

IRINEU ANTUNES JÚNIOR

REDUÇÃO DE RUÍDO EM SINAIS DE VOZ USANDO CURVAS  
ESPECIALIZADAS DE MODIFICAÇÃO DOS COEFICIENTES DA  
TRANSFORMADA EM CO-SENO

Tese apresentada à Escola Politécnica  
da Universidade de São Paulo para  
obtenção do título de Doutor em  
Engenharia Elétrica.

São Paulo  
2006

IRINEU ANTUNES JÚNIOR

REDUÇÃO DE RUÍDO EM SINAIS DE VOZ USANDO CURVAS  
ESPECIALIZADAS DE MODIFICAÇÃO DOS COEFICIENTES DA  
TRANSFORMADA EM CO-SENO

Tese apresentada à Escola Politécnica  
da Universidade de São Paulo para  
obtenção do título de Doutor em  
Engenharia Elétrica.

Área de Concentração:  
Engenharia de Sistemas Eletrônicos

Orientador:  
Phillip M. S. Burt

São Paulo  
2006

Este exemplar foi revisado e alterado em relação à versão original, sob responsabilidade única do autor e com a anuência de seu orientador.

São Paulo, 23 de maio de 2006.

Assinatura do autor

Assinatura do orientador

Antunes Júnior, Irineu

Redução de ruído em sinais de voz usando curvas especializadas de modificação dos coeficientes da transformada em co-seno.

São Paulo, 2006.

112 p.

Tese (Doutorado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Telecomunicações e Controle.

1. Processamento digital de voz 2. Estimaco no-paramtrica. I. Universidade de So Paulo.

Escola Politécnica. Departamento de Engenharia de Telecomunicações e Controle II. t

Ao Professor Max Gerken

*(in memoriam)*

## A G R A D E C I M E N T O S

Ao Professor Max Gerken, pelo incentivo e apoio prestados, anteriormente, durante o mestrado e, recentemente, no primeiro ano deste doutorado.

Ao meu orientador Phillip, pela discussão valiosa e revisão minuciosa do texto, além do apoio, incentivo e atenção prestados durante o curso de doutorado.

Aos Profs. Luiz Cezar Trintinália e Antonio Fischer de Toledo, pela leitura do texto e sugestões fornecidas no exame de qualificação.

Ao meu amigo Claudio José Bordin Júnior, pelo companheirismo e auxílio durante todo o doutorado.

Aos meus grandes amigos Elias Ribeiro de Castro, pela constante amizade e sugestões de revisão gramatical do texto, e Cláudio S. Oliveira, pela companhia e gentil disponibilidade para realizar voluntariamente os testes de audição deste trabalho.

À Ericsson Telecomunicações, pelo apoio financeiro nos primeiros anos do doutorado, no âmbito do projeto de pesquisa sobre métodos de aprimoramento de sinais de voz, desenvolvido junto à Fundação para o Desenvolvimento Tecnológico da Engenharia (FDTE) e junto ao PTC/EPUSP.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pelo financiamento de parte deste trabalho.

À minha mãe Maria das Graças, ao meu pai Irineu, à minha avó Laura, a toda a minha família, pelo auxílio e atenção que sempre me deram.

E a outros que contribuíram direta ou indiretamente para a realização deste trabalho.

# Conteúdo

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Importância e Campo de Aplicação . . . . .	2
1.2	Terminologia Empregada no Texto . . . . .	3
1.3	Organização do Texto . . . . .	4
<b>2</b>	<b>Métodos de Redução de Ruído</b>	<b>5</b>
2.1	Melhoria de Sinais de Voz . . . . .	5
2.2	Redução de Ruído Usando Transformadas . . . . .	7
2.3	Filtro de Wiener . . . . .	10
2.4	Subtração Espectral [12] . . . . .	14
2.5	Funções de Limiar Hard e Soft . . . . .	18
2.6	Eliminação de Ruído no Domínio Wavelet . . . . .	20
2.7	Outras Funções de Limiar . . . . .	25
2.8	Determinação do valor do limiar . . . . .	27
2.9	Seleção dos Coeficientes pelo “Oráculo” . . . . .	32
2.10	Ruído Musical e Outras Distorções . . . . .	33
2.11	Conclusão . . . . .	37
<b>3</b>	<b>Funções Especializadas</b>	<b>39</b>
3.1	Descrição do Problema . . . . .	40

3.2	Outros Métodos de Seleção de Coeficientes . . . . .	41
3.3	Funções de <i>Threshold</i> Propostas . . . . .	42
3.4	Motivação do <i>Threshold</i> com Dois Limiares . . . . .	44
3.5	<i>Threshold</i> como Estimador de Mínimo Erro . . . . .	45
3.6	Obtenção Direta dos Limiares Ótimos . . . . .	57
3.7	Obtenção dos Limiares usando Modelo para $msee(t)$ . . . . .	61
3.8	Obtenção dos Limiares usando a Função GCV . . . . .	71
3.9	Obtenção dos Limiares usando Modelo Estatístico . . . . .	81
3.10	Conclusão e Contribuições . . . . .	89
<b>4</b>	<b>Simulação Computacional</b>	<b>90</b>
4.1	Considerações Iniciais . . . . .	90
4.2	<i>Threshold</i> Aplicado a Sinais de Voz . . . . .	94
4.3	Comparação com Outros Métodos . . . . .	99
4.4	Conclusão . . . . .	106
<b>5</b>	<b>Considerações Finais</b>	<b>107</b>

## R E S U M O

Muitos métodos de redução de ruído se baseiam na possibilidade de representar o sinal original com um reduzido número de coeficientes de uma transformada, ou melhor, obtém-se um sinal com menos ruído pelo cancelamento dos coeficientes abaixo de um valor adequadamente estabelecido de magnitude. Deve-se supor que a contribuição do ruído se distribua de maneira uniforme por todos os coeficientes.

