

# Um Novo Método de Redução de Ruído em Sinais de Voz Baseado em Wavelets

Marco A. Q. Duarte<sup>1,2</sup>, Jozué V. Filho<sup>1</sup> e Francisco Villarreal<sup>3</sup>

**Resumo**—Neste trabalho propõe-se um novo método de redução de ruído em sinais de voz usando wavelets. O método é baseado no cálculo de um limiar nos instantes de silêncio, com aplicação em todo o sinal. A redução do ruído no domínio wavelet é feita comparando-se os coeficientes wavelets do sinal ruidoso com o limiar obtido nos instantes de silêncio, de modo a se obter um novo sinal. Diferentemente de outros métodos, o sinal resultante no domínio wavelet é função do próprio sinal ruidoso, não havendo anulação de coeficientes. Os resultados mostram que o novo método é eficaz na redução do ruído branco e colorido, minimiza o ruído residual característico das técnicas de redução de ruído baseadas em wavelets e não introduz distorções significativas no sinal processado.

**Palavras-Chave**—Transformada de wavelet, redução de ruído, processamento de voz.

**Abstract**—In this work it is proposed a new method for noisy speech enhancement using wavelets. The noise reduction in the wavelet domain is obtained comparing the wavelet coefficients of the noisy speech signal to a threshold estimated in the silence intervals. Differently of other wavelets based methods, in this new method the wavelet coefficient below the threshold is not set to zero, but it is replaced by a new value which is function of the noisy speech coefficient itself. The results show that the new method is efficient to reduce white and colored noise without introducing an annoying residual noise which is common in wavelets based techniques. Also, the results show good noise reduction and low distortions in the processed speech.

**Index Terms**—Wavelet transform, noise reduction, speech processing.

## I. INTRODUÇÃO

A redução de ruído é importante nas mais variadas aplicações que envolvem processamento de sinais de voz, como em sistemas de reconhecimento automático de fala, telefonia, sistemas de codificação, etc. Os métodos mais explorados nas últimas três décadas são baseados na Transformada de Fourier ([1], [2] e [3]) e muitos apresentam bons resultados. Entretanto, ainda existem aplicações para as quais as técnicas atuais não conseguem apresentar bons resultados.

Um exemplo típico é a aplicação em sistemas de codificação usados em comunicações sem fio, onde o nível de ruído pode degradar de maneira significativa o sinal transmitido. Os estudos com base na Transformada de Fourier ainda são importantes ([4], [5], etc.), mas existem muitos estudos que tentam explorar a Transformada de Wavelet Discreta (DWT) ([6], [7] e [8]). Se, do ponto de vista de codificação de sinais de voz com boa relação sinal/ruído (SNR), a Transformada de Wavelet é de fácil aplicação e gera resultados considerados bons, o mesmo não pode ser dito com relação às aplicações de redução de ruído quando a SNR é baixa. As metodologias mais relevantes encontradas na literatura usam como base o cálculo de um limiar no sinal ruidoso, com posterior eliminação de coeficientes wavelets, que é um tipo de codificação básico usando wavelets ([8], [9] e [10]). O problema é que o uso de um limiar resulta em ruídos residuais incômodos ao ouvido humano, apesar de muitos autores tentarem justificar seus métodos usando o nível de redução de ruído obtido. O ruído residual aparece quando o limiar é aplicado aos coeficientes, gerando novos coeficientes que podem ser nulos ou não. Do ponto de vista auditivo, o ruído residual obtido em sistemas baseados na DWT são parecidos com o “ruído musical”, que é um ruído residual característico dos métodos baseados na Transformada de Fourier ([1] e [2]), principalmente aqueles que usam o princípio da subtração espectral [3].

Neste trabalho propõe-se um novo método de redução de ruído usando a DWT, sendo que uma das principais contribuições é a redução do ruído residual sem aumento de distorção no sinal processado. As implementações foram feitas usando-se o software de simulação MATLAB e os resultados foram avaliados formalmente através de testes objetivos e informalmente através de testes subjetivos. Ao longo deste trabalho são apresentados os mecanismos básicos de redução de ruído com base na DWT, a descrição do método proposto e os resultados obtidos.

