INTERNATIONAL JOURNAL FOR ADVANCE RESEARCH IN ENGINEERING AND TECHNOLOGY

WINGS TO YOUR THOUGHTS…..

SPEECH ENHANCEMENT ALGORITHMS: A BRIEF REVIEW

Durgesh[1], Anil Garg[2], Pankaj Bactor[3]

[1] M.Tech Student MMEC, Mullana
M.M. University, Mullana-Ambala-Haryana, India
durgeshsiet786@gmail.com

[2] Asst. Prof. in ECE department MMEC, Mullana
M.M. University, Mullana-Ambala-Haryana, India
agarg001@yahoo.com

[3] Asst. Prof. in ECE department MMEC, Mullana
M.M. University, Mullana-Ambala-Haryana, India

Abstract: *In this paper, the aim of speech enhancement algorithms is to improve the quality or intelligibility of the noisy speech signals by using different enhancement algorithms. Many speech enhancement algorithms are designed to suppress additive background noise. This review paper presented the basic of spectral subtraction algorithm, minimum mean square error, wiener algorithm and TSDD algorithm, and performance evaluation of various modified decision-directed approach. This paper provides valuable hints for analyzing and optimizing noise-reduction algorithm. The speech enhancement methods aimed at suppressing the background noise are based on one way or the other on the estimation of the background noise. If the background noise is evolving more slowly than the speech, i.e., if the noise is more stationary than the speech, it is easy to estimate the noise during the pauses in speech. This paper also reports the subjective and objective tests.*

Keywords: *speech distortions, speech enhancement algorithm, and speech intelligibility improvement.*

1. INTRODUCTION

Speech enhancement is one of the most important topics in speech signal processing. Several techniques have been proposed for this purpose like the spectral subtraction approach, the signal subspace approach, adaptive noise canceling and Wiener filter. The aim of speech enhancement algorithm is to improve the quality of the speech signal [1]. These algorithms have been found to improve the speech quality [2]. From the all available speech enhancement methods, the spectral subtraction technique is one of the first algorithms proposed for background noise reduction. This is done by subtracting the average magnitude of noise signal from the noisy speech to estimate the magnitude of the enhanced speech signal [3].

The Spectral-subtractive algorithms, is the simplest speech enhancement algorithms to implement. They are based on the principle that noise is additive and the noise spectrum can be estimated in the absence of speech and are subtracted from the noisy signal. Statistical model based algorithm, this algorithms, given a set of measurements that is, to the Fourier transform coefficients of the noisy signal, a linear or nonlinear estimator of the parameter of interest, namely the transform coefficients of the clean signal is found. The Wiener algorithm and minimum mean square error (MMSE) algorithms come in this category. Subspace Algorithms, these algorithms are based on the principle the clean signals are not confined to the subspace of the noisy Euclidean space. Given a method of decomposing the vector space of the noisy signal into a subspace that is occupied primarily by the clean signal and a subspace that is occupied primarily by the noise signal, one could estimate the clean signal simply by nulling the component of the noisy vector residing in the noise subspace. These algorithms were evaluated using a developed noisy corpus best for evaluation of speech enhancement algorithm [4].

We have different type of distortion in speech enhancement algorithm can be divided into two categories: - distortion that affects the speech signals itself, and second is distortion that affects the background noise. From these two types of distortion, listeners assume to determine the speech distortion when making knowledge of overall quality [5]. In the previous study, for overall quality and speech distortion, algorithm MMSE, log MMSE, wiener filter performed good in some condition. Subspace algorithms performed poorly for overall quality [6]. In this paper, we address on the subjective and

INTERNATIONAL JOURNAL FOR ADVANCE RESEARCH IN ENGINEERING AND TECHNOLOGY
WINGS TO YOUR THOUGHTS…..

objective comparison and evaluation of all above algorithms using different methods.
The paper is organized as follows: section II gives overview of different methods or algorithms of speech enhancement, Section III presents the subjective and objective tests, and Section IV presents the performance of all above algorithms, and finally section IV present conclusions.

2. SPEECH ENHANCEMENT METHODS

There are various speech enhancement methods or algorithms proposed for noise reduction and to improve the noise quality and intelligibility.

A. SPECTRAL SUBTRACTION ALGORITHM:

Spectral subtraction is one of the first algorithms prefer for speech enhancement. It is simple and easy to implement, based on the principle that one can obtain an estimate of the clean signal spectrum by subtracting an estimate of the noise spectrum from the noisy speech spectrum. The noise spectrum can be estimated and updated, during the time interval when the signal is absent or when only noise is present. Assumption is noise is additive¸ its spectrum does not change with time means noise is stationary or it's slowly time varying signal¸ whose spectrum does not change significantly between the updating periods [7] [4]. With this approach, estimate the enhanced speech spectrum is obtained by subtracting an estimate of the noise spectrum from the noisy speech spectrum during the period when the speech signal is not present. The key advantage of this method of speech enhancement is that it is simple and easy to implement. The spectral subtraction algorithm effectively reduces the noise which is present in the corrupted speech signal [7]. The principle of spectral subtraction algorithm is shown in Fig. 1.

