Survey on Wavelet Based Signal Processing Techniques for Speech Enhancement

Adappa Angadi[1], K. Shridhar[2]

[1]Department of Electronics and Communication Engg, Tontadarya College of Engineering, Gadag, Karnataka, India

[2]Department of Electronics and Communication Engg, Basaveshwar Engineering College, Bagalkot, Karnataka, India

***Abstract*— Information exchange among people occurs with a natural and efficient medium called speech. In speech communication, this speech signal is corrupted by various noises. Hence enhancing the speech is very much essential so that the enhanced speech is made better for human comfortable listening. The speech signal can be enhanced using wavelet based techniques that can reduce the noise efficiently. Enhancing speech involves improving speech quality and speech intelligibility. In our paper we review different wavelet based speech enhancement techniques.**

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

I. INTRODUCTION

This paper deals with the literature survey of wavelet based efficient signal processing techniques for speech enhancement. Most of the interaction among the humans is through speech. Speech processing is essential in many applications. To know the processes and dynamics of speech production and perception, research is going on for many years in this field. Speech processing field is basically a signal processing technique applied to acoustic signals. Noise gets added to the speech signal at any level in speech communication. Speech signal gets distorted due to the noise from various sources like the production organ, various sensors and their location, background noise, channel and reverberation or may be due to the disorder in human ear. From previous research, various techniques are described to enhance the speech particularly from background noise. Improving the quality and the intelligibility is the reason behind enhancing the speech. Various techniques are available for enhancing the speech quality and speech intelligibility. These front end compensation techniques are classified into two categories, single channel techniques and multichannel techniques. For single input channel, user uses single channel techniques like spectral subtractions and statistical model related techniques which are the most popular ones. STFT is used to process the speech signal in the above techniques and it performs well for low additive noise but for high additive noise it does not perform well. Considerable research in recent is focused on wavelet transform instead of STFT. This survey deals with the different wavelet based speech enhancement techniques that are described below.

II. LITERATURE SURVEY

This section describes the literature survey on speech enhancement using wavelets on research carried out by many researches. Quality and intelligibility of noisy speech signal can be improved considerably using wavelet packet decomposition methods.

Wavelet based Speech Signal Enhancement Techniques:

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 of techniques available for better signal enhancement. Survey of the various techniques associated speech signal analysis is explained.

Badiezadegan S and Rose R.C have discussed a new approach, which takes the information about the thresholds from spectrographic masks for removing wavelet domain coefficient noise and making DWT noise removal possible for non stationary noise conditions as well. This Proposed Discrete Wavelet Transform based method reduces the word error rate compared to MMSE method, and reduction of wer compared to cepstral mean-variance normalization method. Selection of threshold values at lower SNR values is challenging and is also a limitation.

Bhowmick A and Chandra M proposed scheme is efficient in noise estimation. Utilizing this advantage, A new approach to enhance the speech using Wavelet Decomposition based on Voiced Speech Probability is presented. Based on the probability calculation by likelihood ratio of two GMMs, Voiced/Unvoiced frames are separated by based improved Voice Activity Detector which is based on VSP . Mean Square Error is reduced by incorporating gain estimator with WD stage. The method which is proposed here results in the significant improvement of speech quality at all SNR values. Speech quality in high noise is shown to have a good improvement and in medium noise level the improvement is about 7.09%. But the drawback of the proposed method is the computational time of multilevel WD which is high.

Ghribi K and et. al. Proposed a technique which uses wavelet decomposition in forward blind source separation (FBSS) structure using a two-channel WFSAD algorithm. The algorithm proposed here gives better trade-off between time and frequency resolution and increase the noise reduction process robustness when set side by side with the two channel FSAD algorithm. The subjects have done there work on the speech signal which is obtained from the input SNR -3, 0 and 5 dB at two noisy sources. For the noise types like white, USASI and street noise sources the performance of the proposed algorithms is good. The efficiency of this method can accepted by averaged values of MOS across subjects. When the DWT scale is chosen high, then a little degradation of CD and segSNR values which is the disadvantage for this proposed WFSAD algorithm.

