International Journal of Engineering Research and General Science Volume 3, Issue 1, January-February, 2015
ISSN 2091-2730

Speech Enhancement through Elimination of Impulsive Disturbance Using Log MMSE Filtering

Sonali N. Malshikare, Prof. V. M. Sardar

Department of Electronics and Telecomm Engineering, JSPMs Jayawantrao Sawant College of Engg ,Hadapsar, Pune 28, India
sonalimalshikare84@gmail.com, 9763321207

Abstract— The 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. It provided that better results in terms of performance parameters, processing time and speech signal quality rather than prior methods.

Keywords— Inventory-style speech enhancement, modified imputation, uncertainty-of-observation techniques.

INTRODUCTION

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 pre processing 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.remaining contents.

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 humanengineered 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.

WAVELET TRANSFORM

Whether we like it or not we are living in a world of signals. Nature is talking to us with signals: light, sounds… Men are talking to each other with signals: music, TV, phones…

The human body is equipped to survive in this world of signals with sensors such as eyes and ears, which are able to receive and process these signals. Consider, for instance, our ears: they can discriminate the volume and tone of a voice. Most of the information our ears process from a signal is in the frequency content of the signal.

Scientists have developed mathematical methods to imitate the processing performed by our body and extract the frequency information contained in a signal. These mathematical algorithms are called transforms and the most popular among them is the Fourier Transform.

The second method to analyze non-stationary signals is to first filter different frequency bands, cut these bands into slices in time, and then analyzes them.

The wavelet transform uses this approach. The wavelet transform or wavelet analysis is probably the most recent solution to overcome the shortcomings of the Fourier transform. In wavelet analysis the use of a fully scalable modulated window solves the signal-cutting problem. The window is shifted along the signal and for every position the spectrum is calculated. Thenthis process is repeated many times with a slightly shorter (or longer) window for every new cycle.

In the end the result is a collection of time-frequency representations of the signal, all with different resolutions. Because of this collection of representations, we can speak of a multiresolution analysis. In the case of wavelets, we normally do not speak about time-frequency.

The discrete wavelet transform (DWT) was developed to apply the wavelet transform to the digital world. Filter banks are used to approximate the behavior of the continuous wavelet transform. The signal is decomposed with a high-pass filter and a low-pass filter. The coefficients of these filters are computed using mathematical analysis and made available to you.

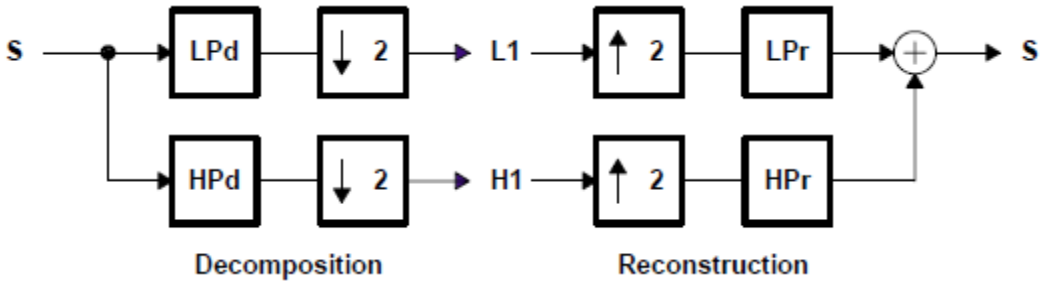

Figure 1. Discrete wavelet transform

Where
LPd: Low Pass Decomposition Filter
HPd: High Pass Decomposition Filter
LPr: Low Pass Reconstruction Filter
HPr: High Pass Reconstruction Filter

The hp[n] coefficients are used as the low-pass reconstruction filter (LPr).
The coefficients for the filters HPd, LPd and HPr are computed from the h[n] coefficients as follows:

- High-pass decomposition filter (HPd) coefficients

$g[n] = (-1)^n h[L-n]$ (L: length of the filter)

- Low-pass reconstruction filter (LPr) coefficients

$h[n] = h[L-n]$ (L: length of the filter)

- High-pass reconstruction filter (HPr) coefficients

$g[n] = g[L-n]$ (L: length of the filter)

The Daubechies filters for Wavelets are provided in the C55x IMGLIB for $2 \leq p \leq 10$. Since there are several sets of filters, we may ask ourselves what are the advantages and disadvantages to using one set or another.

