SPEECH ENHANCEMENT BASED ON SPEECH SPECTRAL COMPLEX GAUSSIAN MIXTURE MODEL

Guo-Hong Ding, Xia Wang, Yang Cao, Feng Ding and Yuezhong Tang

Nokia Research Center, Beijing

ABSTRACT

This paper presents a speech enhancement approach based on speech spectral complex Gaussian Mixture Model (GMM). First, a construction algorithm of speech spectral GMM is introduced and it is based on the distance measure of speech spectral Gaussian probability. Then a noise estimation algorithm based on the GMM is proposed in the Maximum Likelihood criterion using the Expectation-Maximum (EM) algorithm. Speech enhancement experimental results show that the GMM-based MMSE estimators, especially the GMM-based MMSE short-time spectral estimator, can afford better performance than alternative speech enhancement algorithms and the proposed noise estimation algorithm can improve the enhancement performance more, especially at low SNRs.

1. INTRODUCTION

Speech quality and intelligibility might significantly deteriorate in the presence of background noise and suppression of additive noise is crucial to all types of speech applications including speech coding and speech recognition. Speech enhancement has therefore attracted a great deal of research interest for the past two decades.

Many speech enhancement approaches are based on the Gaussian assumption of speech signals. Spectral subtraction is an optimal speech spectral component variance estimator [3], which is developed with the speech spectral Gaussian assumption. With the assumption, speech spectra, amplitude and log-amplitude can be estimated in the minimum mean-square error criterion (MMSE) and the obtained estimators are Wiener filter [2], the MMSE-STSA estimator [5] and the MMSE-LSA estimator [6], respectively.

The performances of the above estimators are greatly dependent on the accuracy of the speech spectral Gaussian model. In fact, the model parameters are updated by a recursive averaging with a time constant comparable to the correlation time of speech variation [5]. The parameter update gives a coarse estimate of the speech parameter and it may constrain the enhancement performance seriously. It is obvious that a more detailed and more accurate description of speech distribution, i.e., Gaussian Mixture Model (GMM), can provide better performance.

In [8], a speech log-spectral GMM is used to construct speech MMSE estimator. In the log-spectral domain the environmental function is nonlinear and the compensation was implemented on a maximum assumption in the production of noisy speech log-spectra. Thus the approximation also constrains the speech enhancement performance. But in the spectral domain, the environmental function is linear and the compensation can be implemented more directly and more accurately.

In [1], a speech spectral HMM is used to construct speech spectral estimators. The HMM is built from the AR-HMM in the time domain. Given the Fourier transform $A(\theta)$ of the AR coefficients, the power spectral densities of the corresponding process can be formulated as

$$S(\theta) = \sigma^2 / \left|A(\theta)\right|^2 \tag{1}$$

where σ^2 is the gain of the AR process. Besides, the covariance matrices of the HMM's can be formulated as

$$\mathbf{\Sigma} = K^{-1}\boldsymbol{USU}^{\#} \tag{2}$$

where $\boldsymbol{U}$ is a $K \times K$ matrix whose (k,n) element is the complex exponential $\exp(-j2\pi kn/K)$ and $\boldsymbol{S}$ is a $K \times K$ diagonal matrix whose k-th diagonal element is given by $S(2\pi k/K)$.

According to equations (1,2), an AR-HMM with mixture components can be easily converted to HMM in the spectral domain. Based on the speech spectral HMM, speech enhancement algorithms can be constructed conveniently [1]. It is well known that GMM is a special case of HMM, with which time correlation is considered. In speech enhancement, since the transcriptions are various, it is difficult to segment the observed speech spectra accurately with just only one HMM. Thus GMM can provide almost the same speech clustering information as HMM in this case.

Note that in [1], the speech model in the spectral domain is heavily dependent on the accuracy of AR coefficients, which are difficult to extract accurately for some consonants. In this paper, the speech model in the spectral domain is constructed more accurately. It is built directly in the spectral domain using an algorithm similar to the K-means clustering algorithm using the Gaussian probability as the distance measure.

