IDIAP RESEARCH REPORT

SPEECH ENHANCEMENT USING BETA-ORDER MMSE SPECTRAL AMPLITUDE ESTIMATOR WITH LAPLACIAN PRIOR

Hamid Reza Abutalebi Mehdi Rashidinejad
Hervé Bourlard Ali Akbar Tadaion

Idiap-RR-24-2011

JULY 2011

Centre du Parc, Rue Marconi 19, P.O. Box 592, CH - 1920 Martigny
T +41 27 721 77 11 F +41 27 721 77 12 info@idiap.ch www.idiap.ch

Speech Enhancement using β-order MMSE Spectral Amplitude Estimator with Laplacian Prior

Hamid Reza Abutalebi, Mehdi Rashidinejad, Hervé Bourlard, and Ali Akbar Tadaion

Abstract—**This report addresses the problem of speech enhancement employing the Minimum Mean-Square Error (MMSE) of β-order Short Time Spectral Amplitude (STSA). We present an analytical solution for β-order MMSE estimator where Discrete Fourier Transform (DFT) coefficients of (clean) speech are modeled by Laplacian distributions. Using some approximations for the joint probability density function and the Bessel function, we also present a closed-form version of the estimator (called β-order LapMMSE). The performance of the proposed estimator is compared to the state-of-the–art spectral amplitude estimators that assume Gaussian priors for clean DFT coefficients. Comparative results demonstrate the superiority of the proposed estimator in terms of speech enhancement/ noise reduction measures.**

Index Terms— **Laplacian speech modeling; spectral amplitude estimation; speech enhancement.**

I. INTRODUCTION

In increasing number of speech processing applications, noise reduction is becoming an essential pre-processing component to improve system performance. The main objective of speech enhancement is to reduce the corrupting noise component from a noisy speech signal with minimum distortion of clean signal. In the past three decades, Minimum Mean Square Error (MMSE)-based single-channel speech enhancement algorithms have received a lot of attention. In [1], Ephraim and Malah proposed a basic estimator for the Short Time Spectral Amplitude (STSA) of clean speech signal. Their work was then extended to a Log Spectral Amplitude (LSA) estimator in [2]. Considering speech presence uncertainty, Cohen [3] proposed the Optimally Modified-LSA (OM-LSA) estimator. Furthermore, as a good trade-off between noise reduction and minimum speech distortion, You *et al.* [4] proposed the β-order MMSE approach for estimating the STSA of a speech signal. In their work [4], You *et al.* investigated the effectiveness of range of fixed-β values in estimating STSA based on the MMSE criterion, and discussed how the β value could be adopted using the frame signal-to-noise ratio (SNR). Moreover, they showed that their approach could achieve a more significant noise reduction and a better spectral estimation of weak speech spectral component from a noisy signal as compared to many existing speech enhancement algorithms.

However, most of the previous works (including those in [1]-[4]) depend on the fundamental assumption that the real and imaginary parts of Discrete Fourier Transform (DFT) coefficients of clean speech signal are modeled by Gaussian distributions. More recently, researchers have searched for adopting a more appropriate statistical model for the speech signal to improve the speech estimators. In this case, non-Gaussian distributions have been used to model the DFT coefficients of the clean speech signals. Generally, Gamma or Laplacian distribution can be used to model real and imaginary parts of the clean DFT coefficients [5]-[10]. In [11], Chen and Loizou proposed an MMSE estimator (named LapMMSE) of the STSA based on Laplacian model for speech probability density function (pdf). The authors followed that work in [13], where an improved version of LapMMSE (called ImpLapMMSE) was presented.

In the present paper, starting with a formulation similar to that of [11], we derive β-order MMSE estimator when the clean speech DFT coefficients are modeled by a Laplacian distribution. However, the derived analytical solution is highly non-linear, computationally complex and very time-consuming for implementation. Hence, we apply here some approximations for the Bessel function as well as for the pdf of the magnitude spectrum of the clean speech, to reduce the complexity of the estimator. This is shown to result in a closed form of the estimator, namely β-order LapMMSE.

Simulation results demonstrate that the proposed method reduces the corrupting noise component in a better way, which results in less residual noise compared to many existing methods (with either Laplacian or Gaussian assumption).

