

# LOG-SPECTRAL AMPLITUDE ESTIMATION WITH GENERALIZED GAMMA DISTRIBUTIONS FOR SPEECH ENHANCEMENT

Bengt J. Borgström and Abeer Alwan\*

University of California, Los Angeles  
Department of Electrical Engineering

## ABSTRACT

This paper presents a family of log-spectral amplitude (LSA) estimators for speech enhancement. Generalized Gamma distributed (GGD) priors are assumed for speech short-time spectral amplitudes (STSAs), providing mathematical flexibility in capturing the statistical behavior of speech. Although solutions are not obtainable in closed-form, estimators are expressed as limits, and can be efficiently approximated. When applied to the Noizeus database [8], proposed estimators are shown to provide improvements in segmental signal-to-noise ratio (SSNR) and COSH distance [14], relative to the LSA estimator proposed by Ephraim and Malah [2].

**Index Terms**— Speech Enhancement, Log-spectral Amplitude Estimator, Generalized Gamma Distribution.

## 1. INTRODUCTION

Single-channel speech enhancement aims to minimize the effect of background noise on speech signals recorded in an acoustically unfavorable environment. In typical speech communication systems, signals can be enhanced at the transmitter-end in order to increase the perceptual quality experienced at the receiver-end.

Bayesian approaches to speech enhancement minimize the conditional expectation of a cost function. In [1], Ephraim and Malah proposed the well-known minimum mean square-error (MMSE) estimator, which minimizes the spectral amplitude mean square-error (MSE). In [2], Ephraim and Malah discuss the use of a perceptually-motivated cost function. Specifically, they minimized the log-domain MSE, resulting in the log-spectral amplitude (LSA) estimator.

The previously mentioned studies assumed independent Gaussian *a priori* distributions for real and imaginary speech and noise components. More recent studies have explored empirical histograms of speech spectral amplitudes, which show a super-Gaussian trend ([7], [6]). In [9], Erkelens et al. utilize the generalized Gamma distribution (GGD) to model speech spectral amplitudes.

Although statistical modeling of speech has been thoroughly explored, few studies have combined the use of super-Gaussian priors and perceptually motivated cost functions. In [12], Hendriks et al. propose an LSA estimator which assumes  $\chi^2$ -distributed speech. Due to the induced mathematical complexity, a closed-form solution is not obtainable, and numerical approximation is instead required. In this paper, we combine statistically flexible GGD speech priors with the perceptually-motivated LSA MMSE cost function. We provide a general framework for determining log-spectral amplitude estimators, of which the solution from [2] is a special case. Although solutions are not obtainable in closed-form, estimators are expressed as limits and can be approximated efficiently.

This paper is organized as follows: In Section 2, the assumed statistical model is discussed. In Section 3, we propose novel LSA estimators assuming GGD priors. Section 4 provides experimental results, followed by conclusions in Section 5.

## 2. STATISTICAL FRAMEWORK

In this study, an additive noise model is assumed, expressed in the short-time frequency domain as  $Y_k = X_k + N_k$ , where  $Y_k$  is the observed speech,  $X_k$  and  $N_k$  are the underlying speech and noise components, respectively, and  $k$  denotes channel index. Here,  $X_k$  and  $N_k$  are random processes, whereas  $Y_k$  is deterministic. Let the frequency representations of observed and clean speech, as well as noise, be decomposed into amplitude and phase components according to

$$Y_k = R_k \exp(j\theta_k) \quad (1)$$

$$X_k = A_k \exp(j\epsilon_k) \quad (2)$$

$$N_k = D_k \exp(j\varphi_k) \quad (3)$$

Here,  $R_k$ ,  $A_k$ , and  $D_k$  refer to the short-time spectral amplitudes of observed speech, clean speech, and noise, respectively, and  $\theta_k$ ,  $\epsilon_k$ , and  $\varphi_k$  refer to the corresponding phases. Real and imaginary components of channel-specific noise processes are modelled as independent zero-mean Gaussian random variables with variances  $\sigma_N^2(k)/2$ , as in ([1], [2]), leading to

$$p(N_k) = \frac{1}{\pi\sigma_N^2(k)} \exp\left(-\frac{D_k^2}{\sigma_N^2(k)}\right) \quad (4)$$

