ISSN 2229-5518

Speech Enhancement in Presence of Near-end Noise by Incorporating Auditory Masking in Frequency Domain

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

Abstract — *In mobile devices, perceived speech signal detoriates significantly in the presence of near-end noise as it reaches directly at the listener's ears. There is an inherent need to improve the clarity and quality of the received speech signal in noisy environments by incorporating speech enhancement algorithms. This paper focuses on speech enhancement method including simultaneous masking properties of the human ear to select only audible samples of speech signals and to improve the intelligibility and quality of the speech signal. Speech enhancement algorithm is 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 that the intelligibility and quality improvement of the speech signal with the proposed approach over the unprocessed far-end speech signal. This particular approach is far more efficient in overcoming the degradation of speech signals in noisy environments.*

Index Terms — *Gain, Maskers, Near-end noise, Speech enhancement, Speech intelligibility index, Speech quality, Psychoacoustic.*

1 Introduction

Mobile phones are the most popular consumer devices in the present day. For a conversation in silent environments, less speech energy is required for the speakers to understand each other. However, for instance, if the vehicle passes by, the conversation is severely disturbed [4]. To overcome this effect, we should either wait until the vehicle passes or raise the signal to produce more speech energy and increase loudness. The external volume control of the mobile phones cannot be used for this purpose as background noise changes dynamically. As the noise signal cannot be fabricated for near-end noise, a reasonable approach is to manipulate the speech signal depending on the surrounding noise [7]. Hence, necessitates the development of speech enhancement algorithms to increase speech perception in adverse listening conditions.

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

The procedures proposed for far-end noise cancelation techniques discussed in the literature [18], [19], [20] are not suitable. It focuses on mitigating background noise on the speaker-end rather at the receiver. Several speech enhancement approaches to mitigate the effect of near-end noise are discussed by Bastian S. et al., [4], [5], [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 with 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. Paper 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 non-linear approximation of the speech intelligibility which is accurate for lower SNRs. Speech enhancement process in [1] increases speech signal above the near-end noise and avoids listener fatigue. In [2], [3] speech samples are given relative weight using absolute threshold hearing but do not include the masking effect of signals. Approaches in [1], [2], [3] do not consider the audible speech samples, rather involves the enhancement of both audible and non-audible samples, and results in waste of speech energy. 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 speech improvement method.

This paper discusses the speech enhancement methods including the masking properties of the human ear. The paper is organized as follows: Section 2 describes NELE in frequency domain with implementation details. In section 3, loudness computation procedure is explained. In section 4 and 5 experimental results and, conclusions are discussed.

- *Premananda B.S., Dept. of Telecommunication, R.V. College of Engineering, Bangalore, India, E-mail:premanandabs@gmail.com*
- *Uma B.V, Professor, Dept. of Electronics & Communication, R.V. College of Engineering, Bangalore, India, E-mail:umabv@rvce.edu.in*

ISSN 2229-5518

2 NELE in Frequency Domain

We propose a frequency domain (including simultaneous masking properties) NELE approach for improving the speech signal perception in a noisy environment. We assume that a clean speech signal with far-end noise removed (using noise-cancellation techniques) is available. 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. Fig. 1 illustrates the overall block diagram of the proposed approach. The speech samples are to be enhanced by multiplying with a dynamic gain by comparing the energies of speech and noise samples. The surrounding noise can be recorded using a dummy microphone in mobile phones. The loudness of the near-end noise and speech samples are calculated (refer Fig. 2) and compared, and gain is computed for improving the speech signal in pre-processor block when the near-end noise dominates the received clean speech signal.

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

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

where n and s are recorded noise and speech signal.

Steps involved in frequency domain approach for enhancement of speech samples when noise dominants are:
Step 1: Record the noise and speech signal for a finite duration with a sampling rate of 8000 samples/sec.
Step 2: Compute loudness of noise and speech samples
The speech loudness of a frame is calculated using (2).

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

where x_i is the sample at the i^{th} location, N is the total number of perceivable samples in a frame.

Repeat the loudness calculation for every frame and the same procedure is used to compute the loudness of the noise samples.