Hsieh C.T et al. have focused color noise filtering and Gaussian noise. Sparse algorithm is more popular in recent days for enhancing the speech signal, This method has a dictionary training part and signal reconstruction part. The sparse coefficients of the clean speech dictionary are made better with orthogonal matching pursuit algorithm and K-SVD(K-means clustering singular value decomposition) trains clean speech algorithm. Then matrices D' and X' are multiplied to construct back the clean speech signal. The results are shown at SNR levels -10,-5,0,5,10 dB by comparing with other traditional methods which shows proposed method is superior. The main drawback of the proposed scheme is its computational time.

Jayakumar E.P et. al presented a system which enhances speech by minimizing the noise from different environments. Spectral subtraction is one of the popular method for enhancing speech, and this method experiences perceptually annoying musical noise. The use of low variance spectrum estimators minimizes musical noise and the speech is enhanced by wavelet thresholding. The fundamental feature of the system proposed here is the mechanism of switching to an optimal wavelet filterbank to the human ear critical bands based on the noise type in the input speech . The system is estimated by noisy speech with various SNR values 0,5,10,15 dB. Results shows improved mean opinion score from proposed method when compared to other traditional method.

Mavaddaty S et.al proposed supervised scenario, takes domain adaptation technique advantage for transformation of a learned noise dictionary to a dictionary which is adjusted according to noise conditions which are reproduced from test environment circumstances Based on the present noisy space , observation data is sparsely coded with minimum sparse approximation error. From the above technique , one can obtain good enhancement results for non stationary noise aswell. In case of semi supervised technique, adaptation of threshold values of wavelet co-efficients is done by taking the approximated noise in every frame of various sub-bands. Experimental results obtained from PESQ , SegSNR and statistical tests for different noise types is less than the supervised enhancement algorithm.

Mourad T et. al proposed a new technique of enhancing the speech which incorporates wavelet transform named as stationary bionic wavelet transform and the maximum a posterior estimator of magnitude-squared spectrum (MSS-MAP). Speech estimation in SBWT domain done using MSS-MAP estimation . Experiments were carried on different types of noise and various speech signals and the results were compared with different well known methods like MSS-MAP estimation in frequency domain and Wiener filtering. Four quality measurement tests [SNR, segSNR, Itakura–Saito distance and perceptual evaluation of speech quality] were done to examine the performance of the proposed technique. The results are obtained for noises such as car, white, pink etc for different SNR levels. But the proposed approach will not perform well for lower SNR values.

Nabi W et. al proposed a dual microphone technique. There are many dual channel speech enhancement techniques for reducing noise but the dual microphone technique specified here for reducing noise has better performance compared to conventional ones. This technique is on the basis of coherence function and bionic wavelet transform which uses kalman filter. When BWT is applied to the input signal , Bionic wavelet coefficients are found and the kalman filter shall apply to these coefficients. The results show that, it can handle noisy speech signals better than traditional methods without distortion of speech at lower SNR. The proposed method achieves 0.9 PESQ score as an average improvement. However the proposed model is time consuming.

Islam M.T et. al have worked on Rayleigh pdf based threshold calculation and Custom thresholding. The Teager energy (TE) is not employed directly instead it is applied to the Perceptual Wavelet Packet coefficients of the noise corrupted speech , and Rayleigh modeling of this is done. The combination of μ-law and semisoft thresholding functions is used to determine a custom thresholding function. Thus we obtain a enhanced speech by applying the threshold derived from custom thresholding function onto PWP coefficients of noise corrupted speech. This technique is used to evaluate car and babble noise corrupted speech signals. Experiments were carried using the NOIZEUS database. The experimental results show that for speech corrupted babble noise, it shows better PESQseg and SNR improvement at higher SNR and lower SNR ranges. However, proposed method works well only for car and babble noise.

