20PM1A  Speech and Audio Signal Processing I
Friday, May 20, 13:20–15:20, Room A

Chair: Makoto Shozakai, Asahi Kasei Corporation, Japan

20PM1A-1  Adaptive Nonlinear Predictive Analysis for Speech Using a Cascaded LMS-VSLMS Predictor
Hirobumi Tanaka, Saitama University, Japan
Tetsuya Shimamura, Saitama University, Japan

When we perform the linear predictive analysis to speech signals, prediction errors are inducted. To suppress these errors we often rely on nonlinear predictors. Nonlinear predictor, however, possesses disadvantages which are high complexity, slow convergence, necessity of using a large number of data samples etc. In this paper, we propose an adaptive nonlinear predictor which has a structure of the cascade of the LMS predictor and the VSLMS predictor. Experiments were conducted on continuous speeches and the proposed predictor provided superior prediction gains compared with the LMS and VSLMS predictors. Furthermore, we investigated the proposed predictor with iterative processing. As a result, we confirmed that the simple cascaded predictor provides sufficient convergence.

20PM1A-2  Waveform Reconstruction from Non-Uniform Samples with Application to Speech Coding
Prasanta Kumar Ghosh, Indian Institute of Science, India
T. V. Sreenivas, Indian Institute of Science, India

For non-stationary signals, such as speech, we show that extrema are useful non-uniform samples for a compact reconstruction. We demonstrate the effectiveness of different interpolating functions for reconstructing the signal from the extrema. We also show that such a non-uniform sample based reconstruction scheme can be used for applications like speech coding. In this context, we explore the quantization properties of the extrema information and design a good quality speech coder at 12-13 kbps.

20PM1A-3  Environmental Warping for In-Car Speech Recognition
Weifeng Li, Nagoya University, Japan
Katunobu Itou, Nagoya University, Japan
Kazuya Takeda, Nagoya University, Japan
Fumitada Itakura, Meijo University, Japan

In this paper, we present an environmental warping approach to reduce the mismatch between the acoustic conditions during training and recognition. The idea of this approach is to map the log mel-filter-bank (MFB) vector obtained from the speech in a test driving condition into the one in the target driving condition, in which the acoustical models are trained. The mapping function is obtained by training multi-layer perceptron (MLP) based neural network. In our in-car isolated word recognition experiments under 12 real car environments, the proposed approach obtained an average relative word error rate (WER) reduction of 47.6% and 17.5%, compared to the original speech and conventional speech enhancement methods, respectively.