CCIT Report #443
October 2003



# Relaxed Statistical Model for Speech Enhancement and *A Priori* SNR Estimation

Israel Cohen

**Abstract**

The widely-used speech enhancement method of Ephraim and Malah is based on a Gaussian statistical model, presuming spectral components are statistically independent. A major drawback is that the model assumptions conflict with the "decision-directed" *a priori* SNR estimation, which heavily relies on the time-correlation of speech spectra. In this paper, we propose a statistical model for speech enhancement that i) takes into account the time-correlation between successive speech spectral components; ii) admits consistent estimators for the *a priori* SNR and the speech spectral components; iii) retains the simplicity associated with the Ephraim-Malah statistical model; iv) provides insight into the decision-directed approach; and v) enables the extension of existing algorithms to noncausal estimation. In the proposed model, the sequence of speech spectral variances is a random process, which is correlated with the sequence of speech spectral components. Causal and noncausal estimators for the *a priori* SNR are derived in agreement with the model assumptions and the estimation of the speech spectral components. We show that a special case of the causal estimator degenerates to a "decision-directed" estimator with a *time-varying* weighting factor. Experimental results demonstrate the improved performance of the proposed algorithms.

## I. Introduction

One of the most popular methods for enhancing speech, degraded by uncorrelated additive noise, is the spectral enhancement algorithm of Ephraim and Malah [1], [2]. This algorithm and its derivatives (*e.g.*, [3]–[5]) have been applied to single-channel and multi-channel speech enhancement in speech recognition systems [6], [7], speech coders [8]–[10], digital hearing-aids [11], [12], voice activity detectors [13]–[15], and hands-free mobile communication systems [16]–[18].

Two decades ago, Ephraim and Malah proposed a statistical model for speech enhancement [2], [19]. Accordingly, the individual short-term spectral components of the speech and noise signals are modeled as statistically independent Gaussian random variables. The assumption that spectral components are statistically independent is clearly unfulfilled. However, it facilitates a mathematically tractable derivation of useful estimators for various distortion measures. In [2], Ephraim and Malah derived a short-term spectral amplitude (STSA) estimator, which

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).