A MULTI-BAND SPECTRAL SUBTRACTION METHOD FOR ENHANCING SPEECH CORRUPTED BY COLORED NOISE

Sunil Kamath; Philipos Loizou, University of Texas at Dallas, United States

The spectral subtraction method is a well-known noise reduction technique. Most implementations and variations of the basic technique advocate subtraction of the noise spectrum estimate over the entire speech spectrum. However, real world noise is mostly colored and does not affect the speech signal uniformly over the entire spectrum. In this paper, we propose a multi-band spectral subtraction approach which takes into account the fact that colored noise affects the speech spectrum differently at various frequencies. This method outperforms the standard power spectral subtraction method resulting in superior speech quality and largely reduced musical noise.

AN EVENT-BASED ACOUSTIC-PHONETIC APPROACH FOR SPEECH SEGMENTATION AND E-SET RECOGNITION

Amit Juneja; Om Deshmukh; Carol Espy-Wilson, University of Maryland, United States

In this paper, we discuss an event-based recognition system (EBS) which is based on phonetic feature theory and acoustic phonetics. First, acoustic events related to the manner phonetic features are extracted from the speech signal. Second, based on the manner acoustic events, information related to the place phonetic features and voicing are extracted. Most recently, we focused on place and voicing information needed to distinguish among the stop consonants /t,d,p,b/. Using the E-set utterances from the TI46 database, EBS achieved 75.7% overall word accuracy. Further, the knowledge-based acoustic parameters (APs) optimized within the EBS framework were compared to the mel-frequency cepstral coefficients in an HMM-based recognition system. The results on the E-set task showed that the APs achieve a higher recognition accuracy.

SPEECH AND MUSIC CLASSIFICATION IN AUDIO DOCUMENTS

Julien Pinquier; Christine Sénac; Régine André-Obrecht, IRIT, France

To index efficiently the soundtrack of multimedia documents, it is necessary to extract elementary and homogeneous acoustic segments. In this paper, we explore such a prior partitioning which consists in detect the two basic components, which are speech and music components. The originality of this work is that music and speech are not considered as two classes and two classification systems are independently defined, a speech/non-speech one and a music/non-music one. This approach permits to better characterize and discriminate each component: in particular, two different feature spaces are necessary as two pairs of Gaussian mixture models. More, the acoustic signal is divided into four types of segments: speech, music, speech-music and other. The experiments are performed on the soundtracks of audio video documents (films, TV sport broadcasts). The performance proves the interest of this approach, so called the Differentiated Modeling Approach.

0-7803-7402-9/02/$17.00 ©2002 IEEE