Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 3, Issue. 8, August 2014, pg.285 – 290

REVIEW ARTICLE

Review on Speech Enhancement Techniques

[1]Sunita Dixit, [2]Dr. MD Yusuf Mulge
[1]Research Scholar, Pacific University, Udaipur
[2]Principal, PDM College of Engineering for Women, Bahadurgarh
bhardwajsunita23@gmail.com

Abstract: Speech is most effective and natural medium to exchange the information among people. The speech system also goes under the process of some software tools to present the information in effective way. This kind of effectiveness is defined as the translation of the information to text or some other speech form. 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. Main objective of speech enhancement is to improve the perceptual aspects of speech such as overall quality, intelligibility, or degree of listener fatigue. In this paper, we present review of various speech enhancement techniques.

I. INTRODUCTION

A major part of the interaction between humans takes place via speech communication. Hence, research in speech and hearing sciences has been going on for centuries to understand the dynamics and processes involved in the production and perception of speech. The field of speech processing is essentially an application of signal processing techniques to acoustic signals using the knowledge offered by researchers in the field of hearing sciences. The speech is the most interactive way to communicate between the humans. This communication can either direct or distance through some electronic medium. Because of this lot of research is been done in area of speech and the hearing science. In speech communication, the speech signal is always accompanied by some noise[1]. The speech signal degradations may be attributed to various factors; viz. disorders in production organs, different sensors (microphones) and their placement (hands free), acoustic non-speech and speech background, channel and reverberation effect and disorders in perception organs. Considerable research recently has examined ways to enhance speech, mostly related to speech distorted by background noise (occurring at the source or in transmission)-both wideband (and usually stationary) noise and (less often) narrowband noise, clicks, and other non-stationary interferences. The goal of speech enhancement is to enhance quality and intelligibility. Except when inputs from multiple microphones are available (in some specially arranged cases), it has been very difficult for speech enhancement systems to improve intelligibility. Thus most speech enhancement methods raise quality, while minimizing any loss in intelligibility[2]. As observed, certain aspects of speech are more perceptually important than others. The auditory system is more sensitive to the presence than absence of energy, and tends to ignore many aspects of phase. Thus speech enhancement algorithms often focus on accurate modeling of peaks in the speech amplitude spectrum, rather than on phase relationships or on energy at weaker frequencies. Voiced speech, with its high amplitude and concentration of energy at low frequency, is more perceptually important than unvoiced speech for preserving quality. Hence, speech enhancement usually emphasizes improving the periodic portions of speech. Good representation of spectral amplitudes at harmonic frequencies and especially in the first three formant regions is paramount for high speech quality. All enhancement algorithms introduce their own distortion and care to be taken to minimize distortion.

II. SPEECH ENHANCEMENT TECHNIQUES

There are many ways to classify speech enhancement methods. It is usually difficult for a typical algorithm to be able to perform homogenously across all noise types. Therefore, usually a speech enhancement system is based on certain assumptions and constraints that are typically dependent on the application and the environment[3]. In general the performance of a speech enhancement algorithm is limited by the following factors: limitations on the number of noise sources available, making different uses of a priori information about the signal of interest and/or the corrupting signal, limitations in the time variations (non-stationary) allowed for the corrupting signal, model based limitations like the restriction of the algorithm to uncorrelated noise. The approach to speech enhancement varies considerably depending upon type of degradation.

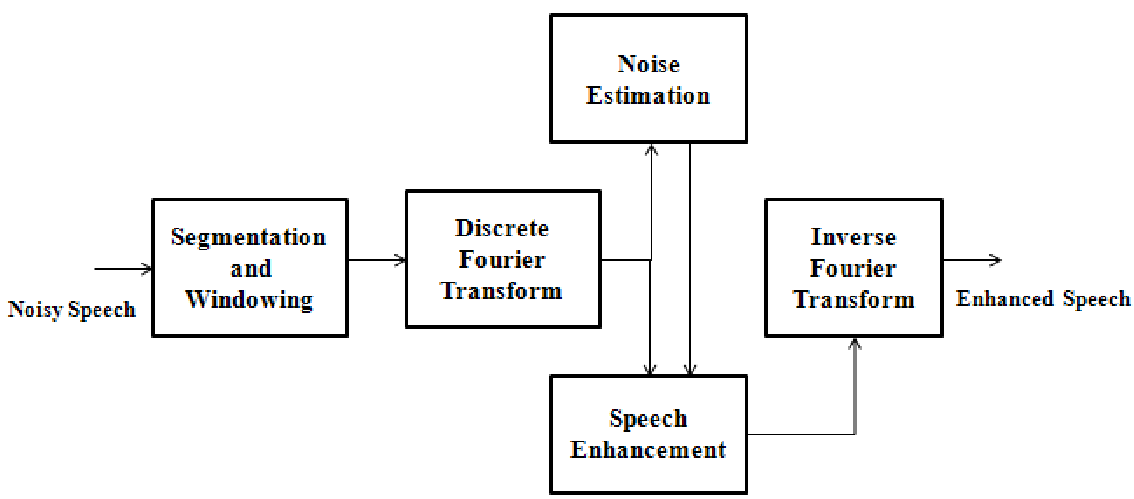

Figure 2.1: Block diagram of speech enhancement

Basic block diagram of speech enhancement is shown in figure.1.In this process first speech signal is segmented for 20-30ms i.e. short term Fourier transform (STFT) is taken and windowed. Hamming window is used for windowing. Then Discrete Fourier transforms (DFT) or Fast Fourier Transform (FFT) of segmented and windowed. Generally in speech enhancement Fast Fourier Transform (FFT) is used. Noisy speech signal is taken. FFT of noisy signal is then given to noise estimation block and speech enhancement block. Noise estimation block estimate the noise during the speech pauses and find the noise spectrum. In most speech enhancement algorithms, it is assumed that an estimate of the noise spectrum is available. The noise estimate can have a major impact on the quality and intelligibility of the enhanced signal[4]. If the noise estimate is too low, unwanted residual noise will be audible, if the noise estimate is too high, speech will be distorted.

