International Journal of Electrical, Electronics and Computer Engineering **7**(2): 86-90(2018)

ISSN No. (Online): 2277-2626

A Review of Various Single Channel and Multi-Channel Transient Noise Reduction Techniques

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 07 August, 2018 Accepted 02 November, 2018)

(Published by Research Trend, Website: www.researchtrend.net)

ABSTRACT: Speech is most natural medium to exchange the information between two medium. In speech communication, the speech signal is always accompanied by some noise. Therefore, speech enhancement is required and it not only involves processing speech signals for human listening but also for further processing prior to listening. The growth of communication device and signaling technique compromised with quality of voice. The quality of voice and speech come in free environment and adds some noise such as pink noise, background noise and adaptive noise. The variety of noise depends on transmission scenario and domain analysis. For the enhancement of speech various authors used transform based technique such as FFT, DCT, Wavelet and some are used adaptive technique. In this paper we discuss different methods for speech improvement and some adaptive technique. The enhancement process of speech signal discusses in two domains one is temporal domain and another one is spectral domain.

Keywords: Speech Enhancement, Transient noise, Transform Function, Adaptive Technique

I. INTRODUCTION

A sound wave is the end product of the speech production mechanism and the primary source of raw material used by the listener to recover the speaker's message. Sound lies at the very center of speech communication. Because of the central role played by sound in speech communication, it is important to have a good understanding of how sound is produced, modified, and measured. Sound is the pressure variations, or oscillations, in an elastic medium like air, water, solids), generated by a vibrating surface, or turbulent fluid flow. Sound always propagates in the form of longitudinal waves, involving a succession of compressions and rarefactions in the medium. The basic paradigm for sound propagation can be understood with a source-path receiver. This can be start with a device that generates sound waves. The source is then connected to a receiver through some propagation media, which is commonly air, but could be a structure, another fluid, or a complex combination of these media. Finally the sound wave reaches an object where it somehow affects the nature of that object. Usually the receiver is a person, but it could be also being a piece of delicate equipment that can be affected by sound waves. Here we focus on how to characterize the source component of the source-path-receiver paradigm. When a sound wave propagates in air, the oscillation in pressure is above and below the ambient atmospheric pressure.

All the speech enhancement methods aimed at suppressing the background noise are based in 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. Finding the pauses in speech is based on checking how close the estimate of the background noise is to the signal in the current window. Voiced sections can be located by estimating the fundamental frequency. Working voice activity detection (VAD) giving values of zero and one as indicators of the voice activity in each frame enable us to update the estimate of the background noise spectrum during the frames [1].

There are also more sophisticated methods for estimating the background noise. One that does not need VAD is called 'minimum statistics'. To do this, it is necessary to make use of prior knowledge about differences between the characteristics of noise and speech. Common assumptions are:

-The short-time power spectrum of noise is more stationary than that of speech

-Within a narrow frequency band, the speech energy frequently falls to a low value

-The frequency of periodic noise sources changes very slowly with time; this is in contrast to voiced speech whose period changes more rapidly.

Its idea is to calculate the flattened spectrum and then to assign to each spectral bin the smallest value of several consequent spectral samples. The estimate will be a little off-biased, but it works better than the first method if the noise is non-stationary. In many speech enhancement algorithms, the first step is to estimate the power spectrum of the noise. The estimation of the noise is almost always performed in a spectral or related domain for several reasons: speech and noise are partially separated in the spectral domain; spectral components of both speech and noise are somewhat de-correlated; psycho-acoustic models are conveniently applied in this domain. In each of the domains, the coefficients are most frequently taken to be Gaussian and uncorrelated; these assumptions are rarely well substantiated. A well-presented evaluation of several noise spectrum estimation techniques is given by Dupont who found that the best performance of the tested algorithms was given by a combination of minimum statistics and harmonic filtering [2].Microphone array based beam forming amplification, various filters like median, nonlocal neighborhood filters, noise suppression, spectral subtraction based enhancement approaches are used to reduce the mismatch between the training and the testing acoustical conditions. This can be complicated while considering the redundancy of speech signal, non-stationary and variable levels of transient noise pulses. An alternate approach used to extract the robust features or to enhance extracted recognition features thus making the speech recognition robust to noise. Recently some machine learning technique is used for the reduction and elimination of noise using clustering technique. These noises are additive generally in nature. Most speech enhancement techniques are devised that handle the additive background noise. Noise which is not additive (e.g., multiplicative or convolutional) can be handled with the application of a logarithmic transformation to the noisy signal, either in time domain (for multiplicative noise) or in frequency domain (for convolution noise), that converts the distortion to an additive one [3]. Variety of technique is devised to handle convoluted distortion and reverberation. The approach to speech enhancement varied considerably depends on the type of degradation.

