The contribution of this paper is to propose a novel approach of evaluating the performance of a noise robust audio-visual speaker identification system in challenging environment. Though the traditional HMM based audio-visual speaker identification system is very sensitive to the speech parameter variation, the proposed hybrid feature and decision fusion based audio-visual speaker identification is found to be stance and performs well for improving the robustness and naturalness of human-computer-interaction. Linear Prediction Cepstral Coefficients and Mel Frequency Cepstral Coefficients are used to extract the audio features and Active Appearance Model and Active Shape Model have been used to extract the appearance and shape based features for the facial image. Principal Component Analysis method has been used to reduce the dimensionality of large feature vector and to normalize, the vector normalization algorithm has been used. Features and decision both are fused in two different levels and finally four
different classifier outputs are combined in parallel fashion to achieve the identification result. The performances of all these uni-modal and multi-modal system performance have been evaluated and compared with each other on VALID audio-visual multi-modal database, containing both vocal and visual biometric modalities.

Reference


**Index Terms**

Computer Science

Bio-informatics

**Key words**

Hybrid Feature and Decision Fusion

Audio-Visual Speaker Identification

Cepstral Base Audio Features

Appearance and Shape Based Facial Features

Likelihood Ratio Based Score Fusion

Discrete Hidden Markov Model