



Christian Dayan Arcos Gordillo

**Reconhecimento de Voz Contínua Combinando
os Atributos MFCC e PNCC com Métodos de
Robustez SS, WD, MAP e FRN**

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
Março de 2013



Christian Dayan Arcos Gordillo

**Reconhecimento de Voz Contínua Combinando
os Atributos MFCC e PNCC com Métodos de
Robustez SS, WD, MAP e FRN**

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. Flávia Magalhães Freitas Ferreira

PUC-Minas

Prof. Marco Antônio Grivet

Departamento de Engenharia Elétrica - PUC-Rio

Prof. José Eugênio Leal

Coordenador Setorial do Centro Técnico Científico - PUC-Rio

Rio de Janeiro, 8 de Março de 2013

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

Christian Dayan Arcos Gordillo

Graduou-se em Engenharia Eletrônica pela Universidade Francisco de Paula Santander (San José de Cúcuta, Colômbia).

Ficha Catalográfica

Arcos Gordillo, Christian Dayan

Reconhecimento de Voz Contínua Combinando os Atributos MFCC e PNCC com Métodos de Robustez SS, WD, MAP e FRN/ Christian Dayan Arcos Gordillo; orientador: Abraham Alcaim. — Rio de Janeiro : PUC–Rio, Departamento de Engenharia Elétrica, 2013.

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

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

Inclui referências bibliográficas.

1. Engenharia Elétrica – Tese. 2. Reconhecimento de Voz Contínua Combinando os Atributos MFCC e PNCC. 3. Métodos de Robustez SS, WD, MAP e FRN. 4. Realce de fala. 5. Compensação de atributos. 6. Pré-extração de atributos. 7. Pós-extração de atributos. I. Alcaim, Abraham. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Elétrica. III. Título.

CDD: 621.3

De forma muito especial quero dedicar este sonho cumprido a meus pais
Guilmar Arcos e Luz Marina Gordillo que tem sido um exemplo de luta,
superação e perseverança. A vocês, amorosamente dedico esta vitória de
minha vida, porque me ensinaram com humildade que nos momentos mais
difíceis é quando devo ser mais forte e persistente.

E Sarita o amor de minha vida e Princesa de todos os meus contos
OS AMO

Agradecimentos

Esta dissertação de mestrado é o resultado de meus últimos dois anos de trabalho no Centro de Estudos de Telecomunicações (CETUC), onde através de luta e sacrifício eu atinjo mais uma etapa neste processo chamado vida. É por isso que desejo primeiramente e acima de tudo, dar infinitas graças a **DEUS** por estar comigo em cada passo que dou, por fortalecer meu coração e iluminar minha mente neste árduo caminho de aprendizado.

Desejo expressar minha mais sincera gratidão ao meu orientador Abraham Alcaim pela amizade, pela orientação, por sempre mostrar boa vontade e pelo tempo dedicado ao projeto.

Eu digo muito obrigado aos meus pais, que levam se todos os créditos porque sacrificaram a maior parte da sua vida para me educar, nunca poderei pagar todos os seu esforços, todas as noites que não conseguiram pegar o sonho visando me tornar em um homem de bem, e por isso que todos os meus triunfos são para vocês.

A minha princesa Sarita, fonte de minha inspiração e de minhas energias inesgotáveis, obrigado por teu apoio incondicional, porque teu amor e ternura, mesmo na distância, foram sempre motivo de esperança e alento.

Aos meus irmãos por todo o apoio e paciência, sempre os levo no meu coração.

A Lorena Chamorro por seu apoio e votos de sucesso sempre desejados e por sua amizade que tem transcendido à irmandade.

Aos meus parceiros do laboratório de sistemas de comunicações, por seu apoio e inestimável ajuda neste longo período de tempo, e por tornar o dia-a-dia mais suportável.

Finalmente, Gostaria de agradecer ao Governo Brasileiro, à Pontifícia Universidade Católica de Rio de Janeiro (PUC-Rio), e o apoio financeiro provido pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES sem os quais este trabalho não poderia ter sido realizado.

Resumo

Arcos Gordillo, Christian Dayan; Alcaim, Abraham. **Reconhecimento de Voz Contínua Combinando os Atributos MFCC e PNCC com Métodos de Robustez SS, WD, MAP e FRN**. Rio de Janeiro, 2013. 101p. Dissertação de Mestrado — Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