The conditional probability of  $Y_k$  given  $A_k$  can be determined via marginalization with respect to  $\epsilon_k$  (see [3])

$$p(Y_k|A_k) = \frac{1}{\pi\sigma_N^2(k)} I_0\left(\frac{2A_k R_k}{\sigma_N^2(k)}\right) \exp\left(-\frac{R_k^2 + A_k^2}{\sigma_N^2(k)}\right) \quad (5)$$

where  $I_\nu(\cdot)$  denotes the  $\nu^{\text{th}}$ -order modified Bessel function of the first kind [4].

Traditional studies such as [1] and [2] model speech spectral (DFT) coefficients as Gaussian processes, corresponding to Rayleigh-distributed spectral amplitudes. In this study, however, we explore the use of generalized Gamma distributed (GGD) speech spectral amplitude priors, which are shown to more accurately approximate empirical histograms of speech ([6],[7])

$$p(A_k) = \frac{\zeta\lambda^\nu}{\Gamma(\nu)} A_k^{\zeta\nu-1} \exp(-\lambda A_k^\zeta), \text{ for } \lambda, \nu, \zeta > 0 \quad (6)$$

where  $\Gamma$  denotes the Gamma function [4]. Here,  $\zeta$  and  $\nu$  serve as shaping parameters, and influence the general behavior of the resulting distribution. In this study,  $\zeta$  is constrained as  $\zeta \in \{1, 2\}$ , as this simplifies the derivation of optimal estimators. Since GGDs

\*Work was supported in part by the NSF.

are defined by multiple shape parameters, they provide mathematical flexibility in capturing the statistical behavior of speech. The scaling parameter  $\lambda$  is related to the noncentral second moment of the distribution:

$$\sigma_X^2 = E[A_k^2] = \begin{cases} \frac{\nu(\nu+1)}{\lambda^2}, & \text{if } \zeta = 1 \\ \frac{\nu}{\lambda}, & \text{if } \zeta = 2 \end{cases} \quad (7)$$

Note that for  $\zeta=2$  and  $\nu=1$ , the GGD reduces to the commonly used Rayleigh prior. We define *a priori* and *a posteriori* SNRs,  $\xi_k$  and  $\gamma_k$  respectively, as in [3]:

$$\xi_k = \frac{E[A_k^2]}{E[D_k^2]} = \frac{\sigma_x^2(k)}{\sigma_N^2(k)} \quad (8)$$

$$\gamma_k = \frac{R_k^2}{E[D_k^2]} = \frac{R_k^2}{\sigma_N^2(k)}. \quad (9)$$

### 3. LOG-SPECTRAL AMPLITUDE ESTIMATION

In this section, we derive LSA estimators with generalized Gamma distributed speech priors. Bayesian spectral amplitude estimators for speech enhancement are derived by minimizing the conditional expectation of a cost function  $C(A_k, \hat{A}_k)$ , where  $\hat{A}_k$  denotes the estimated spectral amplitude and  $\hat{A}_k^o$  denotes the optimal estimate

$$\hat{A}_k^o = \arg \min_{\hat{A}_k} E[C(A_k, \hat{A}_k) | Y_k] \quad (10)$$

In [2], Ephraim and Malah propose a cost function which minimizes the log-domain MSE

$$C(A_k, \hat{A}_k) = (\log A_k - \log \hat{A}_k)^2 \quad (11)$$

In [2], it is shown that manipulation of the moment-generating function of  $\log A_k$  given  $Y_k$  results in the estimated spectral amplitude

$$\hat{A}_k^o = \exp\left(\frac{d}{d\tau} E[A_k^\tau | Y_k]\right) \Big|_{\tau=0} \quad (12)$$