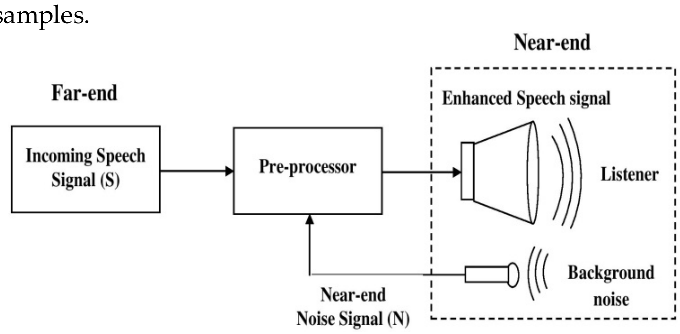

Fig. 1. Proposed block diagram for speech enhancement.

Step 3: Derive 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), then compute Δ,

$$\Delta = (l_n - l_s) \tag{3}$$

For a speech signal to be heard, l_s must be greater (by Γ dB) than l_n in a noisy environment and Γ is set to 3 dB. Then gain can be derived using an empirical formula given in (4).

$$G = 10^{\left(\frac{\Delta+\Gamma}{10}\right)} \tag{4}$$

Gain calculated using (4) for adjacent frames vary more randomly, hence to scale the gain, and it is multiplied by a compensation factor, E . It can be arbitrarily chosen (< 1) depending on noisy environment. Hence, the gain can be computed using (5).

$$G = G \cdot E \tag{5}$$

When l_s is greater than (by 3 dB) l_n then no enhancement is required and gain is set to 1.
Step 4: Gain Smoothening
The computed gain when multiplied with the speech samples results in sudden change in signal levels. Hence gain computed using (5) is to be limited to avoid clicks and pops due to erratic changes in the output level which fatigues the human ear.

The gain obtained in the current frame is averaged with the previous and future frames to make the gain variation smooth using (6). Depending on the delay tolerable by the system, number of pre and post frames can be decided. For example, if gain variations in adjacent frames are minimal, it is suffices to consider the immediate preceding and succeeding frame.

$$G_{avgi} = \frac{G_{i-N} + \cdots + G_{i-1} + G_i + G_{i+1} + \cdots + G_{i+N}}{2N+1} \tag{6}$$

where i, current frame and N, number of frames.
Step 5: Multiply averaged gain with the speech samples
Multiply averaged gain, G_{avg} of every frame with perceivable speech samples of respective frames, to enhance the speech samples. Improved speech samples should not exceed the maximum spectrum level (90 dB [4]). If an enhanced speech sample value exceeds the maximum energy [4] of the mobile speaker [6], then limit the minimum and maximum values computed by normalizing the samples.

3 Loudness Computation

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 masking effects 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:

ISSN 2229-5518

absolute threshold of hearing (ATH) and auditory masking. They provide a method of finding samples of a signal that are not audible.

3.1 Absolute Threshold of Hearing

The ATH is the minimum sound level of a pure tone that an average listener with normal hearing can hear in the absence of extraneous sounds, also known as the auditory threshold or threshold in quiet. The threshold in quiet (dB) [13], is approximately calculated using (6) in [1]. 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 measure of frequency based on critical bandwidths is the Bark.

The relation between frequency and Bark [13] is given using (7), where LHS represents the frequency in Hz, and the RHS represents the equivalent Bark. Bark bandwidth is smaller at low frequencies (in Hz) and larger at high ones. Discard the frequency components that have the power levels below the auditory threshold. The listener will not be able to hear these frequencies of the signal.

$$f = 1.3\arctan(0.00076f) + 3.5\arctan[(\frac{f}{7500})^2] \quad (7)$$

3.2 Auditory Masking

Masking increases the threshold of a sound due to the presence of a masker sound. In the presence of maskers, threshold is changed in its vicinity of time and frequency. The masking [10-13] model preserve the audibility trait of the speech signals from the derived masking thresholds.

For the proposed speech enhancement algorithm, we determine:

- Tone maskers
- Noise maskers
- 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.

Tone Maskers: For a signal frequency component to be a tone, it should be constant for a particular period. The signal frequency with FFT index k is considered to be tone if its power P [k] satisfies the following two conditions:

1. Should be more than P [k-1] and P [k+1], which indicates that it is a local maximum.
2. 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 [k-1], and the one following [k+1] and merge it with the power of [k] to make a tone masker estimation.