Speech enhancement techniques can be divided into two basic categories:

(i) Single channel and

(ii) Multiple channels (array processing)

based on speech acquired from single microphone or multiple microphone sources respectively.

III. SINGLE CHANNEL SPEECH ENHANCEMENT TECHNIQUES

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[5]. Therefore, there is a tradeoff between quality and intelligibility. These the most common real-time scenario algorithms e.g. mobile communication, hearing aids etc. as usually a second channel is not available in most of the applications. 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.

Single-channel speech enhancement techniques are:

A. Spectral Subtraction Method: (SS)

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. 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.

B. Spectral Subtraction with Over subtraction Model: (SSOM)

SSOM procedure was introduced in order to compensate for the ―musical noise effect. It reduces the perception of musical noise. This Method consists of subtracting an overestimate of the noise power spectrum and presenting the resultant spectral components from going below a preset minimum spectral floor value.

C. Non-Linear Spectral Subtraction: (NSS)

NSS is based on combining two different ideas: i) The use of an extended noise and an over subtraction model ii) Non-linear implementation of the subtraction process, taking into account that the subtraction process must depend on the SNR of the frame, in order to apply less subtraction with high SNRs and vice versa[6]. It gives modification in Spectral Subtraction method by making over subtraction factor (α) frequency dependent and subtraction process non-linear. Assumption for NSS is that noise does not affect all spectral component equally. Certain type of noise may affect the low frequency region of spectrum more than high frequency region. NSS suggests the use of a frequency dependent subtraction factor for the different types of noise. Due to the frequency dependent subtraction factor process becomes nonlinear. In NSS larger values are subtracted at frequencies with low SNR level and smaller values are subtracted at frequencies with high SNR level.

IV. MULTI-CHANNEL ENHANCEMENT TECHNIQUES

These 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. 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[7]. These systems tend to be more complex. Multi-channel enhancement techniques are:

A. Adaptive Noise Cancellation

Adaptive noise cancellation is a powerful speech enhancement technique [8] 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* (ANC) cancels the primary unwanted noise $r(n)$ 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.

B. Multisensor Beamforming

Beamforming 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 beamforming, two or more microphones are arranged in an array of some geometric shape. A *beamformer* is then used to filter the sensor outputs and amplifies or attenuates the signals depending on their *direction of arrival* (DOA). 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.

V. CONCLUSION

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. Main objective of speech enhancement is to improve the perceptual aspects of speech such as overall quality, intelligibility, or degree of listener fatigue. 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.

REFERENCES

[1] Yi Hu and Philipos C. Loizou, "Subjective comparison and evaluation of speech enhancement algorithms" IEEE Trans.Speech Audio Proc.2007:49(7): 588–601.

[2] Kim W, Kang S, and ko H.(2000) "Spectral subtraction Based on phonetic dependency and masking effects" IEEE Proc.vision image signal process, 147(5),pp423-427

[3] Phillips C Loizou "Speech enhancement theory & practice" 1st ed. Boca Raton, FL.: CRC, 2007. Releases Taylor Francis.

[4] S.F. Boll, "Suppression of acoustic noise in speech using spectral subtraction", *IEEE Transactions on Acoustics, Speech and Signal Processing,* Vol. 27, No. 2, pp. 113-120, 1979.

[5] Yariv Ephraim, Hanoch Lev-Ari and William J.J. Roberts "A Brief Survey of Speech Enhancement" *IEEE Sig. Proc. Let*., vol. 10,pp. 104-106, April 2003 s.

[6] Yoon. B-Y.; Tashev, I. & Acero, A. (2007) Robust Adaptive Beamforming Algorithm Using Instantaneous Direction Of Arrival With Enhanced Noise Suppression Capability. *IEEE International Conference on Acoustics, Speech and Signal Processing* 1:I-133–I-136.

[7] Reddy, D.R, "Speech recognition by machine: A review", Proceedings of IEEE (Volume: 64, Issue: 4) ISSN: 0018-9219.

[8] Young, S.J, "Robust continuous speech recognition using parallel model", IEEE Transactions on Speech and Audio Processing (Volume: 4, Issue: 5).

[9] Bassam A. Q. Al-Qatab , Raja N. Ainon, "Arabic Speech Recognition Using Hidden Markov Model Toolkit(HTK)", 978-1-4244-6716-711 0/$26.00 ©2010 IEEE.

[10] Chibelushi, C.C. "A review of speech-based bimodal recognition", 2002 IEEE Transactions on Multimedia, (Volume: 4, Issue: 1).

[11] Iosif Mporas,"Dynamic Selection of A Speech Enhancement Method For Robust Speech Recognition In Moving Motorcycle Environment", ICASSP 2011 978-1-4577-0539-7/11©2011 IEEE.

[12] Ciira wa Maina,"Joint Speech Enhancement and Speaker Identification Using Approximate Bayesian Inference", IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING 1558-7916© 2010 IEEE