e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 07
March 2018

Wavelets Transformation Based Speech Signal Enhancement by Removing Impulsive Noise

Dr.S.A.Mohammad Gazni
Principal,Ph.D(Dept. of Physics)
Osmania College,kurnool-518001,India.

Fasiha Shereen
M.Tech.(Dept. of ECE)
G.Pullaiah College of Engineering and technology,Kurnool.

Hunera Tarannum
M.tech. (Dept. of ECE)
Ravindra college of Engg. for women,Kurnool.

Juveria Bughra
Btech (Dept. of ECE)
G.Pullaiah College of Engg and Technology,Kurnool.

Abstract---*The presence of impulse noise in the speech signal has huge impact on the speech quality and on its performance in unprecedented levels. The removal of the impulse noise has been an area of research from over the years, but the work carried out in the past years fails to meet the desired requirements. A new approach based on the wavelets is proposed in this work to accomplish task of impulse noise removal in an accurate manner. The wavelet based impulse noise removal system achieves better results than the traditional short time fourier transform (STFT) system, the utilization of the multi-resolution property of the wavelet transform provides good time resolution at the higher frequencies. It uses two features of speech to discriminate speech from impulse noise: one is the slow time-varying nature of speech and the other is the Lipschitz regularity of the speech components. On the basis of these features, an algorithm has been developed to identify and suppress wavelet coefficients that correspond to impulse noise. The simulation results provide an environment of preserving the speech quality while processing the noisy speech signal in order to remove the impulse noise from the speech.*

Index Terms--- Speech signal, STFT, Wavelets, multi-resolution property, Speech quality, impulsive noise removal, speech enhancement

1. Introduction

The presence of impulse-like noise in speech can signifi- cantly reduces the intelligibility of speech and degrades automatic speech recognition performance.

Impulse noise is characterized by short bursts of acoustic energy having a wide spectral bandwidth and consisting of either isolated impulses or a series of impulses. Typical acoustic impulse noises include sounds of clicks in old phonograph recordings, of rain drops hitting a hard surface like the windshield

of a moving car, of popping popcorn, of typing on a keyboard, of indicator clicks in cars, and so on.

One difficulty with discerning impulse noise from speech is the wide temporal and spectral variation between different parts of speech, such as the periodic and low-frequency nature of vowels and the random and high-frequency nature of consonants. An effective algorithm should, therefore, consistently detect and remove the impulse noise whether it falls in vowels, consonants, or silent portions of speech.

However, these algorithms do not exploit the differences in spectral and temporal characteristics of speech and impulse noise to maximize the detection performance. Classical block processing methods such as the shorttime Fourier transform (STFT) algorithm or the linear prediction (LP) algorithm have also been used to detect or remove impulse-like sounds.4,5 However, two problems may result if classic block processing techniques are used: The first is determining the exact position of the impulse within the analyzed data frame—these methods give no straightforward.

information about the position of the impulse within the analyzed frame. It is possible, however, to reduce the frame size to achieve better resolution in time; but

e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 07
March 2018

doing this leads to the second problem where we lose the frequency resolution needed to effectively analyze the signal. The wavelet transform overcomes both of the difficulties due to its multiresolution property.6 In multi-resolution analysis, the window length or wavelet scale for analyzing the frequency components increases as the frequency decreases. This property enables the wavelet transform to have better time resolution for higher frequency components and better frequency resolution for lower ones. Consequently, by using the wavelet transform we have a relationship between time resolution and frequency resolution that is beneficial for detecting and removing impulse noise.

A wavelet approach for the detection and removal of impulse noise in degraded old analog recordings has been reported, whereby the wavelet coefficient corresponding to the scale where the audio signal is weak in comparison to the impulse noise is rectified, smoothed, and then a peak detector is applied to detect the impulses. However, since the peak detector uses a fixed threshold to detect the impulses, false detection may occur on occasions where the speech signal has high-frequency energy such as during consonants and fricatives; the other possibility is that it may fail to detect the smaller impulses that can be quite audible in regions where there is little or no speech signal. Further, the removal of the impulses in the method is done by substituting with uncorrupted wavelet coefficients from a nearby signal using autocorrelation properties. Although the approach works well if the impulses are sparsely located, substitution of the coefficients can be troublesome if a number of impulses are located in the same vicinity, an issue that is not considered in the method.

