International Journal on Emerging Technologies **9**(2): 70-76(2018)

ISSN No. (Print) : 0975-8364
ISSN No. (Online) : 2249-3255

Noise Reduction for Speech Improvement with Correlation based Method

Shilpi Dwivedi[1] and Sneha Jain[2]
[1]Research Scholar, Department of Electronic and Communication Engineering,
RITS, Bhopal (Madhya Pradesh), India
[2]Assistant Professor, Department of Electronic and Communication Engineering,
RITS, Bhopal (Madhya Pradesh), India

(Corresponding author: Shilpi Dwivedi)

(Received 16 September, 2018 Accepted 08 November, 2018)
(Published by Research Trend, Website: www.researchtrend.net)

ABSTRACT: Speech is most effective and natural medium to exchange of the information among people. Speech enhancement is essential for human listening. Fundamental aspects of acoustics are too related to the understanding the recognition, estimation and control of noise. In speech communication, the speech signal is always accompanied by some noise. Speech enhancement techniques are used to enhance the corrupted signal by reducing noise. It is a technique that improves the quality of the speech signal. The main objective of speech signal enhancement is to improve the overall aspects of speech such as quality and intelligibility to the listener. In current work, we analyzed autocorrelation as a most widely used speech enhancement technique for better noise cancellation and compare it with other techniques.

Keywords: Signal processing, autocorrelation, Cannon noise

I. INTRODUCTION

The speech signal gets degraded because of various types of noise like background noise, reverberation, babble etc. The clean speech signal is necessary for applications such as speech or speaker recognition, hearing aids, mobile communication. It is assumed that the noise is additive. It is assumed that the noise characteristics change very slowly as compared to the signal. The speech enhancement technique can be classified on number of microphones available into single channel, dual or multi-channel speech enhancement technique. The single channel speech enhancement technique uses only one microphone whereas multiple channel speech enhancement technique uses array of more than one microphone. Multi-channel speech enhancement technique provides better performance than single channel but, because of its convenient implementations and ease of computations single channel speech enhancement technique is still a significant area of research. The various types of noise against which speech signal need to process or enhance are periodic noise, wide band noise, interfering speech and impulsive noise. Spectral subtraction is widely used speech enhancement technique. However, the musical noise introduced is the drawback of the spectral subtraction method. Many modifications of the spectral subtraction method to perform better noise cancellation have been proposed in literature. In this paper, we review the modifications made in the spectral subtraction method. When calculating the autocorrelation you get a lot of information of the speech signal. One of that is the pitch period (the fundamental frequency). To make the speech signal closely approximate a periodic impulse train we must use some kind of spectrum flattening. The autocorrelation is calculated and the fundamental frequency is extracted.

In real-world enhancement process will be downgraded by numerous types of background noise, with other speech sources, channel distortion, speech variability. The generation of noise depends on sounding and device used in communication. Most of the existing algorithms on single-channel noise reduction are based on estimation of stationary noise from segments where the desired signal is absent. Traditional approaches of speech enhancement usually consist of two components: first noise power spectrum estimation and then estimation of the desired clean speech signal. Traditional methods do not take into account the repetitive nature of many transient noises. Usually a distinct pattern appears a large number of times at different time locations.

The fact that the same pattern appears multiple times can be utilized for improved de-noising. Specifically, the pattern intervals can be identified, and the transient noise can be extracted by averaging all of these instances. VAD is usually a preprocessing step in speech processing applications such as speech or speaker recognition. A straight forward application of VAD would be an automatic camera steering task. Suppose a scenario in which there exist multiple speakers with a camera assigned to each of them. The speech data can be enhanced with two ways –either by processing the speech signal itself or by enhancing the extracted features.