Noise Maskers: If a signal is not a tone, then 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 eliminate 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.

3.3 Masking Effect

Spreading of masking determines the shape of the masking pattern of a masker to the lower and 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. SF is the spread function which is modelled as described in Table 1 and 'i' is maskee position. The masking thresholds and the masking effect of tone and noise maskers are calculated using (8) and (9) respectively. Here, Pt is the power spectrum of tone at j, Pn is the power spectrum of noise at 'j'. Taking into account the ATH and spectral densities of tone and noise maskers with all masking thresholds, determine the overall global masking threshold.

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

$$tm\,(i, j) = Pt(j) - 0.275\, z(j) + SF(i, j) - 6.025 \quad (8)$$

$$nm\,(i, j) = Pt(j) - 0.175\, z(j) + SF(i, j) - 2.025 \quad (9)$$

In this method we assume that the effects of masking are additive, the masking effect from all the maskers are summed

ISSN 2229-5518

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).

Detailed steps involved in the computation of loudness of the speech samples using above equations using MATLAB are illustrated in Fig. 2. Similarly, loudness of the noise samples is computed. The steps are repeated to compute the loudness of the entire signal.

4 Experimental Results

Noise and speech signals with a sampling frequency 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 grouped into frames of size 256 each, resulting in 125 frames, with each frame corresponding to 32 ms. The power spectral density (PSD) of each frame is computed using 256 point FFT. For multiple speech signals, the proposed algorithm was checked in the presence of different near-end noises. Two types of near-end (train and speech-shaped) noise with different SNRs are generated.

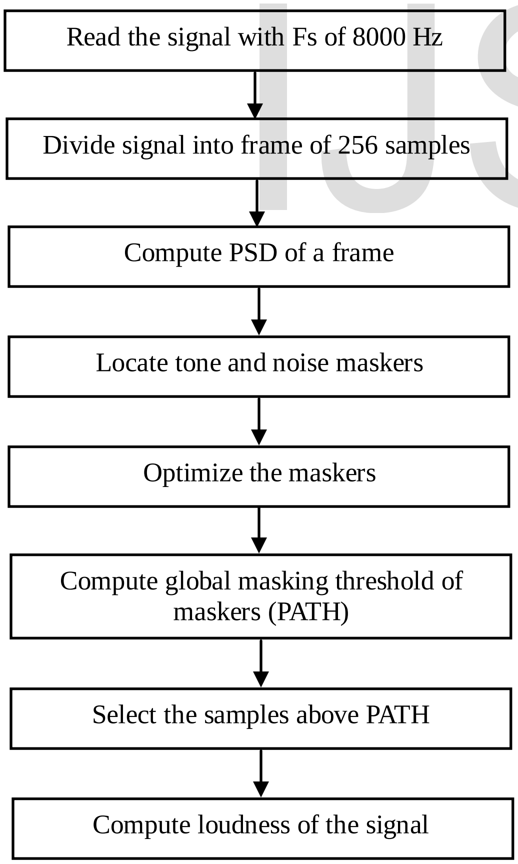

Fig. 2. Loudness computation procedure.

The variance in the recorded noise is adjusted to obtain the desired SNRs in the range of -15 to 20 dB. Verified clean and enhanced speech signals using both MATLAB and GoldWave. The result obtained by incorporating masking model for a frame of the recorded signal is discussed. Obtain the maskers (tone or noise) that are relevant for each frame, and compute the masking thresholds of each masker. Fig. 3 highlights the overall masking threshold for an arbitrary (23rd) frame of the speech signal. It is a cumulative effect of the spread function multiplied with ATH. The samples having power below PATH are unperceivable, and only samples above PATH are audible. Extract and store both types of samples in a separate buffer for every frame. Audible samples are used as input for the enhancement algorithm.

After the samples that are above PATH (audible) are identified, loudness of both noise and speech samples are calculated. The loudness of speech and noise samples are compared frame wise, and the gain is calculated. Optimal/smoothened gain is computed by averaging gain using (6), and abrupt changes can be rectified in smoothed gain.

