



Eduardo Esteves Vale

**Identificação de Locutor Robusta e Independente do Texto
Usando Múltiplos Classificadores em Sub-bandas**

Tese de Doutorado

Documento apresentado ao Programa de Pós-Graduação em Engenharia Elétrica do Departamento de Engenharia Elétrica da PUC-Rio como requisito parcial para obtenção do título de Doutor em Engenharia Elétrica.

Orientador: Prof. Abraham Alcaim
Co-orientadora: Profa. Rosângela Fernandes Coelho

Rio de Janeiro
Setembro de 2010



Eduardo Esteves Vale

**Identificação de Locutor Robusta e
independente do Texto Usando Múltiplos
Classificadores em Sub-Bandas**

Tese apresentada como requisito parcial para obtenção do grau de Doutor pelo Programa de Pós-Graduação em Engenharia Elétrica do Departamento de Engenharia Elétrica do Centro Técnico Científico da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

Prof. Abraham Alcaim
Orientador

Centro de Estudos em Telecomunicações - PUC-Rio

Profa. Rosângela Fernandes Coelho
Co-orientadora
IME

Prof. Marco Antonio Grivet Mattoso Maia
Centro de Estudos em Telecomunicações /PUC-Rio

Prof. Paulo Roberto Rosa Lopes Nunes
IME

Prof. Fernando Gil Vianna Resende Junior
UFRJ

Prof. Rui Seara
UFSC

Prof. José Eugenio Leal
Coordenador Setorial do Centro
Técnico Científico - PUC-Rio

Rio de Janeiro, 17 de setembro de 2010

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e dos orientadores.

Eduardo Esteves Vale

Graduou-se em Engenharia Elétrica com ênfase em Telecomunicações na PUC-Rio (Pontifícia Universidade Católica do Rio de Janeiro) em 2003. Concluiu o Mestrado em Telecomunicações em 2006, também na PUC-Rio. Atualmente é o oficial Encarregado da Divisão de Conectividade do CLTI (Centro Local de Tecnologia da Informação) do Comando-em-Chefe da Esquadra da Marinha do Brasil.

Ficha Catalográfica

Vale, Eduardo Esteves

Identificação de locutor robusta e independente do texto usando múltiplos classificadores em sub-bandas / Eduardo Esteves Vale ; orientador: Abraham Alcaim ; co-orientadora: Rosângela Fernandes Coelho. – 2010. 90 f. : il. (preto e branco.) ; 30 cm

Tese (doutorado)–Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Elétrica, 2010.

Inclui bibliografia

1. Engenharia elétrica – Teses. 2. Identificação de locutor robusta e independente do texto. 3. Múltiplos classificadores. 4. Sub-bandas. 5. Gaussian Mixture Models (GMM). 6. Mel Frequency Cepstral Coefficients (MFCC). 7. Espaço nulo. 8. Atributos dinâmicos. 9. Coeficientes de autocorrelação. I. Alcaim, Abraham. II. Coelho, Rosângela Fernandes. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Elétrica. IV. Título.

CDD: 621.3

À minha querida esposa, Dra. Daniele de Macedo,
minha companheira inseparável.

Agradecimentos

Aos meus orientadores Abraham Alcaim e Rosângela Coelho por terem proporcionado uma orientação segura.

Aos professores que participaram da Banca examinadora.

À PUC-Rio, que forneceu toda a infraestrutura para a realização deste trabalho.

Ao CNPq e CAPES pelo apoio financeiro.

Aos meus companheiros das diversas Organizações Militares da Marinha do Brasil, em especial aos do Comando-Em-Chefe da Esquadra, DCTIM (Diretoria de Comunicações e Tecnologia da Informação da Marinha), CIAW (Centro de Instrução Almirante Wandenkolk), CTIM (Centro de Tecnologia da Informação da Marinha) e BACS (Base Almirante Castro e Silva), pela amizade e reconhecimento.

