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

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.

Index Terms

Computer Science  Signal Processing

Keywords

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