First we need to understand that we will have perfect reconstruction no matter what the filter length is. However, longer filters provide smoother, smaller intermediate results. Thus, if intermediate processing is required, we are less likely to lose information due to necessary threshold or saturation. However, longer filters obviously involve more processing.

4.1.4 Wavelets and Perfect Reconstruction Filter Banks:

Filter banks decompose the signal into high- and low-frequency components. The low-frequency component usually contains most of the frequency of the signal. This is called the approximation. The high-frequency component contains the details of the signal.

Wavelet decomposition can be implemented using a two-channel filter bank. Two-channel filter banks are discussed in this section briefly. The main idea is that perfect reconstruction filter banks implement series expansions of discrete-time signals.

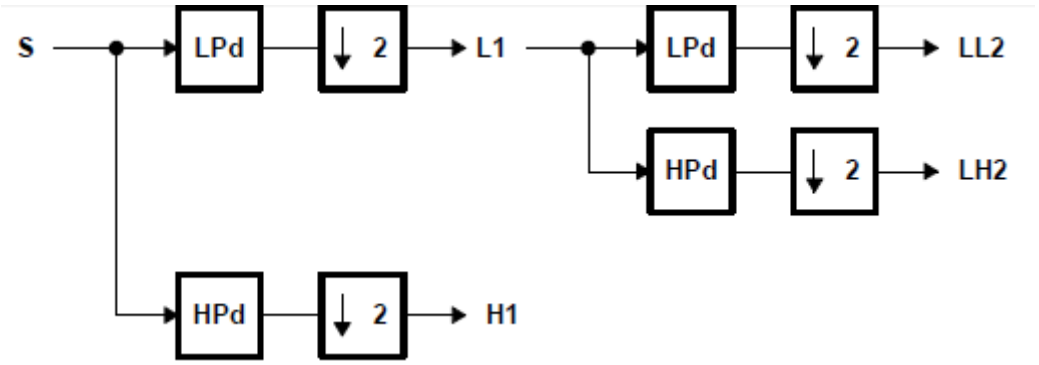

Figure 2. Two level wavelet decomposition

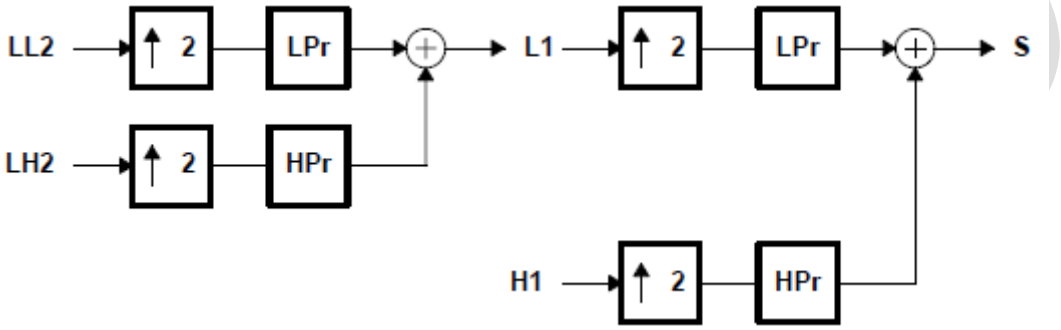

Figure 3. Two level wavelet reconstruction

The input and the reconstruction are identical; this is called perfect reconstruction. Two popular decomposition structures are pyramid and wavelet packet. The first one decomposes only the approximation (low-frequency component) part while the second one decomposes both the approximation and the detail (high-frequency component).

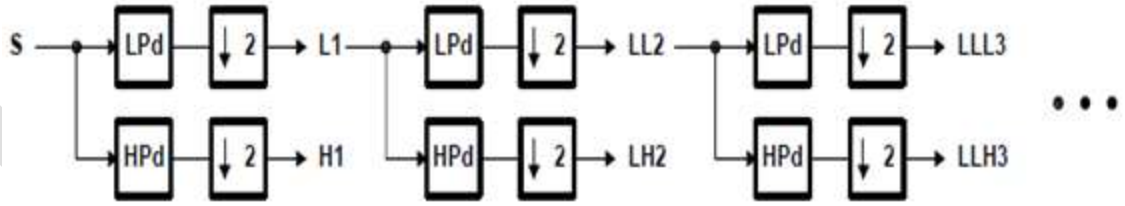

Figure 4. Pyramid packet

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 $W(\cdot)$ and $W^{-1}(\cdot)$ denote the forward and inverse wavelet transform operators. Let $D(\cdot,\lambda)$ denote the denoising operator with soft threshold λ. We intend to wavelet denoised X(t) in order to recover $\hat{S}(t)$ as an estimate of S(t).

