Result Analysis Of Different Wavelet Types Using Speech Enhancement Algorithms

Divya Gupta, Poonam Bansal, Kavita Choudhary

Abstract: The main aim of speech enhancement is to enhance the overall standard and intelligibility of the speech by decreasing the background noise level. This paper proposes a speech enhancement technique to enhance the speech signal in wavelet domain. The technique used is a combined approach of generalized spectral subtraction and MMSE log-STSA. The Experiment is conducted on male and female Database at different SNR levels.

Index Terms: Speech Enhancement, wavelet transform, noise reduction, spectral subtraction

1 Introduction

Speech is considered important mode of communication or interaction between humans or between human and machines. This speech is usually corrupted by background noise or any musical noise. So, the main goal is to reduce the noise. Speech Enhancement becomes useful in such cases as its main purpose is to increase overall quality and intelligibility of speech presented to the listener.

Speech Recognition Approaches

In the beginning of ASR technology dynamic programming techniques are employed to deal with pattern recognition problem.[9].As the technology progresses Artificial Neural Networks (ANN) is used. Currently, Stochastic modeling methodology has been adapted to construct the speech recognition systems. This modeling approach includes Hidden Markov Models. These modeling approaches are discussed below

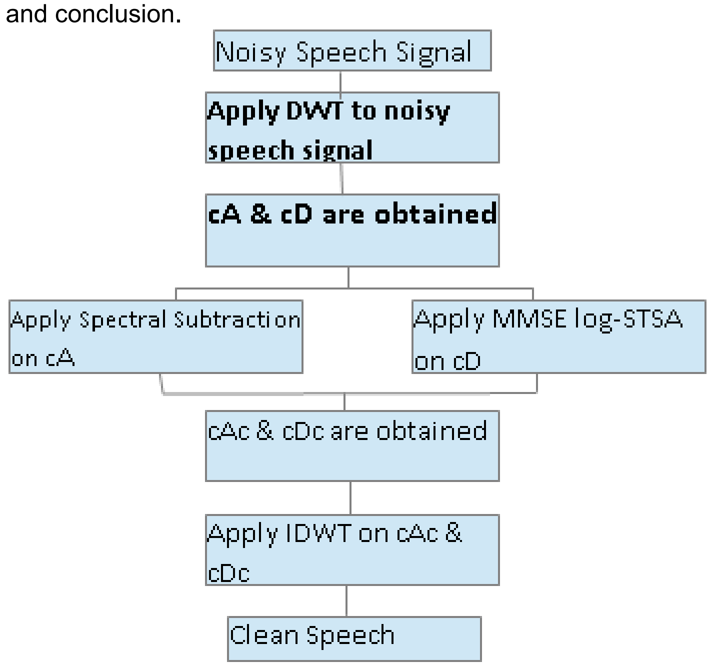
1. Template Based Approach – This approach has given a group of techniques to the field of speech recognition that has made a great progress in this field during the last some years. In this approach, group of acoustic patterns are made to store as reference patterns. Initially the system is trained using these recorded patterns and afterwards when an unknown acoustic signal is given as input to the system then the given utterance is recognized by matching it with the templates stored during the training period of the system. The main idea behind this approach is to derive series of speech frame for each word and then to depend upon their spectral distance measure to compare the utterances.[10]. This approach has the advantage of developing accurate models for given utterances; however the it also suffers from certain demerits as it works with initially stored static patterns, therefore in order to deal with speech variability ,large number of templates are need to be developed per word ,which is practically impossible and leads to high computational cost. Hence, this method was inappropriate both in terms of processing power needed to perform the matching and also it was completely speaker dependent[4].
2. Neural Network Based Approach – The other methodology that can be adopted for classification is the implication of neural networks. They have the ability to deal with complex recognition tasks but does not have the excellence to deal with large size vocabularies like HMM. However, they have the ability to deal with low quality signals and this approach also allows speaker independence [4].These approach allows better performance and recognition rate then other approaches until the training data is less and vocabulary size is small. This methodology is generally adapted for phoneme recognition.
3. Knowledge Based Approach - Deployment of knowledge based processes for developing automatic speech recognition systems was presented by various researchers, speech understanding systems. The “expert” knowledge concern with speech variability has to be encoded within the system. It works with the extracted feature vectors from the speech and then train the system using these vectors that result into the generation of production rules from these vectors. The given utterance is recognized using an inference engine at the frame level. The inference engine is used for executing the decision tree and thus for classifying the rules fired. This modeling approach has the advantage of allowing speech variability. However, expert knowledge is tough for exploitation and hence this methodology was proved to be impractical for implementation.[10]
4. Dynamic Time Warping – This algorithm is adapted to assess the similarities among different sequences which can vary either in terms of time or in terms of speed[4]. Hence, this methodology can be used to develop an ASR that can manage different speaking speeds among various speakers. In other terms, this method allows a machine to search for an optimal match between two considered speech sequences with specific constraints , for instance the series has been "warped" on a non-linear scale thus to allow matching between them. This technique can be easily employed to recognize isolated words and need to be changed to identify the connected word accordingly [4].
5. Vector Quantization – This methodology is adopted for developing speech recognition as it allows good reduction in data. The utility of this approach exist in developing codebooks for acoustic signal modeling. The given test utterance is examined by all codebooks and system opted the word corresponding to which codebook gives the minimum distance measure. In general codebooks generated as a process of VQ does not contain explicit

