

# A Novel Binary Mask Estimation based on Spectral Subtraction Gain-Induced Distortions for Improved Speech Intelligibility and Quality

N. Saleem<sup>1</sup>, M. Shafi<sup>2</sup>, E. Mustafa<sup>3</sup>, A. Nawaz<sup>4</sup>

<sup>1,3,4</sup>*Institute of Engineering & Technology, Gomal University, D. I. Khan-29050, Pakistan*

<sup>2</sup>*University of Engineering & Technology, Campus-3, Mardan, Pakistan*

<sup>1</sup>*nasirsaleem@gu.edu.pk*

**Abstract**-An alternate binary mask is constructed to improve the speech intelligibility and quality based on constraints of the magnitude spectrum. Motivated by previous studies of speech intelligibility obtained using processing strategies based on ideal binary masks, a new method for deriving a mask is proposed that separates noisy speech into time-frequency channels. The binary mask decisions are made on the basis of speech constraints imposed by the spectral subtraction Gain-induced distortions. All time-frequency channels satisfying the constraints are retained while time-frequency channels violating the constraints are discarded. The speech signals degraded at various signal-to-noise levels (-5dB, 0dB, +5dB) using babble and street makers are processed by the proposed mask and are presented to the normal hearing listeners in experiments measuring speech intelligibility. The results revealed significant improvements in speech intelligibility and quality even at low SNR levels.

**Keywords**-Ideal Binary Mask, Noise Estimation, Speech Intelligibility, Spectral Subtraction

## I. INTRODUCTION

Approaches for enhancing the target speech in noisy situations via binary time-frequency masks primarily take advantage of the sparsity and disjointness of speech spectrums in their short-time-frequency illustrations, constructing a mask that only retains the spectro-temporal regions where the target speech is governing. The enhanced speech is then reproduced after imposing this mask to the noisy signal spectrums. The cues for estimating these masks can be attained either using single-microphone techniques [i-ii] or multi-microphone methods [iii-iv]. A comprehensive literature review on the time-frequency masking can be found in [v]. In this framework, efforts have been made to express a so-called ideal binary mask as the objective of binary mask estimation. This all-or-nothing conclusion is established on basis of a local or a fixed signal-to-noise-ratio (SNR) threshold in the time-frequency channels and the prefix *ideal*

illustrates the prior statistics of the speech and noise spectrums. Methods employing binary masks have been revealed to yield substantial intelligibility improvements in extremely low SNR situations in these ideal sceneries. These positive outcomes have encouraged the researchers to estimate/develop the binary masks and proposed as the goal of computational auditory scene analysis (CASA) [iii], [vi], [vii]. Given this obvious evidence of intelligibility improvement using binary masks, work has been done in the recent past in trying to estimate these masks [ii], [viii] and describing the benefits of such masking [ix-xii], all with an interpretation to using these methodologies in auditory prostheses. It is discussed in [v] that binary masks should be favored over soft-masks for the purpose of complexity decline unless a soft-mask method can expressively improve the speech intelligibility. Still intelligibility improvement in the single-channel speech enhancement algorithms remains an indescribable goal [xiii], [xiv] primarily because of inaccurate estimation of the parameters for the filtering. In this paper a new method for deriving a binary mask is proposed that only retains the spectro-temporal regions where the target speech is dominant. The paper is organized as: the signal model and a brief introduction of the proposed approach are presented in the Section 2. In Section 3, we present the tests conducted to evaluate the proposed mask. We wish to evaluate the potential of the mask to improve the speech intelligibility and quality in adverse listening situations.

## II. THE PROPOSED SPEECH ENHANCEMENT ALGORITHM

The prior studies [xv], [xvi] revealed enormous gains in speech intelligibility when proper constraints are enforced on the gain-induced speech magnitude distortions. On basis of the promising outcomes of past studies, a technique to estimate the binary mask is proposed which depends on magnitude spectrum constraints. Fig. 1 shows the block diagram of the proposed speech enhancement (noise-reduction)

algorithm, consists of the noise estimation and gain computation as pre-processing step and ideal binary

masking as the post-processing (intelligibility enhancement) step, which is described next section.

