IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 2 Issue 6, June 2015.
www.ijiset.com
ISSN 2348 – 7968

MMSE STSA Based Techniques for Single channel Speech Enhancement Application

Simit Shah[1] , Roma Patel[2]

[1] Electronics and Communication Department,
Parul institute of engineering and technology,
Vadodara, Gujarat india
[2] Electronics and Communication Department,
Parul institute of engineering and technology,
Vadodara, Gujarat india

Abstract

In speech communication, the occurrence of background noise interference causes the quality as pleasantness or intelligibility as clearness of speech to degrade. The quality of speech can be influenced in data conversion, microphone, noisy data channel, or reproduction like loudspeakers and headphones. Suppressing and Cancelling additive noise in corrupted speech is an important task in speech communication systems. Since it needs to improve the perceptual quality as for human listeners and intelligibility for the machine such system. So various approaches have been defined for speech improvement scheme. In this paper, optimal approaches have been explored for enhancement of voice. Gain estimation of various approaches is designed that reduces residual noise in the corrupted voice signal. So the main objective is to suppress or modify affected noise from real environment noise such as car, airport, and train noise. The algorithm is depended on simulation of objective and subjective evaluation. A software tool MATLAB is used to implement an algorithm for evaluating speech enhancement methods

Keywords: *single channel speech enhancement,MMSE based, additive noise model.*

1. Introduction

The speech signal is always corrupted by various types of noise like babble noise, white noise, colored noise which affected to the quality and intelligibility of the original speech signal in real environment. These degraded speech signal due to additive noise is a severe problem in listening for the human listener. So the main objective of speech enhancement is to minimize the effect of additive noise on speech signals in a real noisy environment, however done through a noise suppression algorithm.
In typical speech communication, speech signals can be enhanced at the transmitter side in according to increase the perceptual quality of speech at the receiver side. Such type of process may be useful in mobile communication, speech recognition and improving performance of aids for hearing impaired devices.

The increasing number of speech enhancement applications has outcome in more advantage for noise cancellation methods. Speech enhancement methods can be categorized into two main classes as single channel and second multichannel. In single channel, there is used only on microphone which is only one noisy source given to spectral information of speech. In the multichannel, there is used multiple microphone, leading more noisy source in real environment. Meanwhile, multi microphone may not be always available for measurement in real environment. Whereas, single microphone can be easier to measure and available in real environment. So focus here single channel speech enhancement methods.

We have to propose here Single channel speech enhancement methods based on short time spectral amplitude (STSA).STSA methods are based on short time Fourier transforms. They have one of the most powerful estimation techniques and extensively used for speech enhancement application. Noisy amplitude must be reduced from observed speech amplitude and get an estimate of the clean speech amplitude; it is called short time spectral amplitude (STSA). STSA methods have simulated and performance comparison based on spectrogram analysis is included after necessary descriptions.

Additive noise model

Firstly, we assumed that an additive noise model. An additive noise model is that observed speech signal is equal to clean speech and noisy signal as shown in fig1 and given by:

$$A(n) = X(n) + D(n)$$

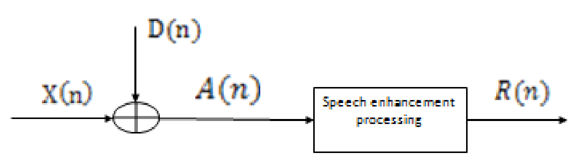


Fig. 1 Additive noise removal

Here, $A(n)$ is observed speech, $X(n)$ and $D(n)$ denote clean speech and noisy signal., we have to be decomposed into magnitude and phase using the STFT as according to figure 2.

$$|A_k|\exp\theta_k = |X_k|\exp\varphi_k + |D_k|\exp\varepsilon_k \quad (1)$$

Here k is frequency bin which conducts a significant portion of the speech energy and, $|X_k|$ and $|D_k|$ refer to the magnitude of observed ` 2qqqqqqspeech, clean speech and noisy signal and θ_k, φ_k, ε_k denote to corresponding phases. But perception of human speech is not sensitive to phase. So, here focus on magnitude of speech signal.
Gain function is estimated using priori and posteriori SNR. Enhanced spectral amplitude R_k is estimated by multiplying gain function with noisy speech amplitude. Enhanced speech magnitude is combined with phase using inverse DFT. Then, Enhanced speech signals have achieved using overlapped add and synthesis methods.