Wei Shi [1] analyzed a voice activity detection algorithm based on a novel long-term metric. The long-term autocorrelation statistics (LTACS) measure is designed as a powerful metric used in VAD. The LTACS measure is calculated among several successive frames around the concerned frame and it represents the significance of harmonics of the signal spectrum over a long term rather than a short term. A novel LTACS-based VAD algorithm is derived by jointly making use of the minimum operator to reduce non-speech variability and of then calculating variance to detect speech. Amol [2] reviewed speech enhancement techniques. Various types of noise and techniques for removal of those noises are presented. Speech enhancement is necessary for many applications in which clean speech signal is important for further processing. The speech enhancement techniques mainly focus on removal of noise from speech signal. Most widely used speech enhancement technique namely, spectral subtraction method is reviewed in this paper with its state-of-art for better noise cancellation. Schasse [3] proposed a solution for the online adaptation of optimal FIR filters for speech enhancement in DFT subbands. An important ingredient to such filters is the estimation of the inter-frame correlation of the clean speech signal. To evaluate two online estimation approaches based on a constant noise inter-frame correlation and on the use of a binary mask. Results show that a filtering of sub-band signals based on these estimated quantities outperforms a conventional, instantaneous spectral weighting, such as the frequency-domain Wiener filter at least for high SNR conditions.

Speech is non-stationary signal where properties change quite rapidly over time [4]. For most phonemes the properties of the speech remain invariant for a short period of time (5-50ms). Thus, for a short window of time, traditional signal processing methods can be applied relatively successfully. Most of speech processing in fact is done in this way: by taking short windows and processing them.

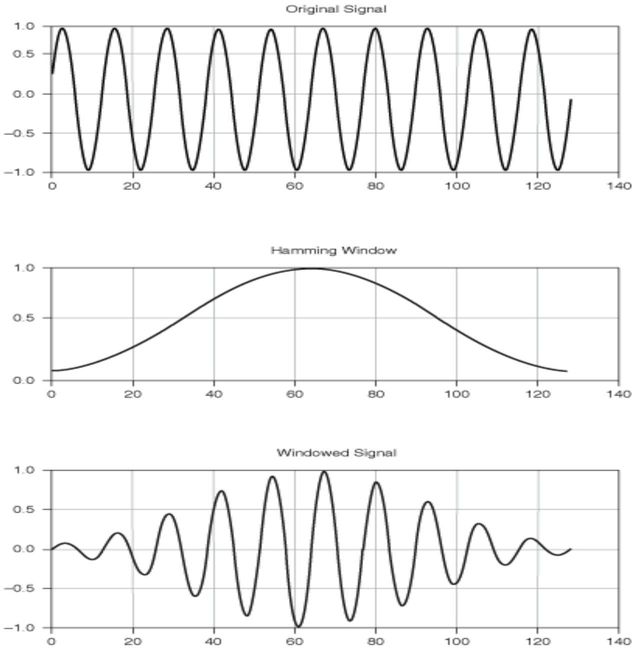

Fig. 1. Application of hamming window on speech signal.

The short window of signal like this is called a frame. In implementations view the windowing corresponds to what is understood in filter design as a window-method [5,6]: a long signal (of speech, for instance, or ideal impulse response) is multiplied with a window function of finite length, giving a finite length weighted (usually) version of the original signal. Based on this, the windowing should be confronted with a certain degree of freedom being prepared to change the window function when necessary. In figure 1 in speech samples will be represented when the hamming window is used [7]. In speech recognition the windows are usually overlapping 10 ms windows, which are analyzed in order to make hypothesis of the current phoneme. Windowing reduces the amplitude of the discontinuities at the boundaries of each finite sequence acquired by the digitizer. Windowing consists of multiplying the time record by a finite-length window with amplitude that varies smoothly and gradually to zero at the edges. This makes the endpoints of the waveform meet and, therefore, results in a continuous waveform without sharp transitions. This technique is also referred to as applying a window.

In our thesis, we tried to specify unwanted transients by correlating the previous power standards and then eliminated it. For this purpose we take canon noise and add it to the speech signal. Thereafter autocorrelation based method is performed on the noisy speech signals. The taken speech signal is a one dimensional signal, and the correlation is performed with its delayed function. The algorithm solves the transient problem of threshold for transient reduction and provides the alternative. Finally the simulated results shows better enhancement. The autocorrelation is calculated and the fundamental frequency is extracted. We reviewed autocorrelation as a most widely used speech enhancement technique for better noise cancellation.

