Incorporating Auditory Masking Properties for Speech Enhancement in presence of Near-end Noise

Premananda B. S., Uma B. V.
R.V. College of Engineering,
Bangalore, India.

ABSTRACT
In mobile devices, perceived speech signal degrades significantly in the presence of background noise as it reaches directly at the listener's ears. There is a need to improve the intelligibility and quality of the received speech signal in noisy environments by incorporating speech enhancement algorithms. This paper focuses on speech enhancement method including auditory masking properties of the human ear to improve the intelligibility and quality of the speech signal in the presence of near-end noise. Implemented by dynamically enhancing the speech signal when the near-end noise dominates. Intelligibility and quality of enhanced speech signal are measured using SII and PESQ. Experimental results show improvement in the intelligibility and quality of the enhanced speech signal with the proposed approach over the unprocessed speech signal. This particular approach is far more efficient in overcoming the degradation of speech signals in noisy environments.

Keywords
Gain, Masking, Near-end noise, Speech enhancement, Speech intelligibility, Speech quality

1. INTRODUCTION
In the communication, most of the times speech signal is accompanied with noise. For instance, in mobile communication, the noise of the environment where the source of the speech lies is the principal component of noise that combines with the speech signal. The effect of this noise makes the listening task laborious for a direct listener. Speech enhancement algorithms not only involve processing speech signals for human listening, but also for further processing prior to listening.

There are two categories of speech enhancement algorithms: one being a noise reduction algorithm and the other speech enhancement algorithm. As the noise signal cannot be manipulated for near-end noise, a reasonable approach is to manipulate the speech signal depending on the near-end noise. Hence, necessitates the development of speech enhancement approaches to improve speech perception in adverse listening conditions. The objective of speech enhancement differs depending on specific applications, for example, to reduce listener fatigue, boost overall speech quality, to enhance intelligibility, or to increase the performance of communication systems [3].

At the receiving end, referred to as *near-end* in literature, the listener may be in a noisy environment. That makes hearing difficult even though the transmitting speech source is in a reticent environment because the near-end noise hits the listener's ear directly. As the quality of speech is lowered, the listener experiences fatigue. The speech enhancement algorithm should provide performance in a wide range of SNRs for both intelligibility and quality.

The presence of noise masks the speech signal and makes it less intelligent or audible. This effect is called masking and it can be of two types, one simultaneous masking (one signal is masked when another signal is present in this case noise mask speech) and other temporal masking (the signal is masked by noise before and after high noise occurs). Hence, the speech signal needs to be improved considering these situations also. The idea of studying masking effects in speech signal enhancement is to remove the non-audible spectral components of speech signals.

The far-end noise (at the transmitter) can be reduced using traditional noise suppressing algorithms like minimum mean-square error [18], spectral subtraction methods [19], etc. The methods proposed for far-end noise cancellation techniques discussed in the literature are not suitable as it focuses on reducing background noise at the speaker-end rather at the receiver. Several approaches to mitigate the near-end noise using speech enhancement are presented by Bastian S. et al., [4-6] and Taal C.H. et al., [7, 8]. [4] investigates listening enhancement under the constraint that the processed signal power is strictly equal to the received signal power.

Near-end listening enhancement (NELE) algorithm by Bastian S. et al., in [5] maximizes the speech intelligibility index (SII) [14] and thus the speech intelligibility by selective frequency enhancing of the speech signal power. Two SII based NELE algorithms are compared by Taal C.H. et al., in [7] to optimize the speech intelligibility in the presence of near-end noise and it focuses on new linear filtering of speech prior to the degradation due to near-end noise. He solved constrained optimization problem of [5] using a nonlinear approximation of the speech intelligibility which is accurate for lower SNRs.

NELE by Premananda B.S. et al., in [2] increases speech signal in presence of the noise and avoids listener fatigue. In [3] speech samples are given relative weight using absolute threshold hearing (ATH) but do not include the masking effect of signals. Approaches in [1-3] do not consider the acoustic speech samples, rather involves the improvement of both audible and non-audible samples, which results in waste of speech energy. The methods do not include auditory masking properties of the human ear. NELE algorithm by Teddy S. et al., [10, 11] provides an operative model of temporal masking, which uses a fractional bark gammatone filter bank related to the changes in NELE method.

