Speech Enhancement With Bivariate Gamma Model

Atanu Saha, Tetsuya Shimamura
Graduate School of Science and Engineering, Saitama University, Saitama, Japan
{saha, shima}@sie.ics.saitama-u.ac.jp

Abstract
In this paper, we propose a family of Bayesian estimators for single channel speech enhancement. Three different kinds of enhancement estimators are derived from the statistics of bivariate Gamma model. In the Gamma model, it is assumed that the spectral components are correlated with each other. The experimental results show that the proposed estimators are effective for speech enhancement compared to state-of-the-art estimators.

Index Terms: speech enhancement, bivariate Gamma model, Gaussian distribution, correlated spectral components

1. Introduction
Speech enhancement algorithms are generally concerned with improving the quality of speech corrupted by additive background noise. Among the existing algorithms for single channel speech enhancement, Bayesian estimators of the amplitude rather than the complex spectrum have received a lot of attention in the last decade.

A well-known Bayesian estimator of the amplitude spectrum is the minimum mean square error (MMSE) estimator that minimizes the conditional expectation of a squared-error cost function [1]. The squared-error cost function in logarithmic domain, resulting in log-spectral amplitude (LSA) estimator [2], has shown itself to be more effective in reducing the annoying musical noise.

A generalization of these cost functions was made in $\beta$-order MMSE estimator [3]. However, the algorithms mentioned above assume that the spectral components of the clean speech are uncorrelated. Unfortunately, this assumption is not exact due to two reasons [4]. Firstly, the finite length of the analysis window used in short-time processing introduces some correlation among frequency components. Secondly, the voiced speech that has inherent harmonics is a significant source of correlation between frequency components.

In an effort for developing enhancement algorithms by assuming correlated frequency components, Bayesian estimators of the amplitude and complex spectrum were derived in [4, 5]. These approaches in discrete-time Fourier transform (DFT) domain assume that the clean speech and noise DFT coefficients are complex Gaussian distributed. The recent studies [6, 7], however, show that the clean speech DFT coefficients have a super-Gaussian such as Laplacian or Gamma distribution behavior.

In this paper, we derive a family of speech enhancement estimators based on bivariate Gamma distribution that assumes the spectral components to be correlated with each other. Three different kinds of estimators are derived from the statistics of the bivariate Gamma model. This is typically done by considering the sum, the product and the ratio of two correlated variables. The experimental results show that the proposed estimators are effective for speech enhancement.

The rest of the paper is organized as follows. Section 2 describes the basic notations. Section 3 derives the proposed estimators utilizing the bivariate Gamma model. Section 4 presents the experimental results while Section 5 concludes the paper.

2. Basic Notations
In this section, we describe the preliminary notations that we will use through out the paper.

Let us assume that the noise signal is additive and uncorrelated with speech signal. The noisy observation can then be expressed as

$$y_i = s_i + d_i$$

where $y_i$, $s_i$, and $d_i$ are the $N$-dimensional noisy speech, clean speech and noise vectors, respectively, for time frame $i$. For notational convenience, we will omit the frame index $i$ and consider the processing of one particular frame. By denoting the $N$-point DFT matrix by $F$, the DFT coefficients of the noisy speech $y$ can then be written as

$$Y = F^H y = F^H s + F^H d = S + D$$

where $S$ and $D$ denote the DFT coefficients of the clean speech vector $s$ and the noise vector $d$, respectively. The elements of $S$ are $S_k = A_k e^{j\theta_k}$, where $A_k$ ($1 \leq k \leq N$) is the spectral amplitude of speech and $\theta \in [-\pi, \pi)$ is the associated phase. We also define $A = [A_1 A_2 \ldots A_N]^T$ and $\theta = [\theta_1 \theta_2 \ldots \theta_N]^T$.

In a similar fashion, the elements of $Y$ are defined as $Y_k = R_k e^{j\psi_k}$, where $R_k$ and $\psi_k$ denote the spectral amplitude and phase of the noisy speech.
The Bayesian spectral amplitude estimator minimizes the conditional expectation of a cost function, \( E[C(\hat{A}, A)] \), where \( \hat{A} \) denotes the estimated spectral amplitudes of \( A \). The estimator is, then, combined with the phase of the noisy speech, for each frequency, to yield the estimator of \( S \), that is,

\[
\hat{A} = [\hat{A}_1 e^{j\psi_1} \ldots \hat{A}_N e^{j\psi_N}]^T. \tag{3}
\]

The corresponding time domain estimator can be obtained by performing an inverse Fourier transform for each frame which is then combined using the overlap-add method. The motivation is, thus, to derive the gain function \( G_k \) so that it satisfies the estimator as \( \hat{A}_k = G_k R_k \).

3. Bivariate Gamma Model

In this section, we derive the gain function \( G_k \). A bivariate Gamma model, which assumes the frequency components to be correlated, is employed in the proposed method to derive the gain function.

