# GMM BASED BAYESIAN APPROACH TO SPEECH ENHANCEMENT IN SIGNAL / TRANSFORM DOMAIN

Achintya Kundu, Saikat Chatterjee, A. Sreenivasa Murthy and T.V. Sreenivas

Department of Electrical Communication Engineering Indian Institute of Science, Bangalore, India 560012 Email: {achintya, saikat, asmuvce, tvsree}@ece.iisc.ernet.in

# ABSTRACT

Considering a general linear model of signal degradation, by modeling the probability density function (PDF) of the clean signal using a Gaussian mixture model (GMM) and additive noise by a Gaussian PDF, we derive the minimum mean square error (MMSE) estimator. The derived MMSE estimator is non-linear and the linear MMSE estimator is shown to be a special case. For speech signal corrupted by independent additive noise, by modeling the joint PDF of time-domain speech samples of a speech frame using a GMM, we propose a speech enhancement method based on the derived MMSE estimator. We also show that the same estimator can be used for transform-domain speech enhancement.

Index Terms- MMSE estimation, GMM, Gaussian noise.

#### 1. INTRODUCTION

To enhance the quality and intelligibility of noisy speech, research in speech enhancement (SE) [1] has focused on better modeling of the speech and noise PDFs, the way the noise contaminates the clean speech, the type of noise source, etc. The most common distortion in speech is due to additive noise, which is independent of the clean speech. The SE algorithms explored can be grouped into two major classes [2]: 1) the class based on hidden Markov model (HMM) ([3], [4]), and 2) the class based on transformation of signals, such as MMSE estimation ([5], [6]), spectral subtraction [1] and subspace based methods ([2], [7], [8]). Ephraim and Van Trees introduced subspace based approach[7], where they assumed the additive noise to be white. This subspace based method of Ephraim and Van Trees [7] is further improved by Rezayee and Gazor [2], and Hu and Loizou [8] to include the case of colored noise also. The generalized subspace based method of Hu and Loizou [8] includes the methods of Rezayee and Gazor [2] and Ephraim and Van Trees [7] as special cases.

In the class of MMSE estimation based processing, many SE algorithms assume that the coefficients of both the clean speech and noise are jointly Gaussian distributed. Such an assumption results in a linear estimator and thus, leads to suboptimal performance where the Gaussian distribution is not the best model, atleast for the PDF of speech signal. It has been shown that the PDFs of speech signal and transform domain signal are better modeled using other distributions, such as Gamma [6] and Laplacian [9]. In [10], the PDF of transform-domain coefficients of speech signal is modeled using Laplacian density and noise component is modeled using a Gaussian density. To achieve better recognition performance, the feature vectors derived from the noisy speech [11] are enhanced by modeling the PDF of clean speech feature vector using a GMM and the noise is modeled using impulse function.

In this paper, we explore the use of non-Gaussian model of clean speech PDF for speech enhancement. We derive the general linear model based MMSE estimator where the PDF of the clean speech signal is modeled using a GMM. The additive noise is assumed to be Gaussian distributed and independent of the clean speech. We show that the developed estimator performs better than the generalized subspace based method of Hu and Loizou [8].

### 2. GENERAL LINEAR MODEL AND MMSE ESTIMATOR

In the general linear model, the observation data vector  $\mathbf{X}$  is modeled as

$$\mathbf{X} = \mathbf{A}\,\mathbf{S} + \mathbf{W},\tag{1}$$

where **X** is a  $q \times 1$  random observation vector, **A** is a known  $q \times p$  observation matrix, **S** is a  $p \times 1$  random vector of parameters to be estimated, and **W** is a  $q \times 1$  random vector of additive noise. **W** and **S** are assumed to be statistically independent of each other. The PDFs of **S** and **W** are assumed to be known and respectively denoted by  $f_{\mathbf{S}}(\mathbf{s})$  and  $f_{\mathbf{W}}(\mathbf{w})$ .

