www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 2 Issue 12, Dec.2013 Page No. 3435-3438

Speech Enhancement through Elimination Of Impulsive Disturbance Using Log MMSE Filtering

D.Koti Reddy[1], T.Jayanandan[2] and C.L Vijay Kumar[3]

[1] Student, Prathyusha Institute Of Technology and Management,
Thiruvallur, TamilNadu, India
Kotireddy.datla@gmail.com

[2] Assistant Professor, Prathyusha Institute Of Technology and Management,
Thiruvallur, TamilNadu, India
jayanandan.t.18@gmail.com

[3] Assistant Professor, Prathyusha Institute Of Technology and Management,
Thiruvallur, TamilNadu, India
Vijai10.7.1988@gmail.com

Abstract:

The project presents an enhancement of the speech signal by removal of impulsive disturbance from noisy speech using log minimum mean square error filtering approach. Impulsive noise has a potential to degrade the performance and reliability of Speech signal. To enhance the speech component from impulsive disturbance we go for emphasis, signal segmentation and log MMSE filtering. In preprocessing of audio signals start with pre-emphasis refers to a system process designed to increase the magnitude of some frequencies with respect to the magnitude of other frequencies. Emphasis refers to a system process designed to increase the magnitude of some frequencies with respect to the magnitude of other frequencies in order to improve the overall signal-to-noise ratio. Then the signal samples are segmented into fixed number of frames and each frame samples are evaluated with hamming window coefficients. Mean-Square Error Log-Spectral Amplitude (MMSE), which minimizes the mean-square error of the log-spectra, is obtained as a weighted geometric mean of the gains associated with the speech signal. The performance of the filtering is measured with signal to noise ratio, Perceptual Evaluation of Speech Quality (PESQ), Correlation

.

Index Terms—**Inventory-style speech enhancement, modified imputation, uncertainty-of-observation techniques.**

I. INTRODUCTION

The fundamental purpose of speech is communication, i.e., the transmission of messages. A message represented as a sequence of discrete symbols can be quantified by its information content in bits, and the rate of transmission of information is measured in bits/second (bps). In speech production, as well as in many human-engineered electronic communication systems, the information to be transmitted is encoded in the form of a continuously varying (analog) waveform that can be transmitted, recorded, manipulated, and ultimately decoded by a human listener. In the case of speech, the fundamental analog form of the message is an acoustic waveform, which we call the speech signal. Speech signals can be converted to an electrical waveform by a microphone, further manipulated by both analog and digital signal processing, and then converted back to acoustic form by a loudspeaker, a telephone handset or headphone, as desired. Signals are usually corrupted by noise in the real world. To reduce the influence of noise, two research topics are the speech enhancement and speech recognition in noisy environments have arose. For the speech enhancement, the extraction of a signal buried in noise, adaptive noise cancellation (ANC) provides a good solution. In contrast to other enhancement techniques, its great strength lies in the fact that no *a priori* knowledge of signal or noise is required in advance. The advantage is gained with the auxiliary of a secondary input to measure the noise source. The cancellation operation is based on the following principle. Since the desired signal is corrupted by the noise, if the noise can be estimated from the noise source, this estimated noise can then be subtracted from the primary channel resulting in the desired signal. Traditionally, this task is done by linear filtering. In real situations, the corrupting noise is a nonlinear distortion version of the source noise, so a nonlinear filter should be a better choice. In the typical speech enhancement methods based on STFT, only the magnitude spectrum is modified and phase spectrum is kept unchanged. It was believed that the magnitude spectrum includes most of the information of the speech, and phase spectrum contains little of that. Furthermore, the human auditory system is phase deaf. For above reason, in typical speech enhancement algorithms, such as Spectral subtraction (SS), MMSE-STSA or MAP algorithm, the speech enhancement process is on the basis of spectral magnitude component only and keep the phase component unchanged.

II. WAVELET BASED DENOISING

Wavelets are mathematical functions defined over a finite interval and having an average value of zero that transform data into different frequency components, representing each component with a resolution matched to its scale. The basic idea of the wavelet transform is to represent any arbitrary function as a superposition of a set of such wavelets or basis functions. These basis functions or baby wavelets are obtained from a single prototype wavelet called the mother wavelet, by dilations or contractions (scaling) and translations (shifts). They have advantages over traditional Fourier methods in analyzing physical situations where the signal contains discontinuities and sharp spikes. Many new wavelet applications such as image compression, turbulence, human vision, radar, and earthquake prediction are developed in recent years. In wavelet transform the basic functions are wavelets. Wavelets tend to be irregular and symmetric. All wavelet functions, $w(2kt - m)$, are derived from a single mother wavelet, $w(t)$. This wavelet is a small wave or pulse like the one shown in Fig. 3.2.