4.1 Enhancement of the Speech Signal

For enhancing the speech signal, we considered the following enhancement algorithm:

1. Overall enhancement of speech signals (Time domain [2])
2. Enhance only speech samples above PATH
3. Enhance only speech samples below PATH

The energies of improved speech signals obtained in above three algorithms are plotted with respect to frames as shown in Fig. 4.

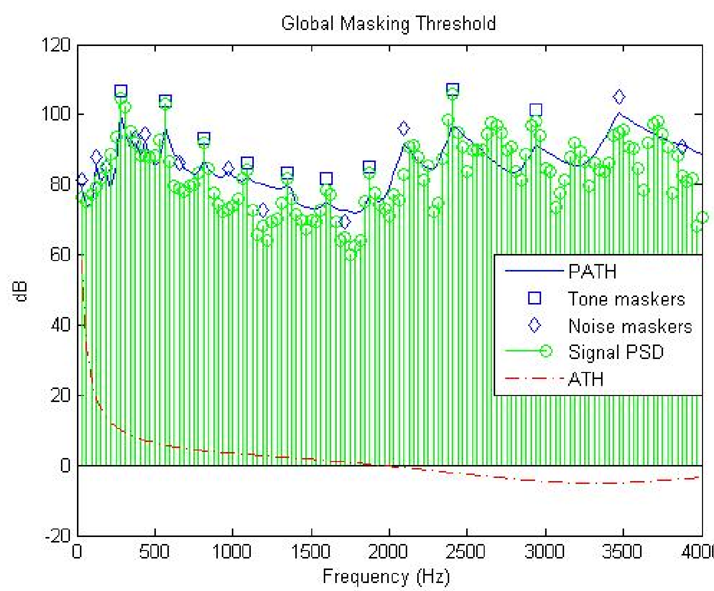

Fig. 3. Overall masking threshold of a frame.

ISSN 2229-5518

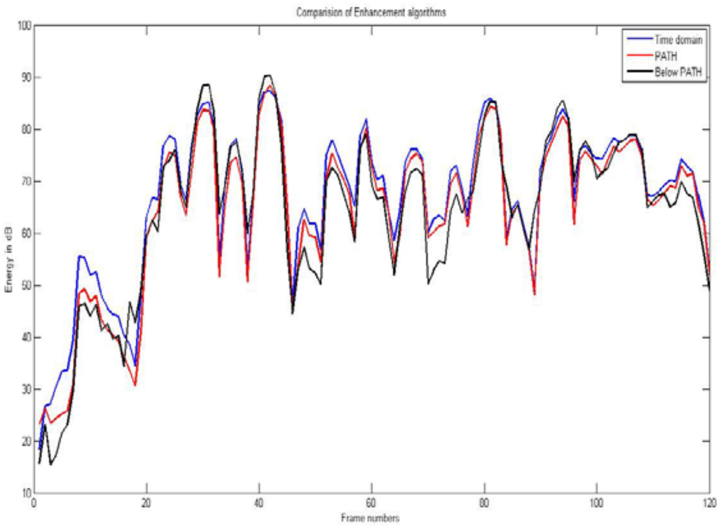

Fig. 4. Speech signal energies for different algorithms.

From results it was later realized that the first approach is not valid because it enhances overall samples, including the samples that are not loud hence waste of mobile power/ battery. Also limits the gain range by unnecessary enhancement of unwanted samples. The second method is more practical than the first, since the samples above PATH that are audible are enhanced. In third one, when we enhance only samples below PATH, may create new maskers that induce change in PATH itself, hence cannot be considered. Spectrograms of the unprocessed speech and enhanced speech signal in presence of train and speech shaped near-end noise is shown in Fig. 5.

4.2 Speech Intelligibility Measurement

Performance of the proposed enhancement algorithms is evaluated in terms of the SII. Intelligibility of the enhanced speech signal is measured based on the standardized SII procedure as described in [14]. For calculating the SII, steps are outlined in [7], [8]. SII predictions are calculated for the unprocessed (original) and processed speech signals in presence of two near-end (train and speech-shaped) noise and is compared with [1] and [8]. SNR is varied in the range between -15 to 20 dB for analyzing.

