IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)
e-ISSN: 2278-2834,p- ISSN: 2278-8735.Volume 9, Issue 2, Ver. III (Mar - Apr. 2014), PP 37-41
www.iosrjournals.org

Single channel speech enhancement techniques for removal of additive noise

Nainesh B Patel[1], Prof. Hardik N Patel[2]
[1]*(EC Department, Gujarat Technological, University, India)*
[2]*(EC Department, Gujarat Technological University, India)*

Abstract : *In addition to the known applications of voice communication , cell phones , hands-free voice recording , automatic speech recognition , interactive voice response systems , man-machine interface , requires at least a microphone and signals are usually corrupted by noise reverberation and background noise . So , the voice signal must be " clean " by digital signal processing tool to be played before storage or transmission . This article describes an speech enhancement technology for single voice channel. Various strategies have been proposed and are described from human to simulate and compare the past in this article . But there are still some obstacles to overcome these methods . Therefore, the aim is to modify or in combination with a single voice channel from the real world , a significant noise, such as airports , cars , restaurants , train stations and other inhibiting characteristics of different algorithms voices improvement efforts improvisation to outperform traditional methods , because the failure is made to develop a new algorithm. The algorithm is based on simulation of objective and subjective evaluation.*

Keywords: *Speech Enhancement ,single channel ,Background noise, noise reverberation*

I. INTRODUCTION

The main purpose is to enhance the voice of one or more improved perception of speech , such as overall quality ,clarity or degree of listener fatigue . Voice processing systems are typically designed noiseless environment , the background in the real world environment, the presence of noise is inevitable . Speech Enhancement different background noise removal algorithm has been applied to problems such as eliminating reverb and multi- voice separator (speaker separation) in modern telephony systems .. various speech enhancement methods have been proposed by researchers over the years . Limitations of these methods is still a considerable challenge , researchers in this area.

Speech enhancement techniques can be divided into two basic categories: (i) based on single-channel voice microphones were obtained from a single source or multiple microphones and (ii) a plurality of channels (array processing). However , single-channel (a microphone) signal can be used to measure or pick up in the real environment, so the focus here on the single-channel speech enhancement methods.

Most single -channel speech enhancement technology is based on transform domain approach. In a very short time discrete Fourier transform (STDFT) as currently used in the majority of technological change tools . They are called short-time spectral amplitude (STSA) method , and the visibility and noise reduction has been a good one plus research techniques . These methods are discussed in the following sections . Faced with the problems of these methods are discussed . STSA simulate various methods , and after the necessary instructions , the performance of spectral analysis based on the comparison is included. Is further scope in this respect also proposed.

II. STSA BASED ADDITIVE NOISE REMOVAL METHODS

These methods are based on the analysis - Modify - synthetic methods. They are based on the fact that human speech perception is not sensitive to the phase spectrum but must be clean and properly extract the amplitude spectrum from the noisy speech to speech output with acceptable quality , and therefore they are called short-time spectral amplitude (STSA) based method . Voice of the phase noise is kept in enhanced voice

STSA based approaches assume that noise is additive white noise and stationary for a frame and changes slowly in comparison with the speech. Most real environmental noise sources such as vehicles, street noise, babble noise etc. are non-stationary and coloured in nature. Therefore complete noise cancellation is more complex as it is not possible to completely track such noises. However, using this assumption it is possible to achieve significant reduction in the background noise levels using simple techniques. The noise statistics are typically characterized during voice-inactivity regions between speech pauses using a voice activity detector (VAD). The VAD always becomes an integral part of any STSA based algorithm.

A. Spectral subtractive algorithm

Spectral subtraction method was first proposed by S.F.Boll [3]. The basic principle of spectral subtraction is to subtract an estimate of the average noise spectrum from noisy speech magnitude spectrum. Degraded speech signal is modelled as

$$y(n) = x(n) + d(n) \tag{1}$$

Taking DFT of (1) gives

$$Y(\omega) = X(\omega) + D(\omega) \tag{2}$$

The estimate of $D(\omega)$ is obtained by using VAD and updated during non-speech or silence periods. For good initial estimate it requires initial silence period of around 0.2 seconds.

