International Journal of Advance Engineering and Research Development (IJAERD)
Volume 1,Issue 6,June 2014, e-ISSN: 2348 - 4470 , print-ISSN:2348-6406

Near-End Perception Enhancement using Dominant Frequency Extraction

Premananda B.S.[1],Manoj[2], Uma B.V.[3]

[1]*Department of Telecommunication, R. V. College of Engineering, premanandabs@gmail.com*
[2]*Department of Telecommunication, R. V. College of Engineering, saranumh@gmail.com*
[3]*Department of Electronics & Communication, R. V. College of Engineering, umabv@rvce.edu.in*

Abstract—In general voice communication, speech is always affected by background noise which reduces perception quality of the speech signal. The incoming speech from far end speaker and near end noise reaches our ear at the same time, noise cannot be cancelled. Hence speech signal is enhanced with respect to variation in the near end noise. In this work, we propose two speech enhancing methods in order to combat near end noise. In first method, suitable gain is derived and smoothened for speech enhancement using weighted hearing curve and overall speech is enhanced to give better quality and perception. In second method we determine dominant frequencies components of speech signal and enhance only those frequencies instead enhancing the entire signal. The energy needed for enhancing is reduced in later approach when compared to former one. So a method of energy conservation is achieved. Quality of enhanced speech signals is measured using perception evaluation of speech quality (PESQ) which is an ITU recommended standard for speech quality assessment.

Keywords-Dominant frequency, gain, near end, weighted hearing curve, speech enhancement

I. INTRODUCTION

Now-a-days, everyone living on planet is linked to each other in one or another way. It is of essential significance to be able to always communicate with others wherever we are. This means that, in most of situations, the place where we are doesn't satisfy the most appropriate requirements to have a conversation. It is not useful to keep the communication if the noise of the subway, train, car, etc., prevents us from understanding. The use of communicating devices is on an increase all over the world and naturally it is becoming more common to see people using their devices in high noise environments. These environments may include driving in cars, socializing in loud gatherings or working in factories. People are facing various difficulties to deal with these noises. For example, cellular phone users will often press the phone to their head and turn up the volume to the maximum which many times is still not enough to understand the speech.

In communication, the speech signal is always accompanied by some noise. For example, mobile communication, in most cases the background noise of the environment where the source of speech lies is the main component of noise that adds to the speech signal. Though the obvious effect of this noise addition is to make the listening task difficult for a direct listener, there are many more far reaching negative effects when we process the degraded speech for some other application [2]. A related problem is processing degraded speech in preparation for coding by a bandwidth compression system. Hence speech enhancement not only involves processing speech signals for human listening but also for further processing prior to listening.

Main objective of speech enhancement is to improve the perceptual aspects of speech such as overall quality, intelligibility, or degree of listener fatigue. The goal of speech enhancement varies according to specific applications, such as to boost the overall speech quality, to increase intelligibility, and to improve the performance of the voice communication device [3]. The

background noise may be noise like such as car noise, aircraft cockpit, other moving vehicle or environmental sounds; or it may be speech-like, comprised of competing speakers.

The far-end noise scenario is tackled by traditional noise suppressing algorithms like minimum mean-square error (MMSE) short-time spectral amplitude (STSA) estimator [1], adaptive filtering algorithms[2, 3], spectral subtraction methods [4, 5], etc. For near-end noise, signal cannot be influenced because the listener is located in the noisy environment and the noise reaches the ears with hardly any possibility to intercept [9]. So, over the past decades for the near-end listening enhancement many approaches such as SNR recovery [6], recursive closed form solution [9], and perceptual distortion measurement [10] are used for the intelligibility and quality improvement. All the approaches are carried out by enhancing overall far end speech signal.

In this work in order to increase quality of speech signal in listener end, we present first method of near-end perception enhancement (NEPE) to amplify the far end speech signal by applying weighted hearing curve which gives perceivable components in speech and multiply derived smoothened gains. In second method we extract dominant frequencies (DF) [12] in speech which are maximum energy carriers and enhance only DF components.The traditional method of determining voice quality is to conduct subjective tests with panels of human listeners. The results of these tests are averaged to give mean opinion scores (MOS) but such tests are expensive and are impractical for testing in the field. For this reason the ITU recently standardized a new model, PESQ (ITU-T recommendation P.862[15] that predict the quality scores that would be given in a typical subjective test. PESQ is the new ITU-T standard for measuring the voice quality of communications networks. Thus, PESQ is used to measure the quality of enhanced signal [14].

