



INSTITUTO SUPERIOR DE ENGENHARIA DE LISBOA
Área Departamental de Engenharia de Electrónica e
Telecomunicações e de Computadores

ISEL

Agrupamento Multi-orador

Hugo Daniel Carolino Alves

Dissertação para obtenção do grau de Mestre em Engenharia de Electrónica e Computadores

Resumo:

Actualmente tem-se observado um aumento do volume de sinais de fala em diversas aplicações, que reforçam a necessidade de um processamento automático dos ficheiros. No campo do processamento automático destacam-se as aplicações de “diarização de orador”, que permitem catalogar os ficheiros de fala com a identidade de oradores e limites temporais de fala de cada um, através de um processo de segmentação e agrupamento.

No contexto de agrupamento, este trabalho visa dar continuidade ao trabalho intitulado “Detecção do Orador”, com o desenvolvimento de um algoritmo de “agrupamento multi-orador” capaz de identificar e agrupar correctamente os oradores, sem conhecimento prévio do número ou da identidade dos oradores presentes no ficheiro de fala.

O sistema utiliza os coeficientes “Mel Line Spectrum Frequencies” (MLSF) como característica acústica de fala, uma segmentação de fala baseada na energia e uma estrutura do tipo “Universal Background Model - Gaussian Mixture Model” (UBM-GMM) adaptado com o classificador “Support Vector Machine” (SVM).

No trabalho foram analisadas três métricas de discriminação dos modelos SVM e a avaliação dos resultados foi feita através da taxa de erro “Speaker Error Rate” (SER), que quantifica percentualmente o número de segmentos “fala” mal classificados.

O algoritmo implementado foi ajustado às características da língua portuguesa através de um corpus com 14 ficheiros de treino e 30 ficheiros de teste. Os ficheiros de treino dos modelos e classificação final, enquanto os ficheiros de foram utilizados para avaliar o desempenho do algoritmo.

A interacção com o algoritmo foi dinamizada com a criação de uma interface gráfica que permite receber o ficheiro de teste, processá-lo, listar os resultados ou gerar um vídeo para o utilizador confrontar o sinal de fala com os resultados de classificação.

Palavras-Chave – Coeficientes MLSF, Modelo de orador, UBM, Classificador SVM, “Agrupamento multi-orador”, Estimação do número de oradores, “Speaker Error Rate”.

Dezembro de 2010