B. Magnitude spectral subtraction (MSS)

From equation 2 taking only magnitude of spectrum we can write

$$|X(\omega)| = \begin{cases} |Y(\omega)| - |\hat{D}(\omega)| \; if \; |Y(\omega)| > |\hat{D}(\omega)| \\ 0 \; else \end{cases} \tag{3}$$

Hence, original speech estimate is given by

$$\hat{X}(\omega) = [|Y(\omega)| - |\hat{D}(\omega)|]e^{j\phi_y(\omega)} \tag{4}$$

The half wave rectification process is only one of many ways of ensuring non-negative $|\hat{X}(\omega)|$. This summarizes the spectral subtraction process. Compute the magnitude spectrum of the noisy speech via the FFT and keep an estimate of the noise spectrum when speech is not present. Subtract the noise magnitude spectrum from the noisy speech magnitude spectrum and finally, take the inverse Fourier transform of the difference spectra to produce the enhanced speech signal.

C. Power spectral subtraction (PSS)

The preceding discussion of magnitude spectrum subtraction can be extended to power spectrum domain as

$$|\hat{X}(\omega)|^2 = \begin{cases} |Y(\omega)|^2 - |\hat{D}(\omega)|^2 \; if \; |Y(\omega)| > |\hat{D}(\omega)| \\ 0 \; else \end{cases} \tag{5}$$

The spectral power subtraction can be generalized with an arbitrary spectral order, called generalized spectral subtraction (GSS), as,

$$|\hat{X}(\omega)|^p = \begin{cases} |Y(\omega)|^p - |\hat{D}(\omega)|^p \; if \; |Y(\omega)| > |\hat{D}(\omega)| \\ 0 \; else \end{cases} \tag{6}$$

D. Berouti spectral subtraction (BSS)

The major problem of the basic spectral subtraction is that, the algorithm may itself introduce a synthetic noise, called musical noise. The half wave rectification is non-linear process and it creates small, isolated peaks in the spectrum occurring at random frequency locations in each frame. In time domain these peaks result in tones with randomly changing frequency from frame to frame. This musical noise is more disturbing to the listener than the original noise. Most researchers suggest that it is difficult to minimize musical noise without affecting the speech signal. So there is always a trade-off between the amount of noise reduction and speech distortion.

Berouti et al.[8] proposed an important variation of the original method, which improves the noise reduction compare to the basic spectral subtraction. It introduces an over subtraction factor(α) and spectral floor parameter (β) and it is defined as

$$|\hat{X}(\omega)|^2 = \begin{cases} |Y(\omega)|^2 - \alpha|\hat{D}(\omega)|^2 \; if \; |Y(\omega)|^2 > (\alpha + \beta)|\hat{D}(\omega)|^2 \\ \beta|\hat{D}(\omega)|^2 \; else \end{cases} \tag{7}$$

The parameter β controls the amount of remaining residual noise and the amount of perceived musical noise. Large β produces audible residual noise but small musical noise and vice versa. The parameter α affects the amount of speech spectral distortion caused by the subtraction in equation (7). Large value of α produces high speech distortion and vice versa. The value of α should vary linearly with SNR in dB on per frame basis as

$$\alpha = \alpha_0 - \frac{3}{20} \times (\mathrm{SNR})$$

Where α_0 is the value of α at SNR=0 dB, and SNR is estimated frame SNR in dB. The optimized value of $\alpha_0$is between 3 to 6 and that of β is in the range of 0.02 to 0.06 for SNR$\leq 0dB$ and in the range of 0.005 to 0.02 for SNR$> 0dB$. Though usage of over subtraction of the noise spectrum and the introduction of a spectral floor serve to minimize residual noise and musical noise, musical noise is not completely avoided.

