

XVIII Encontro de Jovens Pesquisadores Universidade de Caxias do Sul - 2010

PassVox - Detecção Automática da Atividade de Fala

Miguel Angelo Mussoi (PIBIC/CNPq), Andre Gustavo Adami (Orientador(a))

O objetivo do projeto é desenvolver um sistema de reconhecimento automático de locutor para bases de áudio. Neste tipo de aplicação, o sistema deve receber um arquivo de áudio, detectar os períodos de fala, separar os áudios de diferentes locutores e realizar o reconhecimento dos mesmos. Este tipo de aplicação permite a busca de informações relacionadas aos locutores a partir do áudio. Um exemplo de aplicação é a busca de todo áudio que possua um determinado locutor. Este trabalho foca em um dos algoritmos necessários para o desenvolvimento do sistema: detecção da atividade de fala. Dois algoritmos de detecção de fala foram desenvolvidos, um baseado em energia e outro em entropia. O algoritmo baseado energia utiliza um limiar sobre a energia do sinal para determinar os períodos onde existe atividade de voz. Este algoritmo é baseado no princípio que a distribuição da energia do ruído e ou silêncio é centrada (através da média) num valor inferior ao da distribuição da energia do sinal de voz. Assim, pode-se estimar as duas distribuições e utilizar um limiar que divide as duas distribuições para decidir se a energia vem de um sinal de voz ou não. Para estimar as distribuições, utiliza-se um modelo de misturas Gaussianas com dois componentes (ou dois modos). A fim de maximizar a discriminação, o limiar é estimado a partir do ponto onde as duas Gaussianas estimadas se cruzam. O segundo algoritmo, baseado em entropia, utiliza a entropia (medida de incerteza) do espectro de frequências do sinal. O princípio deste algoritmo é que o silêncio/ruído afeta todas as frequências do sinal igualmente, o que resultaria em um valor de entropia alto (maior incerteza). No caso da fala, somente algumas frequências são afetadas produzindo uma entropia baixa (pouca incerteza pois existem frequências dominantes). Um limiar é utilizado para determinar os períodos de fala. O limiar é estimado a partir de um conjunto de dados de desenvolvimento. Os dois métodos foram avaliados utilizando uma base de fala de aproximadamente 60 minutos. Dois tipos de erros podem ocorrer: falso alarme (ocorre quando o sinal é rotulado como fala quando na verdade é silêncio/ruído) e não detecção (ocorre quando o sinal é rotulado como silêncio/ruído quando na verdade é fala). A medida de desempenho utilizada é o ponto onde a taxa de não detecção é igual a taxa de falso alarme. O algoritmo baseado em energia teve taxa de 6% enquanto o baseado em entropia teve uma taxa de 12%.

Palavras-chave: detecção de atividade de fala, processamento de voz, reconhecimento de padrões.

Apoio: UCS, CNPq