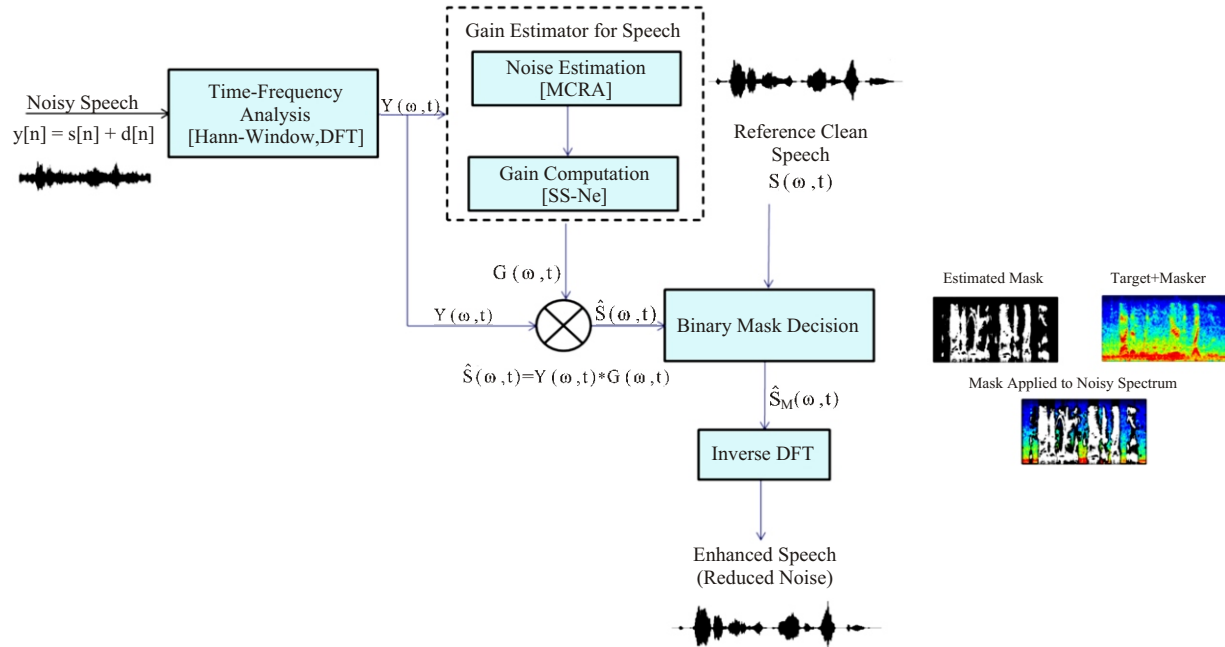


Fig. 1. Block diagram of proposed noise reduction algorithm

#### A. Constraints based Binary Mask Estimation

The noisy speech sentences were processed by conventional spectral subtraction algorithm [xvii]. The spectral subtraction method estimates magnitude spectrum of underlying novel speech by subtracting an estimate of the masker spectrum from noisy speech spectrum in short-time Fourier transform (STFT) domain. The greatest advantage of this technique lies in its simplicity, since all that is required is an estimate of the mean noise power. Let noisy speech, clean speech and noise signal be defined by  $y(n)$ ,  $s(n)$  and  $d(n)$  respectively and expressed as:

$$y[n] = s[n] + d[n] \quad (1)$$

The noisy speech was first segmented into frames with 50% overlap among successive frames. Every frame was Hann windowed and STFT was computed. Let  $Y(\omega, t)$ ,  $S(\omega, t)$  and  $D(\omega, t)$  denotes the noisy, clean and noise signal spectral components at time frame  $t$  and frequency bin  $\omega$ . The estimate of the speech spectrum magnitude  $|\hat{S}(\omega, t)|$  from the noisy observation is achieved by multiplying  $Y(\omega, t)$  with the spectral subtraction gain function  $G(\omega, t)$  as:

$$|\hat{S}(\omega, t)| = |Y(\omega, t)| \times G(\omega, t) \quad (2)$$

The spectral subtraction gain function  $G(\omega, t)$  is computed as;

$$\text{SNR}_{\text{POST}}(\omega, t) = \frac{|Y(\omega, t)|^2}{E\{|D(\omega, t)|^2\}} = \frac{|Y(\omega, t)|^2}{|\lambda(\omega, t)|^2} \quad (3)$$

$$\text{SNR}_{\text{PRIO}}(\omega, t) = \frac{|X(\omega, t)|^2}{E\{|D(\omega, t)|^2\}} \quad (4)$$

The binary mask is constructed by considering only the *a posteriori* SNR information at every frequency bin. The spectral subtraction gain function  $G(\omega, t)$  and *a posteriori* SNR is computed as; From equation 1;

$$|\hat{S}(\omega, t)| = ||Y(\omega, t)| - |D(\omega, t)|| e^{j(f, \omega)} \quad (5)$$

$$|\hat{S}(\omega, t)|^2 \approx |Y(\omega, t)|^2 - \alpha |D(\omega, t)|^2 \quad (6)$$

$$G(\omega, t) = \sqrt{1 - \frac{\alpha |D(\omega, t)|^2}{|Y(\omega, t)|^2}} \quad (7)$$

$$G(\omega, t) = \sqrt{\frac{\gamma(\omega, t) - \alpha}{\gamma(\omega, t)}} \quad (8)$$

$\alpha$  is over-subtraction parameter which is set to 1 in the proposed mask and  $\lambda(\omega, t)$  is the estimate of background noise variance acquired by the noise estimation algorithm proposed in [xviii].  $G(\omega, t)$  is known as gain function in speech enhancement. Note

that  $G(\omega, t)$  is real and, in principle, is always positive, taking values in range of  $0 \leq G(\omega, t) \leq 1$ . Negative values are sometimes obtained owing to inaccurate estimates of noise spectrum. It is important to note that equation 6 is an approximation because of the presence of cross terms [xix]. These cross terms are zero only in the statistical sense assuming that the signals are stationary. Speech, however, is non-stationary but in noise reduction applications, signals are processed on frame-by-frame (20-30 ms windows) basis where we consider the speech to be stationary. We assumed the prior knowledge of the clean speech as this was necessary in order to impose constraints. Therefore; to impose the constraints, the estimated magnitude spectrum  $|\hat{S}(\omega, t)|$  was compared against the true speech spectrum  $|S(j, k)|$  for each time-frequency channel. The time-frequency channels satisfying the constraints ( $|\hat{S}(\omega, t)| < |S(\omega, t)|$ ) were retained whereas time-frequency channels violating the constraints ( $|\hat{S}(\omega, t)| > |S(\omega, t)|$ ) were discarded. The modified magnitude spectrum,  $|\hat{S}_M(\omega, t)|$  was computed as;

$$|\hat{S}_M(\omega, t)| = \begin{cases} |\hat{S}(\omega, t)| & \text{if } |\hat{S}(\omega, t)| \leq |S(\omega, t)| \\ 0 & \text{Elsewhere} \end{cases} \quad (9)$$

