PhD Abstract

Fusing prosodic and acoustic information for speaker recognition

Mireia Farrús

Awarding Institution: Universitat Politècnica de Catalunya

Date of Award: October 2008

Keywords: speaker recognition, prosody, multimodality, imitation, conversion

Automatic speaker recognition is the use of a machine to identify an individual from a spoken sentence. Recently, this technology has been undergone an increasing use in applications such as access control, transaction authentication, law enforcement, forensics, and system customisation, among others (Campbell, 1997).

One of the central questions addressed by this field is what is it in the speech signal that conveys speaker identity. Traditionally, automatic speaker recognition systems have relied mostly on short-term features related to the spectrum of the voice (Rabiner & Juang, 1993). However, human speaker recognition relies on other sources of information; therefore, there is reason to believe that these sources can play also an important role in the automatic speaker recognition task, adding complementary knowledge to the traditional spectrum-based recognition systems and thus improving their accuracy (Peskin et al., 2003).

The main objective of this thesis is to add prosodic information to a traditional spectral system in order to improve its performance. To this end, several characteristics related to human speech prosody – which is conveyed through intonation, rhythm and stress – are selected and combined them with the existing spectral features. Furthermore, this thesis also focuses on the use of additional acoustic features – namely jitter and shimmer – to improve the performance of the proposed spectral-prosodic verification system. Both features are related to the shape and dimension of the vocal tract, and they have been largely used to detect voice pathologies.

Since almost all the above-mentioned applications can be used in a multimodal environment, this thesis also aims to combine the voice features used in the speaker

recognition system together with other biometric identifiers – face – in order to improve the global performance (Bolle, Connell, Pankanti, Ratha, & Senior, 2004). To this end, several normalisation and fusion techniques are used, and the final fusion results are improved by applying different fusion strategies based on sequences of several steps. Furthermore, multimodal fusion is also improved by applying a histogram equalisation to the unimodal score distributions as a normalisation technique.

On the other hand, it is well know that humans are able to identify others from voice even when their voices are disguised. The question arises as to how vulnerable speaker recognition systems are against different voice disguises, such as human imitation (Zetterholm, 2003) or artificial voice conversion, which are potential threats to security systems that rely on automatic speaker recognition. First, some experiments are performed in order to test the influence of foreign accents and dialects – as a sort of imitation – in auditory speaker recognition. Second, the voices of two well-known professional imitators trying to impersonate several well-known politicians are used to analyse the behaviour of some selected and acoustic features in the imitated voices. Finally, automatically converted voices are also used to test the robustness of speaker verification systems.

References

Bolle, R. M., Connell, J. H., Pankanti, S., Ratha, N. K., & Senior, A. W. (2004). *Guide to Biometrics*. New York: Springer.

Campbell, J. P. (1997). Speaker recognition: A tutorial. *IEEE, 85*, 1437-1462.

Peskin, B., Navratil, J., Abramson, J., Jones, D., Klusacek, D., Reynolds, D. A., et al. (2003, April 2003). *Using prosodic and conversational features for high-performance speaker recognition: Report from JHU WS'02.* Paper presented at the ICASSP, Hong Kong.

Rabiner, L. R., & Juang, B. H. (1993). *Fundamentals of Speech Recognition*. Englewood Cliffs, New Jersey: Prentice Hall, Inc.

Zetterholm, E. (2003). *Voice Imitation. A phonetic study of perceptual illusions and acoustic success.* Lund University, Lund.