

Forensic Speaker Identification with Breath

Wenbo Zhao

School of Electrical and

Computer Engineering

Carnegie Mellon University

Pittsburgh, PA 15213

Email: wzhaol@andrew.cmu.edu

Abstract—In forensics, identifying suspects from short period phone recordings is difficult due to several reasons: (1) the rich variations of human voices, (2) spoofing voices, (3) noisy recording channels, etc. In prior work, Hidden Markov Models with Gaussian Mixture Models based methods and i-vector based methods are used to represent the variations and similarities in speech. While effective, these methods require large amount of data, has iterative steps, rely on prior assumptions of the data distributions, thus are inefficient and perform bad when lack of data. In this study, it develops a pipeline to identify speakers using only their breath recordings. First, the breath recordings are transformed to constant-Q spectrograms to leverage pitch variations among speakers. Then, the spectrograms are fed into convolutional network to capture their shift invariant features. Next, these features are fed into a Long Short-Term Memory network to capture the temporal variations. Finally, a feedforward network predicts the multi-class likelihoods for each recording. Compared with conventional methods, the proposed method is (1) simple, without complex feature computation, (2) effective on small amount of data, and (3) independent of assumptions of data distributions.

I. INTRODUCTION

A. Subsection Heading Here

1) Subsubsection Heading Here:

II. CONCLUSION

ACKNOWLEDGMENT

The authors would like to thank...

REFERENCES

- [1] H. Kopka and P. W. Daly, *A Guide to L^AT_EX*, 3rd ed. Harlow, England: Addison-Wesley, 1999.