Intelligent Systems with Applications 16 (2022) 200120

Contents lists available at ScienceDirect

Intelligent Systems with Applications

journal homepage: www.journals.elsevier.com/intelligent-systems-with-applications

Robust intelligibility and quality evaluation of combined temporal and spectral processing for hearing impaired

Hemangi Shinde [a,b,*], Vibha Vyas [a]

[a] Department of Electronics and Telecommunication Engineering, College of Engineering Pune, Pune, Maharashtra, India
[b] Department of Electronics and Telecommunication Engineering, AISSMS Institute of Information Technology, Pune, Maharashtra, India

ARTICLE INFO

Keywords:
Speech enhancement
Hearing impaired
Temporal processing
Spectral processing
Hearing aid speech quality index
Hearing aid speech perception index

ABSTRACT

Hearing-impaired people face numerous challenges with speech perception in the presence of interfering background noise. To suppress interfering background noise, the common approach widely used is speech enhancement. Inspired by the improved results of combined temporal and spectral processing in speech enhancement, this research study proposes temporal enhancement combined with two different spectral enhancement methods, with a novel approach of soft masking using priori and posterior signal to noise ratio uncertainty. The present study investigates quality and intelligibility objective evaluations, namely, hearing aid speech quality index and hearing aid speech perception index, of spectral and a combination of temporal-spectral speech enhancement methods for typical pattern of hearing loss characterized by six audiograms. For evaluation, clean speech files from the NOIZEUS database are mixed with four local noises, namely, cafeteria, traffic, station, and train at -5, -3, 0,3,5, and 10 dB SNRs. These local noises are quite common, which are encountered by people in their day-to-day lives. In most of the testing conditions, the new combined temporal spectral enhancement shows improved results in comparison with the purely spectral processing methods.

1. Introduction

Hearing-impaired people find it quite inconvenient to carry out their daily verbal communication chores. In light of the report by the World Health Organization (WHO), published in 2020 (*Deafness and Hearing Loss*, n.d.), around 466 million people have hearing loss, out of which 432 million (93%) are adults. Due to increased exposure to various kinds of noises and aging, the presence of hearing loss may drastically increase in society. To overcome this issue of hearing loss, a majority of hearing-impaired people require customized hearing aids. Hearing aids use speech enhancement algorithms to improve the audibility of speech, and there is much scope and further requirements for suppressing background noise. Hence, the study results in improved speech in a noisy environment, and proves to be useful especially for hearing impaired.

Several single-channel speech enhancement algorithms have been proposed to overcome this challenge. The speech enhancement algorithms are basically categorized into two types, namely, temporal processing (Ananthapadmanabha & Yegnanarayana, 1979; Yegnanarayana et al., 1999; Yegnanarayana & Satyanarayana Murthy, 2000) and spectral processing algorithms (Berouti et al., 1979; Boll, 1979; Cohen, 2005; Y. Ephraim & Malah, 1985; Ephraim & Malah, 1984). In temporal processing algorithms, background noise is not suppressed as much as it is with the spectral processing, in comparison, however, the features of excitation source such as glottal closure instants in unenhanced speech signal, high signal to noise ratio (SNR) regions, are enhanced. In spectral processing, background noise is modeled and subtracted from the unenhanced signal to get an improved speech.

Several deep learning-based speech enhancement methods have been proposed in the relevant literature by some researchers (Zheng and Zhang, 2019; Tan & Wang, 2020; Lee and Kang, 2019). Different spectral clustering algorithms have been proposed recently (Li, Nie, Chang, Yang, et al., 2018a; Li, Nie, Chang, Yang, et al., 2018b; Li et al., 2019). However, computation becomes more complex while implementing neural network-based methods.

There is one more approach of combining temporal and spectral processing algorithms, which display improved performance in comparison with the individual methods on hearing impaired (HI) listeners (Shinde et al., 2019). The authors evaluated temporal processing (TP) (Krishnamoorthy & Prasanna, 2011; Yegnanarayana et al., 1999),

* Corresponding author at: Department of Electronics and Telecommunication Engineering, AISSMS, Institute of Information Technology, Kennedy Road, Near RTO, Pune 411001, Maharashtra, India.
E-mail addresses: shindehd11.extc@coep.ac.in (H. Shinde), vsv.extc@coep.ac.in (V. Vyas).

https://doi.org/10.1016/j.iswa.2022.200120
Received 2 April 2021; Received in revised form 11 July 2022; Accepted 29 August 2022
Available online 28 September 2022

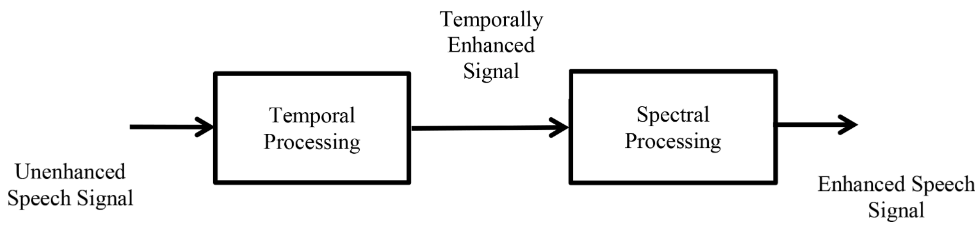

Fig. 1. Basic Block Diagram of the System.

Table 1
Audiograms of six different types of sensorineural hearing loss used for evaluation.

Frequency (kHz)	0.25	0.5	1	2	4	6	Audiogram Number	Audiogram Type
Hearing Loss in dB	70	65	60	50	10	10	1	Reverse sloping loss
	40	40	50	60	65	65	2	Moderately sloping high-frequency loss (Lai et al., 2014)
	10	10	15	65	75	90	3	Steeply sloping high-frequency loss with a normal low-frequency threshold
	0	15	30	60	80	85	4	Mild low-frequency hearing loss with high-frequency loss (Lai et al., 2014)
	14	14	11	14	24	39	5	Mild to moderate sloping high- frequency loss (Lai et al., 2014)
	24	24	25	31	46	60	6	Mild to moderate sloping high- frequency loss (Lai et al., 2014)

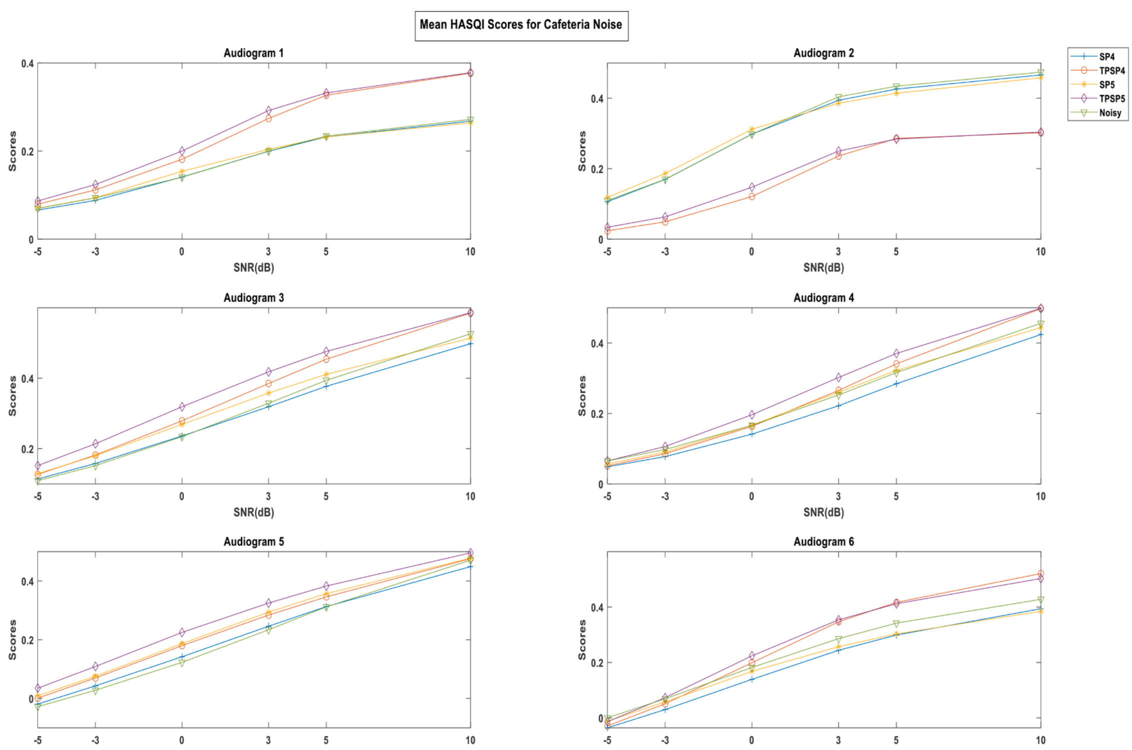

Fig. 2. Mean HASQI scores for cafeteria noise in six audiograms.

spectral processing (SP) (Berouti et al., 1979; Ephraim & Malah, 1984) and combined temporal spectral processing (TPSP) using a subjective evaluation mean opinion score (MOS) method and obtained improvement in quality with combined temporal-MMSE processing compared to the individual algorithms.

The present study focuses on a novel combination of temporal processing and maximum-a-posterior (MAP) estimation of the magnitude squared spectrum (Lu & Loizou, 2011) algorithms. In (Lu & Loizou, 2011), the gain function obtained was similar to the ideal binary mask (IBM) gain function, and the soft masking methods, which were derived by modeling local instantaneous SNR, resulting in better quality in comparison with the standard Minimum Mean Square Error (MMSE) method. In IBM algorithms, it was observed that the quality as well as the intelligibility of the enhanced speech, improved (Li & Loizou, 2008; Wang et al., 2009).

Further, in the temporal processing algorithm proposed in (Krishnamoorthy & Prasanna, 2011; Yegnanarayana et al., 1999), the significant excitation features, like Glottal Closure Instants (GCIs), were computed using the linear prediction (LP) residual signal. However, the accuracy of the computing GCIs was relatively low. In (Deepak & Prasanna, 2015), the proposed Zero Band Filter (ZBF) was observed to offer better results for computing the significant excitation regions.