III. Statistical Model Based Methods

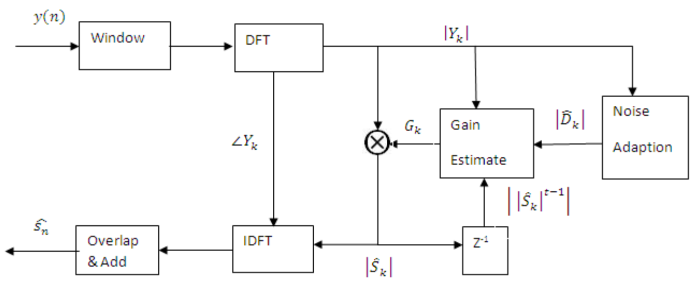

Fig.1 STATISTICAL MODEL

In this method gain is estimated by noise adaption. The noisy speech $\mathrm{y(n)}$,is first converted into STSA, $|\mathrm{Y_k}|$ by a DFT with windowing . the enhanced sprectral amplitude $|\hat{\mathrm{S}}_{\mathrm{k}}|$ is estimated by multiplying the noisy signal spectral components $\mathrm{Y_k}$ with their corresponding estimated gain $\mathrm{G_k}$. Enhanced speech $\widehat{s_n}$, is then reconstructed by applying the inverse DFT to enhance STSA, $|\hat{\mathrm{S}}_{\mathrm{k}}|$,with the noisy speech phase followed by an overlap-add procedure to compensate for the window effect and to alleviate abrupt signal change between two consecutive frames. The most critical part of this process is estimation of the gain, $\mathrm{G_k}$.

From figure.1 enhanced spectrum $\widehat{s_k}$ can be written in terms of the modification factor $\mathrm{G_k}$ and the noisy speech spectrum $\mathrm{Y_k}$ as,

$\hat{\mathrm{S}}_{\mathrm{k}} = \mathrm{G_k Y_k}$, for $0 \leq \mathrm{G_k} \leq 1$

The gain $\mathrm{G_k}$ is a function of a posteriori SNR and a priori SNR.

$$\mathrm{SNR}_{\mathrm{post\ k}} = \frac{|\mathrm{Y_k}|^2}{|\hat{\mathrm{D}}_{\mathrm{k}}|^2}$$

$$\mathrm{SNR}_{\mathrm{prio\ k}} = \frac{|\hat{\mathrm{S}}_{\mathrm{k}}|^2}{|\hat{\mathrm{D}}_{\mathrm{k}}|^2}$$

The function definition of the gain $\mathrm{G_k}$ depends on specific enhancement methods. Based on the relation between gain function and a posteriori SNR , a priori SNR there are several method exist.

$\mathrm{SNR}_{\mathrm{post\ k}}$ can be obtained easily as $|\mathrm{Y_k}|^2$ and $|\hat{\mathrm{D}}_{\mathrm{k}}|^2$ can be obtained by using noise adaptation algorithm. But the speech variance $|\hat{\mathrm{S}}_{\mathrm{k}}|^2$ is not available. Hence, $\mathrm{SNR}_{\mathrm{prio\ k}}$ is not available. As a solution, Ephraim and malah [] proposed the decision directed method given by,

$$\mathrm{SNR}_{\mathrm{prio\ k}}(\mathrm{t}) = \alpha \frac{|\hat{\mathrm{S}}^{(\mathrm{t}-1)}|^2}{|\hat{\mathrm{D}}_{\mathrm{k}}^{(\mathrm{t})}|^2} + (1-\alpha)\mathrm{MAX}(\mathrm{SNR}_{\mathrm{post\ k}}^{(\mathrm{t})} - 1{,}0)$$

Where $0 \leq \alpha \leq 1$ and t is the frame index. Gain estimation can be done by various approaches which is discussed next.

MAXIMUM-LIKELIHOOD SPECTRAL AMPLITUDE ESTIMATION

Gain function based on Maximum-likelihood Spectral Amplitude Estimation can be given as,

$$\mathrm{G_k^{(ML)}} = \frac{1}{2} + \frac{1}{2}\sqrt{1 - \frac{1}{\mathrm{SNR}_{\mathrm{post\ k}}}}$$

WIENER FILTERING

The wiener filter is a minimum mean square error estimate of a desired signal in the time domain. The gain function can be given as,

$$\mathrm{G_k^{(WF)}} = \frac{\mathrm{SNR}_{\mathrm{prio\ k}}}{1 + \mathrm{SNR}_{\mathrm{prio\ k}}}$$

OBJECTIVE EVALUATION OF SPEECH ENHANCEMENT METHODS

Various STSA based speech enhancement methods like magnitude spectral subtraction, power spectral subtraction, berouti, ML based, wiener filter is already discussed. All this methods are implemented and Performance comparison among all this methods on spectrographic analysis is reported in this section.