The rest of the paper is organized as follows. In Section II, we explain our formulation and derivation of the proposed estimator. In Section III, a closed form expression is derived as β-order LapMMSE estimator. In Section IV, we explain our evaluation process and discuss the resulting performance. Finally, Section V concludes the paper.

II. BETA-ORDER LAPMMSE ESTIMATOR

Suppose that $y(n) = x(n) + d(n)$, where $y(n)$, $x(n)$ and $d(n)$ respectively denote noisy signal, clean speech signal, and additive noise. Taking the DFT of $y(n)$, we get:

$$Y_k e^{j\theta_y(k)} = X_k e^{j\theta_x(k)} + D_k e^{j\theta_d(k)}, \tag{1}$$

where $Y_k e^{j\theta_y(k)}$, $X_k e^{j\theta_x(k)}$, and $D_k e^{j\theta_d(k)}$ ($k = 0, 1, 2, ..., N-1$) are the *k*th spectral component of noisy signal, clean speech signal, and additive noise, respectively, and *N* is the frame length. We are looking for $\hat{X}_k$, the estimate of X_k. In [4], You *et al.* considered $J = \mathbf{E}\{(X_k^\beta - \hat{X}_k^\beta)^2\}$ as a cost function that minimizes the mean-square error between the β-order clean speech spectral amplitude and the β-order estimated spectral amplitude. By minimizing the cost function with respect to $\hat{X}_k$, we get [4]:

$$\hat{X}_k = \sqrt[\beta]{\mathbf{E}\left\{X_k^\beta \middle| Y_k\right\}} = \left[\frac{\int_0^\infty \int_0^{2\pi} x_k^\beta f(Y_k|x_k,\theta_k) f(x_k,\theta_k) d\theta_k dx_k}{\int_0^\infty \int_0^{2\pi} f(Y_k|x_k,\theta_k) f(x_k,\theta_k) d\theta_k dx_k}\right]^{\frac{1}{\beta}}, \tag{2}$$

where x_k denotes the sample value of X_k, $\mathbf{E}\{\cdot\}$ is the expectation operator, and $\theta_k = \theta_x(k)$ for convenience; also, $f(x_k,\theta_k)$ is the joint pdf of the magnitude and phase spectra, and $f(Y_k|X_k,\theta_k)$ is given by [1]:

$$f(Y_k|X_k,\theta_k) = \frac{1}{\pi\lambda_d(k)} \exp\left\{-\frac{1}{\lambda_d(k)}|Y_k - X_k|^2\right\}, \tag{3}$$

where $\lambda_d(k)$ denotes the variance of the kth DFT coefficient of the noise.

Following the procedure in [12], it is easy to show that for the Laplacian distribution, the joint pdf $f(x_k,\theta_k)$ is given by [11]:

$$f(x_k,\theta_k) = \frac{x_k}{2\sqrt{\lambda_x(k)}} \exp\left\{-\frac{x_k}{\sqrt{\lambda_x(k)}}\left(|\cos\theta_k| + |\sin\theta_k|\right)\right\}, \tag{4}$$

where $\lambda_x(k)$ is the variance of the kth clean DFT coefficient. Let $\xi_k = \frac{\lambda_x(k)}{\lambda_d(k)}$ and $\gamma_k = \frac{Y_k^2}{\lambda_d(k)}$ respectively denote the *a priori* and *a posteriori* SNR [1]. Substituting (3) and (4) into (2), yields the following form of estimator:

$$\hat{X}_k = \left[\frac{\int_0^\infty x_k^{\beta+1} \exp(-\frac{\gamma_k x_k^2}{Y_k^2}) \int_0^{2\pi} \exp\left[\frac{2x_k\gamma_k\cos\theta_k}{Y_k} - \frac{x_k\sqrt{\gamma_k}}{Y_k\sqrt{\xi_k}}\left(|\cos\theta_k| + |\sin\theta_k|\right)\right] d\theta_k dx_k}{\int_0^\infty x_k \exp(-\frac{\gamma_k x_k^2}{Y_k^2}) \int_0^{2\pi} \exp\left[\frac{2x_k\gamma_k\cos\theta_k}{Y_k} - \frac{x_k\sqrt{\gamma_k}}{Y_k\sqrt{\xi_k}}\left(|\cos\theta_k| + |\sin\theta_k|\right)\right] d\theta_k dx_k}\right]^{\frac{1}{\beta}}. \tag{5}$$