With this motivation, the temporal processing algorithm (TPA) using ZBF to locate significant excitation regions in the unenhanced speech signal and to compute fine weight function (Krishnamoorthy & Prasanna, 2011; Deepak & Prasanna, 2016) was selected. Additionally, TPA was combined with two different MAP estimators of the magnitude squared spectrum; namely, soft masking using posterior SNR

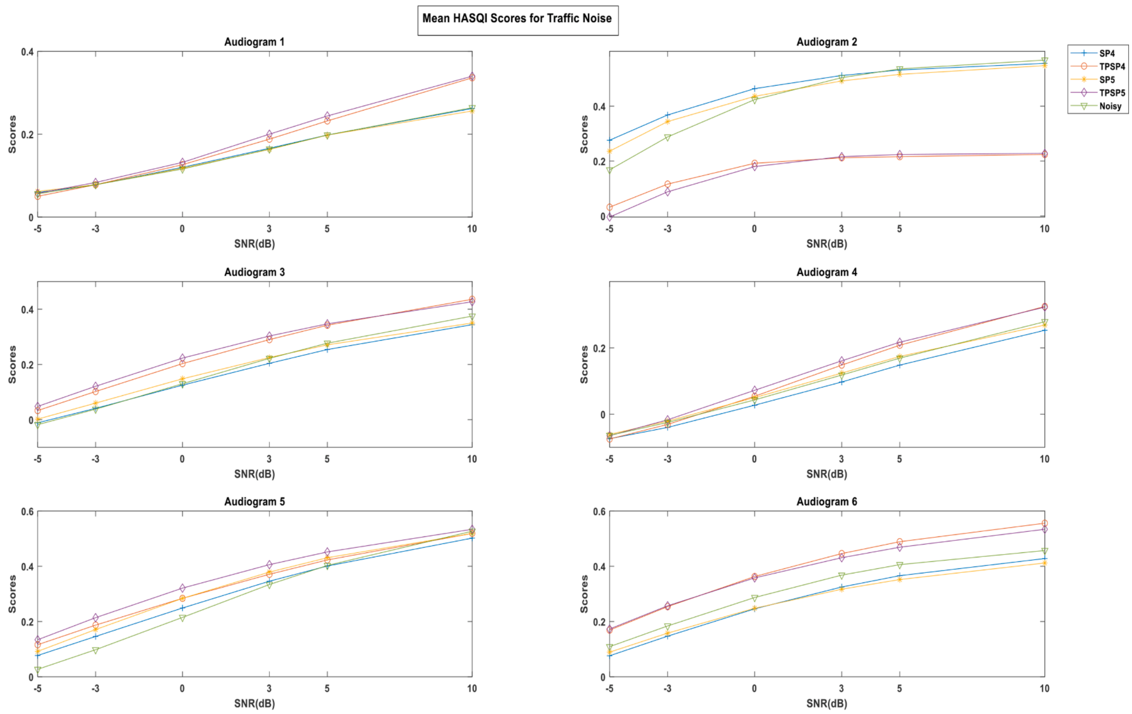

Fig. 3. Mean HASQI scores for traffic noise in six audiograms.

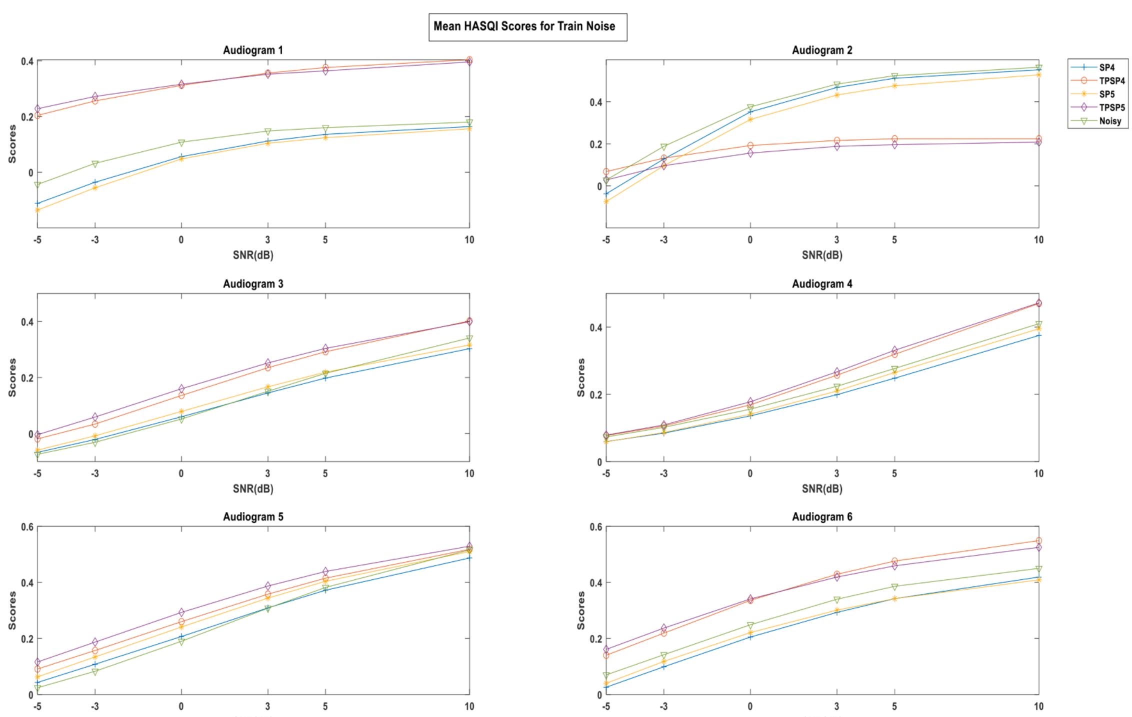

Fig. 4. Mean HASQI scores for train noise in six audiograms.

uncertainty on magnitude squared spectrum (SP4) and soft masking using priori SNR uncertainty on magnitude squared spectrum (SP5). The new combination of temporal and spectral processing algorithm (TPSPA) was evaluated objectively in terms of quality as well as intelligibility for hearing losses characterized by six different audiograms, using hearing aid speech quality index (HASQI) and hearing aid speech perception index (HASPI). For evaluation, the clean speech files used were from the NOIZEUS speech database, and these files were mixed with locally recorded noises, namely; cafeteria, traffic, station, and train at -5,-3, 0,3,5, and 10 dB SNRs.

The significant contribution of this research study involves:

1 Instead of using only spectral processing method for HI, which is seen in most of the speech enhancement methods, temporal processing is combined with spectral processing and a unique new novel combination of temporal spectral processing is proposed for HI.

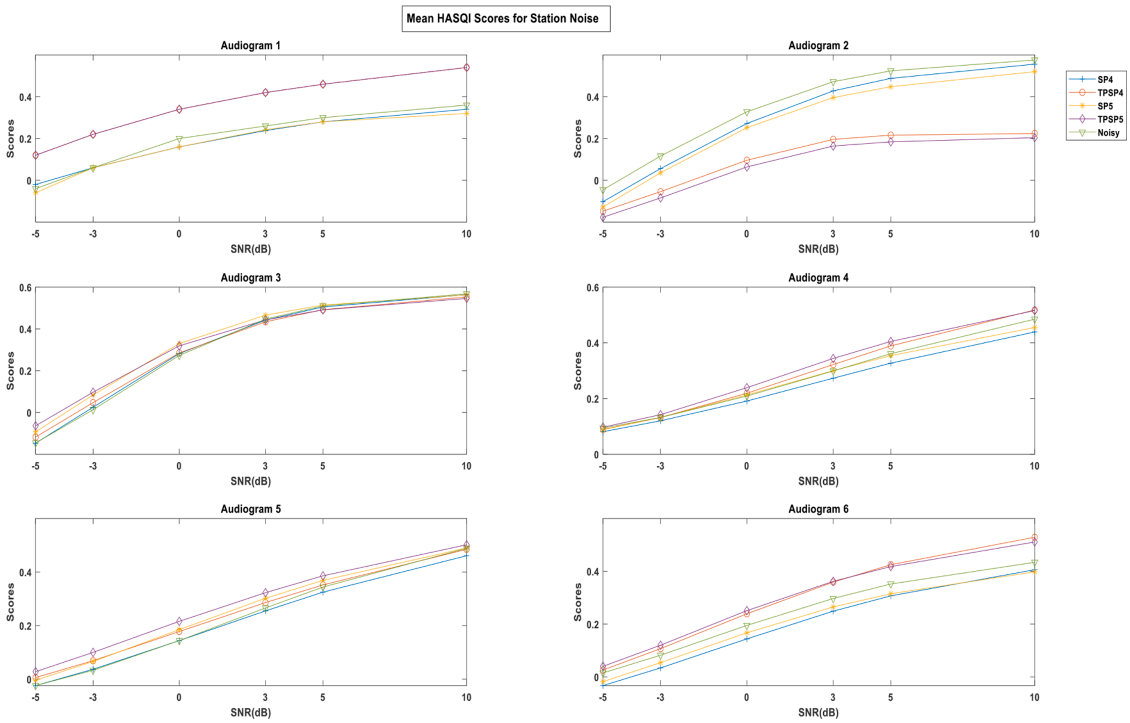

Fig. 5. Mean HASQI scores for station noise in six audiograms.

Table 2
Percentage improvement in HASQI scores for cafeteria noise with reference to the noisy signal and performance of the algorithms ('*' represents significant improvement, '#' represents equal performance, 'LP' represents low performance with reference to the noisy signal).