IV. Simulation Results

The 8kHz sampled input speech signal is applied to the hamming window using a 25ms with a shift percentage of 40%(10ms) and 256 point FFT is used for all the methods. The noise spectrum is estimated during initial silence period 25ms, assuming speaker starts speaking after that.

A speech signal utterance "HELLO" is applied with fan sound as a background noise. AWGN is added into that. Spectrogram of both original speech and noisy speech is shown in fig.2. As it can be seen there is a constant background noise present in the noisy speech. There are several quality measurement techniques available like subjective and objective tests. In this report spectrogram analysis is used.

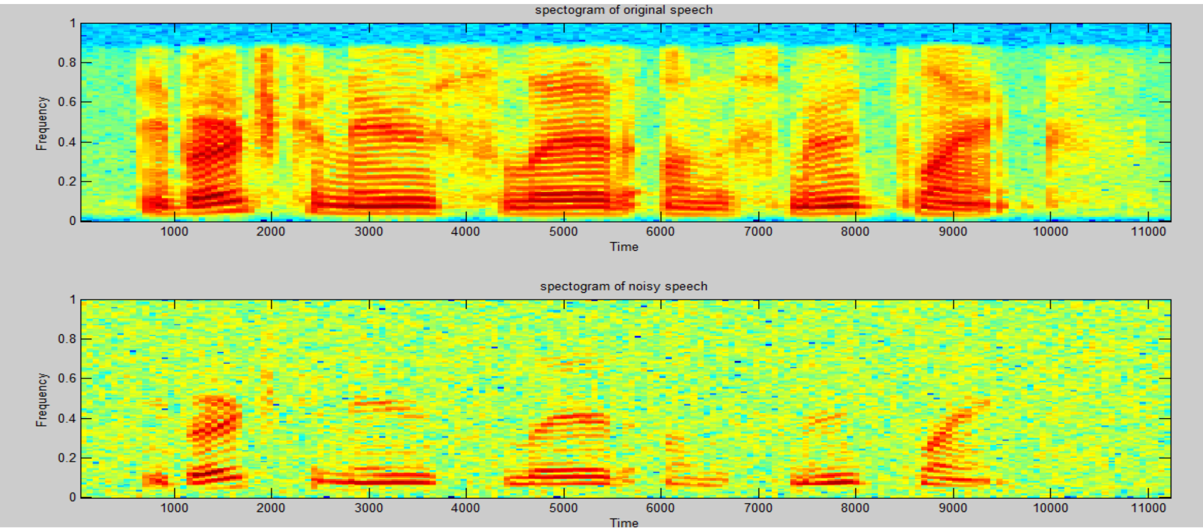

Fig.2

The comparison of above discussed method is shown in figure 3. As it was shown in result STSA based approaches removes the background noise but gives rise to musical noise. It can be stated as optimization problem. The optimization process requires a trade off between noise reduction and speech quality. Performance of wiener filter is outstanding among all them.

