INTERNATIONAL JOURNAL OF CIRCUITS, SYSTEMS AND SIGNAL PROCESSING
DOI: 10.46300/9106.2020.14.45 Volume 14, 2020

Improving Patient Voice Intelligibility by using a Euclidian Distance-Based Approach to Improve Voice Assistant Accuracy

A. M. Mutawa
Computer Engineering Department, College of Engineering and Petroleum, Kuwait University
Kuwait

Received: March 29, 2020. Revised: April 24, 2020. 2nd Revised: May 25, 2020.
Accepted: June 30, 2020. Published: July 20, 2020.

Abstract— Voice assistance (VA) is gaining domestic consumer attention in a variety of products, such as Amazon Alexa, Google Home, Apple's Siri, and Microsoft's Cortana. Furthermore, VA has recently shown its usefulness and ability to improve inpatient experience in hospitals and clinics. Nevertheless, none of the VA products has an accuracy rate greater than 90%. The accuracy decreases even more in noisy or public environments. Hence, improving VA accuracy in noisy environments requires a speech signal algorithm with good quality and intelligibility. There is great interest in developing an objective intelligibility measure that shows maximum correlation with subjective speech intelligibility and that can measure the effect of speech enhancement algorithms on the processing of noisy speech signals. In this paper, Euclidian distance-based speech intelligibility prediction is proposed to measure the correlation with subjective intelligibility in different noisy environments. This paper also presents a comparative analysis and general background research in speech intelligibility improvement. The results show that no single algorithm is effective in improving the intelligibility of speech signals.

Keywords—Euclidian distance, speech enhancement, speech intelligibility, voice assistance

I. INTRODUCTION

VOICE assistance (VA), which is sometimes referred to as a voice user interface (VUI), has become widely used in smart devices and household personal assistants, such as Amazon Alexa, Google Home, Microsoft's Cortana, and Apple's Siri. VA is used to control home smart devices and Internet-of-things (IoT) devices to provide a better overall user experience. Moreover, the applications of VA in hospitals and clinics showed high satisfaction in terms of inpatient experience. Patients can request assistance directly from VA instead of waiting for the nurse to answer basic inquiries, such as when the next meal will be available, the food they are allowed to consume, the time of the doctor's next visit, or help with controlling the heating or lighting systems.

Speech enhancement is the combination of improving both the quality and intelligibility of speech signals. In real-world environments, various noises degrade actual speech signals. Hence, to improve speech quality, various algorithms have been designed and presented in the literature [1-12]. Most of these speech enhancement algorithms improve quality but degrade intelligibility; these algorithms can be classified into four main types: spectral subtractive, statistical model-based, subspace-based, and Wiener-type algorithms. **Spectral subtractive-type algorithms** include Berouti spectral subtraction (Berouti-SS) [1], multiband-SS [9], Boll-SS [2], parametric-SS [11], Scalart-SS [10], and spectral subtraction using reduced-delay convolution (RDC-SS) [6]; the **statistical model-based algorithms** include the log-minimum mean-square error (MMSE) [5], [12], MMSE spectral estimator for the short-time spectral amplitude (STSA) (STSA-MMSE) [4], and Cohen-MMSE [3]. The Karhunen-Loeve transform (KLT) [7] and perceptual KLT (PKLT) [8] are **subspace-based algorithms**.

A speech signal with good quality and intelligibility is required for many applications, such as speech recognition and communication hearing aids [13]. In previous studies, most of the reported algorithms enhance quality and reduce intelligibility [14-17]. Some studies used audio file processing software to generate noisy speech signals for analog communication channels [13]. To measure the intelligibility of speech, an algorithm that produces the actual intelligibility of noisy and processed speech signals must be developed.

As shown in the literature, voice familiarity, among many other factors, improves speech intelligibility [18], and because a subjective intelligibility measure is much more expensive and time consuming, an effective objective intelligibility prediction measure is also required. In the literature, significant attention has been focused on objective speech intelligibility prediction measures [19], [20]. Objective speech intelligibility measures can be classified into two types: measures with a signal-to-noise ratio (SNR)-based design and correlation-based implementations. SNR-based implementations include the articulation index (AI), speech transmission index (STI), frequency-weighted segmental SNR (fw-SNR), and speech intelligibility index (SII); correlation-based implementations include the normalized covariance metric and short-time objective intelligibility (STOI) measures. Methods based on coherence are also provided for objective speech intelligibility predictions, i.e., magnitude squared coherence (MSC), coherence SII (CSII), coherence STI, band importance function-based CSII, and covariance-based STI (CSTI) [19]. All of these measures are useful for only a specific noise environment and are less appropriate for speech enhancement methods in which degraded speech is processed by time-frequency variation-based gain functions. These measures also do not produce objective intelligibility values that are similar to subjective intelligibility values, and most of the recently published measures remain based on SNRs and correlation [21].

The contributions from the proposed study are described as follows. First, a Euclidian distance-based objective speech intelligibility prediction measure is implemented and compared with other commonly used measures. Second, the performance differences in the speech intelligibility values produced by the algorithms are presented. From the comparative evaluation results presented in the tables, it is simple to determine one or more appropriate algorithms that preserve or enhance the speech intelligibility aspect of noisy speech signals.

The remainder of this paper is organized as follows: Section II presents the speech intelligibility evaluation parameters. Section III presents the proposed Euclidian distance-based speech intelligibility measure. A description of single-channel speech enhancement algorithms is provided in Section IV. Simulation and experimental results are discussed in Section V for speech intelligibility evaluations, and Section VI presents future directions and issues. Finally, conclusions are drawn in Section VII.

II. Speech Intelligibility Measures

In this evaluation, five commonly used objective measures for predicting the intelligibility of speech under various noisy conditions are evaluated. A description of these objective intelligibility measures is given along with the proposed Euclidian distance-based objective measure.

A. Frequency-Weighted Segmental SNR (fw-SNRseg)

The frequency-weighted segmental SNR is calculated using equation (1) [22]:

$$\frac{10}{M}\sum_{m=0}^{M-1}\frac{\sum_{j=1}^{k}W(j,m)\log_{10}\frac{X(j,m)^2}{(X(j,m)-\hat{X}(j,m))^2}}{\sum_{j=1}^{k}W(j,m)} \tag{1}$$

where K is the number of bands, M is the total number of frames, $X(j,m)$ is the critical-band magnitude of the clean signal at the j^{th} frequency band at the m^{th} frame, and $\hat{X}(j,m)$ is the corresponding enhanced speech signal. In equation (1), $W(j,m)$ is the weighting function, and p is the power exponent, which varies according to the speech. The weighting function is given in equation (2) as follows:

$$W(j,m)=X(j,m)^p \tag{2}$$

B. Short-Time Objective Intelligibility (STOI)

The STOI is based on short-time segments, i.e., 386 ms. This short segment is selected to obtain maximum correlation with the subjective speech intelligibility. The intelligibility measure is defined as the linear correlation between clean and enhanced time-frequency (TF) units and is given by equation (3) [23]:

$$I_j(m)=\frac{\sum_n (X_j(n)-\frac{1}{N}\sum_l X_j(n))\ (Y_j^{'}(n)-\frac{1}{N}\sum Y_j^{'}(n))}{\sqrt{\sum_n (X_j(n)-\frac{1}{N}\sum_l X_j(n))^2\ \sum_n (Y_j^{'}(n)-\frac{1}{N}\sum_l Y_j^{'}(n))^2}} \tag{3}$$

In equation (3), $X_j(n)$ and $Y_j(n)$ are the clean and enhanced signals, respectively. The overall average of the intelligibility measure from all bands and frames is calculated using equation (4), where M is the total number of frames and j is the number of one-third octave bands.