The mask in the equation (11) was found to be reasonably effective in improving the speech intelligibility [xv]. However, the mask is ideal magnitude-constraint binary mask as it involves the access to true magnitude spectrum, which is not available in real-time applications. Following the above selection of time-frequency channels, an inverse STFT was applied to the modified speech spectrum  $|\hat{S}_M(j, k)|$  using the phase of noisy speech spectrum and the overlap-and-add method was finally used to synthesized noise-suppressed speech.

### III. PERFORMANCE EVALUATION SCALES AND DISCUSSIONS

A new method for deriving a binary mask is proposed which is discussed in section 2. We evaluate the proposed approach on the basis of two performance measures (1) the speech intelligibility enhancement under the low SNR environments and (2) quality of the synthesized speech in terms of the perceptual speech quality and overall quality.

#### A. Speech Intelligibility Tests

We evaluated our novel method to estimate binary mask with intelligibility listening tests in two phases.

##### 1) Phase I: Effect of frame length on intelligibility

In phase I, the intelligibility listening tests were conducted to evaluate the performance of proposed mask for different frame lengths  $N$ , the parameter used in proposed speech enhancement algorithm. Speech sentences were selected from the IEEE database [xx], a database to facilitate the evaluation of the speech enhancement algorithms. Every sentence is sampled at 8 kHz frequency and has an average duration about 2.5 seconds. Sentences were recorded in silent room and were produced by two male and two female speakers. The noisy stimuli were generated by degrading the clean stimuli with the babble and street noise at -5dB, 0dB and 5dB SNRs using ITU-T recommendation P.56 [xxi]. The noise sources were taken from AURORA database [xxii]. Eight normal-hearing listeners were engaged to conduct listening tests. Out of the eight, four listeners participated in the first phase of the listening test. The tests were conducted in the quiet room and the participants were familiar of the listening tasks during a pre-experiment session. We changed the frame length  $N$  from 128-points to 1024-points and computed the mean intelligibility scores, shown in the Fig. 2, where the highest gain in

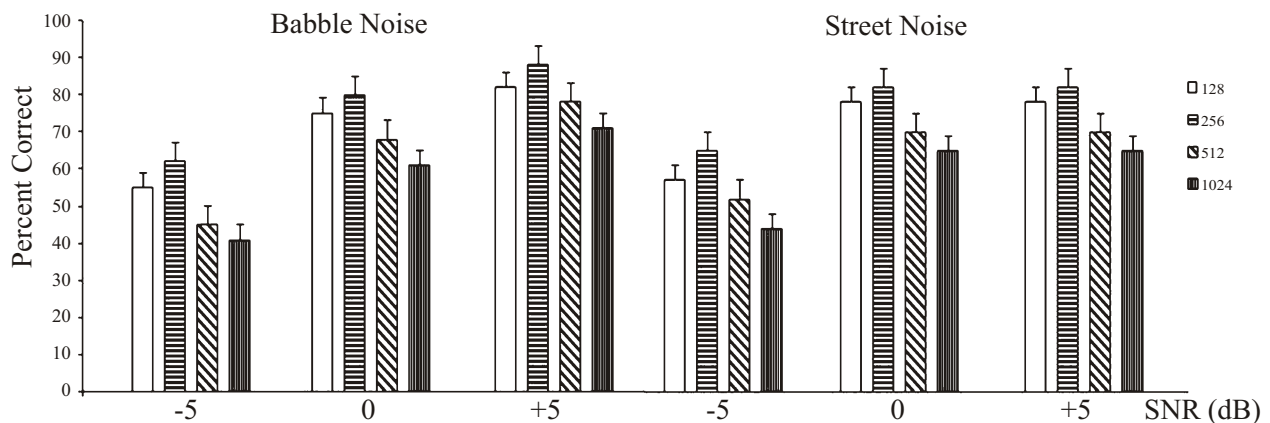


Fig. 2. The mean speech intelligibility scores for the Babble and Street maskers with different Frame lengths  $N$

