# OPTIMIZED IMPULSE NOISE REMOVAL APPROACH BASED ON OPTIMAL WAVELETS

***N.NAGAMANI** (PG Scholar)* [1]

***T. PRASAD M.tech, Associate professor*** [2]

[1]***Department of ECE, Sri Venkateswara Institute of Science & Information***

## Abstract

*The impulse noise removal from the speech signal applications is still area of concern in the field of speech processing. The multi resolution property of the wavelet transform is solution to the problem in an effective way. We propose a new method for designing orthogonal wavelets that are optimized for detecting impulse noise in speech. In the method, the characteristics of the impulse noise and the underlying speech signal are taken into account and a convex optimization problem is formulated for deriving the optimal wavelet for a given support size. Performance comparisons with other well known wavelets show that the wavelets designed using the proposed method have much better impulse detection properties.*

**KEYWORDS:** Impulse noise, Wavelet transform, Speech signal, optimal wavelets

## INTRODUCTION

The presence of impulse-like noise in speech can significantly reduce the intelligibility of speech and degrade 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. For audio signals, several time-domain algorithms have been developed to detect and remove impulse noise.1–3 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 short time 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 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,7 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.

## PROPOSED METHOD

### DETECTION OF IMPULSE NOISE FROM SPEECH

In this section, we summarize the wavelet properties that influence the detection performance and describe a measure for evaluating the detection performance.

**(i) Wavelet properties and features for impulse detection**

A desirable wavelet for impulse detection is one that maximizes the coefficients for the impulse relative to the underlying signal in the finest scale [9]. Such a wavelet will correspondingly have a high pass analysis filter that maximizes the impulse noise relative to the underlying speech and background noise signals. If $Ps(\omega)$ and $Pi(\omega)$ are the power spectrums of the average speech and impulse noise power, respectively, then the ratio between the average impulse noise power and speech power in the finest scale, $Ri$, is dependent on the wavelet high pass analysis filter and given by

$$R_i = \frac{\sigma_i^2}{\sigma_s^2} \quad (1)$$

Where,

$$\sigma_i^2 = \int_{-\pi}^{\pi} \left|G(e^{jw})\right|^2 P_i(w)dw \approx \sum_i \left|G(e^{jw_i})\right|^2 P_i(w_i) \quad (2)$$

$$\sigma_s^2 = \int_{-\pi}^{\pi} \left|G(e^{jw})\right|^2 P_s(w)dw \approx \sum_i \left|G(e^{jw_i})\right|^2 P_s(w_i) \quad (3)$$

and $G(z)$ is the transfer function of the wavelet high pass filter. The design of an optimal wavelet for detecting the impulses should, therefore, seek to maximize $Ri$.

The other factor that influences the detection performance is the size of the wavelet support, which is dependent on the average width and energy of the impulse noise [9]. One way to determine the correct wavelet support for a given application is to design wavelets that maximize $Ri$ at various wavelet support sizes and then select the one with the best detection performance.

**(ii) Metrics to evaluate the detection performance**

To determine the most appropriate wavelet for impulse detection, we evaluate the discriminatory capability of the wavelet coefficients in the finest scale, with respect to the impulse noise. This is done by using a stability criterion derived from the scatter matrices [9]. For a one-dimensional, two-class scenario, the separability criterion for feature $x$ is given by

$$J = \frac{n_1(m_1 - m)^2 + n_2(m_2 - m)^2}{\sum_{x\in w_1}(x - m_1)^2 + \sum_{x\in w_2}(x - m_2)^2} \qquad (4)$$

Where $(m1, n1)$ and $(m2, n2)$ are the means and number of feature samples for classes $\omega 1$ and $\omega 2$, respectively. It has been shown [9] that a wavelet with a higher value of $J$ will correspondingly have better detection performance.

**(iii) DERIVING THE OPTIMALWAVELETS FOR IMPULSE DETECTION**

The optimal wavelets are designed to maximize the ratio of impulse noise power to speech power in the finest scale. At the same time, the necessary constraints required for an orthogonal wavelet need to be imposed.