II. PROBLEM FORMULATION

The signal generation and its storage for further processing, is made by using different types of sensors. These sensors placed at any location may have some noise. For the real time application, it is also affected by the various types of noise. Here we are considering the transient noise and take the step to enhance the quality of signal. The speech signal is generally a one dimensional signal, for that the correlation may be with its delayed function. The transient noise problem is reduced through the autocorrelation function [8].

The noise reduction is performed to the clean speech signal s(k) from the noisy signal observation.

$$S(k) = s(k) + d(k)$$

Where d(k) is the unwanted additive transient noise which is assumed to be a zero-mean random process (white or colored) and uncorrelated with s(k). An estimate of S(k) can be obtained by passing s(k) through a linear filter, i.e.

$$\overline{S}(k) = H^{-1}(s(k) + d(k))$$

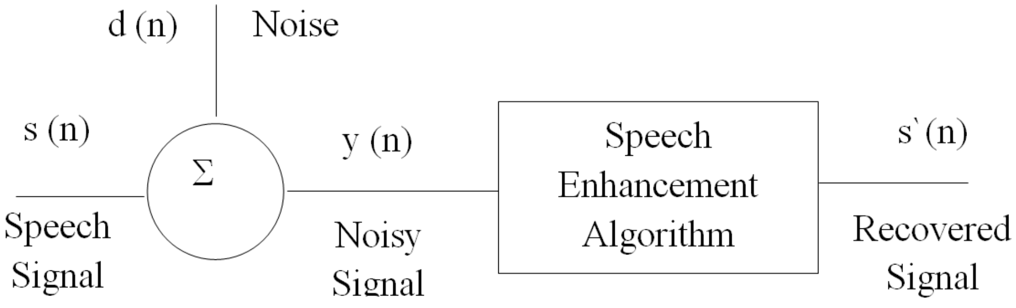

Fig. 2. Noise reduction of speech signals.

With this formulation, the objective of noise reduction is to find an optimal filter that would attenuate the noise as much as possible while keeping the distortion of the clean speech low. To enhance such noisy speech signals, an auto-correlation function based speech enhancement is presented here. Its principle is based on the traditional Spectral Subtraction method. In this Spectral Subtraction method, the degraded speech signals are enhanced by subtracting the estimate of the average noise spectrum from a noisy speech spectrum. The noise spectrum is estimated during the periods when the signal is absent; which is usually very difficult to do in practice. In addition, it is also assumed that speech and noise is additive and uncorrelated. An estimate of the clean signal s(n) is recovered from the noisy signal S(n) by assuming that there is an estimate of the power spectrum of noise, which is obtained by averaging over multiple frames of a known noise segment.

Although the spectral subtraction algorithm can be easily implemented to effectively reduce noise present in the corrupted signal; yet, it has several shortcomings. The major drawback of this method is the resulting musical noise, due to rapid coming and going of speech signals over successive frames. This is why this paper focuses on using the autocorrelation function instead of the power spectrum.

III. METHODOLOGY

Autocorrelation, also known as serial correlation, is the correlation of a signal with a delayed copy of itself as a function of delay. It is a measure of how similar a signal is to itself. Informally, it is the similarity between observations as a function of the time lag between them. Correlation is usually between two signals, we think of it as a system with two inputs and no stored coefficients.

We also know that cross-correlation is a measure of similarity between two signals, while autocorrelation is a measure of how similar a signal is to itself. Autocorrelation is the cross-correlation of a signal with itself. It is the similarity between observations as a function of the time lag between them. It is a mathematical tool for finding repeating patterns, such as the presence of a periodic signal obscured by noise, or identifying the missing fundamental frequency in a signal implied by its harmonic frequencies. It is often used in signal processing for analyzing functions or series of values, such as time domain signals. The idea of autocorrelation is to provide a measure of similarity between a signal and itself at a given lag.

