
Resumo da Dissertação apresentada à UFPE como parte dos requisitos necessários para a obtenção do grau de Mestre em Engenharia Elétrica.

NOVAS ABORDAGENS PARA CODIFICAÇÃO DE VOZ E RECONHECIMENTO AUTOMÁTICO DE LOCUTOR PROJETADAS VIA MASCARAMENTO PLENO EM FREQUÊNCIA POR OITAVA

Roberto Fernando Batista Sotero Filho

Outubro/2009

Orientador: Hélio Magalhães de Oliveira, Docteur

Área de Concentração: Telecomunicações

Palavras-chave: *vocoders*, reconhecimento de locutor, mascaramento em frequência.

Número de Páginas: 144

RESUMO

A área do processamento digital de sinais de voz (PDSV) é uma das áreas mais importantes do processamento digital de sinais. Como subáreas relevantes do PDSV estão a Codificação da Voz, e o Reconhecimento Automático de Locutor (RAL). Esta dissertação propõe uma nova abordagem para um *vocoder* baseado no Mascaramento Pleno em Frequência por Oitavas (MPFO) em adição a uma técnica de preenchimento espectral via distribuição beta de probabilidade. O método do MPFO consiste em simplificar a magnitude do espectro em frequência do sinal, obtido através da transformada rápida de Fourier (FFT), considerando apenas uma amostra por oitava. Tal abordagem, que oferece um compromisso entre taxa de bits (e.g. 2,7 kbits/s), complexidade, inteligibilidade e qualidade dos arquivos de voz, permitiu a criação de um novo formato binário de representação digital da voz: o formato *.voz*. Apresenta-se também um novo método de baixa complexidade computacional para RAL, baseando-se em uma das propriedades-chave da percepção auditiva humana: o mascaramento acústico em frequência. O vetor característico dos quadros do sinal de voz é representado pela fração média das amplitudes dos tons de mascaramento em cada oitava. Ambos os tipos de reconhecimento de locutor (de texto dependente e de texto independente) são estudados. Os resultados confirmam que o algoritmo proposto oferece um compromisso entre a complexidade e a taxa de identificações corretas (típico 85%), sendo atrativo para aplicações em sistemas embarcados.

Abstract of the Master Thesis presented at UFPE as a part of requirements for earning an Electrical Engineering Master degree.

NEW APPROACHES FOR SPEECH CODING AND AUTOMATIC SPEAKER RECOGNITION BY FULL FREQUENCY MASKING PER OCTAVE

Roberto Fernando Batista Sotero Filho

Outubro/2009

Supervisor: Hélio Magalhães de Oliveira, Docteur

Majoring: Telecommunications

Key-words: *vocoders*, automatic speaker identification, frequency masking.

Number of pages: 144

ABSTRACT

Speech digital signal processing (SDSP) is one of the most significant areas of digital signal processing. Voice coding and automatic speaker recognition (ASR) are relevant SDSP subfields. This thesis introduces a new approach for a vocoder scheme, which is based on the full frequency masking per octave (FFMO), jointly with a new spectral stuffing technique through the beta probability distribution. The FFMO method consists of simplifying the magnitude of the voice spectrum (derived by Fast Fourier transform - FFT) retaining just one spectral sample per octave. This approach, which offers a tradeoff between the bit rate (e.g., 2.7 kbits/s), complexity, intelligibility and voice quality, allows the development of a new file format, namely *.voz*. It is also presented a novel and low-complexity ASR technique. It is based on one of the key-properties of the human hearing perception: the auditory frequency masking. The feature vectors of voice frames are merely represented by the average amplitude of the greatest spectral samples within each octave. Both text-dependent and text-independent speaker recognition is investigated. Results support a tradeoff between recognition efficiency (typically 85%) and complexity of this kind of vocoder-based systems, being thereby attractive for embedded systems.