Cafeteria noise Audiogram	SNR (dB)	% Improvement				Performance			
		SP4	SP5	TPSP4	TPSP5	SP4	SP5	TPSP4	TPSP5
A1	-5	0	1.39	13.74	23.84	LP	*	*	*
	-3	0	0.89	19.17	31.24	LP	#	*	*
	0	0	9.05	29.07	42.89	LP	*	*	*
	3	0	0.83	36.14	44.78	LP	#	*	*
	5	0	0	40.03	42.23	LP	LP	*	*
	10	0	0	38.79	39.23	LP	LP	*	*
A2	-5	0	8.05	0	0	LP	*	LP	LP
	-3	0	9.64	0	0	LP	*	LP	LP
	0	0.14	4.64	0	0	LP	*	LP	LP
	3	0	0	0	0	LP	LP	LP	LP
	5	0	0	0	0	LP	LP	LP	LP
	10	0	0	0	0	LP	LP	LP	LP
A3	-5	4.87	19	16.86	39.04	*	*	*	*
	-3	3.78	18.2	19.43	40.16	*	*	*	*
	0	0.85	15.3	19.34	36.5	#	*	*	*
	3	0	8.51	16.76	26.8	LP	*	*	*
	5	0	4.38	15.07	20.73	LP	*	*	*
	10	0	0	11.17	11.33	LP	LP	*	*
A4	-5	0	0	0	0	LP	LP	LP	LP
	-3	0	0	0	8.87	LP	LP	LP	*
	0	0	0	0	18.09	LP	LP	LP	*
	3	0	0	5.14	19.62	LP	LP	*	*
	5	0	0	7.89	17.22	LP	LP	*	*
	10	0	0	9.1	9.34	LP	LP	*	*
A5	-5	5.35	21.5	16.73	36.44	*	*	*	*
	-3	6.76	21.2	18.4	35.5	*	*	*	*
	0	5.85	19.5	17.56	31.24	*	*	*	*
	3	2.89	13.9	11.63	20.93	*	*	*	*
	5	0.15	8.79	6.64	13.8	#	*	*	*
	10	0	0.92	0.63	3.72	LP	#	#	*
A6	-5	0	0	0	0	LP	LP	LP	LP
	-3	0	0	0	1.18	LP	LP	LP	#
	0	0	0	4.74	11.17	LP	LP	*	*
	3	0	0	12.47	13.95	LP	LP	*	*
	5	0	0	13.89	12.97	LP	LP	*	*
	10	0	0	14.87	11.95	LP	LP	*	*

Table 3
Percentage improvement in HASQI scores for traffic noise with reference to the noisy signal and performance of the algorithms ('*' represents significant improvement, '#' represents equal performance, 'LP' represents low performance with reference to the noisy signal).

Traffic noise Audiogram	SNR (dB)	% Improvement SP4	SP5	TPSP4	TPSP5	Performance SP4	SP5	TPSP4	TPSP5
A1	-5	4.42	10.46	0	1.15	*	*	LP	#
	-3	0.72	2.98	0	7.67	#	*	LP	*
	0	3.17	2.87	8.58	13.7	*	*	*	*
	3	0.46	0	14.3	22.4	#	LP	*	*
	5	1.2	0.14	17.1	23.9	#	#	*	*
	10	0	0	27.3	29.6	LP	*	*	*
A2	-5	19.7	12.88	0	0	*	*	LP	LP
	-3	11	7.58	0	0	*	*	LP	LP
	0	4.56	1.25	0	0	*	#	LP	LP
	3	1.06	0	0	0	#	LP	LP	LP
	5	0	0	0	0	LP	LP	LP	LP
	10	0	0	0	0	LP	LP	LP	LP
A3	-5	3.5	10.7	27.8	36	*	*	*	*
	-3	1.21	9	26.5	34.4	#	*	*	*
	0	0	5.36	21.8	28	LP	*	*	*
	3	0	0.67	16	19	LP	#	*	*
	5	0	0	13.5	14.5	LP	LP	*	*
	10	0	0	10.7	9.1	LP	LP	*	*
A4	-5	0	2.4	0	0	LP	*	LP	LP
	-3	0	2.3	0	4.24	LP	*	LP	*
	0	0	2.8	4.52	11.8	LP	*	*	*
	3	0	1.7	9.35	13.4	LP	*	*	*
	5	0	1.3	10.5	13	LP	*	*	*
	10	0	0	9.68	9.2	LP	LP	*	*
A5	-5	21.7	28.6	39	46.8	*	*	*	*
	-3	16.2	24.5	30	39.2	*	*	*	*
	0	8	16.47	16.4	25.3	*	*	*	*
	3	2.25	8.54	6.98	13.4	*	*	*	*
	5	0	4.65	3.26	8.08	LP	*	*	*
	10	0	0	0	1.06	LP	LP	LP	#
A6	-5	0	0	19.4	20.8	LP	LP	*	*
	-3	0	0	18.4	19	LP	LP	*	*
	0	0	0	15.5	14.6	LP	LP	*	*
	3	0	0	13.6	11	LP	LP	*	*
	5	0	0	13.7	10.3	LP	LP	*	*
	10	0	0	15.1	11.8	LP	LP	*	*

2 The combined TPSPA is evaluated objectively in terms of HASQI and HASPI for six different types of audiograms, namely; 1) reverse sloping loss, 2) moderately sloping high-frequency loss, 3) steeply sloping high-frequency loss with normal low-frequency threshold, 4) mild low-frequency hearing loss with high-frequency loss, 5) and 6) mild to moderate sloping high-frequency loss (2 cases) for noisy environments.
3 Local noisy environments, namely; cafeteria, station, traffic, and train at -5 dB, -3 dB, 0 dB, 3 dB, 5dB, and 10dB SNRs are considered for evaluation.

The rest of the paper is organized as follows: Section 2 explains the methodology, Section 3 describes the objective evaluation of the proposed algorithms, results, discussion, Section 4 describes the subjective quality evaluation which is followed by conclusion in Section 5. Finally, a list of references is presented.

2. Methodology

An unenhanced speech signal is passed through a series combination of the temporal processing algorithm and spectral processing algorithm as shown in Fig. 1.

In temporal processing (TP), the gross level processing (eq.2, eq.3, eq.7) (Krishnamoorthy & Prasanna, 2009) is obtained by temporally adding the sum of ten large amplitude peaks of discrete Fourier Transform (DFT) spectrum, Hilbert envelope (HE) of LP residual signal and the modulation spectrum. Significant excitation regions occur at glottal closure instants (GCIs). These GCIs of the unenhanced speech signal are found using a zero-band filter (Deepak & Prasanna, 2015).The fine level processing is carried out by enhancing the region around the calculated GCIs using the fine weight function (eq.18) (Deepak & Prasanna, 2016). The final weight function is obtained by multiplying the gross weight function with the fine weight function. The final weight is multiplied by LP residual signal, and LP synthesis is implemented to get a temporally enhanced signal. The temporally enhanced signal does not suppress background noise; however, it enhances speech-specific features like regions of significant excitation, which have high SNR. Hence, to suppress the background noise, a temporally enhanced speech signal is passed through two different spectral processing methods; namely, soft masking using posterior SNR uncertainty on magnitude squared spectrum (SP4) and soft masking using priori SNR uncertainty on magnitude squared spectrum (SP5) (eq.32, eq,29) (Lu & Loizou, 2011). To get the enhanced speech signal, the basic procedure followed is demonstrated in Algorithm 1.

Algorithm 1

1) Estimate the vocal tract-based features by computing the sum of the ten most prominent peaks of magnitude spectrum of DFT.
2) Estimate the excitation source features by computing HE of LP residual of the unenhanced speech signal.
3) Calculate the modulation spectrum.
4) Sum the parameters obtained in steps 1, 2, and 3 and compute a gross weight function.
5) Estimate GCIs using ZBF.
6) Calculate the fine weight function.
7) Compute the final weight function.

Table 4
Percentage improvement in HASQI scores for station noise with reference to the noisy signal and performance of the algorithms ('*' represents significant improvement, '#' represents equal performance, 'LP' represents no improvement with reference to the noisy signal).

Station noise Audiogram	SNR (dB)	% Improvement SP4	SP5	TPSP4	TPSP5	Performance SP4	SP5	TPSP4	TPSP5
A1	-5	0	0	36.76	40	LP	LP	*	*
	-3	0	0	34.6	35.3	LP	LP	*	*
	0	0	0	31.6	31.7	LP	LP	*	*
	3	0	0	33.9	33.08	LP	LP	*	*
	5	0	0	34.56	32.98	LP	LP	*	*
	10	0	0	35.89	35.97	LP	LP	*	*
A2	-5	0	0	0	0	LP	LP	LP	LP
	-3	0	0	0	0	LP	LP	LP	LP
	0	0	0	0	0	LP	LP	LP	LP
	3	0	0	0	0	LP	LP	LP	LP
	5	0	0	0	0	LP	LP	LP	LP
	10	0	0	0	0	LP	LP	LP	LP
A3	-5	0	8.35	19.32	29.35	LP	*	*	*
	-3	0	10.5	18.73	29.77	LP	*	*	*
	0	0	7.4	14.6	26	LP	*	*	*
	3	0	3	10.4	18.96	LP	*	*	*
	5	0	0	7.73	13.75	LP	LP	*	*
	10	0	0	6.7	8	LP	LP	*	*
A4	-5	0	0	0	3.23	LP	LP	LP	*
	-3	0	0	0	7.45	LP	LP	LP	*
	0	0	1.7	4.67	14.5	LP	#	*	*
	3	0	0.4	7.6	14.9	LP	#	*	*
	5	0	0	7.55	12.09	LP	LP	*	*
	10	0	0	6.76	6.4	LP	LP	*	*
A5	-5	0.41	11.4	16.9	29.88	#	*	*	*
	-3	1.76	14.2	15.87	28.98	*	*	*	*
	0	0.11	12	10.04	20.93	#	*	*	*
	3	0	7.5	4.07	12.18	LP	*	*	*
	5	0	4.55	1.43	7.6	LP	*	#	*
	10	0	0.07	0	1.8	LP	#	LP	#
A6	-5	0	0	5.2	11.82	LP	LP	*	*
	-3	0	0	8.7	13.37	LP	LP	*	*
	0	0	0	10.96	14.02	LP	LP	*	*
	3	0	0	12.43	12.92	LP	LP	*	*
	5	0	0	13.11	11.89	LP	LP	*	*
	10	0	0	15.04	12.07	LP	LP	*	*

8) Multiply LP residual of the unenhanced speech signal with the final weight function to get a temporally enhanced speech signal.
9) Segment the TP speech signal and compute its DFT magnitude spectrum.
10) Apply the gain function of SP4 and SP5 to the TP speech magnitude spectrum to get the final TPSP speech signal.