Based on the constructed speech spectral complex GMM, noise estimation is implemented in the ML framework using the EM algorithm. With the noise estimation, better enhancement performance can be obtained using the proposed GMM-based enhancement algorithms, especially in low SNR environments.

2. GMM-BASED MMSE ESTIMATORS

In [1], the HMM-based MMSE estimators of speech spectra, magnitude, magnitude-squared spectra and log-spectral magnitude are constructed. Here, we will introduce the estimators based on speech spectral GMM.

For stationary stochastic signals, the frequency spectrum can be modeled as the complex Gaussian distribution with zero means. Thus noise spectra can be modeled as single Gaussian model

$$p(\boldsymbol{N}) = \prod_{k=0}^{K} \frac{1}{\pi\lambda_{nk}} \exp\left\{-\frac{|N_k|^2}{\lambda_{nk}}\right\} \tag{3}$$

Speech is a non-stationary process, but over short duration (typically 20-30 ms) speech signals can be considered stationary and speech spectra satisfy Gaussian distribution. To deal with speech non-stationarity, we model the speech distribution with a mixture of Gaussian components, each representing a different class and assume that speech spectra in one class are independent of those in others. So the distribution of the speech spectral vector can be formulated as a weighted sum of Gaussian distributions over all classes

$$p(\boldsymbol{X}) = \sum_{m=1}^{M} c_m \prod_{k=0}^{K} \frac{1}{\pi \lambda_{xmk}} \exp\left\{ -\frac{|X_k|^2}{\lambda_{xmk}} \right\} \tag{4}$$

where M denotes the number of mixture components, c_m is the weighted coefficient of component m and λ_{xmk} is the k-th speech variance of component m.

Since speech spectra of one mixture component are assumed to be independent of those of others, the estimators can be implemented as a weighted sum of the MMSE estimators over all mixture components, i.e.

$$E[\boldsymbol{f}(\boldsymbol{X}) \mid \boldsymbol{Y}] = \sum_{m=1}^{M} P(m \mid \boldsymbol{Y}) E[\boldsymbol{f}(\boldsymbol{X}) \mid \boldsymbol{Y}, m] \tag{5}$$

where $P(m \mid \boldsymbol{Y})$ is the class-conditional probability given $\boldsymbol{Y}$.

2.1 Calculation of the class-conditional probability

Since the speech spectral distribution and the noise spectral distribution are illustrated by Eq. (4) and Eq. (3), respectively, and the environmental function in the spectral domain is linear, the noisy spectral model can be formulated as

$$p(\boldsymbol{Y}) = \sum_{m=1}^{M} c_m \prod_{k=0}^{K} \frac{1}{\pi (\lambda_{xmk} + \lambda_{nk})} \exp\left\{ -\frac{|Y_k|^2}{\lambda_{xmk} + \lambda_{nk}} \right\} \tag{6}$$

Thus the class-conditional probability given the noisy vector can be formulated as

$$P(m \mid \boldsymbol{Y}) = \frac{c_m p(\boldsymbol{Y} \mid m)}{\sum_{m=1}^{M} c_m p(\boldsymbol{Y} \mid m)}, \ 1 \le m \le M \tag{7}$$

2.2 Construction of the MMSE estimators

The last term of equation (5) is the *a posteriori* estimator given the noisy vector and the mixture component, which means that the speech distribution is specified by the component. In this case, the estimator can be constructed in the MMSE criterion.

If we define

$$\boldsymbol{f}(\boldsymbol{X}) = [X_0, X_1, \cdots, X_K]^T$$

then Eq. (5) is the GMM-based MMSE short-time spectral estimator (GMM-STS). The last term of Eq. (5) can be formulated as

$$E[X_k \mid Y_k, m] = \frac{\lambda_{xmk}}{\lambda_{xmk} + \lambda_{nk}} Y_k, \ 0 \le k \le K, \ 1 \le m \le M$$

In fact the above equation is the famous Wiener filter if component index m is ignored. With the GMM assumption, the GMM-STS estimator is a weighted sum of Wiener filters over all mixture components, and the weighted coefficients are the *a posteriori* probabilities of the mixture components given the noisy vector.