The analysis of autocorrelation is a mathematical tool for finding repeating patterns, such as the presence of a periodic signal obscured by noise, or identifying the missing fundamental frequency [9, 10].It is often used in signal processing for analyzing functions or series of values, such as time domain signals. There are several ways to approach it, but for the purposes of pitch/tempo detection, one can think of it as a search procedure. In other words, we step through the signal sample-by-sample and perform a correlation between once reference window and the lagged window. The correlation at "lag 0" will be the global maximum because we are comparing the reference to a verbatim copy of itself. As we step forward, the correlation will necessarily decrease, but in the case of a periodic signal, at some point it will begin to increase again, and then reach a local maximum. The distance between "lag 0" and that first peak gives us an estimate of our pitch/tempo.

Autocorrelation of s(n), which is defined as the sequence

$$r_{ss} = \sum_{n=-\infty}^{\infty} s(n)s(n-1)$$

With finite duration sequences, it is customary to express the auto-correlation in terms of the finite limits on the summation.

$$r_{ss} = \sum_{n=i}^{N-|k|-1} s(n)s(n-1)$$

Computing sample-by-sample correlations can be very computationally expensive at high sample rates, so typically an FFT-based approach is used. Taking the FFT of the segment of interest, multiplying it by its complex conjugate, then taking the inverse FFT will give the cyclic autocorrelation [11]. The main aim of this scheme is to find out the frequency of a complex signal like a speech signal using relatively easily available tools like MATLAB. The scheme presents a design of an interesting experiment for the Digital Signal Processing. In this paper a scheme for finding out the autocorrelation of a sound signal is proposed. The general informative speech signal is not repetitive in nature. That`s why the autocorrelation is giving single peak corresponding the zero time lag. The position of second and other peaks are one fourth of the highest peak and other peaks are one tenth of the highest peak of the speech signal. The coefficient for the musical signal is half and one fourth respectively. The frequency of the periodic signal is found out by considering the first peak on which the autocorrelation has maximum value. This maximum value corresponds to the number of samples (or time duration) after which signal is repeating itself. The autocorrelation function of a signal contains the same information about the signal as its power spectrum. However, the main difference between the power spectrum and auto-correlation domain is that in the power spectrum domain, the information is presented as a function of frequency; while in the latter, it is presented as a function of time [12].

More specifically, the higher-lag auto-correlation coefficients of the speech signal s(n) usually contain information about the signal's power spectrum; whereas, the magnitude of higher-lag auto-correlation coefficients of the noise signal $d(n)$ is relatively small for some noise types.

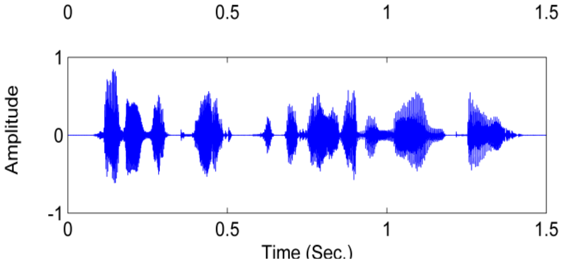

Fig. 3. Speech sample signal.

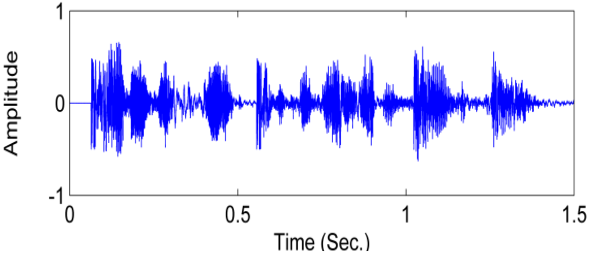

Fig. 4. Noisy signal with cannon transient noise.