3. Objective Evaluation of the Proposed Algorithms

Objective evaluation methods of speech enhancement algorithms have repeatable results, and it is inexpensive in terms of time and money in comparison with the subjective MOS listening test. Quality and intelligibility are two performance metrics for speech enhancement algorithms. It has been observed in the relevant literature that most of the speech enhancement methods are evaluated for quality using the perceptual evaluation of speech quality (PESQ) metric, which was developed for normal hearing (NH) listeners (Rix et al., n.d.). It is extremely important to evaluate speech enhancement algorithms objectively for HI listeners. In this study, the objective evaluation of speech quality and intelligibility are carried out using hearing aid speech quality index (HASQI) version 2 (Kates & Arehart, 2014b) and hearing aid speech perception index (HASPI) (Kates & Arehart, 2014a), respectively. HASQI and HASPI are objective metrics for normal hearing (NH) and hearing-impaired (HI) auditory systems. HASQI is better than PESQ to predict perceived speech quality scores given by HI listeners. HASQI and HASPI evaluation criteria provide accuracy for speech signal corrupted by additive background noise, frequency lowering, dynamic range compression, noise reduction, noise vocoder reproduction, improper acoustic feedback cancellation, peak clipping, and centre clipping (Kates & Arehart, 2014a, 2014b). HASQI index is the product of the cochlear model output signal and the cepstral correlation stage output. Clean and enhanced speech signals are passed through a gammatone analysis filter bank. The hearing loss in dB is defined at six audiometric frequencies [250, 500, 1000, 2000, 4000, 6000] Hz is applied to the auditory model for both the clean and the enhanced speech signals. The enhanced speech signals are amplified using NAL-R equalization, which provide a gain based on the hearing threshold mentioned in the audiogram. However, due to inner hair cell damage, there is a loss in the signal and the envelope of the clean and the enhanced speech signals which is obtained at the output of the cochlear model is attenuated. The envelopes of the clean and the enhanced speech signals are further smoothed using 16 ms Hann window with 50% overlap, which provides a low pass filter cut-off frequency of 62.5 Hz and a smoothed envelope sampling rate of 125 Hz. These smoothed speech envelopes are fitted with a set of cepstral basis functions, and the correlation between these cepstral basis functions is computed for clean and enhanced signals. A nonlinear term Q_{Nonlin} based on the first and second-order terms of the cepstral correlation is calculated and given as

$$Q_{Nonlin} = c^2 v \tag{1}$$

where c is the cepstral correlation and v is the vibration correlation.

A linear term Q_{Linear} which is based on the differences in the clean and enhanced long-term spectrum and slope standard deviations is computed as

$$Q_{Linear} = 1 - 0.579\sigma_1 - 0.421\sigma_2 \tag{2}$$

Table 5
Percentage improvement in HASQI scores for train noise with reference to the noisy signal and performance of the algorithms ('*' represents significant improvement, '#' represents equal performance, 'LP' represents no improvement with reference to the noisy signal).

Train noise Audiogram	SNR (dB)	% Improvement				Performance			
		SP4	SP5	TPSP4	TPSP5	SP4	SP5	TPSP4	TPSP5
A1	-5	0	0	0.68	0.75	LP	LP	#	#
	-3	0	0	0.52	0.56	LP	LP	#	#
	0	0	0	0.39	0.4	LP	LP	#	#
	3	0	0	0.38	0.37	LP	LP	#	#
	5	0	0	0.38	0.37	LP	LP	#	#
	10	0	0	0.39	0.38	LP	LP	#	#
A2	-5	0	0	8.9	0	LP	LP	*	LP
	-3	0	0	0	0	LP	LP	LP	LP
	0	0	0	0	0	LP	LP	LP	LP
	3	0	0	0	0	LP	LP	LP	LP
	5	0	0	0	0	LP	LP	LP	LP
	10	0	0	0	0	LP	LP	LP	LP
A3	-5	5.19	11.62	43.2	54.9	*	*	*	*
	-3	6.2	13.9	38.6	53.3	*	*	*	*
	0	3.17	10.88	33.2	42.6	*	*	*	*
	3	0	4.56	24	28.7	LP	*	*	*
	5	0	0.82	18.6	21.5	LP	#	*	*
	10	0	0	11.3	10.9	LP	LP	*	*
A4	-5	0	0	5.43	7.47	LP	LP	*	*
	-3	0	0	3.95	7.92	LP	LP	*	*
	0	0	0	8.57	13.8	LP	LP	*	*
	3	0	0	14.3	18.9	LP	LP	*	*
	5	0	0	15.3	19.6	LP	LP	*	*
	10	0	0	14.6	15.1	LP	LP	*	*
A5	-5	8.14	17.2	29.8	40.7	*	*	*	*
	-3	8.89	18	26.3	36.6	*	*	*	*
	0	4.18	12.9	17.8	26.2	*	*	*	*
	3	0.18	7.14	9.86	15.6	#	*	*	*
	5	0	3.81	5.77	9.83	LP	*	*	*
	10	0	0	0.38	1.93	LP	LP	#	#
A6	-5	0	0	25.8	33.5	LP	LP	*	*
	-3	0	0	22.7	27.8	LP	LP	*	*
	0	0	0	19.4	20.5	LP	LP	*	*
	3	0	0	16.5	14.7	LP	LP	*	*
	5	0	0	15.4	12.5	LP	LP	*	*
	10	0	0	15.3	11.4	LP	LP	*	*

where σ_1 is the standard deviation of the differences in the spectral shape and σ_2 is the standard deviation of the differences in the spectral slope. Finally, a combined index is calculated which is given as

$$Q_{Combined} = Q_{Nonlin} \times Q_{Linear} \tag{3}$$

In HASPI evaluation criteria, the correlation coefficients which are based on the spectral shape of the clean and the enhanced speech signals are computed. Then the auditory coherence values which are related to the cross-correlation between high portions of the clean and enhanced signals are calculated.

$$p = -9.047 + 14.81c + 4.616a_{High} \tag{4}$$

HASPI index is given as

$$H = \frac{1}{1 + e^{-p}} \tag{5}$$

HASPI and HASQI score display the extent of faithful reproduction of the speech envelope along with the signal temporal fine structure. HASQI score additionally includes changes in the long-term spectrum of the signal. In the case of HI listeners, both metrics require audiogram specifications of listeners. The six different types of sensorineural hearing loss which are used for comparing the algorithms are shown in Table 1.

3.1. Evaluation Procedure

For objective evaluations, clean speech files from the NOIZEUS database are used. Four different background noises from different local noise environments, namely, cafeteria, traffic, station and train were recorded. The noise signal was cut randomly from the recorded noise in such a way that the length of the clean speech signal and the noise signal would be the same. Further the noise signal was scaled as per the required SNR and mixed with the clean speech signal at six SNRs; -5, -3, 0, 3, 5, and 10 dB. A total of 30 clean speech signals from the NOIZEUS database, degraded by four noise environments at six different SNRs, were enhanced by spectral processing (SP4, SP5) and the combined (TPSP4, TPSP5) algorithms. Thus, there were 30 sentences x 6 SNRs x 4 noises x 5 methods (4 speech enhancement methods + 1 unenhanced) x 6 audiograms for quality and intelligibility evaluation. The unenhanced speech signals were first enhanced by the spectral processing and the proposed combined temporal spectral processing algorithms (SP4, SP5, TPSP4, and TPSP5). The enhanced speech signals and the reference unenhanced speech signals were used as an input to HASQI, HASPI evaluation metrics. The hearing losses, as mentioned in Table 1, were used as an input to HASQI and HASPI to evaluate the usefulness of these speech enhancement algorithms for these typical hearing losses. The steps followed are summarized using Algorithm 2.

Algorithm 2

1. Select a clean speech signal.
2. Select a noise signal.
3. Make the length of the noise signal the same as the clean speech signal.
4. Scale a noise signal as per the required SNR.
5. Add a scaled noise signal to the clean speech signal to get a noisy unenhanced signal.

Table 6
Percentage improvement in HASPI scores for cafeteria noise with reference to the noisy signal and performance of the algorithms ('*' represents significant improvement, '#' represents equal performance, 'LP' represents no improvement with reference to the noisy signal).

Cafeteria noise Audiogram	SNR (dB)	% Improvement SP4	SP5	TPSP4	TPSP5	Performance SP4	SP5	TPSP4	TPSP5
A1	-5	0	0	0	15.35	LP	LP	LP	*
	-3	0	0	5.19	17.23	LP	LP	*	*
	0	12.77	27.75	28.39	55.57	*	*	*	*
	3	23.47	40.28	68.44	93.78	*	*	*	*
	5	19.3	32.34	61.94	75.19	*	*	*	*
	10	5.63	8.3	32.49	38.86	*	*	*	*
A2	-5	0	0	0	0	LP	LP	LP	LP
	-3	0	0	0	0	LP	LP	LP	LP
	0	0	0	0	0	LP	LP	LP	LP
	3	0	0	0	0	LP	LP	LP	LP
	5	0	0	0	0	LP	LP	LP	LP
	10	0	0	0	0	LP	LP	LP	LP
A3	-5	32.39	39.02	30.65	43.49	*	*	*	*
	-3	17.4	25.61	20.94	33.33	*	*	*	*
	0	8.99	16.54	12.89	20.37	*	*	*	*
	3	2.32	6.36	4.98	7.73	*	*	*	*
	5	1.72	3.46	2.87	3.95	*	*	*	*
	10	0.23	0.47	0	0	#	#	LP	LP
A4	-5	5.94	19.79	0	0	*	*	LP	LP
	-3	0	7.84	0	0	LP	*	LP	LP
	0	0	4.83	0	4.41	LP	*	LP	*
	3	0	0	0	4.41	LP	LP	LP	*
	5	0	0.24	0.31	3.79	LP	#	#	*
	10	0	0	0	0	LP	LP	LP	LP
A5	-5	0	4.96	2.42	9.59	LP	*	*	*
	-3	0	2	0.91	4.4	LP	*	#	*
	0	0	0.58	0.36	1.1	LP	#	#	*
	3	0	0.06	0.08	0.18	LP	#	#	#
	5	0	0	0.02	0.04	LP	LP	LP	LP
	10	0	0	0	0	LP	LP	LP	LP
A6	-5	0	0	0	0	LP	LP	LP	LP
	-3	0	0	0	0	LP	LP	LP	LP
	0	0	0	0	0	LP	LP	LP	LP
	3	0	0	0	0	LP	LP	LP	LP
	5	0	0	0	0	LP	LP	LP	LP
	10	0	0	0	0	LP	LP	LP	LP

6 Compute HASQI using a clean and enhanced speech signal for an audiogram.
7 Compute HASPI using a clean and enhanced speech signal for an audiogram.