Perceptual Evaluation of Speech Quality (PESQ) is used to determine the quality of enhanced speech signals. The PESQ score ranges from -0.5 to 4.5 in terms of quality of speech signals. [15-17] provide accurate and repeatable estimates of speech quality degraded by noise. Directions are provided in ITU-T recommendations P.800/P.830.

This paper presents a novel speech enhancement method including the auditory masking properties of the human ear. In order to improve intelligibility and quality of speech signals in the near - end side, the work tries to amplify the clean speech signal by adopting a psychoacoustic model which gives actual perceivable components in speech signal.

The paper is organized as follows: Section 2 describes a speech enhancement algorithm in the FFT domain with implementation steps. The loudness computation procedure is explained in section 3. Experimental results and conclusions are discussed in section 4 and 5 respectively.

2. SPEECH ENHANCEMENT IN THE FFT DOMAIN

A frequency domain NELE approach to combating the degradation of the speech signal in a noisy environment is proposed here. Figure 1 illustrates the overall block diagram of the proposed approach. Degradation of intelligibility due to the presence of near-end noise can be reduced by pre-processing the clean speech signal before played in noisy environments or fed to the mobile loudspeakers. The multiplier is used to enhance the speech signal, degraded due to the presence of near-end noise. The speech samples are to be enhanced by multiplying with a varying gain by comparing the loudness/energy of speech and noise samples.

We assume that a clean speech signal with far-end noise suppressed (using noise-cancellation techniques) is available. The near-end noise can be recorded using a dummy microphone in mobile phones. The loudness of the speech and near-end noise samples are calculated and compared and gain is computed for enhancing the speech signal in pre-processor block, as illustrated in Figure 2 when the near-end noise dominates the received clean speech signal.

Steps involved in frequency domain NELE are:

Step 1: The variance in the noise signal is altered to fix the required Signal to Noise Ratio (SNR) using equation (1) as noise and speech signals are dynamic in nature. For illustration, SNR is varied from -15 to +20 dB.

$$n' = \frac{n}{norm(n)} \cdot \frac{norm(s)}{10^{0.05 \cdot SNR}} \quad (1)$$

where n and s are captured noise and speech signal.

Step 2: Compute loudness of speech and noise signals

The speech signal loudness of a frame is computed using equation (2).

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

where x_i is a sample at the i^{th} location and i vary between 1 to N when $x_i > 0$, N is the perceivable samples in a frame.

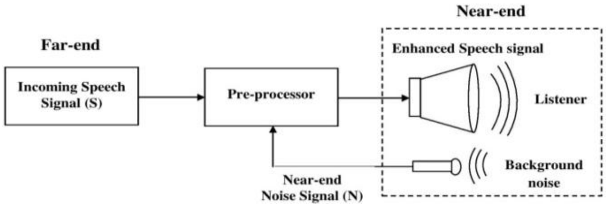

Figure 1. Proposed block diagram for speech enhancement

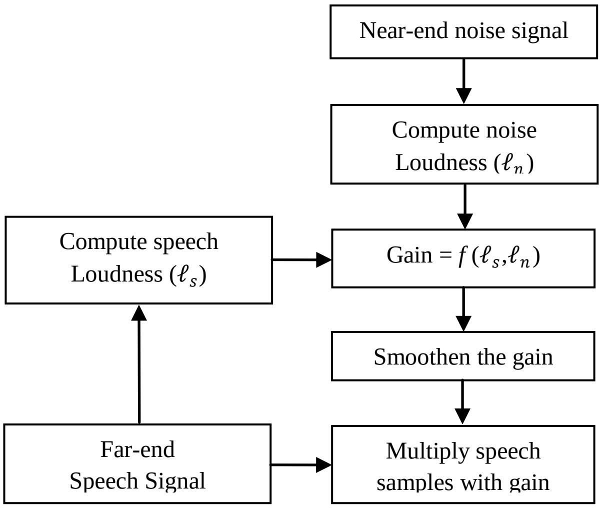

Figure 2. Flow diagram of pre-processor block of Figure 1

Repeat the loudness calculation for every frame. The same procedure is used to obtain loudness of the noise. Let ℓ_s and ℓ_n denote loudness of speech and noise signal for each frame. Computation of loudness of the noise and speech signal is discussed in section 3.

