

Harry Arnold Anacleto Silva

**Reconhecimento de Locutor Independente do
Texto em Presença de Ruído Usando
PAC-MFCC e Classificadores em Sub-Bandas**

Dissertação de Mestrado

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-graduação em Engenharia Elétrica do Departamento de Engenharia Elétrica da PUC-Rio.

Orientador: Prof. Abraham Alcaim

Rio de Janeiro
Abril de 2011

Harry Arnold Anacleto Silva

**Reconhecimento de Locutor Independente do
Texto em Presença de Ruído Usando
PAC-MFCC e Classificadores em Sub-Bandas**

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre 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

Departamento de Engenharia Elétrica — PUC-Rio

Prof. Fernando Gil Vianna Resende Junior

UFRJ-Rio

Prof. Marco Antonio Grivet Mattoso Maia

Centro de Estudos em Telecomunicações — PUC-Rio

Prof. José Eugenio Leal

Coordenador Setorial do Centro Técnico Científico – PUC-Rio

Rio de Janeiro, 1 de Abril de 2011

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

Harry Arnold Anacleto Silva

Graduou-se em Engenharia Eletrônica na Universidad Nacional Pedro Ruiz Gallo do Perú (Chiclayo, Perú).

Ficha Catalográfica

Anacleto Silva, Harry Arnold

Reconhecimento de Locutor Independente do Texto em Presença de Ruído Usando PAC-MFCC e Classificadores em Sub-Bandas/ Harry Arnold Anacleto Silva; orientador: Abraham Alcaim. 2011.

67 f.: il.(color.) ; 30 cm

Dissertação (mestrado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Elétrica, 2011.

Inclui bibliografia.

1. Engenharia Elétrica – Tese. 2. Reconhecimento independente de locutor. 3. PAC-MFCC com classificador em sub-bandas. 4. SSCH. 5. PAC-MFCC. 6. TECC. 7. MFCC. I. Alcaim, Abraham. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Elétrica. III. Título.

CDD: 621.3

Aos meus pais, Victor e Lila, e a minhas irmãs Diana e Astrid pelo carinho,
apoio, incentivo e por estar sempre presentes em minha vida.

Agradecimentos

A meu orientador Abraham Alcaim pela amizade, paciência, compreensão e confiança que me motivaram para a realização do presente trabalho.

Ao amigo Dirceu Silva Gonzaga pelo apoio e ajuda incondicional no desenvolvimento dos programas, que fizeram possível a execução desta dissertação.

À CAPES e à PUC-Rio, pelos auxílios concedidos, sem os quais este trabalho não poderia ter sido realizado.

Ao Presbítero Zamora Alarcon, pela ajuda e apoio oferecida em muitas ocasiões desde minha chegada ao Brasil.

A todos os amigos e familiares que de uma forma ou de outra me estimularam ou me ajudaram.

Resumo

Anacleto Silva, Harry Arnold ; Alcaim, Abraham. **Reconhecimento de Locutor Independente do Texto em Presença de Ruído Usando PAC-MFCC e Classificadores em Sub-Bandas**. Rio de Janeiro, 2011. 67p. Dissertação de Mestrado — Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

O presente trabalho propõe a utilização do atributo PAC-MFCC operando com Classificadores em Sub-Bandas para a tarefa de identificação de locutor independente do texto em ruído. O sistema proposto é comparado com os atributos MFCC (Coeficientes Cepstrais de Frequência Mel), PAC-MFCC (Fase Autocorrelação-MFCC) sem uso de classificadores em sub-bandas, SSCH (Histogramas de Centróides de Sub-Bandas Espectrais) e TECC (Coeficientes Cepstrais da Energia Teager). Nesta tarefa de reconhecimento, utilizou-se a base TIMIT a qual é composta de 630 locutores onde cada um deles falam 10 frases de aproximadamente 3 segundos cada frase, das quais 8 frases foram utilizadas para treinamento e 2 para teste, obtendo-se um total de 1260 locuções para o reconhecimento. Investigou-se o desempenho dos diversos sistemas utilizando diferentes tipos de ruídos da base Noisex 92 com diferentes relação sinal ruído. Verificou-se que a taxa de acerto da técnica PAC-MFCC com classificador em Sub-Bandas apresenta o melhor desempenho em comparação com as outras técnicas quando se tem uma relação sinal ruído menor que 10dB.

Palavras-chave

Reconhecimento independente de locutor; PAC-MFCC com classificador em sub-bandas; SSCH; PAC-MFCC; TECC; MFCC.

