Abstract

The aim of this paper is to measure the recognition capability of composite features extracted from speech signal and compare the result with other individually considered features for both spoken word and speaker based recognitions. Standard features like formants (F1, F2, F3), Linear Predictive Coefficients (LPC) and Mel Frequency Cepstral Coefficients (MFCC) along with various combinations among them are considered for the task to arrive at the conclusion. Six different speakers and six different strings (words) are considered in the present study. The threshold is set through an iterative approach for both spoken word and speaker recognition experiments. The mixing of LPC and MFCC is found to be the most promising combination among all others. Another interesting conclusion that we can draw from the study that the composite feature approach gives accuracy very near to 100% in case of speaker recognition task as compared to spoken word recognition task.

References

- Adjoudj Reda, Boukelif Aoued, "Artificial Neural Network & Mel-Frequency Cepstrum Coefficients-Based Speaker Recognition," 3rd International Conference:
Sciences of Electronic, Technologies of Information and Telecommunications--TUNISIA, March 27-31, 2005


- Braman, J., Kalita, S., Talukdar, P. H., “Features extraction of bodo vowels through lpc-analysis.” Proceedings of Frontiers of Research on Speech and Music (FRMS-2004), 2004


- Mark K. Transtrum and James P. Sethna, “Improvements to the Levenberg-Marquardt algorithm for nonlinear least-squares minimization.” Preprint


Index Terms

Computer Science Speech Processing

Keywords

Feature Extraction and Selection Feed Forward Neural Network Speech and Speaker Recognition