IJE TRANSACTIONS C: Aspects Vol. 27, No. 9, (September 2014) 1367-1376

International Journal of Engineering

Journal Homepage: www.ije.ir

Speech Enhancement Using Laplacian Mixture Model under Signal Presence Uncertainty

Z. Mohammadpoory*, J. Haddadnia

Department of BioMedical Engineering Hakim Sabzevari Univesity, Sabzevar, Iran

PAPER INFO

Paper history:
Received 24 August 2013
Received in revised form 27 October 2013
Accepted 22 May 2014

Keywords:
EM Algorithm
Gaussian Noise
Laplacian Mixture Model
Minimum Statistic
MMSE Estimator
Speech Presence Uncertainty

ABSTRACT

In this paper, an estimator for speech enhancement based on Laplacian Mixture Model (LMM)has been proposed. The proposed method, estimates the complex Discerete Fourier Transform(DFT) coefficients of clean speech from noisy speech using the Minimum Mean Square Error(MMSE) estimator, when the clean speech DFT coefficients are supposed mixture of Laplacians and the DFT coefficients of noise are assumed zero-mean Gaussian distribution. Furthermore, the MMSE estimator under speech presence uncertainty and the Laplacian mixture model were derived. It is shown that the proposed estimator has better performance than three estimators based on single Gaussian and single Laplacian models. Also under speech presence uncertainty the results become better.

doi: 10.5829/idosi.ije.2014.27.09c.06

1. INTRODUCTION

In recent years there has been a lot of interest in the enhancement of noisy speech for digital voice communications, human-machine interfaces, automatic speech recognition systems and many other applications. Because, the presence of noise degrades the performance of these systems. A lot of methods have been proposed for speech enhancement, such as the spectral subtraction[1, 2], the signal subspace[3, 4], the statistical method [5-7] and so on, but it has been reported that the statistical methods have better performances compared with other methods [8]. In these methods, the clean speech and noise are modeled by proper distributions and then the clean speech are estimated by an estimator such as Maximum Likelihood (ML)[9], Minimum Mean Square Error (MMSE) [5, 6, 10] or Maximum A Posteriori (MAP)[10-12]. The first statistical speech enhancement method in Discerete Fourier Transform(DFT) domain was based on the complex Gaussian distribution for DFT coefficients of speech and noise [5, 6]. The Gaussian assumption is motivated by the central limit theorem. The Gaussian assumption is valid when the analysis frame size is long and the DFT length is longer than the span of correlation of signal (more than 100ms for the speech signal).Thus this assumption is not proper for the speech DFT coefficients, estimated using relatively short frames in the range of 10-40 ms, which is a typical frame size inspeech applications, whereas, it might be proper for the noise DFT coefficients [13].

For this reason, Gamma, Laplacian and Chi models of the clean speech DFT coefficients were proposed. Martin[10, 14] has proposed complex DFT coefficients estimator with Laplacian and Gamma distributions for the clean speech. Lotter and Vary [11, 12] have proposed a MAP spectral amplitude estimator with Gamma distribution assumption for speech amplitude. They provided histograms of speech DFT coefficients using their own experiment and confirmed that the Laplacian and Gamma densities provide a reasonable fit to the experimental data. Chen and Loizou presented an analytical solution for MMSE estimating of the magnitude spectrum, when the clean speech DFT

*Corresponding Author's Email: Z.Mohammadpoory@yahoo.com (Z. Mohammadpoory)

Please cite this article as:Z. Mohammadpoory, J. Haddadnia, Speech Enhancement using Laplacian Mixture Model under Signal Presence Uncertainty, International Journal of Engineering (IJE), TRANSACTIONS C: Aspects Vol. 27, No. 9, (September 2014) 1367-1376

coefficients are modeled by single laplacian distribution [15].

Trawickiand Johnson propsed Chi statistical models for the speech prior with Gaussian statistical models for the noise likelihood [16].

Some researchers have proposed mixture of distributions such as Gaussian Mixture Model (GMM), Rayleigh Mixture Model (RMM) and so on for speech spectral coefficients or magnitude [17-19].In some papers the distribution of the noise spectrum are also modeled by non-Gaussian distributions [14].

In many researches, statistical method are used in other domain, for example time domain, DiscreteCosine Transform domain, Canonical Transform domain andso on [20, 21].

In this paper, the Laplacian Mixture Model(LMM), forclean speech DFT coefficients has been proposed, due to its more accurate fit to the distributions of complex speech DFT coefficients than single laplacian, Gaussian or even mixture of Gaussian model. Furthermore, analytical derivation of estimator with proposed distributions is relatively simple. In the proposed method, the real and imaginary parts of the noise DFT coefficients are modeled by Gaussian distribution.

This paper is organized as follows: next section is about signal model and assumptions used in this work and section 3 discusses about LMM distribution. Section 4 presents the new MMSE estimator with proposed models for speech and noise. In Section 5 the explanation of Expectation-Maximization (EM) algorithm for estimating the LMM parameters are given, Section 6is about proposed estimator under signal presence uncertainty (SPU) and Section 7 presents experimental results.

2. BASIC ASSUMPTIONS IN THE PROPOSED METHOD

We assume a signal model of the form:

$$y(i) = s(i) + n(i) \quad (1)$$

in which y(i), s(i) and n(i) denote noisy speech, clean speech and noise signal at the sampling time index i, respectively. It is assumed that s(i) and n(i) are statistically independent. These signals are transformed into the frequency domain by applying them short time Discrete Fourier Transform (DFT) which can be written as:

$$Y(k,\mu) = S(k,\mu) + N(k,\mu) \quad (2)$$

where k is the frequency bin index and μ is the frame index.

Another assumption is the decorrelation of spectral components. Since the spectral components can behaveindependently, the MMSE spectral estimator S(k , μ) can be derived from Y(k,μ) only and MMSE derivation are simplified [22].

