www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 4 Issue 5 May 2015, Page No. 12143-12145

A Survey on Speech Enhancement Techniques

Rohini Kochar [1] **, *Ankur Singhal*** [2] **, *Anil Garg*** [3]

[1] M.Tech Scholar,Geeta Institute of Management and Technology, Kurukshetra
rohinikochar12@gmail.com

[2] Astt.Professor,Geeta Institute of Management and Technology, Kurukshetra
ankursinghal071987@gmail.com

[3] Astt.Professor,Maharishi Markandeshwar University, Mullana

mmuanilgarg@gmail.com

Abstract: ***Speech is the best way to express one's thoughts. The better quality of speech helps in better communication. Speech enhancement is the method to improve the quality of speech by using algorithms. The main aim of the speech enhancement techniques is to provide noiseless communication. This paper discusses the various methods used for improving the quality of speech.***

Keywords: Speech, noise estimation, optimization, speech enhancement.

1. Introduction

Speech is the most natural and important means of communication for humans [2]. Speech can be defined as a mean of delivering thoughts and ideas with the help of vocal sounds [3]. In general, speech production uses a number of organs and muscles such as lungs, larynx and vocal tract [1]. The shape of vocal tract is varied to produce different speech signals according to vocal cords vibrations [3]. The position of tongue, teeth, lips and jaws are used to change the shape of vocal tract [1]. Vocal tract is situated in larynx called Adam's apple [3].

A human ear can hear the sounds with frequencies in between 20 Hz to 20 KHz [3]. The signals with frequencies above 20 KHz are called ultrasonic sounds and below 20 Hz are called infrasonic sounds [3]. The speech signals are usually distorted by background noise and competing interfering signals, which lies in human ear hearing frequency range [2,3]. The main aim of speech enhancement is to remove noise from noisy speech signal keeping the speech component and to reduce the distortion of speech [4]. The quality and intelligibility of the distorted speech determine the performance of these techniques [3].

The speech enhancement techniques can be divided into three categories [5].

1. Time-domain methods

These methods provide trade-off between speech distortion and residual noise. Subspace approach is an example of time-domain methods.

2. Frequency-domain methods

These methods provide the advantage of real-time processing with less computational load. Frequency-domain methods include the spectral subtraction, minimum mean square error (MMSE) estimator and the Wiener filtering.

3. Time-Frequency domain methods

These methods involve the employment of family of wavelets.

Several other techniques can also be used for enhancing the quality and intelligibility of speech such as binaural speech enhancement systems [2], masking techniques [7], voice-conversion based methods [16], Kalman filtering [19], mutual information based methods [11], etc.

The rest of this paper is organized as follows. Section II describes the literature review of some paper based on speech enhancement. Future enhancement and conclusion of this topic has been described in Section III.

1. Literature review

There are several techniques used for the speech enhancement.

Ching-Ta Lu et. al. [6] proposed a single channel speech enhancement method with the use of perceptual decision-directed (TSDD) approach. The two-step decision-directed approach is used to improve the accuracy of approximated speech spectra. This method can also be used to enhance the performance of TSDD approach. Experimental results show that this method enhance the capability of perceptual method in removing the residual noise and also improve the speech quality.

V. Ramakrishnan et. al. [7] introduced a two-stage method to solve the speech enhancement problem in real noisy world. This method comprises of general spectral subtraction method followed by a series of perceptually motivated post-processing algorithms. Subtraction step removes the additive noise but adds some spectral artifacts which are removed by post-processing step. Test results show that performance is effective at SNR greater than 0 db.

A. Narayanan et. al. [8] introduced a SNR estimation system which is based on computational auditory scene analysis (CASA). It is a binary masking scheme. This method cannot be used for short-time SNR estimation. This method involves autocorrelation computation and envelope extraction at each T-F unit. Results of different experiments show that the proposed method works better than other long-term SNR estimation algorithms.

N. Yousefian et. al. [9] proposed a coherence-based dual microphone method for estimation of SNR. This technique can be used for hearing aids and cochlear implant devices. Different experiments have been conducted in different conditions. The results show that the proposed method gives significant performance in anechoic and mildly reverberant conditions.