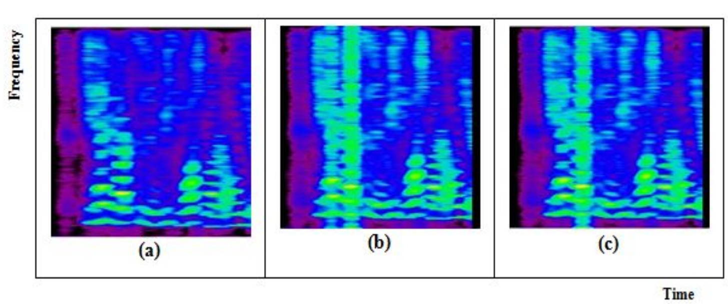

Fig. 5. Spectrogram for (a) Speech signal (b) and (c) Enhanced speech signal in the presence of train and speech-shaped noise.

The obtained results are plotted in Fig. 6 and 7 respectively, in presence of train and speech-shaped noise. In Fig. 6, SII is almost increased by 0.09 (9 %) and 0.12 (12%) when compared to [8] and [1] and around (0.2) 20 % w.r.t. unprocessed speech signal. In Fig. 7, SII is improved by 0.08 (8 %), 0.15 (15%) and (0.18) 18 % when compared to [8], [1] and unprocessed signal. [1] was not efficient in increasing intelligibility as overall samples are increased in number, but the proposed approach increased the intelligibility in all SNR. Hence, the proposed method improves the intelligibility of speech signals as predicted by the speech intelligibility index.

4.3 Speech Quality Measurement

The speech quality of the enhanced speech signal is estimated using Perceptual Evaluation of Speech Quality (PESQ). Guidelines are provided in ITU-T recommendations P.800/P.830.

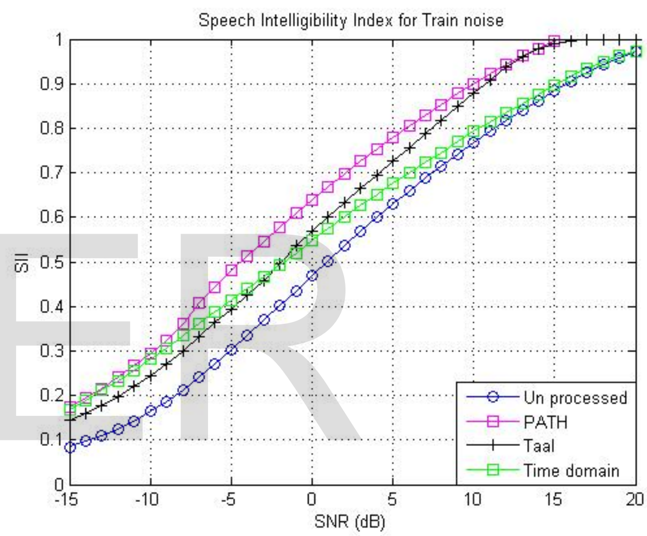

Fig. 6. SII predictions in the presence of train noise.

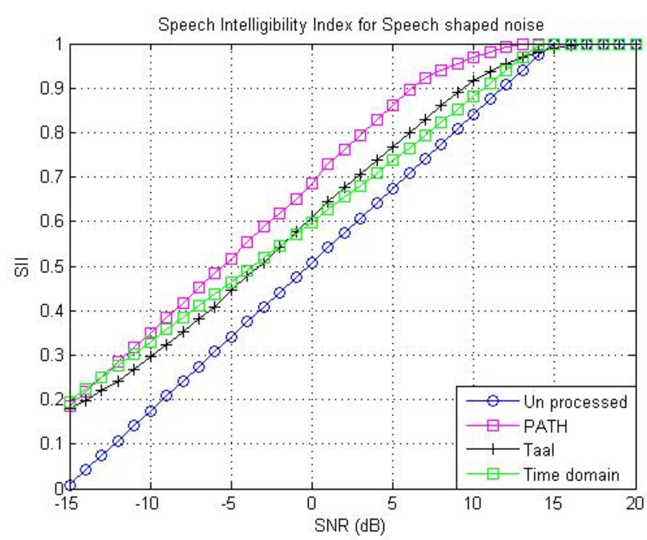

Fig. 7. SII predictions in the presence of speech-shaped noise.