For simplifying the following results, we will omit our notations both the k and μ, thus Y_R, Y_I, S_R, S_I, N_R and N_I denote real and imaginary parts of noisy speech, clean speech and noise signal, respectively.

3. LAPLACIAN MIXTURE MODEL FOR COMPLEX SPEECH DFT COEFFICIENTS

It was confirmed that the probability PDF of the complex DFT coefficients for short frames in the range of 10-40 ms, is much better modeled by a Laplacian, Gamma or Chi density rather than a Gaussian density [10, 14, 16].

We suggest the Laplacian Mixture Model for the PDF modeling of the real and imaginary parts of the DFT coefficients and we show LMM produces better results than single Laplacian model.The Laplacian density is usually represented by:

$$L(x,c,m) = ce^{-2c|x-m|} \quad (3)$$

where m represents the center (mean) and c > 0 controls the width of the density.The LMM is defined as follows:

$$p(s) = \sum_{i=1}^{N} \alpha_i \, L\left(s,c_i,m_i\right) = \sum_{i=1}^{N} \alpha_i \, c_i e^{-2c_i|s-m_i|} \quad (4)$$

where N is the number of Laplacians andα_i, m_i,c_i are the weights, means and variances of each Laplacian, respectively and $\sum_{i=1}^{N} \alpha_i = 1$. A common method used to train a mixture model is the Expectation-Maximization (EM) algorithm [23].

As it is mentioned, the real and imaginary parts of the speech DFT coefficients are modeled byLMM. Thus, they can be written as follow:

$$p(S_R) = \sum_{i=1}^{N} \alpha_i \, c_i e^{-2c_i|S_R-m_i|} \quad (5)$$

$$p(S_I) = \sum_{j=1}^{N} \alpha_j \, c_j e^{-2c_j|S_I-m_j|} \quad (6)$$

The LMM distribution is selected because the histograms of the real and the imaginary parts of the clean speech DFT coefficients are not exactly zero-mean Laplacian distributed, but we can get more accurate fits to these histograms with combination of several nonzero-mean Laplacian distributions.

This result is confirmed by the estimation of Kullback–Leibler discrimination information for the histogram data $p_H(x)$and assumed densities p(x) such as Laplacian, GMM and LMM. Kullback–Leibler discrimination information is defined as follows:

$$I_{KL} = \sum_x p_H(x) \log\left(\frac{p_H(x)}{p(x)}\right) \quad (7)$$

For comparing two distributions (two different p(x)), having smaller I_{KL} means more accurate fit to the histogram of data $p_H(x)$[24].

We find that the Kullback–Leibler discrimination information is smaller for theLMM distribution than the GMM distribution with the same components. Another result is that by increasing the number of LMM's components(N), I_{KL} becomes smaller and it seems reasonable. After N=30, I_{KL}doesn't have significant variations and it shows that N=30 is a proper value for number of the LMM's components.

4. MMSE SPECTRAL ESTIMATION WITH THE PROPOSED DISTRIBUTION

For finding closed form solution to the estimation problem, it is assumed that the real and the imaginary parts are independent and identically distributed(i.i.d). Because of these assumptions, the MMSE estimation for the complex DFT coefficients (E(S|Y)) can be split into the two estimations for the real and the imaginaryparts which can be written as follows [10]:

$$E(S|Y) = E(S_R|Y_R) + jE(S_I|Y_I) \quad (8)$$

At first,$E(S_R|Y_R)$ and $E(S_I|Y_I)$are estimated, separately and then, their results will be combined together.The optimal MMS estimator of the real part is obtained as follows:

$$E(S_R|Y_R) = \frac{\int_{-\infty}^{+\infty} S_R P(Y_R|S_R)P(S_R)dS_R}{P(Y_R)} \quad (9)$$

By modeling the real and imaginary parts of the noise DFT coefficients by zero-mean Gaussian distributions as follow:

$$p(N_R)=\frac{1}{\sqrt{\pi}\sigma_n}\exp\left(\frac{-(N_R)^2}{\sigma_n^2}\right) \quad (10)$$

$$p(N_I)=\frac{1}{\sqrt{\pi}\sigma_n}\exp\left(\frac{-(N_I)^2}{\sigma_n^2}\right) \quad (11)$$

in which $\frac{\sigma_n^2}{2}$denotes the variance of the real and imaginary parts of the noise DFT coefficients, and assuming the LMM speech prior, (Equations (5) and (6)), the MMSE estimation of real part will be provided as follows:

$$E(S_R|Y_R) = \frac{1}{\sqrt{\pi}\sigma_n P(Y_R)}\int_{-\infty}^{+\infty} S_R \exp\left(\frac{-(Y_R-S_R)^2}{\sigma_n^2}\right) \sum_{i=1}^{N}\alpha_i c_i \exp(-2c_i|s_R-m_i|)\,ds_R \quad (12)$$

After some manipulations, and using a theorem described in [25],the MMSE estimation of the real part will be obtained as follows:

$$E(S_R|Y_R) = \frac{1}{2P(Y_R)}\sum_{i=1}^{N}\alpha_i c_i \exp\left(c_i^2\sigma_n^2\right)\{\sigma_n[L_{R1}(Y_R-m_i)\exp\left(2c_i(Y_R-m_i)\right)\operatorname{erfc}\left(L_{R1}(Y_R-m_i)\right) - L_{R2}(Y_R-m_i)\exp\left(-2C_i(Y_R-m_i)\right)\operatorname{erfc}\left(L_{R2}(Y_R-m_i)\right)]+m_i[\exp\left(2c_i(Y_R-m_i)\right)\operatorname{erfc}\left(L_{R1}(Y_R-m_i)\right) + \exp\left(-2c_i(Y_R-m_i)\right)\operatorname{erfc}\left(L_{R2}(Y_R-m_i)\right)]\} \quad (13)$$

where

$$P(Y_R)=\int_{-\infty}^{+\infty} P(S_R)\,P(Y_R|S_R)dS_R \quad (14)$$

Also by using another theorem in [25], $P(Y_R)$ will be calculated as follows:

$$P(Y_R) = \frac{1}{2}\sum_{i=1}^{N}\alpha_i c_i \exp\left(c_i^2\sigma_n^2\right) [\exp\left(2c_i(Y_R-m_i)\right)\operatorname{erfc}\left(L_{R1}(Y_R-m_i)\right) + \exp\left(-2c_i(Y_R-m_i)\right)\operatorname{erfc}\left(L_{R2}(Y_R-m_i)\right)] \quad (15)$$