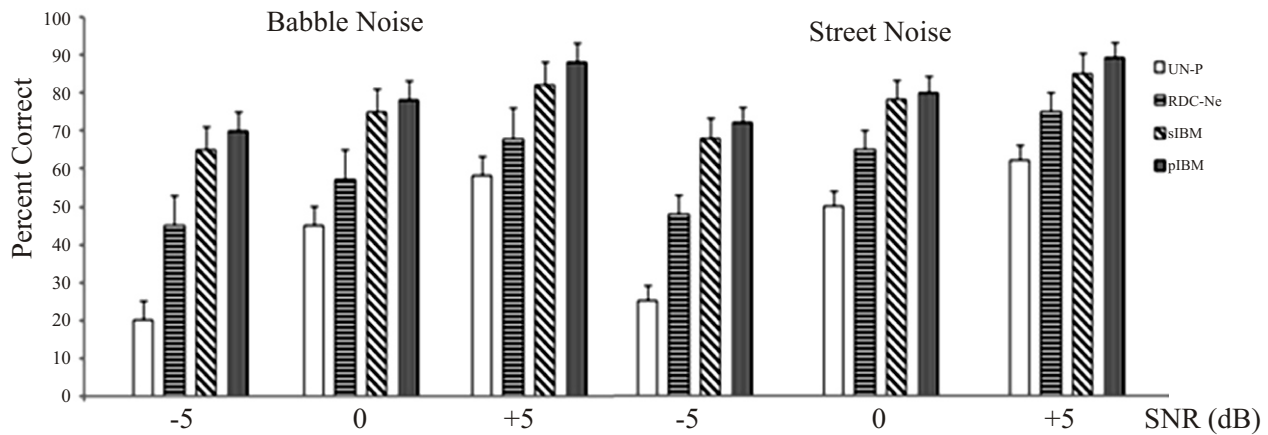


Fig. 3. Mean intelligibility scores as a function of SNR levels for babble and street maskers across the IEEE Sentences for different processing strategies

the speech intelligibility was observed for the frame length  $N=256$ -points in all noisy conditions.

2)Phase II:Comparison with other methods for speech intelligibility

In previous phase, the performance of the proposed binary mask has been evaluated in terms of speech intelligibility for different frame lengths  $N$ . Now we wish to assess the robustness of pIBM in terms of speech intelligibility against different speech enhancement approaches. The listeners participated in the four processing situations including un-processed stimuli (UN-P); speech processed by the SNR based binary mask (sIBM), speech processed by the spectral subtraction using noise estimation (SS-Ne) and speech processed by proposed binary mask (pIBM). In this particular experiment, another 60 IEEE sentences were used to evaluate the performance of the proposed algorithm against other algorithms. The remaining four listeners participated in the particular listening test.

Fig. 3 illustrates the results of the speech intelligibility tests. With pIBM a significant improvement in speech intelligibility was noted compare to that achieved with un-processed speech and SS-Ne. A significant improvement in the speech intelligibility was observed for both the pIBM and sIBM Fig. 3. But question is whether one maskprovides some advantage over the other?, to answer this question; the  $t$ -test based statistical analysis was computed for the mean intelligibility scores obtained with the SS-Ne, sIBM and pIBM respectively. We examine whether improvement in speech intelligibility is statistically significant or not. The results are given in Table I. The computed value of  $t$  determines the acceptance or rejection of the null hypothesis. If the value of  $t$  is found to be greater than a critical value, then the null hypothesis is rejected and concludes that there is statistically significant difference in intelligibility. All the  $t$ -tests in this paper have been carried out at 95% significance level. If  $t < t_{critical}$  and  $p$ -value is smaller than

TABLE I  
THE T-TEST BASED STATISTICAL ANALYSIS

Noise	SNR (dB)	Mean intelligibility scores obtained in % for SS-Ne and pIBM				
		SS-Ne	pIBM	t-test statistical analysis		
				$t_{statistic}$	$t_{critical}$	$P_{value}$
Babble	-5	65.90	75.30	10.03	2.262	$p < 0.00017$
	0	70.90	82.90	15.79	2.262	$p < 0.00017$
	5	75.90	92.90	22.36	2.262	$p < 0.00016$
Street	-5	69.00	78.20	13.29	2.262	$p < 0.00017$
	0	73.92	86.10	17.71	2.262	$p < 0.00014$
	5	78.70	95.70	23.10	2.262	$p < 0.00014$

Noise	SNR (dB)	Mean intelligibility scores obtained in % for sIBM and pIBM				
		sIBM	pIBM	t-test statistical analysis		
				$t_{statistic}$	$t_{critical}$	$P_{value}$
Babble	-5	72.80	75.30	09.30	2.262	$p < 0.00016$
	0	79.10	82.90	10.58	2.262	$p < 0.00017$
	5	90.10	92.90	12.58	2.262	$p < 0.00017$
Street	-5	75.70	78.20	09.47	2.262	$p < 0.00016$
	0	82.21	86.10	10.81	2.262	$p < 0.00017$
	5	92.23	95.70	10.22	2.262	$p < 0.00017$

0.05, then the given results are statistically insignificant. It can be observed from Table I that the  $p$ -values obtained for pIBM against SS-Ne and sIBM in all noisy conditions are smaller than 0.05 and  $t > t_{critical}$

suggesting that the improvement in the speech intelligibility is statistically significant.

