Noise Removal From Pathological Speech Signal, Its Framing And Windowing

Prof. Syed Mohammad Ali
Department of Electronics & Telecommunication Engg.
Anjuman College of Engineering & Technology
Nagpur, India
e-mail: ali_acet2003@yahoo.co.in

Dr. Pradeep Tulshiram Karule
Department of Electronics Engg.
Yeshwantrao chavan College of Engineering
Wanadongri, Nagpur, India
e-mail: ptkarule@rediffmail.com

***Abstract*—Analysis of speech has become a popular non-invasive tool for assessing the speech abnormalities .Acoustic nature of the abnormal speech give relevant information about the type of disorder in the speech production system. In the recent year, the trend towards automated analysis of pathological noise signal has gain momentum. The awkwardness of analog equipment has simulated development of digital computer techniques for processing and analysis of pathological speech signal in patient care system. The given filter design techniques & prepossessing of speech signal can be used in any speech processing application.**

In this paper, speech signals of patients are taken and they are preprocessed .In prepossessing Speech signal is passed through Moving average, High pass filter for removal of noise. After removing the noise, the signals are framed & passed through window. The output of window has a preprocessed signal & can be used for any speech application like voice disorder identification, speaker recognition, etc.

***Keywords*—Pathological speech signal, moving average filter, high pass filter, framing, windowing.**

I. INTRODUCTION

Voice disorder recognition has received a greater attention from researchers in the last decade .Speech processing has proved to be excellent tool for voice disorder detection [3]. Pathological voice signal of patient from Government Medical College & Hospital, Nagpur has been taken. The signal is recorded keeping mic two inch away from mouth using voice recorder of window XP .The sampling frequency is chosen to 11025 samples/sec 8-bit stereo 21 kbps. The patient has pronounced ,'a' which is vowel ,then 'ah' which is consonant and a word 'Hello' for two sec, this signal is noisy & noise needs to be removed. So filters are designed. The signal is passed through filter & then framing & windowing is done. The output of window is called preprocessed output, which can be used for further application like diagnosis of disease or speaker recognition. Physicians often use invasive techniques like endoscopy to diagnose symptoms of vocal fold disorders however, it is possible to diagnose disease using certain feature of speech signal of [3]. Speech signal is sinusoidal signal having different frequency, different amplitude, & different phase. It is given by the expression given below [6].

$$\sum_{i=1}^{N} A_i(t)\sin[2\pi F_i(t)t+\theta_i(t)]$$

Where, $A_i(t)$, $F_i(t)$ & $\theta_i(t)$ are the sets of amplitudes, frequencies & phases respectively, of the sinusoids as shown in fig. 1.

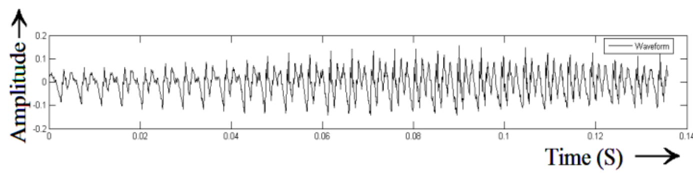

Figure 1. A speech signal

Voice & speech production requires close cooperation of numerous organs which from the phonetic point of view may be divided into organ.

- Lungs, Bronchi, Tracheas (producing expiration air steam necessary for phonation)
- Larynx (amplifying the initial tone)
- Root of the tongue, throat, nasal cavity, oral cavity (forming tone quality & speech sound) [7].

Speech signal in non-intrusive in nature& it has potential for providing. Quantitative data with reasonable analysis time. So study of speech signal of pathological voice has become a important topic for research as it reduces work load in diagnoses of pathological voices [8].

II. FLOW DIAGRAM

The following figure shows flow diagram of preprocessing of speech signal.