N. Madhu et. al. [10] attempted to define a so called binary mask as the objective of binary mask estimation. Here, it is shown that methods using binary masks are able to improve the intelligibility at low SNR values. For relevant results, a low spectral resolution, modeled using the Bark-spectrum scale is to be used. The performance of IBM and IWF has compared. Intelligibility test shows the higher intelligibility values of IWF than IBM.

J. B. Crespo et. al. [11] presented a method for speech reinforcement in a case where there are many play back regions. In such a case, signals from one region go to other resulting in degradation of speech intelligibility. A smooth distortion is used to improve the quality or intelligibility. Results show the advantages of multizone processing over the iterated application of single zone algorithm.

J. Jensen et. al. [12] proposed a method based on mutual information for estimation of average intelligibility of noisy and processed speech signal. This method estimates the mutual information by comparing the critical-band amplitude envelopes of noisy or processed speech signal because mmse can be considered as an indicator for the intelligibility of noisy speech. Simulation results show that the proposed method can predict the intelligibility of speech distorted by both stationary and non-stationary noises.

D. P. K. Lun et. al. [13] presented an improved speech enhancement algorithm based on a novel expectation-maximization (EM) framework. The traditional TCS method is used to initiate the algorithm. The method uses the sparsity of speeches in the cepstral domain. The method performs well when the speech is distorted by the non-stationary noises. Experimental results show that the proposed method outperforms other methods for voiced speech. But for unvoiced case different algorithm along with proposed algorithm is needed.

Seon M. Kim et. al. [14] presented a method for target speech estimation by considering the spatial cues in noisy environments. In this method, SNR is estimated by using the phase difference acquired from dual-microphone signals. As direction-of-arrival (DOA) of target signal is related to the phase difference between multiple microphone signals, so DOA-based SNR is estimated in this method. The performance of proposed method is evaluated in terms of SDR, SIR and SAR. Results show that the Wiener filter using the proposed DOA based SNR estimation performs better than other speech enhancement methods.

R. Bendoumia et. al. [15] introduced two new two channel VSS- FB algorithms for speech enhancement and noise reduction. Least-mean-square (LMS) algorithms have been used with two BSS structures. The two proposed methods are based on optimal step-size estimation with the use of decorrelation criteria. Various experiments have been conducted to check its performance. Final result shows that the proposed 2C-VSSB algorithm is better that 2C-VSSF algorithm because of no need of post filtering correction at output in 2C-VSSB.

M. Ahangar et. al. [16] described voice conversion method using spectral features to improve the speech quality and speaker individuality. Here, four spectral features have studied. Comparative study shows that out of four features, the cepstral features are more suitable for clustering and all-pole features for the analysis/synthesis stage. Each voice conversion consists of three stages: analysis, manipulation and synthesis. Experiment results show that the feature combination method performs better than the original methods.

Sanjay P. Patil et. al. [17] presented a noise estimation method based on spectral sparsity. Speech signal may not be present at all time of communication. Here, the noise is estimated in voiced as well as unvoiced frames. This method can be combined with any traditional speech enhancement method to further improve the performance. The performance of this method has been studied under different types of noise. Results show the better performance of the

proposed method over Martin's and MCRA noise estimation methods.

J. Liu et. al. [18] proposed a practical method for Automatic speech recognition (ASR) in multiple reverberant environments. Here, a multi-model selection method is used to train multiple speech recognisers identified by reverberation time. For estimation of the performance of these models, a Pioneer mobile robot equipped with a binaural microphone is used for different set room IRs. Results show that the attenuated IR model gives the best performance over the non-ideal models. Using this model, real-time processing is confirmed.

T. Mellahi et. al. [19] proposed a new iterative Kalman filtering scheme for speech enhancement distorted by AWGN and colored noises. It has been seen that the formant enhancement method (FEM) improves the formant structure of speech distorted by both types of noise. For objective results, PSEQ, SNR and SegSNR values are compared under various types of noises. In all cases, the result shows better performance.

2. Future scope and Conclusion

Various methods have been studied in literature review used for speech enhancement. There may be many other methods too. But these all methods have some advantages and disadvantages. Some techniques perform well in stationary noise environment while some in non-stationary. The future work is to develop an optimization method for improving the quality and intelligibility of speech signal.

References

[1] Philipos C.Loizou, "Speech Enhancement: Theory and Practice", 1st ed. Boca Raton, FL: CRC, Taylor and Francis, 2007.