**Theorem:** In the general linear model, if S and W are respectively GMM (with M number of mixture components) and Gaussian distributed as

$$f_{\mathbf{S}}(\mathbf{s}) = \sum_{m=1}^{M} \alpha_m \, \mathbf{N}\left(\mathbf{s}; \boldsymbol{\mu}_{\mathbf{S}}^m, \boldsymbol{C}_{\mathbf{SS}}^m\right),\tag{2}$$

$$f_{\mathbf{W}}(\mathbf{w}) = \mathbf{N}\left(\mathbf{w}; \mu_{\mathbf{w}}, C_{\mathbf{w}}\right), \qquad (3)$$

where  $\alpha_m$  ( { $\alpha_m > 0$ } $_{m=1}^M$  and  $\sum_{m=1}^M \alpha_m = 1$  ),  $\mu_s^m$  and  $C_{ss}^m$  are respectively the prior probability, mean vector and covariance matrix of the *m*th Gaussian component of GMM, and,  $\mu_W$  and  $C_W$  are respectively the mean vector and covariance matrix of **W**, then the MMSE estimator,  $\hat{\mathbf{S}} = \mathcal{E} \{ \mathbf{S} | \mathbf{X} \}$ , is given in Eqn. (4).

*Proof:* As S and W are independent, we can write the joint PDF of S and W as

$$f_{\mathbf{S},\mathbf{W}}(\mathbf{s},\mathbf{w}) = \left[\sum_{m=1}^{M} \alpha_m \mathbf{N}\left(\mathbf{s}; \mu_{\mathbf{s}}^m, C_{\mathbf{S}\mathbf{s}}^m\right)\right] \times \mathbf{N}\left(\mathbf{w}; \mu_{\mathbf{w}}, C_{\mathbf{w}}\right) \qquad (6)$$
$$= \sum_{m=1}^{M} \alpha_m \mathbf{N}\left(\mathbf{s}, \mathbf{w}; \mu^m, C^m\right).$$

Thus, the joint PDF of  $\mathbf{S}$  and  $\mathbf{W}$  is GMM distributed as shown in Eqn. (6), where

$$\mu^{m} = \begin{bmatrix} \mu_{\mathbf{S}}^{m} \\ \mu_{\mathbf{W}} \end{bmatrix}, \quad C^{m} = \begin{bmatrix} C_{\mathbf{SS}}^{m} & \mathbf{0} \\ \mathbf{0} & C_{\mathbf{W}} \end{bmatrix}. \tag{7}$$

$$\widehat{\mathbf{S}} = \sum_{m=1}^{M} \beta_m(\mathbf{X}) \left( \mu_{\mathbf{S}}^m + C_{\mathbf{SS}}^m \mathbf{A}^T \left[ \mathbf{A} C_{\mathbf{SS}}^m \mathbf{A}^T + C_{\mathbf{W}} \right]^{-1} \left( \mathbf{X} - \left[ \mathbf{A} \mu_{\mathbf{S}}^m + \mu_{\mathbf{W}} \right] \right) \right), \tag{4}$$

where  $\beta_m(\mathbf{X})$  is defined as

$$\beta_{m}(\mathbf{X}) = \frac{\alpha_{m} \frac{1}{\left(\sqrt{2\pi}\right)^{p} \left|\mathbf{A} C_{\mathbf{SS}}^{m} \mathbf{A}^{T} + C_{\mathbf{W}}\right|^{\frac{1}{2}} \exp\left[-\frac{1}{2} \left(\mathbf{X} - \left[\mathbf{A} \mu_{\mathbf{S}}^{m} + \mu_{\mathbf{W}}\right]\right)^{T} \left[\mathbf{A} C_{\mathbf{SS}}^{m} \mathbf{A}^{T} + C_{\mathbf{W}}\right]^{-1} \left(\mathbf{X} - \left[\mathbf{A} \mu_{\mathbf{S}}^{m} + \mu_{\mathbf{W}}\right]\right)\right]}{\sum_{j=1}^{M} \alpha_{j} \frac{1}{\left(\sqrt{2\pi}\right)^{p} \left|\mathbf{A} C_{\mathbf{SS}}^{j} \mathbf{A}^{T} + C_{\mathbf{W}}\right|^{\frac{1}{2}}} \exp\left[-\frac{1}{2} \left(\mathbf{X} - \left[\mathbf{A} \mu_{\mathbf{S}}^{j} + \mu_{\mathbf{W}}\right]\right)^{T} \left[\mathbf{A} C_{\mathbf{SS}}^{j} \mathbf{A}^{T} + C_{\mathbf{W}}\right]^{-1} \left(\mathbf{X} - \left[\mathbf{A} \mu_{\mathbf{S}}^{j} + \mu_{\mathbf{W}}\right]\right)\right]}, 1 \le m \le M.$$
(5)