Table 2: Survey on Speech Signal Enhancement Algorithms

Title	Year	Algorithm	Advantage	Drawback
A Wavelet-based Thresholding Approach to Reconstructing Unreliable Spectrogram Components [01]	2015	Discrete Wavelet Transform (DWT)	DWT based denoising more suitable for non-stationary noise conditions.	Selection of threshold values at lower SNR values is challenging.
Speech Enhancement using Voiced Speech Probability based Wavelet Decomposition [02]	2017	Wavelet Decomposition (WD)	Easily tracks the quick variations in speech signal while estimating threshold value.	computational time of multilevel WD is high
A Wavelet-based Forward BSS Algorithm for Acoustic Noise Reduction and Speech Enhancement [03]	2015	Forward Wavelet Symmetric Adaptive Decorrelating (WFSAD)	It provides better compromise between noise and frequency resolution.	For high value of DWT scale , there is degradation of the CD and the Seg SNR values
Speech Enhancement is on the basis of Sparse Representation under Color Noisy Environment [04]	2015	Orthogonal Matching Persuit (OMP) and K-SVD	Efficient reconstruction of clean speech signal.	High computational time
Speech Enhancement based on Noise Type and Wavelet Thresholding the Multitaper Spectrum [05]	2015	Wavelet Thresholding	Performs well for different noise conditions.	Proposed method has been compared with only one algorithm.
Speech Enhancement using Sparse Dictionary Learning in Wavelet Packet Transform Domain [06]	2017	Sparse Dictionary based Learning using WPT	Better speech enhancement results under different noise conditions.	Experimental results obtained from PESQ , SegSNR and statistical tests for different noise types is less than that of supervised enhancement algorithm.
Speech Enhancement based on Stationary Bionic Wavelet Transform and Maximum A Posterior Estimator of Magnitude-Squared Spectrum [07]	2016	Stationary Bionic Wavelet Transform (SBWT) and Maximum a Posterior Estimator of Magnitude Squared Spectrum (MSS-MAP)	Sufficient noise reduction without causing signal distortion and musical background noise.	will not perform well when the SNR is low
A Dual-Channel Noise Reduction Algorithm based on the Coherence Function and the Bionic Wavelet [08]	2018	Bionic Wavelet Kalman Filter.	It can treat noisy signal well and provides enhanced signal without distortion.	Time Consuming.
Rayleigh Modeling of Teager Energy Operated Perceptual Wavelet Packet Coefficients for Enhancing Noisy Speech [09]	2016	Wavelet Packet Transform (WPT)	Better results in car noise and multi-talker bubble noise.	Proposed method works well only for car and bubble noise.

III. CONCLUSION

Speech enhancement is used to improvise the quality of the speech with the help of different algorithms. It improves the signal for human comfortable listening and further processing prior to listening. Intelligibility, degree of listener fatigue is also improved. We have studied different types of noise and their removal techniques. Also single channel and multi-channel enhancing techniques are also discussed. Also we have seen speech enhancement carried in different domains. In time domain method we have seen weiner filtering, kalman filtering, and linear predictive coding, In frequency domain method we have seen DCT, MMSE and Spectral Subtraction, And in time-frequency domain we have seen the methods involving family of wavelets. Also in this paper we have presented a review of various speech enhancement techniques.

REFERENCES

[1]. Badiezadegan S and Rose R.C, “A Wavelet-Based Thresholding Approach to Reconstructing Unreliable Spectrogram Components”, Elsevier, Vol. 67, pp. 129-142, 2015.
[2]. Bhowmick A and Chandra M, “Speech Enhancement using Voiced Speech Probability Based Wavelet Decomposition”, Elsevier, Vol. 62, pp. 706-718, 2017.
[3]. Ghribi K, Djendi M and Berkani D, “A Wavelet-Based Forward BSS Algorithm for Acoustic Noise Reduction and Speech Enhancement”, Elsevier, Vol. 105, pp. 55-66, 2016.
[4]. Hsieh C.T, Huang P.Y, Chen T.W and Chen Y.H, “Speech Enhancement based on Sparse Representation under Color Noisy Environment”, IEEE, pp. 134-138, 2015.
[5]. Jayakumar E.P and Sathidevi P.S, “Speech Enhancement based on Noise Type and Wavelet Thresholding the Multitaper Spectrum”, Springer, pp. 187-200, 2016.
[6]. Mavaddaty S, Ahad, S.M and Seyedin S, “Speech Enhancement using Sparse Dictionary Learning in Wavelet Packet Transform Domain”, Elsevier, Vol. 44, pp. 22-47, 2017.
[7]. Mourad T, “Speech Enhancement based on Stationary Bionic Wavelet Transform and Maximum A Posterior Estimator of

Magnitude-Squared Spectrum", Springer, Vol. 20, No. 1, pp. 75-88, 2017.

[8]. Nabi W, et.al. "A Dual-Channel Noise Reduction Algorithm based on the Coherence Function and the Bionic Wavelet", Elsevier, Vol. 131, pp. 186-191, 2018.

