Survey on Efficient Signal Processing Techniques for Speech Enhancement

Adappa S Angadi
Dept. of ECE
Tontadarya College of Engg.
Gadag, India.

Adokshaja Kulkarni
Dept. of ECE
Tontadarya College of Engg.
Gadag, India.

Ravi A Gadad
Dept. of ECE
Tontadarya College of Engg.
Gadag, India.

K. Sridhar
Dept. of ECE
Basaveswar Engg. College
Bagalkot, India

***Abstract*— *Speech is most effective, efficient and natural medium to exchange the information among people. The speech system also goes through the process of some software tools to convey the information in effectual way. This kind of effectiveness is defined as the transformation of the information to signal, text or some other speech form. In speech communication, the speech signal is always accompanied by several noise. Therefore, speech enhancement is essential and it not only involves processing speech signals for human listening but also for further processing prior to listening. The speech signal can be enhanced using wavelet packet based decomposition methods including different thresholds that can reduce the noise efficiently. Main objective of the speech enhancement is to improve the perceptual aspects of speech such as overall intelligibility, quality and degree of listener fatigue. In this paper, we present review of various speech enhancement techniques.*

Keywords— *Speech signal, Signal Quality, Signal Intelligibility and Speech Enhancement Techniques.*

1. INTRODUCTION

A most important part of the interaction among the humans takes place through speech communication. Therefore, research in speech and hearing sciences has been going on for centuries to understand the processes and dynamics involved in the production and perception of speech. The field of speech processing is fundamentally an application of signal processing techniques to acoustic signals utilizing knowledge offered by researchers in the field of hearing sciences. The speech is the most interactive way to converse 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. The speech signal degradations may be attributed to various factors; viz. disorders in production organs, different sensors and their placement, 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 wideband noise and narrowband noise, clicks, and other non-stationary interferences. The main intension of speech enhancement is to intelligibility and quality. Except when inputs from multiple microphones are presented, it has been very complex for speech enhancement systems to improve intelligibility. Thus most speech enhancement technique raises quality, while minimizing any loss in intelligibility. 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 modelling 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.

2. MULTILEVEL WAVELET DECOMPOSITION:

Wavelet decomposition is a basic computer analysis technique used in signal processing. Multilevel wavelet decomposition provides us the information regarding frequency components present in the signal and helps to enhance the signal quality for further process.

Wissam A.Jassim et.al [01] has described the enhancement of speech signals ruined by various types of noise by employing Discrete Tchebichef Transform (DTT) and Discrete Krawtchouk Transform (DKT). These two techniques are originated from discrete orthogonal polynomials. The speech signal representations using a partial number of moment coefficients and their behaviour in domain of orthogonal moments are illustrated. The method involves elimination of noise from signal using a minimum-mean-square error in the domain of the DTT or DKT. According to comparisons with traditional methods, the initial outcomes gives promising results and show that orthogonal moments are applicable in the field of speech signal enhancement.

Kais Khaldi et.al [02] has presented a speech filtering technique to utilize combined effects of empirical mode decomposition (EMD) and the local statistics of the speech signal using the adaptive centre weighted average (ACWA) filter. The innovation lies in integrating frame class (voiced/unvoiced) in conventional filtering using the EMD and the ACWA filter. The speech signal is segmented into number of frames and each one is broken down using EMD into a finite number of intrinsic mode functions (IMFs). The quantity of filtered IMFs depends on whether frame is unvoiced or voiced. An energy criterion is employed to recognize voiced frames while a stationary index distinguishes between transient and unvoiced sequences. Experimental results achieved on signals corrupted by additive noise show that proposed filtering in line with the frame class is very efficient in removing noise components from noisy speech signal.

Marwa A et.al [03] has discussed about the adaptive Wiener filtering algorithm for enhancement of speech. This method lies on the adaptation of filter transfer function from sample to sample based on speech signal statistics; the local variance and the local mean. Rather than in frequency domain it is implemented in the time domain to accommodate for the time-varying nature of the speech signals. The proposed model is evaluated with the traditional spectral subtraction, frequency domain Wiener filtering and wavelet denoising methods using different speech quality metrics. The simulation results give the advantage of the proposed Wiener filtering method in the case of Additive White Gaussian Noise (AWGN) as well as coloured noise.