If $H(z)$ corresponds to the transfer function of a low pass analysis filter of an orthogonal wavelet given by

$$H(z) = h(0) + h(1)z^{-1} + \cdots + h(L-1)z^{-(L-1)} \qquad (5)$$

then the high pass counterpart, $G(z)$, can be obtained by taking the alternating flip of $H(z)$ [10]; that is

$$G(z) = -z^{-(L-1)}H(-z^{-1}) \qquad (6)$$

where $L$ is assumed to be even. To ensure that the wavelet filter bank is orthogonal, the filter coefficients need to satisfy the *double-shift orthogonality* condition [10], given by

$$\sum_n h(n)h(n-2k) = \delta(k), \quad for\ k = 0,1,\dots,(L/2)-1 \qquad (7)$$

where $\delta(k)$ is the delta function. For the existence of the wavelet $\psi(t)$, the following condition must also hold true [11]:

$$H(e^{jw})\big|_{w=0} = \sum_n h(n) = \sqrt{2} \qquad (8)$$

As in the design of signal-adapted filterbanks by Moulin et al[12], the formulation of the optimization problem becomes more tractable if we use the autocorrelation sequence of the filter coefficients given by

$$r_h(l) = \begin{cases} \sum_{n=0}^{L-l-1} h(n)h(n+l) & l \geq 0 \\ r_h(-l) & l < 0 \end{cases} \qquad (9)$$

Therefore, in terms of the auro correlation parameters, the double shift orthogonality condition in (7) can be expressed as

$$r_h(2k) = \delta(k), \quad for\ k = 0,1,\dots,\left[\frac{L-1}{2}\right] \qquad (10)$$

and the necessary condition in (8) as

$$\sum_{m=1}^{L-1} r_h(m) = 0.5 \qquad (11)$$

by exploiting the orthogonality condition in (7) and the symmetry property in (9). Correspondingly, using (6) and (9) in (2) and (3) the average power of the impulse noise and speech in the finest scale are given by

$$\sigma_i^2 \approx \sum_n \left[ r_h(0) + 2\sum_{l=1}^{L-1} (-1)^l\, r_h(l)\cos(w_n l) \right] P_i(w_n) = 1^T C_i A r \quad (12)$$

$$\sigma_s^2 \approx \sum_n \left[ r_h(0) + 2\sum_{l=1}^{L-1} (-1)^l\, r_h(l)\cos(w_n l) \right] P_s(w_n) = 1^T C_s A r \quad (13)$$

Where,

$$r = [r_h(0) \ldots r_h(L-1)]^{\,T} \quad (14)$$

$$A = \begin{bmatrix} a_{00} & \cdots & a_{0(L-1)} \\ \vdots & \vdots & \vdots \\ a_{(N-1)0} & \cdots & a_{(N-1)(L-1)} \end{bmatrix} \quad (15)$$

$$C_i = diag\left(c_0^{(i)}, \ldots, c_{(N-1)}^{(i)}\right) \quad (16)$$

$$C_s = diag\left(c_0^{(s)}, \ldots, c_{(N-1)}^{(s)}\right) \quad (17)$$

$$a_{nl} = 2(-1)^l \cos(w_n l) \quad (18)$$

$$c_n^{(i)} = P_i(w_n), \quad w_n \in [-\pi, \pi] \quad (19)$$

$$c_n^{(s)} = P_s(w_n), \quad w_n \in [-\pi, \pi] \quad (20)$$

and $N$ is the number of samples. The optimization is formulated as the minimization of $\sigma_s^2$ while keeping $\sigma_i^2$ constant so that $Ri$ in (1) is maximized. Consequently, after incorporating the double-shift orthogonality constraint in (10) and the necessary condition in (11), the optimization problem is given by

$$minimize\ \sigma_s^2$$

$$subject\ to:\ \sigma_i^2 = constant$$

$$r_h(2m) = 0, \quad for\ m = 1, \ldots, \left\lceil \frac{L-1}{2} \right\rceil$$

$$r_h(0) = k$$

$$\sum_{m=1}^{L-1} r_h(m) = 0.5k$$

where $k\ and\ rh(m)$ are optimization variables. Note that the last two equality constraints in (21) ensure that the necessary condition in (11) is satisfied when we set $rh(0) = 1$. Replacing $\sigma_s^2$ and $\sigma_i^2$by their matrix representations, (21) can be expressed as a convex optimization problem given by

$$minimize\ 1^T C_s A r$$

$$subject\ to:\ 1^T C_i A r = constant$$

$$r_h(2m) = 0, \quad for\ m = 1, \ldots, \left\lceil \frac{L-1}{2} \right\rceil$$

$$r_h(0) = k$$

$$\sum_{m=1}^{L-1} r_h(m) = 0.5k$$

$$Ar > 0$$

Where $\boldsymbol{r}\ and\ k$ are optimization variables and $\mathbf{0} \in \boldsymbol{R}^{\boldsymbol{N}}$. The inequality constraint in

(22) is a positivity constraint to ensure that the magnitude is always positive. Once we obtain the optimal autocorrelation vector $\boldsymbol{r}_{opt}$ we recover the minimum-phase low-pass wavelet filter coefficients $h_{mp}(n)\ from\ \boldsymbol{r}_{opt}$ using spectral factorization [13]. The filter coefficients obtained are then appropriately scaled so that the necessary condition in (8), or equivalently in (11), is satisfied.