Aos meus familiares, em especial aos meus queridos pais (Prof. Eduardo e Ivanisa), avô (Prof. Eduardo Valle) e avó (Luzia), que sempre me incentivaram para a realização do Doutorado.

À minha dedicada esposa Daniele, que sempre me apoiou.

Principalmente ao meu querido Deus, que está sempre presente me fortalecendo e iluminando durante a caminhada, e por ter criado tão bela Natureza a qual temos a felicidade de poder estudar.

Resumo

Vale, Eduardo Esteves; Alcaim, Abraham (Orientador); Coelho, Rosângela Fernandes (Co-orientadora). **Identificação de Locutor Robusta e Independente do Texto Usando Múltiplos Classificadores em Sub-bandas**. Rio de Janeiro, 2010. 90p. Tese de Doutorado – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

Esta tese destina-se ao desenvolvimento de novas técnicas de combinação de classificadores aplicados em sub-bandas visando melhorar a identificação de locutor robusta e independente do texto. As vantagens observadas nas pesquisas utilizando múltiplos classificadores em sub-bandas para o reconhecimento de locutor robusto motivaram o desenvolvimento de técnicas de combinação desses algoritmos. Nessa tese foram propostas novas abordagens para a combinação das respostas dos classificadores nas sub-bandas. O principal objetivo é melhorar a taxa de acerto em situações onde nada se sabe sobre o tipo de ruído que pode estar corrompendo os sinais de voz usados no teste do sistema. As diferentes propostas consistem no emprego de pesos não-uniformes, espaço nulo, treinamento em múltiplas condições, atributos dinâmicos e coeficientes de autocorrelação – MFCC. A aplicação das novas propostas contribui significativamente para a melhoria da taxa de acerto do sistema de reconhecimento. Obteve-se, por exemplo, um aumento na taxa de reconhecimento, em relação à técnica de combinação Soma apresentada na literatura, de aproximadamente 47% em testes com ruído branco, e de 32% em testes com ruído não-branco em 15 segundos de fala e 10 dB de RSR (Relação Sinal Ruído), apenas utilizando uma nova estratégia que emprega o espaço nulo na combinação de classificadores em sub-bandas. Resultados mais significativos foram obtidos empregando-se as demais propostas apresentadas no presente trabalho.

Palavras-chave

Identificação de locutor robusta e independente do texto; múltiplos classificadores; sub-bandas; *Gaussian Mixture Models* (GMM); *Mel Frequency Cepstral Coefficients* (MFCC); espaço nulo; atributos dinâmicos; coeficientes de autocorrelação.

Abstract

Vale, Eduardo Esteves; Alcaim, Abraham ; Coelho, Rosângela Fernandes.
Robust Text-independent Speaker Identification Using Multiple Classifiers in Sub-bands. Rio de Janeiro, 2010. 90p. DSc. Thesis – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

This Thesis aims to develop new classifier combination techniques applied in sub-bands in order to improve the robustness of text-independent speaker identification systems. The advantages observed in previous experiments using multiple classifiers in sub-bands for robust speaker recognition motivated the development of combination techniques for these algorithms. New strategies to combine the classifiers responses are proposed in this Thesis. The main purpose is to increase the recognition performance in situations when there is no knowledge about the type of noise that corrupts the testing speech signal. The different proposals consist in applying non-uniform weights, null space, multicondition training, dynamic features and autocorrelation based MFCC features. The employment of the new strategies significantly contribute to increase the recognition performance. It was obtained an increase, for instance, compared to the Sum combination technique shown in the literature, of about 47% in tests with white noise, and 32% with non-white noise in 15 seconds of speech in 10 dB of SNR (Signal-to-noise ratio), just using a new strategy which employ the null space to combine the sub-band classifiers. Even better results were obtained by using the other proposals.

Keywords

Robust text-independent speaker identification; multiple classifiers; sub-bands; Gaussian Mixture Models (GMM); Mel Frequency Cepstral Coefficients (MFCC); null space; dynamic features; autocorrelation coefficients.