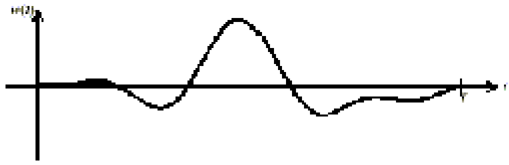
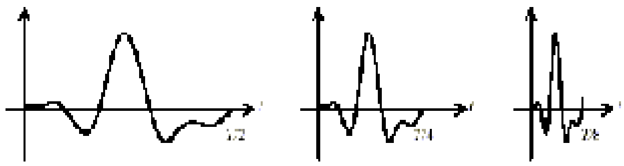


Fig. 1 Mother wavelet w(t)

Normally it starts at time $t = 0$ and ends at $t = T$. The shifted wavelet $w(t - m)$ starts at $t = m$ and ends at $t = m + T$. The scaled wavelets $w(2kt)$ start at $t = 0$ and end at $t = T/2k$. Their graphs are $w(t)$ compressed by the factor of $2k$ as shown in Fig. 3.3. For example, when $k = 1$, the wavelet is shown in Fig 3.3 (a). If $k = 2$ and 3, they are shown in (b) and (c), respectively.



(a)w(2t) **(b)w(4t)** **(c)w(8t)**

Fig. 2 Scaled wavelets

The wavelets are called orthogonal when their inner products are zero. The smaller the scaling factor is, the wider the wavelet is. Wide wavelets are comparable to low-frequency sinusoids and narrow wavelets are comparable to high-frequency sinusoids. The reconstruction of the image is achieved by the inverse discrete wavelet transform (IDWT). The values are first up sampled and then passed to the filters.

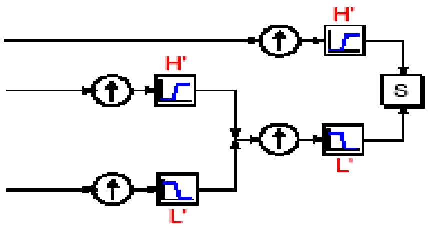


Fig. 3 Wavelet Reconstruction

The wavelet analysis involves filtering and down sampling, whereas the wavelet reconstruction process consists of up sampling and filtering. Up sampling is the process of lengthening a signal component by inserting zeros between samples as shown in fig

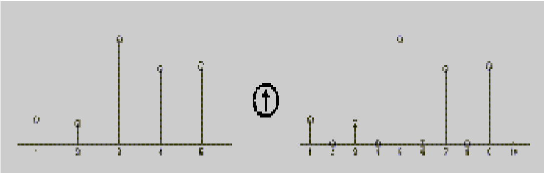


Fig. 4 Reconstruction using up sampling.

Wavelet denoising is considered a non-parametric method. Thus, it is distinct from parametric methods in which parameters must be estimated for a particular model that must be assumed a priori.

$$X(t) = S(t) + N(t)$$

Assume that the observed data contains the true signal S(t) with additive noise N(t) as Functions in time t to be sampled. Let W(·) and W−1 (·) denote the forward and inverse wavelet transform operators. Let D (·,λ) denote the denoising operator with soft threshold λ. We intend to wavelet denoised X(t) in order to recover Ŝ(t) as an estimate of S(t).

III. LOG MMSE FILTERING AND SIGNAL SEGMENTATION

The problem is discussed in more generality than in many other expositions specifically we allow for general filter delays (to accommodate the pitch filtering problem, for instance) and cover both the stochastic case and block-based analyses with a single formalism. For mean-square error computations, we will only need to use at most second order statistical properties (correlations and means). For the case of stochastic signals, these notes look at the derivation of the correlation values required for a minimum mean-square error solution. We also examine systems which involve cyclo stationary signals (interpolation filter, for instance).

The important linear prediction problem is examined in detail. This includes the setup for non-equally spaced delay values. For the equally spaced delay case, we can develop a rich set of results. For the least-squares problem, these notes give a generalized view of windowing: windowing the data and/or windowing the error. This view subsumes the traditional special cases, viz the auto correlation and covariance methods. These notes present a number of examples based on "real" signals. With the background developed, the results are obtained with relatively straightforward MATLAB scripts. The results illustrate the useful insights that can be obtained when minimum mean-square error theory is appropriately fleshed out.

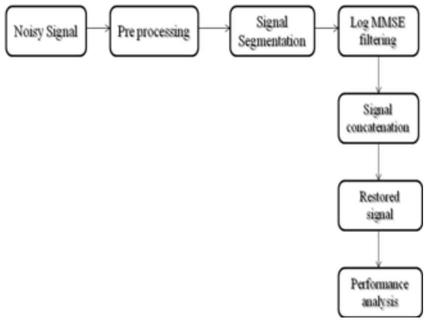


Fig:5. Block Diagram

Consider a filter with an input x[n] and an output y[n] given by

$$y[n] = \sum_{k=0}^{M-1} w_k^* x[n - D_k],$$

Where the W_k values1 weight the samples of the input signal at different delays D_k. We require that the delays be distinct.

A conventional causal FIR filter would simply have Dk = k for k = 0, M − 1. We will keep the delays general until later in this document, when it becomes useful to specialize them to get further results. The goal is to find a set of filter coefficients that minimize the squared error between the output of the filter y[n] and a desired signal d[n].