II. DIFFERENT TECHNIQUES OF SPEECH ENHANCEMENT

It is difficult for a single algorithm to perform evenly across all types of noise. Therefore, there are many ways to classify speech enhancement methods. A speech enhancement system is based on certain assumptions and constraints that depend on the application. In general the performance of a speech enhancement algorithm is limited by the different factors. The approach to speech enhancement varies considerably depending upon type of degradation. Based on speech acquired from single microphone or multiple microphone sources, Speech enhancement techniques can be divided into two basic categories: (I) single channel enhancement and (II) multiple channels enhancement [4,5]. This is based on speech acquired from single microphone or multiple microphone sources respectively.

A. Single Channel Speech Enhancement Techniques

Single-input speech enhancement systems are those where the only available signal is the noise-contaminated speech picked up by a single microphone. Single-input systems do not cancel noise; rather they suppress the noise using estimates of the signal-to-noise ratios of the frequency spectrum of the input signal. These are the most common real-time scenario algorithms and usually a second channel is not available in most of the applications. Mobile communication, hearing aids are the example of single channel. These systems are easy to build and comparatively less expensive than the multiple input systems. They constitute one of the most difficult situations of speech enhancement, since no reference signal to the noise is available, and the clean speech cannot be preprocessed prior to being affected by the noise. Usually single channel systems make use of different statistics of speech and unwanted noise. The performance of these methods are usually limited in presence of non-stationary noise as most of the methods make an assumption that noise is stationary during speech intervals and also, the performance drastically degrades at lower signal to noise ratios. The performance of single channel systems is usually limited because they tend to improve the quality of the noisy signal at the expense of some intelligibility loss. Therefore, there is a tradeoff between quality and intelligibility.

Spectral Subtraction Method. It is very simple method and easy to implement, it based on the principle that we can obtain an estimate of the clean signal spectrum by subtracting an estimate of the noise spectrum from the noisy speech spectrum [6].

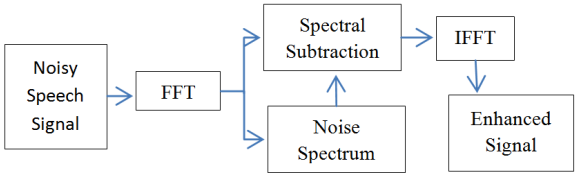

Fig. 1. Spectral Subtraction Method.

The noise spectrum can be estimated, and updated, during the periods when the signal is absent or when only noise is present i.e. during speech pauses. Basic 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.

Weiner Filter. Widely utilized algorithm in speech enhancement research is the Wiener filter. If both the signal and the noise estimates are exactly true, this algorithm will yield the optimal estimate of the clean signal. Because the Wiener filter has a zero phase spectrum the phase from the noisy signal is the output phase for the estimation of the PDS of the clean signal. This was similar to the spectral subtraction algorithms. The Wiener filter assumes that the noise and the signal of interest are ergodic and stationary random processes and thus not correlated to each other. To accommodate the non-stationary of speech signals, the signals can be broken into frames to assume stationarity, as is commonly done in speech signal processing research. Again, similar to spectral subtraction, a prior knowledge of the noise signal is required, but is often difficult to obtain. Incorporating iterative techniques and methods of estimating the noise are therefore important to the Wiener filter algorithm. The iterative techniques re-estimate the Wiener filter with each of the iteration.