$$I=\frac{1}{jM}\sum_{j,\,m} d_j(m) \tag{4}$$

C. Fractional Articulation Index (fAI)

This type of intelligibility measure is based on the SNR values. The fraction or input SNR is calculated using equation (5) [24]:

$$fSNR_j=\begin{cases}\frac{\min(\overline{SNR}_j,\ SNR_j)}{SNR_j} & if \quad SNR_j \geq SNR_L\\ 0 & else\end{cases} \tag{5}$$

where $\overline{SNR}_j$ is the ratio of the output SNR in band j to the noise spectrum and is the true SNR. The lowest SNR is SNR_L, and $fSNR_j$ is bounded from 0 to 1. The weighted average is calculated across all bands to obtain *f*AI in equation (6):

$$fAI=\frac{1}{\sum_{j=1}^{K}W_j}\sum_{j=1}^{K}W_j \times fSNR_j \tag{6}$$

D. Mean Opinion Score (MOS)

The MOS is a listening quality objective measure with a value between 1 and 5. The MOS scale is defined as 5=Excellent, 4=Good, 3=Fair, 2=Poor, and 1=Bad [25].

$$MOS=A+\frac{B}{1+e^{(C.PESQ+D)}} \tag{7}$$

where A, B, C, and D are the variables given in [19] and PESQ is the perceptive evaluation of speech quality calculated from [26]. The range of the PESQ is between -0.5 and 4.5. In this measure, the first step is level equalization to the listening level.

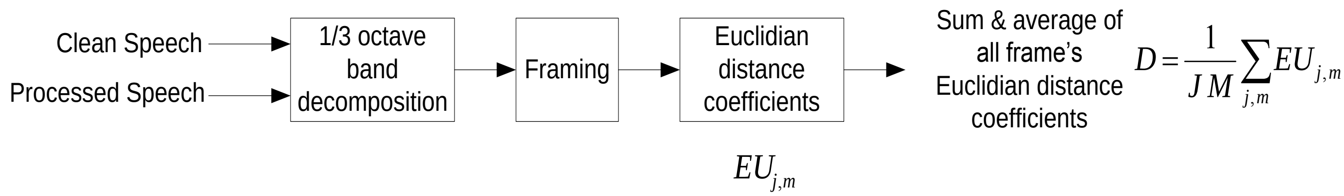

Fig. 1 The Euclidian distance-based speech intelligibility measure is a function of clean and processed speech

III. Euclidian Distance-Based Speech Intelligibility Measure

The Euclidian distance (EU)-based speech intelligibility measure is a function of clean and processed speech, as shown in Fig. 1. The output D is a scalar value that has average intelligibility with processed speech. A sampling frequency of 8 kHz is used to obtain the useful frequency range for speech intelligibility. A new objective speech intelligibility measure that is based on the Euclidian distance function given in equation (8) is proposed.

$$EU=\sqrt{(X_{cl}(j,m)-Y_{enh}(j,m))^2} \tag{8}$$

The sum of the EU coefficients of all frames is averaged using equation (9):

$$D=\frac{1}{JM}\sum_{j,m}EU_{j,m} \tag{9}$$

where EU is the Euclidian distance value and X_{cl} and Y_{enh} are the clean and enhanced speech signals, respectively. Additionally, D is the average value from all frames and bands and is normalized between 0 and 1 by using equation (10).

$$EUI=\frac{1}{(1+D)} \tag{10}$$

The basic procedure begins with a discrete Fourier transform (DFT)-based one-third octave band decomposition. A total of 15 one-third octave bands are selected, where the lowest center frequency is 150 Hz and the highest one-third octave band center frequency is approximately 4.3 kHz. The one-third octave band is defined in equation (11):

$$x_j(m)=\sqrt{\sum_{k=k_1(j)}^{k_2(j)-1}\left|\hat{x}(k,m)\right|^2} \tag{11}$$

where $\hat{x}(k,m)$ denotes the k^{th} DFT bin of the m^{th} frame of clean and processed speech. The one-third octave band edges are given as k_1 and k_2. The Euclidian distance-based intelligibility measure compares the temporal envelopes of the clean and processed speech by using the Euclidian distance coefficients.

IV. Speech Enhancement Algorithms

Noise reduction algorithms for speech can be classified into the following four primary types: (1) spectral subtractive, (2) Wiener (3) statistical model-based, and (4) subspace-based algorithms.

A. Spectral Subtractive-Type Algorithms

These types of algorithms are notably simple and are commonly used in speech enhancement. Spectral subtractive-type algorithms are based on the estimation of the noise spectral amplitude from an observed speech signal, and this estimated noise is subtracted from the noisy speech signal. Some studies have proposed an oversubtraction parameter that compares with other well-established methods [27], and other studies have implemented nonlinear spectral subtraction [28].

A limitation of spectral subtractive-type algorithms is that most of them do not consider the speech spectral property; hence, the estimation error produces isolated peaks in the denoised speech; these peaks are known as musical noise [2]. The basic block diagram of spectral subtractive-type algorithms is given in Fig. 2. To overcome the effect of musical noise, many algorithms based on the spectral subtractive principle have been proposed; these algorithms include spectral oversubtraction, multiband-SS, parametric-SS, Scalart-SS, and RDC-SS. The spectral oversubtraction method assumes that the noise spectrum uniformly affects the speech spectrum; therefore, fixed values of the subtraction parameters are used in this method. Consequently, this method is not practically suitable for all noise environments and results in decreased speech intelligibility.

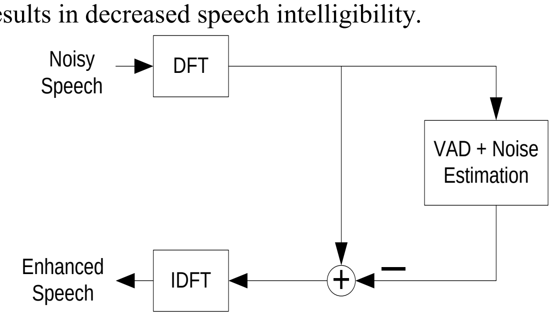

Fig. 2 Block diagram of basic spectral subtractive-type algorithms

To overcome the fixed subtraction parameters used in the spectral oversubtraction method, a multiband spectral subtraction method that divides the noisy speech signal into a number of nonoverlapping bands is proposed, and denoising is performed by readjusting the oversubtraction factors in each band. Because real-world noises are highly nonstationary, the musical noise problem is not removed completely and results in decreased intelligibility. Many other spectral subtractive-type algorithms have also been developed, but no algorithm is highly effective in improving intelligibility.

DOI: 10.46300/9106.2020.14.45 Volume 14, 2020

B. Wiener Methods

The speech and noise spectral probabilistic properties are incorporated in the form of Wiener filtering methods (i.e., the adaptive Wiener, two-stage mel-warped Wiener, and Wiener Scalart) to reduce musical noise [29], [30]. These algorithms assume that speech is a stationary signal and requires a fixed frequency response at all frequencies. Therefore, Wiener filtering methods are also not effective for increasing speech intelligibility.

The basic block diagram of generalized Wiener filtering is shown in Fig. 3. The experimental results prove that, compared with spectral subtractive-type algorithms, Wiener filtering methods are effective for increasing quality (i.e., SNR) but not speech intelligibility.

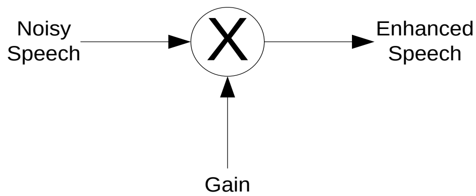

Fig. 3 Block diagram of generalized Wiener filtering

C. Statistical Model-Based Algorithms

The statistical model-based algorithms are highly efficient and historically important for speech enhancement.

