Abstract

Automated Speaker Verification (ASV) systems are extensively used for authentication and verification measures. Countermeasures are developed for ASV systems to protect it from audio replay attacks. This paper describes the ASVspoof2017 database, conceptual analysis of various algorithms and their classification followed by prediction of results. Feature extraction is based on the recently introduced Constant Q Transform (CQT), a perceptually mapped frequency-time analysis tool mainly used with audio samples. The training dataset comprises of 1508 genuine samples and 1508 spoof samples. A training accuracy of 84.4% is achieved for variations of boosted decision tree. Parameters such as learning rate, number of learners and splits were empirically optimized. LogitBoost was found to have outperformed AdaBoost in all metrics. Furthermore, an implementation of a single hidden layer neural network achieved a training accuracy of 92.1%. A comparison of the algorithms revealed that while the neural network achieved a higher overall training accuracy, it had a lower True Negative Rate than LogitBoost. Overall, the paper describes a generalized system capable to detection of replay
Audio Replay Attack Detection in Automated Speaker Verification

attacks in known and unknown conditions.

References

17. S. Novoselov, A. Kozlov, G. Lavrentyeva, K. Simonchik, and V. Shchemelinin, “STC
18. T. B. Patel and H. A. Patil, “Combining evidences from mel cepstral, cochlear filter 
cepstral and instantaneous frequency features for detection of natural vs. spoofed speech,” in 
INTERSPEECH, 2015, pp. 2062–2066.
20. Z. Wu, S. Gao, E. Chng, and H. Li, “A study on replay attack and anti-spoofing for 

Index Terms

Computer Science Security

Keywords

Replay attack detection, Automated speaker verification, Classification of Speech Samples