Therefore, the lower-lag auto-correlation coefficients can be discarded and only higher-lag auto-correlation coefficients of the noisy speech signal can be used for spectral estimate [13]. The spectral estimation uses only the higher-lag auto-correlation coefficients; and, the speech and noise signals can be separated without having to estimate the noise signal directly. The estimated power spectrum can then be used to enhance the corrupted noisy speech signals. It is to be noted that the auto-correlation function of a signal can be computed in time domain and frequency domain. The autocorrelation coefficient is maintained up to 25-35 percent for the speech signal from its second peak at zero delay and up to 45-50 percent for the musical signal except for the zero time delay lag. The iterative median filtering is used for the autocorrelation as a constraint. The median filtering smoothening is applied to eliminate speech like noise components of the speech signal. Smoothing can be non-linear as well as linear.

Figure 3 shows the speech signal taken for practical consideration. It shows the original signal without any processing. The frame length considered is 512 and sampling is 8 KHz. Figure 4 shows the noise which shows that the signal components lies in all the frequency range. Transient noise is added here so that we have a noisy signal [14]. Here we take cannon noise as a noisy signal of transient nature. The noise effect on the speech signal is illustrated. The autocorrelation spectrum is completely disturbed from its original shape as the peaks other than highest and second peaks are following the second peaks. This property is used to reduce the effect of noise in the speech signal [15].

Table 1: Various simulation parameters taken in our experiment.

Parameters	Value
Frame length	512
Sampling Frequency	8 KHz
Noise type	Transient Noise
Windowing technique	Hamming window

As mentioned before feature extraction, signal processing is carried out. Signal pre-processing includes de-noising and end point detection. The process of features extraction of the speech signal starts with: framing the speech signal and windowing it. After that, the smoothening median filter and autocorrelation methods and Spectrogram will be applied to extract the coefficients from the speech signal. The windowing is applied and every frame is multiplied with a window function w(n) with lengthN, where N is the number of samples in each frame (w(n)*y(n), here y(n) is frame signal).

Windowing is used to avoid problems due to truncation of the signal. Here Hamming window is used in speech recognition systems for windowing operation. The signal obtained from the operation of multiplying the frame signal by hamming window is:

$$w(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right), \qquad 0 \le n \le N-1$$

On this noisy signal a sequence of process has been applied to reduce the noise and extract the original speech signal. The proposed method uses a recursive approach and the autocorrelation coefficient as a constraint or stopping criterion for which we used the nonlinear type median filtering. The algorithm solves the transient problem of threshold for transient reduction. It is clearly illustrates that after filtering the signal, noise is reduced above 7 KHz. The time domain waveform also clears the view to understand the effect of noise removal method. The de-noised signals spectrum is like the original signal spectrum.

IV. RESULT ANALYSIS

The simulated result is illustrated through the following mentioned figures. This method simulates in MATLAB 2012a. The graphics are integrated with MATLAB. Since MATLAB is also a programming environment, a user can extend the functional capabilities of MATLAB by writing new modules. MATLAB has a large collection of toolboxes in a variety of domains. In our work we define several functions which are used in our work.

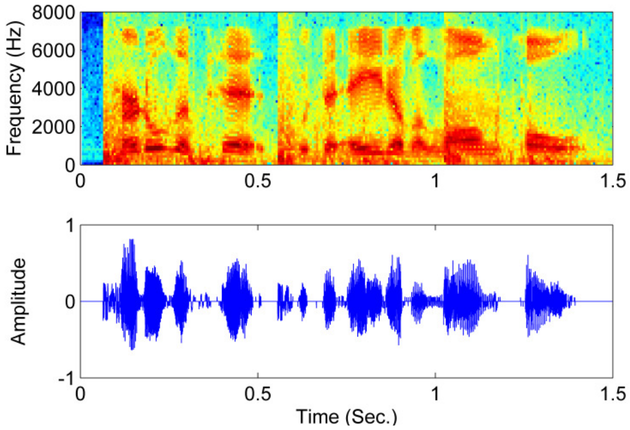

Fig. 5. The frequency spectrum and the time domain waveform of de-noised signal.