- *Divya Gupta is currently doing her PhD at Jaganath University, Jaipur in Computer Science department, E-mail: fromdivya81@gmail.com*
- *Dr Poonam Bansal is a professor in Computer science and engineering at Maharaja Surajmal institute of Technology Janakpuri, Delhi, India, E-mail: Pbansal89@yahoo.co.in*

time information, as codebook entries are not in sequence and can belong to any region of training utterances. Code Words are never generated for stationary frames as these frames would results into large frame distances to the codebook. VQ approach is used for other constructing codebook and are hybrid with other approaches like HMM to allow easy training of data. VQ gives relative emphasis on speech transients that gives an extra advantage over other speech recognition systems for the vocabularies containing similar words[14].The main merit of VQ is that it allows less memory storage requirement for spectral analysis of data and also it provides the discrete representation for input speech utterance. However, VQ has its own demerits as it ignore temporal information and its training size increases linearly with increase in size of vocabulary[11].

6. Support Vector Machines – The other important approach for pattern modeling is support vector machines. This methodology work using non linear as well as linear separating hyper-planes for classifying data. The main issue with SVM is that it cannot be employed for work that involves non-fixed length data classification as this approach can classify non – variable, fixed length data only. Therefore, prior applying SVM, data having non-fixed length need to be converted into a fixed length data. SVM uses maximum- margin fitting function that allows the classifier to generalize better. This approach is not dependent on dimensionality and thus allows large dimension spaces which further allow the development of very large number of non-linear features and then it performs adaptive feature selection for training process.. The major issue with SVM is that its training algorithm are unable to manage large size databases[11]
7. Stochastic Approach - In this approach, automatic learning procedures are used to model speech variability statistically, for instance by using Hidden Markov Model. This methodology shows the current state of art. Latest speech recognition systems are developed using statistical acoustic modeling approach such as HMM. The main demerit of this approach is that they work my making a priori modeling assumptions that could result into inaccurate and false system performance[4].This approach includes the use of probabilistic models to allow the system to deal with uncertain data. Such uncertain information in speech recognition arises due to different reasons including improper sounds, contextual effects or variations among different speakers. The most commonly Stochastic modeling approach used for developing current ASR systems is Hidden Markov Model. HMMs are employed for acoustic signal modeling because they are easy to trained and incur low computational cost. HMM can be divided into different categories including DHMM, CHMM, etc. The attributes that are used to define the model include probabilities for individual state transitions , the means, variances and mixture weights that define the state output distributions. Each word, or each phoneme, will have a different output distribution; a HMM for a sequence of words or phonemes is made by concatenating the individual trained HMM [4] for the separate words and phonemes. Latest HMM-based large vocabulary speech recognition systems has been trained for hundreds of hours with acoustic utterances. It is the main merit of this model that results into reduction of time and complexity of recognition process for training large size vocabulary. Because of the above mentioned advantages; I had use DHMM with VQ for acoustic feature modeling in my project so as to achieve high recognition rate.

Different Speech Enhancement algorithms have been proposed since last decades. The one of the problem is that pitch or spectrum of male voice is different from female voice. Spectral Subtraction Algorithm is the most fundamentally known algorithm. It reduces the noise by subtracting estimated noise spectrum from the noisy signal to generate the clean speech. But this doesn't work for non-stationary noise and musical noise. So to remove such problem parameters are changed and other generalised spectral subtraction algorithms are generated to work in wavelet domain. In this paper the hybrid approach is used for noise reduction. The spectral subtraction is cascaded with MMSE log-STSA algorithm in frequency domain. The time domain is converted into frequency domain used wavelet transform because to implement spectral subtraction algorithm this transformation from time domain to frequency domain is required . Section II gives brief view of proposed system. Section III consist of explanation of Spectral Subtraction followed by section IV explaining System Overview. Lastly Section V shows result and conclusion.