Threshold Detection

The threshold will be selected for shrinking high frequency subband coefficients to remove the noise.

The wavelet threshold will be determined by bayesian shrinkage method and it is given by,

sigmax = sqrt(max(sigma-sigmahat.^2))

Where,

sigma = sum(Coeff.^2)./L ; L = Number of coefficient.

sigmahat = Med(abs(C(var:length(C))))/0.6745

Where,

var=length(C)-S(size(S,1)-1,1)2+1
C - Coefficient Matrix,
S – approximation and detailed coefficient details

Finally , the threshold is based on,

T = max(abs(X)) if sigmax = 0
T = sigmahat.^2 /sigmax is sigmax ~= 0

The threshold is calculated based on the sigmax value and then soft thresholding is used for noise removal.

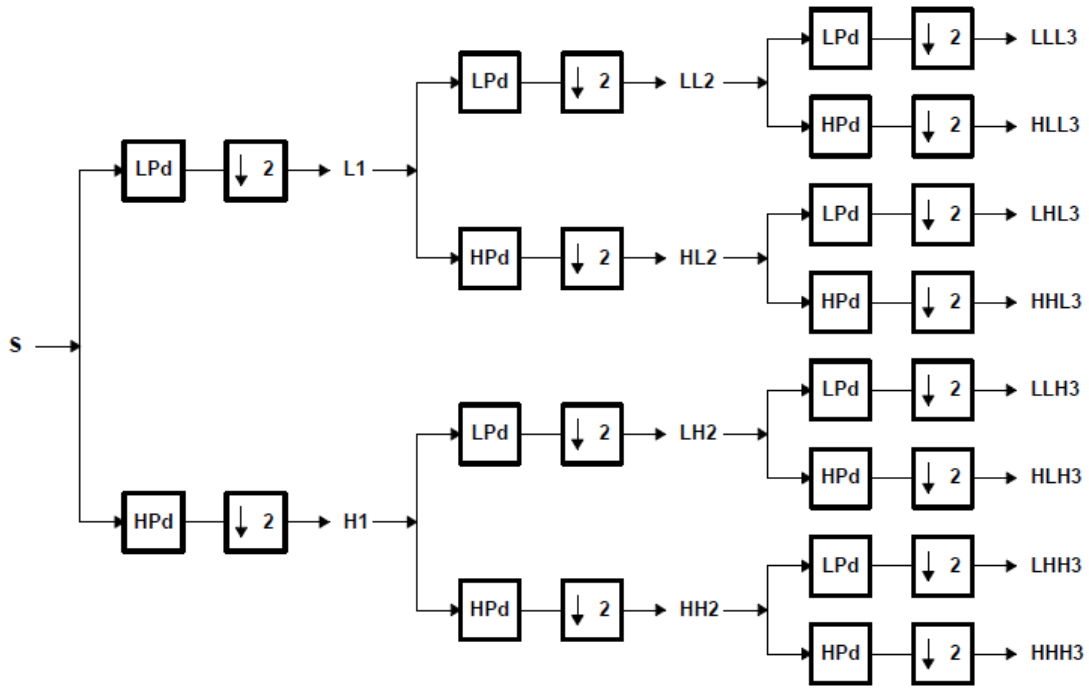

Figure 8 . Wavelet Packet Decomposition

BLOCK DIAGRAM

The proposed system consists of following steps:

1.INPUT SIGNAL:

Input signal is applied to the system which is in .wav format from the database of the system.Input signal is any sample of noisy speech signal which is stored by.wav file.

2. PREPROCESSING:

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 in order to improve the overall signal-to-noise ratio by minimizing the adverse effects of such phenomena as attenuation distortion or saturation of recording media in subsequent parts of the system. The mirror operation is called de-emphasis, and the system as a whole is called emphasis.