Uma desvantagem destes métodos, quando aplicados a sinais de voz, é a distorção introduzida pela eliminação dos coeficientes de pequena magnitude, juntamente com a presença de sinais espúrios, como o “ruído musical” produzido por coeficientes ruidosos isolados que eventualmente ultrapassam o limiar.

Para as transformadas usualmente empregadas, o histograma da distribuição dos coeficientes do sinal de voz possui um grande número de coeficientes próximos à origem. Diante disto, propomos uma nova função de “thresholding” concebida especialmente para redução de ruído em sinais de voz adicionados a AWGN (“Additive, White, and Gaussian Noise”). Esta função, chamada de SoftSoft, depende de dois valores de limiar: um nível inferior, ajustado para reduzir a distorção da voz, e um nível superior, ajustado para eliminar ruído.

Os valores ótimos de limiar são calculados para minimizar uma estimativa do erro quadrático médio (MSE): diretamente, supondo conhecido o sinal original; indiretamente, usando uma função de interpolação para o MSE, levando a um método prático.

A função SoftSoft alcança um MSE inferior ao que se obtém pelo emprego das conhecidas operações de “Soft” ou “Hard-thresholding”, as quais dispõem apenas do limiar superior. Ainda que a melhoria em termos de MSE não seja muito expressiva, a melhoria da qualidade perceptual foi certificada tanto por um ouvinte quanto por uma medida perceptual de distorção (a distância log-espectral).



## A B S T R A C T

Many noise-reduction methods are based on the possibility of representing the clean signal as a reduced number of coefficients of a block transform, so that cancelling coefficients below a certain thresholding level will produce an enhanced reconstructed signal. It is necessary to assume that the clean signal has a sparse representation, while the noise energy is spread over all coefficients.

The main drawback of those methods is the speech distortion introduced by eliminating small magnitude coefficients, and the presence of artifacts (“musical noise”) produced by isolated noisy coefficients randomly crossing the thresholding level.

Based on the observation that the speech coefficient histogram has many important coefficients close to origin, we propose a custom thresholding function to perform noise reduction in speech signals corrupted by AWGN. This function, called SoftSoft, has two thresholding levels: a lower level adjusted to reduce speech distortion, and a higher level adjusted to remove noise. The joint optimal values can be determined by minimizing the resulting mean square error (MSE).

We also verify that this new thresholding function leads to a lower MSE than the well-known Soft and Hard-thresholding functions, which employ only a higher thresholding level. Although the improvement in terms of MSE is not expressive, a perceptual distortion measure (the log-spectral distance, LSD) is employed to prove the higher performance of the proposed thresholding scheme.

# Lista de Abreviaturas

AWGN.....	<i>Additive, White and Gaussian Noise</i>
dB.....	<i>deciBel</i>
DCT.....	<i>Discrete Cosine Transform</i>
DFT.....	<i>Discrete Fourier Transform</i>
DTFT.....	<i>Discrete Time Fourier Transform</i>
DWT.....	<i>Discrete Wavelet Transform</i>
FFT.....	<i>Fast Fourier Transform</i>
FIR.....	<i>Finite Impulse Response (filter)</i>
FWT.....	<i>Fast Wavelet Transform</i>
GCV.....	<i>Generalized Cross Validation</i>
i.i.d. ....	<i>independentes e identicamente distribuídos</i>
IIR.....	<i>Infinite Impulse Response (filter)</i>
LCT.....	<i>Local Cosine Transform</i>
LSD.....	<i>Log-Spectral Distance</i>
MMSEWC ...	<i>Minimum Mean-Square Error Wavelet Coefficient Estimator</i>
MOS.....	<i>Mean Opinion Score</i>
MSCE.....	<i>Mean-Square Complementary Error</i>
MSEE.....	<i>Mean-Square Estimation Error</i>
PDF.....	<i>Probability Density Function</i>
PMF.....	<i>Probability Mass Function</i>
SegSNR.....	<i>Segmented Signal-to-noise Ratio</i>
SNR.....	<i>Signal-to-noise Ratio</i>
SURE.....	<i>Stein's Unbiased Risk Estimate</i>
VAD.....	<i>Voice Activity Detector</i>

## Gracias por visitar este Libro Electrónico

Puedes leer la versión completa de este libro electrónico en diferentes formatos:

- HTML(Gratis / Disponible a todos los usuarios)
- PDF / TXT(Disponible a miembros V.I.P. Los miembros con una membresía básica pueden acceder hasta 5 libros electrónicos en formato PDF/TXT durante el mes.)
- Epub y Mobipocket (Exclusivos para miembros V.I.P.)

Para descargar este libro completo, tan solo seleccione el formato deseado, abajo:

