ISSN 2249 5460 Available online at www.internationaleJournals.com

INTERNATIONAL JOURNAL OF MATHEMATICAL SCIENCES, TECHNOLOGY AND HUMANITIES

www.internationalejournals.com

International eJournals

International Journal of Mathematical Sciences, Technology and Humanities 95 (2013) 1056 – 1062

Different Approaches of Spectral Subtraction Method for Speech Enhancement

Lalchhandami[1] and Rajat Gupta [2]

[1,2]Department of Electronics & Communication Engineering
Maharishi Markandeshwar University, Mullana (Ambala), INDIA

ABSTRACT:

Enhancement of speech signal degraded by several types of noise is a topic of interest for last many years. The main aim of speech enhancement algorithm is to improve the quality and/or intelligibility of the noisy speech signals by using various techniques and algorithms. Among the all available methods, the spectral subtraction algorithm is the historically one of the first algorithm proposed for removing additive background noise. This paper presented the review of basic spectral subtraction algorithm, a short coming of basic spectral subtraction algorithm, different modified approaches of Spectral Subtraction Algorithms such as Spectral Subtraction with over subtraction factor, Non linear Spectral Subtraction, Multiband Spectral Subtraction, Minimum mean square Error Spectral Subtraction and Selective Spectral Subtraction that minimizes the shortcomings of the basic method, then performance evaluation of various modified spectral subtraction algorithms.

Keywords: Speech enhancement, Spectral subtraction, musical noise, SNR, Discrete Fourier Transform

I. INTRODUCTION:

Speech can be expressed as a mechanism of expressing thoughts and ideas using vocal sounds. Voice frequency normally ranges between 30 Hz to 3 KHz, depending upon individuals. However, the human ear can perceive sounds, with frequencies in between 20 Hz to 20 KHz approximately. As the noise produced by various ambient sources such as vehicles normally lies in this frequency range, speech signals get easily distorted by the ambient noises or AWGN. This make the listening task difficult for a direct listener, gives poor performance in automatic speech processing tasks like speech recognition speaker identification, hearing aids, speech coders etc. The degraded speech therefore needs to be processed for the enhancement of speech components. The aim of speech enhancement is to improve the quality and intelligibility of degraded speech signal. Among the all available speech enhancement methods, the spectral subtraction technique is historically one of the first algorithms proposed for background noise reduction. It is a single channel speech enhancement technique for the enhancement of speech degraded by additive background noise. Background noise can be a nuisance a conversation in a noisy environment like in

streets or in a car, and in telephone conversation and can affect both quality and intelligibility of speech. In this paper, a review of speech enhancement method using basic spectral subtraction and modified versions of spectral subtraction has been explained in detail with their performance evaluation.

II. SPEECH ENHANCEMENT METHODS:

There are various speech enhancement methods proposed for noise reduction and to improve the noise quality and intelligibility. The basic spectral subtraction algorithm with its modified version is presented below.

A. Spectral Subtraction algorithm:

Spectral subtraction is historically one of the first algorithms proposed in the field of speech enhancement. To date, it has been modified many times by various scientists, engineers, researchers across the globe [1]. 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 principle of spectral subtraction algorithm is shown in Fig. 1.

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

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

where, $x(n)$ represents the clean speech signal and $d(n)$ 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 $|\widehat{D}_k|$ from the noisy signal magnitude spectrum $|Y_k|$. The noisy signal phase spectrum is used as an estimate of the clean speech phase spectrum, as follows:

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

where, $\varphi(y,k)$ 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 as shown in Fig. 1. In Fig. 1, the clean signal $\hat{x}(n)$ is recovered from the noisy signal $y(n)$, by assuming that there is an estimate of the power spectrum of noise $|\widehat{D}_k|^2$, 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:

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

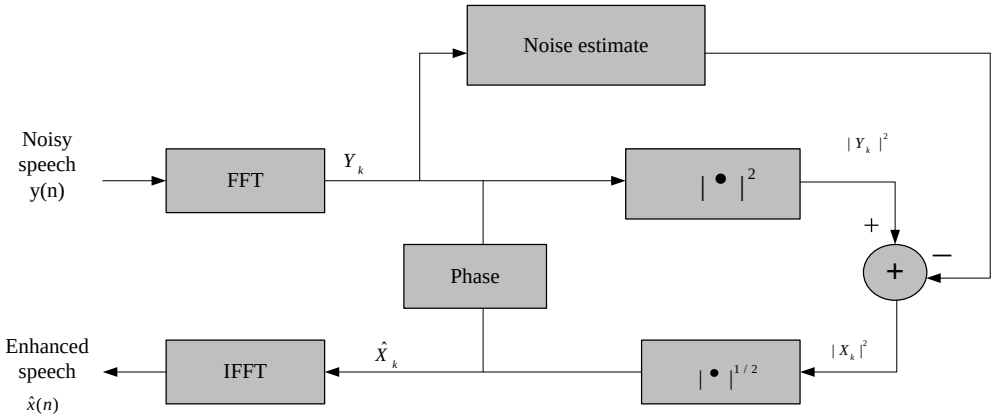

Fig. 1: Block diagram of spectral subtraction technique.

