiente e também simples do ponto de vista computacional, resumido aqui nos parágrafos seguintes.

Suponha que um sinal de voz puro $v(t)$ foi degradado por um ruído aditivo $r(t)$ e resultou no sinal ruidoso $y(t)$. Considerando que os sinais de voz e ruído são dois processos aleatórios independentes e baseando-se na propriedade da linearidade da Transformada de Fourier, tem-se que o espectro de potência do sinal de voz puro, $Pv(\omega)$, pode ser obtido a partir das potências dos sinais ruidoso, $Py(\omega)$, e ruído, $Pr(\omega)$, como segue:

$$Pv(\omega) = Py(\omega) - Pr(\omega) \quad (1)$$

Considerando que a fase do sinal ruidoso pode ser usada na estimação do sinal de voz puro [11], define-se a resposta em amplitude do filtro redutor de ruído como sendo:

$$|H(\omega)| = [\, Pv(\omega)/Py(\omega) \,]^{1/2} \quad (2)$$

Das equações 1 e 2 obtém-se

$$|H(\omega)| = \{[\, Py(\omega) - Pr(\omega)]/Py(\omega)\}^{1/2} \quad (3)$$

Definindo a *relação sinal ruído a posteriori* como sendo

$$SNR\_post(\omega) = Py(\omega)/Pr(\omega) \quad (4)$$

tem-se que a resposta em amplitude do filtro redutor de ruído é dada por

$$|H(\omega)| = \{[SNR\_post(\omega) - 1]/SNR\_post(\omega)\}^{1/2} \quad (5)$$

A equação 5 é a base da subtração espectral original. Nos estudos realizados por [7] foi introduzido um novo termo denominado *relação sinal ruído a priori*, dado por:

$$SNR\_prio(\omega) = Pv(\omega)/Pr(\omega) \quad (6)$$

O termo definido na equação 6 permite eliminar a subtração presente na equação 5, pois de acordo com as equações 1, 4 e 6, tem-se:

$$SNR\_post(\omega) = 1 + SNR\_prio(\omega) \quad (7)$$

Das equações 5 e 7 obtém-se finalmente a resposta em amplitude para o filtro redutor de ruído, que é dada por:

$$|H(\omega)| = \{SNR\_prio(\omega)/[1 + SNR\_prio(\omega)\,]\}^{1/2} \quad (8)$$

O filtro dado pela equação 8 não apresenta subtrações. Isto elimina a possibilidade de obtenção de potências negativas, fisicamente indefinidas e que exigem o uso de recursos para obtenção de um valor positivo. No caso, obtém-se um filtro que permite a redução do ruído sem introduzir o conhecido "ruído musical" [7] no sinal processado. Deve-se lembrar que o ruído musical é resultado das modificações aleatórias do espectro de potência do sinal processado, pois a obtenção de potências negativas está associada às fases dos sinais de voz e ruído, que são processos aleatórios.

O desempenho do filtro proposto na equação 8 depende diretamente da estimação da SNR_prio. Nos estudos

desenvolvidos em [7] e [12] foi utilizado um método denominado de "decisão dirigida", que usa uma parcela da relação sinal ruído a posteriori (SNR_post). Considerando a definição de SNR_prio dada na equação 6, nota-se que não há muitas opções para a estimação da SNR_prio, pois é necessário que se tenha disponível a potência do sinal de voz puro. O uso desse método de estimação para a SNR_prio permite obter uma boa redução de ruído, mas há distorções significativas no sinal de voz processado. Assim, neste trabalho os estudos estiveram concentrados na busca de um novo método de implementação do filtro dado na equação 8 que minimizasse as distorções no sinal processado. Com base nos conceitos de psicoacústica, modificou-se a estimação das potências dos sinais usados no cálculo das relações sinal/ruído, conseguindo-se excelentes resultados.

