Speaker Identification using Spectrograms of Varying Frame Sizes

Abstract

In this paper, a text dependent speaker recognition algorithm based on spectrogram is proposed. The spectrograms have been generated using Discrete Fourier Transform for varying frame sizes with 25% and 50% overlap between speech frames. Feature vector extraction has been done by using the row mean vector of the spectrograms. For feature matching, two distance measures, namely Euclidean distance and Manhattan distance have been used. The results have been computed using two databases: a locally created database and CSLU speaker recognition database. The maximum accuracy is 92.52% for an overlap of 50% between speech frames with Manhattan distance as similarity measure.

Refer
References

- Paul E. Black, "Euclidean distance;" Dictionary of Algorithms and Data Structures [online].
Keywords
Discrete Fourier Transform (DFT)  Row Mean  Euclidean Distance  Manhattan distance