Listening Enhancement in Near End Noisy Environment for Intelligibility Improvement

Premananda B.S.
Department of Telecommunication Engineering,
R.V. College of Engineering, Bangalore, INDIA
premanandabs@gmail.com

Ravisha B.
Department of Telecommunication Engineering,
R.V. College of Engineering, Bangalore, INDIA
ravisha415@gmail.com

***Abstract*— Speech communication plays very important role in day to day life. Noise is unwanted natural data which will always accompany the speech signal. When the clean speech signal received by listener end, located in noisy environment, he perceives not only the received (clean) speech but also the background noise and thus experiences an increased listening effort and possibly reduced speech intelligibility. With this noise effect the intelligibility of original speech signal is degraded. Hence there is a need for speech signal enhancement at listener end to improve the speech intelligibility by digital signal processing. This work focuses on the impact of various background/near-end noises on signal degradation and an approach to overcome noise effect for improved speech signal intelligibility. To enhance the clean far-end speech signal based on the background noise signal strength, we propose time domain near-end listening enhancement (NELE). Speech Intelligibility Index (SII) is used to calculate the intelligibility of the degraded and enhanced speech signal.**

Keywords- ***Background noise, Far-end, Near-end listening enhancement, Speech Intelligibility Index.***

I. Introduction

In speech communication, noise is always accompanied with speech signal. Noise is an unwanted signal that affects the smooth communication between the transmitter and receiver. When the noise is accompanied with speech signal, it makes the listening task difficult for a listener. For example, telephone conversation over cellular phones often takes place in the presence of near-end noise [7]. Background noise makes difficult to understand speech signal [24], as the noise level dominates, listening to speech or audio signals becomes more difficult. The listener end perceives a combination of the clean speech and near-end/background noise and thus experiences an increased listening effort. The speech signals generated at the source are referred to as far-end speech signals and signals received at listener end as near-end signal. The background noise can be originated at source side (the far-end) and also from the environment of the listener end (the near-end) as illustrated in Figure 1.

Considering the situation where a person located in a noisy far-end environment wishes to communicate via transmission line with a person located in a noisy near-end environment. The far-end signal can be affected by background noise from both sides of the transmission line.

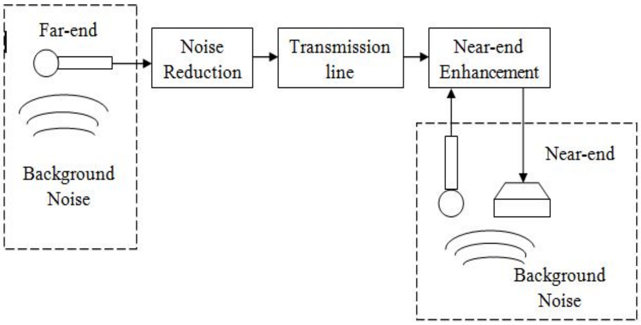

Figure 1: Far-end and Near-end listening scenario

The speech signal which is affected by the noise will results in listeners fatigue as the intelligibility of speech is apparently lowered, thus creates near-end listening problem [7]. Speech intelligibility is the degree to which humans can understand a spoken message [7-9]. Intelligibility of the speech signal is lowered because of the corrupted speech signal.

Speech signal corrupted by the noise at far-end, filtering techniques such as spectral subtraction, wiener filtering can be adopted to get enhanced speech signal at the near end listener [1-6]. For the problem of near end listener, the noise signal cannot be influenced because the person is located in the noisy environment and the noise reaches the ears. Approaches used in [1-6] are not suitable for the near end problem so the reasonable option to improve the intelligibility (by digital signal processing) is to manipulate the far end speech signal.

Bastian Sauert et al., in [7] proposed frequency dependent amplification of speech signal to reestablish the distance between the average measured speech spectrum and the average measured noise spectrum, i.e., to recover a certain signal-to-noise ratio (SNR). The performance of the proposed algorithm was evaluated in terms of SII, defined in [8]. [12] considered near end listening problem with the loudspeaker power constraint to the power of the original signal. A recursive closed-form solution in [13], which maximizes the SII under the constraint of an unchanged average power of the audio signal, was developed. The method has less complexity compared to a previous approach in [8, 10-12] and is thus suitable for real-time processing and found the instrumental evaluation by means of the average SII has shown an identical performance after processing with proposed algorithms, which is noticeably better than without processing.

