An Approach for Noise-Speech Discrimination Using Wavelet Domain

Tayar Myo Tun
University of Computer Studied, Mandalay
taryarmyotun@gmail.com

Abstract

There are many sudden and short period noises in natural envrioments. In this paper, noise reduction is efficiently performed for additive white noise and an automatic thresholding method for discriminating of noise / speech. This modified version of the thresholding method updates the threshold in each frame. In this proposed method, the selection of the threshold value depends on the estimates of the standard deviation and gives it as the input to the super-soft thresholding algorithm. Voice activity detection methods usually work in time or frequency domains. We propose super soft thresholding algorithm based on subband voice activity detection. If clean speech data can be input, it will help prevent system operations errors. These proposed methods are applied in a real-time noise reduction.

Keyword: Noise, Wavelet Transform, Continuous Wavelet Transform, Thresholding algorithm, Threshold value, Power estimator, Voice Activity Detection, Subband

1. Introduction

Speech is considered one of the most natural forms of communications between people. In speech communication, the speech signal is always accompanied by some noise. The background noise of the environment where the source of speech lies is the main component of noise that adds to the speech signal. Sound that is unwanted or disrupts one's quality of life is called as noise. Noise reduction is useful in many applications such as voice communication and automatic speech recognition where efficient noise reduction techniques are required. In the literature, this process is usually known as voice activity detection (VAD) and it becomes an important problem in many areas of speech processing such as real-time noise reduction for speech enhancement, speech recognition, digital hearing aids, and modern telecommunication systems. Noise is a well-known factor which degrades the quality and intelligibility of speech in many applications' areas. To reduce the noise level without affecting the quality of speech signals, a noise reduction algorithm is usually employed. Spectral subtraction is a widely used approach in practical noise suppression schemes.

This scheme usually estimates the noise characteristics from the nonspeech intervals of the signal. In this context, accuracy and reliability of a VAD becomes critical in determining the performance of noise reduction algorithm. Most papers reporting on noise reduction refer to speech pause detection when dealing with the problem of noise estimation. Speech pause detectors are very sensitive and often limiting part of the systems for the reduction of noise in speech. [5]

The rest of the paper is organized as follows: section II will provide the details about the theoretical background, section III will depict the proposed algorithm for noise-speech discrimination in real time, section IV will present conclusion and future work.

2. Related Work

Rajeev Aggarwal , Jai Karan Singh , Sanjay Rathore , Mukesh Tiwari , Dr. Anubhuti Khare and Vijay Kumar Gupta describe discrete-wavelet transform based algorithm are used for speech signal denoising. Here both hard and soft thresholding are used for denoising. Analysis is done on noisy speech signal corrupted by babble Soft thresholding method performs better than hard thresholding at all input SNR levels.

Dr.Moe Pwint and Farook Sattar presents a new method to detect speech/nonspeech components of a given noisy signal. Employing the combination of binary Walsh basis functions and an analysis-synthesis scheme, the original noisy speech signal is modified. From the modified signals, the speech components are distinguished from the nonspeech components by using a simple decision scheme. Minimal number of Walsh basis functions to be applied is determined using singular value decomposition (SVD).

Wahyu KuSuma and Prince Brave Guhyapativ foucus on nalysis of matching process to give a command for multipurpose machine such as a robot with Linear Predictive Coding and HMM. LPC is a method to signals by giving characteristics into LPC coefficients. In the other hand, HMM is a form of signal modeling where voice signals are recognized to find maximum probability and recognize words given by a new input.

3. Theoretical Background

The proposed denoising algorithm computes the discrete wavelet transform for noisy speech signal. The wavelet transform is good at energy compaction. An Initial threshold value is selected by using standard deviation and super soft thresholding algorithm. And then, estimated power is compared with threshold value in each subband. Finally, reconstruct to the inverse discrete wavelet transform.