Step 3: Deriving the Gain

The suitable gain for a couple of speech and noise frames is user specific and depends on multiple constraints. When the speech loudness (ℓ_s) is less than noise loudness (ℓ_n), we should enhance the speech signal more than the difference between them. Hence, gain can be derived using equation (3).

$$G = (\ell_n - \ell_s) \quad (3)$$

Gain computed using equation (3) for adjacent frames vary randomly, hence to control the gain it is multiplied by a compensation factor ∃, which can be arbitrarily chosen (< 1). Hence, the gain equation can be written using equation (4).

$$G = (\ell_n - \ell_s) \exists \quad (4)$$

When ℓ_s is greater than (by 3 dB) ℓ_n then gain, g_0 should be set to 1. Since no enhancement is required, equation 4 becomes negative when added with g_0 resulting in a gain less than 1. To prevent equation 4 from becoming negative, we should set it to zero. Hence, the equation (4) can be modified considering the above constraints to obtain the equation (5).

$$G = g_0 + max(0, (\ell_n - \ell_s)\exists) \quad (5)$$

where max is used to find the maximum. When ℓ_s is approximately equal to ℓ_n, the gain can be appropriately set (e.g. 1.21, to enhance by 1 dB), generally ℓ_s must be 3 dB greater than ℓ_n to be heard clearly.

Step 4: Smoothen the gain (G_{avg})

The calculated gain (in step 3) when multiplied with the speech signal results in sudden changes in the output levels. The gain computed using equation (5) is to be limited to avoid clicks and pops due to sudden changes in the output level that fatigues the listener's ear. The gain obtained in the current frame is averaged with the previous and future frames to make the gain variation smooth using equation (6).

$$G_{avgi} = \frac{G_{i-2} + G_{i-1} + G_i + G_{i+1} + G_{i+2}}{5} \quad (6)$$

where i is the current frame.

Depending on the delay, tolerable by the system the number of pre and post frames can also be selected. For example, if gain variations for adjacent frames are minimal, it is sufficient to consider the immediate preceding and succeeding frame.

Step 5: Multiply averaged gain with the speech samples

When the noise dominates, averaged gain, multiply G_{avg} of every frame with perceivable samples of respective frames, and enhance the speech samples.

Step 6: End-capping

Improved speech samples should not exceed the maximum spectrum level (90 dB [4]). If an enhanced speech sample value exceeds the maximum power [4] of the mobile speaker [6]. Limit the minimum and maximum values computed by normalizing the samples. Speech buffer value can be overridden with a new value, and then fed to the mobile speakers.

3. LOUDNESS COMPUTATION OF SIGNALS

The psychoacoustic studies have revealed that the reception of all the frequencies by a human ear is not the same. Due to the presence of various sounds in the environment along with the drawbacks of the human auditory system leads to evidence that we can remove the inessential data in the speech signal. The two traits of the human auditory system that constitute the psychoacoustic model are an absolute threshold of hearing and auditory masking. They provide a method of finding which samples of a signal are not heard can be removed from the signal.

3.1 Absolute Threshold of Hearing

With normal hearing, the ATH is the minimum sound level of a tone that can be heard in the absence of extraneous sounds. This is also known as the auditory threshold or threshold in quiet, (dB) [13] T_q, is approximately calculated using an empirical equation (7).

$$T_q(f)=3.64\left(\frac{f}{1000}\right)^{-0.8}-6.5e^{-0.6\left(\frac{f}{1000}-3.3\right)^2}+10^{-3}\left(\frac{f}{1000}\right)^4 \quad (7)$$

where f is the frequency (Hz).

