Abstract

In this paper, a Mel-scaled AR (Mel-AR) model based VAD is presented, where likelihood ratio measure is used to classify the input speech frames as speech/non-speech segments. The Mel-AR model parameters have been estimated on the linear frequency scale from the input speech signal without applying bilinear transformation. This has been done by employing a first-order all-pass filter rather than unit delay. The performance of the proposed VAD is evaluated on Aurora-2 database by measuring FAR and FRR. The equal false rate (EFR) at the crossover point is also presented as a merit of VAD. In addition, the performance of the proposed VAD in speech recognition is verified by incorporating it with a Mel-Wiener filter for MLPC based noisy speech recognition.

References

optimized for terminals conforming to recommendation V.70.
3. ETSI. 1999. Voice Activity Detector (VAD) for Adaptive Multi-Rate (AMR) Speech Traffic
Channels. ETSI EN 301 708 Recommendation.
4. ETSI. 2007. Speech Processing, Transmission and Quality Aspects (STQ); Distributed
speech recognition; Advanced front-end feature extraction algorithm; Compression algorithms.
ETSI ES 202 050 v1.1.5.
applied to FF parameters. Proceedings ICASSP’05, 1: 557-560.
Zero-Crossing Rate and Energy. Advanced Techniques in Computing Sciences and Software
static harmonic features. Proceeding ICASSP’10, 4482-4485.
Model and Parallel Non-linear Kalman Filtering. Proceedings ICASSP’07, 4: 797-800.
recognition using statistical model, EURASIP Journal on Audio, Speech, and Music Processing,
16. Tan, L. N.et al. 2010. Voice activity detection using harmonic frequency components in
likelihood ratio test, ICASSP’10, 4466-4469.
Hypothesis Testing for Robust Speech Recognition. IEEE transactions on audio, speech and
language processing, 15(8): 2177-2189.
Proc., 60(6): 681-691.
America, 68(4): 1071-1076.

Index Terms

Computer Science  
Signal Processing

Keywords

VAD, Mel-AR model, Likelihood ratio, Itakura-Saito distortion, Aurora 2 database