

SPEAKERID - TOOLKIT PARA RECONHECIMENTO AUTOMÁTICO DE LOCUTOR

Cristian Keil de Abreu (PIBIC-CNPq), André Gustavo Adami (orientador) - ckabreu@ucs.br

A capacidade humana em identificar os indivíduos (locutores) através das características da voz e as suas aplicações tem inspirado os pesquisadores a desenvolver ferramentas conhecidas como sistemas de Reconhecimento Automático de Locutor (RAL). Mesmo a voz contendo grande quantidade de informações, 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 outras fontes de informação existentes no sinal da fala para melhorar o desempenho desses sistemas. Uma fonte de informação com poder discriminatório é a prosódia, devido às diferenças na produção da entonação, do ritmo e da ênfase no sinal de voz. Tais diferenças na produção de prosódia são refletidas pelas medidas físicas da frequência fundamental e da energia. Um método para representar a identidade do locutor com informações prosódicas é a modelagem das distribuições das medidas correlatas da intensidade e do tom através do uso de misturas de distribuições gaussianas. O uso de misturas de distribuições gaussianas possibilita representar as classes prosódicas que caracterizam a voz de um locutor através da mistura de densidades arbitrárias. Entretanto, a modelagem da distribuição ignora informações temporais que caracterizam os fenômenos prosódicos. Com o objetivo de utilizar a informação temporal da prosódia, foi desenvolvido um método para representar o sinal de voz através de uma seqüência de símbolos discretos que representam padrões das medidas correlatas da intensidade e do tom. Devido à natureza discreta da representação, utilizou-se a modelagem da ocorrência conjunta de símbolos, conhecida como modelos N-gramas, para representar a identidade de um locutor. Este método de modelagem permite capturar a informação temporal entre os padrões prosódicos produzidos por um locutor. Todas estas ferramentas estão disponíveis através do toolkit SpeakerID. Além dos processos de extração e modelagem de características prosódicas, o toolkit permite o desenvolvimento e teste de sistemas de RAL nos padrões definidos pelo órgão americano, análogo ao Inmetro, National Institute of Standards and Technology – NIST, o qual realiza avaliações anuais das tecnologias relacionadas ao processamento de voz.

Palavras-chave: reconhecimento de locutor, reconhecimento de padrões, processamento de voz.

Apoio: UCS.