$$G(\mu_k, \gamma_k) = \frac{|R_k|}{|A_k|} \quad (2)$$

Gain function is related with ratio of estimate of clean speech amplitude to noisy speech amplitude. But For calculating gain function, It is also defined priori and posteriori SNR.
Priori SNR $\mu_k = \frac{|R_k|^2}{|D_k|^2}$ and Posteriori SNR $\gamma_k = \frac{|A_k|^2}{|D_k|^2}$
Posteriori SNR is based on $|A_k|^2$ noisy speech and noisy signal $|D_k|^2$obtained noise reduction algorithm. But priori SNR is not available because noisy speech variance is not known after determining VAD. So Ephraim and malah proposed the decision directed method is given as a function of priori SNR given as

$$\mu_k = \sigma\frac{|R^{(t-1)}{}_k|^2}{|D^{(t)}{}_k|^2} + (1-\sigma)\max(\gamma^{(t)}{}_k - 1, 0 \quad (3)$$

Where $0 \le \sigma \le 1$ is smoothing factor and t is the frame index.

2. MMSE STSA based Methods

STSA based methods are less computation complexity and easy to implement. Noise signal is assumed as additive white Gaussian noise and stationary. They change slowly in comparison with the speech. STSA based methods consist of Short time Fourier analysis, noise estimation (VAD), Gain estimation, short time synthesis as according to fig 2. Magnitude of Noisy speech is applied to noise estimation (VAD). Voice activation detector (VAD) is algorithm to detect active and inactive region in a noisy signal. Here, Gain function is used to modify magnitude of noisy speech and to enhance speech signal. Then, they reconstruct magnitude and phase of speech signal using overlap add synthesis method.

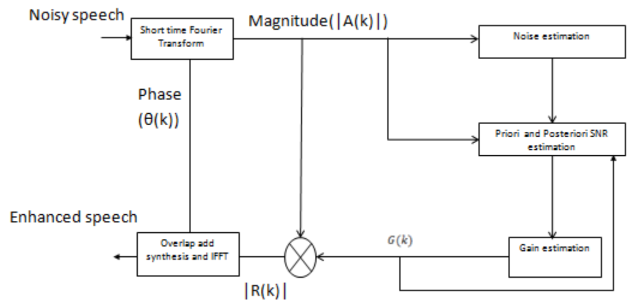


Fig. 2 STSA based statistical models

1. Minimum Mean Square Error (MMSE) Log spectral Amplitude Estimation

MMSE log spectral amplitude (MMSE-LSA) estimator can be obtained by taking the conditional expected value of log of clean speech amplitude. The method minimizes the mean square error between the log of clean speech amplitude and estimate speech amplitude $E[(\log X_k - R_k)^2]$. Optimal MMSE log spectral amplitude estimation is given by the equation

$$G(\mu_k, \gamma_k) = \frac{\mu_k}{1+\mu_k}\exp\left(-\frac{1}{2}\int_{z_k}^{\infty}\frac{e^{-t}}{t}dt\right) \quad (4)$$

The integral in before equation is an exponential integral and can be computed numerically. This method reduces the
remaining noise without affecting the original speech signal. The exponential integral, $Ei(x)$ can be given as below.

$$Ei(x) = \int_x^{\infty}\frac{e^{-x}}{x}dx \approx \frac{e^x}{x}\sum_k\frac{k!}{x^k}$$

2. Minimum Mean Square Error (MMSE) spectral Amplitude Estimation

MMSE-Spectral amplitude is one of best estimation method for estimating clean speech amplitude from corrupted speech signal amplitude. It is statistical model that a distortion measure by mean square error of spectral amplitude of clean speech and estimate speech. $E[(X_k - R_k)^2]$.Gain function of MMSE-spectral amplitude is given by the equation

$$G(\mu_k, \gamma_k) = \frac{\sqrt{\pi}}{2}\frac{\sqrt{z_k}}{\gamma_k}\exp^{-\frac{v_k}{2}}\left[(1+z_k)I_0\left(\frac{z_k}{2}\right) + v_k I_1(\right. \quad (5)$$

Where,

$$z_k = \frac{\mu_k}{1+\mu_k}\gamma_k$$

Where I_0 and I_1denote the modified Bessel functions of zero order and first order. It is assumed that the statistical model of noise is Gaussian model and model is statistical independent zero mean Gaussian random variable. Gain estimation is calculated from estimating priori and Posteriori SNR according to fig 2. Enhanced speech has given colorless remaining noise. The speech distortion has less compared to Wiener filter