In statistical approaches to speech enhancement, estimators can be efficiently described by a gain function  $G(\xi_k, \gamma_k) = \hat{A}_k^o / R_k$ , so that enhanced speech is determined by  $G(\xi_k, \gamma_k) Y_k$ . This leads to the log-spectral amplitude estimator

$$G_{LSA}(\xi_k, \gamma_k) = \frac{1}{R_k} \exp\left(\frac{d}{d\tau} E[A_k^\tau | Y_k]\right) \Big|_{\tau=0} \quad (13)$$

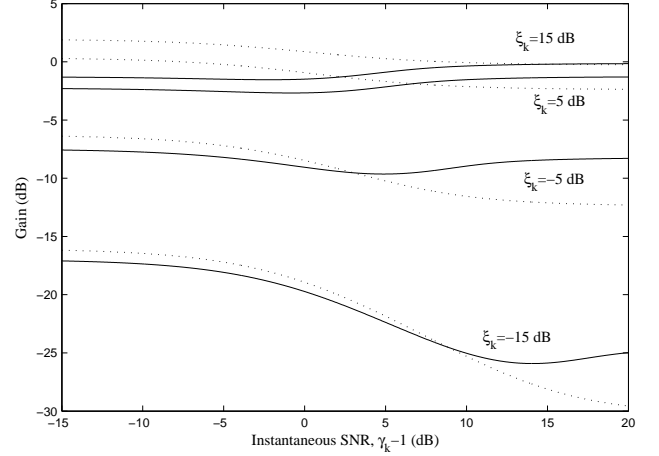
Note that Eq. 13 is equivalent to

$$\begin{aligned} G_{LSA}(\xi_k, \gamma_k) &= \lim_{\tau \rightarrow 0} \frac{1}{R_k} \exp\left(\frac{\frac{d}{d\tau} E[A_k^\tau | Y_k]}{E[A_k^\tau | Y_k]}\right) \\ &= \lim_{\tau \rightarrow 0} \frac{1}{R_k} \exp\left(\frac{d}{d\tau} \log E[A_k^\tau | Y_k]\right) \end{aligned} \quad (14)$$

Eq. 14 can be expressed as

$$\begin{aligned} G_{LSA}(\xi_k, \gamma_k) &= \lim_{\tau \rightarrow 0} \frac{1}{R_k} \exp\left(\frac{\frac{d}{d\tau} \log E[A_k^\tau | Y_k]}{\frac{d}{d\tau} \tau}\right) \\ &= \lim_{\tau \rightarrow 0} \frac{1}{R_k} E[A_k^\tau | Y_k]^{1/\tau} \end{aligned} \quad (15)$$

The second expression in Eq. 15 is obtained by applying L'Hôpital's rule. It is interesting to note that the gain function in Eq. 15 includes



**Fig. 1.** Gain curves for the log-spectral amplitude estimator  $G_{LSA}^{(2)}$  for  $\nu=1.0$  (dotted line) and  $\nu=0.5$  (solid line)

a special case of the  $\beta$ -order MMSE solution proposed by You et al. in [11].

Determining the limit in Eq. 15 is a non-trivial task, and generally requires numerical approximation. For example, in [12], Hendriks et al. utilize a truncated Taylor series to approximate a LSA estimator. In this study, we instead estimate Eq. 15 without relying on numerical approximation

$$G_{LSA}(\xi_k, \gamma_k) = \frac{E[A_k^\tau | Y_k]^{1/\tau}}{R_k} \Big|_{\tau=\delta}, \text{ where } 0 < \delta \ll 1 \quad (16)$$

During implementation,  $\delta=10^{-4}$  was used. Note that the approximation presented in Eq. 16 is given in a generalized form, and is not dependent on specific statistical models.