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.

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

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 [2]. It is clear that spectral subtraction method can lead to negative values, resulting from differences among the estimated noise and actual noise frame, which gives errors in estimating the noise spectrum. The simplest solution is set the negative values to zero and the process is known as negative rectification or half-wave rectification [3]. This non-linear processing of the negative values, however, creates small, isolated peaks in the spectrum occurring at random frequency locations in each frame. When converted in the time domain, these peaks sound like tones with frequencies that change randomly from frame to frame. The new type of noise introduced by the half-wave rectification process is commonly referred to as "musical noise." Musical noise is mostly found in the unvoiced segment of speech where the speech power is low and is comparable to the noise power. This musical noise can sometime be more disturbing to the listener than the distortions caused by other interfering noises.

B. Spectral Subtraction with over subtraction

The spectral subtraction with over-subtraction was introduced to reduce the effect of "musical noise". In this approach, the original spectral subtraction method is modified by subtracting an over-estimate of the noise power spectrum and by preventing the resultant spectrum from going below a preset minimum spectral floor value. This modification minimized the perception of narrow spectral excursions and thus lowers the musical noise effect. This algorithm is given in Eqn. 5.

$$\left|\hat{X}_k\right|^2 = \begin{cases} |Y_k|^2 - \left|\widehat{D}_k\right|^2, & \text{if } |Y_k|^2 \geq (\alpha+\beta)\left|\widehat{D}_k\right|^2 \\ \beta\left|\widehat{D}_k\right|^2 & \text{otherwise} \end{cases} \tag{5}$$

where, α and β is over subtraction factor and spectral floor parameter respectively, with α>1 and $0 < \beta \leq 1$. The parameter α is the function of signal to noise ratio (SNR) given by Eqn. 6.

$$\alpha = \alpha_0 - \frac{3}{20}\mathrm{SNR}, \quad -5\mathrm{dB} < SNR < 20dB \tag{6}$$

where, α_0 is the desired value of α at 0dB SNR. The parameter α affects the amount of speech spectral distortion. If α is too large, the resulting signal will be severely distorted and intelligibility may suffer. On the other hand, if α is too small, then, noise may not be completely removed in enhanced speech signal. Therefore, the appropriate value of α is chosen to prevent both musical and signal distortion. Parameter β controls the amount of musical noise and residual noise. If β is too small, musical noise will become audible but the residual noise will be reduced; but, if β is too large, then the residual noise will be audible but the musical noise related to spectral subtraction reduces.

C. Non–linear Spectral Subtraction (NSS)

The Non–linear Spectral Subtraction algorithm is proposed by Lockwood and Boudy [4]. In this approach, the over subtraction factor is made frequency dependent and the subtraction process non-linear. In NSS algorithm, it is assumed that noise does not affect all spectral components equally. In comparison to the high frequency region, the low frequency region is more affected by the certain type of noise. Therefore, frequency dependent subtraction factor is used for different types of noise. Due to frequency dependent subtraction factor, subtraction process becomes nonlinear. Larger values are subtracted at frequencies with low SNR levels and smaller values are subtracted at frequencies with high SNR levels. The subtraction rule used in the NSS algorithm has the following form.

$$|\hat{X}_k| = \begin{cases} |Y_k| - \alpha_k N_k & \text{if } |Y_k| \geq \alpha_k N_k + \beta\,|\hat{D}_k| \\ \beta\,|Y_k| & \text{otherwise} \end{cases} \tag{7}$$

where β is the spectral floor set to 0.1, $|Y_k|$ and $|\hat{D}_k|$ are the smoothed estimates of noisy speech and noise respectively, α_k is a frequency dependent subtraction factor and N_k is a non-linear function of the noise spectrum given as

$$\mathrm{N_k} = \mathrm{Max}\,|\hat{D}_k| \tag{8}$$

The frequency dependent subtraction factor α_k given as

$$\alpha_k = \frac{1}{r} + p_k \tag{9}$$

where, r is a scaling factor and p_k is the square root of the posteriori SNR estimate given as

$$p_k = \frac{|Y_k|}{|\hat{D}_k|} \tag{10}$$

D. Multiband Spectral Subtraction

In the Multiband Spectral Subtraction method proposed by Kamath and Loizou [5], the speech spectrum divided into N oversampling bands and spectral subtraction is performed independently in each band. In this method, firstly, the signal is windowed and the magnitude spectrum is estimated using FFT. The noise and speech are then divided into different frequency band to calculate the over subtraction factor. In the next stage, the individual frequency bands is processed by subtracting the corresponding noise spectrum from the noisy speech spectrum and finally, the modified frequency bands are recombined and the time signal is obtained by using the noisy phase information and taking the IFFT. The estimate of the clean speech spectrum in the ith band is obtained by Eqn. 11.

