CCIT Report #460
December 2003

Speech Enhancement Using Supergaussian Speech Models and Noncausal *A Priori* SNR Estimation

Israel Cohen

Abstract

Existing supergaussian speech models in the short-time Fourier transform domain are based on the assumption that distinct spectral components are statistically independent. The corresponding minimum mean-square error (MMSE) spectral estimators require an estimator for the *a priori* SNR. Unfortunately, the latter is often obtained by the decision-directed approach of Ephraim and Malah, which relies on the strong time-correlation between successive speech spectral components. In this paper, we extend the supergaussian speech models by taking into consideration the time-frequency correlation between spectral components. We introduce noncausal *a priori* SNR estimators for Gamma and Laplacian speech models, and derive noncausal estimators for the clean speech spectral components. The noncausal *a priori* SNR estimation consists of two major steps, which follow the rational of Kalman filtering: a "propagation" step and an "update" step. Estimates for the speech spectral variances and the instantaneous power from the previous frame are propagated in time to obtain an estimate for the spectral variance in the current frame. Subsequently, the estimate for the spectral variance is updated by computing the conditional variance of the speech spectral component, based on the underlying speech model. Experimental results demonstrate the improved performance of the proposed algorithms.

I. Introduction

Optimal estimators for speech enhancement in the short-time Fourier transform (STFT) domain are often based on a Gaussian statistical model [1]–[5]. Accordingly, the individual short-term spectral components of the speech and noise signals are modelled as statistically independent Gaussian random variables. Using this model, Ephraim and Malah derived a short-term spectral amplitude (STSA) estimator, which minimizes the mean-square error of the spectral magnitude [1], and a Log-Spectral Amplitude (LSA) estimator, which minimizes the mean-square error of the log-spectra. Wolfe and Godsill [6] derived under the same modeling assumptions three alternative suppression rules, which are based on joint maximum a posteriori (MAP) spectral amplitude and phase estimation, MAP spectral amplitude estimation, and minimum mean-square error (MMSE) spectral power estimation. The resulting suppression

The author is with the Department of Electrical Engineering, Technion - Israel Institute of Technology, Technion City, Haifa 32000, Israel (email: icohen@ee.technion.ac.il; tel.: +972-4-8294731; fax: +972-4-8295757).