#### 3.1. The case for GGD priors with $\zeta=2$

If the GGD prior of  $A_k$  is specified such that  $\zeta=2$ , the  $\tau^{th}$  conditional moment is expressed as

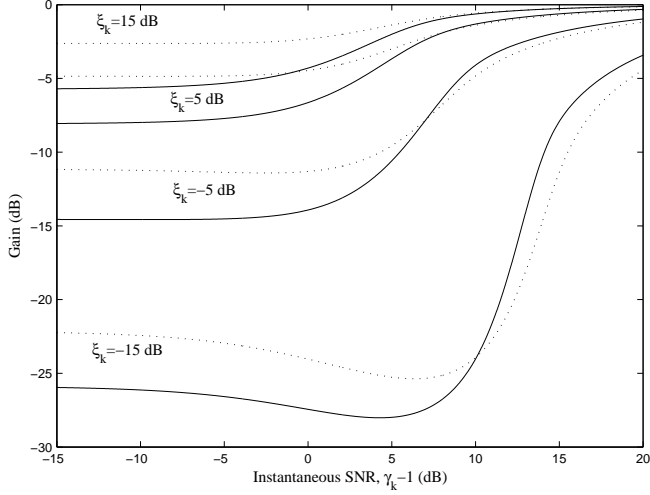
$$\begin{aligned} E[A_k^\tau | Y_k] &= \frac{\int_0^\infty A_k^{2\nu+\tau-1} I_0\left(\frac{2A_k R_k}{\sigma_N^2(k)}\right) \exp\left(\frac{A_k^2 + (\alpha_k/\gamma_k)^2 R_k^2}{(\alpha_k/\gamma_k)\sigma_N^2(k)}\right) dA_k}{\int_0^\infty A_k^{2\nu-1} I_0\left(\frac{2A_k R_k}{\sigma_N^2(k)}\right) \exp\left(\frac{A_k^2 + (\alpha_k/\gamma_k)^2 R_k^2}{(\alpha_k/\gamma_k)\sigma_N^2(k)}\right) dA_k} \end{aligned} \quad (17)$$

where  $\alpha_k = \xi_k \gamma_k / (\nu + \xi_k)$ . Both the numerator and denominator integrals in Eq. 17 are moments of a Rician distribution, allowing the following simplification

$$\begin{aligned} E[A_k^\tau | Y_k] &= \left(\frac{\sqrt{\alpha_k} R_k}{\gamma_k}\right)^\tau \frac{\Gamma(\nu + \tau/2)}{\Gamma(\nu)} \\ &\quad \times \frac{{}_1F_1(-\nu + 1 - \tau/2; 1; -\alpha_k)}{{}_1F_1(-\nu + 1; 1; -\alpha_k)} \end{aligned} \quad (18)$$

where  ${}_1F_1$  denotes the confluent hypergeometric function [4]. Substituting Eq. 18 into Eq. 16 leads to

$$\begin{aligned} G_{LSA}^{(2)}(\xi_k, \gamma_k) &= \frac{\sqrt{\alpha_k}}{\gamma_k} \left(\frac{\Gamma(\nu + \tau/2) {}_1F_1(-\nu + 1 - \tau/2; 1; -\alpha_k)}{\Gamma(\nu) {}_1F_1(-\nu + 1; 1; -\alpha_k)}\right)^{1/\tau} \Big|_{\tau=\delta} \end{aligned} \quad (19)$$



**Fig. 2.** Gain curves for the log-spectral amplitude estimator  $G_{LSA}^{(1)}$  for  $\nu=1.0$  (dotted line) and  $\nu=0.75$  (solid line)

where the superscript is used to denote  $\zeta=2$ . It is interesting to examine the asymptotic behavior of  $G_{LSA}^{(2)}$  for large values of  $\gamma_k$ . Using the following approximation from [4]

$${}_1F_1(a; 1; -z) \Big|_{z \gg 1} \approx \frac{z^{-a}}{\Gamma(1-a)} \quad (20)$$

the gain function  $G_{LSA}^{(2)}$  converges to

$$G_{LSA}^{(2)}(\xi_k, \gamma_k) \Big|_{\gamma_k \gg 1} \approx \frac{\xi_k}{\nu + \xi_k} \quad (21)$$

which can be considered a generalized form of the Wiener Filter.