3.2. Experimental Results and Discussion

In this section, the results of the five methods (unenhanced, SP4, SP5, TPSP4, TPSP5 methods) in terms of objective criteria HASQI and HASPI are reported. Fig. 2, Fig. 3, Fig. 4, Fig. 5 display graphs of the mean scores of HASQI for audiogram 1 to audiogram 6 at -5, -3,0, 3, 5,10 dB SNRs with four different noises; namely, cafeteria, traffic, train, and station. The unenhanced speech signal scores were taken for reference while analyzing the performance of each method. Tables 2–5 and Tables 6–9, summarize the performance of the five methods in terms of objective criteria HASQI and HASPI respectively for audiogram 1 to audiogram 6 at -5, -3,0, 3, 5,10 dB SNRs with four different noises; namely, cafeteria, traffic, train and station. In Tables 2–5 and Tables 6–9, '#' represents that the algorithms performed equally as the unenhanced speech signal, '*' denotes that the performance is better than the unenhanced speech signal and 'LP' represents poor performance with reference to the unenhanced speech signal. Percentage improvement in HASQI and HASPI were calculated as

$$percentage\ improvement = \frac{unenhanced\ speech\ index\ score - enhanced\ speech\ index\ score}{unenhanced\ speech\ index\ score} \times 100$$

A) Objective Evaluation of Speech Quality
It is observed from Fig. 2, Fig. 3, Fig. 4, Fig. 5 and Tables 2–5, that For Audiogram 1,

- The proposed combined TPSP algorithm had a significantly better speech quality performance in the cafeteria and station noise environment at all SNRs in comparison with the unenhanced speech signal as well as individual spectral processing methods.
- For traffic noise, the proposed algorithm performed better at 0, 3, 5,10 dB and had an equal performance at -5, -3 dB SNRs with reference to the unenhanced speech signal.
- For cafeteria and traffic noise, the spectral processing algorithms performed equally well like the unenhanced speech signal for all SNRs; whereas, for the station and train noise, the spectral processing algorithms had low performance.

Table 7
Percentage improvement in HASPI scores for traffic noise with reference to the noisy signal and performance of the algorithms ('*' represents significant improvement, '#' represents equal performance, 'LP' represents no improvement with reference to the noisy signal).

Traffic noise Audiogram	SNR (dB)	% Improvement				Performance			
		SP4	SP5	TPSP4	TPSP5	SP4	SP5	TPSP4	TPSP5
A1	-5	0	0	0	14.65	LP	LP	LP	*
	-3	0	0	4.62	22.19	LP	LP	*	*
	0	0	0	17.24	37.96	LP	LP	*	*
	3	0	0	9.96	24.46	LP	LP	*	*
	5	0	0	10.96	25.27	LP	LP	*	*
	10	0	4.69	18.79	26.7	LP	LP	*	*
A2	-5	15.4	0	0	0	*	LP	LP	LP
	-3	3	0	0	0	*	LP	LP	LP
	0	0	0	0	0	LP	LP	LP	LP
	3	0	0	0	0	LP	LP	LP	LP
	5	0	0	0	0	LP	LP	LP	LP
	10	0	0	0	0	LP	LP	LP	LP
A3	-5	26.5	26.3	33.05	34.31	*	*	*	*
	-3	13.3	15.24	17.17	18.16	*	*	*	*
	0	4.7	5.95	5.65	5.98	*	*	*	*
	3	1.8	1.97	1.31	0.84	*	*	*	#
	5	0.85	0.83	0.15	0	#	#	#	LP
	10	0.06	0	0	0	#	LP	LP	LP
A4	-5	0	0	0	0	LP	LP	LP	LP
	-3	0	0	0	0	LP	LP	LP	LP
	0	0	0	0.1	0.97	LP	LP	#	#
	3	0	0	0.64	0	LP	LP	#	LP
	5	0	0	0.23	0	LP	LP	#	LP
	10	0	0	0	0	LP	LP	LP	LP
A5	-5	4.84	5.29	5.47	5.73	*	*	*	*
	-3	1.4	1.65	1.63	1.76	*	*	*	*
	0	0.17	0.23	0.21	0.25	#	#	#	#
	3	0	0.01	0.02	0.01	LP	#	#	#
	5	0	0	0	0	LP	LP	LP	LP
	10	0	0	0	0	LP	LP	LP	LP
A6	-5	0	0	1.24	0	LP	LP	*	LP
	-3	0	0	0	0	LP	LP	LP	LP
	0	0	0	0	0	LP	LP	LP	LP
	3	0	0	0	0	LP	LP	LP	LP
	5	0	0	0	0	LP	LP	LP	LP
	10	0	0	0	0	LP	LP	LP	LP

For Audiogram 2,

- For cafeteria, traffic, station, and train noise at all SNRs, the proposed algorithms had poor speech quality performance in comparison with the unenhanced speech signal.
- For cafeteria noise, SP5 and for traffic noise SP4 and SP5 at -5, -3, and 0 dB SNRs, showed improved performance in comparison with the unenhanced speech signal.
- SP4 and SP5 showed low performance for the station and train noise at all SNRs and train noise at -5, -3 dB SNRs.

For Audiogram 3,

- For cafeteria, traffic, station, and train noise at all SNRs, the proposed algorithms had a significantly better performance in terms of speech quality in comparison with the unenhanced speech signal.
- The proposed algorithm TPSP5 had a significantly better performance in comparison with SP4 and SP5 for all noises at all SNRs.
- Among the spectral processing algorithms, SP5 showed good performance at -5, -3 0 and 3 dB for all noises, low performance at 5, 10 dB SNRs for traffic and station noise, and at 10 dB SNR for cafeteria and train noise.
- SP5 performed equally as the unenhanced speech signal for traffic noise at 3 dB SNR, for train noise at 5 dB SNR.

For Audiogram 4,

- The proposed algorithm TPSP5 performed better with reference to the unenhanced speech signal for cafeteria, traffic noises at -3, 0, 3, 5, 10 dB SNRs and for the station, train noises at all SNRs but the performance was low for the cafeteria noise at -5 dB SNR.
- SP4 showed low performance for all four noises at all SNRs.
- The performance of SP5 was low at all SNRs for cafeteria and train noise and for station noise at -5, -3, 5,10 dB, whereas the performance was equal to the unenhanced speech signal for traffic noise at 3,5 dB and for the station noise at 0, 3 dB SNRs.

For Audiogram 5,

- The proposed algorithm TPSP4 performed better in terms of speech quality with reference to the unenhanced speech signal for cafeteria, station and train noises at all SNRs and for the traffic noise at -5, -3,0,3,5 dB SNRs. TPSP5 performed better at all SNRs for cafeteria and station noises and for traffic, train noises at -5, -3, 0, 3, 5 dB SNRs.
- Spectral processing algorithm SP5 also performed better with reference to the unenhanced speech signal for cafeteria, station, train noises at all SNRs, for traffic noise at -5, -3,0,3 dB SNRs; however, the performance of TPSP5 was significantly better in comparison with SP5 for all noises at all SNRs.

For Audiogram 6,

- The performance of TPSP5 was better for traffic, station and train noise at all SNRs, for cafeteria noise at 0, 3, 5,10 dB SNRs with

Table 8
Percentage improvement in HASPI scores for station noise with reference to the noisy signal and performance of the algorithms ('*' represents significant improvement, '#' represents equal performance, 'LP' represents no improvement with reference to the noisy signal).

Station noise Audiogram	SNR (dB)	% Improvement				Performance			
		SP4	SP5	TPSP4	TPSP5	SP4	SP5	TPSP4	TPSP5
A1	-5	6.25	0	32.34	50.79	*	LP	*	*
	-3	2.73	0.62	27.1	37.2	*	#	*	*
	0	0	0	19.5	25.05	LP	LP	*	*
	3	0	0	19.1	23.4	LP	LP	*	*
	5	0	0	19.01	23.7	LP	LP	*	*
	10	0	0	19.06	24.8	LP	LP	*	*
A2	-5	0	0	0	0	LP	LP	LP	LP
	-3	0	0	0	0	LP	LP	LP	LP
	0	0	0	0	0	LP	LP	LP	LP
	3	0	0	0	0	LP	LP	LP	LP
	5	0	0	0	0	LP	LP	LP	LP
	10	0	0	0	0	LP	LP	LP	LP
A3	-5	0	21.2	11.05	32.4	LP	*	*	*
	-3	2.7	16.97	7.98	20.13	*	*	*	*
	0	1.3	8.37	1.69	6.8	*	*	*	*
	3	0	2.22	0	0	LP	*	LP	LP
	5	0	0.45	0	0	LP	#	LP	LP
	10	0	0	0	0	LP	LP	LP	LP
A4	-5	0	4.64	2.24	28.14	LP	*	*	*
	-3	0	5.28	4.18	20.29	LP	*	*	*
	0	0	2.92	3.13	12.99	LP	*	*	*
	3	0	0.17	1.55	4.71	LP	#	*	*
	5	0	0	0.3	0.94	LP	LP	#	#
	10	0	0	0	0	LP	LP	LP	LP
A5	-5	0	0	0	2.31	LP	LP	LP	*
	-3	0	0	0	0.39	LP	LP	LP	#
	0	0	0	0	0	LP	LP	LP	LP
	3	0	0	0	0	LP	LP	LP	LP
	5	0	0	0	0	LP	LP	LP	LP
	10	0	0	0	0	LP	LP	LP	LP
A6	-5	0	0	0	0	LP	LP	LP	LP
	-3	0	0	0	0	LP	LP	LP	LP
	0	0	0	0	0	LP	LP	LP	LP
	3	0	0	0	0	LP	LP	LP	LP
	5	0	0	0	0	LP	LP	LP	LP
	10	0	0	0	0	LP	LP	LP	LP

reference to the unenhanced speech signal as well as individual spectral processing algorithms in terms of speech quality.

- TPSP4 followed the same trend as TPSP5 in terms of speech quality except for the cafeteria noise at -5, -3 dB SNRs.
- The spectral processing algorithms showed a low speech quality performance for almost all noises at all the SNRs.

