Speech Denoising Using Different Types of Filters

Kavita Sharma(MTech*) Prateek Haksar (MTech*)

(Department of electronics and communication, MITS, Laxmangarh (UGC-1956))
(Department of electronics and communication, MEC, Bikaner(RTU))

ABSTRACT

Speech Recognition is a broader solution which refers to a technology that can recognize a speech without being targeted at single speaker such call system can recognize arbitrary voice. The fundamental purpose of speech is communication, i.e., the transmission of messages. The problem in speech recognition is the speech pattern variability. The most challenging sources of variations in speech are speaker characteristics including accent, co-articulation and background noise. The filter bank in the front-end of a speech recognition system mimics the function of the basilar membrane. It is believed that closer the band subdivision to human perception better is the recognition results. Filter constructed from estimation of clean speech and noise for speech enhancement in speech recognition systems. In my work, I am using a filter on the frequency spectrum of an input voice, firstly that will record the voice and after that the same will played on simulation and it will give the frequency spectrum. Finally, this spectrum will be filtered to remove noise.

Keywords - **FIR, IIR, WAVELETS, FILTER BANKS.**

I. INTRODUCTION

Speech recognition systems have a wide range of applications from the relatively simple isolated-word recognition systems for name-dialling, automated customer service and voice-control of cars and machines to continuous speech recognition as in auto-dictation or broadcast-news transcription. The speech features are derived from a bank of filers inspired by knowledge of how the co clear of the inner ear performs spectral analysis of audio signals.

The fundamental problem in speech recognition is the speech pattern variability. Speech recognition errors are caused by the overlap of the distributions of the acoustic realizations of different speech units. The sources of speech variability are like Duration variability, Spectral variability, Speaker variability, Accent, Contextual variability, Co-articulation, Noise. Speech recognition methods aim to model, and where possible reduce, the effects of the sources of speech variability. The most challenging sources of variations in speech are speaker characteristics including accent, co-articulation and background noise. The filter bank in the front-end of a speech recognition system mimics the function of the basilar membrane. It is believed that the closer the band subdivision is to human perception, the better the recognition results. Filter constructed from estimation of clean speech and noise for speech enhancement in speech recognition systems. In the current design project a basic speaker identification algorithm has been written to sort through a list of files and choose the 12 most likely matches based on the average pitch of the speech utterance as well as the location of the formants in the frequency domain representation. In addition, experience has been gained in basic filtering of high frequency noise signals with the use of FIR, IIR and WAVELETS, FILTER BANKS.

II. FIR FILTER USING KAISER WINDOW

FIR filters are filters having a transfer function of a polynomial in *z*-plane and is an all-zero filter in the sense that the zeroes in the *z*-plane determine the frequency response magnitude characteristic. The *z* transform of N-point FIR filter is given by

$$H(z) = \sum_{n=0}^{N-1} h(n) z^{-n}$$

FIR filters are particularly useful for applications where exact linear phase response is required. The FIR filter is generally implemented in a non-recursive way which guarantees a stable filter.

III. THE WINDOW METHOD

In this method, from the desired frequency response specification $H_d(\omega)$, corresponding unit sample response h_d (n) is determined using the following relation.

$$h_d(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} H_d(\omega) e^{j\omega n} dw$$

$$H_d(\omega) = \sum_{n=-\infty}^{\infty} h_d(n) e^{-j\omega n}$$

In general, unit sample response $h_d(n)$ obtained from the above relation is infinite in duration, so it must be truncated at some point say n= M-1 to yield an FIR filter of length M (i.e. 0 to M-1). This truncation of $h_d(n)$ to length M-1 is same as multiplying $h_d(n)$ by the rectangular window defined

$$\begin{aligned} w(n) &= 1 \qquad 0 \le n \le M-1 \\ &= 0 \qquad \text{otherwise} \end{aligned}$$

Thus the unit sample response of the FIR filter becomes

$$h(n) = h_d(n) w(n)$$

$$= h_d(n) \qquad 0 \le n \le M-1$$

$$= 0 \qquad \text{otherwise}$$

Now, the multiplication of the window function w(n) with $h_d(n)$ is equivalent to convolution of $H_d(\omega)$ with $W(\omega)$, where $W(\omega)$ is the frequency domain representation of the window function

$$W(\omega) = \sum_{n=0}^{M-1} w(n)e^{-j\omega n}$$

Thus the convolution of $H_d(\omega)$ with $W(\omega)$ yields the frequency responses of the truncated FIR filter.The frequency response can also be obtained using fourier transform.

$$H(\omega) = \frac{1}{2\pi}\int_{-\pi}^{\pi} H_d\,(v)W(\omega\, - v)dw$$

The most challenging sources of variations in speech are speaker characteristics including accent, co-articulation and background noise. It is believed that the closer the band subdivision is to human perception, the better the recognition results. Filter constructed from estimation of clean speech and noise for speech enhancement in speech recognition systems. The work that I have done involves filtering of recorded voice that includes noise .The future scope of speech recognition is implementation of different filters on recorded voice to remove noise .The best filter among all which are implemented will give the best result with respect to filtering the recorded voice

IV.CONCLUSION

.