### 3.2. The case for GGD priors with $\zeta=1$

If  $A_k$  is assumed to follow a GGD with  $\zeta=1$ , the  $\tau^{th}$  conditional moment is expressed as

$$E[A_k^\tau | Y_k] = \frac{\int_0^\infty A_k^{\nu+\tau-1} I_0\left(\frac{2A_k R_k}{\sigma_n^2(k)}\right) \exp\left(-\frac{A_k^2}{\sigma_n^2(k)} - \frac{\sqrt{\nu(\nu+1)}A_k}{\sigma_x(k)}\right) dA_k}{\int_0^\infty A_k^{\nu-1} I_0\left(\frac{2A_k R_k}{\sigma_n^2(k)}\right) \exp\left(-\frac{A_k^2}{\sigma_n^2(k)} - \frac{\sqrt{\nu(\nu+1)}A_k}{\sigma_x(k)}\right) dA_k} \quad (22)$$

Applying the large-value approximation of  $I_0$  from [4]

$$I_0(z) \approx \frac{1}{\sqrt{2\pi z}} \exp(z), \text{ for } z > 0 \quad (23)$$

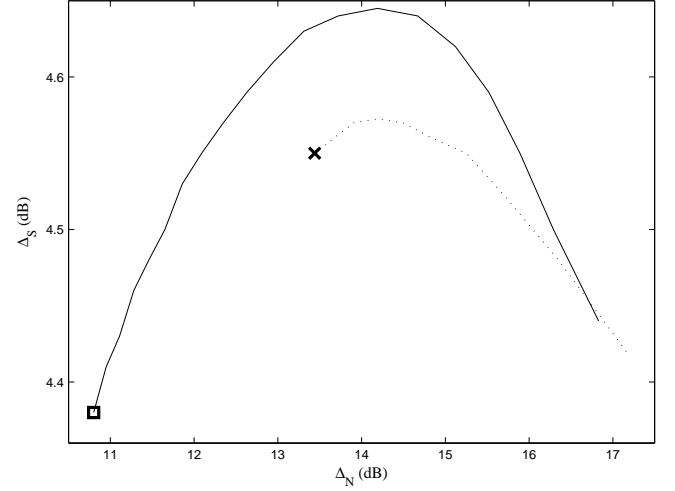
leads to

$$E[A_k^\tau | Y_k] = \frac{\int_0^\infty A_k^{\nu+\tau-3/2} \exp\left(-\frac{A_k^2}{\sigma_n^2(k)} - \mu_k A_k\right) dA_k}{\int_0^\infty A_k^{\nu-3/2} \exp\left(-\frac{A_k^2}{\sigma_n^2(k)} - \mu_k A_k\right) dA_k} \quad (24)$$

where

$$\mu_k = \sqrt{2\gamma_k} - \sqrt{\frac{\nu(\nu+1)}{2\xi_k}} \quad (25)$$

Using the  $\nu^{th}$ -order parabolic cylinder function  $D_\nu$  [4], Eq. 24 becomes



**Fig. 3.** Performance curves for the proposed family of estimators, obtained on the Noizeus database.  $\Delta_S$  and  $\Delta_N$  denote SSNR improvements in active speech and noise frames, respectively. Curves are plotted as a function of GGD shape parameter  $\nu$ , thereby illustrating the range of possible operational points for the proposed family of estimators. Cases for  $\zeta=2$  (solid line) and  $\zeta=1$  (dotted line) are shown. Marked ends ( $\times$  or  $\square$ ) correspond to  $\nu=1.0$ . Note that the operational point denoted by  $\square$  represents the LSA estimator [2].

$$E[A_k^\tau | Y_k] = \left(\frac{R_k}{\sqrt{2\gamma_k}}\right)^\tau \frac{\Gamma(\nu + \tau - \frac{1}{2}) D_{(-\nu-\tau+\frac{1}{2})}(-\mu_k)}{\Gamma(\nu - \frac{1}{2}) D_{(-\nu+\frac{1}{2})}(-\mu_k)} \quad (26)$$