To summarize, the proposed combined TPSP5 showed a significantly improved performance in terms of speech quality in comparison with the unenhanced speech signal as well as the individual spectral processing algorithms SP4, SP5 for almost all audiograms in all the four noise environments at all SNRs. The limitation of TPSP5 was in the case of audiogram 2 which was moderately sloping high frequency loss. For future exploration, hearing loss of audiogram can be considered and deep neural network-based methods could be used to improve speech quality results.

B) Objective Evaluation of Speech Intelligibility

Figs. 6–9 display graphs of the mean scores of HASPI and Table 6–9 summarize the performance of the five methods in terms of objective criteria HASPI for audiogram 1 to audiogram 6 at -5, -3, 0, 3, 5, 10 dB SNRs with four different noises; namely, cafeteria, traffic, train and station respectively. It can be observed from Fig. 6 - Fig. 9, and Tables 6–9, that-

For Audiogram 1,

- The proposed algorithm TPSP5 performed better in terms of speech intelligibility with reference to the unenhanced speech signal for all four noises at all SNRs. TPSP4 showed a good improvement for the station and train noises at all SNRs and for cafeteria and traffic noises at -3, 0, 3, 5, 10 dB SNRs.
- Spectral processing algorithms SP4, SP5 performed better with reference to the unenhanced speech signal for the cafeteria noise at 0,3,5,10 dB SNRs, but for traffic, station, and train noise, their performance in terms of speech intelligibility was low in comparison with the unenhanced speech signal.

For Audiogram 2,

- The spectral processing algorithms and the proposed combined TPSP algorithms showed low performance with reference to the unenhanced speech signal in terms of speech intelligibility. The moderately sloping high-frequency loss can be considered for future exploration.

For Audiogram 3,

- The spectral processing algorithms SP4, SP5 performed better with reference to the unenhanced speech signal in terms of speech intelligibility for all noises at almost all SNRs.
- The proposed combined TPSP showed an improvement in speech intelligibility with reference to the unenhanced speech signal in terms of speech intelligibility for all noises at almost all SNRs. However, the performance of TPSP5 improved in comparison with the individual spectral processing methods for all the noises at almost all SNRs.

Table 9
Percentage improvement in HASPI scores for train noise with reference to the noisy signal and performance of the algorithms ('*' represents significant improvement, '#' represents equal performance, 'LP' represents no improvement with reference to the noisy signal).

Train noise Audiogram	SNR (dB)	% Improvement SP4	SP5	TPSP4	TPSP5	Performance SP4	SP5	TPSP4	TPSP5
A1	-5	0	0	70.2	75.1	LP	LP	*	*
	-3	0	0	52.3	57.8	LP	LP	*	*
	0	0	0	29.1	33.1	LP	LP	*	*
	3	0	0	23.3	26.2	LP	LP	*	*
	5	0	0	22.6	23.97	LP	LP	*	*
	10	0	0	22.76	24.2	LP	LP	*	*
A2	-5	0	0	13.03	0	LP	LP	*	LP
	-3	0	0	0	0	LP	LP	LP	LP
	0	0	0	0	0	LP	LP	LP	LP
	3	0	0	0	0	LP	LP	LP	LP
	5	0	0	0	0	LP	LP	LP	LP
	10	0	0	0	0	LP	LP	LP	LP
A3	-5	36.78	39.28	63.34	70.73	*	*	*	*
	-3	23.68	25.62	38.6	41.63	*	*	*	*
	0	9.67	11.08	15.27	16.64	*	*	*	*
	3	3.19	3.33	4.7	4.38	*	*	*	*
	5	1.34	1.41	1.69	1.04	#	#	*	#
	10	0	0	0	0	LP	LP	LP	LP
A4	-5	3.64	6.72	42.1	49.68	*	*	*	*
	-3	0	2.15	19.88	23.98	LP	*	*	*
	0	0	0	4.07	5.09	LP	LP	*	*
	3	0	0	2.47	0.55	LP	LP	*	#
	5	0	0	1.32	0	LP	LP	#	LP
	10	0	0	0	0	LP	LP	LP	LP
A5	-5	0.39	1.83	5.37	6.44	#	*	*	*
	-3	0.21	0.65	1.95	2.28	#	#	*	*
	0	0.01	0.11	0.42	0.48	LP	#	#	#
	3	0	0	0.07	0.07	LP	LP	LP	LP
	5	0	0	0.02	0.01	LP	LP	LP	LP
	10	0	0	0	0	LP	LP	LP	LP
A6	-5	0	0	7.16	7.59	LP	LP	*	*
	-3	0	0	1.19	0	LP	LP	#	LP
	0	0	0	0	0	LP	LP	LP	LP
	3	0	0	0	0	LP	LP	LP	LP
	5	0	0	0	0	LP	LP	LP	LP
	10	0	0	0	0	LP	LP	LP	LP

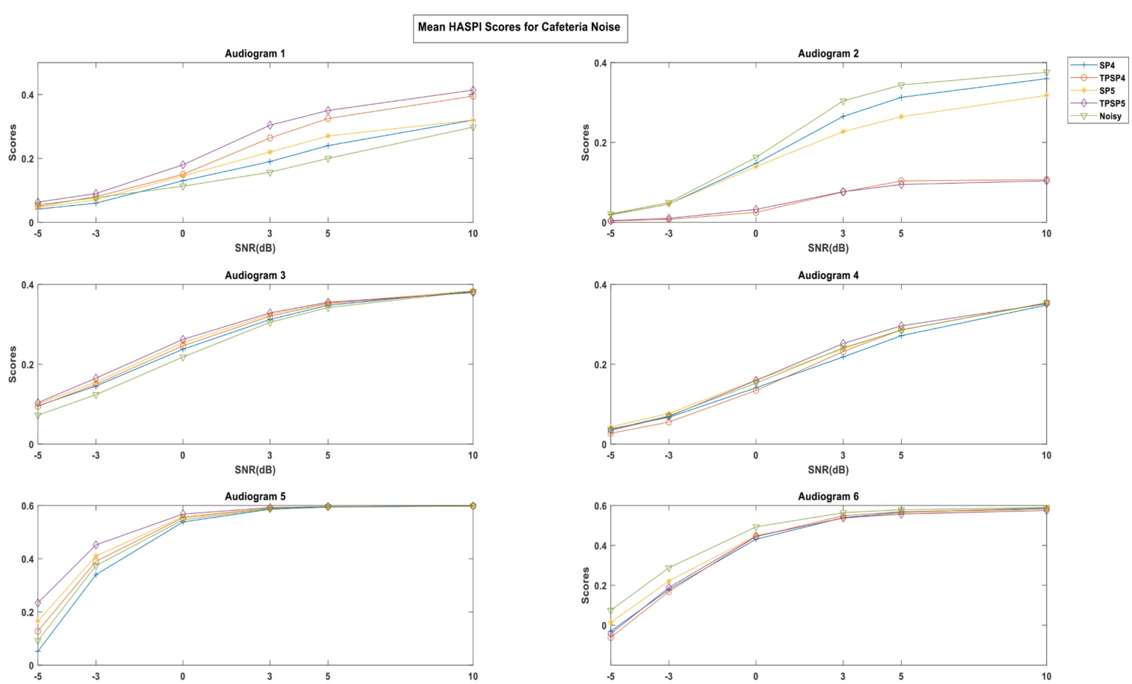

Fig. 6. Mean HASPI scores for cafeteria noise in six audiograms.

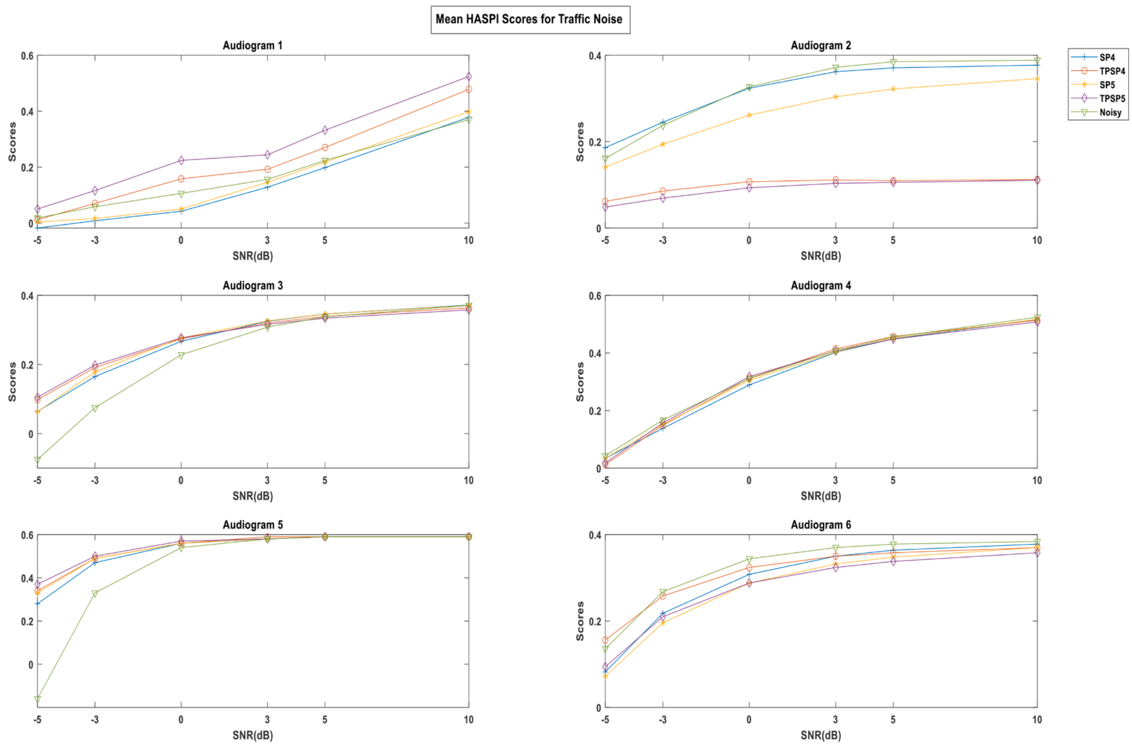

Fig. 7. Mean HASPI scores for traffic noise in six audiograms.