Now, from Eqn. (1) we have

$$\begin{bmatrix} \mathbf{X} \\ \mathbf{S} \end{bmatrix} = \begin{bmatrix} \mathbf{A}\mathbf{S} + \mathbf{W} \\ \mathbf{S} \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{I} \\ \mathbf{I} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{S} \\ \mathbf{W} \end{bmatrix} = \mathbf{G} \begin{bmatrix} \mathbf{S} \\ \mathbf{W} \end{bmatrix}. \quad (8)$$

Property of GMM : If k-dimensional random vector U is GMM distributed as  $f_{\mathbf{U}}(\mathbf{u}) = \sum_{m=1}^{M} \gamma_m \mathbf{N} \left( \mu_{\mathbf{U}}^m, C_{\mathbf{U}}^m \right)$ , then the random vector  $\mathbf{V} = \mathbf{D}\mathbf{U}$  is also GMM distributed according to  $f_{\mathbf{V}}(\mathbf{v}) = \sum_{m=1}^{M} \gamma_m \mathbf{N} \left( \mathbf{D} \mu_{\mathbf{U}}^m, \mathbf{D} C_{\mathbf{U}}^m \mathbf{D}^T \right)$ , where **D** is a  $l \times k$  deterministic matrix.

Thus, using the above property of GMM, from Eqn. (6) and Eqn. (8),  $\mathbf{X}$  and  $\mathbf{S}$  are jointly GMM distributed as

$$f_{\mathbf{X},\mathbf{S}}(\mathbf{x},\mathbf{s}) = \sum_{m=1}^{M} \alpha_m \, \mathbf{N}\left(\mathbf{x},\mathbf{s};\mathbf{G}\,\boldsymbol{\mu}^m,\mathbf{G}\,\boldsymbol{C}^m\,\mathbf{G}^T\right), \quad (9)$$

where [using Eqn. (7) and the **G** matrix of Eqn. (8)]

$$\mathbf{G} \mu^{m} = \begin{bmatrix} \mu_{\mathbf{X}}^{m} \\ \mu_{\mathbf{S}}^{m} \end{bmatrix} = \begin{bmatrix} \mathbf{A} \mu_{\mathbf{S}}^{m} + \mu_{\mathbf{W}} \\ \mu_{\mathbf{S}}^{m} \end{bmatrix}, \text{ and}$$
$$\mathbf{G} C^{m} \mathbf{G}^{T} = \begin{bmatrix} C_{\mathbf{X}\mathbf{X}}^{m} & C_{\mathbf{X}\mathbf{S}}^{m} \\ C_{\mathbf{S}\mathbf{X}}^{m} & C_{\mathbf{S}\mathbf{S}}^{m} \end{bmatrix}$$
$$= \begin{bmatrix} \mathbf{A} C_{\mathbf{S}\mathbf{S}}^{m} \mathbf{A}^{T} + C_{\mathbf{W}} & \mathbf{A} C_{\mathbf{S}\mathbf{S}}^{m} \\ C_{\mathbf{S}\mathbf{S}}^{m} \mathbf{A}^{T} & C_{\mathbf{S}\mathbf{S}}^{m} \end{bmatrix}.$$
(10)

