Abstract

It is well known to enhance the performance of noise robust speaker identification using visual speech information with audio utterances. This paper presents an approach to evaluate the performance of a noise robust audio-visual speaker identification system using likelihood ratio based score fusion in challenging environment. Though the traditional HMM based audio-visual speaker identification system is very sensitive to the speech parameter variation, the proposed likelihood ratio based score fusion method is found to be stancel and performs well for improving the robustness and naturalness of human-computer-interaction. In this paper, we investigate the proposed audio-visual speaker identification system in typical office environments conditions. To do this, we investigated two approaches that utilize speech utterance with visual features to improve speaker identification performance in acoustically and visually challenging environment: one seeks to eliminate the noise from the acoustic and visual
features by using speech and facial image pre-processing techniques. The other task combines speech and facial features that have been used by the multiple Discrete Hidden Markov Model classifiers with likelihood ratio based score fusion. It is shown that the proposed system can improve a significant amount of performance for audio-visual speaker identification in challenging official environment conditions.

Reference

- Claudio Becchetti and Lucio Prina Ricotti, Speech Recognition Theory and C++
Likelihood Ratio Based Score Fusion for Audio-Visual Speaker Identification in Challenging Environment


**Index Terms**

Computer Science

Signal Processing

**Key words**

Audio-Visual Speaker Identification

Cepstral Base Features

Feature Fusion

Decision Fusion

Likelihood Ratio Based Score Fusion

Discrete Hidden Markov Model