4 - ESTIMAÇÃO DAS POTÊNCIAS DOS SINAIS DE VOZ E RUÍDO USANDO CRITÉRIOS PSICOACÚSTICOS

No item 2 foram apresentadas duas definições importantes derivadas do funcionamento do sistema auditivo: a de banda crítica e a de mascaramento. Essas definições sugerem que uma redução de ruído adequada não deve ser baseada em um filtro que apresente grandes variações de ganho em freqüências próximas (muito menor do que a largura de uma banda crítica). Primeiro, de acordo com a definição de mascaramento, um determinado som com freqüência f_1 e potência P_1 será mascarado por um outro som com freqüência f_2 e potência P_2 quando P_2 for maior do que P_1 e f_1 for próxima de f_2. Segundo, pela definição de banda crítica, os sons com freqüência f_1 e f_2 só poderão ser discriminados pelo sistema auditivo a partir de uma certa "distância" Δ_f.
Com essas considerações, propõe-se o uso de um filtro com ganhos próximos dentro de uma determinada banda crítica, que neste trabalho foi alcançado modificando-se a estimação da SNR_prio. Essa estimação passa por uma adequação inicial dos espectros do sinal de voz e do ruído, que também devem ser estimados, aos critérios psicoacústicos e, de acordo com os parâmetros do filtro apresentados no item III, é realizada como segue:

- Os espectros de voz, $Pv(\omega)$, e do ruído, $Pr(\omega)$, são obtidos com base em uma FFT de 512 pontos e uma janela de Hamming com duração de 32 milisegundos.

- Para que voz e ruído se adeqüem ao ouvido médio, faz-se uma filtragem passa-baixas com freqüência de corte em 5 kHz. Para aplicações envolvendo telefonia, que é o caso, a freqüência de corte é de 4 kHz e essa filtragem torna-se desnecessária.

- Após a simulação de passagem pelo ouvido médio, os sinais chegam ao ouvido interno, onde é ativado o sistema de banda crítica. Neste caso são calculadas 18 bandas para cobrir a faixa de 4 kHz. Na implementação, utilizou-se uma fórmula aproximada proposta por [1], dada por:

$$f = 650 \operatorname{senh}(x/7),$$

onde x é o número da banda crítica e f a freqüência superior da banda. Essa aproximação é válida para valores de $x < 20$. Essa equação é usada para obtenção dos espectros em termos de banda crítica: $Pv(x)$ para voz e $Pr(x)$ para o ruído.

- O sinal, após chegar ao ouvido interno, está pronto para a excitação através dos nervos auditivos. Essa excitação é uma função resultante da convolução entre o espectro de potência do sinal e um modelo de espalhamento da membrana basilar [1]. De acordo com [1], um modelo de espalhamento satisfatório, $B(x)$, é:

$$10.\log_{10}[B(x)] = 15.81 + 7.5\,(x + 0.474) - 17.5\,[1 + (x + 0.474)]^{1/2} \quad \text{dB}$$

- Os espectros do sinal de voz e ruído estão finalmente de acordo com o modelo auditivo. Considerando a implementação do filtro com base em uma FFT de N pontos, deve-se calcular os espectros finais de potência baseados no modelo de excitação. Inicialmente, tem-se que os modelos de excitação são:

$Qv(x) = Pv(x)*B(x)$ - excitação do sinal de voz

$Qr(x) = Pr(x)*B(x)$ - excitação do ruído

- Para viabilizar a implementação via FFT, propõe-se o uso de um espectro de potência baseado no modelo de excitação e com valores iguais dentro de cada banda crítica, como segue:

$$Pvx(\omega) = Qv(x),\ \textit{para}\ blx \le \omega < bhx\ e\ 1 \le x \le 18,\ \textit{para voz.}$$

$$Prx(\omega) = Qr(x),\ \textit{para}\ blx \le \omega < bhx\ e\ 1 \le x \le 18,\ \textit{para ruído.}$$

Os valores blx e bhx são os limites inferior e superior de cada banda crítica x. Os valores de ω variam de acordo com o número de pontos da FFT. A implementação prática é feita criando-se um vetor de (1 + N/2) pontos, onde N é o número de pontos da FFT. Para uma freqüência de amostragem ωs, os valores, em termos de índice do vetor, para os limites inferior e superior da banda crítica são dados por:

$$n = \textbf{int}\,[N \cdot (\omega/\omega s)],$$

onde **int** indica a parte inteira do valor obtido entre colchetes e **n** é o índice do vetor FFT.

Por exemplo, considerando uma FFT de 512 pontos e uma freqüência de amostragem de 8 kHz, os limites de *n* para a *banda crítica* número *1* serão ZERO (0 Hz, inferior) e SEIS (≈ 95 Hz, superior).