The above equation gives the β-order Laplacian MMSE estimator of the spectral magnitude. We will refer to this estimator as the β-order Laplacian MMSE estimator (or briefly, β-order LapMMSE). To the knowledge of the authors, (5) has no closed form solution. In [11], by applying some approximations, Chen *et al.* derived a closed form solution for the standard ($\beta = 1$) LapMMSE estimator. Using similar formulation, we generalize the estimator and derive the β-order LapMMSE as:

$$\hat{X}_k = \frac{1}{\sqrt{\gamma_k}} \sqrt[\beta]{\frac{A_k + B_k}{C_k + D_k}}\, Y_k \,, \tag{6}$$

where

$$A_k = \frac{1}{2}\sum_{m=0}^{\infty} \frac{\Gamma(m+\frac{1}{2}\beta+1)}{m!\,\Gamma(m+1)} \left(\frac{\gamma_k}{2\xi_k^2 Y_k^2}\right)^m .F(-m,-m;1;2\xi_k^2 Y_k^2)\,, \tag{6-a}$$

$$B_k = \frac{8}{\pi}\sum_{n=1}^{\infty} \frac{1}{n}\sin\left(\frac{\pi n}{4}\right) \frac{\left(\frac{2\gamma_k}{Y_k}\right)^n \left(\frac{\gamma_k}{Y_k^2}\right)^{-\frac{1}{2}n}}{2^{n+1}\Gamma(n+1)} \sum_{m=0}^{\infty} \frac{\Gamma(m+\frac{1}{2}n+\frac{1}{2}\beta+1)}{m!\,\Gamma(m+1)} \left(\frac{\gamma_k}{2\xi_k^2 Y_k^2}\right)^m .F(-m,-m;n+1;2\xi_k^2 Y_k^2)\,, \tag{6-b}$$

$$C_k = \frac{1}{2}\sum_{m=0}^{\infty} \frac{1}{m!} \left(\frac{\gamma_k}{2\xi_k^2 Y_k^2}\right)^m .F(-m,-m;1;2\xi_k^2 Y_k^2)\,, \tag{6-c}$$

$$D_k = \frac{8}{\pi}\sum_{n=1}^{\infty} \frac{1}{n}\sin\left(\frac{n\pi}{4}\right) \frac{\left(\frac{2\gamma_k}{Y_k}\right)^n \left(\frac{\gamma_k}{Y_k^2}\right)^{-\frac{1}{2}n}}{2^{n+1}\Gamma(n+1)} \sum_{m=0}^{\infty} \frac{\Gamma(m+\frac{1}{2}n+1)}{m!\,\Gamma(m+1)} \left(\frac{\gamma_k}{2\xi_k^2 Y_k^2}\right)^m .F(-m,-m;n+1;2\xi_k^2 Y_k^2)\,, \tag{6-d}$$

and $\Gamma(\cdot)$ is the Gamma function and $F(a,b;c;d)$ is the Gaussian hypergeometric function. As shown in (6-a)-(6-d), the derived β-order LapMMSE estimator is highly non-linear and computationally complex. In Section III, we apply some approximations in (2), which will result in a computationally-feasible estimator.

III. Derivation of Closed Form Approximation for Beta-order LapMMSE Estimator

a) Approximation of Joint pdf

Chen *et al.* [11] have already shown that the magnitude and phase of the complex DFT coefficients of the clean speech signals are statistically independent and also derived $f(x_k, \theta_k)$ [11]. Substituting this approximation into (2) results in the β-order LapMMSE estimator of (6). As stated above, (6) is highly complex and nonlinear. To solve this problem, we exploit here some approximations of $f(x_k, \theta_k)$ and the Bessel function.