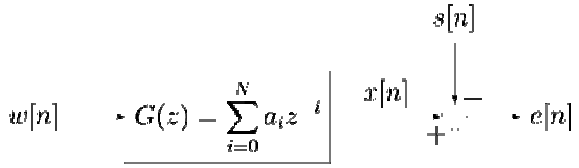

Fig. 2. Wiener filter Algorithm.

Block diagram view of the FIR Wiener filter for discrete series sows that an input signal *w*[n] is convolved with the Wiener filter g[n] and the result is compared to a reference signal *s*[n] to obtain the filtering error *e*[n].

Kalman Filter. The Kalman filter is a Bayesian filter in that it employs the prior probability distributions of the signal and noise processes, the signal is assumed to be a zero-mean Gaussian-Markov process whereas the noise is assumed to be zero-mean independent identically distributed (IID) Gaussian process. The filter also assumes that the parameters of the models of signal and noise generation and channel distortion are known a priori. Kalman filter formulation is based on a state-space approach in which a state Equation models the dynamics of the signal generation process and an observation. The Kalman filter has many applications, for example it is widely used in process control, in radio and other communication devices particularly as phase-lock-loop systems, in GPS position tracking and guidance systems and in signal de-noising and system identification problems.

MMSE. MMSE estimation of the short time spectral amplitude (STSA); its structure is the same as that of spectral subtraction but in contrast to the Wiener filtering motivation of spectral subtraction, it optimizes the estimate of the real rather than complex spectral amplitudes. Central to their procedures is the estimate of SNR in each frequency bin for which they proposed two algorithms: a maximum likelihood approach and a decision directed approach which they found performed better. The maximum likelihood (ML) approach estimates the SNR by subtracting unity from the low-pass filtered ratio of noisy-signal to noise power and half-wave rectifying the result so that it is non-negative. The decision-directed approach forms the SNR estimate by taking a weighted average of this ML estimate and an estimate of the previous frame's SNR determined from the enhanced speech. Both algorithms assume that the mean noise power spectrum is known in advance.

B. Multi-Channel Enhancement Techniques

Multi-Channel systems take advantage of the availability of multiple signal inputs to the system and make use of the noise reference in an adaptive noise cancellation device, the use of phase alignment to reject undesired noise components, or even the use of phase alignment and noise cancellation stages into a combined scheme [7]. By taking into account the spatial properties of the signal and the noise source, the limitations inherent to one-channel systems, particularly non-stationary of noises can be better addressed. These systems tend to be more complex.

Adaptive Noise Cancellation. Adaptive noise cancellation is a powerful speech enhancement technique based in the availability of an auxiliary channel, known as reference path, where a correlated sample or reference of the contaminating noise is present. This reference input will be filtered following an adaptive algorithm, in order to subtract the output of this filtering process from the main path, where noisy speech is present. The adaptive noise cancellation cancels the primary unwanted noise by introducing a canceling anti-noise of equal amplitude but opposite phase using a reference signal. This reference signal is derived from one or more sensors located at points near the noise and interference sources where the interest signal is weak or undetectable.

Multisens or Beam forming. Beam forming is a multiple-input and single-output (MISO) application and consists of multichannel advanced multidimensional (space-time domain) filtering techniques that enhance the desired signal as well as suppress the noise signal. In beam forming, two or more microphones are arranged in an array of some geometric shape. A beam-former is then used to filter the sensor outputs and amplifies or attenuates the signals depending on their direction of arrival. The underlying idea of this scheme is based on the assumption that the contribution of the reflexions is small, and that we know the direction of arrival of the desired signal. Then, through a correct alignment of the phase function in each sensor, the desired signal can be enhanced, rejecting all the noisy components not aligned in phase.

Multi-Input Multi-Output. Multi-input speech enhancement systems include adaptive beam forming, adaptive noise cancellation, multi-input multi-output (MIMO) teleconferencing systems, stereophonic echo cancellation and in-car MIMO communication systems. In a typical multi-input speech enhancement system, there are several microphones. The output of each microphone is a mixture of the speech signal, feedback from loudspeakers, speech reflections from walls and noise. Assuming that there are M microphones and N sets of signal and noise sources, there are N×M different acoustic channels between the sources of signals and the microphones. We can write a system of linear equations to describe the relationship between the signals emitted from different sources and the signals picked up by the microphones.