In order to evaluate the enhancement of the speech, we measure the commonly used signal to noise ratio (SNR). Table 2 summarizes the SNR, calculated only in transient occurrence time frames. The Segmented SNR is computed for both the noisy and de-noised signals.

Table 2: Segmental SNR in dB for the speech signal with transient noise.

Input SSNR (dB)	Output SSNR (dB)	Improvement in SNR
1.103	3.242	2.14 dB

The original speech signal is same for another noise analysis whose waveform in the time domain is illustrated in figure 3. The figure 4 shows the signal with cannon transient noise shows that the signal components lies in all the frequency range above 7KHz also. Figure 5 shows clearly that after filtering the signal, noise is reduced. The de-noised signal spectrum is like the original signal spectrum. We observe that the proposed algorithm obtains better SNR compared to the SS method.

A. Performance evaluation and comparison of result with previous work

Speech signals can be degraded in many ways during their acquisition in noisy environments and they can also be further degraded in the electronic domain. Serious signal degradation, however, is most commonly caused by noise from unwanted acoustic sources in the environment, which may affect the speech quality and/or intelligibility of the wanted signal. In this thesis, we have focused on the enhancement of speech signals that have been corrupted by levels of additive noise that are high enough to affect the intelligibility of the speech. Numerous approaches for speech enhancement have been developed over many years. The majority of algorithms perform the enhancement in a transform domain in which both speech and noise signals are sparse. In our work we implement autocorrelation on speech signal to eliminate the transient noise. Cannon noise is created in a war zone, where it distracts the original speech signal. In terms of SNR we obtain better results and successfully eliminate the cannon noise. As per the performance concern the results are compared with previous work.

In reference paper 2 the author implemented spectral subtraction method on the signal contaminated with some kind of background noise at home like fan noise. In his work he used 3sec and 6sec long duration speech signal sampled at 8 KHz. Experimental results show that, SNR of the enhanced speech signal is improved by 1.72 dB with respect to input SNR. In our work we enhance the speech signal by 2.3 dB with respect to the contaminated signal. In reference paper 10 the author implemented this method on the signal with white noise. Author analyzed the auto-correlation property of speech signal by dividing the speech signal into several component elements and achieved some beneficial conclusions with white noise. They also analyze short term auto-correlation property of speech signal and confirm it through detailed comparing experiment with other kind of signals. By applying the auto-correlation property of the current speech frame and frames nearby, a new feature with voice activity is detecting.

In our work we implement the autocorrelation with a median smoothing filter over the transient noise (cannon noise) and successfully eliminate the noise. The main focus of our work is to implement the autocorrelation method in a war zone where the signals are contaminated with the cannon noise. With 2.3 dB improvement, we successfully eliminate this transient noise and extract the original signals. Results showed that the proposed method outperformed the multiband spectrum subtraction method in enhancing speech corrupted with transient noise.

V. CONCLUSION

A speech enhancement method was proposed for enhancing speech corrupted with cannon noise. Unlike other speech enhancement techniques which assume that speech and noise are uncorrelated, the proposed method takes into account possible correlation between speech and noise. An iterative procedure like mean is an ordinarily jumble-sale manner for assessment of unwanted transient power spectrum. The proposed method uses a recursive approach and the autocorrelation coefficient as a constraint or stopping criterion for which we used the nonlinear type median filtering. The algorithm solves the transient problem of threshold for transient reduction. The Implementation of the algorithm is considered with 50 samples of speech signal.

In this work we enhance the cannon noise contaminated speech signal by 2.14dB.It is also observed that this method is more suitable for the speech signal enhancement. The Segmented SNR is stable for the speech signals. The autocorrelation method had no problem with either type of signal and is robust against noisy signals. The preferred pitch detector is therefore the autocorrelation method. This research aims to reduce or effectively remove the requirement for knowledge of noise. This is significant as it could lead to a technique that is capable of retrieving speech from noisy data without requiring vast amounts of noisy training data or noise statistics. With the help of some studies, and with the experiment, we showed that we could potentially increase significantly noise immunity for speech enhancement without requiring noise knowledge. This thesis describes a realization of this approach to practical use.

