I.J. Image, Graphics and Signal Processing, **2015, 6, 29-37**
Published Online May 2015 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijigsp.2015.06.04

Dominant Frequency Enhancement of Speech Signal to Improve Intelligibility and Quality

Premananda B.S.
Department of Telecommunication, R.V. College of Engineering, Bengaluru, India
Email: premanandabs@rvce.edu.in

Uma B.V.
Department of Electronics & Communication, R.V. College of Engineering, Bengaluru, India
Email: umabv@rvce.edu.in

Abstract—In mobile devices, perceived speech signal deteriorates significantly in the presence of near-end noise as the signal arrives directly at the listener's ears in a noisy environment. There is an inherent need to increase the clarity and quality of the received speech signal in noisier environment. It is accomplished by incorporating speech enhancement algorithms at the receiver end. The objective is to improve the intelligibility and quality of the speech signal by dynamically enhancing the speech signal when the near-end noise dominates. This paper proposes a speech enhancement approaches by inculcating the threshold of hearing and auditory masking properties of the human ear. Incorporating the masking properties, the speech samples that are audible can be obtained. In low SNR environments, selective audible samples can be enhanced to improve the clarity of the signal rather than enhancing every loud sample. Intelligibility and quality of the enhanced speech signal are measured using Speech Intelligibility Index and Perceptual Evaluation of Speech Quality. Experimental results connote the intelligibility and quality improvement of the speech signal with the proposed method over the unprocessed far-end speech signal. This approach is efficient in overcoming the deterioration of speech signals in a noisy environment.

Index Terms—Dominant, Near-end noise, Psychoacoustics, Speech enhancement, Speech intelligibility, Speech quality

I. Introduction

Mobile devices are the most popular consumer devices in the present day. For a conversation in a quiet environment, less speech magnitude is required for the speakers to understand each other. However, for instance, if a train passes by, the conversation is severely disturbed. To overcome this effect, we should either wait until the train passes or raise the signal amplitude to produce more speech energy in order to increase the loudness. The external volume control of the mobile phones cannot be used as background noise changes in a dynamic fashion.

As the noise signal cannot be mended upon, a reasonable approach is to manipulate the far-end speech signal based on the energy of near-end noise. Hence, the problem necessitates the need for the development of speech enhancement algorithms to improve the speech perception in adverse listening conditions. The nature of the speech enhancement differs depending on specific applications.

At the receiving end, referred to as "near-end" in the literature, the listener may be in a noisy environment. It 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. Listener experiences fatigue as the quality of the speech signal deteriorates.

The presence of noise masks the speech signal and makes it less intelligent or audible. This effect is called masking and is of two types, one, simultaneous masking and the other temporal masking. In simultaneous masking, a signal is masked by the presence of another signal (predominantly noise). In temporal masking, the signal is masked by noise before and after the high noise occurs. Hence, the speech signal needs to be enhanced considering these situations in the purview of the problem.

The basic idea, of including masking effects in speech signal enhancement, is to remove the non-audible spectral components of the speech signal and the masked signal. Hence, speech enhancement not only involves increasing speech signal for human listening but also for further improvement prior to listening. The objective of signal enhancement is to increase the perceptual aspects of speech such as overall quality, intelligibility, etc. The speech enhancement algorithms should provide superior performance in a broad range of SNRs for both clarity and quality.

The effect of far-end noise on speech signal can be tackled by using traditional noise suppression algorithms like minimum mean-square error (MMSE), short-time spectral amplitude (STSA) estimator [18], spectral subtraction methods [20], etc. The approaches proposed for far-end noise reduction techniques discussed in the literature [18-20] are not suitable in the present context as they focus on mitigating noise at the speaker end rather than at the receiver end. Near-end noise cannot be influenced because the listener is located in a noisy environment, and the noise reaches the ears with hardly

any possibility to suppress [8].

Several approaches to mitigate the near-end noise using speech enhancement are discussed by Bastian S. et al., in [4-6] and Taal C. H. et al., in [7, 8]. Ref. [4] investigates listening enhancement under the constraint that the processed loudspeaker 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 the novel method of linear filtering of speech prior to the deterioration due to near-end noise. He solved constrained optimization problem of [5] using a non-linear approximation of the speech intelligibility that is accurate for lower SNRs.