[2] J. Li, S. Sakamoto, S. Hongo, M. Akagi and Y. Suzuki, "Two-stage binaural speech enhancement with Wiener filter for high-quality speech enhancement", in Speech Communication, vol. 53, Issue 5, pp. 677-689, 2011.

[3] Savita Hooda and Smriti Aggarwal, "Review of MMSE Estimator for Speech Enhancement", in IJARCSSE, vol. 2, Issue 11, pp. 419-424, November 2012.

[4] F. Deng, F. Bao and Chang-chun Bao, "Speech enhancement using generalized weighted β-order spectral amplitude estimator", in Speech Communication, vol. 59, pp. 55-68, 2014.

[5] Md.T. Islam, C.Shahnaz and S.A Fattah, "Speech Enhancement based on a modified spectral subtraction method", in MWSCAS, pp. 1085-1088, 2014.

[6] Ching-Ta Lu, "Enhancement of Single Channel Speech using Perceptual-Decision-Directed Approach", in C.-T. Lui, Speech Comm., pp. 495-507, December 2010.

[7] Vyass Ramakrishnan, Karthik Shetty, Pawan Kumar G and Chandra Shekhar Seelamantula, "Efficient Post-Proceesing Technique for Speech Enhancemnt", in NCC, pp. 1-5, 2011.

[8] Arun Narayanan and DeLiang Wang, "A CASA-Based System for Long-Term SNR Estimation", IEEE Trans. Audio, Speech, Lang. Process., vol. 20, no. 9, pp. 2518-2527, November 2012.

[9] Nima Yousefian and Philipos C.Loizou, "A Dual-Microphone Algorithm that can Cope with Competing-Talkers Scenarios", IEEE Trans. Audio, Speech, Lang. Process., vol. 21, no. 1, pp. 145-155, January 2013.

[10] Nilesh Madhu, Ann Spriet, Sofie Jansen, Raphael Koning and Jan Wouters, "The Potential for Speech Improvement using the Ideal Binary Mask and the Ideal Wiener Filter in Single Channel Noise Reduction Systems: Application to Auditory Prostheses", IEEE Trans. Audio, Speech, Lang. Process., vol. 21,no. 1, pp. 63-72, January 2013.

[11] Joao B.Crespo and Ricahrd C.Hendriks, "Multizone Speech Reinforcement", IEEE Trans. Audio, Speech, Lang. Process., vol. 22, no. 1, pp. 54-66, January 2014.

[12] Jesper Jensen and Cees H.Taal, "Speech Intelligibility Prediction based on Mutual Information", IEEE Trans. Audio, Speech, Lang. Process., vol. 22, no. 2, pp. 430-440, February 2014.

[13] Daniel P.K.Lun, Tak-Wai Shen and K.C.Ho, "A Novel Expectation-Maximization Framework for Speech Enhancement in Non-Stationary Noise Environment", IEEE Trans. Audio, Speech, Lang. Process., vol. 22, no. 2, pp. 335-346, February 2014.

[14] Seon Man Kim and Hong Kook Kim, "Direction-of-Arrival based SNR Estimation for Dual-Microphone Speech Enhancement", IEEE Trans. Audio, Speech, Lang. Process., vol. 22, no. 12, pp. 2207-2217, December 2014.

[15] Redha Bendoumia, Mohamed Djendi, "Two-channel variable-step-size forward and backward adaptive algorithms for acoustic noise reduction and speech enhancement", in Signal Processing, vol. 108, pp. 226-244, 2015.

[16] M. Ghorbandoost, A. Sayadiyan, M. Ahangar, H. Sheikhzadeh, A. S. Shahrebabaki, J. Amini, "Voice conversion based on feature combination with limited training data", in Speech Communication, vol. 67, pp. 113-128, 2015.

[17] Sanjay P. Patil, John N. Gowdy, "Use of baseband phase structure to improve the performance of current speech enhancement algorithms", in Speech communication, vol. 67, pp. 78-91, 2015.

[18] Jindong Liu, Guang-Zhong Yang, "Robust speech recognition in reverberant environments by using optimal synthetic room impulse response model", in Speech Communication, vol. 67, pp. 65-77, 2015.

[19] Tarek Mellahi, Rachid Hamdi, "LPC based formant enhancement method in Kalman filtering for speech enhancement", in AEU, vol. 69, pp. 545-554, 2015.