Our concern is that this could correspond to one technique used by humans for picking out speech in strong noise to try to make sense of the speech.

REFERENCES

[1]. Wei Shi, Yuexianzou, yiliu, (2014). "Long-Term Auto-Correlation Statistics Based Voice Activity Detection For Strong Noisy Speech", *IEEE/ChinaSIP,* Vol. **2**(14), pp. 100-104.

[2]. Jennifer C Saldanha, Shruthi O R, (2016). "Reduction of Noise for Speech Signal Enhancement Using Spectral Subtraction Method", *IEEE/ICIS,* Vol. **8**(16), pp. 44-47.

[3]. Ji Ming, Danny Crookes, (2017). "Speech Enhancement Based on Full-Sentence Correlation and Clean Speech Recognition", IEEE/TASLP, 2017.

[4]. Amol Chaudhari, S. B. Dhonde, (2015). "A Review on Speech Enhancement Techniques", IEEE/ICPC, 978-1-4799-6272-3/15, 2015.

[5]. Alexander Schasse and Rainer Martin, (2014). "Estimation of Subband Speech Correlations for Noise Reduction via MVDR Processing", *IEEE/ACM Transactions On Audio, Speech, And Language Processing,* Vol. **22**(9), 2014.

[6]. Lalchhandami1, Maninder Pal, (2013). "An Auto-Correlation Based Speech Enhancement Algorithm", *IJERD*, Vol. **7**(5), PP. 23-30.

[7]. Alexander Schasse and Rainer Martin, (2013). "Online Inter-Frame Correlation Estimation Methods For Speech Enhancement In Frequency Subbands", *IEEE/ICASSP,* Vol. **6**(13), pp. 7482-7286.

[8]. Richard C. Hendriks, Timo Gerkmann, (2012). "Noise Correlation Matrix Estimation for Multi-Microphone Speech Enhancement", *IEEE Transactions On Audio, Speech, And Language Processing*, Vol. **20**(1), 2012.

[9]. Pankaj Bactor, Anil Garg, (2012). "Different Techniques for the Enhancement of the Intelligibility of a Speech Signal", *IJERD*, Volume **2**(2), pp. 57-64, 2012.

[10]. Zhang Shuyin, Guo Ying, Wang Buhong, (2009). "Auto-correlation Property of Speech and Its Application in Voice Activity Detection", IEEE/International Workshop on Education Technology and Computer Science, 2009.

[11]. Jesper B. Boldt1, Daniel P. W. Ellis, (2009). "A Simple Correlation-Based model Of Intelligibility For Nonlinear Speech Enhancement And Separation" EURASIP/EUSIPCO, pp. 1849-1853, 2009.

[12]. Jacob Benesty, Jingdong Chen, Yiteng (Arden) Huang, (2008). "On the Importance of the Pearson Correlation Coefficient in Noise Reduction", *IEEE Transactions On Audio, Speech, And Language Processing,* Vol. **16**(4), 2008.

[13]. G. Farahani1, S.M. Ahadi, M.M. Homayounpour, (2007). "Improved Autocorrelation-Based Noise Robust Speech Recognition Using Kernel-Based Cross Correlation And Overestimation Parameters", EURASIP/EUSIPCO, pp. 2355-2359.

[14]. Rab Nawaz, Jonathon A. Chambers, (2005). "A Novel Single Lag Auto-Correlation Minimization (Slam) Algorithm For Blind Adaptive Channel Shortening", *IEEE/ICASSP,* Vol. **7**(5).

[15]. Shujith Ikbul, Hemant Misra, Hewe Bourlurd, (2003). "Phase Autocorrelation (Pac) Derived Robust Speech Features", *IEEE/ ICASSP,* Vol. **3**(3), pp. II-133-II-136.