NELE by Premananda B. S. et al., in [2] increases speech signal when the near-end noise dominates and avoids listener fatigue. In [3] speech samples are given relative weight using threshold of hearing but do not include the masking effect of signals. Approaches in [1-3] do not consider the audible speech samples; rather it enhances both the audible and non-audible spectral components of speech samples, resulting in wastage 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 enhancement method.

The traditional approach used for finding speech quality is to perform subjective tests with a group of listeners. Elucidated directions are provided in ITU-T recommendations P.800/P.830. Perceptual Evaluation of Speech Quality (PESQ) is used to determine the quality of unprocessed and enhanced speech signal. 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.

This paper examines a novel speech enhancement method to improve intelligibility and quality of the speech signal in the near-end side. The frequency domain approach enhances the perceivable components in speech samples, obtained by considering threshold of hearing and simultaneous masking.

The organization of the paper is as follows: section II describes speech enhancement in the frequency domain with an overall block diagram. In section III, loudness computation steps of the samples are explained. In section IV, dominant frequency estimation is described. Implementation details are outlined in section V. In section VI and VII, experimental results and conclusions are discussed.

II. Speech Enhancement In Frequency Domain

The speech enhancement algorithm in the frequency domain (using FFT) is proposed to improve the speech signal perception in a noisy environment. Degradation of intelligibility due to the presence of near-end noise can be reduced by pre-processing the clean far-end speech signal before playing in noisy environments or fed to the mobile speakers. Clean speech signal with far-end noise suppressed (using noise-cancellation techniques) is considered for analysis.

In [2] speech samples are enhanced in the time domain, i.e., both audible and non-audible speech samples are enhanced by comparing the energy of the speech and near-end noise. Redundant power is manifested as non-audible samples are also enhanced. In the proposed method, firstly, only audible speech and noise components are computed. To obtain acoustic samples, a psychoacoustic model has been incorporated. Samples above the threshold of hearing and that are not masked by the neighboring samples are selected. Detailed steps, for choosing audible samples of the signals, are explained in section III. Secondly, only perceivable samples are multiplied by the derived gain to enhance the speech samples. Hence, it requires less multiplication operation and consumes less power.

In a very noisy environment (lower SNR) where noise dominates the speech signals, enhancing just the energy of the audible samples will not be sufficient to improve the speech intelligibility. In these situations, it is desirable to enhance the audible samples that have more energy or are dominant. FFT technique [23] is used to compute dominant frequency components of the speech samples.

Fig. 1 describes the overall block diagram of the proposed approach. The speech samples are enhanced by multiplying a dynamic gain computed by comparing the energy of speech and noise samples. The multiplier is used to enhance the speech samples, degraded due to the presence of near-end noise. The background noise can be recorded using a dummy microphone of mobile phones for analyzing. Energy of the audible signal is termed as signal loudness. The loudness of the near-end noise and speech signal, sampled at 8 kHz are calculated and compared frame-wise. Gain is computed for enhancing the speech samples in a pre-processor block, when the near-end noise dominates the received far-end speech signal. The algorithmic steps realized in pre-processor block of Fig. 1 are illustrated in Fig. 2.

The gain derived for adjacent frames vary drastically. Results in an abrupt increase in speech energy when the gain is multiplied with the adjacent speech samples. To avoid this effect, the gain is smoothened by averaging with pre- and post-frames. The dominant frequency components of the speech samples are extracted and multiplied by the smoothened gain. Procedure, for obtaining dominant samples, is explained in section IV. Implementation steps of the proposed algorithm are discussed in section V.

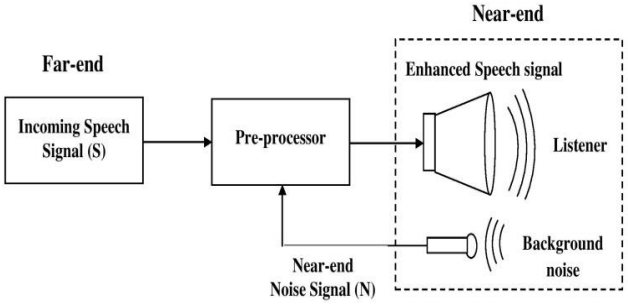


Fig. 1. Proposed block diagram for speech enhancement in the presence of near-end noise.