If we define

$$\boldsymbol{f}(\boldsymbol{X}) = \left[|X_0|^2, |X_1|^2, \cdots, |X_K|^2 \right]^T$$

then Eq. (5) is the GMM-based MMSE magnitude-squared spectral estimator (GMM-MSS). The last term of Eq. (5) can be formulated as

$$E[|X_k|^2 | Y_k, m] = \left| \frac{\lambda_{xmk}}{\lambda_{xmk} + \lambda_{nk}} Y_k \right|^2 + \frac{\lambda_{xmk}\lambda_{nk}}{\lambda_{xmk} + \lambda_{nk}}, \ 0 \le k \le K, \ 1 \le m \le M$$

The estimator with single Gaussian model was published in [10].

If we define

$$\boldsymbol{f}(\boldsymbol{X}) = [|X_0|, |X_1|, \cdots, |X_K|]^T$$

or

$$\boldsymbol{f}(\boldsymbol{X}) = [\log|X_0|, \log|X_1|, \cdots, \log|X_K|]^T$$

then Eq. (5) is the GMM-based MMSE short time spectral amplitude estimator (GMM-STSA) or the GMM-based MMSE log-spectral amplitude estimator (GMM-LSA). The last terms of Eq. (5) were solved in [5] and [6], respectively.

3. CONSTRUCTION OF SPEECH SPECTRAL COMPLEX GAUSSIAN MIXTURE MODEL

We train clean speech spectral complex GMM from high quality speech data using a clustering algorithm, which is different from the K-means clustering algorithm broadly used in speech recognition. Since speech spectra are assumed to be Gaussian distributed, the complex Gaussian probability density is an appropriate distance measure to cluster speech spectra.

The construction of speech spectral GMM is an iterative process. Let the training data consist of N clean speech spectral frames $\mathbf{X} = \{\boldsymbol{X}^1, ..., \boldsymbol{X}^N\}$. $\boldsymbol{\lambda} = \{\lambda_{xmk}, c_m, 1 \le m \le M, 0 \le k \le K\}$ is known initial parameters, $\hat{\boldsymbol{\lambda}}_n = \{\hat{\lambda}_{xmk}, \hat{c}_m, 1 \le m \le M, 0 \le k \le K\}$ is the parameters to be estimated in the current iteration and $\gamma_{n,m}$ denotes the *a posteriori* probability of the component given the observations as

$$\gamma_{n,m} = c_m \prod_{k=0}^{K} \frac{1}{\pi \lambda_{xmk}} \exp\left\{ -\frac{|X_k^n|^2}{\lambda_{xmk}} \right\}, 1 \le n \le N, 1 \le m \le M$$

If we define

$$\alpha_{n,m} = \begin{cases} 1 & if \ \ \gamma_{n,m} = \max\limits_{1 \le m' \le M} \gamma_{n,m'} \\ 0 & otherwise \end{cases}, 1 \le n \le N, 1 \le m \le M$$

the estimated parameters satisfy the following equations

$$\hat{c}_m = \frac{\sum_{n=1}^{N} \alpha_{n,m}}{N}, \ \hat{\lambda}_{xmk} = \frac{\sum_{n=1}^{N} \alpha_{n,m} \cdot \left| X_k^n \right|^2}{\sum_{n=1}^{N} \alpha_{n,m}}, 0 \le k \le K, 1 \le m \le M$$

4. NOISE ESTIMATION BASED ON SPEECH SPECTRAL GAUSSIAN MIXTURE MODEL

For each utterance, an utterance-specific noise model can be constructed on the noisy signals given the speech spectral GMM. In noise model (3), only noise variances need to be estimated and the estimation can be implemented in the ML framework using the EM algorithm.