In Equations (13) and (15) erfc(x) denotes the complementary error function [25] and is defined as follows:

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}}\int_x^{\infty} e^{-t^2}\,dt = 1-\operatorname{erfc}(x) \quad (16)$$

and

$$L_{R1}(Y_R-m_i) = c\sigma_n + \frac{Y_R-m_i}{\sigma_n} \quad (17)$$

$$L_{R2}(Y_R-m_i) = c\sigma_n - \frac{Y_R-m_i}{\sigma_n} \quad (18)$$

The MMSE estimator for the imaginary part is derived in the same method. Finally, the MMSE estimator for the complex DFT coefficients is calculated by$E(S|Y) = E(S_R|Y_R) + jE(S_I|Y_I)$,where $E(S_R|Y_R)$ and $E(S_I|Y_I)$are given in Equations (19) and (20), respectively.

$$E(S_R|Y_R) = \frac{\sum_{i=1}^{N}\alpha_i c_i \exp\left(c_i^2\sigma_n^2\right)\left[\begin{matrix}\sigma_n\left(L_{R1}(Y_R-m_i)\exp\left(2c_i(Y_R-m_i)\right)\operatorname{erfc}\left(L_{R1}(Y_R-m_i)\right)-L_{R2}(Y_R-m_i)\exp\left(-2C_i(Y_R-m_i)\right)\operatorname{erfc}\left(L_{R2}(Y_R-m_i)\right)\right)+\\ m_i\left(\exp\left(2c_i(Y_R-m_i)\right)\operatorname{erfc}\left(L_{R1}(Y_R-m_i)\right)+\exp\left(-2c_i(Y_R-m_i)\right)\operatorname{erfc}\left(L_{R2}(Y_R-m_i)\right)\right)\end{matrix}\right]}{\sum_{i=1}^{N}\alpha_i c_i \exp\left(c_i^2\sigma_n^2\right)[\exp\left(2c_i(Y_R-m_i)\right)\operatorname{erfc}\left(L_{R1}(Y_R-m_i)\right)+\exp\left(-2c_i(Y_R-m_i)\right)\operatorname{erfc}\left(L_{R2}(Y_R-m_i)\right)]} \quad (19)$$

$$E(S_I|Y_I) = \frac{\sum_{i=1}^{N}\alpha_i c_i \exp\left(c_i^2\sigma_n^2\right)\left[\begin{matrix}\sigma_n\left(L_{R1}(Y_I-m_i)\exp\left(2c_i(Y_I-m_i)\right)\operatorname{erfc}\left(L_{R1}(Y_I-m_i)\right)-L_{R2}(Y_I-m_i)\exp\left(-2c_i(Y_I-m_i)\right)\operatorname{erfc}\left(L_{R2}(Y_I-m_i)\right)\right)+\\ m_i\left(\exp\left(2c_i(Y_I-m_i)\right)\operatorname{erfc}\left(L_{R1}(Y_I-m_i)\right)+\exp\left(-2c_i(Y_I-m_i)\right)\operatorname{erfc}\left(L_{R2}(Y_I-m_i)\right)\right)\end{matrix}\right]}{\sum_{i=1}^{N}\alpha_i c_i \exp\left(c_i^2\sigma_n^2\right)[\exp\left(2c_i(Y_I-m_i)\right)\operatorname{erfc}\left(L_{R1}(Y_I-m_i)\right)+\exp\left(-2c_i(Y_I-m_i)\right)\operatorname{erfc}\left(L_{R2}(Y_I-m_i)\right)]} \quad (20)$$

5. TRAINING USING THE EM ALGORTHM

In this section, the EM algorithm [23] for estimating the parameters of the LMM (used for calculating Equations (19) and (20)) from training data are described, based on the literature[26].

Considering Equation (4) and assuming T samples, EM's cost function for LMM is defind as follows:

$$J(c_i,m_i)=\sum_{n=1}^{T}\sum_{i=1}^{N}(\log c_i - 2c_i|s_n - m_i|)\,.\,p(i|s_n) \quad (21)$$

where $p(i|s_n)$ describs the probability of s_n belonging to the ith Laplacian. $p(i|s_n)$and$\alpha_i$are updated as follow:

$$\alpha_i = \frac{1}{T}\sum_{n=1}^{T} p(i|s_n) \quad (22)$$

$$p(i|s_n) = \frac{\alpha_i c_i e^{-2c_i|s_n-m_i|}}{\sum_{i=1}^{N}\alpha_i c_i e^{-2c_i|s_n-m_i|}} \quad (23)$$

For updating $m_i$and c_i, the Equations $\frac{\partial J}{\partial m_i} = 0$ and $\frac{\partial J}{\partial c_i} = 0$ have to be solved. Then the updates are given as follow:

$$m_i = \frac{\sum_{n=1}^{T}\frac{s_n}{|s_n-\theta_i|}p(i|s_n)}{\sum_{n=1}^{T}\frac{1}{|s_n-\theta_i|}\ p(i|s_n)} \quad (24)$$

$$c_i = \frac{\sum_{n=1}^{T}\ p(i|s_n)}{2\sum_{n=1}^{T}|s_n-\theta_i|\ \ p(i|s_n)} \quad (25)$$

For initializing the EM algorithm, at first this algorithm is ran with random initial values and then the results of this EM are considered as initial values for the main EM.