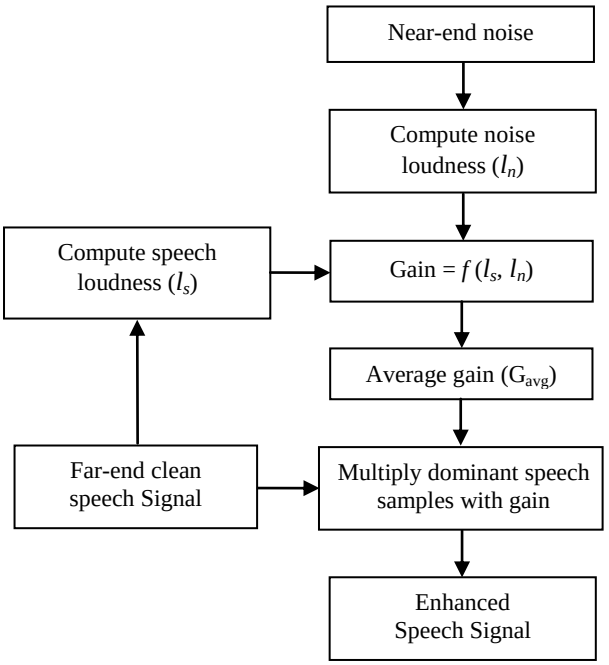


Fig. 2. Flow diagram realized in the pre-processor block.

III. Loudness Computation Of Samples

The psychoacoustic studies have revealed that the reception of all the frequencies by a human ear is not the synonymous [13]. The presence of various sounds in the environment along with the drawbacks of the human auditory system is evidential that non-essential data in the speech samples can be removed.

The two main properties of the human auditory system that constitute the psychoacoustic model are: absolute threshold of hearing (ATH) and auditory masking. They provide a way of finding samples of a signal that are not heard and hence can be removed from the signal.

A. Absolute Threshold Of Hearing

The ATH is the minimal sound level of a pure tone that an average listener with prevalent 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 empirically derived using (6) in [1].

The audio frequency of a human that ranges from 20 to 20 kHz can be split up into critical bandwidths that are non-linear, non-uniform and are dependent on the perceived signal. Samples present in a critical band are indistinguishable for a listener. A uniform measure of frequency based on critical bands is the Bark.

The relation between frequency and Bark [13] is inclined in (1), where LHS represents the frequency in Hz, and the RHS represents the equivalent Bark. Bark bandwidth is smaller at low frequencies and larger at high frequencies. The frequency components that have power levels below the auditory threshold are discarded. The listener will not be able to perceive these frequencies of the signal.

$$f = 1.3\arctan(0.00076f) + 3.5\arctan[(\frac{f}{7500})^2] \quad (1)$$

B. Auditory Masking

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

To use the masking model in the proposed frequency domain speech enhancement algorithm, determine:

- Tone maskers or Tonal components
- Noise maskers or Non – tonal components
- Combined masking effect of tone and noise makers

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

C. Tone Maskers

For a signal frequency component to be a tone, it should be constant for a particular period. It should be a local maximum in the frequency spectrum, indicating that it is higher than the noise component of the signal.

The signal frequency with FFT index m is considered to be a 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 other frequencies in its neighborhood (two).

When such a power level is identified, take the power of 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. The tone may essentially be among the frequency samples.

D. Noise Maskers

If a signal is not a tone, consider all the frequency components that are not in the tone's neighborhood as noise. Humans find it difficult in discriminating signals within 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 of a critical band that do not lie in the vicinity of the tone. Add them into one single entity, keep them at the mean (geometric) location inside the critical band and iterate the process for all critical bands.

Next, remove the maskers that are close to each other to optimize the maskers. Retain the masker possessing power above the ATH and eliminate the remaining maskers because they will not be audible. In the aftermath of the process 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 as it is not perceivable by the human ear.