## II. TRANSFORMADA DE WAVELET: UMA INTERPRETAÇÃO DO PONTO DE VISTA DE PROCESSAMENTO DE SINAIS

A Transformada Wavelet aproxima uma função qualquer através de uma base de funções ortonormais que, diferentemente da Transformada de Fourier, não precisam ser funções com duração infinita (ao contrário das funções seno e cosseno). Sua interpretação ficou mais fácil após os estudos

<sup>1</sup>Departamento de Engenharia Elétrica da FEIS-UNESP-Ilha Solteira/SP/Brasil; <sup>2</sup>Universidade Estadual do Mato Grosso do Sul - UEMS Unidade de Cassilândia/MS/Brasil; <sup>3</sup>Departamento de Matemática da FEIS-UNESP-Ilha Solteira/SP/Brasil. E-mails: marco@dee.feis.unesp.br, jozue@dee.feis.unesp.br, villa@fqm.feis.unesp.br.

de Daubechies ([11], [12]), que propôs uma implementação através de filtros do tipo FIR (Resposta ao Impulso Finita).

A figura 1 apresenta um exemplo de Transformada Wavelet Discreta (DWT) com dois níveis. O sinal de entrada, denominado de  $s[n]$ , é passado por dois filtros: o primeiro representa um filtro passa-baixas, com resposta ao impulso  $h_{LP}[n]$ , e o segundo representa um filtro passa-alta, com resposta ao impulso  $g_{HP}[n]$ . Após a filtragem, os sinais passam por uma dizimação no tempo (no caso, uma subamostragem de ordem 2), gerando as saídas  $c_1[n]$ , que representa as componentes de baixa frequência, e  $d_1[n]$ , que representa as componentes de alta frequência. Do ponto de vista matemático,  $c_1[n]$  contém os chamados coeficientes de aproximação e  $d_1[n]$  os coeficientes de detalhes.

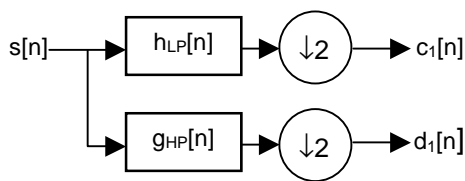


Fig. 1. Decomposição de um sinal em duas faixas

Considerando um sinal de entrada real, as equações que expressam as relações entre  $s[n]$ ,  $c_1[n]$  e  $d_1[n]$  são:

$$c_1[k] = \sum_n h[n-2k].s[n] \quad (1)$$

$$d_1[k] = \sum_n g[n-2k].s[n] \quad (2)$$

A partir do sinal  $c_1[n]$  podem ser obtidas mais duas faixas, e assim sucessivamente, com o processo sendo finalizado quando o comprimento do último sinal  $c_L[n]$  é um. O número de níveis e de faixas que pode ser obtido é proporcional ao comprimento do sinal a ser processado. Definindo  $N$  como sendo o tamanho da seqüência de entrada, o número máximo de níveis e faixa será  $L = \log_2 N$  [9], o que exige um sinal de entrada com comprimento em potência de 2. As várias faixas podem ser vistas como um banco de filtros do tipo passa-faixa.

A reconstrução do sinal é feita através de um processo inverso, ou seja, é feita uma sobreamostragem, uma filtragem e uma combinação final das várias faixas [11].

### III. ASPECTOS GERAIS DE REDUÇÃO DE RUÍDO USANDO A TRANSFORMADA DE WAVELET DISCRETA

Quando um sinal é decomposto em  $L$  faixas via DWT, o número de coeficientes wavelets com energia significativa é pequeno. Isto é uma consequência direta das propriedades de aproximações das wavelets, que assume um número suficiente de momentos nulos [11]. O sinal pode ser representado de modo preciso por um número pequeno de coeficientes. Assim, através de um limiar obtido a partir dos coeficientes wavelets, pode-se eliminar os coeficientes com

menor energia. Essa característica, que é básica na aplicação de wavelets em compressão de sinais [10], pode ser aproveitada também para se reduzir o ruído do sinal de entrada.