The presence of impulse-like noise in speech can significantly reduce the intelligibility of speech and degrade automatic speech recognition (ASR) performance. Impulse noise is characterized by a short burst of acoustic energy of either a single impulse or a series of impulses, with a wide spectral bandwidth. Typical acoustic impulse noises include sounds of machine gun firing, of rain drops hitting a hard surface like the windshield of a moving car, of typing on a keyboard, of indicator clicks in cars, of clicks in old analog recordings, of popping popcorn and so on. One difficulty with removing impulse noise from speech is the wide temporal and spectral variation between different parts of speech, such as the periodic and low-frequency nature of vowel signals and the random and high-frequency nature of consonants. An effective algorithm should, therefore, consistently remove the impulse noise whether it falls in vowels, consonants, or silent portions of speech. For audio signals, several time domain algorithms have been developed to detect and remove impulse noise. However, since these algorithms work in the time domain they do not effectively utilize the frequency information of the signals.

The wavelet transform overcomes both the difficulties due to its multi-resolution property [6]. In multi-resolution analysis, the window length or wavelet scale for analyzing the frequency components increases as the frequency decreases. This property enables the wavelet transform to have better time resolution for higher frequency components and better frequency resolution for lower ones. Consequently, by using the wavelet transform we have the benefit of both time and frequency resolutions to detect and remove impulse noise. In this paper, we utilize the slow time-varying nature of speech relative to the duration of an impulse to detect and suppress the impulses at finer scales. On the basis of the impulse detected in the finer scales, the corresponding wavelet coefficients of the impulse at the coarser scales are attenuated.

2. Discrete Wavelet Transform

Discrete wavelet transform skyline adhere the similar guidelines as discrete cosine transform. To transform digital image in wavelet transform, Wavelet rakes are used. Many rakes are available, the mostly usage rakes for watermarking are daubechies bi-orthogonal rake, haar wavelet rake and daubechies orthogonal rake. Every rake can dissolve the digital image into many frequencies. Now representation of four

International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 07
March 2018

frequencies in digital image obtained by rot of single level

These four frequencies representations are LL, LH, HL, HH sub-bands. The LH, HL, and HH sub-band shows the slender scale wavelet element in which the LL sub-band shows the poor level element that is less frequency element of the digital image. To gain highest level of rot the LL sub-band can be dissolved again. Now for the application this rot can subsist until the desired level of rot is obtained. To maintain the features of digital image the watermark can also be embedded in the three sub-bands LH, HL, and HH. The LL sub-band is milder to human eyes. Discrete wavelet transform is very suitable for dissolving the digital image. The wavelet transform for which waves are differently model, discrete wavelet transform can be defined. The profit of discrete wavelet transform compare to the Fourier transform is its capability of producing provisional conation and now it holds both location information and frequency data. Mother wavelet is responsible for causing the renditions and elucidation of the wavelets. Discrete wavelet transform counts the both high and low frequency elements by dividing the digital image into its separate frequency elements. Now for the edge quest the high frequency elements bequeath. On the other hand the low frequency components are anew divided into both high and low frequency components. The watermarking aim is being by the high frequency elements like as the human eye is mild on the edge diversity.

3. Proposed Method

Removal of Impulse Noise from Speech

Since our objective is the removal of impulse noise from speech, it is not critical that impulses of lower magnitudes that are perceptually inaudible be removed. It is important, however, that impulses of larger magnitudes be suppressed below the just-noticeable level difference (JNLD) to make them inaudible. The JNLD is not a fixed value and varies with the nature of the signal and sound pressure level (SPL). For example, for white noise the JNLD is around 0.7 dB for SPL between 40 and 100 dB, while for a 1 kHz tone the JNLD decreases from 1 to 0.2 dB as the SPL increases from 40 to 100 dB.