3.1. Wavelet Transformation

The theory of the wavelet transform (WT) is based on signal processing and developed from the Fourier transform basis. The wavelet transform is expressed as a series of functions which are related with each other by translation and simple scaling. The original WT function is called mother wavelet and is employed for generating all basis functions. A set of functions is constructed by scaling and shifting the mother wavelet. [1]

3.1.1. Discrete Wavelet Transform

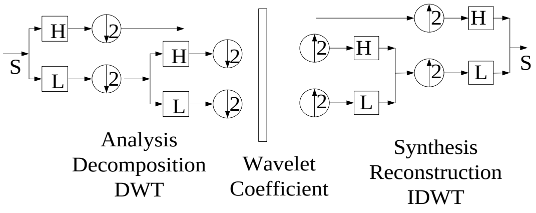

Figure 1. Wavelet decomposition and reconstruction

The Discrete Wavelet Transform (DWT) uses Multi resolution filter banks and special wavelet filters for the analysis and reconstruction of signals. DWT has two major phases, decomposition and reconstruction. In decomposition phase, input signal is decomposed into two sub-bands: low frequency and high frequency sub-bands. These subbands are achieved separately by applying low-pass and high-pass decomposition filters to the input signal. Low frequency and high frequency sub-bands are named as approximation and detail respectively. [2]

3.1.2. Continuous Noise Estimation in Wavelet

In wavelet filter-bank, a set of coefficients are applied to present each sub-band, while in frequency filter-bank, a coefficient is used for presenting each sub-band. In continuous noise estimation, we use a filter including two steps. In the first step, an adaptive filter is applied to noisy signal. After that, powers of input noise and clean signal are estimated in each subband. In the second step, the enhancement filter reduces the noise based on powers of estimated noise and estimated speech signal. The advantage of using two steps filtering is that estimation and reduction filters are separately calculated**.**

The power of signal y, in k-th frame, n-th subband and m is number of coefficients in this subband is calculated as:

$$P_Y(n,k) = \sum_{m=1}^{M} Y^2 \ (m,n,k) \qquad (1)$$

First, the adaptive filter, H(j,k) can be computed as in equation (2) using the estimated power of noise, P_N(j,k-1), and the estimated power of clean signal, P_S(j,k-1). These parameters are estimated in the previous frame.

$$\mathrm{H}(\mathrm{j},\mathrm{k}) = \frac{\mathrm{P_S}\,(\mathrm{j},\mathrm{k}-1)}{\mathrm{P_S}(\mathrm{j},\mathrm{k}-1)+\mathrm{P_N}(\mathrm{j},\mathrm{k}-1)} \qquad (2)$$

Estimation of noise power in current frame, P_N(j,k) is calculated based on power of noisy signal in current frame, X(j,k), and estimated of noise power in previous frame in each sub-band.

$$\mathrm{N}^2(\mathrm{j},\mathrm{n}) = \mathrm{X}^2(\mathrm{j},\mathrm{k}).(1-\mathrm{H}(\mathrm{j},\mathrm{k}))$$

$$\mathrm{P}_{\mathrm{N(j,k)}} = \lambda_\mathrm{N}.\mathrm{P_N}(\mathrm{j},\mathrm{k}-1) + (1-\lambda_\mathrm{N}).\mathrm{N}^2(\mathrm{j},\mathrm{n})(3)$$

where λ_N is the forgetting factor for noise estimation ($0 < \lambda_N \leq 1$). In Equation (4) estimated power of clean signal is approximated based on power of noisy signal in current frame, X(j,k), and power of estimated clean signal in previous frame ,P_S(f,k-1).

$$S_1^2(j,k) = H(j,k) \times X^2(\mathrm{j},\mathrm{k})$$

$$\mathrm{P}_{\mathrm{s(j,k)}} = \lambda_\mathrm{s}.\mathrm{P_s}(\mathrm{j},\mathrm{k}-1) + (1-\lambda_\mathrm{s}).S^2(\mathrm{j},\mathrm{k}) \qquad (4)$$

where λ_S is forgetting factor for speech estimation. In the second step, based on equation (5), main noise reduction filter in each sub-band, A(j,k), is calculated. In this equation, α (1-H) is used to control the amount of remined background noise and speech artefacts. [10]

