© 2013 FIZ Karlsruhe

io-port 02224910

Mak, Man-Wai; Sit, Chin-Hung; Kung, Sun-Yuan

Extraction of speaker features from different stages of DSR front-ends for distributed speaker verification.

Int. J. Speech Technol. 8, No. 1, 67-77 (2005).

Summary: The ETSI has recently published a front-end processing standard for distributed speech recognition systems. The key idea of the standard is to extract the spectral features of speech signals at the front-end terminals so that acoustic distortion caused by communication channels can be avoided. This paper investigates the effect of extracting spectral features from different stages of the front-end processing on the performance of distributed speaker verification systems. A technique that combines handset selectors with stochastic feature transformation is also employed in a back-end speaker verification system to reduce the acoustic mismatch between different handsets. Because the feature vectors obtained from the back-end server are vector quantized, the paper proposes two approaches to adding Gaussian noise to the quantized feature vectors for training the Gaussian mixture speaker models. In one approach, the variances of the Gaussian noise are made dependent on the codeword distance. In another approach, the variances are a function of the distance between some unquantized training vectors and their closest code vector. The HTIMIT corpus was used in the experiments and results based on 150 speakers show that stochastic feature transformation can be added to the back-end server for compensating transducer distortion. It is also found that better verification performance can be achieved when the LMS-based blind equalization in the standard is replaced by stochastic feature transformation.

doi:10.1007/s10772-005-4762-x