E. Masking Effect

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

Table 1 lists the conditions of spreading function. Here, Pt is the power spectrum of the tone at j, Pn is the power spectrum of noise at 'j'. SF is the spread function that is modelled as described in Table 1 and 'i' is maskee position. The masking thresholds and masking effect of tone and noise maskers are calculated (dBSPL) using (2) and (3) respectively. Taking into account the threshold of hearing and spectral densities of tone and noise maskers with all masking thresholds, the overall global masking threshold is determined.

Table 1. Conditions of Spreading function.

Spread function, SF (i, j)	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

$$tm(i,j) = Pt(j) - 0.275z(j) + SF(i,j) - 6.025 \quad (2)$$

$$nm(i,j) = Pt(j) - 0.175z(j) + SF(i,j) - 2.025 \quad (3)$$

Here, it is assumed that the masking effects are additive. The masks of all tone and noise maskers are summed if the masker power is above the threshold of hearing. Global masking threshold is the overall threshold obtained along with the spreading function and is called as practical threshold of hearing (PATH) [21].

IV. Dominant Frequency Computations

The dominant or peak frequency [22 – 23] carries the maximum energy among all frequencies in the speech spectrum. Different methods, for determining dominant frequencies are discussed in [23] results indicate that FFT is the best approach for estimate dominant frequencies of the signal. Hence, for selective sample enhancement, FFT-based dominant frequency estimation is adopted.

Steps involved in computing dominant frequencies and enhancements of the speech samples are:

1. Record speech signal for a finite duration with a sampling frequency of 8000 Hz using audio editor tool, Goldwave.
2. Group the samples into frame size (32 ms) of 256 samples.
3. Apply 256-point FFT for each frame to compute power spectral density (PSD).
4. Find peaks in each frame and sort them to pick highest 'n' number of peaks, these peaks correspond to dominant frequency components of that particular frame.
5. Multiply the dominant speech samples by the derived gain.
6. Apply Inverse FFT (IFFT) and reconstruct enhanced speech signal.

If a large number of peaks are included, the spectral shape of the signal can be retained. However, the power of a large number of frequency components has to be increased by multiplying by the gain and hence requires more power. Trade-off has to be made between them depending on the magnitude of near-end noise.

V. Implementation Details

Steps involved in the implementation of the psychoacoustic model to compute the loudness (that are perceivable) of the samples are:

1. Read the speech signal in .wav format using wavread function with a sampling frequency of 8000 Hz.
2. Group the signal into frames of 256 samples each (32 ms) using a window function.
3. Determine the PSD of a frame using 256-point FFT.
4. Locate the tone and noise maskers within the frame and their positions in each critical band.
5. For optimizing the maskers, check if a masker power is lower than the ATH and if found should be eliminated. If two maskers (tone or noise) are inside the critical band, discard the masker with less power.
6. For the frequency component in the selected frame, compute the masking threshold of every mask and sum the masking thresholds to get the overall global masking threshold.

7. Select the samples that are above the global masking threshold. These correspond to the perceivable samples in that frame (PATH) and are stored in a buffer.
8. Compute the loudness of the speech samples that are stored in the buffer using (4). Similarly, calculate loudness of noise samples.
9. Repeat steps 3-8 for all the frames (entire signal).

Steps involved in the frequency domain approach for enhancement of the speech samples when the near-end noise is dominant 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 (4).

$$l(dB) = 10 * \log \frac{\sum_{i=1}^{N} x_i^2}{N} \quad (4)$$

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

Step 3: Derive the Gain

The suitable gain for a couple of speech and noise frames is user specific and depends on multiple constraints.

When the speech signal loudness (l_s) is less than noise loudness (l_n), then compute Δ,

$$\Delta = (l_n - l_s) \quad (5)$$

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 generally set to 3 dB. The gain can then be derived using an empirical equation given in (6).

$$G = 10^{(\frac{\Delta+\Gamma}{10})} \quad (6)$$

Gain calculated using (6) for adjacent frames has random variation. Thus, in order to scale the gain, it is multiplied by a compensation factor E . It can be arbitrarily chosen (< 1) depending on the noisy environment. Hence, the gain can be computed using (7).

$$G = G \cdot E \quad (7)$$

When l_s is sufficiently greater than l_n (by 3dB) then no enhancement is required, and the gain is set to 1.

$$G = 1 \quad (8)$$

Step 4: Smoothening of the Gain (G_{avg})

The computed gain (in step 3) when multiplied with the speech samples results in sudden changes in the output levels. Gain computed using (7) is to be limited to avoid clicks and pops due to erratic changes in the output level (signal bursts) which fatigues the listener's ear.

