



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° 9

18 Abril 2011 a 24 Abril 2011

Tarefas Realizadas:

- Implementação de sistema de reconhecimento: normalização CMS, modelação GMM, cálculo log-likelihood, cálculo do *score* log-likelihood, normalização do *score* através de Z-norm.
- Os resultados não permitam fazer uma classificação correcta baseada no *score* normalizado, assim foi feita uma pequena alteração à forma de modelação do *background model*, através da soma pesada dos segmentos constituintes. Esta soma foi feita tendo em consideração a energia média de cada segmento, de modo ao *background model* ser equilibrado.
- Faces aos resultados não satisfatórios obtidos, foi feita uma visualização das projecções de coeficientes (dois a dois) de modo a atestar sobre o seu poder discriminatório das características. Foi concluído que de facto em muitas projecções é bastante difícil distinguir os pontos de MFCCs de dois oradores. Em outras projecções já existe um deslocamento do conjunto de pontos de um orador relativo a outro, no entanto esta diferença é também visível para algumas projecções de MFCCs de dois segmentos de voz diferentes pertencentes ao mesmo orador. Conclui-se que os resultados insatisfatórios podem advir em parte das características usadas, por este motivo.
- Em face aos resultados do teste anterior, foram feitas algumas experiências com diferentes códigos de extracção de MFCCs (VoiceBox e código desenvolvido pelo colega Ricardo Sousa). No entanto apenas melhorias ligeiras foram visualizadas em algumas projecções.
- Teste de uma versão base do sistema acima mencionado, com base de dados TIMIT. (Segmentos da base de dados TIMIT foram previamente tratados - foi eliminado um ruído que todas as amostras continham no início, foram subamostradas de 16kHz para 8kHz, e foram retirados os silêncios). Nesta versão foi apenas feita a normalização CMS, a modelação GMM e classificação baseada

em log-likelihood. Não foi considerado assim um *background model*. Era possível fazer uma classificação com taxas de identificação correcta elevada se fosse estabelecido um *threshold* adequado, no entanto os valores de log-likelihood eram demasiado próximos em todos os casos (orador verdadeiros ou impostor), o que não afere confiança suficiente nos resultados.

- Um último teste para verificar se o log-likelihood seria capaz de expressar uma diferença muito grande entre os segmentos que são modelados pelos GMMs foi realizado. Foi criado um orador fictício constituído por segmentos de ruído branco, rosa e castanho, e foram feitos os mesmos testes que anteriormente. Nestes casos os valores de log-likelihood apresentam uma discrepância considerável entre o caso de orador verdadeiro e impostor.

Dificuldades Encontradas:

- Dificuldades foram já mencionadas na descrição das tarefas.

Próximas Tarefas:

- Pesquisar medidas de distância entre GMMs e outras medidas que possam ser usadas como base na classificação.
- Fazer estudo dos modelos GMMs que são criados a partir dos segmentos de oradores, com o objectivo de concluir sobre o quão diferentes ou semelhantes são, na tentativa de verificar se uma classificação seria possível através de uma medida de distância entre modelos, ou se, pelo contrário, os modelos são relativamente indistinguíveis uns dos outros e daí os valores de log-likelihood serem tão próximos.
- Estudar os diferentes parâmetros de extração de MFCCs e a sua influência nos resultados de classificação.