

The influence of intra-speaker variability in automatic speaker identification

Timo Becker

Acoustics Research Institute of the Austrian Academy of Sciences, Austria

timo.becker@oeaw.ac.at

Apart from the known performance-degrading influences on automatic speaker recognition systems, such as channel mismatch or available amount of data, intra-speaker variabilities still have to be investigated. The impact of these variabilities on automatic speech recognition systems must be described if they are to be used in forensic cases, because here, a high degree of intra-speaker variability is usually observed.

In an experiment, the 'Pool 2010' data base, recorded by the Department of Speaker Identification and Audio Analysis, Bundeskriminalamt, Germany (Jessen et al. 2005), was used for automatic speaker identification. The identification system uses Gaussian mixture models (GMM) estimated from feature vectors consisting of Mel frequency cepstral coefficients (MFCC) (Reynolds & Rose 1995).

100 German speakers were used in the experiment. GSM-transmitted recordings of *reading (R)* vs. *spontaneous (S)* speech and *free (F)* vs. *Lombard (L)* speech were cross-identified in all possible 12 mismatch conditions to evaluate the different influences of the speaking styles. Of these 12 mismatch conditions, 4 were 'double mismatch' conditions (neither R/S nor F/L matched) and 8 were 'single mismatch' conditions (only one of R/S or F/L mismatched). R/S mismatch and F/L mismatch cause speaker-individual changes of speech segment durations, segment length relations and spectral properties (Eskénazi 1992, Laan 1997, Karlsson et al. 1998, Steeneken & Hansen 1999, Köster 2002, Holm 2003, Jessen et al. 2005). These changes are included in the feature vectors and the GMMs because MFCC feature vectors represent time-independent and energy-independent smoothed spectra on a cosine base which include little pitch information, while GMMs model the speaker-specific distributions of these vectors.

It could be seen that mismatches in the observed speaking styles showed different identification rates (IR). The lowest IR of 55 % was observed for the double mismatch conditions (4 cases), while the single mismatch conditions had an IR of 79 % (8 cases). Dividing the single mismatch conditions into R/S mismatch conditions (4 cases) and F/L mismatch conditions (4 cases), it was observed that R/S mismatch showed a higher IR of 89 % than F/L mismatch with an IR of 68 %.

References

- Eskénazi, M. (1992). Changing Speech Styles: Strategies in Read Speech and Casual and Careful Spontaneous Speech. *Proc. of the International Conference on Spoken Language Processing*, 1-5.
- Holm, S. (2003). Individual use of acoustic parameters in read and spontaneous speech. *PHONUM*, **9**, 157-160.
- Jessen, M., Köster, O. and Gfroerer, S. (2005). Influence of vocal effort on average and variability of fundamental frequency. *Speech, Language and the Law*, **12**, 174-213.
- Karlsson, I., Banziger, T., Dankovicová, J., Johnstone, T., Melin, H., Nolan, F. and Scherer, K. (1998). Within-speaker variability due to speaking manners. *Proc. of the Int. Conf. on Spoken Language Processing*, 121-129.
- Köster, S. (2002). Acoustic-Phonetic Aspects of Lombard Speech for Different Text Styles. *The Phonetician*, **85**, 9-16.
- Laan, G. (1997). The contribution of intonation, segmental durations, and spectral features to the perception of a spontaneous and a read speaking style. *Speech Communication*, **22**, 43-65.
- Reynolds, D. A. and Rose, R. C. (1995). Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models. *IEEE Transactions on Speech and Audio Processing*, **3**, 72-83.
- Steeneken, H. and Hansen, J. H. L. (1999). Speech under Stress Conditions: Overview of the Effect on Speech Production and on System Performance. *Proc. ICASSP'99*, 2079-2082.