The audio frequency of a human that ranges from 20 to 20,000 Hz can be split up into critical bandwidths, which are non-linear, non-uniform, and are dependent on the perceived sound. Signals present in a critical band are difficult to separate for a listener. A uniform measure of frequency based on critical bandwidths is the Bark. The relation between frequency and Bark [13] is given using equation (8), where LHS represents the frequency in Hz, and the RHS represents the equivalent Bark.

$$f\ (\mathrm{Hz}) = 1.3 \arctan\left(0.00076f\right)+3.5 \arctan\left[\left(\frac{f}{7500}\right)^2\right] \quad (8)$$

Equation (9) gives an analytical expression that describes the variation of critical bandwidth Δf as a function of the masker center frequency fc.

$$\Delta f = 25+75\left[1+1.4\left(\frac{f_c}{1000}\right)^2\right]^{0.69} \quad (9)$$

3.2 Auditory Masking

Masking occurs when the perception of one signal is affected by the presence of another signal. The amount of masking increases the detection threshold of a signal due to the presence of a masker sound. In the presence of a signal, this threshold is elevated in its vicinity of time & frequency. The basic idea of masking [10-13] model is to preserve the audibility trait of the speech signals from the derived masking thresholds.

For the proposed NELE algorithm, we have to determine:

1. Tone maskers or Tonal components
2. Noise maskers or Non – tonal components
3. Combined masking effect of tone and noise maskers

If any frequencies near to these maskers are below the masking threshold, those frequencies are not heard.

3.2.1 Tone Maskers

For a signal frequency component to be a tone, it should be constant for a particular period. It should be local maxima in the frequency spectrum, indicates that it is higher than the noise component of the signal. The signal frequency with FFT index ‘m’ is considered to be tone, if its power P [m] satisfies the following two conditions:

1. It should be more than P [m-1] and P [m+1], which indicates that it is a local maximum.
2. It should be 7 dB greater than the rest of the frequencies in its neighborhood (two).

When found, take the power at one position previous to [m-1] and the one following [m+1] and merge it with the power of [m] to make a tone masker estimation, as the tone may essentially be among the frequency samples.

3.2.2 Noise Maskers

If a signal is not a tone, it should be noise, consider all the frequency components that are not elements of a tone's neighborhood as noise. Humans have a difficulty in discriminating signals inside a critical band the noise inside each of the bands is pooled to appear as one mask. The notion is to find all the frequency components inside a critical band that does not lie in the vicinity of the tone, add them as one. Keep them at the mean (geometric) location inside the critical band and repeat the process for all critical bands.

Next, remove maskers that are close to each other to optimize the maskers. Retain the maskers possessing power above the ATH, and eliminated the remaining maskers because they will not be audible. Then the maskers that have other maskers within their critical bandwidth are located, and if found, the masker having lower power between them is set to zero because the human ear will not hear it.

3.2.3 Masking Effect

Spreading of masking determines the shape of the masking pattern of a masker to the lower frequency and to the higher frequency of the masker. The masking curve shapes are easier to describe in the Bark scale that is linearly related to basilar membrane distances. The models of the spreading of masking are used to approximate simultaneous masking models that work in the frequency domain. The maskers influence the frequencies inside a critical band, as well as those in the neighboring bands.

In literature, it is indicated that the spreading of these maskers has a slope of +25 dB/Bark preceding and -10 dB/Bark following the masker. The spreading of masking can be approximated as a function that relies on the maskee position i, and masker position j. The power spectrum Pt at j, and the difference in Bark scale between masker and maskee as given in equation (10).

$$\delta M = M(f_{maskee}) - M(f_{masker}) \tag{10}$$

Table 1 lists the conditions of spreading function. The masking thresholds and effect of tone and noise maskers are calculated using equation (11) and (12) respectively.

Table 1. Conditions of Spreading function

Spread function, SF (i, j)	Delta conditions, δM
17δM - 0.4Pt(j)+11	-3 <= δM< -1
(0.4Pt(j)+6)δM	-1 <= δM< 0
-17δM	0 <= δM< 1
(0.15Pt(j)-17)δM - 0.15Pt(j)	1 <= δM< 8

Here, Pt is the power spectrum of tone at j, Pn is the power spectrum of noise at 'j'. SF is the spread function which is modelled as described in Table 1 and 'i' is maskee position.

$$tm(i,j) = Pt(j)-0.275\, z(j)+SF(i,j)-6.025 \tag{11}$$

$$nm(i,j)=Pn(j)-0.175\, z(j)+SF(i,j)-2.025 \tag{12}$$

Taking into account the ATH and spectral densities of tone and noise maskers with all masking thresholds, determine the overall global masking threshold. In this method we assume that the effects of masking are additive, so the masks of all maskers are summed and multiplied with the ATH only when the tone masker and noise masker cross the ATH. Global masking threshold is the overall threshold obtained along with the spreading function and is called as the practical threshold of hearing (PATH).

