Speech Enhancement Techniques: A Review

Sandeep Kumar*, Bittu Kumar, Neeraj Kumar****

Abstract

Speech enhancement is a method of improvement in perceptual quality and intelligibility of the speech signal. Speech enhancement is the most important field of the speech processing and used for many applications such as telecommunication, VoIP, speech recognition, hearing aids, etc. In this paper, a review of speech enhancement techniques has been presented. The details provided in this paper will be of interest to those researchers who work in the field of Digital Signal Processing and Speech processing.

Keywords: Speech Processing, Single Channel and Multi-Channel Speech Enhancement, Linear and Non-Linear Spectral Subtraction, Over-Subtraction Factor

INTRODUCTION

Speech is a primary medium of communication between human beings. Now a day's peoples can communicate with each other with the help of telecommunication systems which involves transmission and reception of speech signals from one end to another end. During this communication between transmitter and receiver, speech signal is always affected by some noise [1] (e.g. channel noise, background noise, noise due to microphones etc.). The involvement of noise in the received speech signal leads to reduce the listening quality and intelligibility of the speech signal. The speech enhancement is a technique by which intelligibility and/or overall perceptual quality of the speech signal can be improved. Improving the intelligibility of the speech signal is a difficult and hence most of the available enhancement methods/systems improve the speech quality with minimizing any loss in the intelligibility [2]. Speech enhancement is usually done by improving the spectral peaks at the periodic portion of speech signal.

Several methods such as spectral subtraction approach, Minimum Mean Square Error-Short Time Spectral Amplitude (MMSE-STSA) estimator, the signal subspace approach, adaptive noise canceling and the iterative Wiener filter [3–6], have been proposed for the speech enhancement in the last few decades. Based on speech acquired from single or multiple sources, these methods can be classified into two categories: single channel and multiple channels. In this paper, a review on these speech enhancement techniques has been presented in section II. Finally, concluding remarks for the review is presented in section III.

SPEECH ENHANCEMENT TECHNIQUES

As stated above, speech enhancement is a technique to enhance the perceptual quality of the speech signal. Speech enhancement system is based on certain assumptions and constraints. These assumptions and constraints may vary with application and environment [3].

A block diagram of basic speech enhancement system is shown in Fig.1. In this, the degraded speech signal is framed into 20 to 30 ms and then passed through the hamming window (half-overlapped). After this, FFT is computed to convert the time domain windowed signal into spectral domain signal. The output of *FFT* block is connected to the input of *Noise Estimate* block and *Speech Enhancement* block. *Noise Estimation* block is used to

* Associate Professor, Dept. of ET&T Rungta College of Engineering & Technology, Bhilai, Chhattisgarh, India, Email: sandy_alld@yahoo.com

** Research Scholar, Dept. of Electronics Engg., Indian School of Mines, Dhanbad, Jharkhand, India, Email: kumarbittu135@gmail.com, kumarneeraj.mit@gmail.com

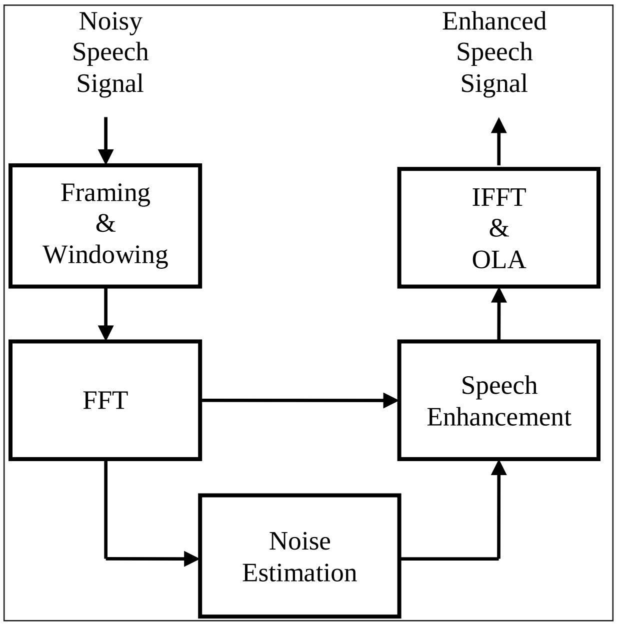

Fig.1. Block Diagram of Speech Enhancement System

find the noise spectrum by estimating the noise during the speech pauses. *Speech Enhancement* block is used to enhanced the noisy speech signal in frequency domain (using chosen algorithm) using the noise estimates and spectral of the windowed noisy speech signal. Finally, enhanced spectrum is passed through the *IFFT & OLA (Overlap-Add)* block which results the enhanced speech in time domain. Quality and intelligibility of enhanced speech are affected by noise estimate [7].

Speech enhancement techniques can be divided into two categories: single channel and multiple channels.

Single Channel

The single channel enhancement system improves the quality of degraded speech signal on the cost of some intelligibility [8]. Due to a single channel, these systems are less expensive and easy to build as compared to the multiple channel systems. Single channel systems rely on the assumption that noise is stationary during speech interval, and hence performance is limited in the presence of non-stationary noise and degraded with lower SNR. Single channel enhancement techniques have carried lots of methods (*e.g. spectral subtraction, MMSE-STSA, MMSE-LSA, wiener filtering*) for separating the speech signal from degraded speech signal with the help of estimated noise which is calculated by noise estimation methods [9]. Among all the spectral subtraction is very simple and easy for implementation. Due to this reason, it received more attention from different authors.