Let *x(t)* denote the pure speech signal and *d(t)* denote the noise signal; then, the input noisy speech signal *y(t)* is given by equation (12) and in the frequency domain given in equation (13), where *i* is the frame index and *k* is the frequency point. The *priori* and *posteriori* SNR are given in equations (14) and (15), respectively, as explained in [3] and [4].

$$y(t) = x(t) + d(t), \quad 0 \le t \le T \tag{12}$$

$$Y(i,k) = X(i,k) + D(i,k) \tag{13}$$

$$\zeta_{i,k} = \frac{\lambda_x(i,k)}{\lambda_d(i,k)} \tag{14}$$

$$\gamma_{i,k} = \frac{|Y(i,k)|^2}{\lambda_d(i,k)} \tag{15}$$

where λ_x and λ_d are the pure speech signal variance and noise signal variance, respectively. The Fourier expansion coefficients of the speech and noise process are statistically independent Gaussian random variables [5]; hence, the amplitude of the speech signal DFT coefficient $\hat{X}$ is derived based on the MMSE criterion and estimated as in equation (16):

$$\hat{X} = \frac{\sqrt{\pi}\sqrt{v_{i,k}}}{2\,\gamma_{i,k}} e^{\frac{-v_{t,k}}{2}} \left[(1 + v_{i,k} I_0\left(\frac{v_{i,k}}{2}\right) + v_{i,k} I_1\left(\frac{v_{i,k}}{2}\right)\right] Y_{i,k} \tag{16}$$

where I_0 and I_1 denote the modified Bessel functions of zero and first order, respectively; and where $v_{i,k}$ is defined by

$$v_{i,k} = \frac{\zeta_{i,k}}{1+\zeta_{i,k}} \gamma_{i,k} \tag{17}$$

Many statistical model-based algorithms (i.e., log-MMSE [5], STSA-MMSE [4], and Cohen-MMSE [3]) have been proposed with an efficient gain and a better method of obtaining the *a priori* SNR. The limitation of these types of algorithms is that the estimation of the *a priori* SNR is difficult and mathematically complex. In terms of improving speech quality, i.e., for SNR improvement, statistical model-based algorithms work better but do not produce significant improvement in speech intelligibility.

D. Subspace-Based Algorithms

Subspace-based algorithms estimate clean speech by canceling the noise subspace signal from the noisy signal subspace. Many algorithms are based on the subspace principle. In these types of methods, either singular value decomposition (SVD) [31], [32] or eigenvalue decomposition (EVD) [29], [30], [33], [34] is used in the signal subspace decomposition. The SVD-based method was proposed by Dendrinos *et al.* [31], and an enhanced signal was reconstructed from the information that has the largest singular values. The limitation of this method is that it is applicable only for white noise. An upgraded version was provided by Jensen *et al.* [32] and is effective for colored noise. Ephraim and Van Trees [34] (EV) proposed a subspace-based method using the Karhunen-Loeve transform (KLT). In this method, the signal subspace containing information was modified using a gain function, and the noise subspace was nullified. The results also show that, compared to other methods, the subspace-based method produces superior speech intelligibility.

V. SIMULATION AND EXPERIMENTS

A. Speech Corpus and Noises

The clean speech patterns are taken from the NOIZEUS database, which is composed of 30 balanced sentences recorded by six speakers (three males and three females) [35]. This database is constructed from various additive noises at different SNR levels (i.e., 0 dB, 5 dB, 10 dB, and 15 dB). In this study, all levels of SNRs are evaluated for intelligibility. The noises used in the evaluation are described as airport, babble, car, exhibition, restaurant, street, train, and station. All patterns of the corpus are sampled at 8 kHz. The performances are compared in terms of speech intelligibility measure parameters, such as MOS, *f*AI and STOI, fw-SSNR, and Euclidian distance (EUI).

B. Experimental Results

The purpose of this study is to assess the ability of noise reduction algorithms to enhance speech intelligibility. The Euclidian distance-based speech intelligibility measure parameter is also evaluated and compared with other parameters. The four major categories of speech enhancement algorithms are evaluated for their performance in enhancing speech intelligibility. These methods are presented according to their category.

1) Spectral subtraction-based methods, such as the Berouti-SS, multiband-SS, Boll-SS, parametric-SS, Scalart-SS, and RDC-SS
2) Statistical model-based methods, such as the log-MMSE, STSA-MMSE, and Cohen-MMSE
3) Subspace-based methods, such as the Karhunen-Loeve transform (KLT) and PKLT
4) Wiener-based methods, i.e., the Wiener Scalart algorithm

1) Intelligibility Evaluation at 0 dB Input

Single-channel speech enhancement methods are evaluated at 0 dB input noises. The experimental results for all noises are given in Table I. The time domain KLT method produces the

