

**Vladimir Fabregas Surigué de
Alencar**

**Atributos e Domínios de
Interpolação Eficientes em
Reconhecimento de Voz
Distribuído**

DISSERTAÇÃO DE MESTRADO

DEPARTAMENTO DE ENGENHARIA ELÉTRICA
Programa de Pós-graduação em
Engenharia Elétrica

Rio de Janeiro
Março de 2005



Vladimir Fabregas Surigué de Alencar

**Atributos e Domínios de Interpolação Eficientes em
Reconhecimento de Voz Distribuído**

Dissertação de Mestrado

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

Orientador: Prof. Abraham Alcaim

Rio de Janeiro
março de 2005



Vladimir Fabregas Surigué de Alencar

**Atributos e Domínios de Interpolação
Eficientes em Reconhecimento
de Voz Distribuído**

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

Centro de Estudos em Telecomunicações - PUC-Rio

Prof. Fernando Gil Vianna Resende Jr.

UFRJ

Prof. Rodrigo Caiado de Lamare

Centro de Estudos em Telecomunicações - PUC-Rio

Prof. José Eugenio Leal

Coordenador Setorial do Centro

Técnico Científico - PUC-Rio

Rio de Janeiro, 18 de março de 2005

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

Vladimir Fabregas Surigué de Alencar

Graduou-se em Engenharia de Telecomunicações na UFF (Universidade Federal Fluminense) em 2003.

Ficha Catalográfica

Alencar, Vladimir Fabregas Surigué de

Atributos e domínios de intepolação eficientes em reconhecimento de voz distribuído / Vladimir Fabregas Surigué de Alencar ; orientador: Abraham Alcaim. - Rio de Janeiro: PUC-Rio, Departamento de Engenharia Elétrica, 2005.

114 f.: il. ; 30 cm

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

Incluí referências bibliográficas.

1. Engenharia elétrica - Teses. 2. Reconhecimento de voz distribuído 3. LSF 4. ITU-T G.723.1. 5. HMM. 6. Redes IP 6. Redes móveis celulares. I. Alcaim, Abraham. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Elétrica. III. Título.

CDD:621.3

Para meus pais Juarez e Laura, minha irmã Tatiana
e minha namorada Daniele
pelo carinho, apoio e confiança.

Agradecimentos

Ao meu orientador, Professor Abraham Alcaim, pela oportunidade, apoio e incentivo para a realização deste trabalho.

Ao corpo docente do CETUC, pelo aprendizado proporcionado.

À CAPES e à PUC-Rio, pelos auxílios concedidos.

Agradeço de forma especial aos meus pais, que mais uma vez foram fundamentais na minha vida, à minha irmã e à Daniele, que estiveram ao meu lado nesta empreitada.

Aos Professores que participaram da minha comissão examinadora.

Ao Aureo por todos os ensinamentos trocados e contribuições que proporcionaram o sucesso deste trabalho.

A todos os amigos que fiz no CETUC, que me proporcionaram não apenas momentos de aprendizagem, mas momentos de companheirismo que espero que se perpetuem.

Resumo

Alencar, Vladimir Fabregas Surigué; Alcaim, Abraham. **Atributos e Domínios de Interpolação Eficientes em Reconhecimento de Voz Distribuído**. Rio de Janeiro, 2005. xxxp. Dissertação de Mestrado - Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

Com o crescimento gigantesco da Internet e dos sistemas de comunicações móveis celulares, as aplicações de processamento de voz nessas redes têm despertado grande interesse. Um problema particularmente importante nessa área consiste no reconhecimento de voz em um sistema servidor, baseado nos parâmetros acústicos calculados e quantizados no terminal do usuário (Reconhecimento de Voz Distribuído). Como em geral estes parâmetros não são os mais indicados como atributos de voz para o sistema de reconhecimento remoto, é importante que sejam examinadas diferentes transformações dos parâmetros, que permitam um melhor desempenho do reconhecedor. Esta dissertação trata da extração de atributos de reconhecimento eficientes a partir dos parâmetros dos codificadores utilizados em redes móveis celulares e em redes IP. Além disso, como a taxa dos parâmetros fornecidos ao reconhecedor de voz é normalmente superior àquela com a qual os codificadores geram os parâmetros, é importante analisar o efeito da interpolação dos parâmetros sobre o desempenho do sistema de reconhecimento, bem como o melhor domínio sobre o qual esta interpolação deve ser realizada. Estes são outros tópicos apresentados nesta dissertação.