O crescente interesse por imitar o modelo que rege o processo cotidiano de comunicação humana através de máquinas tem se convertido em uma das áreas do conhecimento mais pesquisadas e de grande importância nas últimas décadas. Esta área da tecnologia, conhecida como reconhecimento de voz, tem como principal desafio desenvolver sistemas robustos que diminuam o ruído aditivo dos ambientes de onde o sinal de voz é adquirido, antes de que se esse sinal alimente os reconhecedores de voz. Por esta razão, este trabalho apresenta quatro formas diferentes de melhorar o desempenho do reconhecimento de voz contínua na presença de ruído aditivo, a saber: *Wavelet Denoising* e *Subtração Espectral*, para realce de fala e *Mapeamento de Histogramas* e *Filtro com Redes Neurais*, para compensação de atributos. Esses métodos são aplicados isoladamente e simultaneamente, afim de minimizar os desajustes causados pela inserção de ruído no sinal de voz. Além dos métodos de robustez propostos, e devido ao fato de que os reconhecedores de voz dependem basicamente dos atributos de voz utilizados, examinam-se dois algoritmos de extração de atributos, MFCC e PNCC, através dos quais se representa o sinal de voz como uma sequência de vetores que contêm informação espectral de curtos períodos de tempo. Os métodos considerados são avaliados através de experimentos usando os software HTK e Matlab, e as bases de dados TIMIT (de vozes) e NOISEX-92 (de ruído). Finalmente, para obter os resultados experimentais, realizam-se dois tipos de testes. No primeiro caso, é avaliado um sistema de referência baseado unicamente em atributos MFCC e PNCC, mostrando como o sinal é fortemente degradado quando as razões sinal-ruído são menores. No segundo caso, o sistema de referência é combinado com os métodos de robustez aqui propostos, analisando-se comparativamente os resultados dos métodos quando agem isolada e simultaneamente. Constata-se que a mistura simultânea dos métodos nem sempre é mais atraente. Porém, em geral o melhor resultado é obtido combinando-se MAP com atributos PNCC.

Palavras-chave

Reconhecimento de voz; robustez; sinal; realce; compensação; atributos; wavelet denoising; mapeamento de histogramas; subtração espectral; redes neurais; MFCC; PNCC.

Abstract

Arcos Gordillo, Christian Dayan; Alcaim, Abraham (Advisor). **Continuous Speech Recognition by Combining MFCC and PNCC Attributes with SS, WD, MAP and FRN Methods of Robustness.** Rio de Janeiro, 2013. 101p. MSc. Dissertation — Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

The increasing interest in imitating the model that controls the daily process of human communication through machines has become one of the most researched areas of knowledge and of great importance in recent decades. This technological area known as voice recognition has as a main challenge to develop robust systems that reduce the noisy additive environment where the signal voice was acquired. For this reason, this work presents four different ways to improve the performance of continuous speech recognition in presence of additive noise, known as Wavelet Denoising and Spectral Subtraction for enhancement of voice, and Mapping of Histograms and Filter with Neural Networks to compensate for attributes. These methods are applied separately and simultaneously two by two, in order to minimize the imbalances caused by the inclusion of noise in voice signal. In addition to the proposed methods of robustness and due to the fact that voice recognizers depend mainly on the attributes voice used, two algorithms are examined for extracting attributes, MFCC, and PNCC, through which represents the voice signal as a sequence of vectors that contain spectral information for short periods of time. The considered methods are evaluated by experiments using the HTK and Matlab software, and databases of TIMIT (voice) and Noisex-92 (noise). Finally, for the experimental results, two types of tests were carried out. In the first case a reference system was assessed based on MFCC and PNCC attributes, only showing how the signal degrades strongly when signal-noise ratios are higher. In the second case, the reference system is combined with robustness methods proposed here, comparatively analyzing the results of the methods when they act alone and simultaneously. It is noted that simultaneous mix of methods is not always more attractive. However, in general, the best result is achieved by the combination of MAP with PNCC attributes.

Keywords

Speech recognition; robustness; signal; enhancement; compensation; attributes; wavelet denoising; histogram mapping; spectral subtraction; neural network, MFCC, PNCC.

Sumário

| | | |
|-----|--|-----------|
| 1 | Introdução | 13 |
| 1.1 | Motivação | 14 |
| 1.2 | Objetivos da dissertação | 15 |
| 1.3 | Estrutura da dissertação | 16 |
| 2 | Fundamentos do Reconhecimento de Voz | 17 |
| 2.1 | A comunicação oral | 17 |
| 2.2 | O sinal de voz | 18 |
| 2.3 | Problemas do reconhecimento de voz | 24 |
| 2.4 | Estrutura dos sistemas de reconhecimento de voz | 26 |
| 2.5 | Modelos ocultos de Markov - HMM aplicados ao reconhecimento de voz contínua | 31 |
| 2.6 | Reconhecimento de voz contínua | 37 |
| 3 | Os Atributos MFCC e PNCC do Sinal de Voz | 41 |
| 3.1 | Mel-Frequency Cepstral Coefficients (MFCC) | 43 |
| 3.2 | Power-Normalized Cepstral Coefficients (PNCC) | 46 |
| 4 | Robustez do Reconhecimento de Voz | 49 |
| 4.1 | Reconhecimento de voz em presença de ruído | 49 |
| 4.2 | Técnicas de robustez para o reconhecimento de voz em presença de ruído | 51 |
| 5 | Projeto de um Sistema de Reconhecimento de Voz Contínua através de Técnicas de pré-extração de Atributos | 54 |
| 5.1 | Subtração Espectral | 54 |
| 5.2 | Wavelet Denoising | 59 |
| 5.3 | Avaliação das técnicas pré-extração de atributos | 64 |
| 6 | Projeto de um Sistema de Reconhecimento de Voz Contínua através de Técnicas de pós-extração de Atributos | 70 |
| 6.1 | Mapeamento de Histogramas | 70 |
| 6.2 | Filtro com Redes Neurais | 75 |
| 6.3 | Avaliação das técnicas de pós-extração de atributos | 80 |
| 7 | Conclusões e Sugestões para Trabalhos Futuros | 84 |
| 7.1 | Conclusões | 84 |
| 7.2 | Sugestões para trabalhos futuros | 87 |
| A | Algoritmo de Baum-Welch | 94 |
| B | Algoritmo de Viterbi | 96 |
| C | Matlab | 97 |