As is well known, the probability density function of a univariate Gamma distribution for a random variable \( x \) is given by

\[
f(x) = \frac{1}{\beta^\nu \Gamma(\nu)} x^{\nu-1} e^{-x/\beta} \tag{4}
\]

where \( \Gamma(.) \) is the Gamma function, and \( \nu \) and \( \beta \) denote shaping and scaling parameters, respectively. Based on the univariate Gamma distribution, Izawa [8] proposed a bivariate Gamma model that is constructed from Gamma marginals for different scaling and shaping parameters. The marginal distribution of a bivariate Gamma distribution in terms of spectral amplitudes is defined as

\[
f_m(A_m) = \frac{1}{\beta_m^\nu \Gamma(\nu)} A_m^{\nu-1} e^{-A_m/\beta_m} \tag{5}
\]

where \( m \) is the variable specified by 1 and 2. The joint probability density function for two correlated variables can be defined by (6), shown in the top of the next page. In (6), \( I_{\nu-1}(.) \) is the modified Bessel function of the first kind of order \( \nu - 1 \) and \( \rho (0 \leq \rho < 1) \) is the correlation factor.

In order to obtain the gain function, we now consider the sum, the product and the ratio of two correlated variables.

3.1. Sum of two correlated variables

The distribution of the sum of two correlated variables with the bivariate Gamma distribution can be obtained under the condition that \( \beta_1 = \beta_2 \). In (6), substitute \( \beta_1 = \beta_2 = \beta \) and let the transformation be defined by \( l = A_1 + A_2 \) and \( z = A_2 \). Since the probability element is not changed by this linear transformation, the joint distribution of the new variables \( l \) and \( z \) is given by expression (7), shown in the next page. In expression (7), substituting \( z - l/2 = (l/2) \cos \theta \) and integrate with respect to \( \theta \) from 0 to \( \pi/2 \), the distribution of \( l \), \( f(l) \), becomes in (8), shown in the next page. By specifying the Gamma priors of \( A_k \) in (8), and multiplying (8) by \( A_k^\tau \) (where \( \tau \) is a real parameter) and integrating with respect to \( A_k \) from 0 to \( \infty \), we obtain the \( \tau \)th conditional moment for all frequency components as

\[
E[A_k^\tau | Y] = \beta^\tau (1 - \mu_k)^{\nu + \tau} \frac{\Gamma \left( \frac{2\nu + \tau + 1}{2}, \frac{2\nu + \tau + 1}{2}; \nu + 1; \mu_k \right) R_k^\tau}{\Gamma(\nu + 1)} \tag{9}
\]

where \( F(.) \) is the hypergeometric function [9] and \( \mu_k \) is defined as \( \mu_k = \xi_k \gamma_k / (\nu + \xi_k) \). In the definition of \( \mu_k \), \( \xi_k = \lambda_1(k)/\lambda_2(k) \) is the a priori SNR (signal to noise ratio) in which \( \lambda_1(k) \) and \( \lambda_2(k) \) denote the variance of speech and noise, respectively, and \( \gamma_k = R_k^2 / \lambda_2(k) \) is called the a posteriori SNR.

In order to determine the gain function, we now choose the conditional expectation of the estimator proposed in [3] as

\[
\hat{A}_k = E [A_k^\tau | Y]^{1/\tau}. \tag{10}
\]

Note that the above estimator is transformed to the LSA estimator for a small and positive value of \( \tau \) [3]. In our experiment, \( \tau \) is also approximated as a very small and positive value so that it keeps the characteristics of the LSA estimator. Actually, the value of \( \tau \) in our experiment is set to 0.03. Substituting (9) in (10), the gain function \( G_k^{\tau} \) via \( A_k^\tau = G_k^{\tau} R_k \) is determined as

\[
G_k^{\tau} = \beta (1 - \mu_k)^{1+\nu/\tau} \left\{ \frac{\Gamma(\nu + 1) \Gamma \left( \frac{2\nu + \tau + 1}{2}, \frac{2\nu + \tau + 1}{2}; \nu + 1; \mu_k \right) R_k^\tau}{\Gamma(\nu + 1)} \right\}^{1/\tau} \tag{11}
\]

where the superscript is used to denote the gain function \( G_k^{\tau} \) obtained by considering the sum of two correlated variables.