The PESQ score ranges from -0.5 to 4.5 in terms of quality of speech signals. [15], [16], [17] provide accurate and repeatable estimates of speech quality degraded by noise. The quality of enhanced speech signal is measured using PESQ to verify whether increase in intelligibility degrades quality.

Respective PESQ scores of enhanced (using PATH) signal in presence of train and 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 speech and corrupted speech signal (because of near-end noise). PESQ score of the enhanced speech was measured and denoted as $PESQ_{proc}$.

A new value, δ in (10) is used to measure the PESQ improvement achieved by the proposed NELE algorithm.

$$\delta = \frac{(\mathrm{PESQproc} - \mathrm{PESQref})}{\mathrm{PESQref}} 100\% \qquad (10)$$

PESQ improvement obtained in presence of both near-end noises is also listed in the Table 2. The results in Table 2 indicate that the algorithm improves the speech quality for lower values of SNR (-15 dB) than its higher values (15 dB).

Mean opinion score (MOS) in the scale of 1 to 5 was calculated for the SNR in the range -15 to 15 dB. Results tabulated in Table 3 also confirm an increase in speech quality.

Table 2. Comparison of PESQ scores in the presence of train and speech-shaped noise.

SNR (dB)	Train noise			Speech-shaped noise		
	$PESQ_{ref}$	$PESQ_{proc}$	Improvement (%)	$PESQ_{ref}$	$PESQ_{proc}$	Improvement (%)
-15	0.505	2.007	74.83807	0.518	2.065	74.97585
-10	0.626	2.038	69.28361	0.735	2.171	64.7651
-5	0.872	2.186	60.10979	1.072	2.291	51.47125
0	1.4	2.791	49.82079	1.483	2.798	46.20965
5	1.674	3.556	52.92463	1.87	3.634	47.32394
10	2.071	4.241	51.15566	2.225	4.302	46.51442
15	2.427	4.316	43.76738	2.574	4.394	41.2865

Table 3. Comparison of mean opinion scores in the presence of train and speech-shaped noise.

SNR (dB)	Mean opinion scores	
	Train	Speech-shaped
-15	2.4357	2.5392
-10	2.4851	2.7156
-5	2.7173	2.8976
0	3.7371	3.7465
5	4.4935	4.5421
10	4.7715	4.7882
15	4.7909	4.8104

5 Conclusions

In this work, an algorithm is given for improving the intelligibility and quality of speech signals perverted by near-end noise. We presented the experimental results in presence of train and speech-shaped near-end noises. Simulation results were confirmed using an audio editor tool GoldWave and MATLAB. Audible speech samples obtained incorporating masking effects are enhanced. The proposed algorithm has better speech intelligibility, as measured using SII, providing roughly 10 % increase when compared to [1] and 20 % over unprocessed speech signal. The increase is greater at lower SNR where noise dominants the speech signal. From PESQ results, it is revealed that the proposed algorithm increased speech quality also. Results indicate that the NELE leads to a significant increase in intelligibility without compromising on quality.

References

[1] 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.

[2] 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.

[3] Premananda B. S., and Uma B. V., “Speech Enhancement to Overcome the Effect of Near-end Noise in Mobile Phones using Psychoacoustics”, *Fifth IEEE International Conference on Computing, Communication and Networking Technologies* (ICCCNT), Hefei, China, IEEE – 33044, pp. 1-6, July 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 10th Australian International Conference on Speech Science & Technology*, Sydney, pp. 420 - 425, 2004.

ISSN 2229-5518
[11] Gunawan T.S., Khalifa O.O., and Ambikairajah E., "Forward Masking Threshold Estimation using Neural Networks and its Application to Parallel Speech Enhancement", *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] 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.

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

[17] 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.

[18] 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.

[19] 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*, ASSP, Vol. 32, pp. 1109 - 1121, December 1984.

[20] Virag N., "Single Channel Speech Enhancement based on Masking Properties of Human Auditory System", *IEEE Transaction on Speech and Audio Processing*, Vol. 7, pp. 126 - 137, 1999.