Table I. Results for 0 dB noise

Methods	Speech intelligibility	Airport	Babble	Car	Exhibition	Restau-rant	Street	Train	Station
SS	**f-AI**	0.1832	0.1679	0.2254	0.24	0.2459	0.268	0.1954	0.1004
Berouti *et*	**STOI**	0.6144	0.6163	0.7696	0.6385	0.6393	0.6878	0.6847	0.7331
***al.* [1]**	**MOS**	0.3008	0.3063	0.3292	0.284	0.3202	0.3234	0.2932	0.3402
	fw-SSNR	5.8484	5.849	5.9032	5.4941	6.074	5.8725	5.9515	6.3312
	Euclidian	0.3545	0.3288	0.3616	0.3309	0.3446	0.353	0.3388	0.3655
Multiband	**f-AI**	0.1583	0.1446	0.1687	0.209	0.1689	0.2196	0.1133	0.0859
SS	**STOI**	0.5827	0.5914	0.6392	0.6219	0.5818	0.615	0.6363	0.6409
Kamath	**MOS**	0.3167	0.306	0.3273	0.2795	0.3141	0.3184	0.2877	0.3318
and Loizou [9]	**fw-SSNR**	5.3908	5.7629	5.4199	5.7274	5.4241	5.7201	5.7874	5.8013
	Euclidian	0.3406	0.3245	0.3447	0.3228	0.3332	0.3355	0.3304	0.3482
Boll [2]	**f-AI**	0.1698	0.189	0.1461	0.1637	0.2168	0.1866	0.1078	0.0645
SS	**STOI**	0.595	0.63	0.6163	0.6497	0.6199	0.5939	0.5865	0.6029
	MOS	0.28	0.2956	0.2515	0.2476	0.2569	0.2562	0.2552	0.2776
	fw-SSNR	5.3854	5.3667	5.5141	5.3585	5.6331	5.7218	5.2417	5.2576
	Euclidian	0.3357	0.3335	0.3203	0.3212	0.3271	0.3118	0.2961	0.3127
Parametric	**f-AI**	0.1228	0.1931	0.1223	0.1547	0.2026	0.1997	0.0946	0.0623
SS	**STOI**	0.548	0.5566	0.6145	0.5802	0.5167	0.5992	0.5453	0.5977
[11]	**MOS**	0.2745	0.2988	0.2886	0.2511	0.2655	0.2706	0.2783	0.2377
	fw-SSNR	5.2502	5.2123	4.9964	5.4612	4.8499	5.4886	5.4115	4.8369
	Euclidian	0.3185	0.317	0.3146	0.3052	0.3027	0.3152	0.2948	0.3177
Scalart and	**f-AI**	0.1267	0.2236	0.1925	0.292	0.1969	0.2512	0.131	0.0877
Filho [10]	**STOI**	0.6275	0.6247	0.663	0.6615	0.6159	0.65	0.6507	0.6749
SS	**MOS**	0.3169	0.3128	0.324	0.2742	0.3	0.3084	0.2907	0.347
	fw-SSNR	5.9484	5.4959	5.8731	5.4986	6.0493	5.5842	5.6009	6.313
	Euclidian	0.3476	0.3346	0.3487	0.3321	0.3404	0.3465	0.6404	0.3525
	f-AI	0.0977	0.1729	0.1127	0.1412	0.1812	0.1599	0.0846	0.0561
Log	**STOI**	0.5655	0.5549	0.6148	0.5873	0.5446	0.6006	0.6003	0.6178
MMSE	**MOS**	0.2965	0.3088	0.301	0.2603	0.2966	0.2904	0.2932	0.3183
[5]	**fw-SSNR**	4.9237	5.3388	5.1477	5.1832	5.2132	5.4461	5.2379	5.2446
	Euclidian	0.3131	0.317	0.3156	0.3052	0.305	0.3057	0.3071	0.3215
MMSE	**f-AI**	0.1175	0.1956	0.1528	0.1712	0.2035	0.1938	0.0959	0.0609
STSA	**STOI**	0.5651	0.5718	0.6159	0.597	0.5659	0.5923	0.5757	0.6008
	MOS	0.318	0.3147	0.311	0.271	0.2905	0.2995	0.2877	0.2626
	fw-SSNR	5.4813	5.5476	5.54	5.6469	5.7512	5.665	5.4502	5.6958
[4]	**Euclidian**	0.3241	0.3246	0.3244	0.3153	0.318	0.3179	0.3114	0.3256
Cohen	**f-AI**	0.172	0.1775	0.1454	0.2803	0.2404	0.2322	0.1178	0.0719
[3]	**STOI**	0.5905	0.6519	0.6326	0.6888	0.5876	0.6121	0.6211	0.6363
MMSE	**MOS**	0.2864	0.3112	0.3016	0.2757	0.2701	0.2613	0.2699	0.3007
	fw-SSNR	5.5518	5.8875	5.3719	6.0655	5.3119	5.1644	4.9409	5.3829
	Euclidian	0.343	0.3441	0.3322	0.3571	0.3319	0.3295	0.3257	0.3442
Wiener	**f-AI**	0.0906	0.1615	0.1039	0.1313	0.1521	0.1358	0.0784	0.0535
Scalart and	**STOI**	0.5345	0.5415	0.5652	0.5325	0.4939	0.5082	0.5338	0.5725
Filho [10]	**MOS**	0.248	0.2873	0.2468	0.2325	0.2526	0.2351	0.2498	0.2539
	fw-SSNR	4.1757	4.6882	4.2596	4.1412	4.0784	3.9424	4.1532	4.268
	Euclidian	0.2901	0.3081	0.2873	0.2793	0.2841	0.2691	0.2728	0.2874
RDC	**f-AI**	0.0593	0.0985	0.0453	0.0969	0.037	0.0296	0.0414	0.021
SS	**STOI**	0.6923	0.6694	0.6898	0.7297	0.5967	0.6321	0.6109	0.6946
	MOS	0.3245	0.3055	0.3126	0.2795	0.3092	0.2728	0.282	0.3219
[6]	**fw-SSNR**	5.3379	5.7622	4.4967	5.1137	5.1996	3.6984	4.4714	5.1024
	Euclidian	0.3565	0.3368	0.347	0.3559	0.3245	0.3182	0.3143	0.3508
	f-AI	0.1263	0.1864	0.1267	0.2836	0.0749	0.1155	0.0669	0.0443
	STOI	0.6942	0.6908	0.7168	0.7206	0.5738	0.6063	0.5682	0.6466
PKLT [8]	**MOS**	0.2692	0.2863	0.263	0.2461	0.2278	0.243	0.2468	0.2414
	fw-SSNR	4.3131	5.3468	5.1162	4.8712	4.0492	2.5294	2.7648	3.9817
	Euclidian	0.3226	0.3472	0.3174	0.3461	0.2836	0.2822	0.2705	0.2838
KLT	**f-AI**	0.1888	0.2359	0.1108	0.3565	0.0739	0.0968	0.0462	0.04

[7]	**STOI**	0.7311	0.6947	0.662	0.773	0.6217	0.6593	0.668	0.6561
	MOS	0.3228	0.3154	0.3206	0.2853	0.2549	0.2381	0.2668	0.2873
	fw-SSNR	6.1794	5.9173	5.4454	6.2527	4.1912	3.4198	3.074	4.5072
	Euclidian	0.361	0.3519	0.3556	0.3823	0.3112	0.302	0.2997	0.3275
Un-Processed Results	**f-AI**	0.2261	0.2349	0.2669	0.3573	0.2659	0.3429	0.2214	0.111
	STOI	0.6741	0.6551	0.6854	0.7275	0.6841	0.7343	0.7143	0.6809
	MOS	0.3224	0.3179	0.3225	0.2824	0.2958	0.3172	0.2932	0.3277
	fw-SSNR	5.5729	6.125	5.2505	5.1968	6.5424	6.1058	5.9636	5.4357
	Euclidian	0.3641	0.3408	0.353	0.3737	0.3611	0.3783	0.3578	0.3606

maximum improvement in airport, babble, and exhibition noise environments. The Berouti-SS method provides maximum speech intelligibility in the presence of car, restaurant, street, train, and station noises. The Euclidian distance values are greater than unprocessed speech in noise environments. Of the twelve algorithms, only two algorithms perform well in improving intelligibility.

2) Intelligibility Evaluation at 5 dB Input

The intelligibility measure parameter values at 5 dB noises are given in Table II. In car and exhibition noise environments, the KLT method produces maximum intelligibility, comparable to unprocessed speech. Compared to other methods, the Boll-SS method provides greater intelligible speech in airport, babble, train, and restaurant environments.

In the case of street and station noises, the Cohen-MMSE method [3], compared to other methods, results in greater improvement in intelligibility because this method utilizes a log-spectral amplitude estimator to effectively reduce the effect of noise degradation in the signal. Table II shows that, in relation to the unprocessed speech intelligibility values for all

