

## E80 - RALPRO - MODELANDO DIFERENÇAS PROSÓDICAS PARA O RECONHECIMENTO AUTOMÁTICO DE LOCUTOR

Cristian Keil de Abreu (PIBIC/CNPq), Cristian Koliver, André Gustavo Adami - Deptº Informática/UCS - [ckabreu@ucs.br](mailto:ckabreu@ucs.br)

A capacidade humana em identificar os indivíduos (locutores) através das características da fala e as suas aplicações tem inspirado os pesquisadores a desenvolver sistemas de Reconhecimento Automático de Locutor (RAL). Apesar da grande quantidade de informações existente na voz, a maioria destes sistemas utiliza apenas as características físicas do sinal de voz como fonte de informação para o reconhecimento. Isto torna importante o acréscimo de novas fontes de informações existentes no sinal de fala, como a prosódia, para aumentar o desempenho desses sistemas. A prosódia possui um poder discriminatório devido às diferenças na produção da entonação, do ritmo e da ênfase no sinal de voz, caracterizados pelas medidas físicas da frequência fundamental (medida correlata do tom) e da energia (medida correlata da intensidade) do sinal de voz. Um dos métodos para modelagem de características prosódicas é a modelagem das distribuições das medidas correlatas da intensidade e do tom. Tal modelagem é realizada através do uso de misturas de distribuições gaussianas (Modelo de Misturas Gaussianas – MMG) e utilizada para representar a identidade de um determinado locutor. As componentes gaussianas, as quais possibilitam a formação de densidades arbitrárias, representam classes acústicas que caracterizam a voz de um locutor. No processo de reconhecimento dois tipos de modelos são utilizados: modelo independente do locutor (ou dos impostores) e modelo do locutor. O sistema decide se a mensagem de teste foi produzida por um determinado locutor baseado na razão entre a probabilidade que a mensagem foi produzida pelo modelo do locutor (treinado com os dados somente do locutor) e a probabilidade que a mensagem foi produzida por algum impostor (modelo independente do locutor, treinado com dados de diversos locutores). Os parâmetros do modelo independente do locutor são estimados utilizando o algoritmo iterativo Maximização da Expectativa. Devido ao espaço de características esparso do locutor, uma técnica de adaptação do modelo *maximum a posteriori* é utilizada para estimar os parâmetros do modelo do locutor. Estes métodos descritos estão sendo implementados dentro de um toolkit para livre distribuição na internet, os quais estarão brevemente disponíveis no site do projeto. A avaliação dos resultados obtidos está sendo realizada a partir do paradigma de avaliação proposto *National Institute of Standards and Technology* (NIST) para o reconhecimento de locutor, devido a sua aceitação internacional como medida de desempenho para este tipo de tecnologia.

Palavras-chave: prosódia, modelo de locutor, sinal de voz

Apoio: UCS, CNPq