II. DERIVATION OF WEIGHTED THRESHOLD OF HEARING CURVE

The Absolute threshold of hearing (ATH) is minimum sound level of a pure tone that an average listener with normal hearing can hear with no other sound present, this is also known as the auditory threshold or threshold in quiet. The threshold of hearing is given an empirical equation given in equation (1)[11]:

$$T_q(f) = 3.64\left(\frac{f}{1000}\right)^{-0.8} - 6.5e^{-0.6\left(\frac{f}{1000}-3.3\right)^2} + 10^{-3}\left(\frac{f}{1000}\right)^4 \quad (1)$$

where, $T_q(f)$ is Threshold of Hearing (dBSPL), f = frequency (Hz).

ATH is frequency dependent (Figure 1) and it has been shown that the ear's sensitivity is best at frequencies between 1 kHz to 4 kHz [11].Simulated results of ATH curve is as shown in Figure 1. It can be easily observed that for perception of sound it needs lesser sound pressure level perception for frequencies between 2 kHz to 4 kHz.The weighed curve is the inverse of threshold of hearing curve shown in Figure 2, normalized between 0 and 1.In the measurement of loudness, weighted curve is used to emphasize frequencies around 2 to 4 kHz where the human ear is most sensitive, while attenuating very low and high frequencies to which the ear is insensitive [11]. After that perception tend to decrease gradually as it moves towards higher frequencies. Deriving a weighted curve is considered as one of the most important factors which facilitates in measuring noise loudness.

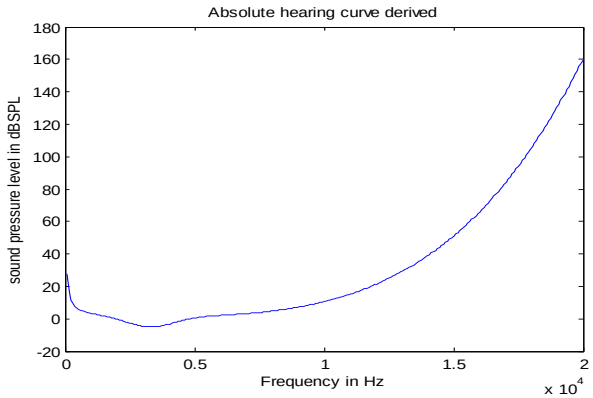

Figure 1: Derived Threshold of hearing curve.

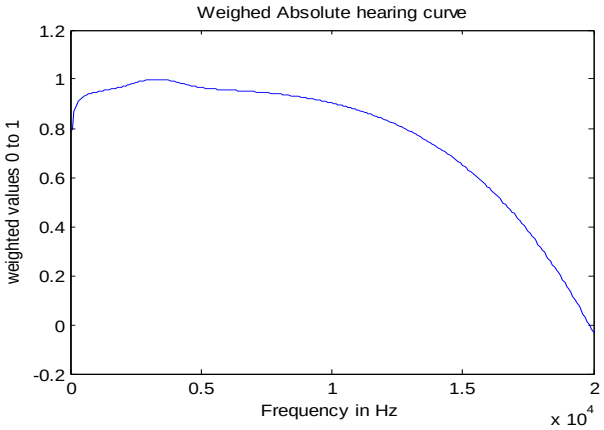

Figure 2: Derived weighted curve.

III. NEAR END PERCEPTION ENHANCEMENT USING WEIGHTED CURVE

After deriving weighted curve, real time speech and noise signals of finite length are captured and divide them into framesize of 256 samples each. The magnitude of every frequency sample is obtainedby applying 256 point Fast Fourier Transformer (FFT) to every frame.The magnitude of frequency samples is multiplied by 256 point weighed curve(Figure 2) on sample basis, signal and noise loudness is calculated using the equation 2 [10]:

$$l(dB) = 10 * \log \frac{\sum_{i=1}^{N} x_i^2}{N} \tag{2}$$

where,$x_i$sample ati^{th}location, 'i'varies from 1 to N,'N' is the number of samples in a frame.