Table 1: Survey on Different Speech Signal enhancement Techniques

Title	Year	Algorithm	Advantage	Improvement
Enhancing Noisy Speech Signals using Orthogonal Moments [01]	2014	Discrete Tchebichef (DTT) and Discrete Krawtchowk Transform (DKT)	It removes noise from the signal using minimum mean square error.	Enhanced speech and removal of car noise.
Voiced/unvoiced speech classification based adaptive filtering of decomposed empirical modes for speech enhancement [02]	2015	Emperical mode decomposition (EMD) and Adaptive centre weighted average (ACWA) filter	Frame class is very effective in removing noise components from noisy speech signal.	Improves the performance of speech signal in terms of SNR and perceptual evaluation of speech quality (PESQ) score.
Speech enhancement with Adaptive wiener filter [03]	2014	Adaptive Wiener Filter	Methods depends on adaption of the filter transfer function from sample to sample speech signal statistics: local mean and local variance.	It reveals the superiority of weiner filtering methodin case of adaptive white Gaussian nose (AWGN) and coloured noise.
Comprehensive sensing based speech enhancement in non sparse noisy environments [04]	2013	Compressive sensing based speech enhancement in adverse environments (CS-SPEN)	It solve the theoretic difficulty of CS-SPEN on the treatment of non-sparse noise by using a relaxed upper bound for the constraint governing data consistency and a relaxed estimation error bound	It handles car internal and F16 cockpit noises.
Wavelet Speech enhancement based on Non-Negative Matrix Factorization [05]	2015	Discrete Wavelet Packet Transform (DWPT) and Non-Negative Matrix Factorization (NMF).	Splits time domain speech signal into subbands without distortion then it highlights the speech component for each band.	Provides noise corrupted speech with significant quality and intelligibility.

Dalei Wu et.al [04] has studied to resolve theoretic complexity of compressive sensing speech enhancement (CS-SPEN) on treatment of non-sparse noise by utilizing a relaxed upper bound for constraint governing data consistency and a relaxed estimation error bound. Their foremost outcome is mathematically proved. In addition, the effectiveness of the proposed method is executed by computational simulations, showing assured progresses to the previous method for both non stationary and stationary white Gaussian noises across diverse segmental signal-noise-ratios (SNRs). In these cases, employed technique is shown to have comparable results to state-of-the-art SE algorithms and some advantages over them at low SNRs. CS-SPEN without the sparse noise assumption works evenly with CS-SPEN with the sparse noise assumption for car internal and F16 cockpit noises.

Syu Siang Wang et.al [05] has proposed a novel speech enhancement algorithm that adopts the nonnegative matrix factorization (NMF) and discrete wavelet packet transform (DWPT) algorithm in order to conquer the aforementioned limitation. In brief, the DWPT is first applied to split a time domain speech signal into a sequence of sub band signals without adding any distortion. Then we utilize NMF to emphasize speech component for each sub band. Finally, the enhanced sub-band signals are integrated together via inverse DWPT to reconstruct a noise reduced signal in time domain. We experimented the proposed DWPTNMF based speech enhancement model on MHINT task. Experimental results show that this new method performs very well in prompting speech quality and intelligibility and it outperforms the conventional STFT-NMF based method.

3. WAVELET PACKET BASED DECOMPOSITION:

Wavelet packet based decomposition is most basic signal enhancement algorithm used in signal processing. It is mainly used to reduce the noisy components from input speech signal. There are different types techniques available for better signal enhancement. Survey of the various techniques associated speech signal analysis is explained below.