SEEK DIGITAL LIBRARY

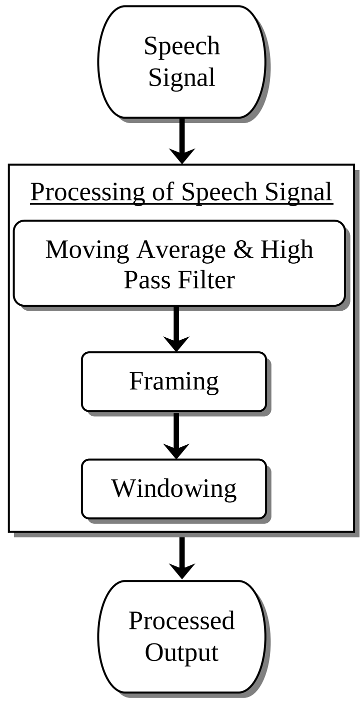

Figure 2. Flow diagram of preprocessing of speech signal.

III. FILTER DESIGN

The noisy speech signal is passed through filters like moving average filter, prepossessing filter. The moving average filter takes average of samples for filtering the noise from signal. The expression for output of such filter is given below [6].

$$Y(n)=\frac{X(n)+X(n-1)+X(n-2)}{3}$$

Where, X(n) is the input speech sample.

Analysis with different averages of samples are taken and it is found that with average of 3samples proper filtering is done. The system function H(z) & the impulse response h(n) of filter are given below:

$$Y(z)=\frac{1}{3}[X(z)+z^{-1}X(z)+Z^{-2}X(z)]$$

$$H(z)=\frac{Y(z)}{X(z)}=\frac{1}{3}[1+z^{-1}+Z^{-2}]$$

$$\therefore\ h(n)=[\frac{1}{3},\frac{1}{3},\frac{1}{3}]$$

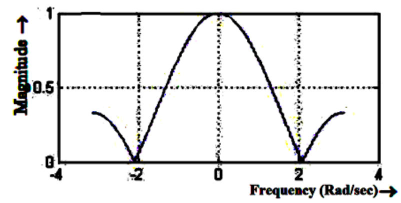

Figure 3. Magnitude v/s Frequency plot of moving average filter

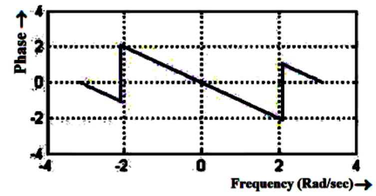

Figure 4. Phase v/s Frequency plot of moving average filter

This FIR filters shown by magnitude & phase plots in fig. 3, fig. 4, fig. 5 & fig. 6 are more stable.

IV. PRE-EMPHASIS FILTER

The Pre-emphasis filter is used to offset the negative spectral slope of voiced speech signal to improve the efficiency of spectral analysis [2]. Also Pre-emphasis is used to flatten speech signal spectrum & to make the speech signal less sensitive to finite precision effects later in speech signal processing [4].

Here we have used single coefficient FIR filter. The system function H(z) of filter is

$$H(z)=1-\lambda z^{-1} \tag{1}$$

$$\frac{Y(z)}{X(z)}=1-\lambda z^{-1}$$

$$Y(z)=X(z)-\lambda z^{-1}X(z)$$

The time domain representation of filter will be

$$Y(n)=X(n)-\lambda X(n-1)$$

Where, λ is the filter coefficient and the value of pre-emphasis coefficient is λ ε (0.9 - 1.0) with λ = 0.9375, best optimum result of filtering is received [5]. The pre-emphasis filter serves to offset this natural slope before spectral analysis, thereby improving the efficiency of the analysis secondly the hearing is more sensitive above 1 KHz region of spectrum. The pre-emphasis filter amplifies the area of spectrum. Thus improving the efficiency of spectral analysis [2].

SEEK DIGITAL LIBRARY

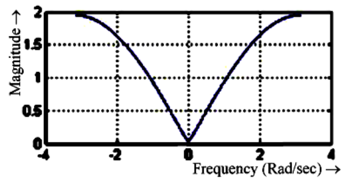

Figure 5. Magnitude v/s Frequency plot of Pre-emphasis Filter

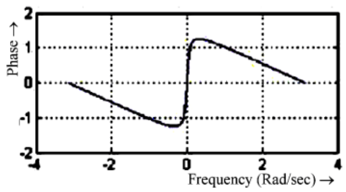

Figure 6. Phase v/s Frequency plot of Pre-emphasis Filter

V. FRAMING

The speech signal is slowly varying over time (quassi stationary) that is when the signal is examined over a short period of time (5 msec to 100 msec). The signal is fairly stationary.