Table II. Results for 5 dB noise

Methods	Speech intelligibility	Airport	Babble	Car	Exhibition	Restau-rant	Street	Train	Station
Berouti *et al*. [1] SS	**f-AI**	0.2981	0.2576	0.3003	0.3261	0.3859	0.3569	0.3157	0.3031
	STOI	0.726	0.7049	0.742	0.704	0.7884	0.7191	0.7199	0.7497
	MOS	0.3491	0.3639	0.3667	0.3156	0.3736	0.3523	0.3367	0.3657
	fw-SSNR	7.4247	7.3508	7.5977	6.9752	8.3515	8.2473	7.8207	7.8366
	Euclidian	0.3984	0.3899	0.4044	0.3688	0.4373	0.3946	0.383	0.4087
Multiband SS [9]	**f-AI**	0.2693	0.2157	0.2611	0.2918	0.3799	0.3177	0.2865	0.2822
	STOI	0.6982	0.6701	0.7101	0.6738	0.7719	0.6836	0.7032	0.7116
	MOS	0.3447	0.355	0.3671	0.3219	0.3743	0.3481	0.3442	0.3736
	fw-SSNR	7.1173	6.8852	7.0273	6.9815	7.996	7.7419	7.5199	7.5129
	Euclidian	0.3769	0.3705	3797	0.3517	0.4285	0.376	0.3786	0.3882
Boll [2] SS	**f-AI**	0.3714	0.3559	0.3684	0.4754	0.394	0.3458	0.3923	0.2672
	STOI	0.838	0.8332	0.7979	0.7788	0.8462	0.7268	0.8158	0.7606
	MOS	0.351	0.3942	0.4097	0.3198	0.3783	0.3465	0.3757	0.3486
	fw-SSNR	7.6301	7.8255	8.382	8.0021	8.6628	7.6049	8.775	7.2864
	Euclidian	0.4329	0.4448	0.4328	0.4356	0.445	0.3913	0.4367	0.3946
Parametric SS [11]	**f-AI**	0.2853	0.22	0.2891	0.4392	0.3423	0.2677	0.3484	0.2347
	STOI	0.6923	0.6558	0.7016	0.6909	0.717	0.6281	0.7078	0.6926
	MOS	0.3101	0.3601	0.325	0.3198	0.3627	0.351	0.36	0.3527
	fw-SSNR	6.2644	6.4045	6.8467	6.9423	7.1432	6.6055	7.1647	6.4718
	Euclidian	0.3524	0.3584	0.3563	0.3614	0.3803	0.3481	0.3697	0.3484
Scalart and Filho [10] SS	**f-AI**	0.341	0.3417	0.3575	0.4586	0.3834	0.3755	0.3797	0.3323
	STOI	0.7599	0.7427	0.7602	0.7592	0.7993	0.7111	0.7681	0.7604
	MOS	0.3345	0.3833	0.3906	0.3311	0.3765	0.3582	0.5556	0.3895
	fw-SSNR	6.9587	7.444	7.8284	7.4987	8.1603	8.116	8.2573	7.9824
	Euclidian	0.3946	0.402	0.4051	0.408	0.426	0.3917	0.4012	0.4033
Log MMSE [5]	**f-AI**	0.2346	0.1997	0.2591	0.4001	0.3157	0.24	0.3109	0.1975
	STOI	0.6927	0.6757	0.7034	0.6801	0.7148	0.6479	0.7063	0.6843
	MOS	0.3139	0.3761	0.3567	0.3326	0.3709	0.361	0.359	0.3758
	fw-SSNR	5.8245	6.1859	607038	6.5913	7.0438	6.5409	7.2585	6.2999
	Euclidian	0.3464	0.3507	0.3474	0.3483	0.3689	0.3418	0.3634	0.3421
MMSE STSA [4]	**f-AI**	0.2788	0.2771	0.2882	0.4301	0.3432	0.3072	0.3417	0.2637
	STOI	0.688	0.6683	0.6901	0.6938	0.7406	0.6604	0.7109	0.6906
	MOS	0.3379	0.3917	0.3817	0.3348	0.3783	0.3665	0.3671	0.3928
	fw-SSNR	6.6531	7.1336	7.1794	7.2237	7.6425	7.714	7.7188	7.2274
	Euclidian	0.3584	0.3629	0.3576	0.3676	0.389	0.3603	0.3727	0.3574
Cohen [3] MMSE	**f-AI**	0.3644	0.3559	0.3011	0.4803	0.3798	0.3832	0.3756	0.3274
	STOI	0.766	0.7282	0.79	0.7536	0.8139	0.798	0.781	0.8368
	MOS	0.3429	0.3766	0.3812	0.3253	0.3761	0.3705	0.3594	0.4041
	fw-SSNR	7.4707	7.3481	7.3879	7.709	8.5288	8.31	7.9965	8.0464

DOI: 10.46300/9106.2020.14.45 Volume 14, 2020

	Euclidian	0.4226	0.4121	0.4225	0.4204	0.4436	0.4037	0.417	0.4262
Wiener Scalart and Filho [10]	**f-AI**	0.1691	0.1865	0.198	0.3765	0.3051	0.2281	0.2627	0.1792
	STOI	0.6544	0.6719	0.6701	0.6548	0.6883	0.6383	0.7092	0.6859
	MOS	0.2834	0.3401	0.2903	0.301	0.3501	0.3243	0.3541	0.3145
	fw-SSNR	4.3201	5.3549	5.565	5.5364	6.0506	5.4237	6.3391	5.311
	Euclidian	0.322	0.3369	0.3199	0.3293	0.3548	0.3262	0.3536	0.3253
RDC SS [6]	**f-AI**	0.1442	0.1494	0.1118	0.2084	0.1388	0.171	0.1751	0.1353
	STOI	0.8118	0.7904	0.7873	0.8011	0.8169	0.7676	0.7746	0.7943
	MOS	0.338	0.3544	0.3519	0.3193	0.3616	0.3329	0.3288	0.3468
	fw-SSNR	604586	6.956	6.2522	7.1371	7.2741	7.5126	7.5813	7.1566
	Euclidian	0.4092	0.3988	0.3975	0.4003	0.4061	0.3822	0.3875	0.3985
PKLT [8]	**f-AI**	0.304	0.3106	0.3189	0.4539	0.2901	0.2478	0.3703	0.2713
	STOI	0.8145	0.8151	0.8223	0.8079	0.8186	0.788	0.8048	0.8034
	MOS	0.3296	0.3455	0.3314	0.3033	0.2921	0.2903	0.3327	3322
	fw-SSNR	6.1607	6.4768	6.9921	6.439	7.5189	7.3937	7.9275	6.9152
	Euclidian	0.4006	0.4136	0.3922	0.4283	0.4083	0.3686	0.4273	0.3852
KLT [7]	**f-AI**	0.2857	0.3142	0.3771	0.4812	0.2858	0.2152	0.3917	0.2539
	STOI	0.8007	0.8332	0.8553	0.8374	0.8289	0.7108	0.8082	0.7554
	MOS	0.3365	0.3558	0.413	0.6406	0.3586	0.3195	0.3618	0.3843
	fw-SSNR	6.8598	7.5021	8.4308	8.015	8.2439	7.1799	8.775	7.6065
	Euclidian	0.4294	0.4448	0.4373	0.4625	0.444	0.3926	0.4367	0.4114
Un-Processed Results	**f-AI**	0.3598	0.3634	0.3896	0.4805	0.2869	0.3848	0.3834	0.336
	STOI	0.7899	0.78	0.7763	0.8035	0.8297	0.7471	0.7903	0.7559
	MOS	0.3416	0.3693	0.3537	0.3115	0.3612	0.3375	0.3304	0.3429
	fw-SSNR	7.1762	8.1636	7.2454	7.0135	8.2481	8.3451	8.0673	7.3281
	Euclidian	0.4193	0.4181	0.4124	0.4428	0.4417	0.3985	0.4086	0.3928