Let $y(n)$ be the noisy speech signal given by

$$z(n) = x(n) + w(n) \tag{1}$$

Where, $x(n)$ represents the clean speech signal and $w(n)$ is the uncorrelated additive noise. In spectral subtraction algorithm, it is assumed that the noise and clean signal are uncorrelated so as to estimate the noise spectrum. Initially, the spectral subtraction approach was used to estimate the short term magnitude spectrum of the clean signal $|X_k|$. This is done by subtracting the estimated noise magnitude spectrum $|\widehat{W}_k|$ from the noisy signal magnitude spectrum $|Z_k|$. The noisy signal phase spectrum is used as an estimate of the clean speech phase spectrum, as follows:

$$\hat{X}_k = \left(|Z_k| - |\widehat{W}_k|\right)\mathrm{e}^{j\varphi(y,k)} \tag{2}$$

Where, $\varphi(z,k)$ is the phase of noisy signal Y_k. The estimated time-domain clean speech signal is obtained by taking the inverse Fourier Transform of$\hat{X}_k$. However, this approach has several shortcomings. Therefore, another enhanced version of spectral subtraction algorithm is proposed, the clean signal $\hat{x}(n)$ is recovered from the noisy signal $z(n)$, by assuming that there is an estimate of the power spectrum of noise $|\widehat{W}_k|^2$, which is obtained by averaging over multiple frames of a known noise segment. An estimate of the short-time squared magnitude spectrum of the clean signal using this method can be obtained as follows:

$$|\hat{X}_k|^2 \begin{cases} |Z_k|^2 - |\widehat{W}_k|^2, & \text{if } |Z_k|^2 - |\widehat{W}_k|^2 \geq 0 \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

To recover the signal, the magnitude spectrum estimate is combined with the phase of the noisy signal as shown in Eqn. 4 and the
Clean speech can be obtained with the Inverse Fourier Transform.

$$\hat{x}(n) = |\hat{X}_k|\mathrm{e}^{j\varphi(y,k)} \tag{4}$$

Although the spectral subtraction algorithm can be easily implemented; yet, it has several shortcomings. The subtraction process needs to be done carefully to avoid any speech distortion. If too little is subtracted, much of the interfering noise remains¸ but if too much is subtracted, then some speech information might be removed [4].

B. Wiener–type filtering algorithm:

This algorithm proposed for noise reduction. Principle lies to obtain an estimate of clean signal from that corrupted by additive noise [4]. This estimate is obtained by minimizing the mean square error between the desired signal and the estimated signal. In this the input signal goes through a LTI system to produce an output signal z (n). We design this in such a way that the output signal $\hat{d}(n)$ is close to desired signal d(n).This can be done by computing the estimation error, the optimal filter that minimize the estimation error is called the wiener filter.
A linear discrete-time filter for estimating a desired signal z(n) based on an excitation x(n).We assume that both x(n) and d(n) are random processes. The filter output is $\hat{d}(n)$ and e (n) is the estimation error.

INTERNATIONAL JOURNAL FOR ADVANCE RESEARCH IN ENGINEERING AND TECHNOLOGY
WINGS TO YOUR THOUGHTS…..

Performance function is defined as

$$£=E\ [|e(n)|^2] \tag{5}$$

Where e (n) = d (n)-z (n)
The Wiener filter is used to reduce the amount of noise presented in a signal by comparison with an estimation of the desired noiseless signal. As can be observed in the results, the image restoration is not absolutely perfect but it achieves a very close image to the original one [4].
The goal of the Wiener filter is to filter out noise that has corrupted a signal. It is based on a statistical approach, and a more statistical account of the theory is given in the MMSE estimator article.