A linear time-invariant filter was designed in [16] to improve the speech understanding when the speech signal is played back in a noisy environment. To accomplish this, the SII is maximized under the constraint that the speech energy is held constant.

The work provides solution to the near-end listening problem, by enhancing the far-end speech signal which also increases the intelligibility of speech signal. A simple time domain approach with includes temporal masking is implemented for enhancing the speech signals. SII is used to measure the intelligibility of the speech signal. MATLAB and audio editor tool, GoldWave is used for verifying the results.

II. NELE USING TIME DOMAIN ANALYSIS

We propose a simple time domain NELE approach to overcome the degradation of the speech signal in the noisy environment. In order to increase the intelligibility of the speech signal different algorithms have been proposed and implemented in [7-25], but the complexity involved in deriving optimum gain is not resolved. Block diagram of the proposed time domain approach is shown in the Figure 2.

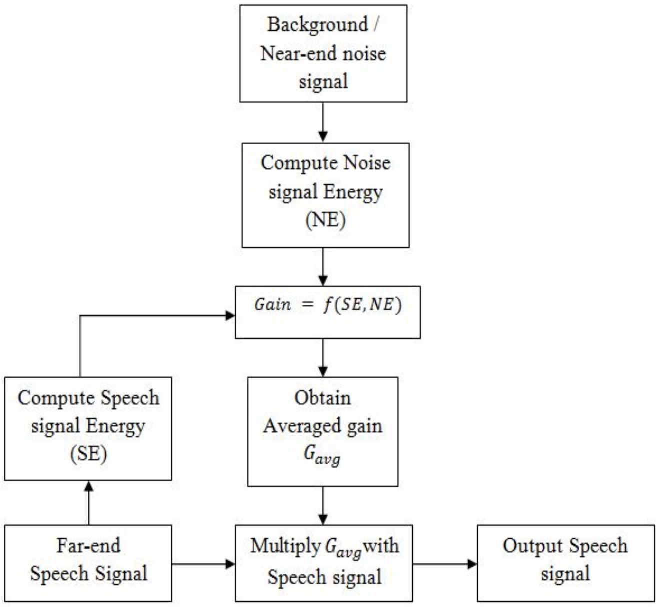
Background / Near-end noise signal
Compute Noise signal Energy (NE)
$Gain = f(SE, NE)$
Compute Speech signal Energy (SE)
Obtain Averaged gain G_{avg}
Far-end Speech Signal
Multiply G_{avg} with Speech signal
Output Speech signal

Figure 2: Block diagram of proposed time domain approach.

Energy of the incoming speech signal and near-end noise signal is computed using equations (1).

$$E\,(dB\,) = 10\;\log\sum\frac{x}{N}^{2} \qquad (1)$$

where, x is the samples within a frame and N total number of samples in a frames.

Steps involved in time domain near-end listening enhancement are:

Step 1: Incoming speech and noise signal are captured using an audio editor tool, GoldWave with sampling rate of 8 kHz.

Step 2: For NELE it is mandatory to fix the SNR of speech and noise signals. Since speech and noise signal are dynamic in nature the variance of the noise signal is varied to get required SNR using equation 2. SNR is varied from -18 to +21 dB for experimenting.

$$n = \frac{n}{norm(n)} * \frac{norm(x)}{10^{0.05*SNR}} \qquad (2)$$

where, n and x are captured noise and speech signal.

Step 3: Computing energy of speech and noise signal

Individual frame energy is calculated using equation 3.

$$E(dB) = 10 * \log\frac{\sum_{i=1}^{N} {x_i}^2}{N} \qquad (3)$$

where x_i is the sample at i^{th} location and *i* varies from 1 to N, N are is the total number of samples in a frame. Same procedure is used to obtain energy of the noise signal. Let SE denote energy of speech signal obtained for each frame and NE is the energy obtained for each frame of noise signal.