## SIMULATION RESULTS

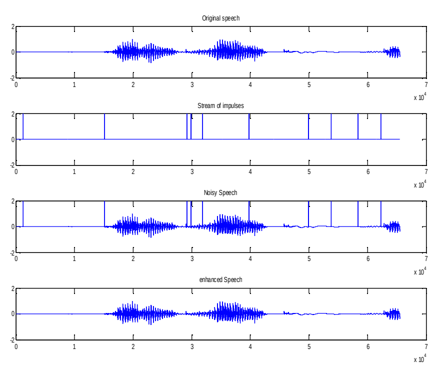


Figure 1: Noisy, original and enhanced speech signal

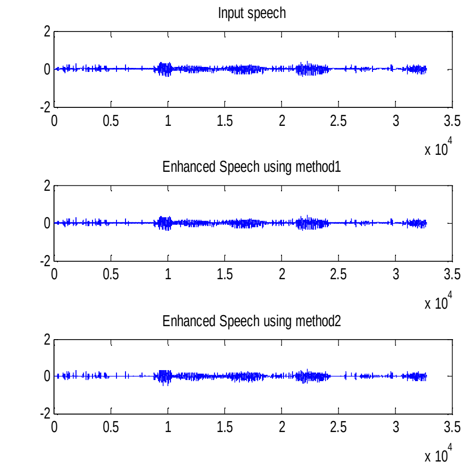


Figure 2: Enhanced speech signal method 1 and method 2

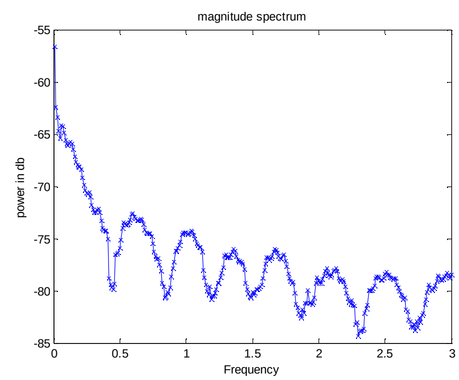


Figure 3: Magnitude Spectrum of speech signal

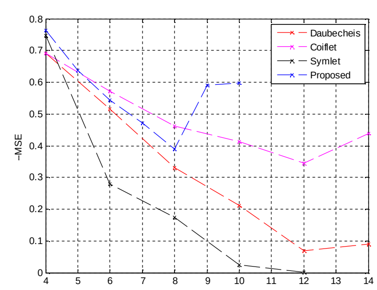


Figure 4: Scaling level by different levels

## CONCLUSION

A new method for designing orthogonal wavelets that are optimized for detecting impulse noise in speech has been described. In the method, the characteristics of the impulse noise and the underlying speech signal are taken into account and a convex optimization problem was formulated for deriving the optimal wavelet for a given support size. Performance comparisons with other well-known wavelets showed that the wavelets designed using the proposed method have superior impulse detection properties.

## REFERENCES

[1] P. Sprechmann, A. Bronstein, J.-M. Morel, and G. Sapiro, “Audio restoration from multiple copies,” Proceedings of ICASSP 2013, pp. 878-882.

[2] C. Zheng, X. Chen, S. Wang, R. Peng, and X. Li, “Delayless method to suppress transient noise using speech properties and spectral coherence,” Proceedings of the 135th AES Convention, New York, USA (2013).

[3] Z. Liu, A. Subramanya, Z. Zhang, J. Droppo, and A. Acero, “ Leakage model and teeth clack removal for air- and boneconductive integrated microphones,” in Proceedings of ICASSP 2005, vol. 1, pp. 1093-1096.

[4] S. V. Vaseghi and R. Frayling-Cork, “Restoration of old gramophone recordings,” J. Audio Eng. Soc., 40, 791-801 (1992).

[5] J. A. Moorer, “Dsp restoration techniques for audio,” Proceedings of ICIP 2007, vol. 4, pp. 5-8.

[6] S. Mallat, A Wavelet Tour of Signal Processing, 2nd ed. (Academic, San Diego, 1998), pp. 221-228.

[7] S. Montresor, J. C. Valiere, J. F. Allard, and M. Baudry, “The restoration of old recordings by means of digital techniques,” in Proceedings of the 88th AES Convention, Montreux, Switzerland (1990).

[8] R. C. Nongpiur, “Impulse noise removal in speech using wavelets,” Proceedings of ICASSP 2008, pp. 1593-1597.

[9] R. C. Nongpiur and D. J. Shpak, “Impulse-noise suppression in speech using the stationary wavelet transform,”J. Acoust. Soc. Am., 133(2), 866-879 (2013).

[10] G. Strang and T. Nguyen, Wavelets and filter banks, WellesleyCambridge Press (1997).