The temporal and spectral envelop of speech is slow time-varying in comparison to an impulse. This property is used to detect and suppress the wavelet coefficients that correspond to an impulse. Therefore, what is needed is a dynamic threshold for each wavelet level that varies in proportion to the smooth envelop of the absolute wavelet coefficients values, but, at the same time, not affected by impulse noise. That is, for scale s and sample n, such a dynamic threshold, τ(s, n), can be defined as

where operator Env[·] is the envelop of the signal that is unaffected by impulse noise and ks is a factor that is determined empirically for each level on the basis of the JNLD and the nature of the impulse noise. A median filter is known to possess the property where stepfunction type signals are preserved while at the same time robust to impulse noise . As such, the operator Env[·] in (8) can be replaced by a median filter of length N = 2K + 1 so that τ(s, n) becomes

The length of the median filter needs be adjusted so that it is sufficiently long in comparison to an impulse but short in comparison to a vowel or consonant. A wavelet coefficient would be considered to be that of an impulse if it greater than

e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 07
March 2018

τ(s, n); to suppress the impulse the coefficient is attenuated to a new coefficient, , given by

For impulses that occur in a consonant or in the non-voice portion of speech, suppression of the wavelet coefficients at coarser scales is as important as in the finer scales; this is because an impulse has a Lipschitz exponent that is usually greater than a consonant or background noise and, as such, its coefficients will not decay as fast at coarser scales. And if these coefficients at coarser scales are not suppressed adequately, they will be audible as low frequency thuds. However, for an impulse that occurs in the middle of a vowel, suppression of the corresponding coefficients at coarser scales is not as critical as in the finer scales. This is because the Lipschitz exponent of a vowel is much greater than that of an impulse, and at coarser scales the contribution to the coefficients comes mainly from the vowel. This also implies that the vowel will usually mask out the low-frequency portion of an impulse.

At larger scales, the use of (8) and (9) to detect the coefficients that correspond to an impulse become less effective if a discrete wavelet transform is used since the number of sample points decreases by half for the next increase in wavelet scale, thereby reducing the time resolution by half; furthermore, at larger scales an impulse will have much lesser contributions on a wavelet coefficient since other portions of the signal within the wavelet support length will also contribute to the coefficient. Therefore, for wavelet coefficients at larger scales it becomes more effective if the attenuation is done on the basis of the impulses detected in the smaller scales. If is the average decay of an impulse and = 1, where is the finest scale, then the attenuated wavelet coefficient,, for the coarser scale, , is given by

Where

is the absolute wavelet coefficient of the detected impulse in the finer scale sf , and kc is a constant that is empirically determined depending on the type of impulse noise and the JNLD at that scale.

4. Results

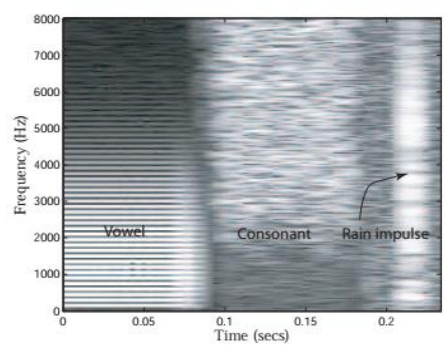

Fig. 4.1 (a) A typical spectrogram of a vowel, a consonant, and a rain impulse

International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 07
March 2018

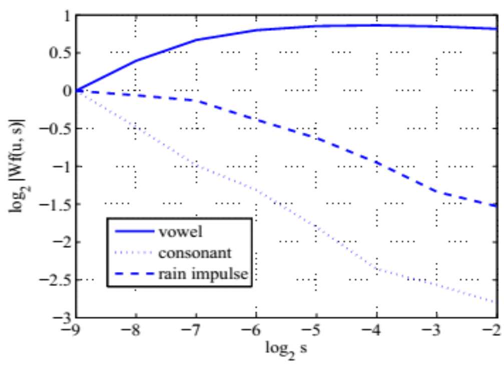

Fig. 4.3: Corresponding normalized wavelet magnitude variation across scales.