Since, S and X are jointly GMM distributed with known parameters, the conditional PDF of S given X can be evaluated as

$$f_{\mathbf{S}|\mathbf{X}}(\mathbf{s}|\mathbf{x}) = \sum_{m=1}^{M} \beta_m(\mathbf{x}) \mathbf{N}\left(\mathbf{s}; \mu_{\mathbf{S},c}^m(\mathbf{x}), C_{\mathbf{SS},c}^m\right),$$
where
$$\mu_{\mathbf{S},c}^m(\mathbf{x}) = \mu_{\mathbf{S}}^m + C_{\mathbf{S}\mathbf{X}}^m \left[C_{\mathbf{X}\mathbf{X}}^m\right]^{-1} \left(\mathbf{x} - \mu_{\mathbf{X}}^m\right),$$

$$C_{\mathbf{SS},c}^m = C_{\mathbf{SS}}^m - C_{\mathbf{S}\mathbf{X}}^m \left[C_{\mathbf{X}\mathbf{X}}^m\right]^{-1} C_{\mathbf{X}\mathbf{S}}^m,$$

$$\beta_m(\mathbf{x}) = \frac{\alpha_m f_{m,\mathbf{x}}(\mathbf{x})}{\sum_{m=1}^{M} \alpha_m f_{m,\mathbf{x}}(\mathbf{x})},$$

$$f_{m,\mathbf{X}}(\mathbf{x}) = \mathbf{N}\left(\mathbf{x}; \mu_{\mathbf{X}}^m, C_{\mathbf{X}\mathbf{X}}^m\right).$$
(11)

Hence, the MMSE estimator  $\widehat{\mathbf{S}} = \mathcal{E} \{ \mathbf{S} | \mathbf{X} \}$  is given as

$$\widehat{\mathbf{S}} = \mathcal{E} \left\{ \mathbf{S} | \mathbf{X} \right\} = \sum_{m=1}^{M} \beta_m(\mathbf{X}) \ \mu_{\mathbf{S},c}^m(\mathbf{X}) \,. \tag{12}$$

Substituting the values from Eqn. (10) and Eqn. (11) in Eqn. (12), the expression for the MMSE estimator is shown in Eqn. (4).

From Eqn. (5), we note that  $\beta_m(\mathbf{X})$  is a non-linear function of  $\mathbf{X}$  and thus, the estimator of Eqn. 4 is non-linear. The non-linear

weights,  $\{\beta_m(\mathbf{X})\}_{m=1}^M$ , satisfy  $\beta_m(\mathbf{X}) > 0$  and  $\sum_{m=1}^M \beta_m(\mathbf{X}) = 1$ . We mention that the usual MMSE estimator for the case of Gaussian distributed **S** is a special case of our estimator. For this case, M = 1 and thus, the estimator of Eqn. (4) reduces to the well-known MMSE estimator of linear form shown in Theorem 11.1 of [12]. In case of modeling non-Gaussian PDF, GMM is a generalization of the Gaussian model. Hence, we expect better performance of the non-linear estimator, for M > 1, compared to M = 1.

# 3. SPEECH ENHANCEMENT FRAME-WORK

The speech signal is processed as frames of p samples. For the *n*th frame we denote observed noisy speech vector by  $\mathbf{Y}(n)$  and corresponding clean speech vector by  $\mathbf{S}(n)$ . For simplicity, we drop the subscript 'n'. We model the PDF of  $\mathbf{S}$  using a GMM with M number of mixture components. Now, we propose different model of speech degradation, which fits into general Bayesian linear model frame-work of section 2.

Model 1: Observed noisy speech vector Y is is modeled as

$$\mathbf{Y} = \mathbf{S} + \mathbf{Z} \tag{13}$$

where the clean speech vector  $\mathbf{S}$  is degraded by additive noise vector  $\mathbf{Z}$ . We assume the PDF of  $\mathbf{Z}$  to be Gaussian. Therefore, the degraded speech model of Eqn. 13 fits into the general linear model frame-work of Section 2, where  $\mathbf{A}$  is a  $p \times p$  identity matrix. Now, we can use the MMSE estimator of  $\mathbf{S}$  given by Eqn. 4 to get the enhanced speech vector  $\hat{\mathbf{S}}$ .