Table III. Results for 10 dB noise

Methods	Speech intelligibility	Airport	Babble	Car	Exhibition	Restau-rant	Street	Train	Station
Berouti *et al.* [1] SS	**f-AI**	0.5513	0.5508	0.5135	0.6282	0.5689	0.5624	0.5302	0.36
	STOI	0.9251	0.9054	0.8452	0.9323	0.9058	0.9237	0.8198	0.8053
	MOS	0.4776	0.4512	0.4319	0.4268	0.4595	0.4758	0.4035	0.3855
	fw-SSNR	11.0156	11.273	10.0212	11.919	11.5887	11.7822	10.3666	8.5436
	Euclidian	0.5552	0.5403	0.5082	0.576	0.5663	0.5396	0.4749	0.4351
Multiband SS [9]	**f-AI**	0.5489	0.5462	0.4974	0.6235	0.5674	0.5612	0.5169	0.3096
	STOI	0.8938	0.8832	0.8207	0.9107	0.8969	0.8909	0.8075	0.7661
	MOS	0.462	0.4337	0.4336	0.4155	0.4338	0.474	0.4173	0.3855
	fw-SSNR	10.9623	11.0839	9.5678	11.7767	11.2932	11.4364	908560	7.9983
	Euclidian	0.5508	0.5359	0.4832	0.5679	0.5616	0.538	0.4702	0.4061
Boll [2] SS	**f-AI**	0.5432	0.5406	0.5428	0.6167	0.5514	0.5473	0.5457	0.4598
	STOI	0.9004	0.8906	0.9124	0.9067	0.8911	0.8973	0.9107	0.8422
	MOS	0.4681	0.4271	0.4696	0.4192	0.4544	0.4747	0.4379	0.3872
	fw-SSNR	10.8375	10.7561	10.653	10.6965	10.981	11.0703	10.8549	7.9674
	Euclidian	0.5275	0.5263	0.5271	0.5422	0.5348	0.5267	0.5297	0.4289
Parametric SS [11]	**f-AI**	0.5015	0.4769	0.495	0.5596	0.5026	0.5205	0.5007	0.3226
	STOI	0.8274	0.8102	0.837	0.8109	0.8059	0.8279	0.8011	0.7343
	MOS	0.4509	0.4014	0.4665	0.3928	0.4435	0.4531	0.4154	0.3888
	fw-SSNR	9.3585	8.8349	0.9878	8.9306	9.8533	9.4227	901996	7.1446
	Euclidian	0.4549	0.4367	0.4365	0.4398	0.4588	0.4588	0.4405	0.3716
Scalart and Filho [10] SS	**f-AI**	0.53	0.507	0.527	0.5841	0.5277	0.5352	0.5259	0.4864
	STOI	0.8761	0.8567	0.874	0.8707	0.8563	0.879	0.8646	0.8516
	MOS	0.4478	0.4178	0.4432	0.4075	0.4333	0.4622	0.4198	0.4203
	fw-SSNR	10.1897	9.3795	10.0616	10.1054	10.5164	10.6086	10.8242	8.3029
	Euclidian	0.5063	0.4824	0.4927	0.4898	0.5007	0.5028	0.4881	0.4428
Log MMSE [5]	**f-AI**	0.4642	0.4276	0.4584	0.5206	0.458	0.4831	0.4723	0.2859
	STOI	0.8069	0.7954	0.8118	0.7984	0.7856	0.816	0.7939	0.7453
	MOS	0.4766	0.4198	0.4649	0.4063	0.4571	0.4611	0.437	0.4141
	fw-SSNR	806058	8.0562	8.3401	806492	9.2284	9.0828	8.9789	6.88
	Euclidian	0.4272	0.4143	0.408	0.4123	0.4288	0.4328	0.4233	0.3606
MMSE STSA [4]	**f-AI**	0.4853	0.4542	0.4788	0.5478	0.4852	0.5005	0.4889	0.358
	STOI	0.8209	0.8008	0.8125	0.8086	0.8012	0.8322	0.8012	0.7609
	MOS	0.4641	0.4249	0.4662	0.413	0.4499	0.4627	0.4206	0.4396
	fw-SSNR	9.5272	8.9744	9.2097	9.4014	9.7576	9.8324	9.8627	7.7526
	Euclidian	0.4511	0.4336	0.4288	0.4332	0.4512	0.4542	0.4389	0.3812

Cohen [3] MMSE	**f-AI**	0.5426	0.5326	0.5387	0.6141	0.555	0.5536	0.5434	0.4224
	STOI	0.881	0.8626	0.8734	0.8757	0.8624	0.8763	0.8581	0.7988
	MOS	0.4772	0.436	0.4586	0.4159	0.4328	0.4649	0.4285	0.3946
	fw-SSNR	10.7271	10.4701	10.2181	10.4266	10.6736	10.875	10.595	7.6441
	Euclidian	0.5357	0.514	0.5179	0.5253	0.5273	0.5272	0.507	0.4372
Wiener Scalart and Filho [10]	**f-AI**	0.4527	0.4113	0.4481	0.5071	0.4398	0.4743	0.462	0.2698
	STOI	0.8107	0.7896	0.8072	0.7972	0.7738	0.8116	0.7917	0.7202
	MOS	0.474	0.3864	0.4432	0.3762	0.4244	0.4335	0.411	0.3213
	fw-SSNR	7.6494	6.9945	7.1947	7.4519	0.5794	8.0682	7.8255	5.6371
	Euclidian	0.416	0.4003	0.3928	0.4004	0.4129	0.4208	0.413	0.3402
RDC SS [6]	**f-AI**	0.2513	0.3321	0.2553	0.3567	0.2496	0.2939	0.2749	0.2569
	STOI	0.8852	0.8741	0.8776	0.9021	0.8674	0.8853	0.8636	0.8544
	MOS	0.4135	0.4106	0.4232	0.3981	0.4189	0.4358	0.3721	0.3717
	fw-SSNR	9.1437	9.7213	9.1842	10.5557	10.0973	9.8525	9.4949	7.6391
	Euclidian	0.4738	0.4781	0.4654	0.4892	0.467	0.4712	0.4529	0.4377
PKLT [8]	**f-AI**	0.4904	0.4867	0.4894	0.5872	0.4834	0.4834	0.4956	0.4868
	STOI	0.9063	0.8954	0.8943	0.9166	0.8939	0.9101	0.8937	0.8709
	MOS	0.4278	0.3907	0.3811	0.3853	0.3768	0.4307	0.4226	0.3858
	fw-SSNR	9.2342	9.8517	9.1675	10.1836	9.2893	11.056	9.9423	7.3121
	Euclidian	0.5044	0.5058	0.4891	0.5497	0.5043	0.5145	0.5087	0.4669
KLT [7]	**f-AI**	0.5111	0.4967	0.4796	0.5936	0.4821	0.4798	0.49	0.4884
	STOI	0.8986	0.8852	0.9024	0.9323	0.9056	0.9119	0.9027	0.8921
	MOS	0.4715	0.3996	0.433	0.4253	0.4121	0.4612	0.4167	0.4427
	fw-SSNR	10.9226	10.0507	9.8856	11.1779	10.2893	11.1513	10.7352	8.5959
	Euclidian	0.5525	0.5391	0.5158	0.5693	0.5347	0.5325	0.5278	0.4838
Un-Processed Results	**f-AI**	0.5295	0.5369	0.5362	0.63	0.5547	0.5313	0.5334	0.535
	STOI	0.8875	0.8775	0.8719	0.9152	0.9047	0.8861	0.881	0.8738
	MOS	0.4028	0.3976	0.3977	0.3869	0.4215	0.4362	0.3817	0.3887
	fw-SSNR	9.8534	10.5963	10.0502	10.4018	11.6193	11.1714	10.8161	9.8693
	Euclidian	0.5166	0.5146	0.4998	0.5711	0.5534	0.5123	0.502	0.4867

Table IV. Results for 15 dB noise

Methods	Speech intelligibility	Airport	Babble	Car	Exhibition	Restau-rant	Street	Train	Station
Berouti *et al.* [1] SS	**f-AI**	0.6485	0.605	0.6277	0.6447	0.6577	0.5958	0.5958	0.6813
	STOI	0.9598	0.9521	0.959	0.9557	0.9452	0.9369	0.9369	0.9569
	MOS	0.5635	0.5374	0.598	0.5043	0.5292	0.488	0.488	0.4961
	fw-SSNR	14.3136	13.7854	14.0339	13.8722	14.4848	13.5211	13.5211	14.1595
	Euclidian	0.6591	0.6354	0.6514	0.6468	0.6457	0.6075	0.6075	0.6611
Multiband SS [9]	**f-AI**	0.6438	0.603	0.6257	0.6414	0.6557	0.5296	0.5926	0.6763
	STOI	0.9444	0.9518	0.9456	0.9496	0.9436	0.9289	0.9289	0.9446
	MOS	0.5397	0.5404	0.5545	0.5319	0.5325	0.5082	0.5082	0.4921
	fw-SSNR	14.0158	13.3274	13.7511	13.4899	14.1079	12.8859	12.8859	13.6903
	Euclidian	0.6514	0.629	0.6387	0.636	0.6396	0.5996	0.5996	0.6419
Boll [2] SS	**f-AI**	0.6288	0.5949	0.6187	0.632	0.6534	0.6012	0.6012	0.6765
	STOI	0.9469	0.9481	0.9463	0.9482	0.9482	0.9484	0.9484	0.9388
	MOS	0.5592	0.5599	0.5791	0.5479	0.5437	0.5192	0.5192	0.4884
	fw-SSNR	12.7015	12.5655	12.4138	12.032	13.1254	12.6851	12.6851	12.0906
	Euclidian	0.5949	0.5894	0.5964	0.6006	0.6009	0.6088	0.6088	0.5983
Parametric SS [11]	**f-AI**	0.5836	0.547	0.5684	0.5776	0.6014	0.547	0.547	0.6001
	STOI	0.8638	0.8626	0.8609	0.8523	0.864	0.862	0.862	0.8365
	MOS	0.5272	0.5145	0.5817	0.5107	0.504	0.5196	0.5196	0.4385
	fw-SSNR	10.8461	10.7994	10.834	10.4685	11.1463	11.2727	11.2727	9.9228
	Euclidian	0.5111	0.5056	0.4967	0.4878	0.517	0.5097	0.5097	0.4697
Scalart and Filho [10] SS	**f-AI**	0.604	0.5623	0.5918	0.5974	0.6181	0.5685	0.5685	0.6287
	STOI	0.9116	0.9192	0.9126	0.9106	0.9121	0.9138	0.9138	0.8947
	MOS	0.513	0.5097	0.5357	0.5133	0.5089	0.5077	0.5077	0.4551
	fw-SSNR	12.0919	11.5968	12.0918	11.7517	12.1052	12.3957	12.3957	10.306
	Euclidian	0.5696	0.5558	0.5599	0.5486	0.565	0.5616	0.5616	0.5199
Log MMSE [5]	**f-AI**	0.532	0.5093	0.5309	0.5418	0.5584	0.5088	0.5088	0.5574
	STOI	0.8531	0.8499	0.8428	0.8391	0.8588	0.8455	0.8455	0.8183
	MOS	0.5521	0.5369	0.544	0.5249	0.5018	0.5349	0.5349	0.4396
	fw-SSNR	9.9556	9.9003	9.808	9.7071	10.2661	10.3662	10.3662	8.7207
	Euclidian	0.482	0.4745	0.4594	0.4548	0.4908	0.4765	0.4765	0.4379