Sumário

1. Introdução	15
1.1. Sistemas de Reconhecimento de Locutor	15
1.2. Considerações Iniciais Sobre as Novas Propostas	20
1.3. Organização da Tese	22
2. Reconhecimento Automático de Locutor em Ambientes Ruidosos	24
2.1. Os Atributos Utilizados	26
2.2. Os Classificadores	35
3. Identificação de Locutor Usando Técnicas de Múltiplos Classificadores em Sub-bandas Com Pesos Não-Uniformes	42
3.1. Estratégias Empregadas em Identificação de Locutor Usando Múltiplos Classificadores em Sub-bandas	43
3.2. O Método Proposto que Utiliza uma Combinação de Pesos Não Uniformes	44
3.3. Resultados de Simulação	45
3.4. Conclusões	49
4. Técnica de Combinação de Medidas de Verossimilhança Baseada no Espaço Nulo	50
4.1. O Espaço Nulo	51
4.2. A Proposta Baseada no Cálculo do Espaço Nulo	52
4.2.1 Resultados de Simulação	57
4.3. Combinação de Classificadores em Sub-bandas Usando o Espaço Nulo e Treinamento Com Múltiplas Condições	63
4.3.1 Resultados de Simulação	65
4.4. Combinação de Classificadores em Sub-bandas Usando o Espaço Nulo e Atributos Dinâmicos	66

4.4.1 Resultados de Simulação	67
4.5. Conclusões	70
5. A Utilização da Técnica do Espaço Nulo e dos Atributos Baseados na Escolha de Coeficientes de Autocorrelações	72
5.1. Resultados de Simulação	72
5.2. Conclusões	78
6. Conclusões e Sugestões Para Trabalhos Futuros	79
6.1. Trabalhos Futuros	80
Referências Bibliográficas	81

Lista de Figuras

Figura 1- Etapa de treinamento de um sistema automático de reconhecimento de locutor	16
Figura 2 – Etapa de teste do sistema automático de reconhecimento de locutor	17
Figura 3 - Sistema de reconhecimento usando múltiplos classificadores em sub-bandas	19
Figura 4 – Espectro de potência dos ruídos	25
Figura 5 – Ilustração simplificada do modelo do aparelho vocal	28
Figura 6 – Banco de filtros na escala Mel	31
Figura 7 – Cruzamentos de nível	33
Figura 8 – Regressão usada para obter os parâmetros de Hurst	35
Figura 9 – Sistema de classificação baseado na verificação de locutor	36
Figura 10 – Sistema de classificação baseado na identificação de locutor	36
Figura 11 – Histograma dos padrões	39
Figura 12 – Esquema de combinação de pesos não-uniformes	45
Figura 13 – Esquema de combinação de medidas de verossimilhança	52
Figura 14 – Desempenho de identificação (%) em 15s de teste para ruído de Fábrica em RSR=10dB	57
Figura 15 – Desempenho de identificação (%) em 15s de teste para ruído de Fábrica em RSR=0dB	58
Figura 16 – Desempenho de identificação (%) em 15s de teste para ruído de Falatório em RSR=10dB	58
Figura 17 – Desempenho de identificação (%) em 15s de teste para ruído de Falatório em RSR=0dB	59
Figura 18 – Desempenho de identificação (%) em 15s de teste	

para ruído de Carro em RSR=10dB	59
Figura 19 – Desempenho de identificação (%) em 15s de teste para ruído de Carro em RSR=0dB	60
Figura 20 – Desempenho de identificação (%) em 15s de teste para ruído Branco em RSR=10dB	60
Figura 21 – Desempenho de identificação (%) em 15s de teste para ruído Branco em RSR=0dB	61
Figura 22 – Esquema de combinação utilizando o treinamento com múltiplas condições	64