3.3 Implementation of the Psychoacoustic Model

Steps involved in the implementation of the psychoacoustic model in MATLAB to compute the loudness of the perceivable samples are:

1. Read the speech signal in .wav format using wavread function with a sampling frequency of 8 kHz.
2. Divide the signal into frame of 256 samples each (32 ms) using a window function.
3. Determine the power spectral density (PSD) of the frame using 256 point FFT.
4. Locate the tone and noise maskers within the frame and their positions inside each critical band.
5. For optimizing masker's check if a masker is lower than the ATH, it should be eliminated. Remove the masker with less power, if two maskers (tone or noise) are inside the critical band.
6. Compute the masking threshold for every masker and add the masking thresholds to obtain the global masking threshold of other frequencies in that frame.
7. Select the samples that are above the global masking curve these correspond to the perceivable samples in that frame (PATH) and store them in a buffer.
8. Compute the loudness of the speech samples which are stored in the buffer using equation (2).
9. Repeat steps 3-8 for all the frames of the signal.

4. EXPERIMENTAL RESULTS

Noise and speech signals both with a sampling frequency (Fs) of 8 kHz each is recorded for the duration of 4 seconds using GoldWave, an audio editor tool and is saved in .wav format for analyzing. The recorded signals have 32000 samples, and the samples are divided into frames of size 256 each, resulting in 125 frames, with each frame corresponding to 32 ms. The variance in the speech-shaped noise is adjusted to obtain the desired SNR using equation (1) in the recorded signal in the range -15 to 20 dB. Verified enhanced speech signals using proposed NELE methods using MATLAB and GoldWave.

The result obtained by applying a Psychoacoustic model for a frame of the recorded signal is discussed. After determining the maskers that are necessary for each frame, obtain the masking threshold of each masker using equations (11) and (12).

Figure 3 (blue line) shows the overall masking threshold for the arbitrary (eleventh) frame of the speech signal. It is a cumulative effect of the spread function multiplied with ATH (red line). PSD of the frame is denoted in the green line. The samples having power below PATH are unperceivable, and only samples above PATH are audible, as shown in Figure 4. Extract and store both types of samples in a separate buffer for every frame. Use them as input for the enhancement algorithms. After the samples that are above PATH are identified, loudness of both speech and noise samples is calculated using equation (2).

The loudness of speech and noise samples are compared frame wise, and the gain is calculated using equation (5). Optimal/smoothened gain is calculated by averaging gain (equation 6) with two pre and two post frames gain. Original and the smoothed gain are plotted in Figure 5, red line represents the original gain and the blue line represents the updated/ smooth gain. By observing the Figure 5, it is clear that abrupt changes are rectified in the smoothened gain.