Palavras-chave

Reconhecimento de Voz Distribuído; LSF; ITU-T G.723.1; HMM; Redes IP; Redes Móveis Celulares

Abstract

Alencar, Vladimir Fabregas Surigué; Alcaim, Abraham (Advisor). **Efficient Features and Interpolation Domains in Distributed Speech Recognition.** Rio de Janeiro, 2005. xxxp. Msc. Dissertation - Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

The huge growth of the Internet and cellular mobile communication systems has stimulated a great interest in the applications of speech processing in these networks. An important problem in this field consists in speech recognition in a server system, based on the acoustic parameters calculated and quantized in the user terminal (Distributed Speech Recognition). Since these parameters are not the most indicated ones for the remote recognition system, it is important to examine different transformations of these parameters, in order to allow a better performance of the recogniser. This dissertation is concerned with the extraction of efficient recognition features from the coder parameters used in cellular mobile networks and IP networks. In addition, as the rate that parameters supplied for the speech recogniser must be usually higher than that generated by the codec, it is important to analyze the effect of the interpolation of the parameters over the performance of the recognition system. Moreover, it is paramount to establish the best domain over which this interpolation must be carried out. These are other topics presented in this dissertation.

Palavras-chave

Distributed Speech Recognition; LSF; ITU-T G.723.1; HMM; IP Networks; Cellular Mobile Networks

Sumário

1 Introdução	13
1.1. Sistemas de Reconhecimento de Voz em Ambiente Celular/Redes IP	13
1.2. Organização do Trabalho	15
2 Sistemas de Reconhecimento de Voz	17
2.1. Características e Modelo de Produção do Sinal de Voz	20
2.1.1. Sons da Fala	20
2.1.2. Modelo de Produção da Fala	27
2.2. Reconhecimento Automático de Voz através do Modelo de Markov Escondido	28
2.2.1. Elementos de um HMM	31
2.2.2. Os Três Problemas Básicos dos HMMs	35
2.3. Conclusão	45
3 Reconhecimento de Voz Distribuído	46
3.1. Atributos mais Utilizados em Reconhecimento de Voz Distribuído	47
3.2. Estrutura do Sistema de Reconhecimento Distribuído Objeto da Dissertação	52
3.3. Conclusão	55
4 Atributos de Voz para Reconhecimento	56
4.1. Pré-Processamento	57
4.2. Atributos Extraídos de Voz Reconstruída	61
4.2.1. Coeficientes Mel-Cepstrais (MFCC)	61
4.3. Atributos Extraídos dos Parâmetros LPC	67
4.3.1. LPC <i>Cepstrum</i> (LPCC)	68
4.3.2. <i>Mel-Frequency</i> LPCC (MLPCC)	69
4.4. Atributos Extraídos das LSFs	71
4.4.1. Pseudo-Cepstral Coeficientes (PCC)	73

4.4.2. Pseudo-Cepstrum (PCEP)	75
4.4.3. Mel-Frequency PCC (MPCC)	75
4.4.4. Mel-Frequency PCEP (MPCEP)	76
4.5. Conclusão	76
 5 Análise dos Atributos de Voz em Reconhecimento Distribuído	 77
5.1. Obtenção dos Atributos de Voz em 10ms e 20ms e o Desempenho do Reconhecimento para Cada Taxa	80
5.2. Interpolação de Atributos de Voz para Reconhecimento	83
5.3. Conclusão	84
 6 Análise dos Atributos de Voz em Reconhecimento Distribuído com a Utilização do <i>Codec</i> de Voz ITU-T G.723.1	 86
6.1. Características do <i>Codec</i> ITU-T G.723.1	87
6.2. Desempenho do Sistema de Reconhecimento com o Uso do <i>Codec</i> ITU-T G.723.1	90
6.3. Conclusão	92
 7 Conclusões e Sugestões para Trabalhos Futuros	 93
7.1. Conclusões	93
7.2. Sugestões para Trabalhos Futuros	96
 Referências Bibliográficas	 98
 Apêndice	 101
A.1. Preparação dos Dados	102
A.2. Treino	104
A.3. Teste	110
A.4. Análise dos resultados	112