$$\mathrm{A(j,k) = H(j,k) + \alpha\,(1\text{-}H\,(j,k))} \qquad (5)$$

3.2. Thresholding

3.2.1. Initial Value of Threshold Selection

This propose method involves the estimating the standard deviation of the noise, σ, for every subband and time frame. When σ is given, we will calculate the threshold, λ , again for each subband and time frame. We will start by segmenting each i-th subband of decomposed coefficients, $c^i_{m,}$, into frames of length $L^i{}_{frm}$. $\sigma^{i,p}$ is denoted as the corresponding estimated noise level of the p-th frame in the i-th subband. These are estimated using the segment of previous data { $c^{i,p}{}_{m,, m=0,\ldots,}$ $L^i{}_{seg-1}$ } , where $L^i{}_{seg-1} > L^i{}_{frm}$. The noise estimate is then given as

$$\sigma^{i,p} = \beta \;.\; \sum_{J=0}^{int(q.L^i_{seg})} c_j^{i,p} \,/L^i_{seg} \qquad (6)$$

where the constant β is an appropriate scale factor. Nominal values: q = 0.2, β = 0.38. The corresponding time lengths of L_{seg} and Lf_{rm} are 512ms and 64ms respectively, and the frame shift is 32ms. Finally, the threshold for each subband at the p-th frame,

$$\lambda^{i,p} = \sqrt{2\log(L^i_{seg}\;log_2\,L^i_{seg})}\;.\;\sigma^{i,p} \qquad (7)$$

3.2.2. Thresholding Algorithm

The soft and hard thresholding methods are used to estimate the wavelet coefficients in wavelet threshold denoising. Hard thresholding can be described as the usual process of setting to zero the elements whose absolute values are lower than the threshold. The hard threshold signal is x if x ≥ thr and is 0 if x < thr, where „thr is a threshold value. Soft thresholding is an extension of hard thresholding, first setting to zero the elements whose absolute values are lower than the threshold, and then shrinking the nonzero coefficients towards 0. If x ≥ thr, soft threshold signal is (sign (x). (x - thr)) and if x < thr , soft threshold signal is 0.Donoho and Johnstone derived a general optimal universal threshold for the Gaussian white noise.

Super Soft thresholding algorithm instead of setting some wavelet coefficients to zero, the algorithm attenuates the coefficients depending on their distance from the threshold. Super-Soft thresholding algorithm avoids forcing the wavelet coefficients smaller than the threshold to zero but instead replaces them by a fraction of their original values. The Super-Soft thresholding algorithm is defined as follow:

$$\delta_\lambda^S(x) = \begin{cases} sign(x)(a|x|) & |x| \leq \lambda \\ sign(x)(|x| - \lambda_1) & |x| \succ \lambda \end{cases} \qquad (8)$$

where a is the line slope for the values smaller than threshold, so it should be a small value and λ is the threshold value.

3.3. Voice Activity Detection (VAD)

The VAD detects the presence of speech in a noisy signal. So these methods are suitable in order to apply to real-time applications. In thresholding methods, some parameters such as zero-crossing rate and power are compared with threshold values. In this paper, we propose a subband VAD algorithm based on power of estimated speech comparing with estimation threshold. There are two types of VAD: Fullband VAD and Subband VAD.

3.3.1. Fullband VAD

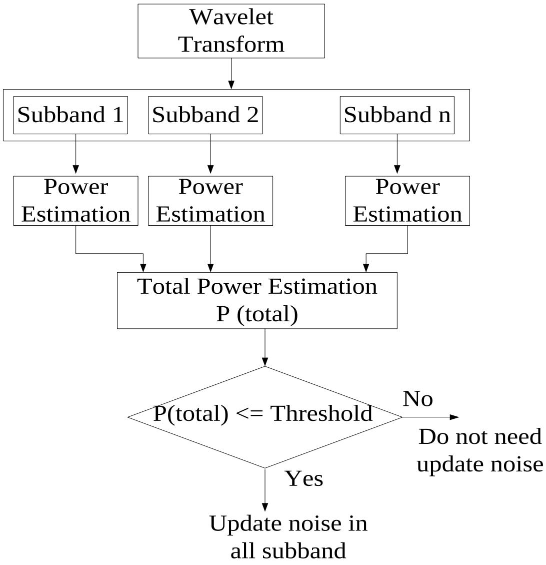

Figure 2. Fullband VAD diagram

In Fullband VAD, powers of speech for all sub-bands are accumulated together. This value is used as estimation of speech power in current frame. If this value is less than a determined threshold, current frame is labelled as noise frame; otherwise current frame is labelled as speech frame. In noise frames, power of estimated noise is updated. In speech frames, power of estimated noise is not changed.[2]