The parameters are estimated by maximizing the following auxiliary function

$$Q(\boldsymbol{\lambda}_n,\overline{\boldsymbol{\lambda}}_n)=E[\log p(\boldsymbol{Y},\boldsymbol{X},\boldsymbol{M};\overline{\boldsymbol{\lambda}}_n)\,|\,Y,\boldsymbol{\lambda}_n]$$
$$=\sum_{t=1}^{T}\sum_{m=1}^{M}\int p(\boldsymbol{X}_t,m\,|\,\boldsymbol{Y}_t,\boldsymbol{\lambda}_n)\log p(\boldsymbol{Y}_t,\boldsymbol{X}_t,m;\overline{\boldsymbol{\lambda}}_n)d\boldsymbol{X}_t$$

where $\boldsymbol{\lambda}_n$ and $\overline{\boldsymbol{\lambda}}_n$ are known and unknown noise parameters, respectively, $\boldsymbol{M}$ is mixture component series and $\boldsymbol{Y}=\{\boldsymbol{Y}_1,...,\boldsymbol{Y}_T\}$, $\boldsymbol{X}=\{\boldsymbol{X}_1,...,\boldsymbol{X}_T\}$ are known corrupted and unknown clean speech spectral vectors series.

Consider the following equations

$$p(\boldsymbol{X}_t,m\,|\,\boldsymbol{Y}_t;\boldsymbol{\lambda}_n)=p(\boldsymbol{X}_t\,|\,m,\boldsymbol{Y}_t;\boldsymbol{\lambda}_n)p(m\,|\,\boldsymbol{Y}_t;\boldsymbol{\lambda}_n)$$
$$p(\boldsymbol{Y}_t,\boldsymbol{X}_t,m;\overline{\boldsymbol{\lambda}}_n)=p(\boldsymbol{Y}_t\,|\,\boldsymbol{X}_t,m;\overline{\boldsymbol{\lambda}}_n)p(\boldsymbol{X}_t\,|\,m)p(m)$$

we can rewrite the auxiliary function as

$$Q(\boldsymbol{\lambda}_n,\overline{\boldsymbol{\lambda}}_n)=\sum_{t=1}^{T}\sum_{m=1}^{M}p(m\,|\,\boldsymbol{Y}_t;\boldsymbol{\lambda}_n)\int p(\boldsymbol{X}_t\,|\,m,\boldsymbol{Y}_t;\boldsymbol{\lambda}_n)\log p(\boldsymbol{Y}_t\,|\,\boldsymbol{X}_t,m;\overline{\boldsymbol{\lambda}}_n)dX_t$$

where

$$p(\boldsymbol{X}_t\,|\,m,\boldsymbol{Y}_t;\boldsymbol{\lambda}_n)=N(\boldsymbol{X}_t;\boldsymbol{\mu}_{x|y,m},\boldsymbol{\sigma}^2_{x|y,m})$$
$$\mu_{x|y,m,k}=\frac{\lambda_{xmk}}{\lambda_{xmk}+\lambda_{nk}}y_{tk}\,,\ \sigma^2_{x|y,m,k}=\frac{\lambda_{xmk}\lambda_{nk}}{\lambda_{xmk}+\lambda_{nk}}\,,\ k=0,...,K$$
$$\log p(\boldsymbol{Y}_t\,|\,\boldsymbol{X}_t,m;\overline{\boldsymbol{\lambda}}_n)=C-\frac{1}{2}\sum_{k=0}^{K}\log\overline{\lambda}_{nk}-\sum_{k=0}^{K}\frac{(y_{tk}-x_{tk})^2}{2\overline{\lambda}_{nk}}$$

The auxiliary function can be expanded as

$$Q(\boldsymbol{\lambda}_n,\overline{\boldsymbol{\lambda}}_n)=-\frac{1}{2}\sum_{t=1}^{T}\sum_{m=1}^{M}\sum_{k=0}^{K}p(m\,|\,\boldsymbol{Y}_t;\boldsymbol{\lambda}_n)\int\left[\log\overline{\lambda}_{nk}+\frac{(y_{tk}-x_{tk})^2}{\overline{\lambda}_{nk}}\right]$$
$$N(x_{tk};\mu_{x|y,m,k},\sigma^2_{x|y,m,k})dx_{tk}$$
$$=-\frac{1}{2}\sum_{t=1}^{T}\sum_{m=1}^{M}\sum_{k=0}^{K}p(m\,|\,\boldsymbol{Y}_t;\boldsymbol{\lambda}_n)\left[\log\overline{\lambda}_{nk}+\frac{1}{\overline{\lambda}_{nk}}\left(\frac{\lambda_{xmk}\lambda_{nk}}{\lambda_{xmk}+\lambda_{nk}}+\frac{\lambda_{nk}}{\lambda_{xmk}+\lambda_{nk}}|\,y_{tk}\,|^2\right)\right]$$