| | | |
|-----|----------------------------|-----|
| D | HTK | 98 |
| E | Ruído branco gaussiano | 100 |
| E.1 | Confecção do Sinal Ruidoso | 100 |

Lista de figuras

| | | |
|------|---|----|
| 2.1 | Processo de comunicação oral | 18 |
| 2.2 | Estrutura do aparelho fonador | 19 |
| 2.3 | Cordas vocais (a) Glótis aberta e cordas vocais separadas gerando sons surdos. (b) Glótis fechada e cordas vocais em vibração gerando sons sonoros | 20 |
| 2.4 | Triângulo vogais de Hellwag. | 20 |
| 2.5 | Formas de onda dos sons sonoros e surdos (a) fonema /sh/ (b) fonema /ix/. | 22 |
| 2.6 | (a) Aspecto de sinal de voz no domínio do tempo; (b) e (c) análise com janelas de 25 ms, sinal quasiestacionario. | 22 |
| 2.7 | Diagrama de blocos geral de um sistema de reconhecimento. | 27 |
| 2.8 | Segmento janelado com hamming | 29 |
| 2.9 | Representação de esquerda a direita do HMM | 32 |
| 2.10 | Sequência de operações para qualquer variável $\alpha(i)$ para frente (forward). | 36 |
| 3.1 | Comparação dos métodos de extração de atributos | 42 |
| 3.2 | Banco de filtros usado na técnica MFCC | 44 |
| 3.3 | Banco de filtros Gammatone. | 47 |
| 4.1 | Diagrama de blocos do modelo de ambiente acústico. | 50 |
| 4.2 | Restauração do sinal através de técnicas de de realce de fala | 51 |
| 4.3 | Restauração de atributos através de técnicas de compensação. | 52 |
| 4.4 | Distribuição das técnicas para o reconhecimento robusto de voz. | 53 |
| 5.1 | Diagrama de blocos do processo de subtração espectral. | 55 |
| 5.2 | Diagrama de fluxo do método VAD. | 58 |
| 5.3 | Diagrama de blocos da técnica wavelet denoising. | 59 |
| 5.4 | Transformada <i>Wavelet</i> no domínio tempo- frequência. | 60 |
| 5.5 | Diagrama de decomposição de sinais através de transformadas <i>wavelet</i> . | 61 |
| 5.6 | Comparação do sinal senoidal e sinal <i>wavelet</i> . | 62 |
| 5.7 | Comparação dos gráficos da função <i>hard thresholding</i> e a função <i>soft thresholding</i> . | 62 |
| 6.1 | Distorção do espaço de representação com ruído branco a 10 db. (a) MFCC (b) PNCC. | 71 |
| 6.2 | Mapeamento de histogramas do coeficiente C_0 dos atributos MFCC (a) <i>fdp</i> do coeficiente cepstral original (b) <i>fdp</i> do coeficiente cepstral mapeado. | 72 |
| 6.3 | Processo de mapeado por através da estatística ordenada. | 74 |
| 6.4 | Modelo não linear de um neurônio. | 77 |
| 6.5 | Rede neural <i>feedforward</i> , com 4 camadas formadas por conexões entre neurônios artificiais. | 79 |
| 6.6 | Modelo de aprendizado. | 79 |

| | | |
|-----|--|----|
| A.1 | Diagrama de fluxo do método EM. | 95 |
| C.1 | Interface gráfica do programa Matlab. | 97 |
| D.1 | Interface gráfica ao invocar a ferramenta Hslab. | 99 |

Lista de tabelas

| | | |
|-----|---|----|
| 2.1 | Parâmetros típicos que caracterizam o sistema de reconhecimento de voz. | 24 |
| 5.1 | Taxas de acerto do sistema de referência. | 67 |
| 5.2 | Taxas de acerto utilizando as técnicas pré-extração de atributos. | 67 |
| 5.3 | Taxas de acerto utilizando a mistura de técnicas pré-extração de atributos | 68 |
| 6.1 | Relação entre o cérebro e as redes neurais artificiais. | 75 |
| 6.2 | Configuração da rede neural. | 80 |
| 6.3 | Taxas de acerto utilizando as técnicas pós-extração de atributos. | 81 |
| 6.4 | Taxas de acerto utilizando a mistura das técnicas pós-extração de atributos. | 82 |
| D.1 | Grupo de ferramentas de HTK utilizadas nas aplicações de reconhecimento de voz. | 99 |