C. MMSE Estimator
MMSE estimation is also known as Ephraim and Malah's estimator [in 1984]. MMSE estimation producing colourless residual noise, this is the advantage of this method [8]. To overcome the problem of the musical noise distortion present in the above algorithm method, Ephraim and Malah, proposed the MMSE method which reduces the distracting musical noise to a considerable extent, and thus improved the quality of the resulting enhanced speech. Mainly MMSE based algorithms are Minimum Mean Square Error Short-Time Spectral Amplitude (MMSE-STSA) estimator and MMSE Logarithm Spectral Amplitude (MMSE-LSA) estimator. In various MMSE of power spectrum have been proposed. Some power spectrum estimator in decision-directed approach used for the calculation of a priori SNR [9].
The aim of MMSE-STSA method to minimize the mean square error between the short-time spectral magnitude of the clean and enhanced speech signal. In the previous, wiener method can derived by minimizing the error between a linear model of clean spectrum and real spectrum [4]. The MMSE-STSA method gives good results in reducing the musical noise; however, it suffers a drawback of not taking into consideration the non-linear characteristics observable in human perception. Therefore, MMSE-LSA enhancement method was proposed to minimize the mean square error between the logarithm of the STSA of the clean and enhanced speech. The MMSE-LSA is often favored because of its psychoacoustic considerations and provides a better quality of the enhanced speech.

D. TSDD ALGORITHM:
For improve the performance of a gain factor for noise reduction, a perceptual-decision directed approach, is proposed. Initially the decision-directed method (Ephraim and Malah), is performed to enhance a noisy and corrupted speech signal. Whereas, the decision-directed method is more suitable, to reduce the effect of musical residual noise. Therefore, a decision-directed method is performed again to improve the estimated a priori SNR by removing the frame relay. These procedures specify a two-step-decision directed approach algorithm. The drawback of decision-directed approach delay inherent in speech transients. This delay version of gain factor will generate a repetition. To compensate this we use the TSDD algorithm to improve the estimate of a priori SNR [10].
The gain factor of TSDD algorithm is given as:

$$g_{TSDD(m,w)}=\frac{g_{DD*}y_{post}(m,w)}{1+g_{DD*}y_{post}(m,w)} \tag{6}$$

Where $y_{post}(m,w)$ is the posteriori SNR, and g_{DD} is the gain factor used to estimate a priori SNR.

3. SUBJECTIVE AND OBJECTIVE *EVALUATION*

This review paper shows the previous result from the comparative analysis of the subjective and objective tests. Enhanced speech files were sent to Dynastat, Inc for subjective evaluation for evaluating noise suppression algorithm based on ITU-T p.835 [6].

[a] Subjective testing- include the methods which focused on speech intelligibility and overall quality. The subjective test include, the goodness test, Mean opinion score tests [MOS]. Another test which evaluates the speech and background signal quality across the multiple scales is diagnostic acceptability measure [10]. Subjective tests were arranged according to ITU-T P.835 methodology. In terms to assess perceived quality, a subjective mean opinion score test be performed, this test allows for overall quality.

[i] Test methodology
This method informs the listener to respectively attend to and rate the enhanced speech signal on [6].
A. SIG
Speech signal alone using a five-point scale of signal distortion (Table 1).
B. BAK
Background noise alone is using a five-point scale of background intrusiveness (BAK) (Table 2).
C.OVRL
Overall effect using the scale of the (MOS) – [1=

INTERNATIONAL JOURNAL FOR ADVANCE RESEARCH IN ENGINEERING AND TECHNOLOGY
WINGS TO YOUR THOUGHTS…..

bad, 2= poor, 3= fair, 4= good, 5= excellent]

Table 1(SIG) Speech signal
5 - Very natural, no degradation
4 - Fairly natural, little degradation
3- Somewhat natural, somewhat degraded
2- Fairly unnatural, fairly degraded
1- Very unnatural, very degraded

Table 2(BAK) Background noise)
5- Not noticeable
4- Somewhat noticeable
3- Noticeable but not intrusive
2- Fairly conspicuous, somewhat intrusive
1- Very conspicuous, very intrusive

[b] Objective test - depend on mathematically based measure between original and degraded speech.

[i] Ltakura-saito measure: - The Itakura-Satio distance is a measure of the perceptual difference between an original spectrum and an approximation of that spectrum. The distortion measure is given by,

$$d_{IS}(a_{d,}a_{\emptyset}) = \left[\frac{\sigma_{\emptyset}^2}{\sigma_d^2}\right]\left[\frac{a_d\ R_{\emptyset}\ a_d^T}{a_{\emptyset}\ R_{\emptyset}\ a_{\emptyset}^T}\right] + log\left(\frac{\sigma_d^2}{\sigma_{\emptyset}^2}\right) - 1 \quad (7)$$

Where σ_d^2 is speech with linear prediction coefficient vector and $\sigma_{\emptyset}^2$ is processed speech coefficient vector which represents the all-pole gains for processed and clean speech.