O cálculo do limiar e a metodologia de aplicação definem a qualidade na redução de ruído. De acordo com [10], a aplicação de um limiar pode ser feita de duas formas: “dura”(LD) ou “suave”(LS). Definindo  $x(t)$  como sendo o sinal na saída de cada faixa da DWT, o sinal obtido com a aplicação de um determinado limiar  $\lambda$  será:

$$y_{hard}(t) = \begin{cases} x(t), & |x(t)| > \lambda \\ 0, & |x(t)| \leq \lambda \end{cases} \quad \text{aplicação LD} \quad (3)$$

$$y_{soft}(t) = \begin{cases} \text{sign}[x(t)].[|x(t)| - \lambda], & |x(t)| > \lambda \\ 0, & |x(t)| \leq \lambda \end{cases} \quad \text{aplicação LS} \quad (4)$$

De forma mais genérica, pode-se afirmar que qualquer aplicação de limiar que não resulte nos próprios coeficientes abaixo ou acima do valor do limiar pode ser considerada como uma aplicação do tipo “limiar suave”.

A obtenção do limiar  $\lambda$  não é simples e o método mais usado é o proposto por [10], que é dado por:

$$\lambda = \sigma \sqrt{2 \log(N)}, \quad (5)$$

onde  $\sigma$  representa uma estimativa da potência do ruído (ou perfil do ruído).

Com a escolha de um  $\lambda$  apropriado, é possível melhorar a qualidade de um determinado sinal degradado por um ruído aditivo. Tipicamente, o sinal de voz contaminado por ruído é dado por:

$$s[n] = v[n] + \sigma.r[n] \quad (6)$$

onde  $\sigma.r[n]$  e  $v[n]$  representam o ruído aditivo e o sinal de voz limpo, respectivamente.

Essa possibilidade de redução de ruído tem gerado muitos artigos científicos [6], [11], etc., mas o ruído considerado na maioria deles tem sido, predominantemente, do tipo ruído branco, enquanto que as aplicações reais envolvem sinais com outros tipos de ruído, como os produzidos por carros, aviões, etc. Além disso, muitos ruídos não são estacionários, o que também causa sérias restrições nas aplicações desses métodos. Nos métodos clássicos [10], [11], calcula-se e aplica-se um único limiar para todas as faixas de frequência. Essa solução não parece ser adequada, pois o sinal de voz possui energia diferenciada nas diferentes faixas de frequência. Por exemplo, nos segmentos de voz do tipo “unvoiced”, para os quais a energia é mais significativa nas frequências altas, a aplicação de um único limiar gera alguns problemas. Por exemplo, quando o ruído é branco, pode-se melhorar a qualidade do sinal processado reduzindo-se de maneira mais significativa o ruído nas frequências altas. Já para o caso de um ruído colorido, como o obtido dentro de um carro, o efeito será o contrário, pois o ruído de carro tem energia mais significativa nas frequências baixas. Alguns métodos propõem a detecção desses trechos e a aplicação de limiares diferentes para as frequências altas e baixas [6]. Essa

solução pode até gerar algum resultado satisfatório, mas cria uma dependência com relação ao tipo de ruído, o que pode inviabilizar muitas aplicações.

Com relação ao limiar, seja ele do tipo duro ou suave, o sinal processado irá apresentar sempre descontinuidades no tempo e na frequência. Como consequência, o sinal de saída, apesar de apresentar um nível de ruído mais baixo, pode também apresentar um nível de degradação significativo. Dessa forma, torna-se necessário encontrar uma solução que minimize tais descontinuidades e, conseqüentemente, as distorções no sinal processado.

#### IV. O MÉTODO PROPOSTO

Nos estudos sobre a aplicação da DWT na redução de ruído, verificou-se que um dos principais problemas é o ruído residual obtido após o processamento do sinal. Esse ruído residual é algo semelhante ao "ruído musical" obtido em técnicas baseadas na Subtração Espectral [2] e é mais significativo quando o ruído aditivo é do tipo colorido e não-estacionário. O método proposto neste trabalho apresenta duas inovações com relação ao limiar: uma no cálculo e outra na aplicação.