$$\left|\hat{X}_{ik}\right|^2 = |Y_{ik}|^2 - \alpha_i \delta_i |D_{ik}|^2 \tag{11}$$

where, $k=\frac{2\pi n}{N}, n = 0,1,2, \ldots N-1$ are the discrete frequencies, $|D_{ik}|^2$are the estimated noise power spectrum when speech is absent, α_i is the over subtraction factor of the ith band and δ_i is the additional band. The main difference between the Multiband Spectral Subtraction and the Non-linear Spectral Subtraction algorithm is in the estimation of the over subtraction factors. The Multiband approach estimates one subtraction factor for each frequency band, whereas the Non-linear Spectral Subtraction algorithm estimates one subtraction factor for each frequency bin.

E. MMSE Estimator

To overcome the problem of the aforementioned musical noise distortion present in the spectral subtraction method, Ephraim and Malah [6] in 1984 proposed the MMSE method which reduces the distracting musical noise to a considerable extent, and thus improved the quality of the resulting enhanced speech. The key MMSE based algorithms are Minimum Mean Square Error Short-Time Spectral Amplitude (MMSE-STSA) estimator and MMSE Logarithm Spectral Amplitude (MMSE-LSA) estimator.

The MMSE-STSA method aims to minimize the mean square error between the short-time spectral magnitude of the clean and enhanced speech signal. This method assumes that each of the Fourier expansion coefficients of the speech and noise process can be modeled as independent, zero mean, Gaussian random variables [6]. 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.

F. Selective Spectral Subtraction Algorithm

The methods mentioned in the previous sections made no distinction between voiced and unvoiced segments. However, due to the spectral differences between vowels and consonants [3] several algorithms that treated the voiced and unvoiced segment differently have been proposed. The resulting spectral subtractive algorithms were therefore selective for different classes of speech sounds [3]. In the two band spectral subtraction algorithm, the incoming speech frame was first classified into voiced or unvoiced by comparing the energy of the noisy speech to a threshold. Voiced segments were then filtered into two bands, one above the determined cutoff frequency (high pass speech) and one below the determined cutoff frequency (low pass speech). Different algorithms were then used to enhance the low passed and high passed speech signals accordingly. The over subtraction algorithm was used for the low passed speech based on the short term FFT. The subtraction factor was set according to short term SNR as per [7]. For high passed voiced speech as well as for unvoiced speech, the spectral subtraction algorithm was employed with a different spectral estimator [3].

A dual excitation Model was proposed for speech enhancement; where, speech is decomposed into two independent components voiced and unvoiced components. The first step is to perform voiced component analysis which is done by extracting the fundamental frequency and the harmonic amplitudes. The noisy estimates of the harmonic amplitudes are adjusted according to some rule for any noise that might have leaked to the harmonics and the unvoiced component spectrum is then computed by subtracting the voiced spectrum from the noisy speech spectrum. Finally, a two pass system, which included a modified Wiener Filter, is used to enhance the unvoiced spectrum. As a result, the enhanced speech consists of the sum of the enhanced voiced and unvoiced components. The major challenge with such

algorithms is making accurate and reliable voiced, unvoiced decisions particularly at low SNR conditions.

III. PERFORMANCE OF SPECRAL SUBTRACTION ALGORITHMS

The spectral subtraction algorithm was evaluated in many studies, primarily using objective measures such as SNR improvement, spectral distances and subjective listening tests. The intelligibility and speech quality measures reflect the true performance of speech enhancement algorithms in real life scenarios. Ideally, the Spectral Subtraction algorithm should improve both intelligibility and quality of speech in noise. Results from the literature were mentioned as follows. Boll [8] performed intelligibility and quality measurement tests using the Diagnostic Rhyme Test (DRT). Result indicated that Spectral subtraction did not decrease speech intelligibility but improved speech quality particularly in the area of pleasantness and inconspicuousness of the background noise. Kang and Fransen [9] evaluated the quality of noise processed by the SS algorithm and then fed to a 2400 bps LPC recorder. Here SS algorithm was used as a pre-processor to reduce the input noise level. The Diagnostic Acceptability Measure (DAM) test [10] was used to evaluate the speech quality of ten sets of noisy sentences, recorded actual military platforms containing helicopter, tank, and jeep noise results indicated that SS algorithm improved the quality of speech. The largest improvement in speech quality was noted for relatively stationary noise sources [3], [11]. The NSS algorithm was successfully used in [4] as a pre-processor to enhance the performance of speech recognition systems in noisy environment. The performance of the multiband spectral subtraction algorithm [12] was evaluated by Hu Y. and Loizou [2], [10] using formal subjective listening tests conducted according to ITU–T P.835 [13]. The ITU T P.835 methodology is designed to evaluate the speech quality along with three dimensions signal distortion, noise distortion and overall quality. Results indicated that the MBSS algorithm performed the best consistently across all noise conditions, [3] in terms of overall quality. In terms of noise distortion the MBSS algorithms performed well, except in 5dB train and 10dB street conditions. The algorithm proposed by Virag was evaluated in [14] using objective measures and subjective tests, and found better quality than the NSS and standard SS algorithms. The low energy segments of speech are the first to be lost in the subtraction process; particularly when over subtraction is used. Overall most studies confirmed that the SS algorithm improves speech quality but not speech intelligibility.