Abstract

Anacleto Silva, Harry Arnold ; Alcaim, Abraham(Advisor). **Independent Text Robust Speaker Recognition in the Presence of Noise Using PAC-MFCC and Sub-Band Classifiers**. Rio de Janeiro, 2011. 67p. MSc. Dissertation — Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

In this work is proposed the use of the PAC-MFCC feature with Sub-band Classifiers for the task of text-independent speaker identification in noise. The proposed scheme is compared with the features MFCC (Mel-Frequency Cepstral Coefficients), PAC-MFCC (Phase Autocorrelation MFCC) without sub-band classifiers, SSCH (Subband Spectral Centroid Histograms) and TECC (Teager Energy Cepstrum Coefficients). In this recognition task, we used the TIMIT database which consists of 630 speakers, where every one of them speak 10 utterances of 3 seconds each one approximately, of which eight utterance were used for training and two for testing, thus obtaining a total of 1260 test utterance for the recognition. We investigated the performance of these techniques using different types of noise from the base Noisex 92 with different signal to noise ratios. It was found that the accuracy rate of the PAC-MFCC feature with Sub-band Classifiers performs better in comparison with other techniques at a lower signal noise (less than 10dB).

Keywords

Speaker Independent Reconigition; PAC-MFCC with Sub-band Classifier; SSCH; PAC-MFCC; TECC; MFCC.

Sumário

1	Introdução	13
1.1	Motivação e Objetivos da Dissertação	14
1.2	Organização da Dissertação	14
2	Reconhecimento Automático De Locutor	15
2.1	Pré-Processamento	17
2.2	Extração de Características	18
2.3	Sistemas de Classificação	28
3	Proposta do Atributo PAC-MFCC com Classificador em Sub-Bandas para o Reconhecimento de Locutor Independente do Texto em Presença de Ruído	29
3.1	Fase AutoCorrelação com Coeficientes Cepstrais de Frequência Mel (PAC-MFCC)	30
3.2	PAC-MFCC com Classificador em Sub-Bandas	33
4	Resultados de Simulação e Avaliação das Técnicas Utilizadas para o Reconhecimento de Locutor Independente do Texto em Presença de Ruído	36
4.1	Caracterização da Base de Dados	36
4.2	Base de Ruídos Noisex-92	37
4.3	Avaliação Experimental	38
4.4	Comparação dos porcentagem de acerto obtidos de cada um dos atributos utilizados	47
5	Conclusões e Sugestões para Trabalhos Futuros	60
5.1	Conclusões	60
5.2	Sugestões para trabalhos futuros	61
	Referências Bibliográficas	62
A	Resultados Experimentais Detalhados	65
B	Tempo de Simulação por Atributo	67

Lista de figuras

2.1	Etapas de um sistema genérico para reconhecimento de locutor	16
2.2	Janelamento no domínio do tempo	19
2.3	Banco de Filtros triangulares	22
2.4	Diagrama em blocos para o cálculo dos MFCCs	22
2.5	diagrama do SSCH	24
2.6	Banco de filtros de 30 filtros Gammatone	27
3.1	FFT de um quadro da palavra "one"	32
3.2	Espectro PAC de um quadro da palavra "one"	33
3.3	Diagrama em blocos do atributo PAC-MFCC	33
3.4	Diagrama em blocos da etapa de treinamento do atributo PAC-MFCC com Classificador em Sub-bandas	34
3.5	Diagrama em blocos da etapa de teste do atributo PAC-MFCC com Classificador em Sub-bandas	35
4.1	Forma de onda para 60 mseg de ruído: a)fábrica, b)pink, c)branco, d)falatório	38
4.2	Comparação do atributo MFCC com o atributo TECCs em presença de diferentes tipos de ruído e de relação sinal ruído SNR	48
4.3	Comparação do atributo SSCH com o atributo MFCC em presença de diferentes tipos de ruído e de relação sinal ruído SNR	49
4.4	Comparação do atributo SSCH com o atributo TECC em presença de diferentes tipos de ruído e de relação sinal ruído SNR	50
4.5	Comparação do atributo PAC-MFCC com o atributo MFCC em presença de diferentes tipos de ruído e de relação sinal ruído SNR	51
4.6	Comparação do atributo PAC-MFCC com o atributo SSCH em presença de diferentes tipos de ruído e de relação sinal ruído SNR	52
4.7	Comparação do atributo PAC-MFCC com o atributo TECC em presença de diferentes tipos de ruído e de relação sinal ruído SNR	53
4.8	Comparação do atributo PAC-MFCC com Classificador em Sub-bandas com o atributo MFCC em presença de diferentes tipos de ruído e de relação sinal ruido SNR	54
4.9	Comparação do atributo PAC-MFCC com Classificador em Sub-bandas com o atributo SSCH em presença de diferentes tipos de ruído e de relação sinal ruído SNR	55
4.10	Comparação do atributo PAC-MFCC com Classificador em Sub-bandas com o atributo TECC em presença de diferentes tipos de ruído e de relação sinal ruido SNR	56
4.11	Comparação do atributo PAC-MFCC com Classificador em Sub-bandas com o atributo PAC-MFCC em presença de diferentes tipos de ruído e de relação sinal ruido SNR	57
4.12	Comparação das técnicas MFCC, SSCH, TECC, PAC-MFCC e PAC-MFCC com Classificador em Sub-bandas em ruído de fábrica com diferentes relação sinal ruído SNR	58