Substituting Eq. 26 into Eq. 16 leads to

$$G_{LSA}^{(1)}(\xi_k, \gamma_k) = \frac{1}{\sqrt{2\gamma_k}} \left(\frac{\Gamma(\nu + \tau - \frac{1}{2}) D_{(-\nu-\tau+\frac{1}{2})}(-\mu_k)}{\Gamma(\nu - \frac{1}{2}) D_{(-\nu+\frac{1}{2})}(-\mu_k)}\right)^{1/\tau} \Big|_{\tau=\delta} \quad (27)$$

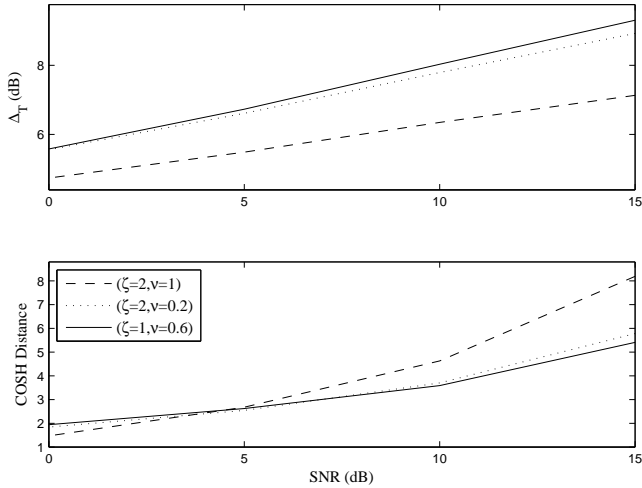
Note that due to the  $\Gamma$  term in the denominator, GGD shape parameters must be chosen according to  $\nu > 0/5$ . The asymptotic behavior for the case of  $\zeta_k=1$  is given by

$$G_{LSA}^{(1)}(\xi_k, \gamma_k) \Big|_{\gamma_k \gg 1} \approx 1 \quad (28)$$

Figures 1 and 2 illustrate gain curves for the  $G_{LSA}^{(2)}$  and  $G_{LSA}^{(1)}$  estimators, respectively, for various values of  $\nu$ . It can be observed that relative to the case for  $\zeta=2$ , the estimator corresponding to  $\zeta=1$  provides significantly decreased attenuation for large  $\gamma_k$  and increased attenuation for small  $\gamma_k$ . Furthermore, in both cases, reducing the value of the shape parameter  $\nu$  leads to decreased attenuation for large  $\gamma_k$  and increased attenuation for small  $\gamma_k$ .

## 4. EXPERIMENTAL RESULTS

In order to assess the success of the proposed estimators, we apply them to the Noizeus database [8], which is comprised of phonetically balanced utterances, and includes 8 types of non-stationary additive



**Fig. 4.** Speech enhancement results for log-spectral amplitude estimators as a function of input SNR. The top panel illustrates SSNR scores ( $\Delta_T$ ), whereas the bottom panel illustrates COSH distance measures.

noise. Speech enhancement techniques are applied to the same 30 utterances and results were averaged across noise types. Estimators are embedded within the speech enhancement system provided by [15]. A gain floor of -26 dB was applied. As a quantitative metric to assess speech quality, we measure improvements in global segmental SNR ( $\Delta_T$ ). To study the effect of enhancement on active and inactive speech frames separately, we measure the SSNR improvement of active ( $\Delta_S$ ) and inactive ( $\Delta_N$ ) speech frames, as in [7]. Additionally, we measure the distortion of enhanced speech signals using the COSH distance proposed in [14].

Figure 3 illustrates SSNR-related measures obtained by proposed enhancement solutions on the Noizeus database. Performance curves are plotted as a function of GGD shape parameter  $\nu$ , thereby illustrating the range of possible operational points. Cases for  $\zeta=2$  (solid line) and  $\zeta=1$  (dotted line) are shown. Marked ends ( $\times$  or  $\square$ ) correspond to  $\nu=1.0$ . The operational point denoted by  $\square$  represents the LSA estimator from [2].

