

XVIII Encontro de Jovens Pesquisadores

Universidade de Caxias do Sul - 2010

PASSVOX - Segmentação Automática de Locutor.

Cristian Keil de Abreu (BIC/FAPERGS), Andre Gustavo Adami (Orientador(a))

Com o desenvolvimento tecnológico, cada vez mais diversas formas de informação, entre elas o áudio, estão sendo coletadas para diferentes fins. Programas de televisão, programas de rádio e sessões das câmaras dos deputados federais e do senado são registradas através de gravações em meio digital para futuras referências de seu conteúdo. A informação a ser obtida não é o próprio áudio em si, mas o que está contido nele. Uma das informações que podem ser obtidas é a informação do locutor, que é extraído através de um sistema de reconhecimento automático de locutor para bases de áudio. Esse sistema requer a utilização de diversos algoritmos para realizar diferentes atividades, como detecção de atividade de fala, definir os intervalos onde um locutor está falando, e identificar um determinado locutor. Neste trabalho focamos no desenvolvimento de algoritmos para estimar os intervalos de tempo quando um locutor está falando. Esta tarefa é comumente chamada de segmentação automática de locutor. Um fator complicador neste tipo de tarefa é que não existe dados sobre os locutores de um áudio. Isto requer que o algoritmo realize a segmentação de forma não supervisionada (isto é, sem conhecer a identidade dos locutores). O processo de segmentação de locutores pode ser dividido em duas etapas: a detecção de mudança de locutor e o agrupamento. A primeira etapa visa detectar os pontos onde existe uma mudança de locutor. Assim, o áudio é segmentado de tal forma que cada segmento possua somente um locutor. A etapa seguinte realiza o agrupamento dos segmentos do mesmo locutor (já que os locutores falam de forma intercalada). Existem diversos métodos para estimar quando ocorreu a mudança de um locutor. Neste trabalho utilizamos os períodos de silêncios para assumir que ocorreu uma mudança de locutor. Os períodos de silêncio são estimados através da energia da voz (a energia do sinal quando existe silêncio ou ruído é menor do que nos períodos de fala). Para a avaliação da segmentação de locutor será usado o paradigma de avaliação proposto pela agência NIST (National Institute of Standards and Technology) devido a sua aceitação internacional como medida de desempenho para este tipo de tecnologia. Uma das avaliações disponíveis pelo NIST e que será usada para o projeto é para áudios com somente dois locutores.

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

Apoio: FAPERGS - UCS