4.13	Comparação das técnicas MFCC, SSCH, TECC, PAC-MFCC e PAC-MFCC com Classificador em Sub-bandas em ruído falatório com diferente relação sinal ruído SNR	58
4.14	Comparação das técnicas MFCC, SSCH, TECC, PAC-MFCC e PAC-MFCC com Classificador em Sub-bandas em ruído de branco com diferentes relação sinal ruído SNR	59
4.15	Comparação das técnicas MFCC, SSCH, TECC, PAC-MFCC e PAC-MFCC com Classificador em Sub-bandas em ruído pink com diferentes relação sinal ruído SNR	59

Lista de tabelas

4.1	Estruturas dos subdiretórios da base TIMIT.	37
4.2	Desempenho das características do SSCH para diferentes valores do parâmetro de faixa dinâmica γ . A avaliação foi feita com a base TIMIT tanto para o sinal limpa e para o sinal com ruído de fábrica com diferentes SNR.	40
4.3	Desempenho das características do SSCH para diferentes valores de numero de filtros e número de bins. A avaliação foi feita com a base de dados TIMIT para locuções limpas, com ruído branco e ruído de fábrica para diferente relação sinal ruído SNR.	41
4.4	Desempenho das características SSCH utilizando a primeira e segunda derivada (delta e delta-delta). A avaliação foi feita com a base de dados TIMIT para locuções limpas, com ruído pink e ruído de fábrica para diferente relação sinal ruído SNR.	42
4.5	Desempenho das características MFCC para diferentes valores de número de filtros. A avaliação foi feita com a base de dados TIMIT para locuções limpas, com ruído pink e ruído de fábrica para diferente relação sinal ruído SNR.	42
4.6	Comparação entre os distintos número de coeficientes cepstrais utilizados para a obtenção dos porcentagem de acertos da técnica MFCC. A avaliação foi feita com a base de dados TIMIT para locuções limpas, com ruído branco e ruído de fábrica para diferente relação sinal ruído SNR.	43
4.7	Desempenho das características MFCC utilizando a primeira e segunda derivada (delta e delta-delta) com um número de filtros igual a 24. A avaliação foi feita com a base de dados TIMIT para locuções limpas, com ruído pink e ruído de fábrica para diferente relação sinal ruído SNR.	43
4.8	Desempenho das características da técnica TECC para diferentes valores de número de filtros. A avaliação foi feita com a base de dados TIMIT para locuções limpas, com ruído branco e ruído de fábrica para diferente relação sinal ruído SNR.	44
4.9	Comparação entre os distintos número de coeficientes cepstrais utilizados para a obtenção dos porcentagem de acertos da técnica TECC. A avaliação foi feita com a base de dados TIMIT para locuções limpas, com ruído branco e ruído pink para diferente relação sinal ruído SNR.	45
4.10	Desempenho das características TECC utilizando a primeira e segunda derivada (delta e delta-delta) com um número de filtros igual a 34. A avaliação foi feita com a base de dados TIMIT para locuções limpas, com ruído branco e ruído de fábrica para diferente relação sinal ruído SNR.	45

4.11	Comparação entre os distintos número de coeficientes cepstrais utilizados para a obtenção dos porcentagem de acertos da técnica PAC-MFCC. A avaliação foi feita com a base de dados TIMIT para locuções limpas, com ruído branco e ruído pink para diferente relação sinal ruído SNR.	46
4.12	Desempenho das características PAC-MFCC utilizando a primeira e segunda derivada (delta e delta-delta). A avaliação foi feita com a base de dados TIMIT para locuções limpas, com ruído branco e ruído pink para diferente relação sinal ruído SNR.	46
4.13	Desempenho das características PAC-MFCC com Classificador em Sub-bandas para 2 e 4 sub-bandas. A avaliação foi feita com a base de dados TIMIT para locuções limpas, com ruído de fábrica e ruído pink para diferente relação sinal ruído SNR.	47
A.1	Comparação dos porcentagens de acertos obtidas pelos atributos utilizados para o reconhecimento automático del locutor independente do texto, a avaliação foi feita com a base de dados TIMIT para locuções limpas, e com ruído de fábrica para diferente relação sinal ruído SNR.	65
A.2	Comparação dos porcentagens de acertos obtidas pelos atributos utilizados para o reconhecimento automático del locutor independente do texto, a avaliação foi feita com a base de dados TIMIT para locuções limpas, e com ruído falatório para diferente relação sinal ruído SNR.	65
A.3	Comparação dos porcentagens de acertos obtidas pelos atributos utilizados para o reconhecimento automático del locutor independente do texto, a avaliação foi feita com a base de dados TIMIT para locuções limpas, e com ruído branco para diferente relação sinal ruído SNR.	66
A.4	Comparação dos porcentagens de acertos obtidas pelos atributos utilizados para o reconhecimento automático de locutor independente do texto, a avaliação foi feita com a base de dados TIMIT para locuções limpas, e com ruído de pink para diferente relação sinal ruído SNR.	66
B.1	Comparação dos tempos de simulação obtidos pelos atributos utilizados para o reconhecimento automático del locutor independente do texto.	67