Figure 4 provides enhancement results for the proposed estimators, as a function of input SNR. The top panel illustrates SSNR scores, whereas the bottom panel illustrates COSH distance measures. The shape parameter pairs ( $\zeta=2, \nu=0.2$ ) and ( $\zeta=1, \nu=0.6$ ) correspond to the optimal points obtained from Figure 3. Note that the shape parameter combination ( $\zeta=2, \nu=1$ ) corresponds to the LSA estimator from [2].

The benefit of the proposed family of estimators can be observed in Figures 3 and 4. Specifically, the flexibility of assuming GGD speech priors is seen to provide improvements in segmental SNR and COSH distance, when applied to the Noizeus database. Informal listening tests proved consistent with quantitative results.

## 5. CONCLUSION

This paper proposed a family of log-spectral amplitude estimators for speech enhancement. Generalized Gamma distributions were assumed for speech spectral amplitudes, due to their mathematical flexibility. Solutions were expressed as limits, thereby allowing ef-

ficient approximation. When applied to the Noizeus database, proposed estimators provided improvements in SSNR and COSH distance, relative to the solution form [2].

## 6. REFERENCES

- [1] Y. Ephraim and D. Malah, *Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator*, IEEE Trans. on Acoustics Speech, and Signal Processing, Vol. 32, No. 6, pp. 1109-1121, 1984.
- [2] Y. Ephraim and D. Malah, *Speech Enhancement Using a Minimum Mean-Square Error Log-Spectral Amplitude Estimator*, IEEE Trans. on Acoustics, Speech, and Signal Processing, Vol. 33, No. 2, pp. 443-445, 1985.
- [3] R. J. McAulay and M. L. Malpass, *Speech Enhancement Using a Soft-Decision Noise Suppression Filter*, IEEE Trans. on Acoustics, Speech, and Signal Processing, Vol. 28, No.2, pp. 137-145, 1980.
- [4] M. Abramowitz, I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, New York: Dover, 1965.
- [5] I. Cohen, *Optimal speech enhancement under signal presence uncertainty using log-spectral amplitude estimator*, IEEE Signal Processing Letters, Vol. 9, Issue 4, pp. 113-116, 2002.
- [6] R. Martin, *Speech Enhancement Based on Minimum Mean-Square Error Estimation and Supergaussian Priors*, IEEE Trans. Speech and Audio Processing, Vol. 13, Issue 5, pp. 845-856, 2005.
- [7] T. Lotter, P. Vary, *Speech enhancement by map spectral amplitude estimation using a super-Gaussian speech model*, EURASIP Journal on Applied Signal Processing archive Vol. 2005, pp. 1110-1126, 2005.
- [8] Y. Hu and P. Loizou, *Subjective evaluation and comparison of speech enhancement algorithms*, Speech Communication, vol. 49, pp. 588-601, 2007.
- [9] J. S. Erkelens et al., *Minimum Mean-Square Error Estimation of Discrete Fourier Coefficients with Generalized Gamma Priors*, IEEE Trans. Audio, Speech, and Language Processing, Vol. 15, No. 6, pp. 1741-1752, 2007.
- [10] E. Plourde and B. Champagne, *Auditory-Based Spectral Amplitude Estimators for Speech Enhancement*, IEEE Trans. on Audio, Speech, and Language Processing, Vol. 16, No. 8, pp. 1614-1623, 2008.
- [11] C. H. You et al.,  *$\beta$ -order MMSE spectral amplitude estimation for speech enhancement*, IEEE Trans. on Speech and Audio Processing, vol. 13, no. 4, pp. 475-486, 2005.
- [12] R. C. Hendriks et al., *Log-Spectral Magnitude MMSE Estimators under Super-Gaussian Densities*, Interspeech, pp. 1319-1322, 2009.
- [13] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*, Wiley, 1949.
- [14] A. H. Gray, Jr., and J. D. Markel, *Distance Measures for Speech Processing*, IEEE Trans. on Acoustics, Speech, and Signal Processing, Vol. 24, No. 5, pp. 380-391, 1976.
- [15] <http://webee.technion.ac.il/Sites/People/IsraelCohen/>