Abstract

The most important challenges in AVSR and the focus of most research are the features that are extracted, and when combined give better results. The other challenge is the resulted feature here of nature are large in size, then prefers here to reduce the features by use of an appropriate way to reduce these data with ensure have their properties after downsizing. The System that is presented in this research is for recognition a group of Arabic words voices, from one to ten words. In the acoustic parts the features were extracted of coefficients MFCC, LPC,FFT to be determined which type of these features is efficient in AVSR. All these types of feature are showed efficient results but MFCC is the best. The visual features are calculated of DCT matrix, and the features are extracted by applying the zigzag scan. In the reduction features stage, several methods of data reducing have been implemented; they are LDA, PCA and SVD. Each method are applied to the data separately. The KNN models are used in the stage of recognition, where the testing is implemented on dependent and independent database of words from one to ten. The final results that obtained are efficient and encouraging.
References


Index Terms

Computer Science  Signal Processing
Keywords

Audio-Video Speech Processing, Automatic Speech recognition, Mouth detection, Discrete cosine transformation, Visual Features