Using the Tailor series expansion, authors have already shown in [13] that $f\,(x_k, \theta_k)$ can be approximated as:

$$f\,(x_k, \theta_k) \approx \frac{x_k}{\pi \lambda_x} \exp(-\sqrt{\frac{2}{\lambda_x}} x_k), \quad x_k \geq 0\,; \tag{7}$$

Substituting (7) into (2) yields:

$$\hat{X}_k = \left[\frac{\int_0^\infty x_k^{\beta+1} \exp(-\frac{1}{\lambda_d(k)} x_k^2 - \sqrt{\frac{2}{\lambda_x(k)}} x_k)\, I_0(\frac{2x_k Y_k}{\lambda_d(k)})\, dx_k}{\int_0^\infty x_k \exp(-\frac{1}{\lambda_d(k)} x_k^2 - \sqrt{\frac{2}{\lambda_x(k)}} x_k)\, I_0(\frac{2x_k Y_k}{\lambda_d(k)})\, dx_k} \right]^{\frac{1}{\beta}} . \tag{8}$$

Since there is no closed form solution for these integrals (nominator and denominator of (8)), we suggest the following approximations for the Bessel function to reach a closed-form low-complexity estimator.

b) Approximation of the Bessel Function

To reach a closed form solution for (8), we first propose the use of Tailor series expansion of $I_0(.)$ around $x = 0$, i.e. $I_0(x;M) \approx \sum_{m=0}^{M-1} (x/2)^{2m} (1/m!)^2$. Considering this expansion and using [14] and [15, Thm.3.462.1], we get:

$$\hat{X}_k = \frac{1}{\sqrt{2\gamma_k}} \left[\frac{\sum_{m=0}^{M-1} (\frac{\gamma_k}{Y_k})^{2m} \frac{1}{(m!)^2} (\frac{2}{Y_k} \sqrt{\frac{\gamma_k}{\xi_k}})^{-m} \Gamma(2m+\beta+2) D_{-(2m+\beta+2)}(T)}{\sum_{m=0}^{M-1} (\frac{\gamma_k}{Y_k})^{2m} \frac{1}{(m!)^2} (\frac{2}{Y_k} \sqrt{\frac{\gamma_k}{\xi_k}})^{-m} \Gamma(2m+2) D_{-(2m+2)}(T)} \right]^{\frac{1}{\beta}} Y_k\,. \tag{9}$$

where $T = \sqrt{\gamma_k}$ and $D_\gamma(z)$ is parabolic cylinder function. Increasing the number of summation terms (M), (9) presents a good approximation of (8) [13], However, the resulting estimator is still computationally demanding.

As another solution, we consider the well-known approximation of the Bessel function, $I_0(x) \approx (1/\sqrt{2\pi x}) \exp(x)$. Again, using [14] and [15, Thm.3.462.1], this results in:

$$\hat{X}_k = \frac{1}{\sqrt{2\gamma_k}} \left(\frac{\Gamma(\beta+\frac{3}{2}) D_{-(\beta+\frac{3}{2})}(P)}{\Gamma(\frac{3}{2}) D_{-(\frac{3}{2})}(P)} \right)^{\frac{1}{\beta}} . Y_k \,, \tag{10}$$

where $P = \left(\sqrt{\frac{1}{\xi_k}} - \sqrt{2\gamma_k} \right)$.

Equation (10) now presents a closed-from low-complexity version of the proposed estimator (which is referred to as the β-order LapMMSE). Finally, the clean speech component is obtained using the inverse STFT and the weighted overlap-add method.

IV. Implementation and Performance Evaluation

To evaluate the performance of the proposed estimator (β-order LapMMSE), we have compared its performance with those for MMSE-STSA [1], LSA [2], OM-LSA [3], β-order MMSE [4], and our recently proposed one, ImpLapMMSE [13]. Unlike the first four reference methods [1-4] that consider Gaussian pdf for speech, ImpLapMMSE (and β-order LapMMSE) assumes Laplacian priors. In this experiment, we considered $\beta = 0.4$. This was done after many trails on different values of β and evaluating output quality.

For simulation, eight (clean) speech signals (sampled at 16 kHz) were selected from the TIMIT database. We corrupted these signals with white Gaussian noise, covering a wide range of input SegSNRs (-10 dB, -7 dB, -2 dB, -5 dB, 0 dB, 2 dB, 5 dB, and 10 dB). To evaluate the performance of the estimators in speech enhancement task, we have used two basic measures: SegSNR, and PESQ. The reported values are the averages over eight input signals.