SPECTRAL SUBTRACTION ALGORITHM

Spectral subtraction is one of the oldest and fundamental algorithm known for speech enhancement. It is very effective for stationary noise. This algorithm is well known due to its easy implementation and simplicity this algorithm estimated noise spectrum is subtracted from noisy speech signal to obtain clean speech signal.

$$x(n)=y(n)-z(n)...........[8]$$

x(n) represent clean speech signal. y(n) represent noisy speech signal and z(n) represent additive noise[8]. It uses magnitude spectral subtraction. One major draw back of subtraction of additive noise spectrum is that if too little is subtracted then there are chances of of remains of noise that can cause interference and on other hand if large magnitude is subtracted then speech information might be lost. So in the proposed work spectral subtraction is applied on approximation coefficient cA that is low pass spectrum. On applying we get cAc which is cleaned approximation coefficient.

MMSE LOG-STSA

MMSE log-STSA means Minimum Mean Squared Error. This algorithm is used in amplitude estimator in log spectral domain. This algorithm is used for spectrum which is above average noise level. It decreases this level. Basically it is used for musical noise. It is used for high pass filters. It is applied on detailed coefficient cD to get clean detailed coefficient cDc.

COMBINED APPROACH USED IN THE SYSTEM

A combined approach is used in wavelet de-noising system. Spectral Subtraction and MMSE log-STSA techniques are combined in wavelet domain. Wavelet transform is a new and good set of tools and technique in speech enhancement area. DWT is simply a filter being used or it can be said as decomposition of signals. It divides the the signal in two different bands:- high filter band and low filter band. The low filter band coefficient is denoted by cA which is known as approximation coefficient. On the other hand high pass filter band coefficient is denoted by cD which is commonly known as detailed coefficient. The important part of algorithm is reconstruction of signals without the loss of signal information or speech. So first cA is changed into clean cA denoted by cAc by applying spectral subtraction on this low band coefficient. And cD is changed into clean cDc by applying MMSE log-STSA technique on it. Once this part is done, reconstruction of signal is done in order to protect the loss of important information. This reconstruction is done by IDWT(Inverse Discrete Wavelet Transform).More over there are many types of wavelets such as Haar, Daubechies, Coiflets, Symlet and etc. Here we are using all possible types of wavelets to analyse the result and find the best wavelet for removing the noise at different levels.

Results and Conclusion

Experimental results are analysed for all different types of wavelets. Noise level varies from minimum where no noise is added to maximum where noise level is too high. The best results were seen using coif5 wavelet. The .wav file was completely denoised. Whereas while using Haar wavelet some noise is left and speech not clean. Other wavelets also didnt produced good cleaned speech. Some amount of noise was still left after denoising. The results were also analysed for hard and soft thresholding. Soft thresholding came out to yield better result and better SNR ratio. Our experimental results show that our proposed system is able to remove noise and improving the speech using coif5 wavelet.