By making use of (9), the gain obtained in the current frame is averaged with the previous and future frames to make the gain variation smooth. Depending on the delay tolerable by the system, the number of pre and post frames is selected. For example, if gain variation computed per frame are minimal, it suffices just to consider the immediate preceding and succeeding frames.

$$Gavgi = \frac{G_{i-M} + \cdots + G_{i-1} + G_i + G_{i+1} + \cdots + G_{i+M}}{2 \cdot M + 1} \quad (9)$$

where i is the current frame and M is the number of adjacent frames.

Step 5: Multiply averaged gain with the speech samples

When the noise dominates, multiply G_{avg}, the average gain 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 energy [4] of the mobile speaker [6], then limit the minimum and maximum values computed by normalizing the samples.

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 (10).

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

where n and s are recorded noise and speech signal.

For illustration, SNR is varied from -25 to -5 dB.

VI. Experimental Results

Noise and speech signals are recorded with a sampling frequency (Fs) of 8000 Hz for the duration of 4 seconds using GoldWave, an audio editor tool and saved in .wav format for the purpose of analysis. 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.

For multiple speech signals, the proposed algorithm was verified in the presence of different near-end noises. In this scenario, results are indicated only for the train noise. The variance in the recorded near-end (train) noise is adjusted to obtain the desired SNR using (10). Simulation results of the original and the enhanced speech signals are verified using both MATLAB and audio editor tool GoldWave.

The results obtained by including the simultaneous masking for a frame of the recorded signal are discussed here. After obtaining the tone or noise maskers for each frame, the masking thresholds of each masker are computed. Fig. 3 highlights the overall masking threshold for an arbitrary (23^{rd}) frame of the speech signal. It is a

cumulative effect of the spread function multiplied with threshold of hearing. The samples having power below PATH are unperceivable, and only samples above the PATH are audible, extract and store both types of samples in a separate buffer for every frame. Audible samples are the input for the proposed enhancement algorithm. The obtained audible speech sample in an arbitrary frame is as shown in Fig. 4. After the samples that are above the PATH are identified, loudness of both speech and noise samples (that are audible) are calculated using (4). The loudness of speech and noise samples is compared frame-wise, and the gain is calculated using (7) and (8).

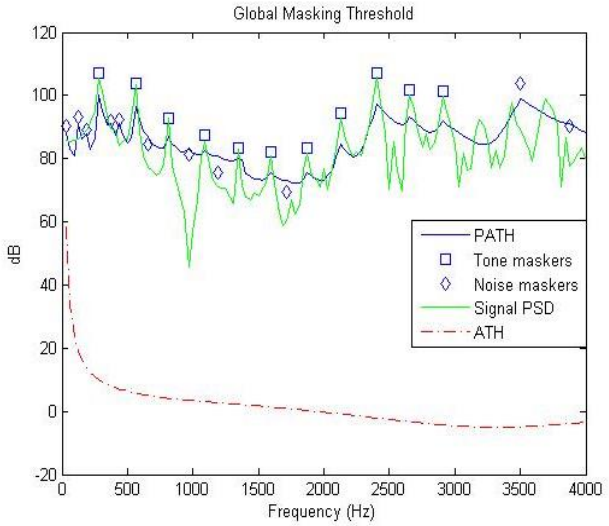


Fig. 3. Overall masking threshold of a frame.

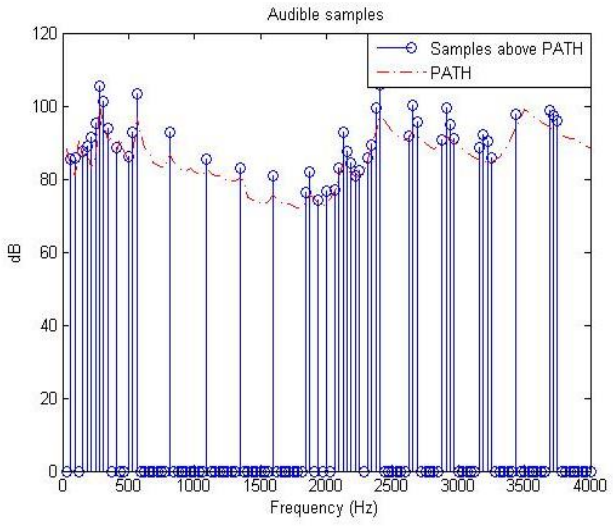


Fig. 4. Audible samples of the signal in a frame.