Step 4: Deriving Gain: The correct gain for a pair of speech and noise frames (selected parameters) is user specific. The gain is derived using the equation 4,

$$Gain = initialvalue + (A \times B) \qquad (4)$$

where,

A = max(0, NE-SE)

Initial values is set to 1 so that Gain =1, when no enhancement for the speech signal is required, B is used to control gain (B is always <1).

Step 5: To derive averaging gain (G_{avg}): To avoid signal saturation and signal bursts due to sudden gain changes, optimal gain computed must be characterized by slow and configurable response time for the gain variations. Obtained gain is averaged over five frames gain values obtained from step 4.

Step 6: Multiply averaged gain with speech signal:

Once averaged gain is derived next is to multiply (**G_{avg}**) with every speech sample of respective frame, resulting in enhanced speech signal.

Step 7: End capping: To avoid overflow of enhanced values, positive values must not be above +1 and negative values must not be below -1.

The intelligibility measure which will be used for the project is based on the standardized SII [8]. Speech and noise are presented are assumed to be above the threshold in quiet. Masking effects are excluded from ANSI SII procedure [8]. Let us consider s is input signal and y is noise signal, then corrupted signal received is given by equation 5,

$$z(n) = s(n) + y(n) \qquad (5)$$

S_n windowed versions of *S*, where n is the window frame-index. A Hann-window is used with 50% overlap, 32 ms length.

The impulse response of the i^{th} auditory filter is denoted by h_i , where $i \in \{1,......,m\}$ where m is the total number of auditory filters.

Energy within one time frequency (TF) unit is calculated by equation 6,

$$S^2{}_{n,i} = \sum_k |S_n(k)|^2 * |H_i(k)|^2 \qquad (6)$$

The averaged energy within one critical band is based on the long term sample mean over many short time frames and is denoted by equation 7,

$$E^2{}_{S_i} = \frac{1}{N} \sum_n S^2{}_{n,i} \qquad (7)$$

where, N is the total number of short time frames.

Say $E^2{}_{Y_i}$ the average noise energy within critical band, let E_i be SNR within one critical band is denoted by equation 8,

$$E_i = \frac{E^2{}_{S_i}}{E^2{}_{Y_i}} \qquad (8)$$

E_i is used to calculate an intermediate measure to determine the audibility of the speech in presence of noise within one band. In Speech intelligibility Index (SII) (ANSI, 1997), for instance, the SNR calculation is limited to the range of [-15, 15] dB, prior to the mapping of the computed SNR to the range of [0, 1]. Thus, the SNR is log transformed and clipped between -15 dB, SNR_{MIN} to +15 dB, SNR_{MAX} and normalized such that its range is between 0 and 1 and SII is calculated from equation 9,

$$SII = \sum_i \frac{\max(\min(10 * \log(E_i), SNR_{MAX}), SNR_{MIN})}{30} + \frac{1}{2} \quad (9)$$

III. RESULTS

The speech and noise (babble) signals are captured using an audio editor tool, GoldWave and saved in uncompressed.wav format for duration of 3.968 seconds (16-bit PCM) with sampling rate of 8 kHz. The captured speech signal has 31744 (8000*3.968) samples, total samples are divided into frame size of 256 each, resulting in 124 frames. Energy plots for speech and noise signals are shown in Figure 3.

Comparing the energy of speech and noise signals frame wise, gain is computed using equation 4 and optimal gain is computed by averaging gain with pre and post frame gains and is plotted in Figure 4. Averaged gain obtained, plotted in blue line is smooth compared to red line, hence enhanced speech signal vary smoothly and does not fatigue listeners ears. Enhanced speech signal is obtained by averaged gain with every speech sample for respective frames. The enhanced speech truncated to maximum values to saturation of the signals. Energy comparison plot for speech, noise and enhanced signals are plotted in Figure 5. It is observed that the speech signals energy is enhanced wherever noise dominates