Therefore, speech signals are often analyzed in short time segment, which are referred as short-term spectral analysis. This practically means the signal is blocked in frames of typically 20-30 msec. Adjacent frames typically over lap each other with 30% to 50 %. This is done in order not to lose any information due to the windowing.

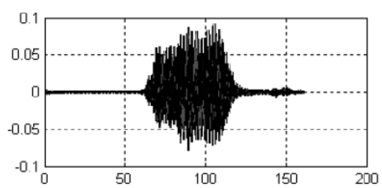

Figure 7. A Framed Signal

VI. WINDOWING

After being partitioned into frames, each frame is multiplied by a window function prior to the spectral analysis to reduce the effect of discontinuity introduced by the framing process by attenuating the values of the samples at the beginning and end of each frame. Commonly used windows include Hamming & Hanning windows. If no window is used, the case can be treated as rectangular window. Each window has its own pros & cons. Compared to rectangular window, the Hamming & Hanning windows decrease the frequency resolution of the spectral analysis while reducing the side lobe level of the window transfer function [9].

In this paper, after the signal has been framed, each frame is multiplied by window function W(n) with length N. The frame has N= 256 samples & adjacent frame separated by M=128 samples [4]. Where N is the length of the frame. Typically hamming window W(n) is used & given by

$$W(n) = 0.5 - 0.46\cos(2\pi n / N - 1) \qquad 0 \le n \le N-1$$

Where, N is total number of samples. The windowing is done to avoid problems due to truncation of signal as window helps in smoothing of signal [1].

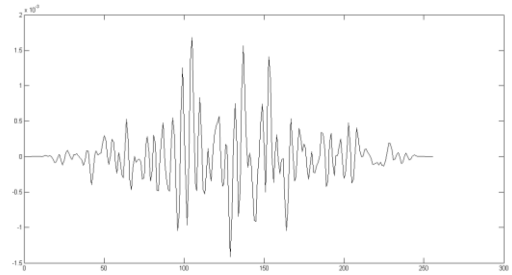

Figure 8. A Windowed Signal

VII. RESULT & CONCLUSION

The table on the last page (fig. 9) shows that, signals are perfectly filtered using both ma and pre-emphasis filter. The output of ma filter needs to be scaled to get amplified output. From the table it is clear that this system can be used for any speech preprocessing applications. Filtered signal is framed and passed through window.

REFERENCES

[1] Manjot Kaur Gill, Vector, “Quantization based Speaker identification”, International Journal of Computer Application (0975-8887) Volume 4-No.2, pp1-4 July 2010.

[2] J. W. Picone, “Signal modeling techniques in speech recognition”, Proceedings of the IEEE, Vol.81, No.9, Sept.1993.

[3] Lotfi Salhi, Talbi Mourad, and Adnene Cherif, “Voice Disorders Identification Using Multilayer Neural Network”, The International Arab Journal of Information Technology Volume 7-No.2, pp.177-185 April 2010.

[4] Antanas Lipeika, Joana Lipeikiene, Laimutis Telksnys, “Development of Isolated speech Recognition System”, INFORMATICA, 2002, Vol 13, No.1, 37-46, 2002.

[5] Milan Sigmund, “Voice Recognition By Computer”, Tectum Verlag publication, pp no20-22.

[6] John G. Proakis, Dimitris, G. Manolakis, “Digital Signal Processing, principles, Algorithm and Applications”, Prentice Hall India, pp.309, Third Eition.

[7] Orzechowski, A. Izworski, R. Tadeusiewiez, K. Chmunzynska, P. Radkowski, I. Gotkowska, “Processing of pathological change in speech caused by Dysarthria”, IEEE Proceedings of 2005 International symposium on intelligent signal processing & communication system, pp 49-52, Dec. 13-16, 2005.

