1 OBJECTIVE

Objectives of the proposed Research

1. The objective of audio segmentation for classifying the audio components into Speech, Music, Non-speech, noise, silence along with the other major features such as MFCC, SF, LPC, SNR, HZCRR etc and can be transferred over the network and by analysing these audio features the reconstruction of audio signal should be more accurate.

2. We proposed to use thirteen features in time, frequency, and cepstrum domains and model-based (MAP, GMM, KNN, etc.) classifier, which achieved an accuracy rate over 90% on real-time discrimination between speech and music. As in general, speech and music have quite different spectral distribution and temporal changing patterns, it is not very difficult to reach a relatively high level of discrimination accuracy.

3. Further classification of audio data may take other sounds into consideration besides speech and music.

4. We also proposed an approach to detect and classify audio that consists of mixed classes such as combinations of speech and music together with environment sounds. The accuracy of classification is more than 80%.

5. An acoustic segmentation approach was also proposed where audio recordings to be segmented into speech, silence, laughter and non-speech sounds.

6. We have to use cepstral coefficients as features and the hidden Markov model (HMM) as the classifier. We propose a MGM-based (Modified Gaussian Modelling) hierarchical classifier for audio stream classification. Compared to traditional classifiers, GM can automatically optimize the weights of different kinds of features based on training data. It can raise the discriminative capability of audio classes with lower computing cost.
System Flow

Figure 1: The flowchart of segmentation and classification algorithm

Figure 1 shows the flowchart of proposed audio segmentation and classification algorithm. It is a hierarchical structure. In the first level, a long audio stream can be segmented into some audio clips according to the change of background sound by MBCR based histogram modelling. Then a two level MGM (Modified Gaussian modeling) classifier is adopted to hierarchically put the segmented audio clips into six pre-defined categories in terms of discriminative background sounds, which is pure speech (PS), pure music (PM), song (S), speech with music (SWM), speech with noise (SWN) and silence (SIL).

SEGMENTATION ALGORITHM

Since background sounds always change with the change of scenes, the acoustic skip point of an audio stream may be checked by background sounds. As shown in Figure 2, the MBCR feature vectors are firstly extracted from the audio stream. We set a sliding window which consists of two sub-windows with equal time length. The window on input signal is
shifted with a range of overlapping. Then two histograms are created from each sliding sub-windows. The similarity between two sub-windows can be measured by histogram matching. The skip point can thus be detected by searching the local lowest similarity below a threshold.

Fig 2. Block diagram of Segmentation algorithm