5 - ESTIMAÇÃO DA SNR_PRIO

A estimação da SNR_prio depende da estimação das potências dos sinais de voz e ruído. Na maioria das aplicações o sinal ruidoso é obtido em um único canal, tornando-se necessária a separação entre trechos de voz e ruído. Essa tarefa não é fácil e muitos algoritmos já foram propostos. Neste trabalho foi utilizado um algoritmo baseado na SNR, que funciona bem para sinais com SNR média maior do que 14 dB.

No desenvolvimento apresentado no item IV supôs-se que as potências já estavam disponíveis e que o processamento era baseado na STFT (Short Time Fourier Transform). De forma análoga, no procedimento apresentado a seguir supõe-se o uso da STFT e define-se $\mathbf{w_i}$ como sendo a i-ésima janela de sinal analisada/processada. Define-se também $Pvx(\omega)_A$ como sendo a excitação do sinal de voz estimado na janela anterior, adequada ao índice correto do vetor da FFT. Assim, a estimação da SNR_prio é dada por:

$$SNR\_prio\_x(\omega) = (\alpha)\frac{P^*vx(\omega)_A}{P^*rx(\omega)} + (1-\alpha)abs\left[\frac{P^*yx(\omega)_A}{P^*rx(\omega)} - 1\right] \quad (9)$$

O símbolo ***** na equação anterior indica que as potências não estão em dB, ao contrário do que está definido no modelo de excitação.

O fator α permite que se calcule a SNR_prio não somente com base na informação da janela anterior, usando a potência do sinal de voz estimado, mas também com base na janela atual. Essa contribuição é feita de acordo com as definições de SNR_prio e SNR_post dadas nas equações 4, 6 e 7. Observa-se na equação anterior o uso do operador **abs,** que elimina a possibilidade de uma SNR_post negativa[1]. Para evitar a presença do ruído musical, a contribuição da SNR_post deve ser moderada. Baseado em testes subjetivos [12], [7], deve-se usar $\alpha > 0.9$.

6 - RESULTADOS OBTIDOS

Para avaliar o desempenho do novo método, foram feitos testes objetivos baseados em dois critérios: *nível de redução de ruído* e *distância cepstral*. O primeiro permite uma visualização direta da quantidade de ruído removida e o segundo permite uma avaliação das distorções no sinal de voz. Segundo [13], a distância cepstral é um método de avaliação objetiva com boa correlação com o sistema auditivo. Para comparações, são apresentados resultados obtidos com o método original proposto em [7]. Uma avaliação subjetiva informal também foi realizada.

[1] *Valores negativos na expressão (SNR_post-1) são resultantes dos mesmos problemas que levam a valores negativos na expressão dada na equação 1, determinantes para o aparecimento do ruído musical na subtração espectral clássica.*

6.1 - Nível de redução de ruído

Para avaliar o nível de redução de ruído, foram obtidas as médias de potência para os sinais ruidoso e processado, em cada janela, de acordo com a equação:

$$P_w = \frac{1}{N}\sum_{i=0}^{N-1} x^2(i) \quad (10)$$

onde **w** representa a janela processada e **N** é o número de amostras (no caso, igual ao número de pontos da FFT).

Na figura 2 são apresentadas as formas de onda do sinal puro, do sinal ruidoso e dos sinais processados pelos métodos original e modificado. A relação sinal/ruído (SNR) média do sinal ruidoso é de 15 dB.

Na figura 3 são apresentadas as curvas de potência média para os sinais puro, ruidoso e para os sinais processados pelo métodos original (proposto em [7]) e modificado (proposto neste trabalho). Nota-se claramente que os dois métodos apresentam praticamente o mesmo nível de redução de ruído, tanto nos trechos de silêncio como também nos trechos de voz.

6.2 - Distorções

Na figura 4 são apresentadas as curvas de distância cepstral com histogramas para os sinais processados com os dois métodos. Verifica-se que há uma redução considerável nas distorções, bem como uma melhor distribuição. Observando os intervalos de voz na figura 2, nota-se que o método original (figura 4-A) apresenta distorções mais significativas nos trechos de voz do que nos trechos de silêncio. Já com o método proposto (figura 4-B), as distorções estão melhor distribuídas nos diferentes trechos (voz e silêncio).