3. Generalized gamma based Minimum Mean Square Error (MMSE-GG)

Generalized gamma distribution is used for determining estimator instead of Gaussian distribution.. The use of GGDs allows the optimal estimator to be determined in a generalized form.The optimal estimator is used thid distribution for noise prior as well as speech prior.

$$G(\mu_k, \gamma_k) = \phi_k[\frac{1}{\gamma_k} exp(-\frac{\gamma_k(1+\xi_k^2)}{4\xi_k(1+\xi_k)}) sinh(\frac{\gamma_k(\xi_k-1)}{4\xi_k})]$$

Where

$$\phi_k = \frac{\xi_k}{1+\xi_k}$$

3. Prapose workflow and implemantation

For evaluating of speech enhancement methods, input signal considering as speech signal is taking from NOIZEUS database. NOIZEUS database is a noise speech corpus consists of a speech database and one of application for assessing performance of speech enhancement methods.

The clean speech signal has sampled at 8 kHz and is applied with SNR in the range of 5 dB. AWGN is added into the clean speech signal. The STFT is implemented with Hamming window for 25ms.Here; a frame size of 25ms with 40% is used. Then Voice activity detection is determined noise spectrum.VAD is used to separate active and inactive (noise) speech signal. A Gain of MMSE estimation is evaluated with decision directed priori SNR. And lastly reconstructs enhanced speech signal use of overlap add and inverse DFT.

Spectrogram of both original speech and noisy speech is shown in fig.4. As it can be seen there is present a background noise in the noisy speech. Different quality measurement techniques are available like subjective and objective tests. But In this study spectrogram analysis is done.

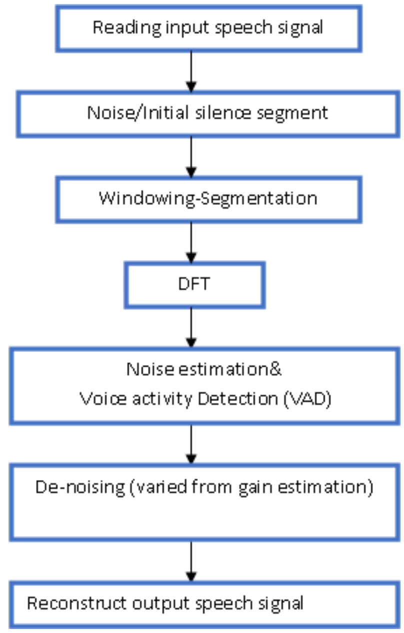
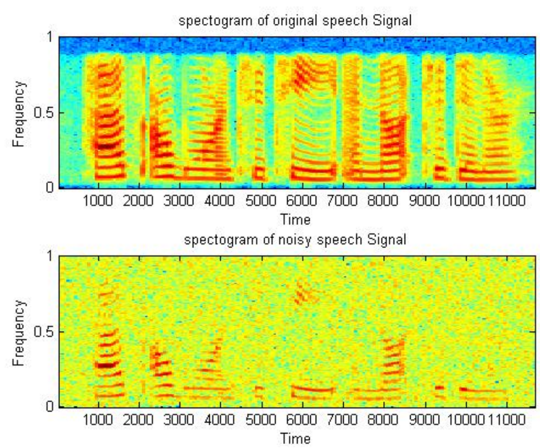


Fig. 3 Implementation of STSA algorithms

The spectrogram result of proposed, discussed method is shown in above figure. The spectrogram is a visual representation of an acoustic speech signal and describes speech signal's relative energy. Color is represented energy at a particular time and frequency. Red to blue color is high and low energy. Wiener based approaches removes the background noise, but remain speech distortion and some portion of speech attenuates.Whereas;In MMSE-LSA method, the speech distortion is less and speech is a high accuracy compared to Wiener filter.MMSE-LSA is given the best performance compare to other methods as according to the spectrogram.



IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 2 Issue 6, June 2015.
www.ijiset.com
ISSN 2348 – 7968

Fig 4 Spectrogram of original speech signal and corrupted speech signal.