The results have been listed in Tables I and II for SegSNR and PESQ, respectively, showing that for both high and low input SNRs, the proposed method yields excellent performance, demonstrating the superiority of Laplacian assumption for speech priors. The comparative results have also been validated through some informal listening tests. These tests show that the β-order LapMMSE produces lower residual noise than state-of-the-art estimators.

Furthermore, we have examined the effect of order parameter (β) in the performance of the proposed estimator. Figure 1 shows the SegNSRs of β-order LapMMSE output versus different values for β (0<β<4). The evaluation has been repeated for a wide range of input SegSNRs (-10 dB, -7 dB, -2 dB, -5 dB, 0 dB, 2 dB, 5dB, and 10 dB). Roughly speaking, the speech enhancement performance is decreased when the considered order of STSA (β) is increased from 0 to 4. Motivated by previous experiences with β-order estimators [4, 16], we are continuing this research to propose an adaptive procedure for calculating optimum value of β in each frame.

V. Summary and Conclusion

In this paper, we focus on speech enhancement using β-order STSA MMSE estimation where the clean speech DFT coefficients are modeled by a Laplacian prior. The resulting analytical solution is highly complex. So, considering some approximations for joint pdf as well as the Bessel function, we derive β-order LapMMSE estimator. Comparing the proposed method with alternative state-of-the-art approaches shows that the proposed estimator is much more effective in reducing the additive noise. Also, similar to what is done in [9], this method has less residual noise, less distortion speech signal and finally better performance in results.

References

[1] Y. Ephraim, and D. Malah, “Speech enhancement using a minimum mean square error short-time spectral amplitude estimator,” *IEEE Trans. Acoust, Speech, Signal Process*., vol. ASSP-32, no. 6, pp. 1109–1121, Dec. 1984.

[2] Y. Ephraim, and D. Malah, “Speech enhancement using a minimum mean square error log-spectral amplitude estimator”, *IEEE Trans. Acoust., Speech, Signal Process*., vol. ASSP-33, no. 2, pp. 443–445, April 1985.

[3] I. Cohen, “On speech enhancement under signal presence uncertainty,” in *Proc. IEEE ICASSP*, Salt Lake City, Utah, USA, May 2001.

[4] C. H. You, S. N. Koh, and S. Rahardja, "β-Order MMSE spectral amplitude estimation for speech enhancement", *IEEE Trans. Speech, Audio Process.*, vol. 13, no. 4, pp. 475-486, Jul. 2005.

[5] J. Porter, S. Boll, "Optimal estimators for spectral restoration of noisy speech," in *Proc. IEEE ICASSP*, San Diego, California, USA, pp. 18A.2.1–18A.2.4, 1984.

[6] R. Martin, "Speech enhancement using MMSE short time spectral estimation with Gamma distributed priors," in *Proc. IEEE ICASSP*, Orlando, Florida, USA, pp. 504–512, 2002.

[7] T. Lotter, and P Vary, "Noise reduction by maximum a posteriori spectral amplitude estimation with Supergaussian speech modeling," in *Proc. IWAENC*, Kyoto, Japan, 2003.

[8] C. Breithaupt, and R. Martin, "MMSE estimation of magnitude squared DFT coefficients with supergaussian priors," in *Proc. IEEE ICASSP*, Hong Kong, vol. I, pp. 896–899, 2003.

[9] B. Chen, and P. Loizou, "Speech enhancement using a MMSE short time spectral amplitude estimator with Laplacian speech modeling," in *Proc. IEEE ICASSP*, Philadelphia, Pennsylvania, USA, vol. 1, pp. 1097–1100, 2005.

[10] R. C. Hendrix, and R. Heusdens, "On linear versus non-linear magnitude-DFT estimators and the influence of the supergaussian priors," in *Proc. IEEE ICASSP*, Dallas, Texas, USA, Mar. 2010.

[11] B. Chen, and C. Loizou, "A Laplacian-based MMSE estimator for speech enhancement," *Speech Communication*, vol. 49, pp. 134–143, 2007.

[12] A. Papoulis, and U. Pillai, *Probability, Random Variables and Stochastic Processes*, Fourth ed., McGraw-Hill, 2001.

[13] M. Rashidinejad, H. R. Abutalebi, and A. A. Tadaion, "Speech Enhancement using an Improved MMSE Estimator with Laplacian Prior," in *Proc. 5th International Symposium on Telecommunication*, Tehran, Iran, Dec. 2010.

