Title of paper: A novel use of residual noise model for the Parallel Model Combination

Abstract: The robust speech recognition system available for all environments has become a hotspot in speech research. The environment adaptive methods play an important part in improving the system robustness including PMC.

In this paper, PMC is investigated deeply and further developed to achieve the better performance. In general noisy environments, the channel distortion and noise corruption are both present and time-varying, the signal-to-noise ratio (SNR) often be low. PMC can get around additive noise as well as convolutive noise. However, it has some shortcomings: Firstly, even with the precise measured convolutive noise model, it is hard to get high recognition accuracy since the speech recognition is based on short-frame windowing. The effect of convolutive noise on the speech is equivalent to periodic convolution in time domain; however, actual distorted speech results from linear convolution. The combined noisy speech models using PMC cannot model the actual distorted speech precisely. Secondly, the SNR lower, the additive noise model is more dominant in combined noisy speech models. It is impossible to achieve good performance with such additive noise model. On the other hand, it is known that some efficient noise-reduction schemes, such as the spectral subtraction (SS) schemes, CMN, etc can achieve good performance in some extents. In these cases, it is assumed that the residual noises are sufficiently small so that no quantitative modeling for them is needed. However, in our experiments with noise reduction schemes, it is easy to observe the deviations of the distributions of the enhanced speech from those of the corresponding clean speech. This suggests it is useful to model the residual noises. Motivated by these reasons, we propose a new PMC: using the noise-reduced observation data, we can model the residual noise model, combine such new noise model with the clean speech models, we can construct the pseudo-clean speech models that closely match the noise-reduced test data. The new approach can incorporate with noise-reduction schemes innovatively and improve the recognition accuracy in adverse environments.

In our experiment, Cambridge’s HTK toolkit 3.0 was used as test platform with suitable modification embedding PMC algorithms implement the continuous Mandarin digit recognition. The training data were collected in clean office environment while the test data included the data contaminated by white Gaussian noise at different SNR levels and also the noisy speech collected in real environment. The results of experiment clarify the effectiveness of the proposed schemes.

The paper is organized as following: In Section 1, the influence of additive and convolutive noise on the speech is discussed respectively. In Section 2, the residual noise models are introduced based on the different noise-reduction schemes; the details of the residual noise modeling strategy are described and the PMC is modified to combine the residual models. In Section 3, the experiment results are provided to show the performance of the proposed schemes. In Section 4, the future work is suggested and the conclusion is drawn.

Keywords: Parallel Model Combination; Additive noise; Convolutive noise; Residual noise models; Pseudo clean speech models
Author’s contact details:
Cailian Miao
National Laboratory of Pattern Recognition
Institute of Automation, Chinese Academy of Sciences
Phone: +86-010-62527940-250
Fax: +86-010-62552171
Yangsheng Wang
National Laboratory of Pattern Recognition
Institute of Automation, Chinese Academy of Sciences
Phone: +86-010-62542941
Fax: +86-010-62552171