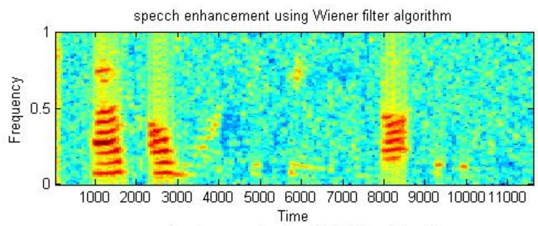


Fig 5 Spectrogram of output speech based on wiener filter method

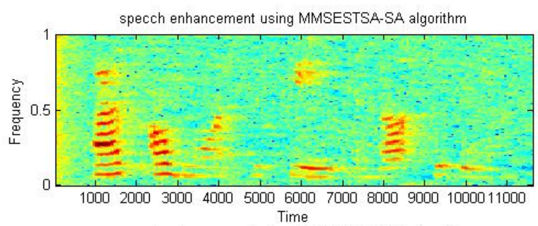


Fig 7Spectrogram of output speech based on MMSE-SA method

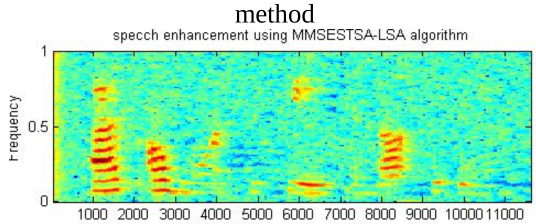


Fig 8 Spectrogram of output speech based on MMSE-LSA method

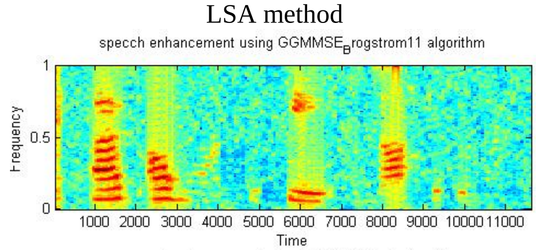


Fig 8 Spectrogram of output speech based on MMSE-GG method

STSA algorithms are evaluated using objective measures SSNR. The evaluation is done using NOIZEUS database. The measures have been observed over 0-10dB range of SNRs with different colored noises included in NOIZEUS database. The objective quality measures SSNR observed over 0-10 dB SNRs using NOIZEUS database are given following table.

NOISE	MMSE-SA	MMSE-LSA	MMSE-GG
BABBLE NOISE -0db	-3.2938	-2.6609	-1.7098
BABBLE NOISE -5db	1.0154	1.1069	-1.708
BABBLE NOISE -10db	0.2845	0.8869	1.1361
CAR NOISE-10db	1.0084	1.7412	2.3322
CAR NOISE-5db	-0.647	-0.0036	0.8788

4. Conclusion

In this paper, we have presented MMSE based various approaches on a single channel for speech processing application. The proposed system is implemented and evaluated using MATLAB simulation tools. The main purpose of speech enhancement is to improve the perceptual aspects of speech like quality, intelligibility or degree of listener fatigue. So, these methods are to improve quality and intelligibity and increasing performance of human perception and speech production system.using objective test in term of SSNR, MMSE-GG provides best performance compare to other algorithm.

REFERENCES

1. R.C. Hendriks, J.S. Erkelens, J. Jensen and R. Heusdens,"Minimum mean-square error amplitude Estimators for speech enhancement under Generalized Gamma distributed." IWAENC 2006
2. The NOIZEUS database Available: http://ecs.utdallas.edu/loizou/speech/noizeus
3. John R. Deller, Jr., John H. L. Hansen, John G. Proakis,"Discrete-Time Processing of Speech Signals" by A Johnwiley & Sons, inc., Publcation.
4. Philipos C. Loizou "Speech Enhancement Theory and Practice" by.CRC press 2007
5. Y. Ephraim and D. Malah,"Speech enhancemnt using a minimum mean-square error short-time spectral amplitude estimator,"IEEE Trans.vol. ASSP-32,no. 6,pp.1109-1121,Dec.1984

ISSN 2348 – 7968

6. 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,Apr. 1985
7. Yoshihisa Uemura, Yu Takahashi,Hiroshi Saruwatari,Kiyohiro Shikano, and Kazunobu Kondo,"Musical noise generation analysis for noise reduction methods based on spectral substraction and MMSE STSA estimation",icasssp 2009
8. R. J. McAulay and M. L. Malpass, "Speech enhancement using a soft-decision noise suppression filter," IEEE Trans Acoust., Speech, Signal Processing, vol. ASSP-28, no. 2, pp. 137–145, April 1980.
9. J. Benesty,"Springer series on signals and communication technology Speech enhancement"
10. Yi Hu and Philipos C. Loizou,"Evaluation of objective Measures for Speech enhancement"IEEE Trans. Audio,speech and Language processing, Jan 2008