

Comparing speaker recognition techniques

abstract

D. Bloembergen, P. Hanckmann, A. Lautenbach, B. Mehlkop, F. Schadd

February 6, 2007

This abstract is taken from the report *Comparing speaker recognition techniques* written at January 24, 2007. The report is written under supervision of dr. ir. Ralf Peeters, dr. ir. Jos Uiterwijk and drs. Jan Paredis. It represents the results of the research done in block 3.3 at the Maastricht University Faculty of Humanities and Sciences.

The research question that is tried to answer in this report reads as follows:

'How well do certain speaker recognition techniques perform under different circumstances?'

Three algorithms for speaker recognition and three different methods of extracting features from speech samples are described. The algorithms discussed are k-Nearest Neighbour, Neural Network and Gaussian Mixture Models. Methods for feature extraction used are Cepstral Coefficients, Mel Frequency Cepstral Coefficients and Linear Predictive Coding.

The algorithms and methods are compared qualitatively. Tests are performed to analyse the accuracy of the algorithms using different feature extraction methods in order to find the best combination. Furthermore the algorithms are tested under various circumstances and purposes, including the use of disguised voices and the analysis of multi speaker samples.

The results achieved give indication that each algorithm works best with one particular feature extraction method. Moreover the algorithms show differences in performance when varying training and validation data. Gaussian Mixture Models benefit from having longer speech samples, whereas k-Nearest Neighbour prefers shorter samples and for the neural network the length of the speech sample does not matter much.

The best performances of the algorithms range between 88% and 94%. The tested models are not able handle unknown speakers, but a method for determining a degree of certainty for a given classification has been devised. The best performance for recognition in multi-speaker samples can be achieved by a word-by-word classification approach.

Further research could be conducted in order to improve the multi-speaker classification and to test the influence of the recording environment.