3.2. Product of two correlated variables

The distribution of the product of two correlated variables can be obtained from (6) by considering the transformation defined by \( l = A_1 A_2 \) and \( z = A_2 \). The distribution of the new variables \( l \) and \( z \) is, then, given by expression (12), as shown in the next page. The expression (12) provides that \( \beta_m = 1 \). The distribution of \( l, f(l) \), can now be obtained by integrating (12) with respect to \( z \) as shown in (13). In (13), \( K_0 \) is the modified Bessel function of the second kind. The \( \tau \)th conditional moment can now be obtained by specifying the Gamma priors of \( A_k \) in (13) and integrating with respect to \( A_k \) as

\[
E[A_k^\tau | Y] = \frac{\Gamma(\nu + \tau)^2}{\Gamma(\nu)^2} F (-\tau, -\tau; \nu; \mu_k) R_k^\tau. \tag{14}
\]

INTERSPEECH 2012 936
\[ f(A_1, A_2) = \frac{(A_1 A_2)^{\frac{\nu}{2}}}{\Gamma(\nu)(\beta_1 \beta_2)\sqrt{1 - \rho}} \exp \left[ -\frac{1}{2} \left( \frac{A_1}{\beta_1} + \frac{A_2}{\beta_2} \right) \right] I_{\nu-1} \left( \frac{2\sqrt{\rho}}{\beta_1 \beta_2 (1 - \rho) \sqrt{A_1 A_2}} \right). \]  

\[ f(l) = \frac{1}{\Gamma(\nu)(1 - \rho)^{\nu + 1}} \exp \left[ -\frac{l}{\beta(1 - \rho)} \right] \left\{ \frac{l^2}{4} - \left( \frac{z - 1}{2} \right)^2 \right\}^{(\nu - 1)/2} I_{\nu - 1} \left( \frac{2\sqrt{\nu}}{\beta(1 - \rho) \sqrt{4 - \left( \frac{z - 1}{2} \right)^2}} \right). \]  

Substituting (14) into (10) leads to the gain function \( G_k^p \) via \( A_k^p = G_k^p R_k \) as

\[ G_k^p = \left\{ \frac{\Gamma(\nu + 1/2)^2}{\Gamma(\nu)^2} F(-\tau, -\tau; \nu; \mu_k) \right\}^{1/2}. \]  

where the superscript is used to denote the product of two correlated variables.

### 3.3. Ratio of two correlated variables

In (6), we now consider the transformation that is defined by \( l = \frac{A_1}{A_2} \) and \( z = \frac{A_2}{A_1} \). The distribution of the new variables \( l \) and \( z \) is given by the expression (16) (provides that \( \beta_m = 2 \)). To obtain the \( \tau \)-th conditional moment, by specifying the Gamma priors of \( A_k \) in (16), multiplying by \( A_k^p \) and integrating, then we have

\[ E[A_k^p|Y] = \frac{\Gamma(\nu + \tau)\Gamma(\nu - \tau)}{\Gamma(\nu)^2} F(-\tau, -\tau; \nu; \mu_k) R_k^p. \]  

Substituting (17) into (10) leads to the estimator \( A_k^g = G_k^g R_k \), where the gain function \( G_k^g \) is defined as

\[ G_k^g = \left\{ \frac{\Gamma(\nu + 1/2)^2}{\Gamma(\nu)^2} F(-\tau, -\tau; \nu; \mu_k) \right\}^{1/2} \]  

where the superscript is used for denoting the ratio of two correlated variables.

The asymptotic behavior for large values of \( \gamma_k \) in all cases is given by \( G_k^g |_{\gamma_k \gg 1} \approx 1 \). As can be observed that the obtained gain functions are functions of both the \( a \) priori SNR \( \xi_k \) and a \( a \) posteriori SNR \( \gamma_k \). Fig. 1 illustrates the gain functions \( G_k^p, G_k^p, \gamma_k \) and \( G_k^g \) as the functions of instantaneous SNR, \( \gamma_k - 1 \), for a fixed value of \( \xi_k (\xi_k = 0 \text{ dB}) \) for several values of \( \nu \). The amount of suppression seems to be dependent on the value of the shaping parameter \( \nu \). Small values of \( \nu \) provide higher attenuation while large values of \( \nu \) provide relatively lower attenuation. However, speech distortion may be introduced at very small values of \( \nu \). For better compromise between noise reduction and speech distortion, \( \nu = 0.6 \) has been used in the experiment.

### 4. Experimental Results

In order to assess the effectiveness of the proposed estimators, the NOIZEUS speech corpus [10] is used in the experiments. The corpus comes from different types of
\[ \frac{1}{\Gamma(\nu)2^{\nu+1}(1-\rho)\rho^{\nu-1}/2}z^{\nu l(\nu-1)/2}\exp \left[ -\frac{1}{2(1-\rho)}(l-1) z \right] I_{\nu-1} \left( \sqrt{\frac{\rho}{1-\rho}} z \right). \]  

\section{5. Conclusions}

In this paper we have proposed a family of Bayesian estimators based on bivariate Gamma model. The experimental results show that the proposed estimators provide marked improvements in overall quality. This is mainly due to the use of Gamma distribution which assumes that the correlation exists among spectral components. The use of the power exponent in the cost function as a small value keeps the characteristics of the LSA estimator. These facts make the proposed method to perform well by reducing the certain amount of noise while keeping the speech components as undistorted as possible.

\section{6. References}