[ii] Log-Likelihood Ratio Measure: - Likelihood ratio test is used to compare the fit of two models one of which is nested within the other. The LLR measure is also referred to as takura distance. The LLR measure will be: -

$$d_{LLR}(a_{d,}a_{\emptyset}) = \log\left(\frac{a_d\ R_{\emptyset}\ a_d^T}{a_{\emptyset}\ R_{\emptyset}\ a_{\emptyset}^T}\right) \quad (8)$$

[iii] Segmental SNR Measure: - Correlation of SNR with subjective quality is not good. The time-domain segmental SNR measure was computed as: -

$$d_{SEGSNR} = \frac{10}{M}\sum_{m=0}^{M-1} log\frac{\sum_{n=Nm}^{Nm+N-1} s_{\phi}^2(n)}{\sum_{n=Nm}^{Nm+N}\left[s_d(n) - s_{\phi}(n)\right]^2} \quad (9)$$

Where $s_d(n)$ is the input signal, $s_{\phi}(n)$ is the enhanced signal, N is the frame length, and M is the number of frames in the signal. Only frames with SNRseg in the range of [-10, 35] dB were consider in the computation of the average [10].

4. PERFORMANCE OF ALL ALGORITHMS

The performance of spectral subtraction, MMSE, and wiener algorithm is good and also improve the quality of speech signal and TSDD gives the much better result than these entire algorithms it improves the quality as well as intelligibility of the signal. Noise estimation algorithms were assessed using both objective and subjective measures. The algorithms that performed the best in terms of low speech distortion were also the algorithms yielding the highest overall quality. This suggests that listeners were affected for the most part by the distortion imparted on the speech signal than on the background noise when making knowledge of overall quality [6]. In the two step decision-directed algorithm, the decision-directed algorithm is utilized to estimate the priori SNR. In turn, the estimated a priori SNR is refined again by the TSDD algorithm [11]. The spectra of enhanced speech are obtained by multiplying the spectra of noisy speech with this perceptual gain factor.

5. CONCLUSION

This is a review paper and various approaches are used in this. The subjective evaluation, in terms of overall quality and speech distortion of speech, the algorithm performed best are: logMMSE, MMSE-SPU, pMMSE and MMSE-ne. Wiener algorithm also performed well in some cases. In the case of speech with weak energy, the values of gain factor for the perceptual and the proposed method are larger than those of the TSDD method. TSDD algorithm employs the Wiener filter twice to estimate the spectra of speech. Two step decision-directed approaches are better able to reduce greater amount of residual noise than the perceptual algorithm. Evaluation of TSDD algorithms revealed that these algorithms improve speech quality and also improve the intelligibility of speech signals. TSDD also improve the performance of perceptual method approach in removing maximum residual noise.

REFERENCES

[1] **Hong-yan Li, Qing Zhao, Guang-ling Ren**

INTERNATIONAL JOURNAL FOR ADVANCE RESEARCH IN ENGINEERING AND TECHNOLOGY
WINGS TO YOUR THOUGHTS…..
and Bao-jin Xiao, "Speech enhancement algorithm based on independent component analysis", ICNC, vol.76, pp.598-602, 2009.

[2] Gibak Kim and Philipos C. Loizou, "Gain-induced speech distortion and the absence of intelligibility benefit with existing noise-reduction algorithm", J. Acoust. Soc. Am, vol.103, pp. 1501-1595, September 2011.

[3] P. Krishnamoorthy and S.R. M. Prasanna, "Enhancement of noisy speech by temporal and spectral processing", ScienceDirect vol.53, vol.154-174, 2011.

[4] P.C. Loizou, "Speech Enhancement: Theory and Practice (CRC Press, Taylor Francis Group, Florida)," pp. 1-394, 2007.

[5] Yi Hu and C. Loizou, "Evaluation of objective quality measure for speech enhancement", IEEE trans. Audio and language processing, vol.6, no.1, January 2008.

[6] Yi Hu and Philipos C. Loizou, "Subjective comparison and evaluation of speech enhancement algorithm", ScienceDirect, Speech Communication, vol.49, pp.588-601, 2007.

[7] Yang Lu and Philipos C. Loizou, "A geometric approach to spectral subtraction", ScienceDirect, Speech Communication,vol 50,pp.453-466, January 2008.

[8] Teddy Surya Gunwan, Eliathamby Ambikairajah and Julien Epps, "Perceptual speech enhancement exploiting temporal masking properties of human auditory system", ScienceDirect, Speech Communicationvol.52, pp.381-393, 2010.

[9] Yang Lu and Philipos C. Loizou, "Estimator of the magnitude-squared spectrum and method for incorporating SNR uncertainty", IEEE trans. Audio and lang. processing, vol.19, no.5, pp.1123-1137, 2011.

[10] John H.L.Hansen and Bryan L. Pellom, "An objective quality evaluation protocol for speech enhancement algorithm", The international conference on speech and language processing processing, 1998.

[11] Ching-Ta Lu, "Enhancement of single channel speech using perceptual-decision-directed approach", Science Direct, Speech Communication, vol.53, pp.495-507, December 2010.