[14] M. Abramowitz, and I. A. Stegun, *Handbook of Mathematical Functions*, Dover, New-York, 1964.

[15] I. Gradshteyn, and I. Ryzhik, *Table of Integrals, Series and Products*, New York: Academic, 6th edition, 2000.

[16] H. R. Abutalebi, and B. Dashtbozorg, "Speech dereverberation in noisy environments using an adaptive minimum mean square estimator," Accepted for publication in *IET Signal Processing*, 2011.

TABLE I
COMPARATIVE PERFORMANCE, IN TERMS OF SEGMENTAL SNR, OF THE GAUSSIAN-BASED MMSE-STSA, LSA, OM-LSA, BETA-ORDER MMSE, AND LAPLACIAN-BASED IMPLAPMMSE, AND BETA-ORDER LAPMMSE ESTIMATORS (FOR BETA=0.4).

Estimator	Input SegSNR							
	-10 dB	-7 dB	-5 dB	-2 dB	0 dB	2 dB	5 dB	10 dB
MMSE-STSA	-4.8	-3.7	-3.0	-1.8	-1.1	-0.8	0.5	2.2
LSA	-4.2	-3.0	-2.4	-1.4	-0.6	0.4	1.1	2.9
OM-LSA	-2.6	-2.1	-1.6	-0.7	0.2	1.0	3.2	5.1
β-order MMSE	-4.2	-3.3	-2.6	-1.6	-0.8	-0.5	0.1	1.2
ImpLapMMSE	-3.8	-2.7	-2.0	-0.8	-0.1	0.2	1.7	3.8
β-order LapMMSE	**-2.5**	**-1.8**	**-1.2**	**-0.3**	**0.6**	**1.2**	**3.5**	**5.3**

TABLE II
COMPARATIVE PERFORMANCE, IN TERMS OF PESQ OF THE GAUSSIAN-BASED MMSE-STSA, LSA, OM-LSA, BETA-ORDER MMSE, AND LAPLACIAN-BASED IMPLAPMMSE, AND BETA-ORDER LAPMMSE ESTIMATORS (FOR BETA=0.4).

Estimator	Input SegSNR							
	-10 dB	-7 dB	-5 dB	-2 dB	0 dB	2 dB	5 dB	10 dB
MMSE-STSA	0.8	1.1	1.3	1.6	1.8	2.0	2.2	2.5
LSA	0.9	1.2	1.5	1.8	2.0	2.2	2.4	2.6
OM-LSA	1.1	1.3	1.6	1.8	2.1	2.3	2.5	2.7
β-order MMSE	0.9	1.2	1.3	1.4	1.6	1.7	1.8	2.1
ImpLapMMSE	1.1	1.4	1.6	1.9	2.1	2.5	2.5	2.7
β-order LapMMSE	**1.2**	**1.5**	**1.7**	**2.0**	**2.2**	**2.4**	**2.6**	**2.8**

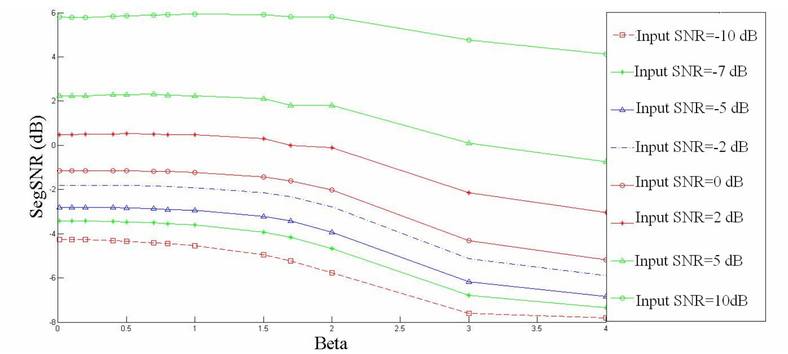

Fig. 1. Output SegSNR values of β–order LapMMSE estimator for 0< β <4 and wide range of input SegSNRs (-10 dB, -7 dB, -5 dB, -2 dB, 0 dB, 2 db, 5 dB, and 0 dB).