Para dar uma idéia das modificações no filtro, na figura 5 são apresentados os ganhos para um trecho de silêncio (figura 5-A) e outro de voz (figura 5-B). Observa-se que a curva de ganho do novo método é mais suave nos dois trechos, mas a mudança é ainda mais significativa no trecho de voz. Nota-se uma limitação do ganho do método original em –14 dB, que foi implementada de modo a minimizar as distorções [7].

Para finalizar as avaliações de distorções, bem como do nível de redução de ruído, foram feitos alguns testes subjetivos, sem uso de metodologia formal. Foram ouvidas 10 pessoas e todas optaram, de maneira quase imediata, pelo sinal processado com o método proposto. Também foram unânimes em afirmar que o sinal ruidoso era pior do que os outros sinais. Na realização desses testes, as pessoas tinham disponíveis todos os sinais e ouviram quantas vezes acharam necessário.

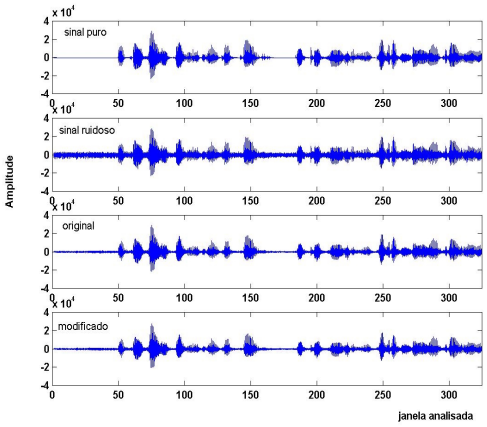

Figura 2 – Formas de onda

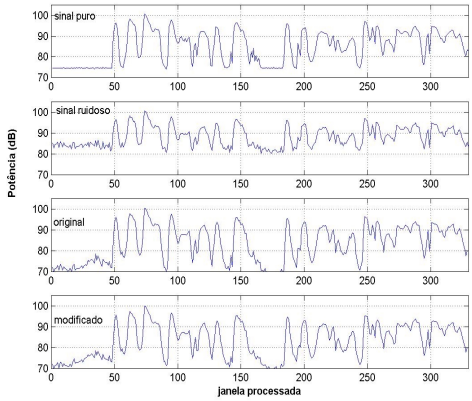

Figura 3 – Curvas de potência

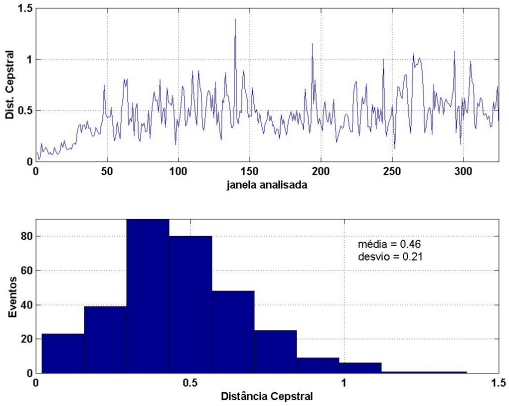

Figura 4 A– Distorções (método de referência)

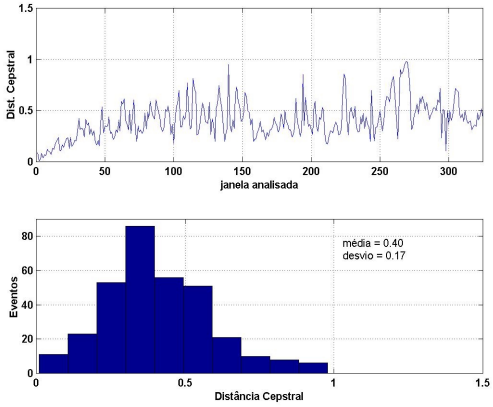

Figura 4 B– Distorções (método proposto)

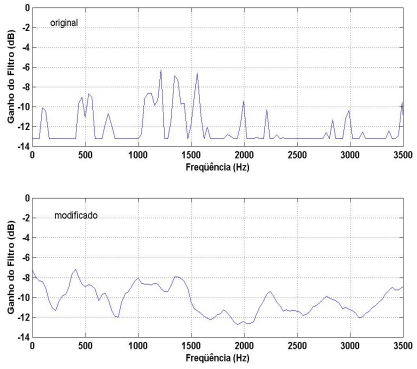

Figura 5A – Curvas de ganho (trecho de silêncio)