##### A. Método para Cálculo do Limiar

Nos métodos clássicos de redução de ruído usando limiar [6] e [7], o cálculo do limiar é feito com base em todos os coeficientes wavelets, não havendo diferenciação de faixa de frequência. Esses métodos apresentam resultados satisfatórios quando o ruído é do tipo branco, que apresenta uma potência média praticamente igual em todas as frequências. O mesmo não ocorre com um ruído colorido como, por exemplo, o ruído dentro de um veículo, que tem energia predominante nas frequências baixas.

Considerando a  $i$ -ésima janela de silêncio do sinal de voz e a  $j$ -ésima faixa de frequência no domínio wavelet, obtém-se inicialmente o perfil do ruído de acordo com [8], que é dado por:

$$\sigma_{i,j} = (1/0.6745) \cdot \text{mediana}(|c_{i,j}|) \quad (7)$$

com  $c_i$  sendo o conjunto de coeficientes wavelets em cada banda  $j$ .

O limiar a ser aplicado é dado por:

$$\lambda_q = (1 - \alpha) \lambda_{q,j} + \alpha \lambda_{q-1,j}, \quad (8)$$

onde os limiares da janela atual e anterior são calculados de acordo com a equação (5).

Verifica-se que o limiar utilizado é uma média entre os limiares da janela atual  $i$  e da janela anterior  $i-1$ , para cada faixa de frequência  $j$ . O fator  $\alpha$  indica o grau de variação dessa média. Testes práticos mostram que é possível obter bons resultados quando  $\alpha=0,8$ .

##### B. Método para Aplicação do Limiar

A metodologia usada na aplicação do limiar é um dos pontos que diferenciam as diferentes técnicas de redução de ruído baseadas na DWT.

Para um melhor entendimento da metodologia proposta, considere o uso da DWT em sistemas de compressão de sinais de voz [12]. Nesses sistemas, é feito um cálculo de limiar para cada segmento de voz no domínio wavelet e anulam-se todos os coeficientes com valores abaixo do limiar. Isso significa a eliminação da redundância no sinal (coeficientes com baixa energia) e, por conseguinte, a compressão do mesmo. Porém, quando o sinal é ruidoso, como apresentado na equação (6), o limiar obtido será, considerando a linearidade da DWT, um valor que definirá um nível de redundância para o ruído e, quando aplicado, levará a uma redução de ruído praticamente desprezível. Isto significa que o limiar deve ser aumentado de acordo com a potência do ruído. Porém, considerando que as fases do sinal de voz limpo e do ruído são aleatórias, é possível que alguns valores de limiares já estejam acima de valores do sinal. Assim, o aumento no valor do limiar deve manter um compromisso entre o nível de redução de ruído que se deseja e as distorções introduzidas no sinal processado pois, quando se aplica o limiar, pode-se eliminar de forma aleatória coeficientes wavelets importantes para o sinal de voz. Essa eliminação aleatória faz surgir um ruído residual incômodo ao ouvido humano que lembra o "ruído musical". Para minimizar o problema, alguns autores propõem a aplicação de um limiar suave, mantendo uma relação direta com o sinal, o que significa o uso de funções não lineares. Os principais métodos mencionam o uso de funções não lineares para calcular os novos coeficientes wavelets, abaixo e acima do limiar obtido. Por exemplo, em [7], é usada uma função não linear aproximada por funções exponenciais. Os resultados obtidos são, sem dúvida, melhores, mas o ruído residual ainda persiste.

As propostas para aplicação de limiar neste trabalho são: 1 - o limiar obtido é aumentado por um fator  $\beta$  e 2 - os coeficientes abaixo do limiar são substituídos pela sua sigmóide multiplicada pelo próprio coeficiente. Desta forma, o conjunto de coeficientes de cada faixa  $j$  é alterado como segue:

$$\hat{c}_{j,k} = \begin{cases} c_{j,k} & , \text{ se } |c_{j,k}| > \beta\lambda \\ c_{j,k} \cdot \text{SIGMOIDE}(c_{j,k}) & , \text{ se } |c_{j,k}| \leq \beta\lambda \end{cases} \quad (9)$$

onde  $k$  varia de acordo como número máximo de coeficientes de cada faixa de frequência  $j$ .