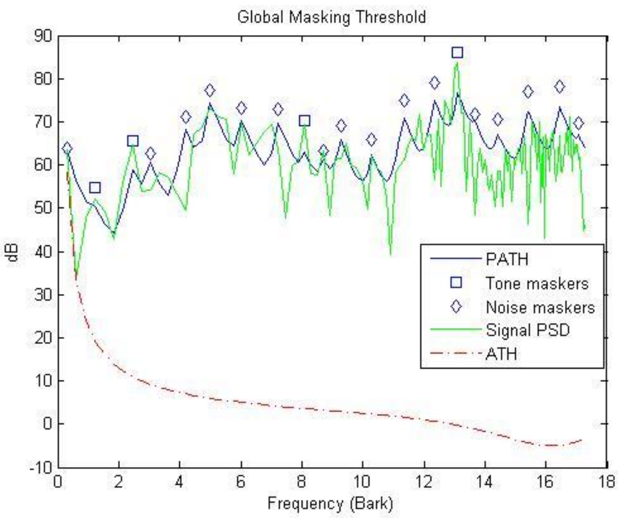

Figure 3. Overall masking threshold of a frame

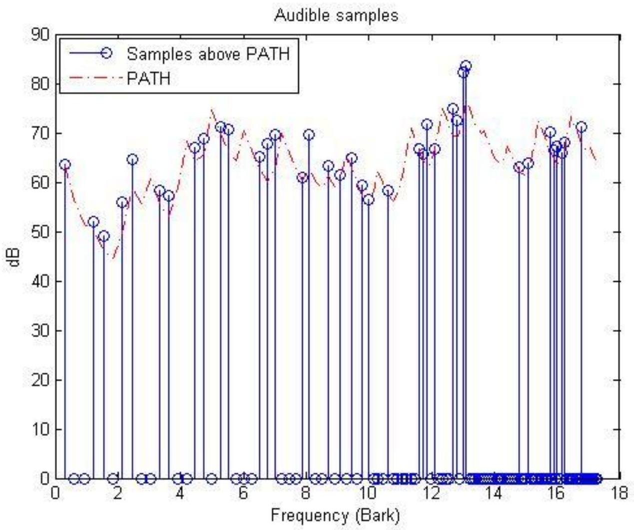

Figure 4. Audible samples of the signal in a frame

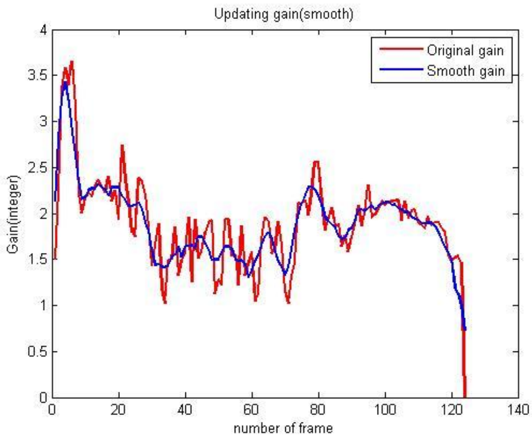

Figure 5. Smoothened gain

4.1 Enhancement of Speech Signal

For enhancing the speech signal we considered following enhancement algorithms:

1. Overall enhancement of speech samples (Time domain [2])
2. Only speech samples above PATH to be enhanced
3. Speech samples below PATH to be enhanced

Based on the results it was realized that among implemented three algorithms first one is not valid because it enhances overall samples, including the samples that are not audible also hence waste of mobile power/battery. Also limits the gain range by unnecessary enhancement of unwanted or non-audible samples. The second method is more practical than the first, since the samples above PATH that are audible are enhanced. In third approach, when we enhance samples below PATH may create new maskers that induce change in PATH itself, hence cannot be considered.

4.2 Speech Intelligibility Measurement

SII is used to check the performance of the proposed algorithm. Intelligibility of enhanced signal is measured based on the standardized SII procedure as outlined in [14]. The procedure for calculating the SII is given in [7, 8]. SII predictions are calculated for the unprocessed (original) and processed speech signals in the presence of speech-shaped near-end noise. SNRs in the range between -15 to 20 dB and is compared with [2] and [8]. The obtained results are plotted in Figure 6. It is evident that SII is almost increased by 0.09 (9 %) and 0.12 (12%) when compared to [8] and [2] and around (0.2) 20 % w.r.t. unprocessed signal.

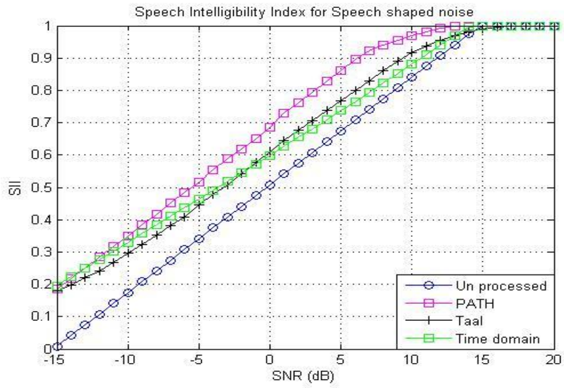

Figure 6. SII predictions in presence of speech-shaped noise

The proposed approach enhanced only the audible samples to improve the intelligibility in all the ranges of SNR. Hence, the proposed method improves the intelligibility of speech signals as predicted by the SII.