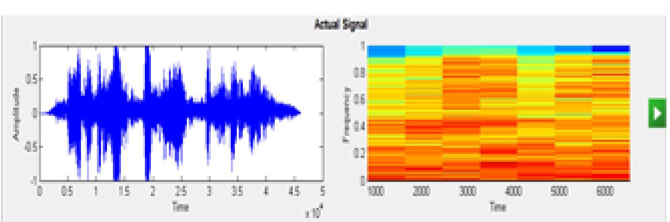


Fig:-*Actual Signal*

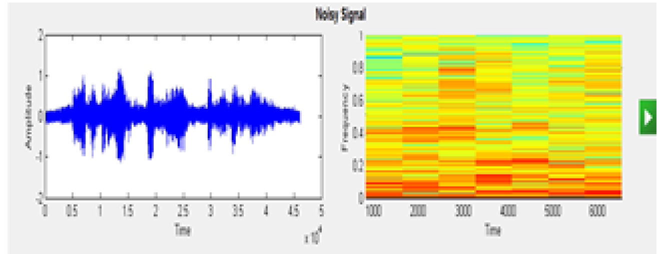


Fig:-*Noisy Signal*

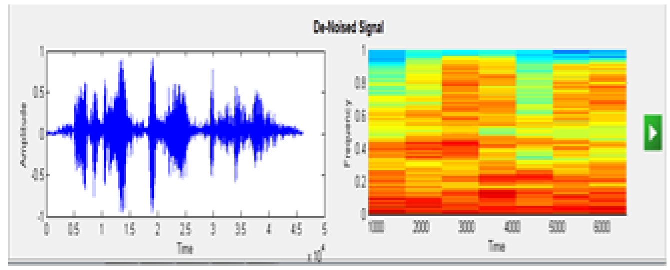


Fig:-*Result after denoising using Coif5 Wavelet*

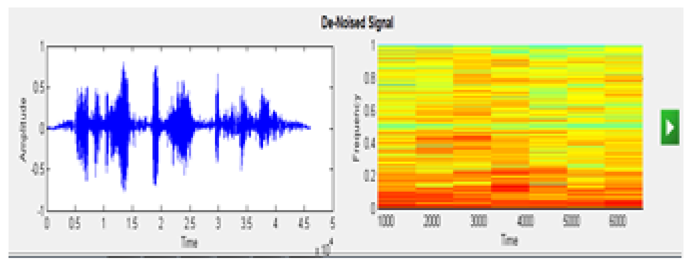


Fig:-*Result after denoising using Haar Wavelet*

REFERENCES

[1] Nikesh Bajaj and Rahul Kashyap “Extension of Wavelet Family in Fractional Fourier Domain” December 2012

[2] Hai-rong Jia, Xue-ying Zhan, Jing Bai, “A continuous differentiable wavelet threshold function for speech enhancement”,2013

[3] Paruav Goel and Anil Garg “Review of spectral subtraction Technique for speech enhancement” IJECT Vol 2 December 2011.

[4] Liuyang Gao and Yunfei “Speech enhancement algorithm based on improved spectral suntraction “ IEEE 2009.

[5] Anuradha R. Fukane, Shashikant L. Sahare, "Enhancement of Noisy Speech Signals for Hearing Aids", 2011

[6] Anuprita P. Pawar , Kirtimalini B. Choudhari and Madhuri A Joshi , “Review of Single Channel Speech Enhancement Methods in Spectral Domain” International Journal of Applied Engineering Research, ISSN 0973-4562 Vol. 7 No.11 ,2012.

[7] Ganga Prasad Surender, “A Review of Different Approaches of Spectral Subtraction Algorithms for Speech Enhancement”, Current Research in Engineering, Science and Technology (CREST) Journals,

[8] Paruav Goel and Anil Garg ”Develpoments in Spectral Subtraction for Speech Enhancement”IJERA Vol 2 February 2012.

[9] Tahsina Farah Sanam, Celia Shahnaz “An Approach for Noisy Speech Enhancement through soft thresholding function by Employing the Teager Energy Operation on Wavelet Packet” March 2012

[10] Yao Ren, Michael T. Johnson, and Jidong Tao "Perceptually motivated wavelet packet transform for bioacoustic signal enhancement"April 2008.
[11] Ruchismita Tripathy, Hrudaya Kumar Tripathy, "Unalike Methodologies of Feature Extraction & Feature Matching in Speech Recognition", published in,2014.
[12] M.A.Anusuya, S.K.Katti, "Speech Recognition by Machine: A Review,"(IJCSIS) International Journal of Computer Science and Information Security, Vol. 6, No. 3, 2009, pp.181-205.
[13] Paul A.K., Das D., Kamal M.M.,2009."Bangla Speech Recognition System Using LPC and ANN", Seventh International Conference on Advances in Pattern Recognition, IEEE Xplore, (Kolkata, Feb. 4-6 2009), 171 – 174.
[14] Sonia Sunny, David Peter S and K Poulose Jacob, "A Comparative Study of Parametric Coding and Wavelet Coding Based Feature Extraction Techniques in Recognizing Spoken Words", CUBE 2012,Published in September 3–5, 2012, Pune, Maharashtra, India.
[15] Thiang, Suryo Wijoyo,,. "Speech Recognition Using Linear Predictive Coding and Artificial Neural Network for Controlling Movement of Mobile Robot", Proc. of. Int. Conf. on Information and Electronics Engineering, IPCSIT vol.6, (IACSIT Press, Singapore), 2011.
[16] C. Demiroglu and T. Barnwell, "A Missing-Data Approach to Noise-Robust LPC Extraction for Voiced Speech Using Auxiliary Sensors", Published in ICASSP 2005.
[17] L. S. Chee, Ooi Chia Ai, and S. Yaacob, "Overview of Automatic Stuttering Recognition System," in International Conference on Man- Machine Systems (ICoMMS 2009) Penang, Malaysia, 2009.
[18] Preeti Saini , Parneet Kaur, "Automatic Speech Recognition: A Review", Published in International Journal of Engineering Trends and Technology-Volume4Issue2- 2013.