TABLE II  
PESQ AND  $C_{OVL}$  RESULTS FOR NOISE REDUCTION ALGORITHM WITH DIFFERENT FRAME LENGTHS  $N$  (WITH 95% CONFIDENCE INTERVAL)

Noise	Frame Lengths (N-points)	PESQ-MOS			Overall quality ( $C_{OVL}$ )		
		-5dB	0dB	5dB	-5dB	0dB	5dB
Babble	128 samples	2.18±0.05	2.53±0.05	2.81±0.05	2.09±0.05	2.41±0.05	2.72±0.05
	256 samples	2.31±0.04	2.71±0.04	3.02±0.05	2.48±0.04	2.94±0.04	3.21±0.05
	512 samples	2.15±0.05	2.49±0.05	2.73±0.06	2.11±0.05	2.38±0.05	2.67±0.06
	1024 samples	2.03±0.04	2.19±0.04	2.31±0.05	1.97±0.04	2.08±0.04	2.22±0.05
Street	128 samples	2.26±0.05	2.58±0.05	2.92±0.05	2.19±0.05	2.49±0.05	2.84±0.05
	256 samples	2.52±0.05	2.82±0.05	3.24±0.04	2.63±0.05	2.87±0.05	3.29±0.04
	512 samples	2.32±0.06	2.54±0.05	2.88±0.05	2.12±0.06	2.41±0.05	2.77±0.05
	1024 samples	2.12±0.05	2.35±0.05	2.58±0.04	2.01±0.05	2.21±0.05	2.51±0.04

TABLE III  
PESQ AND  $C_{OVL}$  RESULTS AND COMPARISON BETWEEN THE UNPROCESSED STIMULI (UN-P), SPECTRAL SUBTRACTION WITH MCRA NOISE ESTIMATION (SS-NE), SNR BASED BINARY MASK (sIBM) AND PROPOSED BINARY MASK (pIBM) IN THE VARIOUS MASKER CONDITIONS (WITH 95% CONFIDENCE INTERVAL)

Noise	Algorithms	PESQ-MOS			Overall quality ( $C_{OVL}$ )		
		-5dB	0dB	5dB	-5dB	0dB	5dB
Babble	UN-P	1.46±0.04	1.68±0.04	2.01±0.04	1.30±0.04	1.52±0.04	2.02±0.04
	SS-Ne	1.51±0.04	1.81±0.04	2.33±0.04	1.40±0.05	1.82±0.05	2.37±0.05
	sIBM	2.21±0.05	2.61±0.05	3.01±0.05	2.43±0.05	2.81±0.05	3.09±0.05
	pIBM	2.31±0.03	2.71±0.05	3.02±0.05	2.48±0.04	2.94±0.04	3.21±0.05
Street	UNP	1.39±0.04	1.77±0.04	2.16±0.05	1.34±0.04	1.71±0.04	2.11±0.05
	SS-Ne	1.50±0.05	1.93±0.06	2.34±0.04	1.63±0.05	2.11±0.05	2.63±0.05
	sIBM	2.41±0.05	2.71±0.04	3.27±0.04	2.58±0.05	2.79±0.05	3.31±0.05
	pIBM	2.52±0.04	2.82±0.05	3.24±0.05	2.63±0.05	2.87±0.05	3.29±0.04

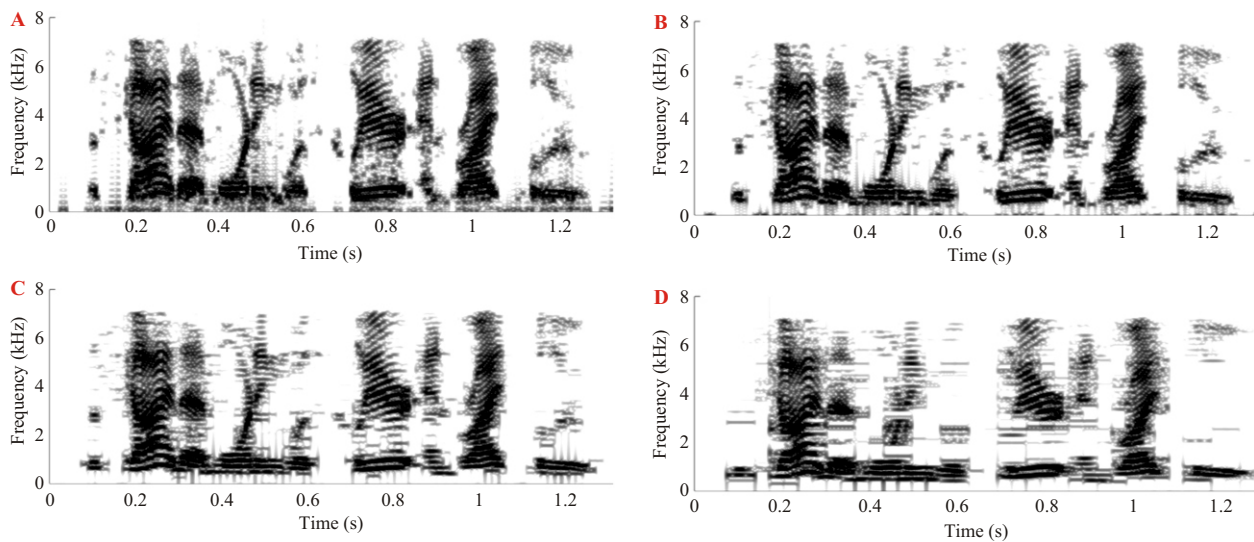


Fig. 4. Spectrograms of utterance by a male speaker from IEEE database. The speech enhanced by proposed binary mask at different frame lengths  $N$  (A)  $N= 128$ -Points, (B)  $N= 256$ -Points, (C)  $N= 512$ -Points and (D)  $N= 1024$ -Points

