Generalized multi-microphone spectral amplitude estimation based on correlated noise model

Lae-Hoon Kim, Mark Hasegawa-Johnson and Koeng-Mo Sung

Abstract

Enhancing speech contaminated by uncorrelated additive noise, when the degraded speech alone is available, has received much attention. In recent years many systems have used multi-microphone arrays for the task of speech enhancement and robust speech recognition. In this paper we introduce a generalized multi-microphone spectral amplitude estimation approach based on a model with non-negligible inter-microphone noise coherence. We demonstrate that our approach is a general foundation that can support any classical single-channel spectral estimation technique. We also present experimental results using data from the AVICAR database, spanning a wide range of noise conditions in a real moving car.

Précis

In recent years many systems have used multi-microphone arrays for the task of speech enhancement and robust speech recognition. A microphone array is usefully described as a spectral estimation tool: spectral estimation accuracy depends on the SNR (Signal to Noise Ratio), and by using beamforming we can improve this SNR. Most theoretical estimates of SNR improvement, however, are wrong, because they incorrectly assume that noises recorded at different microphones are uncorrelated. In practice, at frequencies and microphone spacings of interest in practical systems, noises are correlated. One of the very few published systems that considers inter-microphone noise coherence is that of McCowan and Bourlard, which extends the Wiener estimator for the case of multi-microphone estimation. In this paper we would like to introduce a generalized multi-microphone spectral amplitude estimation approach based on a correlated noise model. Our estimator is general in the sense that the proposed method can be used to extend any classical single-channel spectral estimation technique to an array with correlated noise; we demonstrate extension, in particular, of the Ephraim & Malah amplitude and log-amplitude estimators. We first review a derivation showing that mutual information between the source and the multi-channel measured signal is equal to mutual information between the source and a one-channel sufficient statistic. The one-channel sufficient statistic has the form of an MVDR (Minimum Variance Distortionless Response) beamformer. The form of the sufficient statistic depends on knowledge of the channel and the noise correlation matrix; accuracy of the method depends on our ability to estimate the noise correlation matrix as accurately as possible. Previous work based on the MVDR has ignored the noise correlation, and assumed that the noise spectra are independent of each other over the whole frequency range. To consider this correlation in

the estimation process, we introduce a coherence function. In the ideal isotropic noise field, this coherence function is known to be simply described as a *sinc* function of the frequency and the distance between microphones. In a real situation, noise is not isotropic, therefore we have to estimate this correlation matrix as well as the noise spectral amplitude. In this paper we use the ideal *sinc* isotropic noise coherence as an initial value, and then we update it during every non-speech frame in a recursive manner using observations of the cross-power spectral matrix. We present experimental results using data from the AVICAR database recorded in various noise conditions in a real moving car. Inside of a car, low frequency noise is severely boosted, therefore correlated noise in the lower frequency range should not be ignored. The proposed method is used to compute the minimum mean-squared error estimates of the spectral amplitude (MMSE-SAE) and of the log spectral amplitude (MMSE-logSA). MMSE-logSA is particularly applicable to speech recognition, because the cepstral features used in speech recognition are a linear transform of the log spectrum. Estimation results and isolated digit recognition results are compared using single-microphone input, using an MVDR with the typical uncorrelated-noise assumptions, using the ideal isotropic noise coherence, and using the proposed recursive coherence estimate, demonstrating the superiority of each of these systems over its predecessor.