Pontifícia Universidade Católica
do Rio de Janeiro

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

Pontifícia Universidade Católica
do Rio de Janeiro

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 Cientifico - PUC-Rio

Rio de Janeiro, 8 de Março de 2013

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 infintas 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 poque 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 trough 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; enhacement; compensation; attributes; wavelet denoising; histogram mapping; spectral subtraction; neural network, MFCC, PNCC.

Sumário

Lista de figuras

Lista de tabelas