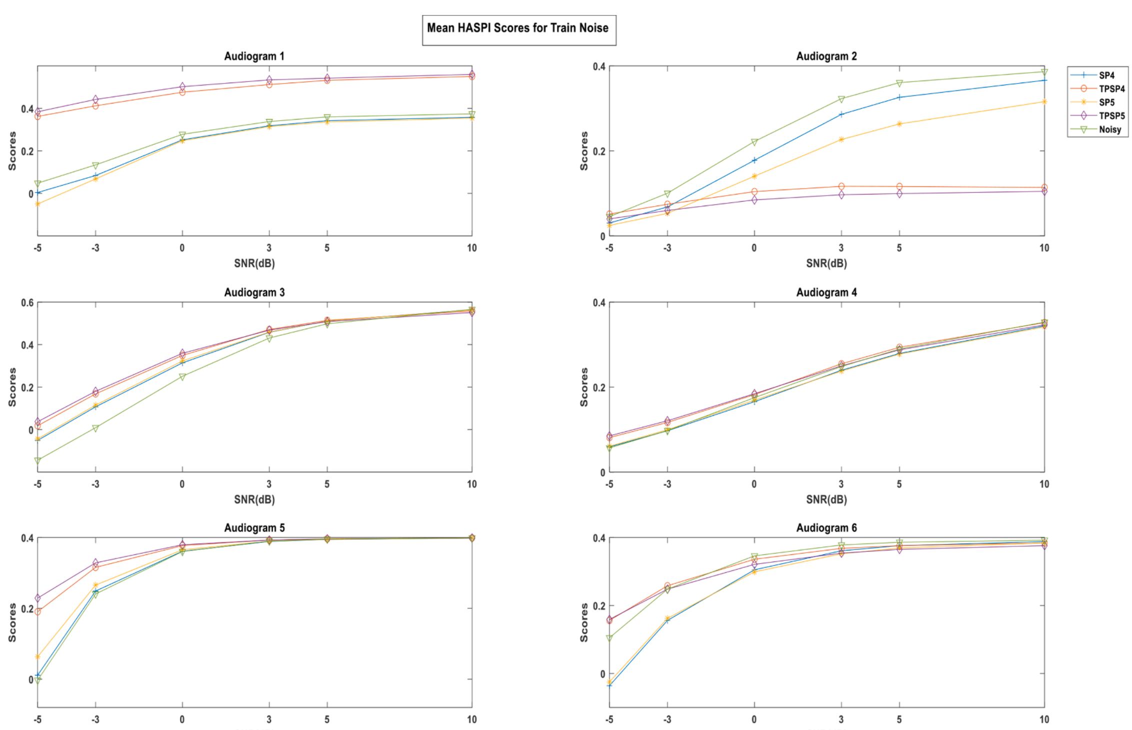

Fig. 8. Mean HASPI scores for train noise in six audiograms.

For Audiogram 4,

- The proposed TPSP5 algorithm showed better results in terms of speech intelligibility for the station, train noises at -5, -3, and 0 dB SNRs and for cafeteria noise at 0, 3, 5 dB SNRs, whereas low performance for the cafeteria and traffic noise at -5, -3 dB SNRs.
- The spectral processing algorithms SP4, SP5 showed low performance in terms of speech intelligibility with reference to the unenhanced speech signal for traffic, station and train noises at all SNRs.
- Though the performance of the combined TPSP was not at par, it showed improved results in comparison with the individual spectral processing methods.

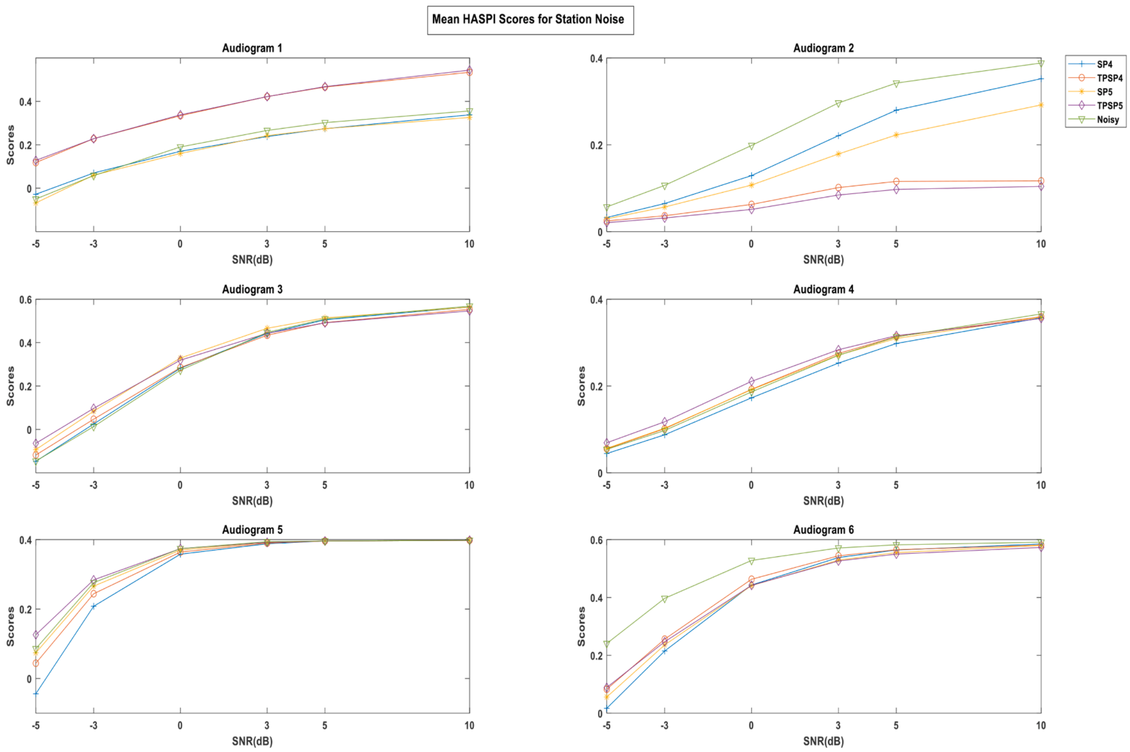

Fig. 9. Mean HASPI scores for station noise in six audiograms.

Table 10
Processing time of proposed algorithms for different noisy unenhanced speech files.

Proposed Algorithm	Cafeteria noise	Traffic noise	Station noise	Train noise
TPSP4	5.22 sec	4.73 sec	7.9 sec	15.4 sec
TPSP5	6.15 sec	5.95 sec	6.22 sec	4.68 sec

Table 11
Audiograms of the HI listeners participated in the listening test.

Frequency (kHz)	0.25	0.5	1	2	4	6	8	Listener number
Hearing Loss in dB	35	35	40	70	80	65	65	1
	15	15	20	25	50	45	55	2
	20	15	20	30	35	40	60	3
	25	30	30	25	35	40	40	4
	25	30	25	30	35	35	35	5

For Audiogram 5,

- The combined TPSP5 had better results with reference to the unenhanced speech signal in terms of speech intelligibility for the cafeteria, station, train noise at -5, -3, 0 dB SNRs and for the station noise at -5 dB SNR.
- Spectral processing algorithms SP4, SP5 performed better with reference to the unenhanced speech signal in terms of speech intelligibility for the traffic and train noise at -5, -3,0 dB SNRs.
- The combined TPSP had an equal performance as the unenhanced speech signal in terms of speech intelligibility for all noises at 3,5,10 dB SNRs.
- Though the performance of the combined TPSP algorithm was not at par with some SNRs, the performance improved in comparison with the individual spectral processing algorithms.

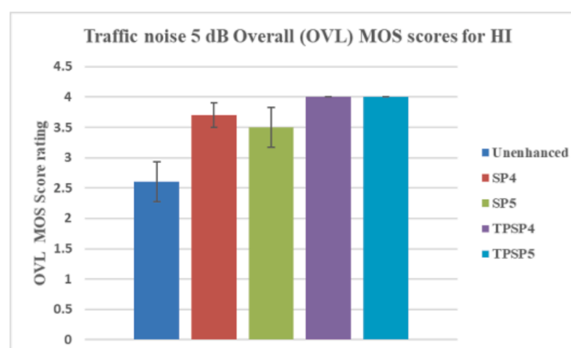

Fig. 10. Bar graph of the overall (OVL) quality MOS scores for traffic noise at 5 dB SNR.

For Audiogram 6,

- The spectral processing and the combined TPSP, both the algorithms had low performance in terms of speech intelligibility for all the noises.

To summarize, the proposed combined TPSP5 algorithm had a better performance for Audiogram 1, Audiogram 3, and Audiogram 5 in terms of speech intelligibility in comparison with the unenhanced speech signal as well as the individual spectral processing algorithms. For Audiogram 2 that was moderately sloping high-frequency hearing loss, the spectral processing, as well as the combined TPSP algorithms, showed a poor performance. It was further observed from the graphs that SP4 was introducing less distortion to the processed speech signal in comparison with other algorithms. For Audiogram 4, the combined TPSP5 had better results for the station, train noises at negative SNRs, whereas the equal performance as the unenhanced speech signal for positive SNRs. For Audiogram 6, which was again with high-frequency loss, had a poor performance.

The running time of the proposed algorithms TPSP4 and TPSP5 is different for different noises. For the same input unenhanced wav speech file, the running times are mentioned in Table 10.

It was observed that the algorithms improved speech quality in comparison with the unenhanced speech signal in terms of HASQI criteria, but the intelligibility improvement did not follow the same trend in terms of HASPI criteria in some cases. Though the spectral processing algorithms estimate and suppress the background noise accurately and they contribute to the speech intelligibility improvements, they cannot provide a considerable improvement in intelligibility (Loizou et al., 2011). It was reported that the nonlinear compensate processing (NAL-NL2) could degrade the SNR of the enhanced speech in hearing aids (Naylor & Burmand Johannesson, 2009; Souza et al., 2006). Hence, for future analysis and exploration, the hearing threshold, comfortable level of hearing of the subjects can be used, and a deep neural network-based algorithm can be used to improve performance. The proposed algorithm still can be improved if the different spectral clustering algorithm ideas (Li, Nie, Chang, Yang, et al., 2018a; Li, Nie, Chang, Yang, et al., 2018b; Li et al., 2019) are explored.