*Model 2:* If the speech signal undergoes linear filtering distortion, characterized by a matrix  $\mathbf{H}$  then, the observed speech vector  $\mathbf{Y}$  is given by

$$\mathbf{Y} = \mathbf{H}\,\mathbf{S} + \mathbf{Z} \tag{14}$$

where  $\mathbf{Z}$  is defined as in *Model 1*. Now, if we have some prior information about  $\mathbf{H}$ , then using the estimator of section 2, speech enhancement can be done in a similar way as in *Model 1*.

*Model 3:* Many SE techniques use orthogonal transforms (such as Discrete Cosine Transform or Karhunen Love Transform) and then process speech for enhancement. We can easily extend our Bayesian MMSE estimator to work in a transform domain by modeling the degraded speech in the transform domain. If we denote the transformation matrix by  $\mathbf{T}$ , then transformed noisy speech vector  $\mathbf{X} = \mathbf{T} \mathbf{Y}$  can be modeled as

$$\mathbf{X} = \mathbf{A}\mathbf{S} + \mathbf{W} \tag{15}$$

where  $\mathbf{A} = \mathbf{T} \mathbf{H}$  and  $\mathbf{W}$  is additive noise vector in transform domain. Assuming the PDF of  $\mathbf{W}$  to be Gaussian, we use the MMSE



**Fig. 1**. SNR of enhanced speech signal at different vector dimension (p) and number of Gaussian mixtures (M) for optimum choice of p and M. (a) Input SNR = 0 dB. (b) Input SNR = 5 dB.

estimator of Eqn. 4 for speech enhancement.

It is clear that the SE models described above can handle a more general case of speech distortion than just additive noise. Also, the estimator can operate in time-domain as well as in the transform domain. We use Expectation Maximization (EM) algorithm to estimate the GMM parameters of clean speech from training data. For achieving optimum speech enhancement performance, the number of Gaussian components (M) of the GMM and the vector dimension (p) should be determined experimentally. Further, although the clean speech GMM parameters are found offline using training data, noise statistics (mean vector, covariance matrix) are estimated online.

#### 4. EXPERIMENTS AND RESULTS

# 4.1. Experimental Setup

The speech data used in the experiments are taken from the TIMIT database, where the speech is sampled at 16 kHz. For our experiments, the speech signal is first low pass filtered (3.4 kHz cut-off frequency) and then down-sampled to 8 kHz. We have used 40 minutes of speech data for training and a separate 3 minutes of speech data for testing. The training data is used for estimating the clean speech GMM parameters employing EM algorithm. The test speech is generated by adding noise to clean speech at the required level. Therefore, all our experiments are based on *Model 1* of section 3. We have considered two types of noise: white Gaussian noise (generated by MATLAB) and aircraft cockpit noise (taken from Duke University database). In all our experiments we have assumed the noise to be zero mean. Assuming the noise to be stationary, covariance matrix of noise is estimated only once from the initial 200 msec segment (contains only noise) of the test speech.

To measure the speech enhancement performance, we have used the following objective measures as used in [4]: signal-to-noise ratio (SNR), average segmental SNR (SSNR) and average log spectral distortion (SD). The SSNR and SD measures are perceptually motivated objective measures widely used in speech enhancement experiments.

 Table 1. Performance of the proposed GMM based MMSE estimator

 for white Gaussian noise at different input SNR

SNR (dB)		Avg. SSNR (dB)		Avg. SD (dB)	
Input	Output	Input	Output	Input	Output
-5	6.07	-14.47	-0.24	9.47	7.11
-2.5	7.58	-11.97	1.09	9.19	6.76
0	9.12	-9.47	2.43	8.87	6.41
2.5	10.65	-6.97	3.65	8.50	6.13
5	12.17	-4.48	4.74	8.09	5.86
7.5	13.76	-1.98	5.85	7.63	5.59
10	15.39	0.51	6.89	7.15	5.30