MMSE STSA [4]	**f-AI**	0.5617	0.5286	0.5496	0.5678	0.5838	0.5344	0.5344	0.5826
	STOI	0.8594	0.8587	0.8467	0.8524	0.8676	0.8578	0.8578	0.832
	MOS	0.5482	0.5359	0.5399	0.5091	0.5063	0.5125	0.5125	0.452
	fw-SSNR	11.0855	11.1488	10.8957	10.6958	11.2715	11.6964	11.6964	9.7919
	Euclidian	0.5073	0.4986	0.4845	0.4813	0.5123	0.5041	0.5041	0.4597
Cohen [3] MMSE	**f-AI**	0.6306	0.5907	0.6131	0.6254	0.6386	0.5985	0.5985	0.66
	STOI	0.9212	0.9256	0.9228	0.9161	0.9135	0.918	0.918	0.9025
	MOS	0.5407	0.5445	0.5763	0.5216	0.5136	0.5315	0.5315	0.4616
	fw-SSNR	12.7844	12.3432	12.2711	11.6659	12.5415	12.8568	12.8568	11.6487
	Euclidian	0.6112	0.5936	0.6024	0.5876	0.599	0.6034	0.6034	0.5732
Wiener Scalart and Filho [10]	**f-AI**	0.5144	0.4994	0.5186	0.5284	0.5452	0.4946	0.4946	0.5385
	STOI	0.8479	0.8449	0.8417	0.8325	0.88517	0.8381	0.8381	0.8083
	MOS	0.5136	0.4985	0.5223	0.483	0.47	0.4981	0.4981	0.3909
	fw-SSNR	8.6798	8.8244	8.5455	8.4896	9.3418	9.0199	9.0199	8.0677
	Euclidian	0.4659	0.4592	0.4433	0.4384	0.4769	0.459	0.459	0.4251
RDC SS [6]	**f-AI**	0.3606	0.3946	0.3593	0.3705	0.3894	0.4161	0.4161	0.4391
	STOI	0.9287	0.9356	0.9277	0.9357	0.9392	0.9354	0.9354	0.9361
	MOS	0.4984	0.4906	0.4901	0.4626	0.5147	0.4736	0.4736	0.4817
	fw-SSNR	11.913	12.2716	11.6392	11.2421	12.8564	11.5022	11.5022	12.1295
	Euclidian	0.5362	0.547	0.531	0.5295	0.5432	0.5392	0.5392	0.5459
PKLT [8]	**f-AI**	0.5744	0.5818	0.5961	0.6186	0.6292	0.593	0.593	0.6267
	STOI	0.9466	0.9516	0.943	0.9442	0.9556	0.9501	0.9501	0.9436
	MOS	0.4752	0.5329	0.4679	0.4642	0.4963	0.479	0.479	0.4335
	fw-SSNR	12.4443	12.8537	12.1614	11.5235	12.8886	11.0246	11.0246	11.9245
	Euclidian	0.5853	0.6185	0.5942	0.6106	0.6207	0.6195	0.6195	0.6118
KLT [7]	**f-AI**	0.5934	0.6097	0.6012	0.6543	0.6619	0.6071	6089	0.6423
	STOI	0.9585	0.9568	0.957	0.9644	0.9603	0.9592	0.9592	0.95
	MOS	0.5357	0.5615	0.5943	0.561	0.5451	0.5392	0.5439	0.445
	fw-SSNR	13.0993	13.8292	13.2884	13.9148	14.6575	13.6047	13.6547	13.2314
	Euclidian	0.6264	0.6454	0.6353	0.6495	0.6481	0.6402	0.6402	0.6376
Un-Processed Results	**f-AI**	0.6426	0.5891	0.6146	0.6423	0.6462	0.6048	0.6048	0.6822
	STOI	0.9498	0.9427	0.9366	0.9511	0.9505	0.9521	0.9521	0.9494
	MOS	0.4871	0.4881	0.4672	0.4453	0.5018	0.4546	0.4546	0.4833
	fw-SSNR	13.3299	13.5476	12.8686	12.7238	14.8712	14.1625	14.1625	14.5885
	Euclidian	0.64	0.609	0.6154	0.6242	0.6359	0.6169	0.6169	0.6821

noise environments, none of the methods is highly effective in terms of speech intelligibility improvement.

3) Intelligibility Evaluation at 10 dB Input

Table III shows the intelligibility values of the parameters before and after processing the speech signal in the presence of various noises. Two types of algorithms, i.e., the spectral subtraction and subspace methods, are important for intelligibility improvement. For airport, babble, exhibition, restaurant, and street noises, the Berouti-SS method produces the maximum improvement in intelligibility, and for other noises, the KLT method is more effective at a 10 dB input.

4) Intelligibility Evaluation at 15 dB Input

The values provided in Table IV show that the KLT method produces more intelligible speech at the 15 dB input than other methods, except in the airport, car, and station environments. For these three noise environments, the Berouti-SS method is more effective.

VI. Future Directions and Issues

Developing a method that can improve speech intelligibility to a greater extent than unprocessed speech intelligibility is needed because none of the available speech enhancement methods produce a speech intelligibility value that is better than the unprocessed speech intelligibility value.

In this analysis, the phase information of the signal is not considered in any algorithm. Therefore, we must design an effective algorithm that also considers the phase information in speech enhancement.

Because the subjective intelligibility measure is highly expensive and time consuming, an effective objective speech intelligibility measure is also required.

VII. Conclusions

This study presents the intelligibility measure parameters and speech intelligibility values produced by thirteen widely used speech enhancement algorithms for eight noises (airport, babble, car, exhibition, restaurant, street, train, and station) at four input SNR levels (0 dB, 5 dB, 10 dB, and 15 dB). From the speech intelligibility parameters values, the following conclusions are obtained:

- Most of the single-channel speech enhancement algorithms cannot significantly improve speech intelligibility.
- Only two types of algorithms, i.e., spectral subtractive and subspace, significantly improve intelligibility.
- The processed speech signal intelligibility values are not significantly higher than the unprocessed speech signal intelligibility values.