3.3.2. Subband VAD

In case of Subband VAD, different threshold values are computed for each sub-band. If estimation of speech power in each sub-band is less than these thresholds, then these subband is labelled as noise, otherwise it is labelled as signal. If a sub-band is considered as noise, the estimation of noise power is updated in the sub-band; otherwise the estimation of noise power is not changed in the sub-band. [2]

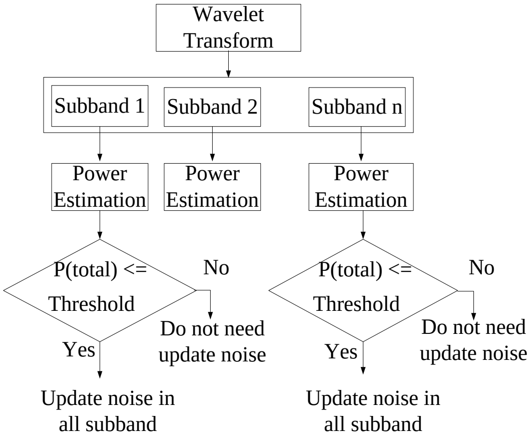

Figure 3. Subband VAD diagram

4. Proposed Algorithm for Noise-Speech Discrimination

In this proposed method, some reference signal for recording are created (clean signal). The reason for choosing white noise is that white noise is one of the most difficult types of noise to remove, because it does not have a localized structure either in the time domain or in the frequency domain. [3] White noise with different SNR values is added to reference signals to generate noisy signal. Speech segments are labelled by Hamming window.

Wavelet filter-bank will decompose reference signals at first. Number of decomposition levels will set and that gives the resolution in the approximation. Wavelet Transform is defined for sequences with length of some power of 2, and for other sizes we need to use signal extension to a size that is in the power of 2.After then creating reference signals in each sub-band, we will calculate Subband VAD classification rate in each sub-band.

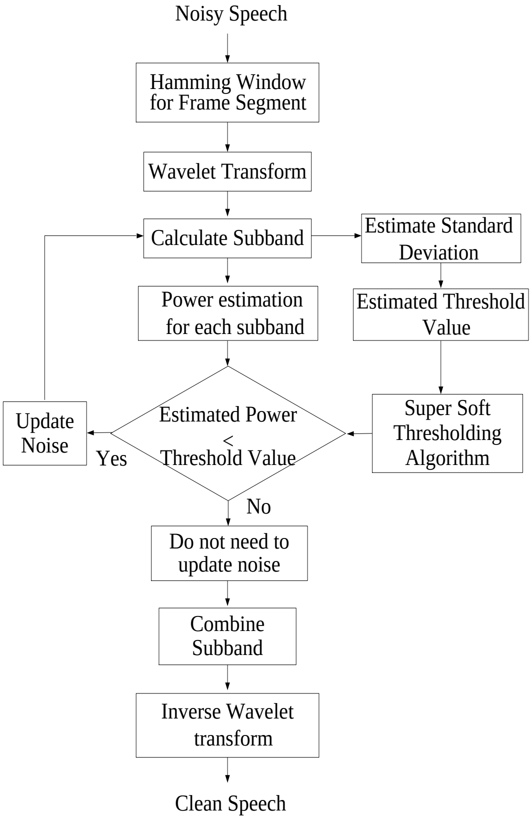
Noisy Speech
Hamming Window for Frame Segment
Wavelet Transform
Calculate Subband
Estimate Standard Deviation
Power estimation for each subband
Estimated Threshold Value
Update Noise
Estimated Power < Threshold Value
Super Soft Thresholding Algorithm
Yes
No
Do not need to update noise
Combine Subband
Inverse Wavelet transform
Clean Speech

Figure 4. Proposed System

The standard deviation of the noise σ is estimated. When σ is given, we will calculate the threshold, λ , again for each subband and time frame. We will start by segmenting each i-th subband of decomposed coefficients, $c^{i}_{m,}$, into frames of length $L^{i}{}_{frm}$. The proposed system uses an estimation of the standard deviation for the each frame to select the threshold of the current frame and gives it as the input to the super-soft thresholding algorithm. Different threshold values are computed for each sub-band.