Optimal/smoothened gain is calculated by averaging the gain using (9) with the gain of the pre and post frames. Original and the smoothed gain are plotted in Fig. 5. Red line represents the original gain, and the blue line the averaged/smooth gain. From the observation of Fig. 5, it is clear that abrupt changes are rectified in smoothed gain.

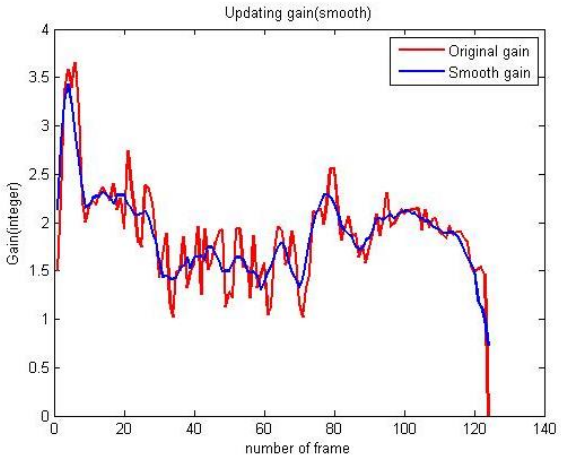


Fig. 5. Variation of gain.

A. Enhancement of the Speech Signal

For enhancing the speech samples the following three approaches are considered:

1. Overall enhancement of speech signals (Time domain [2])
2. Enhancing the speech samples above the PATH
3. Enhancing the dominant frequency components of the speech samples above the PATH

Based on the results it is inferred that the first approach is not practical. It enhances the overall power of the samples including the samples that are not audible hence resulting in the wastage of mobile power/battery. Also, it limits the gain range by unnecessary enhancement of unwanted or non-audible samples. The second method is more efficient than the first since the audible samples above the PATH are enhanced. In the third approach, only the dominant or selective frequencies of the speech samples that have more power are enhanced. Here, 20 (n) dominant frequency components are selected in a frame. If more number of peaks are selected, the spectral shape of the speech signal can be retained but requires more power to increase those n peaks. This method is more useful in lower SNRs than the second method when it is not possible to improve all the audible samples then enhance the audible speech samples having high energy.

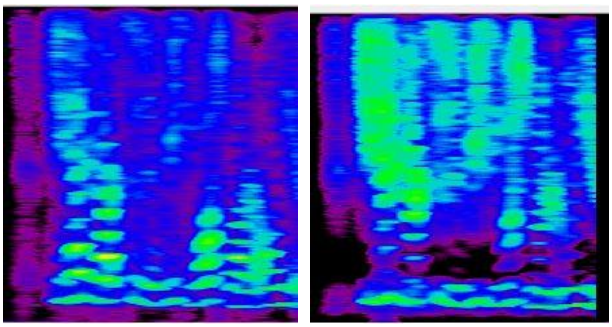
In a noisier environment (lower SNR), it is desirable to improve the speech signal using the last two methods. In the first stage, the smoothened gain is multiplied with all the audible samples and in the second only the specific frequency components of samples above the PATH are enhanced by providing the gain particularly for those frequencies. By this approach, speech samples can be increased to obtain higher intelligibility.

Spectrogram of the input and enhanced speech signals in the presence of train noise is shown in Fig. 6 (a) and (b) respectively. The frequency time elements are increased, resulting in the darker portion. The dark contrasting part in the spectrogram highlights the enhanced energy of the signal.

B. Speech Intelligibility Measurement

Performance of the proposed enhancement algorithms is evaluated in terms of the SII. Intelligibility of the enhanced signal is measured based on the standardized SII procedure as described in [14]. For calculating the SII, steps are outlined in [7, 8]. For unprocessed and processed speech signals, SII is computed in the presence of near-end (train) noise and compared with [2] and [8] for SNR in the range -25 to -5 dB and obtained results are plotted in Fig. 7.