By taking the gradient of $Q(\boldsymbol{\lambda}_n,\overline{\boldsymbol{\lambda}}_n)$ with respect to noise parameter $\hat{\lambda}_{nk},k=0,...,K$ and letting the gradient zero, we obtain the estimates of noise parameters, for $k=0,...,K$

$$\overline{\lambda}_{nk}=\frac{\sum_{t=1}^{T}\sum_{m=1}^{M}p(m\,|\,\boldsymbol{Y}_t;\boldsymbol{\lambda}_n)\left(\frac{\lambda_{xmk}\lambda_{nk}}{\lambda_{xmk}+\lambda_{nk}}+\left(\frac{\lambda_{nk}}{\lambda_{xmk}+\lambda_{nk}}\right)^2|\,Y_{tk}\,|^2\right)}{\sum_{t=1}^{T}\sum_{m=1}^{M}p(m\,|\,Y_t;\boldsymbol{\lambda}_n)}$$
$$=\frac{1}{T}\sum_{m=1}^{M}\left[\frac{\lambda_{xmk}\lambda_{nk}}{\lambda_{xmk}+\lambda_{nk}}\sum_{t=1}^{T}p(m\,|\,\boldsymbol{Y}_t;\boldsymbol{\lambda}_n)+\left(\frac{\lambda_{nk}}{\lambda_{xmk}+\lambda_{nk}}\right)^2\sum_{t=1}^{T}pp(m\,|\,\boldsymbol{Y}_t;\boldsymbol{\lambda}_n)\,|\,Y_t\,|^2\right]$$

5. EXPERIMENTAL EVALUATION

In this section, speech enhancement experiments are performed to evaluate the performance of the proposed algorithms.

5.1 Experimental Settings

In the experiments, all utterances are sampled at 8kHz with a 16bit resolution. The database is composed of two independent sections. Section-1, which is used to train the speech spectral complex GMM, includes 4880 Mandarin utterances from 80 speakers (40 males and 40 females). Section-2, which is used for performance evaluation, consists of speech files collected from 10 speakers (5 males and 5 females), each delivering 10 Mandarin utterances. We obtain the noisy speech by adding white noise and factory noise to the clean speech with noise amplitudes being adjusted to achieve SNRs of 0, 5, 10, 15 and 20dB, respectively. The noise data are chosen from the NOISEX-92 database [7] resampled to 8kHz.

In all the experiments, the frame length is 256, which corresponds to $K=128$. Frame overlapping of 50% is used so that the reconstructed speech signals are synthesized as in [4].

A diagonal-variance GMM with 256 components for 129-d speech spectral vectors is trained using the high quality speech data of Section-1 to represent the distribution of clean speech spectral vectors.

5.2 Experimental evaluation on the proposed approach

Speech enhancement experiments are performed on the proposed algorithms in comparison with the alternative ones.

The alternative algorithms include MMSE-STS (wiener filter), MMSE-MSS, MMSE-STSA and MMSE-LSA, all of which have been published in [2], [10], [5] and [6], respectively. In all the algorithms, noise statistics are estimated using a rough Voice Activity Detector.

	0dB	5 dB	10 dB	15 dB	20 dB
MMSE-STS	5.60	9.82	14.26	18.58	22.79
MMSE-MSS	4.05	8.07	12.6	17.19	21.78
MMSE-STSA	5.72	9.95	14.36	18.64	22.79
MMSE-LSA	6.17	10.42	14.77	18.95	22.99
GMM-STS	8.91	12.65	16.25	19.73	23.26
GMM-MSS	7.50	11.47	15.27	18.94	22.56
GMM-STSA	7.16	11.32	15.32	19.04	22.68
GMM-LSA	7.66	11.74	15.62	19.23	22.79
GMM-STS-NE	9.22	12.66	16.16	19.42	22.77
GMM-MSS-NE	8.45	12.04	15.63	18.97	22.38
GMM-STSA-NE	7.33	11.28	15.24	18.78	22.23
GMM-LSA-NE	7.81	11.66	15.49	18.91	22.28

