Abstract

Speech recognition of disorder people is a difficult task due to the lack of motor-control of the speech articulators. Multimodal speech recognition can be used to enhance the robustness of disordered speech. This paper introduces an automatic speech recognition system for people with dysarthria speech disorder based on both speech and visual components. The Mel-Frequency Cepstral Coefficients (MFCC) is used as features representing the acoustic speech signal. For the visual counterpart, the Discrete Cosine Transform (DCT) Coefficients are extracted from the speaker’s mouth region. Face and mouth regions are detected using the Viola-Jones algorithm. The acoustic and visual input features are then concatenated on one feature vector. Then, the Hidden Markov Model (HMM) classifier is applied on the combined feature vector of acoustic and visual components. The system is tested on isolated English words spoken by disorder speakers from UA-Speech data. Results of the proposed system indicate that visual features are highly effective and can improve the accuracy to reach 7.91% for speaker dependent experiments and 3% for speaker independent experiments.

Refer
ences

- Heidi Christensen, Stuart Cunningham, Charles Fox, Phil Green, and Thomas Hain. 2012. A comparative study of adaptive, automatic recognition of disordered speech. INTERSPEECH. ISCA.
- Chikoto Miyamoto, Yuto Komai, Tetsuya Takiguchi, Yasuo Arikii, and Ichao Li. 2010. Multimodal speech recognition of a person with articulation disorders using AAM and MAF. In proceeding of Multimedia Signal Processing (MMSP).
- Paul Viola and Michael J. Jones. 2001. Rapid Object Detection using a Boosted Cascade of Simple Features. IEEE CVPR.
- P. Scanlon and G. Potamianos. 2005. Exploiting lower face symmetry in

Index Terms

Computer Science Pattern Recognition

Keywords

AVSR HMM HTK MFCC DCT UA-Speech OpenCV.