From Fig. 7, it is evident that dominant frequency enhancement of speech samples increases the speech intelligibility in the lower SNRs. Using this approach, few frequency components that have comparatively more energy have to be enhanced. Hence, it requires less battery power to increase these samples. Hence, the proposed method improves the intelligibility of speech signals as predicted by the speech intelligibility index. The quality of the speech signal is also measured to check whether enhancing the speech samples has degraded the quality.



(a) (b)

Fig. 6. Spectrograms for (a) Unprocessed and (b) Enhanced speech signal.

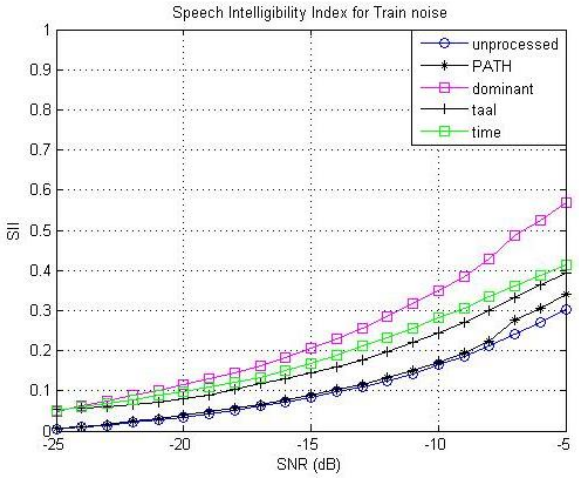


Fig. 7. SII predictions in the presence of train noise.

C. Speech Quality Measurement

The quality of the enhanced speech signal is estimated using PESQ. An accurate and repeated estimation of speech quality perturbed by noise is provided in [9]. PESQ scores of unprocessed and enhanced speech signal (using both PATH and dominant approaches) in the presence of train noise for SNR in the range -15 to 0 dB is as shown in Fig. 8. PESQ scores show an improvement in the dominant method as compared to PATH.

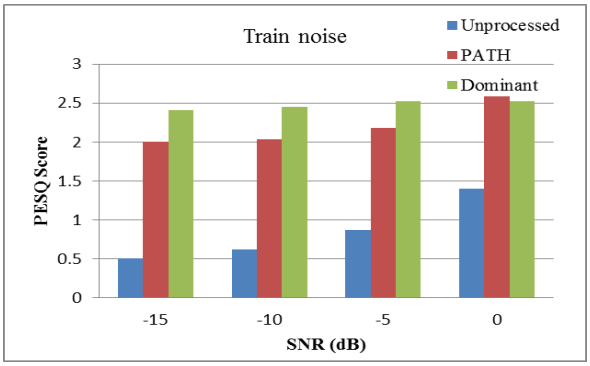


Fig. 8. PESQ scores in the presence of train noise

If $PESQ_{ref}$ is the PESQ score for the clean and corrupted speech signal (because of the near-end noise), then PESQ score of the enhanced speech signal is represented as $PESQ_{proc.}$ A new value, δ, in (11) is used to measure the PESQ improvement achieved by the proposed algorithm.

$$\delta = \frac{(PESQ_{proc} - PESQ_{ref})}{PESQ_{ref}} 100\% \qquad (11)$$

PESQ improvement δ, are listed in Table 2. Results in Table 2 indicate that the algorithm improves the speech quality for lower values of SNR (-15 dB) than its higher values (0 dB). Percentage of increase, δ, in quality of the speech signal using dominant frequency approach, is good in lower SNR (noisier environment) as compared to PATH.

Table 2. PESQ improvement obtained in the presence of train noise.

SNR	Improvement, δ (%)	
	PATH	Dominant
-15	74.838	79.002
-10	69.284	74.491
-5	60.109	65.397
0	44.821	44.554

Mean opinion score (MOS) in the scale of 1 to 5 were also computed for the SNR in the range -15 to 0 dB for unprocessed and enhanced speech signal (using both PATH and dominant approaches). Obtained results are tabulated in Table 3. MOS results also confirm an increase in quality of enhanced speech signal.

