Single Channel Speech Enhancement using Wiener Filter and Compressive Sensing

Amart Sulong[1], Teddy Surya Gunawan[2], Othman O. Khalifa[3], Mira Kartiwi[4], Hassan Dao[5]

[1,2,3]Department of Electrical and Computer Engineering, International Islamic University Malaysia, Malaysia
[4]Department of Information Systems, International Islamic University Malaysia, Malaysia
[5]Institute of Information Technology, University Kuala Lumpur, Malaysia

Article Info	ABSTRACT
	The speech enhancement algorithms are utilized to overcome multiple limitation factors in recent applications such as mobile phone and communication channel. The challenges focus on corrupted speech solution between noise reduction and signal distortion. We used a modified Wiener filter and compressive sensing (CS) to investigate and evaluate the improvement of speech quality. This new method adapted noise estimation and Wiener filter gain function in which to increase weight amplitude spectrum and improve mitigation of interested signals. The CS is then applied using the gradient projection for sparse reconstruction (GPSR) technique as a study system to empirically investigate the interactive effects of the corrupted noise and obtain better perceptual improvement aspects to listener fatigue with noiseless reduction conditions. The proposed algorithm shows an enhancement in testing performance evaluation of objective assessment tests outperform compared to other conventional algorithms at various noise type conditions of 0, 5, 10, 15 dB SNRs. Therefore, the proposed algorithm significantly achieved the speech quality improvement and efficiently obtained higher performance resulting in better noise reduction compare to other conventional algorithms.

Corresponding Author:

Teddy Surya Gunawan,
Department of Electrical and Computer Engineering,
International Islamic University Malaysia,
Jalan Gombak, 53100 Kuala Lumpur, Malaysia.
Email: tsgunawan@iium.edu.my

1.　INTRODUCTION

In the modern era, the advancement in technology has brought about great benefits to human beings and their daily lives. Today innovation technologies, signal processing is one of the most powerful sources for modern designed engineering that capable of realizing various applications in their real implementations from theoretical aspect point of view to counterpart in different application areas. There are always a trade-off between noise reduction and signal distortion. Most of research found that more noise reduction is always accompanied by more signal distortion [1], [2]. The main challenge of the speech enhancement process is to design effective algorithms to suppress the noise without introducing any possibility of perceptual distortion into the speech signal [1], [3]. Research and investigations on speech enhancement problem have been growing at a rapid rate that cover a broad spectrum of constrains, application, and issues. The challenging work for enhancing noisy speech is on single microphone and the speech problem that was degraded by the noise and remains widely open for investigation [3], [4]. Such problem is well known as single-channel speech enhancement and considered as the most difficult task [1], [5]. This is because of fact that the noise

and speech are perceived as within the same channel by assuming no access to reference noise where the improvement of the speech signal-to-noise ratio (SNR) is target of most techniques.

Most of the speech enhancement techniques have concentrated principally on statistically uncorrelated and independent additive noise [3], [5]. However, the design of effective algorithms that can combat additive noise while producing high quality and improved speech signal is limited. Thus, the studies of additive noise in various types of applications and their related behavior are crucial endeavors. Most of the literatures focus on the difference of the noise sources in terms of temporal and spectral characteristics, and the range of the noise levels that may be encountered in real life [1]. Many existing researches on speech enhancement have based relatively on samples of speech quality measurements which has made it impossible to carry out satisfactory studies. This aspect of study may suggest a better understanding of the related characteristics with a great number of the noisy speech date available for the speech at various dB SNR environments [1].

Concerns have been expressed about speech enhancement approaches. However, there has been a few researches so far that seek possible solution to the speech enhancement based on compressive sensing (CS) technique. Consequently, the question remain whether it can achieve suitable high improvement in both its performance and quality. Thus, it may be useful to investigate and analyze this new approach of data acquisition which is known as compressive sensing (CS) technique [6]. Its theory assert that one can recover certain signals from far fewer samples or measurements than conventional method that is based on the well-known Shannon/Nyquist sampling theorem [7, 8]. In turn, new type of sampling theory can predict from the sparce signals and be constructed from what previously believed to be incomplete information [6]. This method also provides efficient algorithm which can be used for perfect recovery of the sparse signal [9]. Majority of researches in the CS techniques have been introduced in image processing to provide compressed version of the original image with noiseless distortion [6, 9]. This technique relies mainly on empirical observation that many signals can be well-approximated by sparse expression in terms of suitable basis [6].

2.   LITERATURE SURVEY OF SPEECH ENHANCEMENT

Many literatures have been report [1], [3], [11], [28] and mentioned a widely used single channel speech enhancement based on the short-time spectral magnitude (STSM). In real processing speech enhancement techniques, the algorithm employed a simple principle in which the spectrum of the clean speech estimation signal can be obtained by subtracting a noise estimation spectrum from the noisy speech spectrum conditions.