### B. Objective Speech Quality Evaluation

To evaluate the quality of synthesized speech objectively, we use the quality evaluation measure, the perceptual evaluation of speech quality (PESQ), which is suggested by the ITU-T [xxiii] for speech quality evaluation. We also used the objective speech quality evaluation method recommended by the ITU-T [xxiv], which was referred to as the composite measure ( $C_{OVL}$ ). Again, we changed the frame length  $N$  from 128-points to 1024-points and computed the PESQ and  $C_{OVL}$  scores respectively. The highest gain in the PESQ and  $C_{OVL}$  scores was recorded for 256-points in all noisy conditions shown in Table II. By observing the results of the experiments for the different frame lengths  $N$ , optimum frame length of 256-points was chosen as the best compromise for the proposed speech enhancement algorithm. At longer frame lengths  $N$ , a slurring effect was noticed which is visible in the spectrograms shown in Fig. 4 (D). Since the proposed mask is applied to the spectral subtraction processed spectrums instead of the degraded spectrums, we supposed that the pIBM produces better speech quality. To test our supposition, the PESQ and  $C_{OVL}$  measures were used to evaluate the speech quality of processed speech against SS-Ne and sIBM. Table III compares the PESQ and  $C_{OVL}$  results

for three algorithms and un-processed condition. For pIBM, we observed high PESQ and  $C_{OVL}$  scores in all noisy situations compared to un-processed noisy speech and speech processed by the SS-Ne. Moreover, the PESQ and  $C_{OVL}$  scores for speech processed by the pIBM were found to be consistently higher in all SNR conditions than those processed by the sIBM. The highest gain in PESQ (0.17) was recorded for the street noise at 5dB and lowest gain in PESQ (0.01) was obtained for babble noise at 5dB. Similarly, the highest gain in  $C_{OVL}$  (0.13) was observed for the babble noise at 0dB and lowest gain in  $C_{OVL}$  (0.01) was obtained for babble noise at 5dB. We believe that higher PESQ and  $C_{OVL}$  scores were recorded with pIBM technique can be endorsed to the fact that better noise reduction was attained since spectral subtraction gain function was enforced to noisy stimuli before binary masking. The performance difference is observable which implies that the pIBM approach is efficient in the noise reduction at acceptable speech distortion. The spectrograms for the clean, noisy and processed speech by different techniques are illustrated in the Fig. 5. It is seen that the speech spectra by proposed binary mask are well preserved while residual noise spectra are effectively reduced as shown in Fig. 5(E) and 6(E).

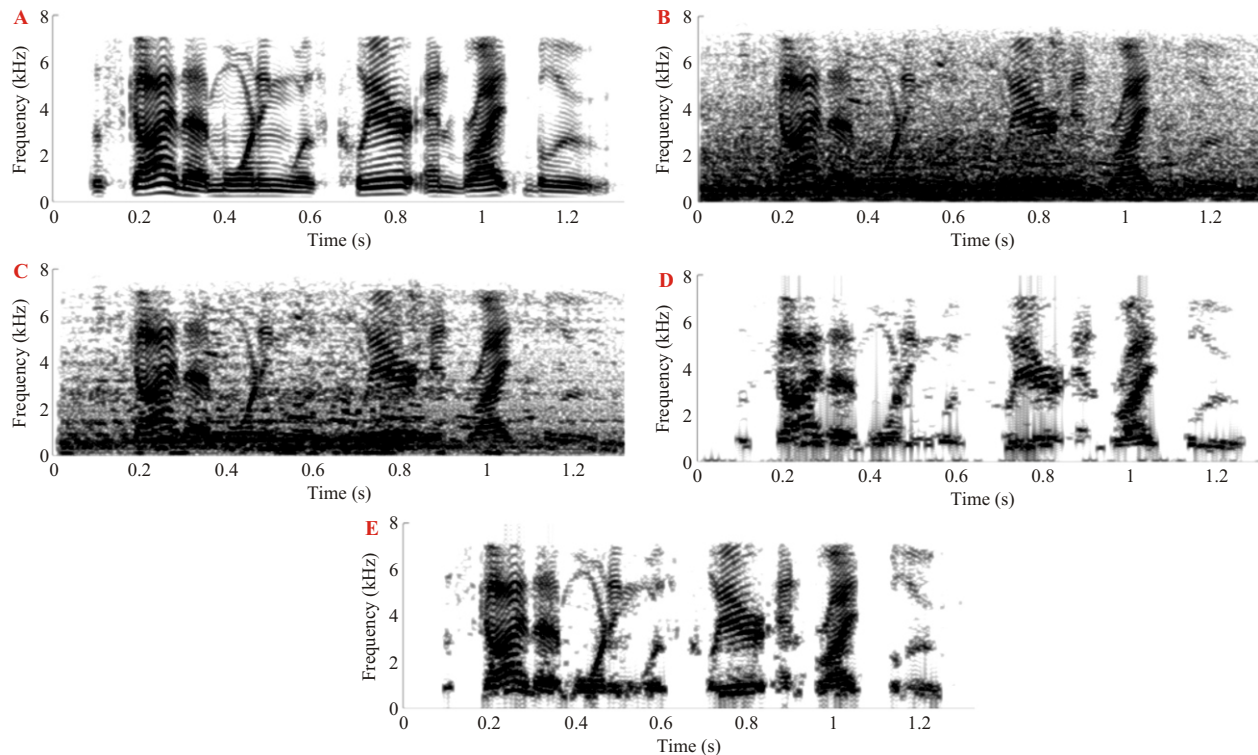


Fig. 5. The spectrograms of (A) clean speech, (B) noisy speech and processed speech by (C) Spectral subtraction with noise estimation (SS-Ne), (D) SNR based BM (sBM) and (E) Proposed binary mask (pBM)