[8] Martinez Cesar E, Refiner Hugo L, “Acoustic Analysis of speech for detection of Laryngeal pathologies”, IEEE proceeding of the 22nd Annual EMBS international conference chicago IL, pp.2369-2372, 2000.

[9] S. K. Mitra, “Digital signal processing: a computer-based approach”, 2nd Edition, McGraw Hill, 2002.

SEEK DIGITAL LIBRARY

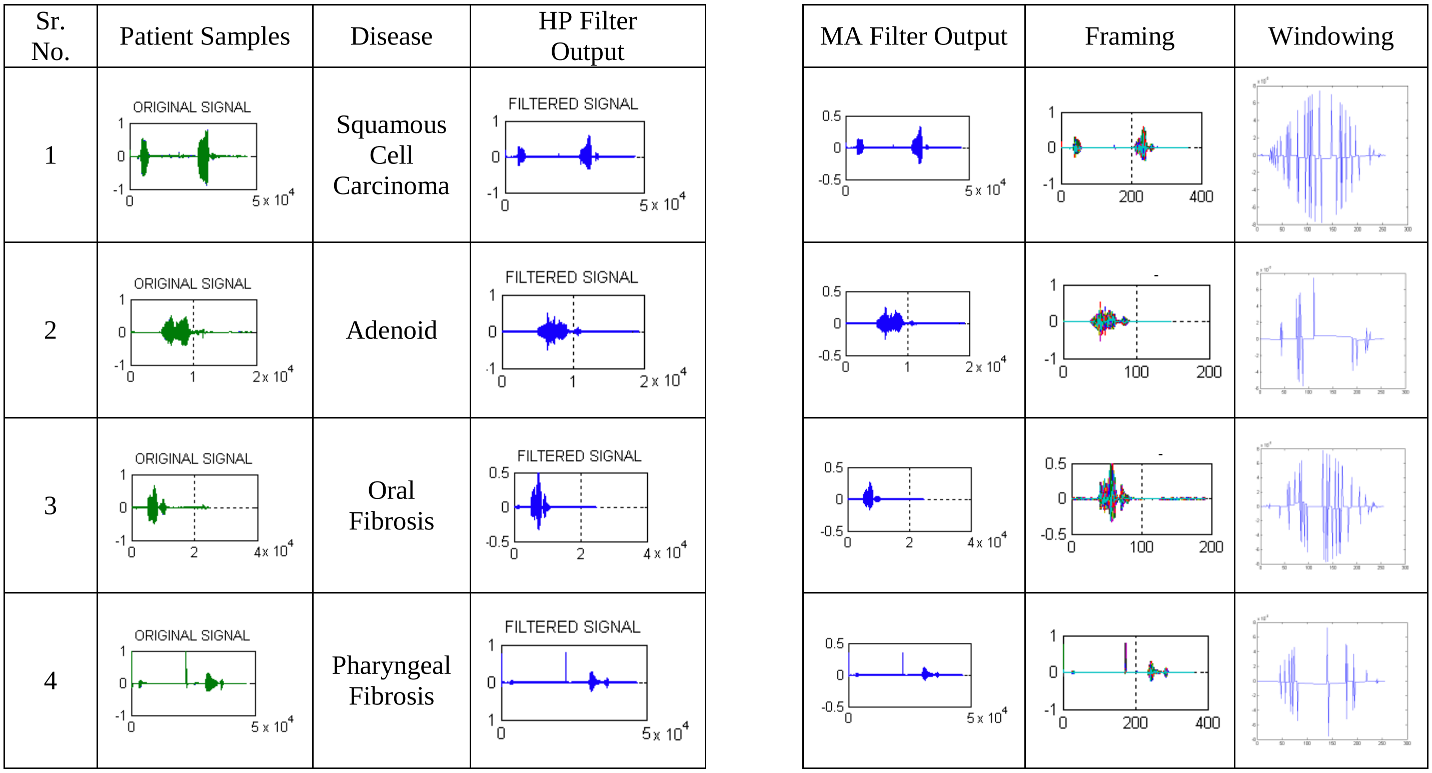

Figure 9. Examples showing complete preprocessing analysis of patient's speech samples.

SEEK DIGITAL LIBRARY