Repeating the loudness calculation for every frame, we get 124 loudness values. Same procedure is used to obtain loudness of the noise signal.

Let 'l_s' and 'l_n' denote the loudness obtained for a frame of speech and noise signal, gain, g is derived using equation 3 [11]:

$$g = 1 + max\left(0, \left(\ell_n - \ell_s\right)\right)\exists \tag{3}$$

where,'$\exists$' is Enhancing factor.

When signal loudness is sufficiently greater than noise loudness gain is set to 1, because no enhancement is required. When 'l_s' is approximately equal to 'l_n' then gain is selected such that enhanced speech signal is increased nearly by 3 dB. When 'l_s' is less than 'l_n' then gain is calculated

using equation 3. The gain obtained is dynamic and changes' rapidly from frame to frame, do avoid sudden change in signal levels, smoothening of gain is required. For this, obtained gain is averaged over six frames one pre- frame gain and four post frame gains and present frame gain, as given in equation 4.

$$g = (g_{-1} + g + g_{+1} + g_{+2} + g_{+3} + g_{+4})/6 \quad (4)$$

here, '*i*' is position of frame under consideration.

A plot showing derived original gains and smoothened gain is shown in Figure 3. Optimized gain is multiplied with speech signal in frequency domain to get enhanced speech signal. If the enhanced samples of speech signal exceeds the maximum loudness of the loud speaker, then end capping can be performed depending on minimum and maximum values of enhanced speech signal.

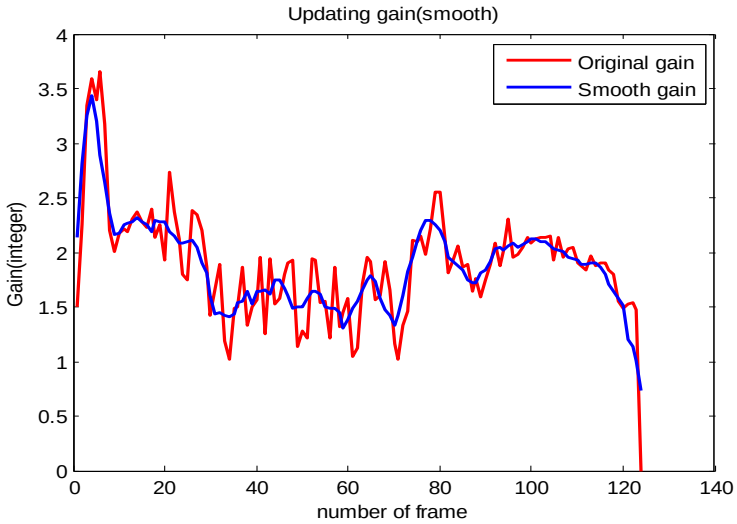

Figure 3: Smoothening of gain.

IV. DOMINANT FREQUENCY EXTRACTION AND ENHANCEMENT

There are situations when the observed speech signal show a periodic behavior due to one frequency, called the dominant frequency, which carries the maximum energy among all frequencies found in the spectrum. Block diagram of dominant frequency extraction and enhancement module is shown in Figure 4, steps involved are as follows:

1. Capture the speech signal of finite duration with sampling frequency of 8000Hz using audio editor tool, Goldwave and samples are divided into frame size (duration 32ms) of 256 samples.
2. Take fast Fourier transformer (FFT) for each frame to plot samples of frame in terms of frequency.
3. Find peaks in each frame and sort them to pick highest 'N' number of peaks, these peaks corresponds to dominant frequency components of that particular frame.
4. Each dominant frequency components magnitude will be multiplied with derived gain which is a scalar quantity.
5. Take Inverse FFT (IFFT) and reconstruct signal to get enhanced speech signal.
6. Measure the quality of enhanced signal using PESQ.