Pre-emphasis is achieved with a pre-emphasis network which is essentially a calibrated filter. This network composed of two resistors and one capacitor. The frequency response is decided by special time constants. The cutoff frequency can be calculated from that value.

Pre-emphasis is commonly used in telecommunications, digital audio recording, record cutting, in FM broadcasting transmissions, and in displaying the spectrograms of speech signals.

De-emphasis is the complement of pre-emphasis, in the anti noise system called emphasis. Emphasis is a system process designed to decrease, (within a band of frequencies), the magnitude of some (usually higher) frequencies with respect to the magnitude of other (usually lower) frequencies in order to improve the overall signal-to-noise ratio by minimizing the adverse effects of such phenomena as attenuation differences or saturation of recording media in subsequent parts of the system.

3.SIGNAL SEGMENTATION:

The signal samples are segmented into fixed number of frames and each frame samples are evaluated with hamming window coefficients.

The total frames are calculated by,

Fn = (Ls – Ns)/ (Ns*Sp) +1

Where, Ls = length of signal, Ns = Length of each frame

Sp = Shift Percentage

Finally the samples of each frames are separated from input signal using Fn and Sp and its scaled by the hamming window coefficients.

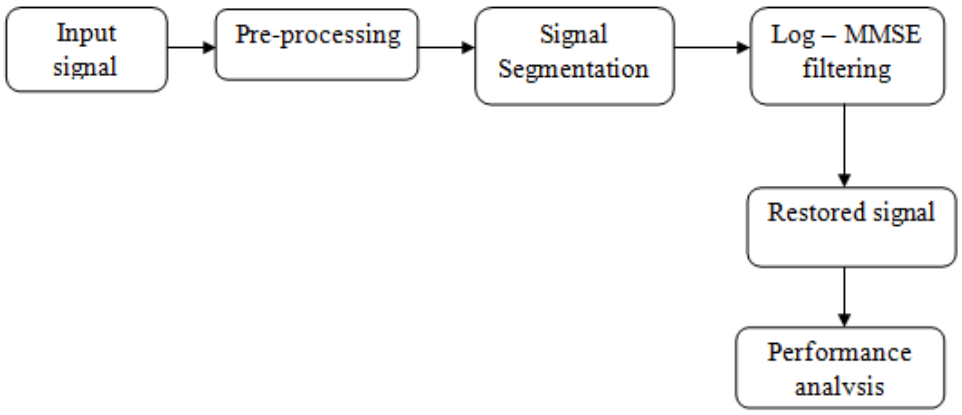

Figure 9. Block Diagram

4. LOG-MMSE FILTERING:

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.

After the signal segmentation, the magnitude and phase spectrum from noisy signal are computed by applying fast fourier transform. The magnitude of noisy signal spectrum are further utilized for filtering process and signal phase kept same.

The restored signal magnitude spectra is obtained by,

Rs = G .* Y

Where, G – Log spectral amplitude Gain function

Y – magnitude response of noisy signal

The log spectral gain function is defined by,

G = x./(1+x) exp(eint (v))

Where, v = x./(1+x) * r

x and r– priori and posteriori signal to noise ratio

eint – exponential integral

The posteriori snr is defined by, r = (Y.^2)/ lamda

lamda = E[(Y).^2]

Where, lamda - noise power spectrum variance

E – Mean value

Complex spectrogram obtained by Filtered magnitude spectrum is combined with noisy signal phase spectrum. The restored signal is reconstructed by applying inverse fast fourier transform to this complex spectrogram.The performance of filtering is measured with SNR evaluation and it is defined by,

SNR = 10log10(Msig^2./(sum((inp - output).^2)./Ls))

Where, Msig = Maximum amplitude of signal

inp, output = Noisy input signal and restored output

5.PERFORMANCE ANALYSIS:

The performance of log spectral filtering will be measured based on,

Correlation

Perceptual Evaluation of Speech Quality

Log Likelihood ratio

Correlation Coefficient: It is used to find the similarity between two different speech signals. It will be described by,

```
Cor_coef = [sum(sum(u1.*u2))] /
           [sqrt(sum(sum(u1.*u1))*sum(sum(u2.*u2)))];
```

Where, u1 = F1 – mean of F1, u2 = F2 – mean of F2

F1 – Original signal and F2 – Restored signal

PESQ: Perceptual evaluation of speech quality predicts with high correlation subjective mean opinion score listening tests and it is computed by,

PESQ = 4.5 – 0.1Dind – 0.0309Aind

Where, Dind and Aind are known as the average disturbance and the average asymmetrical disturbance values.