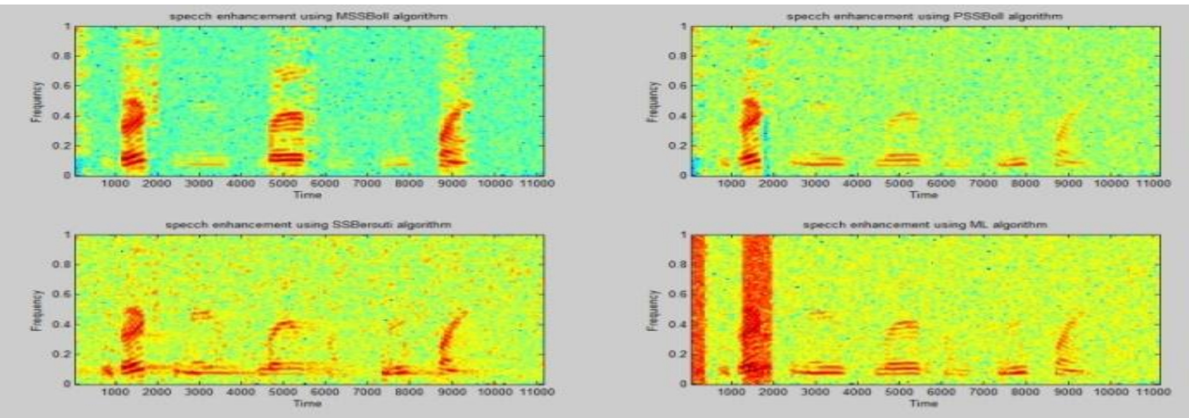

FIG.3

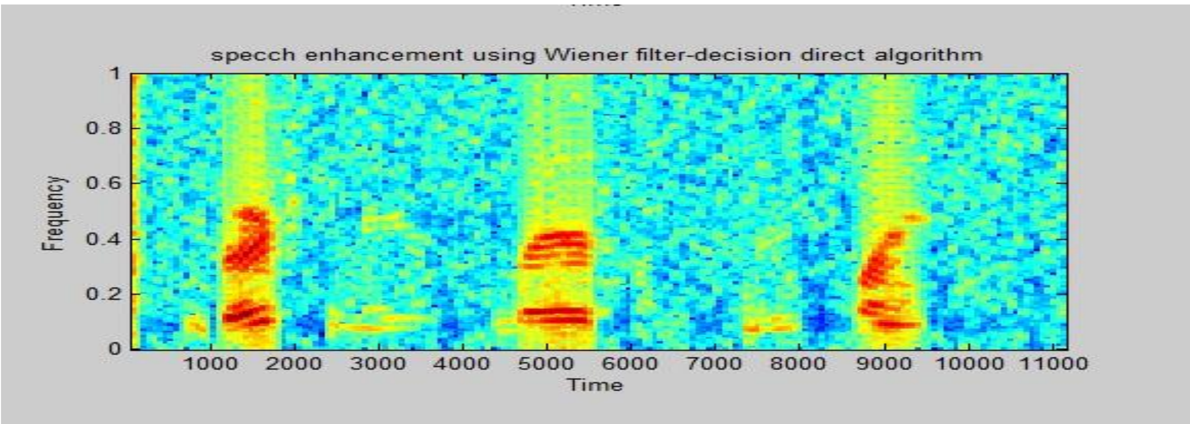

FIG.4

IV. CONCLUSION

In this paper, speech enhancement technology for single channel based on different methods have been given. Snapshots results show that the method has been implemented on matlab. Ideally there should be no degradation in the quality of the original speech and/or human subjects have normal speech production and perception systems and /or clarity, In reality there is degradation in the quality and/or intelligibility and/or damage to human subjects speech production and perception systems. Thus, given techniques to improve the quality and intelligibility.

ACKNOWLEDGEMENTS

I would like to extend my sincere thanks to all of them. I am highly indebted to ***Prof. Hardik Patel, PIET, Limda*** for her guidance and constant supervision as well as for providing necessary information regarding the project. I would like to thank IOSR journal for the support to develop this unique document.

REFERENCES

[1] S.K.Shah, J.H.Shah, N.N.Parmar , "Performance Evaluation of STSA based Speech Enhancement Techniques for Speech Communication Systems" NCWCVD-2010,IEEE MP subsection 2010.
[2] "The NOIZEUS database (2009)." Available: http://www.utdallas.edu/~loizou/speech/noize
[3] A.M.Kondoz, Digital Speech, 2nd Ed., Wiley India Pvt. Ltd., 2007.
[4] Thomas F. Quatieri, Discrete-time Speech Signal Processing, 1st Indian reprint, Pearson education signal processing series, 2004.
[5] Suppression of acoustic noise in speech using spectral subtraction, Steven F. Boll, IEEE transactions on acoustics, speech and signal processing , vol. ASSP-27,1979,page-113-119.
[6] N.Virag, "Single channel speech enhancement based on masking properties of the human auditory system," IEEE Trans. Speech and Audio processing, Vol. 7, pp. 126-37, Mar. 1999
[7] Mingyang Wu and DeLiang Wang, "A Two stage algorithm for enhancement of reverberant speech," Proc. IEEE ICASSP 2005, pp. 1085-88, 2005.
[8] RASTA processing of speech Hynek Hermansky. IEEE transactions on acoustics, speech and signal processing Vol.2 Oct.1994.
[9] M. Berouti "enhancement of speech corrupted by acoustic noise" ICASSP'79 vol.4
[10] Springer series on Signals and communication technology,speech enhancement by J. Benesty