A função sigmóide é calculada de acordo com a equação:

$$\text{SIGMOIDE}(x) = \frac{1 - e^{-x}}{1 + e^{-x}} \quad (10)$$

O método proposto na equação (9) possibilita reduzir mais ruído sem introduzir distorções significativas no sinal.

#### V. IMPLEMENTAÇÃO E RESULTADOS

O algoritmo de redução de ruído foi implementado com o software de simulação MATLAB e a DWT foi implementada usando a wavelet de Daubechies de ordem 10 [12]. O sinal

ruidoso foi obtido a partir de um sinal de voz limpo, em português, amostrado a 16 kHz, com posterior adição de ruído branco e de ruído de carro, também amostrados a uma taxa de 16 kHz. O número de amostras de cada segmento de voz foi de 256. Isto significa que a decomposição no domínio wavelet foi feita em oito faixas de frequência. A separação entre trechos de voz e silêncio, necessária para o cálculo do ruído, foi feita de forma ideal, ou seja, identificou-se todos os trechos de silêncio a partir do sinal limpo. A relação sinal/ruído dos sinais avaliados foi de, aproximadamente, 16 dB (ruído branco e colorido). O método proposto foi implementado assumindo  $\beta=6$  na equação (9) e o ruído de carro foi gravado no interior de um veículo de passeio a uma velocidade de 90 km/h em um asfalto semi-ruçoso [2].

Na figura 2 são apresentadas as formas de onda dos sinais originais (com ruído branco e ruído de carro) e na figura 3 são apresentadas as formas de onda dos sinais limpo e processados. Comparando as figuras 2 e 3 verifica-se, de modo imediato, que há uma redução significativa do ruído após o processamento. Já nas figuras 4 e 5 são apresentados os respectivos espectrogramas, com frequência limitada a 4 kHz (isto permite uma melhor visualização do espectro de interesse do sinal de voz). Observa-se, além da redução de ruído, que os espectros dos sinais após o processamento (figura 5) estão mais próximos do espectro do sinal original do que os espectros dos sinais ruidosos (figura 4). Isto significa, naturalmente, uma melhor qualidade auditiva. Esses resultados foram comprovados em testes subjetivos informais (a grande diferença entre os sinais processados e os sinais ruidosos justifica a não aplicação de uma avaliação subjetiva formal), onde 3 voluntários indicaram 100% de preferência pelos sinais processados, quando comparados com os respectivos sinais ruidosos.

Como mencionando no início deste trabalho, os métodos de redução de ruído baseados em wavelets apresentam bons resultados quando o ruído é branco. No método proposto, verifica-se que é possível obter bons resultados com ruído colorido. Usando a distância cepstral [13] como medida objetiva de distorção, foram avaliadas as distorções nos dois sinais processados. O resultado é apresentado na figura 6. Verifica-se que o sinal contaminado por ruído de carro apresenta menos distorções nos trechos de fala do que o sinal contaminado por ruído branco. Esse resultado parece ser óbvio, considerando que a potência do ruído de carro é menor nas frequências altas, enquanto que o ruído branco apresenta, teoricamente, a mesma potência em todas as frequências. Isto é, de fato, uma explicação. Porém, só foi possível obter tal resultado porque o método proposto calcula e aplica o limiar por faixa e, também, gera coeficientes finais que são parcelas dos coeficientes do sinal ruidoso (aplicação da sigmóide). Para uma avaliação comparativa, na figura 7 é apresentado o resultado de medida de distorção para o método proposto por Hamid [7], considerando os mesmos sinais.

Verifica-se que as distorções para os dois ruídos têm o mesmo perfil, sendo que as distorções para o ruído de carro são ligeiramente inferiores nos trechos de fala. Uma comparação entre as figuras 6 e 7 mostra que método

proposto apresenta menos distorções para os dois tipos de ruído (na verdade, considerando o tratamento dado para os trechos de sinal do tipo “unvoiced”, era de se esperar que o método proposto por Hamid apresentasse um desempenho superior na redução do ruído branco).