Lista de Tabelas

Tabela 3.1 - Desempenho de identificação (%) em 15s de teste para ruído de Fábrica em RSR=10dB	46
Tabela 3.2 - Desempenho de identificação (%) em 15s de teste para ruído de Fábrica em RSR=0dB	47
Tabela 3.3 - Desempenho de identificação (%) em 15s de teste para ruído de Falatório em RSR=10dB	47
Tabela 3.4 - Desempenho de identificação (%) em 15s de teste para ruído de Falatório em RSR=0dB	47
Tabela 3.5 - Desempenho de identificação (%) em 15s de teste para ruído de Carro em RSR=10dB	47
Tabela 3.6 - Desempenho de identificação (%) em 15s de teste para ruído de Carro em RSR=0dB	47
Tabela 3.7 - Desempenho de identificação (%) em 15s de teste para ruído Branco em RSR=10dB	47
Tabela 3.8 - Desempenho de identificação (%) em 15s de teste para ruído Branco em RSR=0dB	47
Tabela 4.1 - Desempenho de identificação (%) em 15s de teste para ruído de Fábrica, Falatório, Carro e Branco em RSR=10dB	65
Tabela 4.2 - Desempenho de identificação (%) em 15s de teste para ruído de Fábrica, Falatório, Carro e Branco em RSR=15dB	67
Tabela 4.3 - Desempenho de identificação (%) em 5s de teste para ruído de Fábrica, Falatório, Carro e Branco com RSR=15dB	68
Tabela 4.4 - Desempenho de identificação (%) em 15s de teste para ruído de Fábrica, Falatório, Carro e Branco com RSR=10dB	68
Tabela 4.5 - Desempenho de identificação (%) em 5s de teste para ruído de Fábrica, Falatório, Carro e Branco em RSR=10dB	69
Tabela 5.1 - Desempenho de identificação (%) em 15s de teste para ruído de Fábrica, Falatório, Carro e Branco em RSR=15dB	73
Tabela 5.2 - Desempenho de identificação (%) em 5s de teste	

para ruído de Fábrica, Falatório, Carro e Branco em RSR=15dB	74
Tabela 5.3 - Desempenho de identificação (%) em 15s de teste	
para ruído de Fábrica, Falatório, Carro e Branco em RSR=10dB	75
Tabela 5.4 - Desempenho de identificação (%) em 5s de teste	
para ruído de Fábrica, Falatório, Carro e Branco em RSR=10dB	76
Tabela 5.5 - Desempenho de identificação (%) em 15s de teste	
para ruído de Fábrica, Falatório, Carro e Branco em RSR=5dB	77
Tabela 5.6 - Desempenho de identificação (%) em 5s de teste	
para ruído de Fábrica, Falatório, Carro e Branco em RSR=5dB	77

Siglas

AMFCC - *Autocorrelation Mel Frequency Cepstral Coefficients*

CMS - *Cepstral Mean Subtraction*

dB - *Bhattacharyya Distance*

DCT- *Discrete Cosine Transform*

DFT - *Discrete Fourier Transform*

DWT - *Discrete Wavelet Transform*

EM - *Expectation-Maximization*

EIH - *Ensemble Interval Histogram*

GMM - *Gaussian Mixture Models*

pH - *parâmetro de Hurst*

LDC - *Linguistic Data Consortium*

LFCC - *Linear Frequency Cepstrum Coefficients*

LPC - *Linear Predictive Coefficients*

LPCC - *Linear Prediction Cepstrum Coefficients*

M_dim_fBm - *Multidimensional fractional Brownian motion*

MFCC - *Mel Frequency Cepstral Coefficients*

PLP - *Perceptual Linear Predictive*

RSR - *Relação Sinal Ruído*

SNR - *Signal- to-noise ratio*

SVD - *Singular Value Decomposition*

UBM - *Universal Background Model*

ZCPA - *Zero Crossings with Peak Amplitudes*

ZCPAC - *Zero Crossing with Peak Amplitude Cepstrum*