

ACOUSTICS2008/2025

Speaker adaptation combined with missing data reconstruction

Ulpu Remes, Kalle Palomäki and Mikko Kurimo

Adaptive Informatics Research Centre, Helsinki University of Technology, P.O. Box 5400, 02015 Espoo, Finland

Missing data methods offer an attractive framework for automatic speech recognition (ASR) under time-varying and unpredictable noise conditions. The motivation for the methods comes from the human speech perception and auditory scene analysis, studies on which have suggested that the more intense sound components in an auditory scene dominate the neural response to the scene. The missing data methods developed for ASR divide the noisy speech signal to speech and noise dominated regions. Speech components in the noise dominated regions are viewed as missing but may be reconstructed based on the observed components. Missing data methods have performed well in previous studies with noise corrupted speech but have not become a standard in ASR. Especially considering ASR on large vocabularies, the methods have a disadvantage: it is not straightforward to combine them with common speaker adaptation methods such as constrained maximum likelihood linear regression (CMLLR). In this work, we investigate efficient ways to combine missing data reconstruction with speaker adaptation in large vocabulary ASR task with speech data recorded in noisy real-world environments. The experiments show that speaker adaptation combined with the missing data reconstruction can improve the missing data approach in noisy large vocabulary ASR.