For enhancing a scalar gain which is generated for each frame is multiplied only to magnitude of dominant frequency components. IFFT is applied for this enhanced frame of signal. These steps are repeated for each of 124 frames and finally all frames are combined to reconstruct the speech signal which is enhanced.

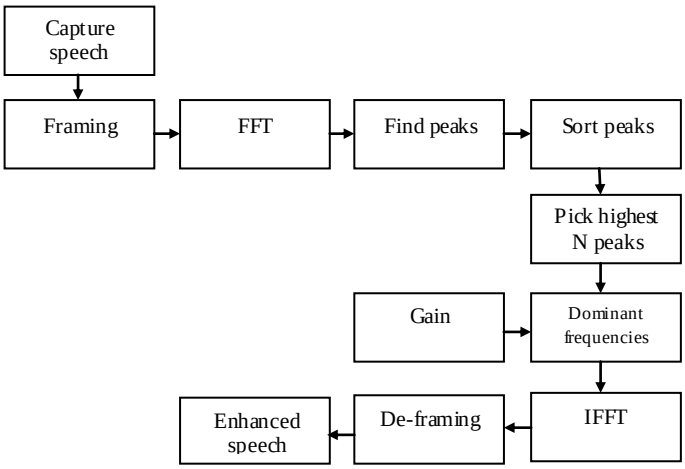

Figure 4: Block diagram of Dominant frequency extracting and enhancing module.

V.RESULTS AND DISCUSSION

Absolute threshold of hearing and weighted threshold of hearing curve are modeled as explained in section II and implemented in MATLAB. The speech and noise signals are captured using audio editor tool, GoldWave and saved in .wav format for duration 3.968 seconds (16-bit PCM) with sampling rate of 8 kHz, total samples are divided into frame size of 256 each, resulting in 124 frames.

As explained in section IV, the speech signal played at noisy environment is enhanced by multiplying derived optimized gain for dominant frequency components. The original speech and enhanced speech also plotted in MATLAB with respect to time are shown in Figure 5. The plots of original and enhanced speech signal for NEPE method and dominant method shown in Figure 6 and Figure 7 respectively. Pink line represents energy speech shaped noise with Signal to Noise Ratio (SNR) of 5 dB, blue line represents original speech energy and red line represents enhanced signal energy plot for frame 1 to 124. To obtain the SNR of 5 dB, variance of noise is adjusted in the recorded signals using equation 17 [18].

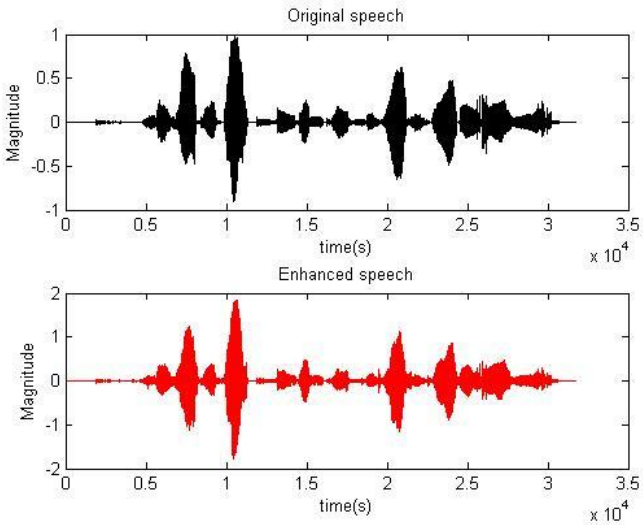

Figure 5: Original signal (black) and enhanced signal (red).

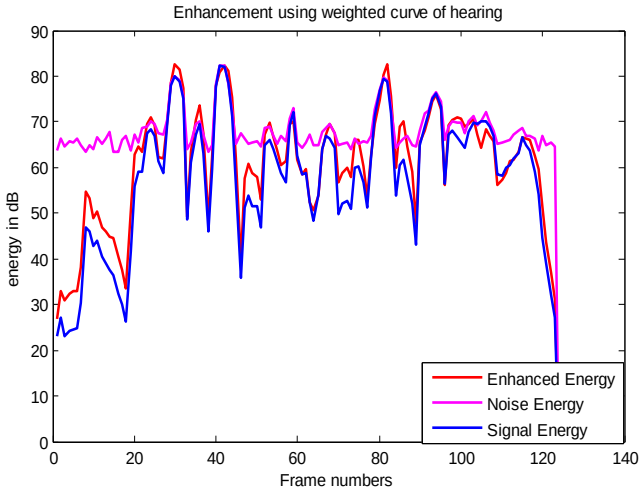

Figure 6: Energy plot of NEPE using weighted curve.