6.DERIVATION OF MMSE ESTIMATOR UNDER SPEECH PRESENCE UNCERTANTY

In order to further performance improvement of the estimator, we alsoderive a Speech Presence Uncertainty (SPU) estimator and incorporate it into the spectral estimator. SPU addresses the speech/silence detection problem in terms of probability, and is derived using the Bayes' rule. After the spectral coefficients are estimated by the method proposed above, the SPU estimator refines the estimate of the spectral coefficients by scaling them by the SPU probability.

We consider a two-state model for each frequency bin of the speech, 1- Speech is present at a particular frequency bin (k)(hypothesis H_1^k) 2- Speech is not present (hypothesis H_0^k).This is expressed mathematically using the following binary hypothesis model:

$$H_0^k : \text{speech absence: } Y_k = N_k \quad (26)$$

$$H_1^k : \text{speech present: } Y_k = S_k + N_k \quad (27)$$

To incorporate the above binary model to an MMSE estimator, we can use a weighted average of two estimators: one that is weighted by the probability that speech is present, and one that is weighted by the probability that speech is absent. So, if the original MMSE estimator had the form $E(S_k|Y_k)$, then the new estimator has theform

$$\hat{S_k} = E(S_k|Y_k,H_1^k)P(H_1^k|Y_k) + E(S_k|Y_k,H_0^k)P(H_0^k|Y_k) \quad (28)$$

where $P(H_1^k|Y_k)$ denotes the conditional probability that speech is present in frequency bin k given the noisy speech spectrum. Similarly, $P(H_0^k|Y_k)$ denotes the conditional probability that speech is absent given the noisy speech spectrum. The term $E(S_k|Y_k,H_0^k)$ in the above equation is zero since it represents the average value of S_k given the noisy spectrum Y_k and the fact that speech is absent. Therefore, the MMSE estimator in Equation (28) reduces to:

$$\hat{S_k} = E(S_k|Y_k,H_1^k)P(H_1^k|Y_k) \quad (29)$$

The MMSE estimator of the spectral component at frequency bin k is weighted by the probability that speech is present at that frequency. Bayes' rule can be used to compute $P(H_1^k|Y_k)$:

$$P(H_1^k|Y_k)=\frac{p(Y_k|H_1^k)p(H_1^k)}{p(Y_k|H_1^k)p(H_1^k)+p(Y_k|H_0^k)p(H_0^k)} \quad (30)$$

where $p(H_0^k)$ denotes the a priori probability of speech absence and $p(H_1^k)$is the a priori probability of speech presence, for frequency bin k. It is clear that $p(H_1^k) = 1 - p(H_0^k)$.

Under hypothesis H_0^k, $Y_k = N_k$and given that the noise is complex Gaussian with zero mean and variance σ_n^2, it follows that $p(Y_k|H_0^k)$ will also have a Gaussian distribution with the same variance.

$$p(Y_k|H_0^k) = \frac{1}{\sqrt{\pi}\sigma_n}\exp\left(\frac{-(Y_k)^2}{2\sigma_n^2}\right) \quad (31)$$

If S_k follows a Mixture of Laplacian distribution, we need to compute $p(Y_k|H_1^k)$. Assuming independence between real and imaginary components, we have:

$$p(Y_k|H_1^k) = P_{Y_k}(y_k) = P_{Y_r}(y_r)P_{Y_i}(y_i) \quad (32)$$

where $y_r = \text{Re}\{Y_k\}$, and $y_i = \text{Im}\{Y_k\}$.Under hypothesis H_1^k, we need to derive the PDF of $Y_k = S_k + N_k$, where $S_k = S_r + jS_i$ and $N_k = N_r + jN_i$. The PDFs of S_r and S_i are assumed to be Mixture of Laplacian and the PDFs of N_r and N_i are assumed to be Gaussian with variance $\sigma_n^2/2$ and zero mean. The derivation of Equation (32) is given in Appendix A. The solution for Equation (32) is given by:

$$P_{Y_r}(y_r) = \sum_{i=1}^{N}\alpha_i c_i \exp\left(\sigma_n^2(c_i^2 + 2m_i c_i)\right) \left[\begin{matrix}\exp(-c_i y_r) + \exp(c_i y_r) + \exp(-c_i y_r) \\ \text{erf}(c_i(y_r - \sigma_n)) + \exp(c_i y_r)\text{erf}(c_i(y_r + \sigma_n))\end{matrix}\right] \quad (33)$$

$$P_{Y_i}(y_i) = \sum_{i=1}^{N}\alpha_i c_i \exp(\sigma_n^2(c_i^2 + \quad (34)$$

$$2m_i c_i)\Big)\begin{bmatrix}\exp(-c_i y_i)+\exp(c_i y_i)+\exp(-c_i y_i)\\ \operatorname{erf}\big(c_i(y_i-\sigma_n)\big)+\exp(c_i y_i)\operatorname{erf}\big(c_i(y_i+\sigma_n)\big)\end{bmatrix}$$

Simply substituting Equations (33) and (34) into Equation (32), and Equations (32) and (31) into Equations (30) and (29),we obtain the Speech Presence Uncertainty (SPU) estimator [27].

7. EXPRIMENTS AND RESULTS

7. 1. Experimental Setup The TIMIT database has been used in our experiments. We used 200 sentences from 100 male and 100 female (about 11 minutes) for training the LMM models with EM algorithm and 10 sentences from 5 males and 5 females for evaluating our proposed method.

White noise, babble noise, and F-16 cockpit noiseare added at 0, 5 and 10 dB SNR to 10 mentioned sentences. Then the proposed method is applied to noisy sentences. To determine the variance of these noises (used for calculating Equations (19) and (20)), a Voice Activity Detection (VAD) method is employed to noisy sentences. Indeed, the clean speech LMM model parameters are found off-line using training data and EM algorithm and noise variance are estimated on-line using noisy test data VAD method.VAD refers to the ability of distinguishing speech from noise, and estimating the parameter of noise from noise frames[28].