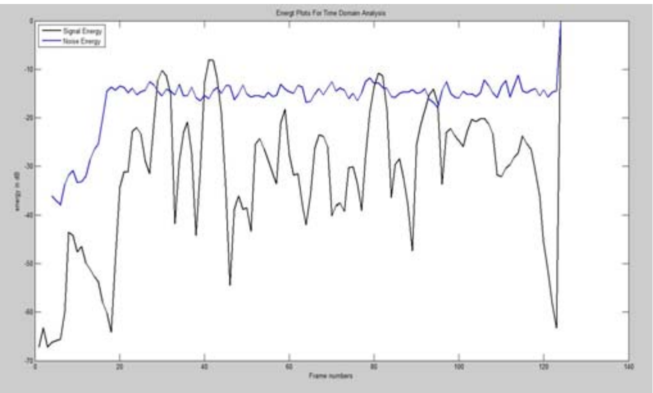

Figure 3: Energy plot for Time domain analysis.

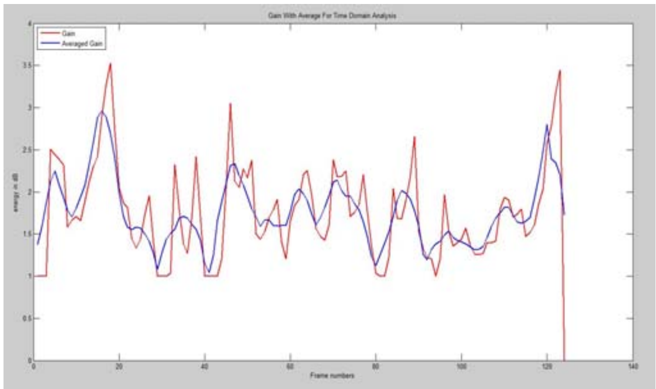

Figure 4: Gain with average plot for time domain analysis.

Spectrogram is a photographic or other visual or electronic representation of a spectrum. Spectrogram plots for time domain analysis is shown in Figure 6. Spectrogram for input speech signal is shown in Figure 6(a), and spectrogram for enhanced speech signal is shown in Figure 6(b) and 6(c) for babble noise and speech shaped noise respectively, darker portion shown in Figure 6(b) and 6(c) is enhanced signal. Intelligibility of the speech signal before and after enhancement is calculated using equation 9, and plotted in Figure 7. SII is calculated with respect to the varying SNR from -18 to +21 dB.

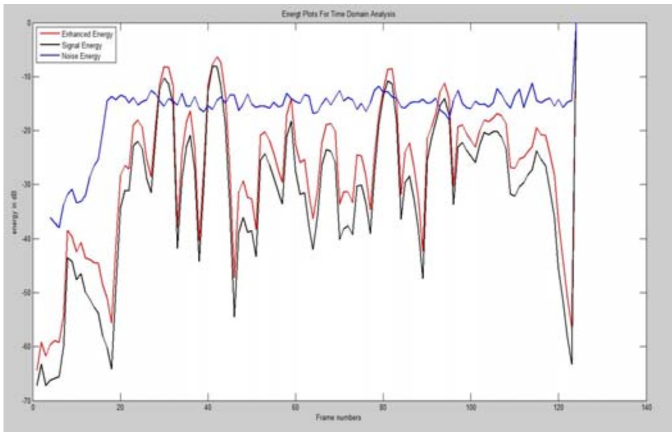

Figure 5: Energy Comparison Plot for Time Domain analysis.

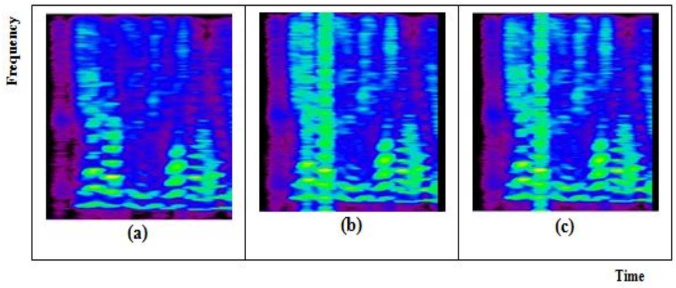

Figure 6: Spectrogram plots for (a) input signal, (b) and (c) Enhanced signal for babble noise and speech shaped noise from time domain analysis.