Ritwik Giri et.al [08] has proposed 2 advances to progress deep neural network (DNN) acoustic models for speech recognition in reverberant environments. Both techniques use auxiliary information in training the DNN but diverge in the type of information and manner in which it is used. The initial technique utilizes parallel training data formulti-task learning, in which network is trained to perform both a primary senone classification task and a secondary feature enhancement task using a shared representation. The second technique utilize a parameterization of reverberant environment extracted from observed signal to train a room-aware. The proposed approach obtained a word error rate of 7.8%on the SimData test set, which is lower than all reported systems using the same training data and evaluation conditions, and 27.5% on the mismatched Real Data test set, which is lower than all but two systems.

L Zao et.al [09] has presented a speech enhancement method for corrupted signals from non stationary acoustic noises. The proposed approach utilizes empirical mode decomposition (EMD) to the noisy speech signal and gets a sequence of intrinsic mode functions (IMF). The main role of proposed procedure is espousal of Hurst exponent in the selection of IMFs to reconstruct original speech. This EMD and Hurst-based (EMDH) approach is tested in speech enhancement experiments considering environmental acoustic noises with various indices of non stationarity. The results show that EMDH improves segmental SNR ratio and an overall quality composite measure, encompassing perceptual evaluation of speech quality (PESQ). Moreover, the short-time objective intelligibility (STOI) compute reinforces superior performance of EMDH. Finally, the EMDH is also inspects in a speaker identification task in noisy conditions. Highest speaker identification rates when compared to the baseline speech enhancement.

Yu Tsao et.al [10] has presented a generalized maximum a posteriori spectral amplitude (GMAPA) algorithm in designing a gain function for enhancement of speech signal. The proposed GMAPA algorithm vigorously mentions the weight of prior density of speech spectra with respect to SNR of testing speech signals to determine the optimal gain function. Whenever the SNR is high, GMAPA accepts a small weight to avoid overcompensations that may result in speech distortions. On the other hand, whenever the SNR is low, GMAPA utilizes a large weight to avoid disturbance of restoration caused by measurement noises. In The technique exposed to compute weight with consideration of time–frequency correlations that result in a more precise estimation of gain function. Results of subjective listening tests indicate that GMAPA gives significantly higher sound quality than other speech enhancement algorithms.

Wei Xue et.al [11] have derived a multi-channel Kalman Filter (MKF) that mutually utilizes both inter frame temporal correlation and inter channel spatial correlation for speech enhancement. The method presents LP in modulation domain, and by integrating spatial information, derives an optimal MKF gain in the short-time Fourier transform domain. They show that exposed MKF diminishes to conventional multichannel Wiener filter if LP information is discarded. Furthermore, they show that, under an appropriate supposition, the MKF is corresponding to a concatenation of the minimum variance distortion response beam former and a single-channel modulation-domain KF and therefore present an alternative implementation of MKF.

Kisoo Kwon et.al [12] has given a speech enhancement algorithm concatenating non-negative matrix factorization (NMF) and statistical models with on-line update of noise and speech bases. The statistical model-based enhancement techniques have been acknowledged to be less effective to non-stationary noises while template-based enhancement techniques can deal with them quite well. However, the template-based enhancement methods typically rely on *a priori* information. To defeat shortcomings of both approaches, the method adopted a novel speech enhancement technique that incorporating statistical model-based enhancement scheme with the NMF-based gain function. For a better performance in time-varying noise environments, both the noise and speech bases of NMF are adapted simultaneously with the help of estimated speech presence probability. Performed results showed that proposed model outperformed not only the statistical model-based and NMF approaches, but also their combination in various noise environments.