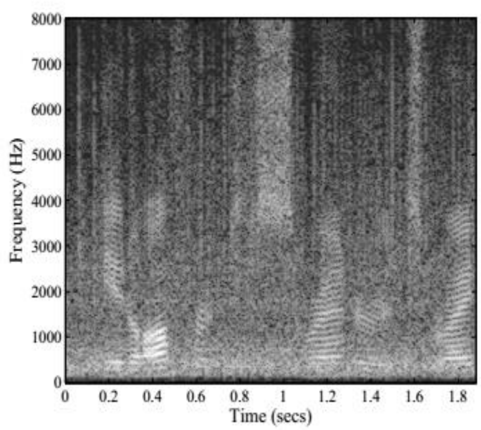

Fig. 4.3: Spectrogram of speech corrupted with sounds of raindrops hitting the windshield and roof of a car, which is moving at 60 km/hr

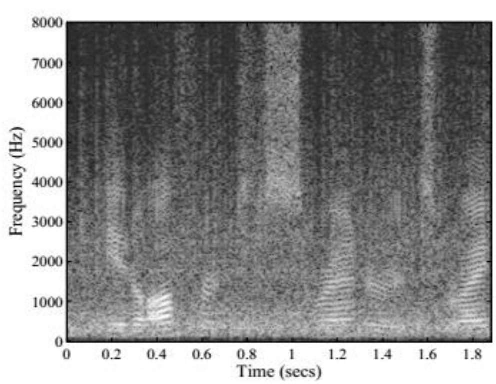

Fig. 4.4: Spectogram of speech after suppression of the rain impulses by the proposed algorithm.

5. Conclusion

The multi-resolution property based wavelet system is deployed in this paper to resolve the issue of impulse noise in the speech signal in an accurate manner. To discriminate the impulse from speech, it uses the slow time-varying nature of speech relative to an impulse, and the difference in regularity between an impulse and various parts of speech. On the basis of these differences, an algorithm was developed to identify and suppress wavelet coefficients that correspond to impulse noise. The simulation results shows that the proposed wavelet based produced good results compare to the existing SIFT based system.

References

[1] S. V. Vaseghi and P. J. W. Rayner, "Detection and suppression of impulse noise in speech communication systems," IEE Proceedings, vol. 137, pt. 1, no. 1, Feb. 1990.

[2] P. Esquef, M. Karjalainen, and V. Valimaki, "Detection of clicks in audio signals using warped linear prediction," 14th International Conference on Digital Signal Processing 2002, vol. 2, pp. 1085-1088, July 2002.

[3] C. Chandra, M. S. Moore, S. K. Mitra, "An efficient method for the removal of impulse noise from speech and audio signals," Proceedings of ISCAS 1998, vol. 4, pp. 206-208, Jun 1998.

[4] Z. Liu, A. Subramanya, Z. Zhang, J. Droppo, and A. Acero, " Leakage model and teeth clack removal for air- and bone-conductive integrated microphones," Proceedings of ICASSP 2005, vol. 1, pp. 1093-1096, Mar 2005.

[5] S. V. Vaseghi and R. Frayling-Cork, "Restoration of old gramophone recordings," Journal of Audio Eng. Soc., vol. 40, No. 10, pp. 791-801, 1992.

e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 07
March 2018

[6] S. Mallat, A Wavelet tour of signal processing, Second Edition, Academic Press 1998.

[7] S. Mallat and W. L. Hwang, “Singularity detection and processing with wavelets,” IEEE Trans. Inform. Theory, vol. 38, no. 2, pp. 617-643, 1992.

[8] E. Zwicker and H. Fastl, Psychoacoustics - Facts and Models, Second Edition, Springer 1999.

[9] L. Yin, R. Yang, M. Gabbouj, and Y. Neuvo, “Weighted median filters: a tutorial,” IEEE Trans. Circuits Syst, vol. 43, no. 3, pp. 157-192, 1996.

[10] S. F. Boll, “Suppression of acoustic noise in speech using spectral subtraction,” IEEE Trans. Acoust., Speech, Signal Processing, vol. 27, pp. 113-120, Apr. 1979.