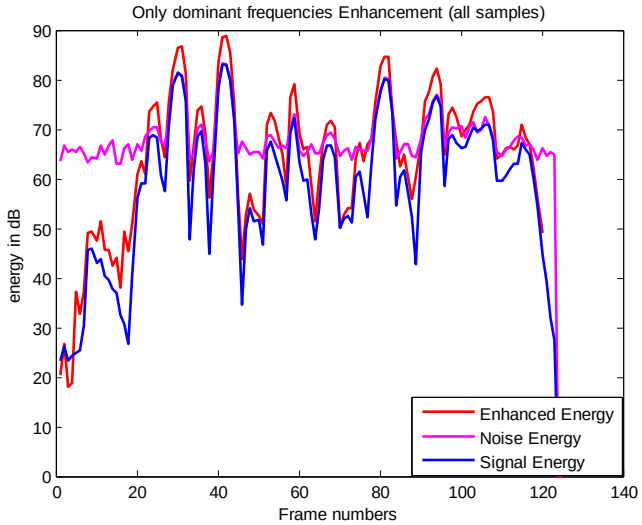

Figure 7: Energy plot of only dominant frequency enhancing method.

In former method, algorithm enhance overall signal power whenever noise signal is greater then it. In later method, whenever noise signal energy is greater than speech signal effective enhancement can be observed. For calculating quality scores original speech signal is compared with enhanced speech signal. PESQ scores are calculated for enhanced signal with respect to speech shaped noise (SSN), SNR varied -15 dB to 15 dB tabulated in table I.

TABLE I: Quality PESQ scores and MOS for SSN noise.

SSN SNR	Dominant method (Method II)		Weighted method (Method I)	
	PESQ	MOS	PESQ	MOS
-15	2.3743	1.9888	3.7259	3.8492
-10	2.4479	2.0729	3.7207	3.8427
-5	2.4433	2.0676	3.7839	3.9188
0	2.5090	2.1462	3.6734	3.7837
5	2.6264	2.2950	3.7001	3.8172
10	2.8286	2.5723	3.7576	3.8876
15	2.6866	2.3750	3.7744	3.9075

VI. CONCLUSION

This work presents novel methods of near end perception enhancement using smoothened gain. In first method derived weighted curve is mutipled with frequency samples to trunct them to thier actual perception, and then enhanced using gain updating so perception is increased. In second method only dominant frequencies are enhanced istead of enhancing over all signal their by reducing energy required for enhancement. Quality is measured using ITU standard PESQ scores. PESQ score for both methods are tabulated for speech shaped noise SNR varying from -15 to +15. It obseved that second method is a better when compared to first method. But quality of first method is better then second one.In future auditory masking properties can be considered for enhancement of perception.

REFERENCES

[1] Yariv Ephraim and David Malah, "Speech Enhancement using A Minimum Mean Square Error Short-Time Spectral Amplitude Estimator", *Proceedings of IEEE Transactions on Acoustic Speech and Signal Processing (ASSP)*, Vol. 32, DOI: 10.1109/TASSP.1984.1164453, pp. 1109–1121, Dec. 1984.

[2] Jagan Naveen V., Prabakar, Venkata Suman T. J. and Devi Pradeep P., "Noise Suppression in Speech Signals using Adaptive Algorithms", *International Journal of Signal Processing ImageProcessing and Pattern Recognition*, Vol. 3, No. 3, pp. 87-95, September 2010.

[3] Md Zia Ur Rahman, Murali Krishna K., Karthik G. V. S., John Joseph M., and Ajay Kumar M., "Non Stationary Noise Cancellation in Speech Signals using an Efficient Variable Step Size Higher Order Filter", *International Journal of Research and Reviews in Computer Science (IJRRCS)*, Vol. 2, No. 2, pp. 414-422, April 2011.