Figure 7(a) shows the SII for babble noise. Blue line corresponds to unprocessed speech signal and Red line corresponds to processed signal intelligibility. Results indicate that intelligibility is good for lower SNRs compared to higher SNRs. Lower SNR behavior for babble noise is shown in figure 8(a). Figure 7(b) shows the SII for speech shaped noise. Figure 8(b) highlights the improvement in intelligibility in lower SNRs.

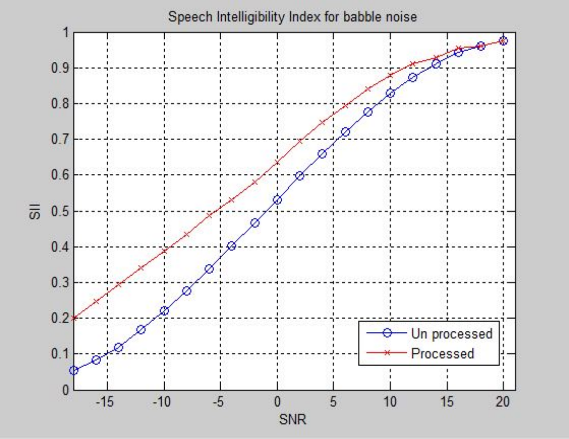

Figure 7(a): SII for babble noise.

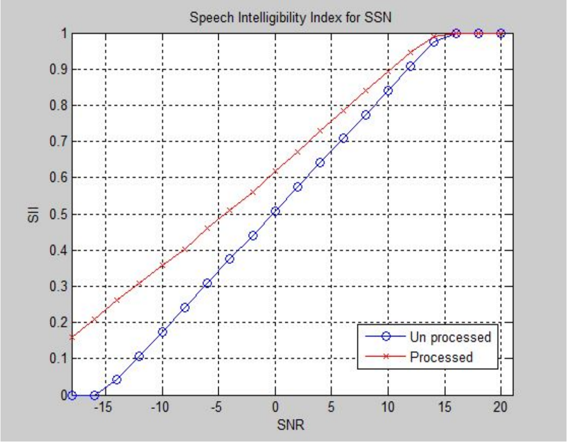

Figure 7(b): SII for SSN.

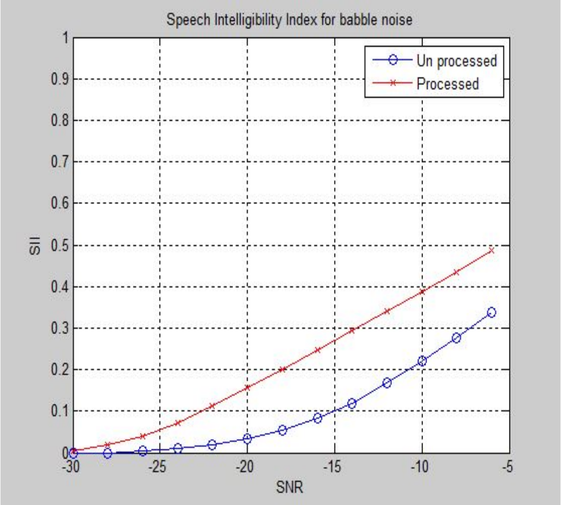

Figure 8(a): SII for babble noise at lower SNR levels.

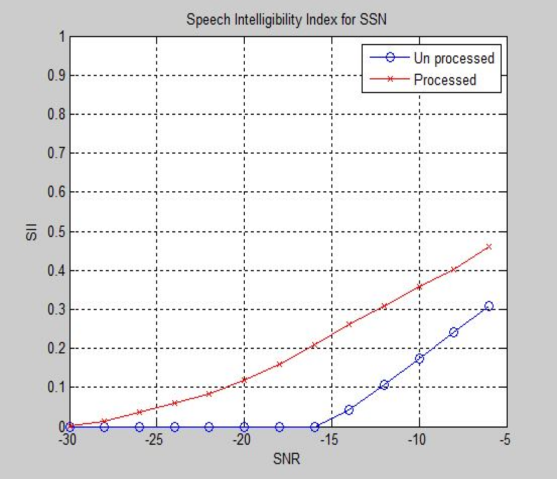

Figure 8(b): SII for SSN at lower SNR levels.