4. Subjective quality evaluation

Although, the proposed algorithms were evaluated objectively for six different types of audiograms, it was tedious and a very time-consuming process to search the HI listener with the same type of audiogram and evaluate the proposed algorithms subjectively. Hence, subjective quality evaluation discussed in this section for the sake of giving vision to the future detailed analysis of the application. For conducting listening test, HI listeners with mild to moderate hearing loss were considered. The audiograms of HI listeners participated in the quality evaluation test are shown in Table 11.

While conducting the listening test for HI listeners, the hearing aid which was worn regularly by them was removed. The output speech signal files were amplified as per the requirement of the HI listener's audiogram. The amplification was adjusted before the test and kept undisturbed during the test. Ordinary headphones which are usually used with computers were used. Firstly, the listeners were given training to become aware of the listening test procedure. The unenhanced speech, clean speech, and the enhanced output speech files were presented to the listeners in a random sequence. The listeners were asked to rate the enhanced output, unenhanced, and clean speech file in terms of the overall quality criteria on a scale of 1 to 5. 5 being the excellent quality and 1 being poor quality of speech (Hu and Loizou, 2006). Fig. 10 shows the bar graph of overall quality MOS scores with error bars computed using 95% confidence interval. As observed in the bar graph, TPSP shows improved performance in comparison with the unenhanced and spectral processing algorithms.

5. Conclusion

In this study, the authors objectively evaluated the performance of a new and unique combination of temporal-spectral processing (TPSP) for hearing-impaired people in terms of HASQI and HASPI parameters. From the experimental results, it can be concluded that the TPSP5 algorithm improved the speech quality performance in comparison with the individual spectral processing methods in different cases of background noise (i.e., cafeteria, traffic, station and train) listening environments in terms of HASQI criteria. The proposed combined TPSP5 algorithm has better speech intelligibility in terms of HASPI criteria than individual spectral processing methods for most of the typical cases of audiograms. In some cases, like Audiogram 2 and Audiogram 6, the results in terms of HASPI were not significant. The subjective quality evaluation displayed that TPSP4 and TPSP5 improved the speech quality. In future analysis, hearing threshold criterion can be used, and deep neural network algorithms can be trained for further improvements. A limitation of this research is, only typical varied cases of audiograms were considered for objective and subjective evaluations. To support the results in future, the algorithms can be evaluated in terms of quality and intelligibility in real hearing aids.

Credit Author Statement

Hemangi Shinde: Methodology, Evaluation, Data analysis, Draft preparation

Vibha Vyas: Supervision, Review, Analyzing results

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank Dr. S. R. M. Prasanna and Dr. Vikram C.M. for their timely guidance and support in implementing temporal processing method

References

Ananthapadmanabha, T.v., & Yegnanarayana, B. (1979). Epoch Extraction from Linear Prediction Residual for Identification of Closed Glottis Interval. *IEEE Transactions on Acoustics, Speech, and Signal Processing*. https://doi.org/10.1109/TASSP.1979.1163267

Berouti, M., Schwartz, R., & Makhoul, J. (1979). Enhancement of speech corrupted by acoustic noise.

Boll, S. F. (1979). Suppression of Acoustic Noise in Speech Using Spectral Subtraction. *IEEE Transactions on Acoustics, Speech, and Signal Processing*. https://doi.org/10.1109/TASSP.1979.1163209

Cohen, I. (2005). Relaxed statistical model for speech enhancement and a priori SNR estimation. *IEEE Transactions on Speech and Audio Processing*. https://doi.org/10.1109/TSA.2005.851940

Deafness and hearing loss. (n.d.). Retrieved April 1, 2021, from https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss.

Deepak, K. T., & Prasanna, S. R. M. (2015). Epoch Extraction Using Zero Band Filtering from Speech Signal. *Circuits, Systems, and Signal Processing*. https://doi.org/10.1007/s00034-014-9957-4

Deepak, K. T., & Prasanna, S. R. M. (2016). Foreground Speech Segmentation and Enhancement Using Glottal Closure Instants and Mel Cepstral Coefficients. *IEEE/ACM Transactions on Audio Speech and Language Processing*. https://doi.org/10.1109/TASLP.2016.2549699

Ephraim, Y., & Malah, D. (1984). Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator. *IEEE Transactions on Acoustics, Speech, and Signal Processing*. https://doi.org/10.1109/TASSP.1984.1164453

Ephraim, Y., & Malah, D. (1985). Speech enhancement using a minimum mean-square error-log-spectral amplitude estimator. *IEEE Transactions on Acoustics, Speech, and Signal Processing*. https://doi.org/10.1109/TASSP.1985.1164550

Hu, Y., & Loizou, P. C. (2006). Subjective comparison of speech enhancement algorithms. In *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*. https://doi.org/10.1109/icassp.2006.1659980

Kates, J. M., & Arehart, K. H. (2014a). The hearing-aid speech perception index (HASPI). *Speech Communication, 65*, 75–93. https://doi.org/10.1016/j.specom.2014.06.002

Kates, J. M., & Arehart, K. H. (2014b). The hearing-aid speech quality index (HASQI) version 2. *AES: Journal of the Audio Engineering Society, 62*(3), 99–117. https://doi.org/10.17743/jaes.2014.0006

Krishnamoorthy, P., & Prasanna, S. R. M. (2009). Application of combined temporal and spectral processing methods for speaker recognition under noisy, reverberant or multi-speaker environments. *Sadhana - Academy Proceedings in Engineering Sciences, 34*(5), 729–754. https://doi.org/10.1007/s12046-009-0043-8

Krishnamoorthy, P., & Prasanna, S. R. M. (2011). Enhancement of noisy speech by temporal and spectral processing. *Speech Communication*. https://doi.org/10.1016/j.specom.2010.08.011

Lai, A., Lai, Y.-H., Tsao, Y., & Chen, F. (2014). Issued Date 2014 Rights Creative Commons: Attribution 3.0 Hong Kong License. In *Citation International Journal of Speech & Language Pathology and Audiology*. http://hdl.handle.net/10722/194756.

Lee, J., & Kang, H. G. (2019). 'A joint learning algorithm for complex-valued t-f masks in deep learning-based single-channel speech enhancement systems'. *IEEE/ACM Transactions on Audio Speech and Language Processing*. https://doi.org/10.1109/TASLP.2019.2910638

Li, N., & Loizou, P. C. (2008). Factors influencing intelligibility of ideal binary-masked speech: Implications for noise reduction. *The Journal of the Acoustical Society of America, 123*(3), 1673–1682. https://doi.org/10.1121/1.2832617

Li, Z., Nie, F., Chang, X., Nie, L., Zhang, H., & Yang, Y. (2018a). Rank-Constrained Spectral Clustering With Flexible Embedding; Rank-Constrained Spectral Clustering

With Flexible Embedding. *IEEE Transactions on Neural Networks and Learning Systems*, (12), 29. https://doi.org/10.1109/TNNLS.2018.2817538

Li, Z., Nie, F., Chang, X., Yang, Y., Zhang, C., & Sebe, N. (2018b). Dynamic Affinity Graph Construction for Spectral Clustering Using Multiple Features; Dynamic Affinity Graph Construction for Spectral Clustering Using Multiple Features. *IEEE Transactions on Neural Networks and Learning Systems*, (12), 29. https://doi.org/10.1109/TNNLS.2018.2829867

Li, Z., Yao, L., Chang, X., Zhan, K., Sun, J., & Zhang, H. (2019). Zero-shot event detection via event-adaptive concept relevance mining. *Pattern Recognition, 88*, 595–603. https://doi.org/10.1016/J.PATCOG.2018.12.010

Loizou, P. C., Member, S., & Kim, G. (2011). Reasons why Current Speech-Enhancement Algorithms do not Improve Speech Intelligibility and Suggested Solutions. *IEEE Transactions on Audio, Speech, and Language Processing, 19*(1), 47. https://doi.org/10.1109/TASL.2010.2045180

Lu, Y., & Loizou, P. C. (2011). Estimators of the Magnitude-Squared Spectrum and Methods for Incorporating SNR Uncertainty. *IEEE Transactions on Audio, Speech and Language Processing*. https://doi.org/10.1109/TASL.2010.2082531

Naylor, G., & Burmand Johannesson, R. (2009). Long-Term Signal-to-Noise Ratio at the Input and Output of Amplitude-Compression Systems. *J Am Acad Audiol, 20*, 161–171. https://doi.org/10.3766/jaaa.20.3.2

Rix, A. W., Beerends, J. G., Hollier, M. P., & Hekstra, A. P. (n.d.). Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs.

Shinde, H., Sapkal, A. M., Phatak, A., Vikram, C. M., & Prasanna, S. R. M. (2019). Quality evaluation of combined temporal and spectral processing for hearing impaired. In *2019 IEEE 16th India Council International Conference, INDICON 2019 - Symposium Proceedings*. https://doi.org/10.1109/INDICON47234.2019.9030289

Souza, P. E., Jenstad, L. M., & Boike, K. T. (2006). Measuring the acoustic effects of compression amplification on speech in noise. *The Journal of the Acoustical Society of America, 119*(1), 41–44. https://doi.org/10.1121/1.2108861

Tan, K., & Wang, D. (2020). Learning Complex Spectral Mapping with Gated Convolutional Recurrent Networks for Monaural Speech Enhancement. *IEEE/ACM Transactions on Audio, Speech, and Language Processing, 28*. https://doi.org/10.1109/TASLP.2019.2955276

Wang, D., Kjems, U., Pedersen, M. S., Boldt, J. B., & Lunner, T. (2009). Speech intelligibility in background noise with ideal binary time-frequency masking. *The Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.3083233

Yegnanarayana, B., Avendano, C., Hermansky, H., & Satyanarayana Murthy, P. (1999). Speech enhancement using linear prediction residual. *Speech Communication*. https://doi.org/10.1016/S0167-6393(98)00070-3

Yegnanarayana, B., & Satyanarayana Murthy, P. (2000). Enhancement of reverberant speech using LP residual signal. *IEEE Transactions on Speech and Audio Processing*. https://doi.org/10.1109/89.841209

Zheng, N., & Zhang, X. L. (2019). Phase-aware speech enhancement based on deep neural networks. *IEEE/ACM Transactions on Audio Speech and Language Processing*. https://doi.org/10.1109/TASLP.2018.28707