Table 2: Survey on Speech Signal Enhancement Algorithms

Title	Year	Algorithm	Advantage	Improvement
Improving Speech Recognition in Reverberation using a Room-Aware Deep Neural Network And Multi-Task Learning [08]	2015	Deep Neural Network (DNN)	It enhances the speech signal in reverberant environments.	DNN Training with multitask learning significantly improves speech recognition performance over conventional DNN.
Speech Enhancement with EMD and Hurst-Based Mode Selection [09]	2014	Empirical Mode Decomposition Hurst (EMDH) approach	Adaption of the hurst exponent in the selection of intrinsic mode functions (IMF) to reconstruct speech.	It improves the main objectives that are highly correlated with speech quality and intelligibility.
Generalized maximum a posteriori spectral amplitude estimation for speech enhancement [10]	2015	Generalized Maximum a Posteriori Spectral Amplitude (GMPSA) algorithm.	It gives more reliable gain function by incorporating a prior density.	It can provide higher speech quality for listening under various noisy conditions.
Modulation-Domain Multichannel Kalman Filtering for Speech Enhancement [11]	2018	Multichannel Kalman Filter (MKF)	It derives optimal MKF gain in short time Fourier transform domain.	It estimates clean signal in noisy and reverberant environments.
NMF-Based Speech Enhancement Using Bases Update [12]	2015	Non-Negative Matrix factorization (NMF)	Performs better in time varying noisy environments.	It enables efficient suppression of non-stationary noises even with mismatched data.

4. 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. We have studied different types of noise and its removal techniques. Also we have seen speech enhancement methods like single channel and multi-channel enhancing techniques and their sub types. Also in time domain method we have seen weiner filtering, kalman filtering, and linear predictive coding. And in transform domain method DFT Based (STSA Methods), signal subspace method. Also in this paper we have presented a review of various speech enhancement techniques.

References

[1] Jassim W. A, Paramesran R and Zilan, M.S, "Enhancing Noisy Speech Signals using Orthogonal Moments", IET Signal Processing, Vol. 8, No. 8, pp. 891-905, 2014.

[2] Khaldi K, Boudraa A.O and Turki M, "Voiced/unvoiced Speech Classification-Based Adaptive Filtering of Decomposed Empirical Modes for Speech Enhancement", IET Signal Processing, Vol. 10, No. 1, pp. 69-80, 2016.

[3] El-Fattah M.A.A, Dessouky M.I, Abbas A.M, Diab S.M, El-Rabaie E.S.M, Al-Nuaimy, W, Alshebeili S.A and El-Samie F.E.A, "Speech Enhancement with an Adaptive Wiener Filter", Springer, Vol. 17, No. 1, pp. 53-64, 2014.

[4] Wang J.C, Lee Y.S, Lin C.H, Wang S.F, Shih C.H. and Wu C.H, "Compressive Sensing-based Speech Enhancement", IEEE/ACM Transactions on Audio, Speech, and Language Processing, Vol. 24, No. 11, pp. 2122-2131, 2016.

[5] Wang S.S, Chern A, Tsao Y, Hung J.W, Lu X., Lai Y.H and Su B, "Wavelet Speech Enhancement based on Nonnegative Matrix Factorization", IEEE Signal Processing Letters, Vol. 23, No. 8, pp. 1101-1105, 2016.

[6] Tavares R and Coelho R, "Speech Enhancement with Non stationary Acoustic Noise Detection in Dime domain", IEEE Signal Processing Letters, Vol. 23, No. 1, pp. 6-10, 2016.

[7] Tavares R and Coelho R, "Robust Speech Processing using Local Adaptive Non-Linear Filter", IEEE Signal Processing Letters, 2013.

[8] Giri R, Seltzer M.L, Droppo J. and Yu D, "Improving Speech Recognition in Reverberation using a Room-Aware Deep Neural Network and Multi-Task Learning", IEEE, pp. 5014-5018, 2015.

[9] Zao L, Coelho R and Flandrin P, "Speech Enhancement with Emd and Hurst-based Mode Selection. IEEE, Vol. 22, No. 5, pp. 899-911, 2014.

[10] Tsao Y, and Lai Y.H, "Generalized Maximum a Posteriori Spectral Amplitude Estimation for Speech Enhancemen", Springer, Vol. 76, pp. 112-126, 2016.

[11] Xue W, Moore A.H, Brookes M and Naylor P.A, "Modulation-Domain Multichannel Kalman Filtering for Speech Enhancement", IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2016.

[12] Kwon K, Shin J.W and Kim N.S, "NMF-based Speech Enhancement using bases Update", IEEE Signal Processing Letters, Vol. 22, No. 4, pp. 450-454, 2015.