IV. CONCLUSIONS

In this work we presented an efficient near-end listening enhancement algorithm for the speech enhancement when listener is located in noisy environment. A time domain analysis is presented to improve the speech intelligibility in noisy environment for the near-end listener by gain adjustment of speech signal according to the background noise variation. Work is carried out by considering two types of background noise such as, babble and speech shaped noise. Results indicate that gain obtained is adaptive and varies with respect to change in speech and noise signals. Speech intelligibility index (SII) is measured for degraded speech signal and enhanced speech signal with respect to SNR varying from -18 to +21 dB and results have shown that the proposed algorithm has better speech intelligibility and higher advantages are found at lower SNR where noise is more dominant than speech signal.

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 Image Processing and Pattern Recognition*, Vol. 3, No. 3, pp. 87-95, Sept. 2010.

[3] 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.

[4] 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.

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

[6] Martin R., “Spectral Subtraction Based on Minimum Statistics”, *in Europe Signal Processing Conference*, Scotland, pp. 1182-1185, 1994.

[7] 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: 10.1,109/ICASSP. 2006.1660065, pp. 493-496, May 2006.

[8] ANSI, “Methods for Calculation of the Speech Intelligibility Index”, *S3.5-1997 (American National Standards Institute, New York)*, 1997.

[9] 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.

[10] Bastian Sauert and Peter Vary, “Near-end Listening Enhancement Optimized with Respect to Speech Intelligibility Index*”, Proceedings of European Signal Processing Conference* (EUSIPCO), Vol. 17, pp. 1844-1848, August 2009.

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

[12] 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, 2013, © Springer. DOI: 10.1007/ 978-1-4614-6154-8_68, pp. 699-707, Feb. 2013.

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

[14] 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, 2012.

[15] Taal C. H., Jesper Jensen and Arne Leijon, “On Optimal Linear Filtering Of Speech for Near-End Listening Enhancement”, *IEEE Signal Processing Letters*, Vol. 20, DOI: 10.1109/ LSP.2013.2240297, ISSN 1070-9908, pp. 225-228, 2013.

[16] Taal C.H., Jesper Jensen, “SII-based Speech Preprocessing for Intelligibility Improvement in Noise”, *Proceeding of Interspeech*, Lyon, France, August 2013.

[17] Westerlund N., “Applied Speech Enhancement for Personal Communication”, *Thesi*s, Blekinge Institute of Technology, 2003.

[18] 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.

[19] Gunawan T. S., Khalifa O. O. and Ambikairajah E., “Forward Masking Threshold Estimation using Neural Networks and its Application to Parallel Speech Enhancement”, *in International Conference on Computer and Communication Engineering*, Vol. 11, No 1, DOI 10.1109/ICCCE.2008.4580598, pp. 15-26, 2010.

[20] Gunawan T. S. and Ambikairajah E., “A New Forward Masking Model and its Application to Speech Enhancement”, *in Acoustics, Speech and Signal Processing,* 2006, DOI: 10.1109/ICASSP.2006.1659979, pp. 149-152, 2006.

[21] Jong Won Shin, YuGwang Jin, SeungSeop Park and Nam Soo Kim, “Speech Reinforcement Based on Partial Masking Effect”, *in IEEE International Conference on Acoustics, Speech and Signal Processing,* DOI: 10.1109/ICASSP.2009.4960605, pp. 4401 – 4404, 2009.

[22] Goldin A. A., Budkin A. and Kib S., “Automatic Volume and Equalization Control in Mobile Devices,” in *Audio Engineering Society 121th Convention*, Preprint No. 6960, Oct. 2006.

[23] Jong Won Shin and Nam Soo Kim, “Perceptual Reinforcement of Speech Signal Based on Partial Specific Loudness”, *IEEE Signal Processing Letters*, Vol. 14, No. 11, 2007. DOI: 10.109 /LSP.2007.900222, ISSN 1070-9908, pp 897-890.

.