



Dissertação: **Reconhecimento de Orador em Dois Segundos**

Mestrado Integrado em Engenharia Electrotécnica e de Computadores, ramo Telecomunicações

Orientador: Prof. Aníbal Ferreira

Relatório realizador por: Diana Rocha Mendes

Relatório N° 10/11

25 Abril 2011 a 8 Maio 2011

Tarefas Realizadas:

- Repetição de testes feitos com a base de dados TIMIT, agora com a base de dados Mocha TIMIT. A Mocha TIMIT apresenta uma maior quantidade de dados para cada orador (460 frases são lidas por cada orador, ou contrário das 10 lidas por cada orador da TIMIT). No entanto, só existem 3 oradores masculinos.
- Estudo do *likelihood ratio test*, e experimentação prática do mesmo através da função "lratiotest" do Matlab.
- Comparação entre os Gaussian Mixture Models construídos pela função da VoiceBox e pela função da Statistics Toolbox. Foi concluído que, em condições idênticas, os valores de log-likelihood são muito próximos e como tal, o uso das duas funções é equivalente para o estudo em questão.
- Análise gráfica dos valores de log-likelihood obtidos para cada ponto, através da função "gaussmixp" da toolbox Matlab. Surgiu a ideia para o teste descrito seguidamente.
- Teste de uma diferente medida de classificação: a variância dos valores de log-likelihood para cada ponto dos dados, ao invés da soma. Os resultados mostraram que essa variância pode atingir valores muito distintos entre o caso de "verdadeiro" orador e impostor, mas nem sempre isso se verifica e nem sempre há uma diferença considerável.
- Continuação de teste com a base de dados MOCHA TIMIT: foi fixado o tempo máximo de treino (70% dos dados disponíveis para cada orador) e variado o tempo de teste (2 minutos, 1 minutos, 30 segundos, 15 segundos, 7 segundos e 3 segundos), com o intuito de estabelecer um limite no tempo de teste que permite uma diferença entre os valores de log-likelihood razoável.
- Revisão com o orientador da calendarização das tarefas a realizar até ao final da dissertação.

- Pesquisa de bases de dados que possuam anotações com os fonemas ou anotações semelhantes, que permitam a segmentação da parte vozeada da voz numa fase posterior do trabalho. Concluiu-se que a TIMIT possui essas anotações, e da MOCHA TIMIT só foram encontradas essas *labels* para 2 dos 3 oradores masculinos.

Dificuldades Encontradas:

- Poucas bases de dados disponíveis com os dados que pretendemos. Visto que só a TIMIT e a MOCHA TIMIT podem ser usadas, tem que haver um compromisso entre usar poucos oradores ou poucos dados para cada orador, visto que nenhuma das duas apresenta simultaneamente um número elevado de oradores e uma grande quantidade de voz para cada orador.

Próximas Tarefas:

- Correção de um erro cometido no *resample* dos segmentos de voz (passagem de 16kHz para 8kHz), repetição dos testes afectados por esse erro.
- Agendar uma breve conversa com o professor Jaime Cardoso, com o intuito de discutir os resultados obtidos até à data.
- Após obtido o tempo de teste mínimo (de acordo com o ponto 3 das tarefas realizadas), fazer um estudo sobre o tempo de treino mínimo adequado, para o tempo de teste estabelecido.