Abstract

Automatic Speech Recognition (ASR) by machine is an attractive research topic in signal processing domain and has attracted many researchers to contribute in this area of signal processing and pattern recognition. In recent year, there have been many advances in automatic speech reading system with the inclusion of audio and visual speech features to recognize words under noisy conditions. The objective of audio-visual speech recognition system is to improve recognition accuracy. In order to develop robust AVSR systems under Human Computer Interaction an appropriate simultaneously recorded speech and video data are needed. This paper describes a ‘vVISWa’ (Visual Vocabulary of Independent Standard Words) database consists of audio visual data of 48 native speakers and 10 nonnative speakers. These speakers have contributed towards development of corpus in three profiles that is full frontal, 450 profile and side pose. This database was primarily designed to deal with Multi-pose Audio Visual Speech Recognition system for three languages that is, ‘Marathi’ (The Native language of Maharashtra), ‘Hindi’ (National Language of India) and ‘English’ (Universal
language). This database is multi-pose, multi-lingual database formed in Indian context. This database available by request from http://visbamu.in/viswaDataset.html.

References

‘vVISWa’ – A Multilingual Multi-Pose Audio Visual Database for Robust Human Computer Interaction


Index Terms

Computer Science  Databases

Keywords

Automatic Speech Recognition (ASR), Visual Speech Reading (VSR), Multi-pose Audio Visual Speech Recognition (AVSR) and ‘vVISWa’.