The proposed estimator are applied to 32ms frams, with 50% overlap, which are multiplied by Hamming window. For obtaining the enhanced speech signal, these frames are transform to the time domain by IDFT and the Overlap-Add method is applied to them.

7. 2. Performance Evaluations Three objective measures, segmental SNR, log-likelihood ratio (LLR) and PESQ (Perceptual Evaluation of Speech Quality),were applied for performance evaluation of proposed method. The segmental SNR is computed as follows:

$$SNR_{seg} = \frac{10}{M}\sum_{f=1}^{M} log_{10}\left[\frac{\sum_{n=L.F}^{L.f+L-1} s^2(n)}{\sum_{n=L.F}^{L.f+L-1}[s(n)-\hat{s}(n)]^2}\right] \quad (35)$$

where M is the total number of frames, L is the frame size, s(n)is the clean signal and ŝ(n) is the enhanced signal. As SNR_{seg} does not have strong correlation with subjective evaluation methods, we use LLR and PESQ measures which have stronger correlations with subjective evaluation methods.

The LLR measure is one of the most common all-pole-based measure for evaluating speech enhancement algorithms. The log-likelihood ratio (LLR) for each frame is computed as follows:

$$LLR=\log\left(\frac{a_e^T R_x a_e}{a_x^T R_x a_x}\right) \quad (36)$$

wherea_x and a_e are the prediction coefficients of the clean and enhanced signals, respectively, and R_x is the autocorrelation matrix of the clean signal. For LLR values, being lower shows that the enhanced signal is more similar to the clean signal [29-31]. The PESQ is an objective measurement tool thatpredicts the results of subjective listening tests. PESQ uses a sensory model to compare the original, unprocessed signal with the enhanced signal. The PESQ scores are calibrated using a large database of subjective tests and it is between -0.5-4.5.The higherPESQ means the higher quality of enhanced signal [30]. Figure1 shows the output SNR_{seg}(in dB) measure of the enhanced speech, for different values of N(number of Laplacian components), under different input SNR conditions and white noise.

It is important to find a reasonable N that the proposed method provides an acceptable performance andcomplexity. It is clear that larger N has better performance.Beacause higher number of Laplacian components leads to better PDF matching of the clean signal. But larger N causes increasing the computational complexity. As it was mentioned before, after N=30, $I_{KL}$and subsequently the assumed LMM distributions, do not have significant changes. Thus the results based on assumed distributions, do not have considerable variations. For this reasons, N=30 is selected as proper value.As Figure 1 shows, N has been increased until N=50, and after N=30 the results do not havesignificant variations, while computational complexity is increasing and proposed method becomes very time consuming.

For comparative purposes, the performance of the Gaussian-based MMSE estimator [5], Log-MMSE estimator [6] and Laplacian based MMSE spectral estimator [10] are evaluated. They are indicated as MMSE, Log-MMSE and Lap-MMSE, respectively. The proposed method is also indicated as LMM-MMSE andLMM-MMSE-SPU estimator.

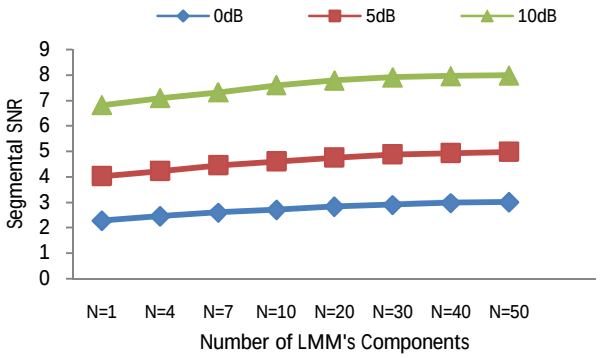

Figure 1. SegSNR of the enhanced speech signalat different Ns and different input SNRs for additive white noise

Table 1 and 2 show the segmental SNR and LLR values obtained by mentioned estimators for babble and F-16 cockpite noise at different SNRs, and Table 3 shows the PESQ values. As can be seen, the SNRseg values (Table 1) obtained withthe Laplacian estimators (LMM-MMSE, Lap-MMSE) are significantly higher than the SNRseg values obtained with the Gaussian estimators (MMSE and Log-MMSE), in all SNRconditions and two noises. It confirms that Laplacian estimators have better performance in reducing the additive noise than Gaussian estimtors.Table 1 shows that for the babble noise at different SNRs, theproposed methods provide more than 1dB improvement in SNRseg compared to Gaussian-based estimators. As Table 2 shows, the LLR values obtained with theLaplacian-based estimators (LMM-MMSE, Lap-MMSE)were also significantly smaller than the values obtained by the Gaussian estimators (MMSEand Log-MMSE) in all SNR conditionsand noises.

Higher PESQ values (Table 3) are obtained by theLaplacian estimators compared to the Gaussian-based estimators for all noises and SNR conditions. It means that the enhanced signal with Laplacian estimators has higher quality than Gaussian estimators. Thus, better results (higher segmental SNR, higher PESQ, and lowerLLR at different SNRs and differentnoises) were obtained by the Laplacian-based estimators.

Comparing the results of two Laplacian-based methods, Lap-MMSE (based on single Laplacian) and LMM-MMSE (based on Mixture of thirty laplacians),shows that better results in term of allmeasures, at different SNRs andnoises, areobtained by the LMM-MMSE estimator. Beacause the use of higher number of Laplacian components causes better PDF matching of the clean signal and thusbetter results. Also comparing the results of two LMM-basedmethods, LMM-MMSE and LMM-MMSE-SPU, shows that better results in term of all objective measures, at different SNRs and different noises, areobtained by the LMM-MMSE-SPU estimator. It showsproposed estimator under SPU has better performance, while the complexity increases.