Table 1. SNR of different algorithms with noise statistics estimated before speech begins in additive white noise

	0dB	5 dB	10 dB	15 dB	20 dB
MMSE-STS	4.51	8.71	13.07	17.38	21.95
MMSE-MSS	3.59	7.57	12.00	16.54	21.36
MMSE-STSA	4.44	8.70	13.04	17.37	21.89
MMSE-LSA	4.68	8.99	13.30	17.56	22.01
GMM-STS	4.87	9.33	13.45	17.41	21.58
GMM-MSS	3.97	8.29	12.50	16.60	20.89
GMM-STSA	4.04	8.59	12.90	17.02	21.14
GMM-LSA	4.32	8.90	13.18	17.22	21.29
GMM-STS-NE	5.10	9.48	13.49	17.33	21.04
GMM-MSS-NE	4.46	8.83	12.82	16.74	20.55
GMM-STSA-NE	4.44	8.84	13.07	16.99	20.74
GMM-LSA-NE	4.70	9.10	13.28	17.14	20.83

Table 2. SNR of different algorithms with noise statistics estimated before speech begins in additive factory noise

The proposed GMM-based enhancement algorithms with the rough Voice Activity Detector are implemented and they are denoted as GMM-STS, GMM-MSS, GMM-STSA and GMM-LSA, respectively. Besides, the proposed GMM-based enhancement algorithms with the noise estimation algorithm described in

Section 4 are also evaluated. They are denoted as GMM-STS-NE, GMM-MSS-NE, GMM-STSA-NE and GMM-LSA-NE, respectively.

Table 1 and Table 2 display the SNRs of enhanced speech data using the 12 algorithms in additive white noise and additive factory noise, respectively. It is obvious that all GMM-based estimators are superior to the corresponding MMSE estimators at almost all noise levels in both noise types except at 20dB SNR in factory noise.

Furthermore, from Table 1 and Table 2, it is also showed that the noise estimation algorithm described in Section 4 can give more performance improvement in low SNR environments. But at high SNRs the noise estimation algorithm does not take its advantage, since the rough VAD can give enough accurate noise estimates at high SNRs.

Besides, it is obvious that GMM-STS and GMM-STS-NE achieve the best performance in all the GMM-based estimators and in all the GMM-based ones with noise estimation, respectively. The results are also confirmed in [1]. Clearly, the objective function of spectral estimators is consistent with the SNR test, which is a wave comparison between the tested sentences and the clean ones.

Similar results also appear in the subjective quality evaluation. Even if the MMSE estimators are applied, the remained noises are still distinct and annoying. But when the GMM-based estimators are applied to noisy data at 5dB input SNR, for the enhanced speech data, especially those processed by GMM-STS and GMM-STS-NE, a significant reduction in the noise level is perceived and the remained noise is little. The phenomena are also testified in the next section.

5.3 Data analysis

Sonograms of the clean, noisy, and several enhanced speech using MMSE-STS, GMM-STS and GMM-STS-NE at 5dB SNR are showed in Figure 1. From the figure, it is obvious that although MMSE estimators are applied, speech signals are still drowned in the remained noises and the GMM-based estimators can give greater noise reduction than the MMSE ones. Besides, the figure shows that GMM-based estimators with noise estimation can give even more noise reduction while main speech information is kept.

From the clean wav plot it is clear that the original speech is still degraded very lightly by noises, which make the sonogram a little dark when speech is absent. In the sonogram of “GMM-STS-NE”, it is clear that all these dark areas are removed while the main light areas are remained. In comparison of the sonogram of “clean”, a little speech distortion exists in the one of “GMM-STS-NE” since the sonograms in the high frequency domain is a little different when speech is present. It can be concluded that with GMM-STS-NE, almost all noises are removed, while almost all main speech information is remained though a little speech distortion is introduced.