Spectral subtraction method uses the principle that the spectrum of original speech can be obtained by subtraction of degraded speech spectrum and estimated noise spectrum. Noise spectrum can be estimated and updated at the time when signal (or noise) is absent (or present). In this method, it is assume that the noise is additive and stationary and its spectrum does not change between the updating periods. In spectral subtraction method, we subtract the magnitude of the degraded speech spectrum from noise spectrum for obtaining the desired speech signal. If the magnitude of the degraded speech spectrum is too much subtracted from noise spectrum, then information of real speech might be removed, whereas if degraded speech is little subtracted then effected noise is available as it is in obtained speech signal. To solve this problem we can use one parameter which is called '*over-subtraction factor*'.

In spectral subtraction with over subtraction method the spectrum of clean speech can be obtained by subtracting the resultant spectral component and an overestimate of noise power spectrum. Spectral subtraction methods are used either for magnitude or power of degraded speech spectrum. Due to over subtraction factor and negative spectral components (*it comes if magnitude/power of degraded speech signal is less than noise spectrum*), a new noise introduce in enhanced speech spectrum which is known as 'musical noise'. Musical noise is one of the major limitations of spectral subtraction methods. For reducing the musical noise, one more factor is introduced in spectral subtraction method that is '*spectral floor factor*'. A generalized equation of spectral subtraction is:

$$|\bar{E}_i(k)| = \begin{cases} \beta^{\frac{1}{\gamma}} D_i(k), & if\,|S_i(k)| < (\alpha-\beta)^{\frac{1}{\gamma}} D_i(k) \\ [\,S_i(k)^{\gamma} \quad \alpha(D_i(k))^{\gamma}]^{1/\gamma}, & otherwise \end{cases}$$

where, $E_i(k)$ –Enhanced Speech Spectrum

$D_i(k)$ –Noise spectrum

$S_i(k)$ –Degraded Speech Spectrum

α - Over Subtraction Factor (α = 2 to 6)

β - Spectral Floor Factor (β = 0.002 approx.)

γ - Exponent Factor (*1-magnitude spectral subtraction, 2- power spectral subtraction*)

In spectral subtraction, over-subtraction factor is fixed. If we can vary the α then it is called as non-linear spectral subtraction. In non-linear spectral subtraction, it is assume that the effect of noise on the entire spectral component is not equal (low frequency region may be more affected by noise as compared to high frequency region). This method is the modification of spectral subtraction method. Modification has been made by making the subtraction factor frequency dependent which results the non-linear subtraction process. In this method, larger/smaller values are subtracted with frequency having low/high SNR.

Multiple Channels

These systems have more than one input signals. For adaptive noise cancellation and rejection of undesired noise component, these systems use reference noise signal and phase alignment, respectively. Due to multiple channels, these systems are more complex. Multiple channel speech enhancement techniques are: adaptive noise cancellation and multisensory beamforming.

In adaptive noise cancellation technique [8], an auxiliary channel is available for the reference noise. This reference signal will be filtered and then subtracted by adaptive algorithm with degraded speech signal. To cancel the unwanted noise, adaptive noise cancellation introduces an anti-noise signal using reference noise signal (equal amplitude, opposite phase). The reference noise signal can be obtained from sensors located at point near the noise and interference sources.

Multisensory beamforming is a multiple input and single output system. To enhance the desired signal, this system use multichannel multidimensional filtering techniques. The system has two or more microphones. A beamformer is used to filter and amplify/attenuate the sensors output depending on their direction of arrival. This system works on the assumption that the direction of arrival of desired signal is known.

CONCLUSION

Speech enhancement is a process by which the perceptual quality and intelligibility of speech signal can be improved. In this paper, different speech enhancement techniques into two major categories (single channel and multiple channels) have been reviewed. The single channel speech enhancement systems are less expensive and easy to build as compared with the multiple channel speech enhancement systems. Most of the available speech enhancement systems improve the speech quality with minimizing any loss in the intelligibility.

REFERENCES

Hu, Y., & Loizou, P. C. (2007). Subjective comparision and evaluation of speech enhancement algoriths. *IEEE Transaction on Speech and Audio Processing*, 49(7), 588-601.

Kim, W., Kang, S., & Ko, H. (2000). Spectral subtraction based on phonetic dependency and masking effects. *IEEE Proceedings, Vision Image And Signal Processing*, 147(5), 423-427.

Boll, S. F. (1979). Suppression of acoustic noise in speech using spectral subtraction. *IEEE Transaction on Acoustics, Speech and Signal Processing*, 27(2), (pp. 113-120).

Berouti, M., Schwartz, R., & Makhoul, J. (1979). Enhancement of speech corrupted by acoustic noise. *Proc. IEEE Int. Conf. Acoust., Speech Signal Processing*, 208-211.

Ephriam, Y., & van Trees, H. L. (1993). A signal subspace approach for speech enhancement. *Proc. International Conference on Acoustic, Speech and Signal Processing*, 2, 355-358, Detroit, MI, USA.

Lim, J. S., & Oppenheim, A. V. (1978). All-pole modelling of degraded speech. *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP-26.

Phillips C. L. (2007). Speech enhancement theory & practice. *1st ed. Boca Raton, FL. CRC, Taylor Francis.*

Yariv, E., Hanoch, L. A., & William, J. J. R. (2003). A brief survey of speech enhancement. *IEEE Signal Processing Letter*, 10, 104-106.

Kumar, B. (2015). Spectral Subtraction using Modified Cascaded Median based Noise Estimation for Speech Enhancement. *ACM, 6th International Conference on Computer and Communication (ICCCT-15)*, MNNIT, Allahabad, India.