[4] Malihe Hassani and Karami Mollaei M. R., "Speech Enhancement Based on Spectral Subtraction in Wavelet Domain", *IEEE 7th International Colloquium on Signal Processing and its Applications(CSPA)*,DOI: 10.1109/CSPA.2011.5759904, pp.366-370, March 2011.

[5] Boll S. F., "Suppression of Acoustic Noise in Speech Using Spectral Subtraction", *IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol. 27, pp. 113-120, 1979.

[6] Bastian Sauert and Peter Vary, "Near-end listening Enhancement: Speech Intelligibility Improvement in Noisy Environments", *Proceedings of IEEE international Conference on Acoustics, Speech and Signal Processing*, France DOI: 10.1109/ICASSP.2006.1660065, pp. 493-496, May 2006.

[7] Bastian Sauert and Peter Vary, "Near-End Listening Enhancement Optimized with respect to Speech Intelligibility Index and Audio Power Limitations", *Proceedings of European Signal Processing Conference*, Aalborg, Denmark, ISSN 2076-1465, pp. 1919-1923, August 2010.

[8] Bastian Sauert, Heinrich Lollmann, and Peter Vary, "Near End Listening Enhancement by Means of Warped Low Delay Filter Banks", *Proceedings of ITG-symposiumvoice communication*, Vol. 8, Aachen, Germany, VDE Verlag GmbH, ISBN: 978-3-8007-3120-6, October 2008.

[9] Bastian Sauert and Peter Vary, *"*Recursive Closed Form Optimization of Spectral Audio Power Allocation for Near-end Listening Enhancement*", Proceedings of ITG-* symposiumvoice communication, Vol. 9. Berlin: VDE-Verlag, ISBN: 978-3-8007-3300-2, October 2010.

[10] Taal C. H., Hendriks R. C. and Heusdens R., "A Speech Preprocessing Strategy for Intelligibility Improvement in Noise Based on a Perceptual Distortion Measure", *IEEE International Conference on Acoustics Speech and Signal Processing*, Kyoto, pp. 4061–4064, Japan, 2012,

[11] Premananda B. S. and Uma B. V., "Low Complexity Speech Enhancement Algorithm for Improved Perception in Mobile Devices", *International Workshop on Wireless and Mobile Networks, WiMoNe-2012*, Lecture Notes in Electrical Engineering 131, Vol. 131, Springer, DOI: 10.1007/978-1-4614-6154-8 68, pp. 699-707, Feb. 2013.

[12] Rastislav Telgarsky, "Dominant Frequency Extraction," *eprint arXiv: 1306.0103*, June 2013.

[13] Rix A.W., Beerends J.G., Hollier M. P., and Hekstra A. P., "Perceptual Evaluation of Speech Quality (PESQ)-A New Method for Speech Quality Assessment of Telephone Networks andCodecs", *IEEE International Conference on Acoustics, Speech, and SignalProcessing*, Vol. 2, DOI: 10.1109/ICASSP.2001.941023, pp.749-752, May 2001.

[14] "PESQ: An Introduction", *Psytechnics Limited, White paper*, September 2001.

[15] Yi Hu, P.C. Loizou, "Evaluation of Objective Quality Measures for Speech Enhancement", IEEE Trans. on Audio, Speech, and Language Processing, Vol.16, No.1, DOI: 10.1109/TASL.2007.911054, ISSN: 1558-7916, pp.229-238. Jan. 2008.

[16] ITU-T P.862, Perceptual evaluation of speech quality (PESQ), and objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs, ITU-T Recommendation P.862, 2000.

[17]Gunawan T. S. and Ambikairajah E., "Speech Enhancement using Temporal Masking and Fractional Bark Gammatone Filters", *in 10th International Conference on Speech Science & Technology*, Sydney, pp. 420-425, 2004.

[18] ITU-T P.835, Subjective test methodology for evaluating speech communication systemsthat include noise suppression algorithm, ITU-T Recommendation P.835, 2003.

