Abstract

Large speech databases such as Television broadcasts, TV programs, radio broadcasts, CDs and DVDs are available online these days. Research related to speech indexing and retrieval has received much attention in recent days due to the huge multimedia data storage capabilities. The goal of speech indexing and retrieval system is to provide the user with capabilities to index and retrieve the speech archives in an efficient manner. In this paper, we propose a method for indexing and retrieval of the speech. The speech activity is identified using voice activity detection and each complete speech dialogue is separated into individual words by marking each word’s segment through the Root Means Square (RMS) energy envelope. Then the features namely Perceptual Linear Prediction (PLP), Power Normalized Cepstral Coefficient (PNCC), Subband Coding (SBC) and Sonogram extracted from each of the individual word. For retrieval, a novel method is proposed using Gaussian mixture models. The probability that the query feature vector belongs to the Gaussian is computed. The average Probability density function is computed for each of the feature vectors in the database and the retrieval is based on the highest probability.
References


22. Menaka Rajapakse and Lonce Wyse, 2005, Generic Audio Classification using a Hybrid Model Based on GMMs and HMMs, IEEE International Multimedia Modelling Conference, pp. 53-58.

**Index Terms**

Computer Science

Signal Processing

**Keywords**

Voice Activity Detection, Root Mean Square, Gaussian Mixture Model, Probability density function