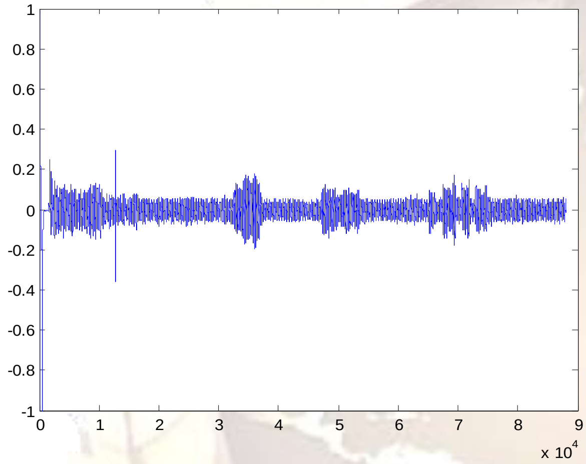

Figure 1 Spectrum of Recorded voice

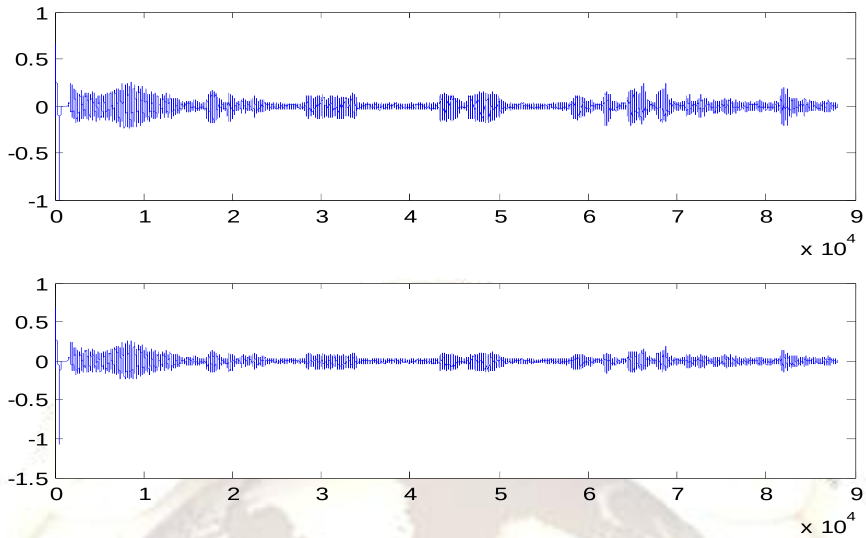

Figure 2 Filtered Spectrum

REFERENCES

1. Claudio Garreton and Nestor Becerra Yoma," Telephone Channel Compensation in Speaker Verification Using a Polynomial Approximation in the Log-Filter-Bank Energy Domain" IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 20, NO. 1, JANUARY 2009.
2. C. P. Chen and J. Bilmes, "MVA processing of speech features," IEEE Trans. Audio, Speech, Lang. Process., vol. 15, no. 1, pp. 257–270, Jan.2007.
3. P. Kenny, G. Boulianne, P. Ouellet, and P. Dumouchel, "Speaker and session variability in GMM-based speaker verification," IEEE Trans. Audio, Speech, Lang. Process., vol. 15, no. 4, pp. 1448–1460, May 2007.
4. X. Cui and A. Alwan, "Noise robust speech recognition using feature compensation based on polynomial regression of utterance SNR," IEEE Trans. Speech Audio Process., vol. 13, no. 6, pp. 1161–1172, Nov. 2005.
5. N. Roman and D. Wang, "Binaural sound segregation for multisource reverberant environments," in Proceedings ICASSP 2004, Volume 2, 2004, pp. 373 – 376.
6. Jurafsky and James H. Martin, Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, Prentice-Hall, 2000.
7. J.G. Proakis and D.G. Manolakis, Digital Signal Processing-Principles, Algorithms and Applications New Delhi: Prentice-Hall, 2000.