FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Reconhecimento de Orador em Dois Segundos

Relatório de Progresso

Diana Rocha Mendes

8 de Dezembro de 2010

Orientador e Proponente: Prof. Dr. Aníbal Ferreira

Índice

Introdução

O presente relatório tem por objectivo documentar o trabalho efectuado no âmbito da disciplina de Preparação da Dissertação, após a atribuição do tema de dissertação Reconhecimento de Orador em Dois Segundos e até à data corrente.

Após um breve estudo do tema mencionado e do estado da arte da área científica em que se insere foi possível delinear com mais detalhe o domínio abrangido pela solução a desenvolver e os objectivos que se pretendem atingir, parâmetros que serão descritos neste documento.

Por fim descrevem-se brevemente os fundamentos teóricos já estudados desde o início do trabalho.

Âmbito e Objectivos do Trabalho

O trabalho a desenvolver tem por objectivo implementar uma solução de reconhecimento de orador. Esta solução deverá incorporar técnicas inovadoras que permitam atingir níveis de robustez elevados através de segmentos de voz de curta duração (cerca de dois segundos).

Existem duas áreas principais em reconhecimento de orador: identificação de orador e verificação de orador. Nesta última pretende-se confirmar que o segmento de voz em análise foi produzido por determinada pessoa, cuja identidade é conhecida de antemão, tomando-se apenas uma decisão binária de confirmação ou rejeição. Em identificação de orador, por contraste, o objectivo é seleccionar o orador de um universo de oradores conhecidos, sem qualquer indicação prévia da sua identidade. O trabalho que será desenvolvido insere-se neste último segmento.

O reconhecimento de orador abrange também outros dois métodos distintos: dependente e independente de texto, conforme as gravações de voz usadas correspondem ou não a uma frase específica (texto) comum a todas. O método que será utilizado é a identificação de voz independente de texto.

As soluções actualmente em uso apresentam já níveis de fiabilidade muito elevados, pelo que não é o objectivo principal deste trabalho aumentar a robustez absoluta na solução a implementar. O desafio em que este trabalho se foca primariamente consiste em diminuir o tempo das amostras de voz utilizadas para a identificação robusta do orador. O objectivo é, como já foi mencionado, atingir um tempo de cerca de dois segundos. Para isto serão analisadas as técnicas mais promissoras do estado da arte e estudadas formas de optimizar os algoritmos utilizados. Serão também estudadas novas abordagens ao tema através da exploração de características da voz menos utilizadas, como o impulso glótico.

Plano de Trabalhos

Como guia geral do trabalho a efectuar durante a unidade curricular Dissertação, apresenta-se abaixo o plano de trabalhos. É de notar que esta estrutura e calendarização poderão ser alteradas mais tarde.

- Familiarização com as técnicas de processamento digital de sinal tipicamente aplicadas no reconhecimento de orador (2 semanas);
- Ensaio e avaliação de desempenho de técnicas do estado da arte identificadas como mais promissoras (3 semanas);

- Investigação de novas características permitindo melhorar a robustez no reconhecimento de orador assim como diminuir o tempo de teste necessário para o reconhecimento (4 semanas);
- Caracterização de desempenho da solução encontrada em relação a técnicas e referência e construção de uma plataforma de demonstração funcionando em tempo-real (4 semanas);
- Escrita da dissertação, preparação da apresentação final (4 semanas).

Trabalho Desenvolvido

Durante esta fase inicial o trabalho desenvolvido focou-se sobre pesquisa bibliográfica e estudo teórico dos fundamentos de reconhecimento de orador. Este foi feito em grande parte através de artigos científicos publicados, com o objectivo de compreender os principais blocos constituintes das soluções actuais. Compreendido nesta fase de ambientação esteve também o estudo da fisiologia envolvida na produção de voz e estudo dos conceitos matemáticos e estatísticos básicos aplicados ao reconhecimento de orador.

É de salientar a leitura do artigo escrito por Sten Ternström e a frequência do seminário ministrado pelo mesmo na Faculdade de Engenharia da Universidade do Porto, com o tema “Does the acoustic waveform mirror the voice?”. Neste artigo é proposta uma análise da comunicação oral com base nos protocolos de comunicação utilizados em sistemas de engenharia, oferecendo assim uma nova perspectiva para o estudo da voz falada.

Foi feita também uma pesquisa geral sobre as ferramentas tipicamente usadas para o desenvolvimento de uma solução de reconhecimento de orador. Algumas das ferramentas que poderão ser utilizadas são: Adobe Audition, Cool Edit, Praat (orientado especificamente para fala), Matlab (com auxílio da *toolbox* VoiceBox, utilizada também para processamento da fala), e por fim *software* de tratamento e análise estatística de dados, como o Weka.

Estado da Arte

Ao longo da pesquisa efectuada foi possível identificar as principais técnicas utilizadas no estado da arte e os blocos funcionais tipicamente comuns entre estas.

O primeiro destes blocos é a selecção e extracção de características da voz. Nesta fase cada intervalo do sinal (geralmente entre 10 a 30 milissegundos) é mapeado num espaço multidimensional de características, de forma a permitir comparar os dados extraídos através de simples medidas de semelhança. Esta fase de comparação com modelos previamente obtidos designa-se *pattern matching*. Várias técnicas estatísticas são usadas nesta fase de

processamento, que serão estudadas mais aprofundadamente nas próximas semanas. Para referência futura, enumeram-se de seguida algumas destas: *Dynamic time warping* (DTW), *Vector-Quantized Source Models* (VQ) e *Nearest Neighbour* (modelos determinísticos); *Hidden Markov Model* (modelo probabilístico). Desta fase de *pattern matching* resulta uma pontuação referente à semelhança entre os modelos analisados, que é utilizada como entrada para a fase seguinte - classificação. Nesta inicia-se um processo de decisão (aceitação/rejeição) para concluir sobre a identidade do orador. Umas das ferramentas mais importantes neste processo são as curvas ROC para análise das probabilidades de falsa rejeição e falsa aceitação.

Esta descrição breve do funcionamento geral de um sistema de reconhecimento de orador serve para ilustrar alguns dos tópicos estudados e referir algumas das técnicas mais relevantes que serão estudadas em profundidade ao longo da disciplina de Preparação da Dissertação.

Conclusão

Desde o início do plano de trabalhos foi possível fazer um estudo geral do tema proposto – Reconhecimento de Orador em Dois Segundos – de forma a haver uma melhor compreensão do âmbito do trabalho a desenvolver, da metodologia a ser utilizada e dos objectivos que se pretendem atingir. No entanto, só após um estudo mais aprofundado dos trabalhos já realizados nesta área e possivelmente efectuação de testes práticos será possível delinear com maior precisão em que aspectos se poderão oferecer melhorias às soluções existentes e que técnicas se deverão explorar para esse atingir esse fim.

Bibliografia

[1] Campbell, J. P. (1997). Speaker Recognition: A Tutorial. *Proceedings of the IEEE*, Vol. 85, nº 9, 1437-1462.

[2] Ternström, S. (2005). Does the acoustic waveform mirror the voice?. *Logopedics Phoniatrics Vocology*, Vol. 30, 100-107.