**Table 2**. Performance of the proposed GMM based MMSE estimator for cockpit noise at different input SNR

SNR (dB)		Avg. SSNR (dB)		Avg. SD (dB)	
Input	Output	Input	Output	Input	Output
-5	6.29	-14.49	-0.09	9.49	7.09
-2.5	7.77	-11.99	1.23	9.21	6.74
0	9.25	-9.49	2.49	8.89	6.43
2.5	10.74	-6.99	3.67	8.52	6.13
5	12.26	-4.49	4.79	8.10	5.85
7.5	13.84	-1.99	5.87	7.65	5.57
10	15.44	0.50	6.92	7.16	5.28

### 4.2. Performance of the GMM based MMSE method

It is clear that larger frames of speech carry more information enabling the statistical speech enhancement. However, larger vectors demand larger M and increasing the computational complexity. Thus, it is important to find the reasonable values of p and Msuch that the estimator provides an acceptable trade-off between improvement in performance and complexity. For the additive white Gaussian noise case, the output SNR performance (in dB) of the enhanced speech is shown in Fig. 1, for different values of p and M, under different input SNR conditions. We note that the output SNR increases as M increases for any particular value of p. This behavior is attributed to the fact that the use of higher number of Gaussian components leads to better PDF matching of the source signal and thus results in better enhancement. Further, for a fixed M, if p is increased (keeping the same amount of training data) the performance increases in general upto moderate values of p, but starts to droop for higher p. This can be attributed to two reasons: (1) The initial increase with p is due to exploitation of higher signal correlation between samples, and (2) The drooping for higher p could be due to modeling approximation for fixed training data. From Fig. 1, we choose the optimum values as p = 40 and M = 256 and keep these parameters fixed for our rest of the work.

For performance evaluation, we have processed the noisy speech as frames of 40 samples with 50% overlap between successive frames. Successive enhanced frames are overlap-added using Hamming window as in [8]. The performance of GMM based MMSE estimator is shown in Table 1 for the white Gaussian noise case. At a low input SNR of -5 dB, we note that the MMSE estimator provides more than 10 dB, 14 dB and 2 dB performance improvements respectively in terms of output SNR, average SSNR and average SD measures. The uniform improvement of 2 dB in average SD for all the 4 input SNR conditions, indicates that MMSE estimator is providing better intelligibility enhancement, in addition to quality improvement in terms of SNR and SSNR. It is observed that the

The und Bolzou for white Guessian house at anterent input State							
SNR (dB)		Avg. SSNR (dB)		Avg. SD (dB)			
Input	Output	Input	Output	Input	Output		
-5	5.01	-14.47	-1.55	9.47	9.28		
-2.5	6.49	-11.97	-0.13	9.19	8.71		
0	8.08	-9.47	1.30	8.87	8.16		
2.5	9.75	-6.97	2.76	8.50	7.57		
5	11.51	-4.48	4.23	8.09	6.99		
7.5	13.33	-1.98	5.78	7.63	6.44		
10	15.24	0.51	7.35	7.15	5.88		

**Table 3**. Performance of the generalized subspace based method of Hu and Loizou for white Gaussian noise at different input SNR

**Table 4**. Performance of the generalized subspace based method of

 Hu and Loizou for cockpit noise at different input SNR

SNR (dB)		Avg. SSNR (dB)		Avg. SD (dB)	
Input	Output	Input	Output	Input	Output
-5	5.12	-14.49	-1.74	9.49	8.98
-2.5	6.61	-11.99	-0.27	9.21	8.43
0	8.20	-9.49	1.20	8.89	7.87
2.5	9.87	-6.99	2.69	8.52	7.30
5	11.62	-4.49	4.22	8.10	6.73
7.5	13.42	-1.99	5.79	7.65	6.18
10	15.32	0.50	7.37	7.16	5.65

improvements in all the performance measures decrease as the input SNR increases. This observation is consistent with the general results for speech enhancement in the literature. Table 2 shows the performance of the MMSE estimator where the speech signal is contaminated with aircraft cockpit noise and the noise PDF is modeled using Gaussian distribution. The same trend in improvements of performance measures are observed like in the case of Gaussian noise.

