Multi-Environment Linear Normalization For Robust Speech Analysis in Cars

Luis Buera, Eduardo Lleida, Antonio Miguel, and Alfonso Ortega

Abstract

In this paper Phoneme-Dependent Multi-Environment Models based Linear feature Normalization, PD-MEMLIN, is presented. The target of this algorithm is learning the mismatch between clean and noisy feature vectors associated to a pair of Gaussians of the same phoneme (one for a clean model, and the other one for a noisy model), for each basic defined environment. These differences are estimated in a previous training process with stereo data. In order to compensate some of the problems of the independence assumption of the feature vectors components and the mismatch error between perfect and proposed transformations, two approaches have been proposed too: a multienvironment rotation transformation algorithm, and the use of transformed space acoustic models. The behavior of this technique was studied for speech recognition and speaker verification and identification in a real acoustic environment. The experiments were carried out with SpeechDat Car database and the results show an average improvement in speech recognition of more than 77% using PD-MEMLIN, and more than 85% using transformed space acoustic models and multi-environment rotation transformation. In speaker verification and identification, PD-MEMLIN is applied as a previous phase to clean the signal, with an average improvement in Equal-Error Rate of more than 70%, and 48.69%, respectively.