#### IV. CONCLUSIONS

We proposed a novel noise reduction algorithm to reduce background noise and improve speech intelligibility and quality using time-frequency mask which was based on the spectral subtraction gain induced distortions. The binary mask retains all time-frequency channels satisfying the constraints ( $|\hat{S}(j,k)| < |S(j,k)|$ ) while discard all time-frequency channels violating constraints ( $|\hat{S}(j,k)| > |S(j,k)|$ ). The results of listening tests with the normal-hearing listeners showed a remarkable gain in the speech

intelligibility. The statistical analysis suggests that the improvement in the speech intelligibility is statistically significant. Moreover, the speech synthesized using proposed mask revealed high quality even at very low SNR than that attained by speech processed with instantaneous SNR based mask and spectral subtraction with noise estimation. By observing the results of experiments for different frame lengths  $N$ , the optimum frame length of 256-points was chosen as the best compromise for the proposed noise reduction algorithm. We conclude that the proposed mask provides an improved benchmark for future research.

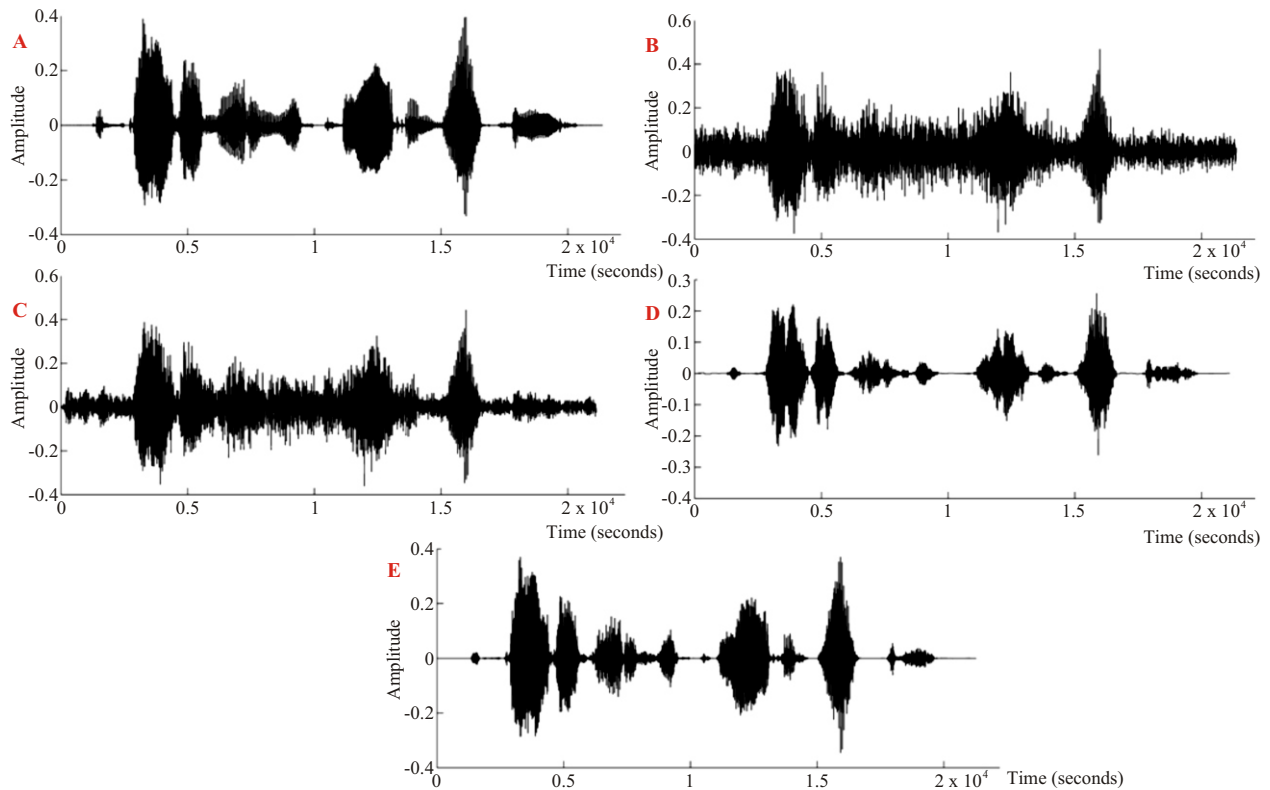


Fig. 6. The Time-domain waveforms of (A) clean speech, (B) noisy speech and processed speech by (C) Spectral subtraction with noise estimation (SS-Ne), (D) SNR based BM (sBM) and (E) Proposed binary mask (pBM)