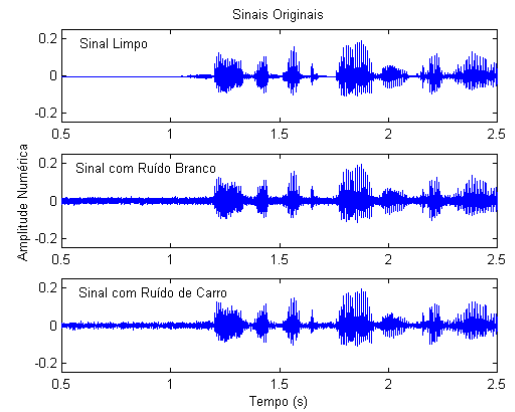


Fig. 2 – Formas de ondas dos sinais original e ruidosos

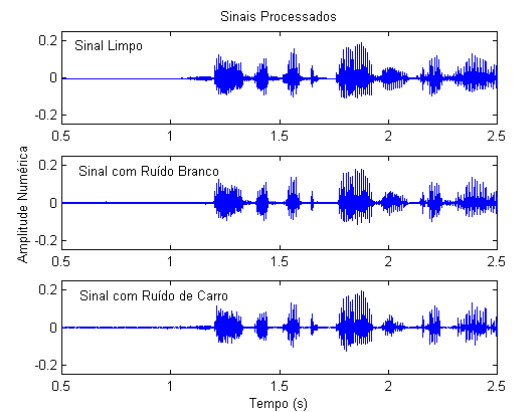


Fig. 3– Formas de ondas dos sinais original e processados

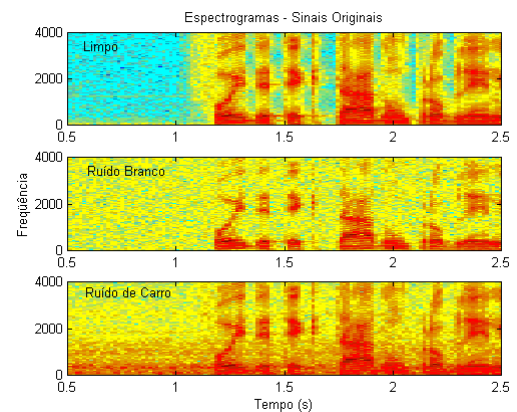


Fig. 4 – Espectrogramas dos sinais original e ruidosos

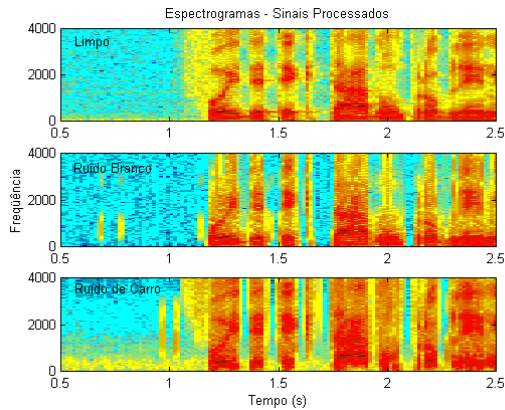


Fig. 5 – Espectrogramas dos sinais original e processados

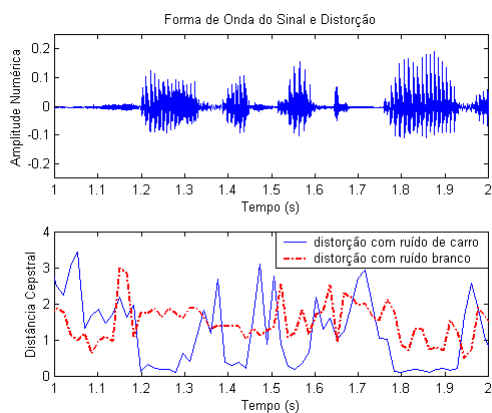


Fig. 6 – Distorções com o Método Proposto

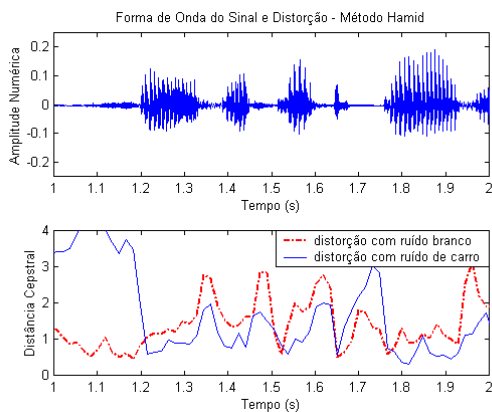


Fig. 7 – Distorções com o Método de Hamid

## VI CONCLUSÕES

Neste trabalho foi apresentado um novo método de redução de ruído baseado em wavelets. As inovações básicas foram feitas no cálculo de limiar e na aplicação do mesmo. O método de limiar aplicado é do tipo “suave”, onde os coeficientes wavelets não são zerados. Os resultados mostram que o método proposto possibilita a redução de ruído branco e

colorido sem introduzir distorções significativas no sinal processado. Além disso, o sinal processado com o novo método reduz significativamente um incômodo ruído residual que é típico de métodos que usam wavelets e similar ao “ruído musical” (característico de metodologias baseadas na Transformada de Fourier).

## REFERÊNCIAS

- [1] Y. Ephraim and D. Malah, “Speech Enhancement Using Minimum Mean Square Error Short-Time Spectral Amplitude Estimator,” *IEEE Trans. On Acoustics, Speech and Signal Processing*, v. 32, NO 6 - December-1984.
- [2] J. Vieira Filho, P. Scalart e J. G. Chiquito, “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.
- [3] S. F. Boll, “Suppression of Acoustic Noise in Speech using Spectral Subtraction,” in: *IEEE Trans. On Acoustic, Speech and Signal Processing (TASSP), Proceedings*, vol. 29, pp: 113-120, April-1979.