LLR: Log likelihood ratio expresses that how many times more likely the data are under one model than the other and is the ratio of the likelihood function varying the parameters over two different sets.

EXPERIMENTAL RESULTS

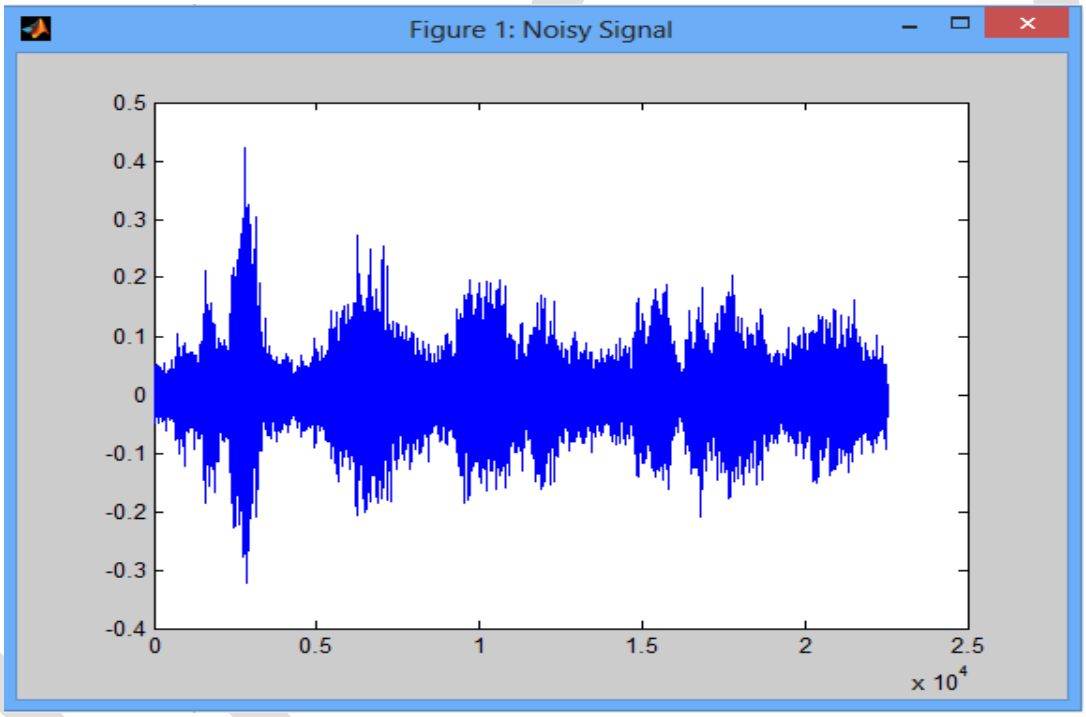

Figure 10. Noisy signal

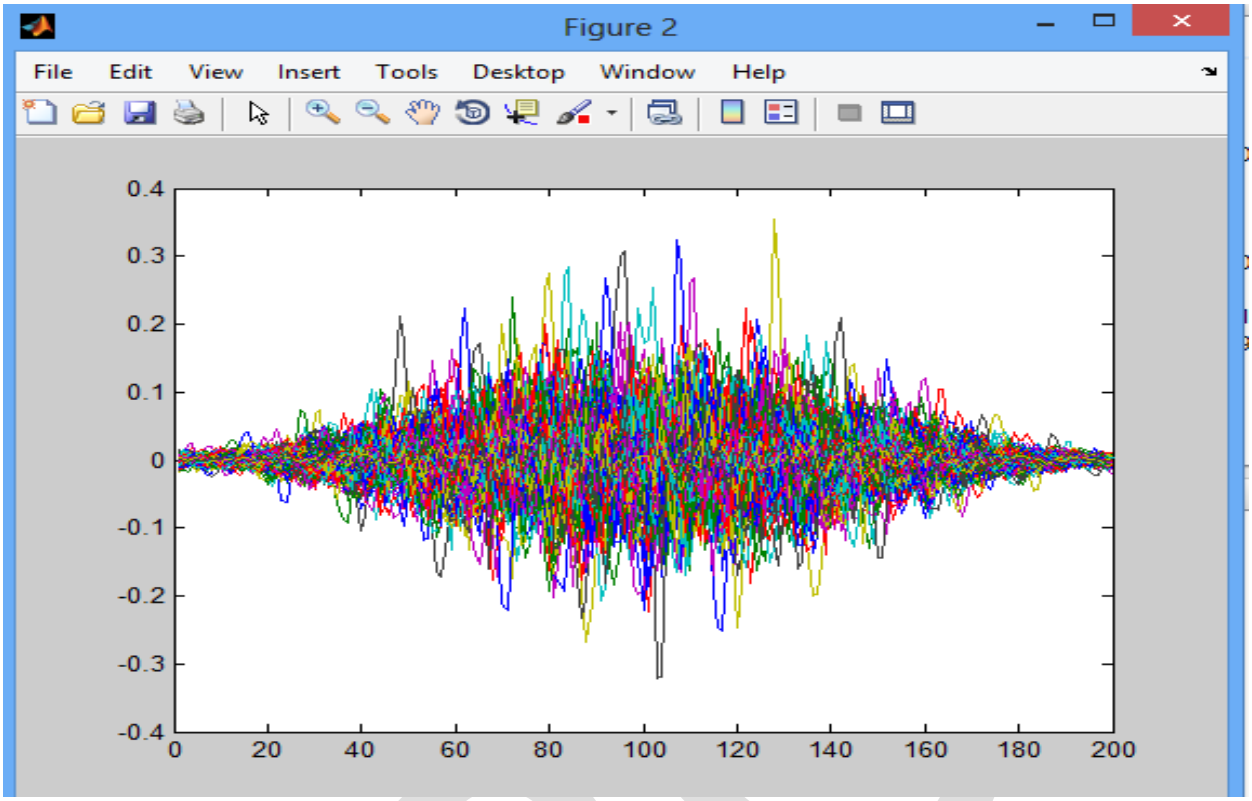

Figure 11. Segmented signal

1. Select any input signal:input signal is any sample of noisy speech signal which is stored by.wav file.
Given specifications of input signal:
Sample frequency:Fs=8000Hz,
Number of Samples:Ns=200,
Length of Input signal:Ls=22529 samples.

2. Create symmetric Hamming window in a column vector.
The Hamming window coefficients are expressed as:

$$w[n] = 0.54 - 0.46(1 - \cos(\frac{2\Pi n}{N-1}); 0 \le n \le N-1$$

N=Ns=Number of samples,
Wcoefficients={1,2,.....upto 200}
Wcoefficients values={.0800,.0802,.809,.....,.0800}.

3. Signal segmenting into Frame:
The total number of frames are calculated by,

Nframes = (Ls – Ns)/ (Ns*Sp) +1

Where, Ls = length of signal, Ns = Length of each frame

Sp = Shift Percentage=0.4

Nframes = {(22529 – 200)/ (200*0.4)} +1
Nframes =280.11

CONCLUSION

The paper present 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 is 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 provide 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.

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.
[7] D. Kolossa, A. Klimas, and R. Orglmeister, "Separation and robust recognition of noisy, convolutive speech mixtures using time-frequency masking and missing data techniques," in Proc. Workshop Applicat. Signal Process. Audio Acoust. (WASPAA), Oct. 2005, pp. 82–85.
[8] C. Breithaupt, T. Gerkmann, and R. Martin, "Cepstral smoothing of spectral filter gains for speech enhancement without musical noise," IEEE Signal Process. Lett., vol. 14, no. 12, pp. 1036–1039, Dec. 2007.
[9] R. Nickel, R. F. Astudillo, D. Kolossa, S. Zeiler, and R. Martin, "Inventory- style speech enhancement with uncertainty-of-observation techniques," in Proc. ICASSP, Kyoto, Japan, Mar. 2012, pp. 4645–4648.
[10] I. Cohen, "Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging," IEEE Trans. Speech Audio Process., vol. 11, no. 5, pp. 466–475, Sep. 2003

IJERGS