[9]. Islam, Shahnaz C, Zhu and Ahmad M, "Rayleigh Modeling of Teager Energy Operated Perceptual Wavelet Packet Coefficients for Enhancing Noisy Speech", Elsevier, Vol. 86, pp. 64-74, 2017.

[10]. Singh S, Tripathy M and Anand R.S, "A Wavelet based Method for Removal of Highly Non-Stationary Noises from Single-Channel Hindi Speech Patterns of Low Input SNR", Springer, Vol. 18, No. 2, pp. 157-166, 2015.

[11]. Singh S and Mutawa A.M, "A Wavelet-Based Transform Method for Quality Improvement in Noisy Speech Patterns of Arabic Language", Springer, Vol. 19, No. 4, pp. 677-685, 2016.

[12]. Singh S, Tripathy M and Anand R.S, "A Wavelet Packet Based Approach for Speech Enhancement Using Modulation Channel Selection", Springer, Vol. 95, No. 4, pp. 4441-4456, 2017.

[13]. Soleymani R, Selesnick I.W and Landsberger D.M, "SEDA: A Tunable Q-Factor Wavelet-based Noise Reduction Algorithm for Multi-Talker Babble", Springer, Vol. 96, pp. 102-115, 2018.

[14]. Swami P.D, Sharma R, Jain A and Swami D.K, "Speech Enhancement by Noise Driven Adaptation of Perceptual Scales and Thresholds of Continuous Wavelet Transform Coefficients", Elsevier, Vol. 70, pp. 1-12, 2015.

[15]. Bouzid A and Ellouze N, "Speech Enhancement based on Wavelet Packet of an Improved Principal Component Analysis. Elsevier, Vol. 35, pp.58-72, 2016.

[16]. Oktar M.A, Nibouche M. and Baltaci Y, "Speech Denoising using Discrete Wavelet Packet Decomposition Technique", IEEE, pp. 817-820, 2016.

[17]. Ming J, Crookes D, Ming J and Crookes D, "Speech Enhancement based on Full-Sentence Correlation and Clean Speech Recognition", IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP), Vol. 25, No. 3, pp. 531-543, 2017.

[18]. Cees H, Taal, Jesper Jensen and Arne Leijon. "On Optimal Linear Filtering of Speech for Near End Listening Enhancement", IEEE, Vol, 20, No. 3, 2013.

[19]. Joseph Szurler,, Alexander Bertrand, and Marc Moonen, "On the Time-Domain widely Linear Filtering for Binaural Speech Enhancement", IEEE, Vol. 20, No. 7, 2013.

[20]. Tabibian S, Akbari A and Nasersharif B, "Speech Enhancement using a Wavelet Thresholding Method based on Symmetric Kullback–Leibler Divergence", Elsevier, Vol. 106, pp. 184-197, 2015.

[21]. Robert M Nickel, Ramon Fernandez Astudillo, Dorothea Kolossa and Rainer Martin, "Corpus-based Speech Enhancement with Uncertainty Modelling and Cepstral Smoothing", IEEE, Vol. 21, No. 5, 2013.

[22]. Salehi H, and Parsa V, "Nonintrusive Speech Equality Estimation based on Perceptual Linear Prediction", IEEE, pp. 1-4, 2016.

[23]. Tang H.S, Yang C.Y, Liu C.W and Chien C.C, 2016, "Binaural-Cue-based Noise Reduction using Multirate Quasi-ANSI Filter Bank for Hearing aids", IEEE, pp. 21-24, 2016.

[24]. Necciari T, Holighaus N, Balazs P and Prusa Z., "A Perceptually Motivated Filter Bank with Perfect Reconstruction for Audio Signal Pprocessing", 2016.

[25]. Nikhil G.V, Keerthi A.M and Premananda B.S, "Impact of ERB and Bark Scales on Perceptual Distortion based Near-end Speech Enhancement", IEEE, pp. 364-370, 2017.

[26]. Taal C.H, Hendriks, R.C and Heusdens R, "Speech Energy Redistribution for Intelligibility Improvement in Noise based on a Perceptual Distortion Measure. Elsevier, Vol. 28, No. 4), pp. 858-872, 2014.

[27]. Zhang Y, Sun L, Wang P.P and Luo Y, "Robust Speech Recognition based on Speech Enhancement and Improved Perceptual Non-Uniform Spectral Compression", IEEE, pp. 3077-3082, 2016.