III. REVIEW ON PREVIOUSLY MODIFIED SPEECH ENHANCEMENT WORK

In this section we discuss the related work of noise reduction process based on spectral and frequency domain. The process of reduction improved the quality of speech. The reduction of noise used in recognition process and improved the recognition ratio of speech. Some work discuss here used in current research trend.

In this paper Wei Shi [8] proposed 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. Simulative comparisons with four standardized VAD algorithms (ETSI adaptive multi-rate option 1 and 2, ETSI advanced front-end and G.729 Annex B) as well as three former proposed VAD algorithms show that the proposed LTACS-based VAD algorithm achieves the best performance under all SNR conditions, especially in strong noisy environments.

In this paper Saldanha [9] improve the intelligibility and quality of the speech signal. Speech signal when recorded in the noisy environment degrades the clean speech signal. Speech enhancement reduces the noise without distorting the original signal. In this concept, the method of Multi- Band Spectral Subtraction (MBSS) is used to enhance the noisy speech signal. The fact that the speech spectrum is affected by colored noise differently at various frequencies is taken into account. After simulation the properties are derived in terms of input and output SNR.

In this paper Ming [10] analyzed Conventional speech enhancement methods, based on frame, multi-frame or segment estimation and after presents a new method which aims to reduce or effectively remove this requirement. By using the Zero-mean Normalized Correlation Coefficient (ZNCC) as the comparison measure, and by extending the effective length of speech segment matching to sentence long speech utterances, it is possible to obtain an accurate speech estimate from noise without requiring specific knowledge about the noise. It could be used to deal with unpredictable noise or noise without proper training data. For evaluation, they build new system on one training data set and test it on two different test data sets across two databases, for a range of different noises including highly nonstationary ones. It is shown that the new approach, without any estimation of the noise, is able to significantly outperform conventional methods.

In this paper Amol [11] 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. In this paper Schasse [12] proposed to use the minimum-variance distortion-less response (MVDR) approach in single-channel speech enhancement in the short-time frequency domain. By applying optimal FIR filters to each sub-band signal, these filters reduce additive noise components with less speech distortion compared to conventional approaches. An important ingredient to these filters is the temporal correlation of the speech signals. They derive algorithms to provide a blind estimation of this quantity based on a maximum-likelihood and maximum *a-posteriori* estimation. To derive proper models for the inter-frame correlation of the speech and noise signals, they investigate their statistics on a large dataset. Results show significantly.

IV. CONCLUSION

Main objective of speech enhancement is to improve the perceptual aspects of speech such as overall quality, intelligibility, or degree of listener fatigue. The speech signal is degraded due to various types of noise. In this paper we have presented a review of various speech enhancement techniques by classifying them into two main categories: single-channel speech enhancement methods and multi-channel speech enhancement methods. Speech enhancement aims to improve speech quality by using various algorithms. Speech enhancement not only involves processing speech signals for human listening but also for further processing prior to listening.

REFERENCES

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

[2]. 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, 2013.

[3]. Shimpy Rai and J. Singh (2015). Low Power, Noise-Free 4/5 Prescalar Using Domino Logic. *International Journal of Electrical, Electronics and Computer Engineering,* **4**(2): 154-161.

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

[5]. 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).

[6]. 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

[7]. 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.

[8]. Jesper B. Boldt1, Daniel P. W. Ellis, (2009). "A Simple Correlation-Basedmodel of Intelligibility For Nonlinear Speech Enhancement and Separation" EURASIP/EUSIPCO, pp. 1849-1853.

[9]. 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.

[10]. Wei Shi, Yuexianzou, Yiliu, (2014). "Long-Term Auto-Correlation Statistics Based Voice Activity Detection for Strong Noisy Speech", IEEE/China SIP, Vol. **2**(14), pp. 100-104, 2014.

[11]. 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.

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

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

[13]. 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.