TABLE 1. Comparative performance, in terms of segmental SNR of the Gaussian-based MMSE, Gaussian-based LogMMSE , Laplacian based MMSE spectral, LMM-based MMSE and LMM-based MMSE with SPU estimators

Noises	**Babble noise**			**F-16**	**Cockpitnoise**	
Estimators	-9.48/0 dB	5.48/5 dB-	-1.4/10 dB	9.1/0 dB-	6.1/5 dB-	-2.1/10 dB
MMSE [5]	1.341	3.892	6.42	1.57	4.132	6.731
Log-MMSE [6]	1.773	4.251	7.123	1.873	4.052	6.69
Lap-MMSE [11]	2.874	4.808	7.582	2.81	4.613	7.288
LMM-MMSE	2.912	4.915	7.916	2.849	4.671	7.623
LMM-MMSE-SPU	3.043	4.992	7.98	2.976	4.703	7.692

TABLE 2. Comparative performance, in terms of LLR of the Gaussian-based MMSE, Gaussian-based LogMMSE , Laplacian based MMSE spectral, LMM-based MMSE and LMM-based MMSE with SPU estimators

Noises	**Babble noise**			**F-16**	**Cockpitnoise**	
Estimators	0 dB	5 dB	10 dB	0 dB	5 dB	10 dB
MMSE [5]	0.981	0.751	0.572	1.021	0.812	0.619
Log-MMSE [6]	1.213	0.984	0.852	1.196	0.935	0.734
Lap-MMSE [11]	0.81	0.644	0.507	0.879	0.696	0.57
LMM-MMSE	0.793	0.612	0.503	0.861	0.663	0.532
LMM-MMSE-SPU	0.734	0.594	0.487	0.806	0.637	0.515

TABLE 3. Comparative performance, in terms of PESQ of the Gaussian-based MMSE, Gaussian-based LogMMSE, Laplacian based MMSE spectral, LMM-based MMSE and LMM-based MMSE with SPU estimators

Noises	**Babble noise**			**F-16**	**Cockpit noise**	
Estimators	0 dB	5 dB	10 dB	0 dB	5 dB	10 dB
MMSE [5]	2.114	2.456	2.781	2.081	2.396	2.712
Log-MMSE [6]	2.172	2.487	2.833	2.035	2.19	2.585
Lap-MMSE [11]	2.183	2.52	2.857	2.136	2.314	2.508
LMM-MMSE	2.225	2.576	2.89	2.17	2.51	2.841
LMM-MMSE-SPU	2.481	2.612	2.911	2.198	2.524	2.849

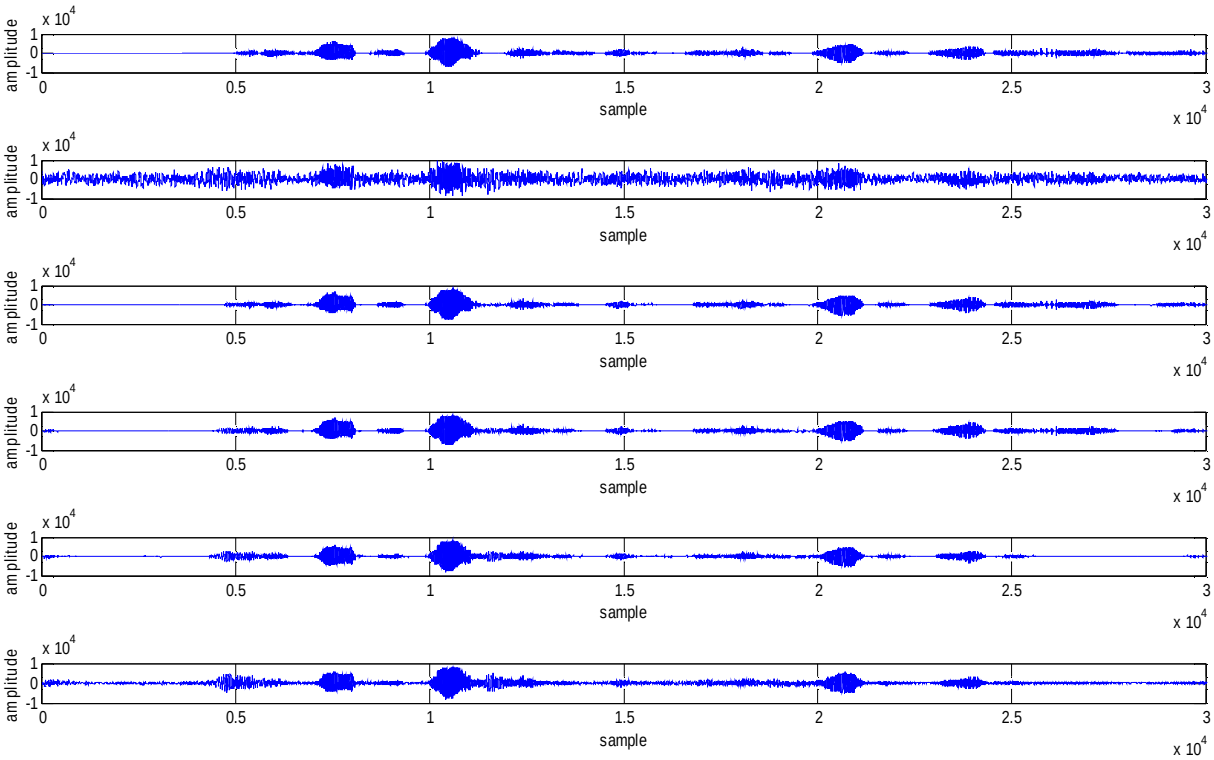

Figure2. A TIMIT sentence corporates with babble noise at SNR=-5dB enhanced by MMSE, Lap-MMSE, LMM- MMSE LMM-MMSE-SPU estimators. From top to bottom, the clean signal, the noisy signal,the signal enhanced by the LMM- MMSE-SPU, the signal enhanced by the LMM- MMSE, the signal enhanced by the Lap-MMSE and the signal enhanced by the MMSE estimator.