First we write the filtering operation in vector form,

$$y[n] = \mathbf{w}^H \mathbf{x}[n],$$

Where,

$$\mathbf{w} = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_{M-1} \end{bmatrix}, \quad \mathbf{x}[n] = \begin{bmatrix} x[n - D_0] \\ x[n - D_1] \\ \vdots \\ x[n - D_{M-1}] \end{bmatrix}.$$

To be able to handle both the stochastic case and block-based squared-error cases with a single formulation, we define an "averaging" operation with a bar over an expression. For the case of ensemble averaging, this is the expectation operator. For other cases, it will signify a sum of squares. In many of the cases we study, the averaging operation will remove the dependency on n. For the case of wide-sense stationary processes, results are reviewed in Appendix B.

The error is,

$$e[n] = d[n] - y[n]$$
$$= d[n] - \mathbf{w}^H \mathbf{x}[n].$$

IV. SIMULATION RESULTS

We analyzed the performance of the proposed method within a variety of noise scenarios. Results were compared to four established reference techniques. Two of these reference techniques were log-MMSE enhancers after Ephraim and Malah. One of these methods, referred to as **log-MMSE(MS)**, employed the *Minimum Statistics* technique developed by Martin [11] to estimate the underlying noise power. The other method, referred to as **log-MMSE(RA)**, employed the same VAD-supported *Recursive Averaging* that was also used in the "log-MMSE Filter" block of Fig. 5. The performance gains between the **log-MMSE(RA)** method and the proposed method are, therefore, directly attributable to the inventory search and the subsequent cepstral smoothing. As a third reference method we chose the *Multiband Spectral Subtraction* (**MBSS**) technique proposed by Kamath and Loizou [2] and lastly, we also implemented a slightly modified version of the inventory-style baseline system.

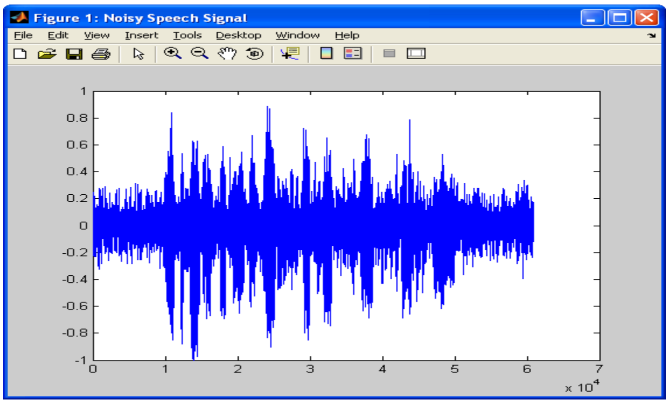


Fig:6. Noisy Speech Signal

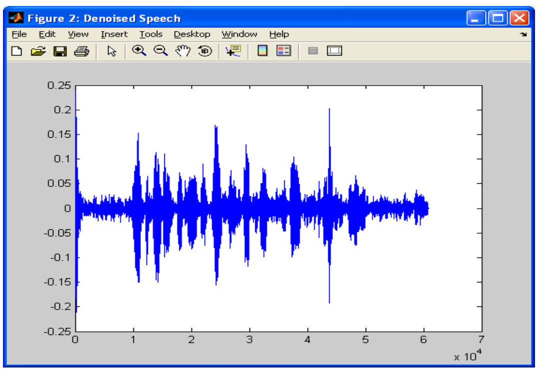


Fig:7.Denoised Signal

V. CONCLUSION

The project presented that an enhancement of the speech signal by removal of impulsive disturbance based on log spectral gain filtering approach. Here, Mean-Square Error Log-Spectral Amplitude was used to minimize the mean-square error of the log-spectra, is obtained as a weighted geometric mean of the gains associated with the speech signal effectively. It provided that better results in terms of performance parameters, processing time and speech signal quality rather than prior methods. This system will be enhanced with a modified filtering method to restore signals with better accuracy rather than Log spectra.

VI. REFERENCES

[1] P. Vary and R. Martin, Digital Speech Transmission: Enhancement, Coding and Error Concealment. Chichester, U.K.: Wiley, 2006.

[2] P. C. Loizou, Speech Enhancement—Theory and Practice. Boca Raton, FL, USA: CRC, Taylor and Francis, 2007.

[3] X. Xiao and R. M. Nickel, "Speech enhancement with inventory style speech resynthesis," IEEE Trans. Audio, Speech, Lang. Process., vol. 18, no. 6, pp. 1243–1257, Aug. 2010.

[4] J. Ming, R. Srinivasan, and D. Crookes, "A corpus-based approach to speech enhancement from nonstationary noise," IEEE Trans. Audio, Speech, Lang. Process., vol. 19, no. 4, pp. 822–836, May 2011.

[5] J. Ming, R. Srinivasan, and D. Crookes, "A corpus-based approach to speech enhancement from nonstationary noise," in Proc. INTERSPEECH, Makuhari, Japan, Sep. 2010, pp. 1097–1100.

[6] R. M. Nickel and R. Martin, "Memory and complexity reduction for inventory-style speech enhancement systems," in Proc. EUSIPCO, Barcelona, Spain, Sep. 2011, pp. 196–200.