6. CONCLUSION

In this paper, the MMSE estimation based on speech spectral complex GMM is proposed to estimate uncorrupted speech spectra, magnitude-squared spectra, spectral amplitudes and log-spectral amplitudes. Based on the speech spectral GMM, noise estimation from noisy data is introduced using the EM algorithm in the ML framework. Experimental results show that the GMM-based MMSE estimators, especially the GMM-based MMSE short-time spectral estimator, can afford better performance than alternative speech enhancement algorithms and with the noise estimation the enhancement algorithms can introduce better enhancement performance.

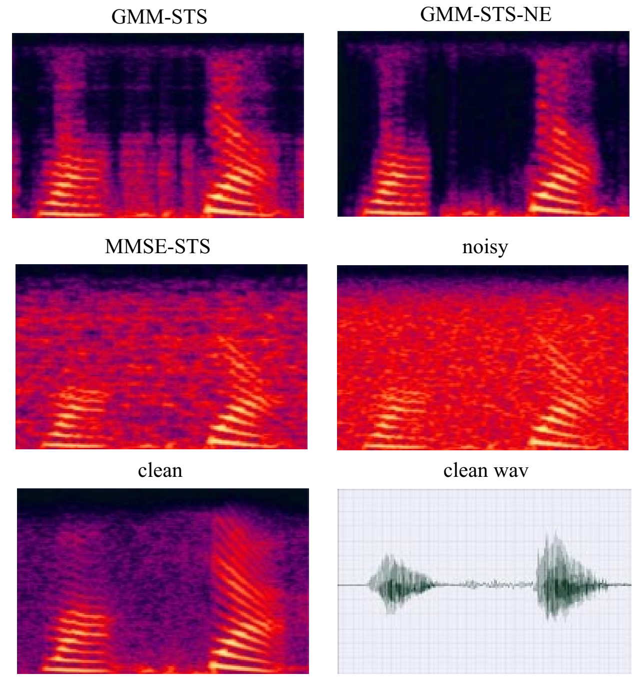

Figure 1. Sonograms of the clean, noisy, and several enhanced speech , along with the clean wav plot

7. REFERENCES

[1] Y. Ephraim, A Bayesian Estimation Approach for Speech Enhancement Using Hidden Markov Models, *IEEE Trans. Signal Proc.*, 40:725-735, 1992.

[2] J. S. Lim, A. V. Oppenheim, Enhancement and Bandwidth Compression of Noisy Speech, *Proc. of IEEE*, 67(12): 1586-1604, 1979.

[3] R. J. McAulay, M. L. Malpass, Speech Enhancement Using a Soft-Decision Noise Suppression Filter, *IEEE Trans. Acoust. Speech and Signal Proc.*, 28(2): 137-145, 1980.

[4] D. W. Griffin, J. S. Lim, Signal Estimation from Modified Short-Time Fourier Transform, *IEEE Trans. Acoust. Speech and Signal Proc.*, 32(2): 236-242, 1984.

[5] Y. Ephraim, D. Malah, Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator, *IEEE Trans. Acoust. Speech and Signal Proc.*, 32(6): 1109-1121, 1984.

[6] Y. Ephraim, D. Malah, Speech Enhancement Using a Minimum Mean-Square Error Log-Spectral Amplitude Estimator, *IEEE Trans. Acoust. Speech and Signal Proc*., 33(2): 443-445, 1985.

[7] Varga and H. J. M. Steeneken, Assessment for Automatic speech Recognition: II. NOISEX-92: A Database and an Experiment to Study the Effect of Additive Noise on Speech Recognition Systems, *Speech Communication*, 12: 247-251, 1993.

[8] D. Burshtein, S. Gannot, Speech Enhancement Using a Mixture-Maximum Model, *IEEE Trans. Speech and Audio Proc.*, 10(6): 341-351, 2002.

[9] P. Dempster, N. M. Laird and D. B. Rubin, Maximum Likelihood Estimation from Incomplete Data, *J. Royal Statist. Soc.*, 39(1): 1-38, 1977.

[10] J. Accardi, R. V. Cox, A Modular Approach to Speech Enhancement with an Application to Speech Coding, *In Proc. ICASSP*, 1999.