ACKNOWLEDGMENTS

This research was funded by Kuwait University under project grant number QE09/16.

References

[1] M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," in *ICASSP'79. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Washington, DC, USA, 1979, pp. 208–211.

[2] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 27, no. 2, pp. 113–120, Apr. 1979.

[3] I. Cohen, "Speech enhancement using a noncausal a priori SNR estimator," *IEEE Signal Process. Lett.*, vol. 11, no. 9, pp. 725–728, Sep. 2004.

[4] Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 32, no. 6, pp. 1109–1121, Dec. 1984.

[5] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 2, pp. 443–445, Apr. 1985.

[6] H. Gustafsson, S. E. Nordholm, and I. Claesson, "Spectral subtraction using reduced delay convolution and adaptive averaging," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 8, pp. 799–807, Nov. 2001.

[7] H. Yi and P. C. Loizou, "A generalized subspace approach for enhancing speech corrupted by colored noise," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 4, pp. 334–341, Jul. 2003.

[8] F. Jabloun and B. Champagne, "Incorporating the human hearing properties in the signal subspace approach for speech enhancement," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 6, pp. 700–708, Nov. 2003.

[9] S. Kamath and P. Loizou, "A multi-band spectral subtraction method for enhancing speech corrupted by colored noise," in *2002 IEEE Int. Conf. Acoust., Speech, Signal Process.*, Orlando, FL, USA, 2002, pp. IV–4164.

[10] P. Scalart and J. V. Filho, "Speech enhancement based on a priori signal to noise estimation," in *1996 IEEE Int. Conf. Acoust., Speech, and Signal Process. Conf. Proc.*, Atlanta, GA, USA, 1996, pp. 629–632.

[11] S. B. Lim, T. Y. Chow, J. S. Chang, and T. C. Tuan, "A parametric formulation of the generalized spectral subtraction method," *IEEE Trans. Speech Audio Process.*, vol. 6, no. 4, pp. 328–337, Jul. 1998.

[12] R. K. Kandagatla and P. Subbaiah, "Speech enhancement using MMSE estimation of amplitude and complex speech spectral coefficients under phase-uncertainty," *Speech Commun.*, vol. 96, pp. 10–27, Feb. 2018.

[13] M. J. Cevette, J. Stepanek, M. C. Killion, and C. J. Aldous, "Speech intelligibility enhancement system," Google Patents, 2020.

[14] Y. Zhou, C. Bao, and R. Cheng, "GSC based speech enhancement with generative adversarial network," in *2019 Asia-Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA ASC)*, Lanzhou, China, 2019, pp. 901–906.

[15] P. Pertilä, "Data-dependent ensemble of magnitude spectrum predictions for single channel speech enhancement," in *2019 IEEE 21st Int. Workshop Multimedia Signal Process. (MMSP)*, Kuala Lumpur, Malaysia, 2019, pp. 1–5.

[16] B. M. Mahmmod, A. R. Ramli, T. Baker, F. Al-Obeidat, S. H. Abdulhussain, and W. A. Jassim, "Speech enhancement algorithm based on super-Gaussian modeling and orthogonal polynomials," *IEEE Access*, vol. 7, pp. 103485–103504, Jul. 2019.

[17] Y. Wakabayashi, T. Fukumori, M. Nakayama, T. Nishiura, and Y. Yamashita, "Phase reconstruction method based on time-frequency domain harmonic structure for speech enhancement," in *2017 IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, New Orleans, LA, USA, 2017, pp. 5560–5564.

[18] E. Holmes, G. To, and I. Johnsrude, "How do voices become familiar? Speech intelligibility and voice recognition are differentially sensitive to voice training," *PsyArXiv*, 2020.

[19] J. Ma, Y. Hu, and P. C. Loizou, "Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions," *J. Acoust. Soc. Am.*, vol. 125, no. 5, pp. 3387–3405, May 2009.

[20] Y. Tang, B. M. Fazenda, and T. J. Cox, "Automatic speech-to-background ratio selection to maintain speech intelligibility in broadcasts using an objective intelligibility metric," *Appl. Sci.*, vol. 8, no. 1, p. 59, Jan. 2018.

[21] A. M. Mutawa, "Single channel speech enhancement using a complex spectrum method," *Int. J. Circuits, Syst. Signal Process.*, vol. 13, pp. 667–677, 2019.

[22] Y. Hu and P. C. Loizou, "Evaluation of objective quality measures for speech enhancement," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 1, pp. 229–238, Jan. 2007.

[23] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "An algorithm for intelligibility prediction of time–frequency weighted noisy speech," *IEEE Trans. Audio, Speech, Lan. Process.*, vol. 19, no. 7, pp. 2125–2136, Oct. 2011.

[24] P. C. Loizou and J. Ma, "Extending the articulation index to account for non-linear distortions introduced by noise-suppression algorithms," *J. Acoust. Soc. Am.*, vol. 130, no. 2, pp. 986–995, Aug. 2011.

[25] International Telecommunication Union, "Itu-t recommendation p. 800.1: Mean opinion score (mos) terminology," Technical Report, 2006.

[26] International Telecommunication Recommendation, "Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs," Rec. ITU-T P. 862, 2001.

[27] H. Carvajal and N. Garcia, "Evaluation of the parameters of spectral subtraction techniques and its effects on speech signal quality," in *Symp. Signals, Images Artif. Vision - 2013: STSIVA - 2013*, Bogota, Colombia, 2013, pp. 1–5.

[28] N. Danthi and A. Aswatha, "Single channel speech enhancement technique—an implementation on

MATLAB," in *2017 Int. Conf. Innovative Mechanisms Industry Appl. (ICIMIA)*, Bangalore, India, 2017, pp. 369–374.

[29] U. Mittal and N. Phamdo, "Signal/noise KLT based approach for enhancing speech degraded by colored noise," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 2, pp. 159–167, Apr. 2000.

[30] A. Rezayee and S. Gazor, "An adaptive KLT approach for speech enhancement," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 2, pp. 87–95, Mar. 2001.

[31] M. Dendrinos, S. Bakamidis, and G. Carayannis, "Speech enhancement from noise: A regenerative approach," *Speech Commun.*, vol. 10, no. 1, pp. 45–57, Feb. 1991.

[32] S. H. Jensen, P. C. Hansen, S. D. Hansen, and J. A. Sorensen, "Reduction of broad-band noise in speech by truncated QSVD," *IEEE Trans. Speech Audio Process.*, vol. 3, no. 6, pp. 439–448, Nov. 1995.

[33] Y. Hu and P. C. Loizou, "A subspace approach for enhancing speech corrupted by colored noise," in *2002 IEEE Int. Conf. Acoust., Speech, Signal Process.*, Orlando, FL, 2002, pp. I-573-I-576.

[34] Y. Ephraim and H. L. Van Trees, "A signal subspace approach for speech enhancement," *IEEE Trans. Speech Audio Process.*, vol. 3, no. 4, pp. 251–266, Feb. 1995.

[35] P. Loizou, "NOIZEUS: A noisy speech corpus for evaluation of speech enhancement algorithms," *Speech Commun.*, vol. 49, pp. 588–601, 2017.

Dr. Mutawa received his Ph.D. from Syracuse University, New York (1999) in artificial intelligence, and his current research interests are in robotics, expert systems artificial intelligence, signal processing, pattern recognition, deep learning and e-learning. Dr. Mutawa served as Director of the Office of Engineering Education at the College of Engineering and Petroleum, Kuwait University, from 2004-2012 and as Assistant Vice President for Academic Services for computer systems and distant learning from 2012-2014. Currently, Dr. Mutawa is Vice President of the Arab Robotic Association from 2016-present; Founder and President of ROBOTEC consultations in IT and Technologies 2016-present; and Director of the Office of E-Services and Engineering Education at the College of Engineering and Petroleum, Kuwait University, 2017-present.