To estimate the power of each subband, the adaptive filter calculates the power of noisy speech. In the second step, the enhancement filter reduces noise based on powers of estimated noise and estimated speech signal. If estimation of speech power in each sub-band is less than these estimate thresholds, then the subband is labelled as noise, otherwise it is labelled as speech. If a sub-band is considered as noise, the estimation of noise power is updated in the sub-band; otherwise the estimation of noise power is not changed in the sub-band.After discrimination of noise and speech, subbands are combined. Finally, the decomposed components will be assembled back into the original signal without loss of information. This process is called reconstruction, or synthesis. This synthesis is called the inverse discrete wavelet transform (IDWT).

5. Conclusion and Future Work

We propose a method to reduce noise in recorded signal by using super soft thresholding algorithm in subband voice activity detection. The proposed system uses an estimation of the standard deviation for the each frame to select the threshold of the current frame and gives it as the input to the super-soft thresholding algorithm. We will calculate the most suitable threshold value that will use to remove noise from noisy signal, but also recover the original signal efficiently. If the threshold value is too high, it will also remove the contents of original signal and if the threshold value is too low, denoising will not work properly. In the future, there is still needs some improvement of the simulation model in order to provide a more resemble compared to the real world and another algorithm.

Acknowledgement

The author would like to thank all my teachers, fiends and my husband who taught and help me during the period of studies and thesis in Computer University, Mandalay.

References

[1] Prof. Dr. Ir. M. Steinbuch, Dr. Ir. M.J.G. van de Molengraft, “Wavelet Theory and Applications”, a literature study, R.J.E. Merry, DCT , Eindhoven University of Technology, Control Systems Technology Group Eindhoven, June 7 2005.53

[2] Siamak Rasoolzadeh, Moshen Rahimani,” Sub-band VAD Based on Continuous Noise Estimation in Wavelet Domain”, Department of Computer Engineering, Malayer Branch, Islamic Azad University, Malayer, Iran, Przegladelektrotechniczny (Electrical Review), ISSN 0033-2097, 2/2012

[3] S. V. Vaseghi, “Advanced Digital Signal Processing and Noise Reduction”, Wiley, third edition, 2005.

[4] Wahyu Kusumar,Prince Brave Guhyapativ,” Simulation Voice Recognition System for Controlling Robotic Applications” , Journal of Theoretical and Applied Information Technology,15 May 2012.

[5] Moe Pwint and Farook Sattar, “Speech/Nonspeech Detection Using Minimal Walsh Basis Functions” EURASIP Journal on Audio, Speech, and Music Processing, Mark Clements, Volume 2007

[6] N. Mokhtar*, H. Arof, F. R. Mahamd Adikan and M. Mubin,” Real Time Noise-Speech Discrimination in Time Domain for Speech Recognition Application”, Malaysia, Scientific Research and Essays, 4 January, 2011, Vol. 6(1), pp. 18-22

[7] Rajeev Aggarwal , Jai Karan Singh, Vijay Kumar Gupta, Sanjay Rathore Mukesh Tiwari and Dr. Anubhuti Khare,” Noise Reduction of Speech Signal using Wavelet Transform with Modified Universal Threshold”, International Journal of Computer Applications (0975 – 8887) , April 2011

[8] Sumithra M G, Member, IACSIT and Thanushkodi K,” Performance Evaluation of Different Thresholding Methods in Time Adaptive Wavelet Based Speech Enhancement “,IACSIT International Journal of Engineering and Technology Vol.1, December 2009, No.5

[9] Qiang Fu, Eric A. Wan,” Perceptual Wavelet Adaptive Denoising of Speech”, Geneva, Eurospeech 2003

[10] M. Rahmani, M. Mohammadi, A. Akbari, "Back ground noise control for speech enhancement," in Proc.14th ICEE, IRAN, Tehran, 2006.