- [4] Yi Hu and P.C. Loizu, “A Perceptually Motivated Approach for Speech Enhancement,” *IEEE Trans. On Speech and Audio Processing*, Vol. 11, No. 5, September – 2003, Elsevier Signal Processing, Vol. 83, 2003.
- [5] T. Gülzow, T. Ludwig and U. Heute, “Spectral-Subtraction Speech Enhancement in Multirate Systems with and without non-uniform and Adaptive Bandwidth,” *Elsevier Signal Processing*, Vol. 83, 2003.
- [6] J. Seok, and K. S. Bae., “Speech Enhancement with Reduction of Noise Components in the Wavelet Domain,” in: *Proceedings of the 1997 IEEE International conference on Acoustics, Speech and signal processing (ICASSP'97)*, Vol. II, pp. 1323-1326, Munich, April 1997.
- [7] H. Sheikzadeh and H. R. Abutalebi, “An Improved Wavelet-Based speech Enhancement System,” in: *Proceedings of the 7<sup>th</sup> European Conference on Speech Communication and technology (Eurospeech 2001)*, Aalborg, Denmark, September 2001.
- [8] L. D. Donoho, and I. M. Johnstone, “Ideal Spatial Adaptation via Wavelet Shrinkage”, *Biometrika*, Vol. 81: No. 3, pp: 425-455, 1994.
- [9] O. Rioul, and M. Vetterli., “Wavelets and Signal Processing”, *IEEE Signal Processing Magazine*, vol.8 No.4, pp. 14-38, October 1991.
- [10] K. Kinsner and A. Langi, “Speech and Image Signal Compression with Wavelets,” in: *IEEE Wescanex Conference Proceedings*, pp: 368-375, New York, 1993.
- [11] T. K. Sarkar, C. Su, R. Adve, M. S. Palma, L. G. Castilho and R. R. Boix, “A Tutorial on Wavelets from an Electrical Engineering Perspective, Part 1: Discrete Wavelet Techniques,” *IEEE Antennas and Propagation Magazine*, Vol. 40, No. 5, October 1998.
- [12] I. Daubechies, *Ten Lectures on Wavelets*, SIAM Books, 1992.
- [13] A. H. Gray Jr & J. D. Markel, “Distance Measures for Speech Processing,” *IEEE Trans. On Acoustics, Speech and Signal Processing*, Vol.. 24, No. 5 - October - 1976.