Lista de figuras

Figura 2.1 – Formas de onda de sons sonoro e surdo	21
Figura 2.2 – Fases sucessivas em um período de vibração das cordas vocais	22
Figura 2.3 – Diagrama de blocos do modelo de produção da fala	27
Figura 2.4 – As duas fases de um sistema de reconhecimento de fala	29
Figura 2.5 – Exemplo de HMM usado para modelar o sinal de voz	31
Figura 2.6 – Esquema de um sistema de reconhecimento de palavras isoladas que emprega HMM.	36
Figura 2.7 – Detalhamento da fase de reconhecimento	36
Figura 3.1 Sistemas de Reconhecimento Distribuído – Diagrama Básico	46
Figura 3.2 – Sistema de reconhecimento de voz distribuído baseado nos parâmetros de voz do codificador	47
Figura 3.3 – Parâmetros do codificador de voz	48
Figura 3.4 – Atributos extraídos dos coeficientes LPC	48
Figura 3.5 – Atributos PCC e MPCC obtidos de parâmetros LSF quantizados	49
Figura 3.6 – Atributos PCEP e MPCEP obtidos de parâmetros LSF quantizados	49
Figura 3.7 – Sistema de reconhecimento de voz distribuído baseado em voz decodificada	50
Figura 3.8 – Atributos obtidos de voz reconstruída	50
Figura 3.9 – Sistema de reconhecimento de voz distribuído com codificação dos atributos de reconhecimento no <i>front-end</i> local	51
Figura 3.10 - Sistema de reconhecimento de voz distribuído com codificação dos atributos de reconhecimento e codificação de voz no <i>front-end</i> local	51
Figura 3.11 – Sistema de reconhecimento distribuído objeto da dissertação	52
Figura 4.1 – Divisão em quadros do sinal de voz.	58

Figura 4.2 – Percepção subjetiva da frequência fundamental de sons sonoros.	62
Figura 4.3 – Magnitude do espectro dos filtros de banda crítica.	63
Figura 4.4 – Aproximação da escala mel pela transformação bilinear, escala mel em preto e aproximação em cinza.	71
Figura 5.1 – Sistema para análise sem quantização	77
Figura 6.1 – Sistema de reconhecimento de voz distribuído a ser utilizado para o teste com uso do <i>codec</i> ITU-T G.723.1	86
Figura 6.2 – Diagrama de blocos do codificador de voz	88
Figura 6.3 – Diagrama de bloco do decodificador de voz	90
Figura A.1 – Conjunto de ferramentas HTK	101

Lista de tabelas

Tabela 2.1 – Fonemas do Português brasileiro	21
Tabela 4.1 – Freqüências dos centros e banda crítica dos filtros utilizados para cálculo dos coeficientes mel-cepestrais.	64
Tabela 5.1 – Resultado do teste de reconhecimento de voz para os atributos obtidos a cada 10 ms em porcentagem de acertos	79
Tabela 5.2 – Resultado do teste de reconhecimento de voz para os atributos obtidos a cada 20 ms em porcentagem de acertos	80
Tabela 5.3 – Resultado do teste de reconhecimento de voz para os atributos obtidos a cada 10 ms em porcentagem de acertos	81
Tabela 5.4 – Resultado do teste de reconhecimento de voz para os atributos obtidos a cada 20 ms em porcentagem de acertos	81
Tabela 5.5 – Resultado do teste de reconhecimento de voz para HMMs de cinco estados	82
Tabela 5.6 – Resultado do teste de reconhecimento de voz com ou sem interpolação.	83
Tabela 6.1 – Tabela de alocação de bits para o codificador operando a 6,3 kb/s	89
Tabela 6.2 – Tabela de alocação de bits para o codificador operando a 5,3 kb/s	89
Tabela 6.3 – Resultados dos testes de reconhecimento com o <i>codec</i> ITU-T G.723.1	91