IV. CONCLUSIONS

This is a review paper and various spectral subtraction algorithms are described for speech enhancement. These algorithms are computationally simple to implement as they involve a forward and an Inverse Fourier Transform. However, the major drawback of this algorithm is that subtraction of the noise spectra from the noisy spectrum introduces a distortion in the signal known musical noise and different techniques that mitigated the musical noise distortion are presented in this paper. The spectral subtraction were modified a number of times over the years. The most common variation involved the use of an over subtraction factor that controlled to some amount of speech spectral distortion caused by subtraction process. Use of spectral floor parameter prevents the resultant spectral components from going below a preset minimum value. The spectral floor value controlled the amount of remaining residual noise and the amount of musical noise. Different methods proposed for computing the over subtraction factor are based on different criteria that includes linear and non-linear functions of the spectral SNR of individual frequency bins or bands. Evaluation of

spectral subtractive algorithms revealed that these algorithms improve speech quality and not affect much more on intelligibility of speech signals.

REFERENCES:

[1] M. Berouti, R. Schwartz, & J. Makhoul, "Enhancement of Speech Corrupted by Acoustic Noise," Proc. ICASSP, pp. 208-211, 1979.
[2] Yi Hu & Philipos C. Loizou, "Subjective comparison and evaluation of speech enhancement algorithms," IEEE Trans. Speech Audio Proc., Vol. 49, No. 7, pp. 588–601, 2007.
[3] Phillips C Loizou, "Speech enhancement theory and practice" 1st ed. Boca Raton, FL, CRC, 2007.
[4] P. Lockwoord, & J. Boudy, "Experiments with a Nonlinear Spectral Subtractor (NSS), Hidden Markov Models and the projection, for robust speech recognition in cars," Speech Communication, Vol. 11, pp. 215-228, 1992.
[5] K. Lebart, & J. M. Boucher, "A New method based on spectral subtraction for speech enhancement," Acustica, Vol. 87, pp. 359-366, 2001.
[6] Y.Ephraim & D. Malah, "Speech Enhancement using a minimum mean square error short-time spectral amplitude estimator," IEEE Trans. Acoust., Speech , Signal Process., Vol. ASSP-32, pp. 1109-21, Dec. 1984.
[7] M. Berouti, R. Schwartz & J. Makhoul, "Enhancement of Speech Corrupted by Acoustic Noise," Proc ICASSP, pp. 208-211, 1979.
[8] S. F. Boll, "Suppression of Acoustic Noise in Speech using Spectral Subtraction," IEEE Trans ASSP Vol. 27, No. 2, pp. 113-120, April 1979.
[9] W. Kim, S. Kang & H. Ko, "Spectral subtraction based on phonetic dependency and masking effects," IEEE Proc. vision image signal process, Vol. 147, No. 5, pp. 423-427, 2000.
[10] Yi Hu & Philipos C. Loizou, "Evaluation of Objective Quality Measures for Speech Enhancement," IEEE Trans. on Audio, Speech, and Language processing, Vol.16, 2008.
[11] Gustafsson, Nordhohm & Claesson, "Spectral subtraction using reduced delay convolution and adaptive averaging," IEEE. Trans. Speech Audio Process, Vol. 9 No. 8, pp. 799-805, 2001.
[12] S. Kamath & P. Loizou, "A multiband spectral subtraction method for enhancing speech corrupted by colored noise" Proc. IEEE Intl. Conf. Acoustics, Speech, Signal Processing, 2002.
[13] ITU-T, "subjective test Methodology for evaluating speech communication system that include noise suppression algorithm." ITU-T recommendation p.835, 2003.
[14] N. Virag, "Single channel speech enhancement based on masking properties of the human auditory system". IEEE. Trans. Speech Audio Process, Vol. 7, No. 3, pp. 126-137, 1997.