Figure 2 shows a TIMIT sentence enhanced by the LMM-MMSE-SPU, LMM-MMSE, Lap-MMSE and MMSE methods. The Babble noise is added to this sentence at -5 dB SNR. It is clear that the residual noise in the sentence enhanced by the proposed methods is less than the others. Also proposed methods, especially LMM-MMSE-SPU, remove main signal instead of noise, less than the others. Therefore this method does not create significant perceptible distortion in the speech signal comparing the others.

As the results of experiment, it is credible that the LMM based MMSE estimator can be considered as an effective method for speech enhancement.

8. CONCLUSION

An MMSE estimator was derived for the speech spectral estimation from noisy signal based on the LMM for the speech DFT coefficient and the Gaussian model for the noise DFT coefficients.Results, in term of the objective measures, indicate that the better performance are obtained with the increment ofN, but after N=30 there is not significant difference between results. Also, the proposed LMM-based MMSE estimator, provides better performance than Gaussian-basedMMSE, Log MMSE and Laplacian-based MMSE spectral estimators. Also under speech presence uncertainty the results become better. The improvement in performance shows that the PDF of clean speech DFT coefficients in MMSE clean speech estimation is better modeled using the LMM.

9. REFERENCES

1. Boll, S., "Suppression of acoustic noise in speech using spectral subtraction", ***Acoustics, Speech and Signal Processing, IEEE Transactions on***, Vol. 27, No. 2, (1979), 113-120.
2. Paliwal, K., Wójcicki, K. and Schwerin, B., "Single-channel speech enhancement using spectral subtraction in the short-time modulation domain", ***Speech Communication***, Vol. 52, No. 5, (2010), 450-475.
3. Ephraim, Y. and Van Trees, H.L., "A signal subspace approach for speech enhancement", ***Speech and Audio Processing, IEEE Transactions on***, Vol. 3, No. 4, (1995), 251-266.
4. Borowicz, A. and Petrovsky, A., "Signal subspace approach for psychoacoustically motivated speech enhancement", ***Speech Communication***, Vol. 53, No. 2, (2011), 210-219.
5. Ephraim, Y. and Malah, D., "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator", ***Acoustics, Speech and Signal Processing, IEEE Transactions on***, Vol. 32, No. 6, (1984), 1109-1121.
6. Ephraim, Y. and Malah, D., "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator",

Acoustics, Speech and Signal Processing, IEEE Transactions on, Vol. 33, No. 2, (1985), 443-445.

7. Ephraim, Y., Malah, D. and Juang, B.-H., "On the application of hidden markov models for enhancing noisy speech", ***Acoustics, Speech and Signal Processing, IEEE Transactions on***, Vol. 37, No. 12, (1989), 1846-1856.
8. Loizou, P.C., "Speech enhancement: Theory and practice, CRC press, (2013).
9. McAulay, R. and Malpass, M., "Speech enhancement using a soft-decision noise suppression filter", ***Acoustics, Speech and Signal Processing, IEEE Transactions on***, Vol. 28, No. 2, (1980), 137-145.
10. Martin, R., "Speech enhancement based on minimum mean-square error estimation and super-gaussian priors", ***IEEE Trans. Speech Audio Proc.***, Vol. 13, No. 5, (2005), 845–856.
11. Lotter, T. and Vary, P., "Speech enhancement by map spectral amplitude estimation using a super-gaussian speech model", ***EURASIP Journal on Applied Signal Processing***, (2005), 1110-1126.
12. Wolfe, P.J. and Godsill, S.J., "Efficient alternatives to the ephraim and malah suppression rule for audio signal enhancement", ***EURASIP Journal on Advances in Signal Processing***, Vol. 2003, No. 10, (1900), 1043-1051.
13. Papoulis, A. and Pillai, S.U., "Probability, random variables, and stochastic processes, Tata McGraw-Hill Education, (2002).
14. Martin, R. and Breithaupt, C., "Speech enhancement in the dft domain using laplacian speech priors", in Proc. of IWAENC, (2003).
15. Chen, B. and Loizou, P.C., "A laplacian-based mmse estimator for speech enhancement", ***Speech Communication***, Vol. 49, No. 2, (2007), 134-143.
16. Trawicki, M.B. and Johnson, M.T., "Speech enhancement using bayesian estimators of the perceptually-motivated short-time spectral amplitude (stsa) with chi speech priors", ***Speech Communication***, Vol. 57, No., (2014), 101-113.
17. Huang, Q., Yang, J. and Zhou, Y., "Variational bayesian method for speech enhancement", ***Neurocomputing***, Vol. 70, No. 16, (2007), 3063-3067.
18. Erkelens, J., Jensen, J. and Heusdens, R., "Speech enhancement based on rayleigh mixture modeling of speech spectral amplitude distributions", in Proc. EUSIPCO. (2007), 65-69.
19. Kundu, A., Chatterjee, S., Sreenivasa Murthy, A. and Sreenivas, T., "Gmm based bayesian approach to speech enhancement in signal/transform domain", in Acoustics, Speech and Signal Processing, ICASSP. IEEE International Conference on, (2008), 4893-4896.
20. Qiu, W., Li, B.-Z. and Li, X.-W., "Speech recovery based on the linear canonical transform", ***Speech Communication***, Vol. 55, No. 1, (2013), 40-50.
21. Laska, B., Bolić, M. and Goubran, R., "Discrete cosine transform particle filter speech enhancement", ***Speech Communication***, Vol. 52, No. 9, (2010), 762-775.
22. Martin, R., Statistical methods for the enhancement of noisy speech, in Speech enhancement., Springer, (2005). 43-65.
23. Bilmes, J.A., "A gentle tutorial of the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models", ***International Computer Science Institute***, Vol. 4, No. 510, (1998), 126.
24. Kullback, S., "Information theory and statistics, Courier Dover Publications, (1997).
25. Jeffrey, A. and Zwillinger, D., "Table of integrals, series, and products, Academic Press, (2007).
26. Mitianoudis, N. and Stathaki, T., "Overcomplete source separation using laplacian mixture models", ***IEEE Signal Processing Letters***, Vol. 12, No. 4, (2005), 277-280.
27. Stark, A. and Paliwal, K., "Use of speech presence uncertainty with mmse spectral energy estimation for robust automatic speech recognition", ***Speech Communication***, Vol. 53, No. 1, (2011), 51-61.
28. Eshaghi, M. and Karami Mollaei, M., "A new algorithm for voice activity detection based on wavelet packets", in Electrical Engineering, ICEE Second International Conference on, (2008), 1-4.
29. Davari, P. and Hassanpour, H., "A robust feedforward active noise control system with a variable step-size fxlms algorithm: Designing a new online secondary path modelling method", ***International Journal of Engineering-Transactions A: Basics***, Vol. 21, No. 3, (2008), 231.
30. Hu, Y. and Loizou, P.C., "Evaluation of objective measures for speech enhancement", in Interspeech, Citeseer. (2006).
31. Hansen, J.H. and Pellom, B.L., "An effective quality evaluation protocol for speech enhancement algorithms", in ICSLP, Citeseer. Vol. 7, (1998), 2819-2822.