### 4.3. Comparison with generalized subspace based method

We compare the performance of developed method with the generalized subspace based method (Time Domain Constrained estimator) of Hu and Loizou [8], which has been shown to be an improvement over the subspace based method of Ephraim and Van Trees [7]. Table 3 and Table 4 show the performances of generalized subspace based method, respectively for the Gaussian noise and aircraft cockpit noise cases. For the white Gaussian noise, comparing Table 1 and Table 3, we note that our method provides better performance than the generalized subspace based method in all the aspects. At low values of input SNR, the GMM based MMSE method provides more than 1 dB improvement compared to the generalized subspace based method in the sense of output SNR and average SSNR. Also, the GMM based MMSE method provides considerable improvement in performance in the sense of average SD for all the input SNR conditions considered. For the aircraft cockpit noise also, comparing Table 2 and Table 4, we observe same trend in performance improvement for the GMM based MMSE method over the generalized subspace based method. Thus, the GMM based MMSE estimator can be regarded as an effective method for speech enhancement.

## 5. CONCLUSION

For the general linear model of signal distortion, we derive a Bayesian MMSE estimator where the PDF of clean signal vector to be estimated is modeled using GMM instead of the usual Gaussian distribution. The estimator is shown to be a non-linear one which encompasses the well-known MMSE estimator for the Gaussian case. For speech enhancement, the estimator is shown to perform better than the generalized subspace based method of Hu and Loizou [8], and the improvement in performance is attributed to the fact that the PDF of clean speech is better modeled using the GMM.

### 6. REFERENCES

- P.C. Loizou, "Speech enhancement: Theory and practice," CRC Press, 2007.
- [2] A. Rezayee and S. Gazor, "An adaptive KLT approach for speech enhancement," *IEEE Trans. Speech and Audio Processing*, vol. 9, no. 2, pp. 87-95, Feb. 2001.
- [3] Y. Ephraim, "A Bayesian estimation approach for speech enhancement using hidden Markov models," *IEEE Trans. Signal Processing*, vol. 40, no. 4, pp. 725 - 735, April 1992.
- [4] D.Y. Zhao and W.B. Kleijn, "HMM-Based Gain Modeling for Enhancement of Speech in Noise," *IEEE Trans. Audio, Speech* and Language Processing, vol. 15, Issue 3, pp. 882-892, March 2007.
- [5] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean square error short-time spectral amplitude estimator," *IEEE Trans. Acoust., Speech and Signal Processing*, vol. ASSP-32, pp. 1109-1121, Dec. 1984.
- [6] C. Breithaupt and R. Martin, "MMSE estimation of magnitudesquared DFT coefficients with superGaussian priors," *Proc. ICASSP*, pp. 896-899, Apr. 2003.
- [7] Y. Ephraim and H.L. Van Trees, "A signal subspace approach for speech enhancement," *IEEE Trans. Speech and Audio Processing*, vol. 3, no. 4, pp. 251-266, July 1995.
- [8] Y. Hu and P.C. Loizou, "A generalized subspace approach for enhancing speech corrupted by colored noise," *IEEE Trans. Speech and Audio Processing*, vol. 11, no. 4, pp. 334 - 341, July 2003.
- [9] S. Gazor and W. Zhang, "Speech probability distribution," *IEEE Signal Processing Letters*, vol. 10, no. 7, pp. 204-207, July 2003.
- [10] S. Gazor and W. Zhang, "Speech enhancement employing Laplacian-Gaussian mixture," *IEEE Trans. Speech and Audio Processing*, vol. 13, no. 5, pp. 896-904, Sept. 2005.
- [11] L. Deng, J. Droppo and A. Acero, "Estimating cepstrum of speech under the presence of noise using a joint prior of static and dynamic features," *IEEE Trans. Speech and Audio Processing*, vol. 12, no. 3, pp. 218-233, May 2004.
- [12] S.M. Kay, "Fundamentals of statistical signal processing: Estimation theory," Prentice Hall, vol. II, pp. 364-365, 1993.