Table 3. Comparison of MOS in the presence of train noise.

SNR	MOS		
	Unprocessed	PATH	Dominant
-15	1.0317	2.4357	2.9883
-10	1.0438	2.4851	3.0894
-5	1.0843	2.7173	3.216
0	1.4342	3.6371	3.553

VII. Conclusions

In this work, speech enhancement algorithms are proposed to improve the speech intelligibility and the quantum of quality degraded in the effect of near-end noise. Speech signal corrupted by the presence of train noise is enhanced. Simulation results are verified using MATLAB and an audio editor tool, GoldWave. Audible speech samples obtained by considering masking effects are only enhanced. Selective frequency boosting will be a real solution in situations where the entire samples cannot be increased. Proposed algorithm has better speech intelligibility, as measured using SII, providing roughly 10 % increase when compared to [2] and 20 % over unprocessed speech signal. The improvement is greater at lower SNR where noise dominants the speech signal. From PESQ results, it is revealed that the proposed method increases speech quality in lower SNR as well. Results show that the proposed method leads to a significant rise in the intelligibility without compromising on quality. In the future, one can even incorporate pre- and post-masking properties of the psychoacoustic model to obtain audible speech samples.

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., and Uma B. V., "Speech Enhancement to Overcome the Effect of Near-end Noise in Mobile Phones using Psychoacoustics", 5th *IEEE International Conference on Computing, Comm. and Networking Technologies* (ICCCNT), Hefei, China, IEEE – 33044, DOI:10.1109/ICCCNT.2014.6963017, pp. 1-5, 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] C. H. Taal and R. Heusdens, "A Low-complexity Spectro-temporal based Perceptual Model," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 153–156, 2009.

[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.

[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, Feb. 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-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 (PESQ), 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*, ASSP, Vol. 32, pp. 1109 - 1121, December 1984.

[19] 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.

[20] 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.

[21] Premananda B. S., and Uma B. V., "Speech Enhancement in Presence of Near-end Noise by Incorporating Auditory Masking in Frequency Domain", *International Journal of Scientific & Engineering Research*, IJSER, Vol. 5, Issue 11, ISSN 2229-5518, pp. 621-627, Nov. 2014.

[22] Rastislav Telgarsky, "Dominant Frequency Extraction," *eprint arXiv: 1306.0103*, pp. 1 – 12, June 2013.

[23] Anita Ahmad, Fernando Soares Schlindwein, and Ghulam Andre N., "Comparison of Computation Time for Estimation of Dominant Frequency of Atrial Electrograms: Fast Fourier Transform, Blackman Tukey, Autoregressive and Multiple Signal Classification", in *Journal of Biomedical Science and Engineering*, JBiSE, pp. 843 – 847, September 2010.

[24] 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.

Authors' profiles

Premananda B. S. has completed B.E. in Electronics & Communication Engg., in 2000 from Bangalore University and his M.Tech. in Digital Electronics in 2004 from Visveswaraya Technology University, Belagavi, India and pursuing Ph.D. in speech enhancement for mobile devices at the same University.

Currently, he is working as Assistant Professor in the Dept. of Telecommunication Engg., at R.V. College of Engineering, Bengaluru, India. He has published 20 papers in national, international conferences and journals. His main areas of interests are in the field of digital signal processing, communication, and VLSI.

Uma B. V. has obtained her M.E. in Digital Tech. & Instrumentation in 1995 and Ph.D. from Visveswaraya Technology University, Karnataka in 2009. Currently, she is working as Professor & Associate Dean in the Dept. of Electronics & Communication Engg., at R.V. College of Engineering, Bengaluru, India.

She has published 30 papers in national, international conferences and journals. Her areas of interests are in the field of signal processing, broadband communication, underwater video compression and signal integrity in high-speed VLSI circuit. She is also working in the area of thin film transistor for flexible electronics. Executed project on video compression funded by DRDO, India.

How to cite this paper: Premananda B.S., Uma B.V.,"Dominant Frequency Enhancement of Speech Signal to Improve Intelligibility and Quality", IJIGSP, vol.7, no.6, pp.29-37, 2015.DOI: 10.5815/ijigsp.2015.06.04