APPENDIX

In this appendix, we derive the PDF of $Y_k = S_k + N_k$, where $S_k = S_r + jS_i$ and $N_k = N_r + jN_i$. The PDFs of S_r and S_i are assumed to be Mixture of Laplacian and the PDFs of N_r and N_i are assumed to be Gaussian with variance $\sigma_n^2/2$ and zero mean. Let $Y = Y_r + jY_i$, then $Y_r = S_r + N_r$ and $Y_i = S_i + N_i$. The PDF of Yr can be computed by the convolution of the Mixture of Laplacian and Gaussian densities, and is given by:

$$P_{Y_r}(y_r)=\int_{-\infty}^{\infty} P_{S_r}(y_r - n_r)\, P_{N_r}(n_r)\,dn_r = \sum_{i=1}^{N}\Big(\int_{-\infty}^{y_r} \frac{2\alpha_i c_i}{\sqrt{\pi}\sigma_n} \exp\Big(-\frac{n_r^2 - c_i n_r \sigma_n^2 + 2\sigma_n^2 y_r c_i - 2\sigma_n^2 m_i c_i}{\sigma_n^2}\Big)\, dn_r + \int_{y_r}^{\infty} \frac{2\alpha_i c_i}{\sqrt{\pi}\sigma_n} \exp\Big(-\frac{n_r^2 + c_i n_r \sigma_n^2 - 2\sigma_n^2 y_r c_i + 2\sigma_n^2 m_i c_i}{\sigma_n^2}\Big)\, dn_r\Big) \quad (37)$$

After using a theorem based on the literature [26], we get:

$$P_{Y_r}(y_r) = \sum_{i=1}^{N} \alpha_i c_i \exp\big(\sigma_n^2(c_i^2 + 2m_i c_i)\big)\big[\exp(-c_i y_r) + \exp(c_i y_r) + \exp(-c_i y_r)\ \mathrm{erf}\big(c_i(y_r - \sigma_n)\big) + \exp(c_i y_r)\mathrm{erf}\big(c_i(y_r + \sigma_n)\big)\big] \quad (38)$$

The probability density for the imaginar y part, has exactly the same form as that of $P_{Y_r}(y_r)$. Assuming independence between y_r and y_i we get the following expression for the conditional density $p\big(Y_k \big| H_1^k\big)$ at frequency bin k:

$$p\big(Y_k \big| H_1^k\big) = P_{Y_r}(y_r) P_{Y_i}(y_i) \quad (39)$$

Speech Enhancement Using Laplacian Mixture Model under Signal Presence Uncertainty

Z.Mohammadpoory, J.Haddadnia

Department of BioMedical Engineering Hakim Sabzevari Univesity, Sabzevar, Iran

PAPER INFO

Paper history:
Received 24 August 2013
Received in revised form 27 October 2013
Accepted 22 May 2014

Keywords:
EM Algorithm
Gaussian Noise
Laplacian Mixture Model
Minimum Statistic
Mmse Estimator
Speech Presence Uncertainty

چکیده

در این مقاله یک روش بهسازی گفتار آماری با فرض توزیع مخلوط لاپلاس برای گفتار، برای تخمین سیگنال گفتار تمیز (بدون نویز) از سیگنال گفتار نویزی ارائه شده است. در روش پیشنهادی، ضرایب تبدیل فوریه زمان کوتاه گسسته سیگنال گفتار با استفاده از تخمین گر کمترین میانگین مربعات خطا، بدست می آید. در این تخمین، فرض می شود که تابع چگالی احتمال ضرایب تبدیل فوریه سیگنال تمیز و نویز به ترتیب، مخلوط لاپلاس و گوسی با میانگین صفر می باشد. همچنین برا بهبود نتایج تخمین طیف با الحاق عدم قطعیت گفتار محاسبه شده است. نتایج حاصل از معیارهای SNRقطعه ای، LLRو PESQ نشان می دهد که روش پیشنهادی عملکرد بهتری نسبت به دو روش مبتنی بر توزیع گوسی و روش مبتنی بر توزیع لاپلاس دارد و با الحاق عدم قطعیت گفتار به تخمین گر، نتایج بهتر می شوند.

doi: 10.5829/idosi.ije.2014.27.09c.06