#### REFERENCES

- [i] S. T. Roweis, "One microphone source separation," in *Advances in Neural Information Processing Systems (NIPS'00)*. Cambridge, MA: MIT Press, 2001, vol. 13, pp. 793-799.
- [ii] Kim, Gibak and Lu, Yang and Hu, Yi and Loizou, Philipos C. "An algorithm that improves speech intelligibility in noise for normal-hearing listeners", *The Journal of the Acoustical Society of America*, 126, 1486-1494 (2009),
- [iii] N. Roman, D. Wang, and G J. Brown, "Speech segregation based on sound localization," *J. Acoust. Soc. Amer.*, vol. 114, no. 4, pp. 2236-2252, Oct. 2003.
- [iv] Jourjine, A.; Rickard, Scott; Yilmaz, O., "Blind separation of disjoint orthogonal signals: demixing N sources from 2 mixtures," *Acoustics, Speech, and Signal Processing, 2000. ICASSP '00. Proceedings. 2000 IEEE International Conference on*, vol.5, no., pp.2985,2988 vol.5, 2000.
- [v] D. Wang, "Time-frequency masking for speech separation and its potential for hearing aid design," *Trends in Amplificat.*, pp. 332-353, Oct. 2008.

[vi] Hu, Guoning; DeLiang Wang, "Speech segregation based on pitch tracking and amplitude modulation," Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the, vol., no., pp.79,82, 2001.

[vii] D. Wang, "On ideal binary mask as the computational goal of auditory scene analysis," in Speech Separation by Humans and Machines, P. Divenyi, Ed. Norwell, MA: Kluwer, 2005, pp. 181-197.

[viii] J. B. Boldt, U. Kjems, M. S. Pedersen, T. Lunner, and D. Wang, "Estimation of the ideal binary mask using directional systems," in Proc.Int. Workshop Acoust. Echo and Noise Control (IWAENC), 2008, pp. 1-4.

[ix] U. Kjems, M. S. Pedersen, J. B. Boldt, T. Lunner, and D. Wang, "Speech intelligibility of ideal binary masked mixtures," in Proc. Eur.Signal Process. Conf. (EUSIPCO), 2010, pp. 1-5.

[x] D. Wang, U. Kjems, M. S. Pedersen, J. B. Boldt, and T. Lunner, "Speech perception of noise with binary gains," J. Acoust. Soc. Amer., vol. 124, no. 4, pp. 2303-2307, 2008

[xi] D. Wang, U. Kjems, M. S. Pedersen, J. B. Boldt, and T. Lunner, "Speech intelligibility in background noise with ideal binary time-frequency masking," J. Acoust. Soc. Amer., vol. 125, no. 4, pp. 2336-2347, 2009

[xii] N. Li and P. C. Loizou, "Factors influencing intelligibility of ideal binary- masked speech: Implications for noise reduction," J. Acoust. Soc. Amer., vol. 123, no. 3, pp. 1673-1682, 2008

[xiii] Y. Hu and P. Loizou, "A comparative intelligibility study of single microphone noise reduction algorithms," J. Acoust. Soc. Amer., vol. 122, no. 3, pp. 1777-1786, 2007

[xiv] H. Luts, K. Eneman, J. Wouters, M. Schulte, M. Vormann, M. Buechler, N. Dillier, R. Houben, W. A. Dreschler, M. Froehlich, H. Puder, G. Grimm, V. Hohmann, A. Leijon, A. Lombard, D. Mauler, and A. Spriet, "Multicenter evaluation of signal enhancement algorithms for hearing aids," J. Acoust. Soc. Amer., vol. 127, no. 3, pp. 1491-1505, 2010

[xv] Gibak Kim; Loizou, P. C., "Why do speech-enhancement algorithms not improve speech intelligibility?," Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on , vol., no., pp.4738,4741, 14-19 March 2010

[xvi] Gibak Kim; Loizou, P.C., "Improving Speech Intelligibility in Noise Using a Binary Mask That Is Based on Magnitude Spectrum Constraints," Signal Processing Letters, IEEE , vol.17, no.12, pp.1010,1013, Dec. 2010

[xvii] Berouti, M., Schwartz, M., and Makhoul, J., Enhancement of speech corrupted by acoustic noise. Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, 1979, pp. 208-211

[xviii] Cohen, I. Noise estimation by minima controlled recursive averaging for robust speech enhancement. IEEE Signal Processing Letters, 2002, 9(1), pp. 12-15

[xix] P. Loizou, Speech Enhancement: Theory and Practice. Boca Raton, FL: CRC, 2007

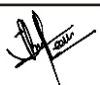
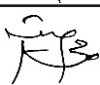
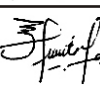
[xx] IEEE Subcommittee, IEEE recommended practice for speech quality measurements. IEEE Trans. Audio, Electroacoustic, 1969, pp. 225-246.

[xxi] Objective measurement of active speech level. ITU-T Recommendation P.56, 1993

[xxii] Hirsch, H., Pearce, D., The aurora experimental framework for the performance evaluation of speech recognition systems under noisy conditions. In: ISCA ITRW ASR2000, Paris, France, 2000.

[xxiii] 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).

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

<b>Authorship and Contribution Declaration</b>			
	<b>Author-s Full Name</b>	<b>Contribution to Paper</b>	
1	Engr. Nasir Saleem (Assistant Professor)	Proposed topic, basic study Design, methodology and manuscript writing	
2	Dr. Muhammad Shafi (Associate Professor)	Statistical analysis, interpretation of results and manuscript writing	
3	Engr. Ehtasham Mustafa (Lecturer)	Data Collection, Literature review	
4	Engr. Aamir Nawaz (Lecturer)	Literature review & Referencing and quality insurer	