4.3 Speech Quality Measurement

The speech quality of the enhanced speech signal is estimated using PESQ. [9] provided accurate and repeatable estimates of speech quality degraded by noise. Respective PESQ scores of enhanced (using PATH) signal in the presence of speech-shaped noise for SNR in the range -15 to +15 dB are tabulated in Table 2. $PESQ_{ref}$ is the PESQ score for the clean and corrupted speech (because of near-end noise) signals then PESQ score of the enhanced speech was also measured and denoted as $PESQ_{proc}$. It is observed from the Table 2 that the PESQ scores are improved when compared to perverted speech signal.

Table 2. Comparison of PESQ scores

SNR	Speech-shaped noise	
	$PESQ_{ref}$	$PESQ_{proc}$
-15	0.518	2.065
-10	0.735	2.171
-5	1.072	2.291
0	1.483	2.798
5	1.87	3.634
10	2.225	4.302
15	2.574	4.394

5. CONCLUSIONS

In this work, an approach is presented to enhance the intelligibility and quality of speech signals corrupted due to the presence of near-end noise. Simulation results were tested using MATLAB and an audio editor tool, GoldWave. Enhanced audible speech samples in the presence of simultaneous masking effects. The proposed algorithm has better speech intelligibility, as measured using SII, providing roughly 10 % improvement when compared to [2] and 20 % over unprocessed speech signal. From PESQ results, it is revealed that the proposed algorithm increases the speech quality also. The NELE method leads to a significant increase in intelligibility without compromising on quality.

6. REFERENCES

[1] Premananda B.S., and Uma B.V., "Speech Enhancement Algorithm to Reduce the Effect of Background Noise in Mobile Phones", International Journal of Wireless and Mobile Networks (IJWMN), Vol. 5, No. 1, pp. 177 - 189, Feb. 2013.

[2] 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, Vol. 131, Springer, pp. 699 - 707, Feb. 2013.

[3] Premananda B.S., Manoj and Uma B.V., "Near-End Perception Enhancement using Dominant Frequency Extraction", International Journal of Advanced engineering and Research Development (IJAERD), Vol. 1, No. 6, pp. 351-358, June 2014.

[4] Bastian Sauert and Peter Vary, "Near End Listening Enhancement Considering Thermal Limit of Mobile Phone Loudspeakers," Proceedings of Elektronische Sprach Signal Verarbeitun *(*ESSV), Vol. 61, Germany, pp. 333–340, Sept. 2011.

[5] 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, pp. 493 - 496, 2006.

[6] 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, pp. 1919 - 1923, August 2010.

[7] Taal C.H., Jensen J., and Leijon A., "On Optimal Linear Filtering of Speech for Near-End Listening Enhancement", IEEE Signal Processing Letters, Vol. 20, No. 3, pp. 225 - 228, March 2013.

[8] Taal C.H., and Jesper Jensen, "SII-based Speech Pre-processing for Intelligibility Improvement in Noise", Proceeding of Interspeech, Lyon, France, pp. 3582 - 3586, August 2013.

[9] Jeon Yu-young and Lee Sang-min, "A Speech Enhancement Algorithm to Reduce Noise and Compensate for Partial Masking Effect", Journal of Central South University of Technology, Vol. 18, issue 4, pp. 1121 - 1127, August 2011.

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

[11] 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, pp. 15 - 26, 2010.

[12] Yi Hu and Philipos C. Loizou, "Incorporating a Psychoacoustic Model in Frequency Domain Speech Enhancement", IEEE signal processing letters, Vol. 11, No. 2, pp. 270 - 273, February 2004.

[13] Eberhard Zwicker and Hugo Fastl, Psychoacoustics, Facts and Models. New York: Springer, 2007.

[14] American National Standard. Methods for the Calculation of the Speech Intelligibility Index. ANSI S3.5-1997, 1997.

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

[16] 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 and Codecs", IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 2, pp.749 - 752, May 2001.

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

[18] Ephraim Y., and Malah D., "Speech Enhancement using a Minimum Mean Square Error Short-Time Spectral Amplitude Estimator", IEEE Transaction on Acoustic, Speech, and Signal Processing, Vol. ASSP-32, pp. 1109 - 1121, December 1984.

[19] Malihe Hassani and Karami Mollaei M. R., "Speech Enhancement based on Spectral Subtraction in Wavelet Domain", Seventh IEEE International Colloquium on Signal Processing and its Applications, pp. 366 - 370, March 2011.