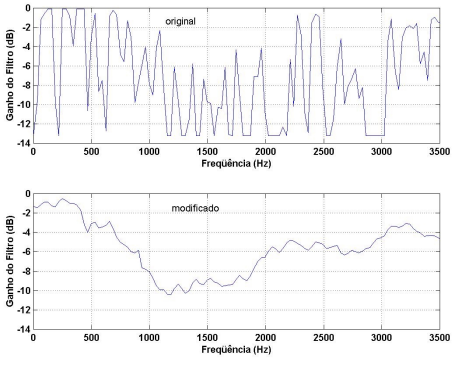

Figura 5B – Curvas de ganho (trecho de voz)

7 - CONCLUSÕES

Neste trabalho apresentou-se uma nova proposta para redução de ruído em sinais de voz. O método usa critérios psicoacústicos para estimar parâmetros do filtro redutor de ruído. Os resultados foram avaliados através de medidas objetivas e testes subjetivos informais e comparados com os obtidos com outros métodos (particularmente, com o método proposto em [7]). A avaliação final mostra que o novo método permite obter excelente redução de ruído, sem introduzir distorções significativas no sinal de voz processado.

AGRADECIMENTOS

O autor agradece ao pesquisador Edson José Nagle (Fundação CPqD), pela valiosa revisão, e à FAPESP, pelo apoio financeiro.

REFERÊNCIAS

[1]- SCHROEDER, M.R, ATAL, B.S and HALL, J.,L, *"Optimizing Digital Speech Coders by Exploiting Masking properties of the Human Ear*", in Journal of Acoustical Soc. of America, 1979, pp. 1647-1652.

[2]- WANG, S. et al. *"An Objective Measure for Predicting Subjective Quality of Speech Coders*", IEEE Journal on Selected Areas in Communications, Vol.10 No.5, pp.819-829, June/1992

[3]- CHENG, Y.M & O'SHAUGHNESSY, *"Speech Enhancement Based Conceptually on Auditory Evidence*", in Trans. on Acoust. Sigal Processign, vol.39, No.9, sept-1991.

[4] - TSOUKALAS, D. et al., "*Speech Enhancement Using Psychoacoustic Criteria*", in Proc. IEEE ICASSP, 1993, pp. II.359-II.362.

[5]- JOHNSTON, J.D. "*Transform Coding of Audio Signals Using Perceptual Noise Criteria*", IEEE Journal on Selec. Areas in Comm., Vol.6. No. 2, February, 1988.

[6]- AZIRANTI, A. A. et al*." Optimizing Speech Enhancement by Exploiting Masking Properties of the Human Ear*", IEEE Proc. ICASSP, pp.800-803, 1995.

[7]- SCALART, P., & VIEIRA FILHO, J*., "Speech Enhancement Based on a Priori Signal-to-Noise Ratio Estimation*", in: 1996 IEEE Int. Conf. on Acoust., Speech and Signal Processing, pp: 629-632, Atlanta-USA, May, 1996.

[8]- HOWARD, D.M. & ANGUS, J. "*Acoustics and Psychoacoustics*", Music Technology Series of FRANCIS RUMSEY editor, Great Britain, 1996.

[9]- BOLL, S.F. "*Suppression of Acoustic Noise in Speech using Spectral Subtraction*", in IEEE Trans. Acoustic, Speech, and Signal Processing (TASSP), Proc., vol. 29, pp113-120, April-1979.

[10]- VIEIRA FILHO, J., SCALART, P., e CHIQUITO, J.G. "*Redução de Ruído em Sinais de Voz: análise e avaliação das técnicas clássicas baseadas na subtração espectral a curto-termos*", in: Anais do XIII Simpósio Brasileiro de Telecomunicações (SBT), pp: 685-689, Águas de Lindóia-SP, Setembro de 1995.

[11]- FLANAGAN, J.L. "*Speech Analysis Synthesis and Perception"*, Ed. Springer-Verlag, 1972.

[12]- EPHRAIM, Y. and MALAH, D. "*Speech Enhancement Using Minimum Mean Square Error Short-Time Spectral Amplitude Estimator*", IEEE Trans. on Acoustics, Speech and Signal Processing, vol.32, NO. 6 - December-1984.

[13]- GRAY JR, A.H & MARKEL, J.D, "*Distance Measures for Speech Processing*", in IEEE Trans. on Acoust. Speech and Signal Processing, vol.24, No.5, october-1976.