In general, speech enhancement [1], [12] was contaminated and degraded with additive noise. It is typically attacked by the background noise of uncorrelated speech. This signal was known as *noisy speech* and its spectrualrum can be expressed as follow;

$$y(n) = s(n) + d(n) \text{ and } Y(\omega, k) = S(\omega, k) + D(\omega, k) \tag{1}$$

where $y(n)$, $s(n)$, and $d(n)$ are noisy speech, clean speech, and additive noise respectively with n sample number of the discrete time signal. It is often computed on a frame-by-frame basis. The noisy speech is then calculated in the discrete time domain of the short-time Fourier transform (STFT) in which it is generally non-stationary in nature. Its noisy speech spectrum $|Y(\omega, k)|$, clean speech spectrum $|S(\omega, k)|$, and noise spectrum $|D(\omega, k)|$ are calculated depend on ω and k (ω and k are denoted as frequency response and the frame number respectively). For simplicity, the k term throughout the assumption of a frame segment are dropped. Hence the noisy speech power spectrum can be expressed as follows

$$|Y(\omega)|^2 = |S(\omega)|^2 + |D(\omega)|^2 \tag{2}$$

The enhanced speech estimation in short-time magnitude $|\hat{S}(\omega)|$ can be obtain by subtracting a noise estimate during speech pause, which formulated as follow:

$$|\hat{S}(\omega)|^2 = \begin{cases} |Y(\omega)|^2 - |\hat{D}(\omega)|^2 & \text{if } |Y(\omega)|^2 > |\hat{D}(\omega)|^2 \\ 0 \end{cases} \tag{3}$$

The noise estimation spectrum $\left|\hat{D}(\omega)\right|^2$ is calculated from the averaging frames of the recent speech pauses:

$$\left|\hat{D}(\omega)\right|^2 = \frac{1}{M}\sum_{j=0}^{M-1}\left|\hat{Y}_{SP}(\omega)\right|^2 \tag{4}$$

where M is the number of speech pauses in consecutive frames. Equation (4) is not taken into action when the background noise is stationary and coverage to optimal estimate of noise power spectrum. In addition, Equation (3) can also be consider as filter when its product of the noisy speech spectrum is represented as

$$\left|\hat{S}(\omega)\right|^2 = \left(1 - \frac{\left|\hat{D}(\omega)\right|^2}{\left|Y(\omega)\right|^2}\right)\left|Y(\omega)\right|^2 = \Gamma(\omega)\left|Y(\omega)\right|^2 \text{ and } \Gamma(\omega) = \left\{\max\left(0, 1 - \frac{\left|\hat{D}(\omega)\right|^2}{\left|Y(\omega)\right|^2}\right)\right\}^{\frac{1}{2}} \tag{5}$$

where $\Gamma(\omega)$ is the gain function of spectral subtraction and also known as filter. This gain function $\Gamma(\omega)$ is defined as the magnitude response of $0 \le \Gamma(\omega) \le 1$, therefore it is zero phase filter as shown $\Gamma(\omega)$ in Equation (6). To synthesise results, the enhanced speech signal needs reconstruction. This phase is done by using the noisy phase as the clean speech estimation signal, due to insensitivity of the human auditory system [5], [13]. Subsequently, the enhanced speech in a frame is estimated and the clean speech estimation is then synthesis as $\hat{s}(n) = IFFT\left(\left|\hat{S}(\omega)\right|e^{j\phi(Y(\omega))}\right)$. It synthesis will recover speech estimation waveform by inverse Fast Fourier transforming $IFFT(\bullet)$ using an overlap and add method. Moreover, the subtractive-type algorithms can also be estimated using filter approach dependent on the noisy speech's characteristics and on the noise estimation spectrum that can be expressed as $\hat{S}(\omega) = \Gamma(\omega)\left|Y(\omega)\right|$. This gain function $\Gamma(\omega)$ combine the noise reduction of the proposed method [14] with [15]. In extensive studied [16] and [17] reported that the gain improvement relatively used the parameters (i.e., α, β, and γ respectively). The following gain function is as follow:

$$\Gamma(\omega) = \begin{cases} \left(1 - \alpha\left[\left|\frac{\hat{D}(\omega)}{Y(\omega)}\right|\right]^{\gamma_1}\right)^{\gamma_2} & , if \left[\left|\frac{\hat{D}(\omega)}{Y(\omega)}\right|\right]^{\gamma_1} < \frac{1}{\alpha + \beta} \\ \left(\beta\left[\left|\frac{\hat{D}(\omega)}{Y(\omega)}\right|\right]^{\gamma_1}\right)^{\gamma_2} & otherwise \end{cases} \tag{6}$$

The gain function from Equation (6) is the designed parameter to deal with the tradeoff in noise reduction, residual noise and speech distortion signal. These variation parameter can be described as the free parameter and can be described as follows: a). Over-subtraction factor α $(\alpha > 0)$: to avoid the attenuation of the spectrum more than necessary which leads to the reduced residual noise peaks even though the distortion to the speech signal increased (reducing auditable distortion).; b). Spectral flooring β $(0 \le \beta \ll 1)$: to reduce the background noise whereas the background noise is added but only remaining the minimum value of the background noise to be taken.; c). Exponent γ_1 and γ_2: to determine the sharpness of the transition from changing the gain function, by assigning $\Gamma(\omega) = 1$ (modified spectral component). The modification of the exponent γ_1 and γ_2 parameters of the algorithm and its results are described as follow: in case of magnitude subtraction ($\gamma_1 = 1$ and $\gamma_2 = 1$), in case of power spectral subtraction ($\gamma_1 = 2$ and $\gamma_2 = 0.5$), and in case of Wiener filter ($\gamma_1 = 2$ and $\gamma_2 = 1$) respectively.

In [5], it is mentioned the advantages of the spectral subtraction algorithms as follow; 1) simple and only requiring noise estimation spectrum, and 2) variation of subtraction parameters with highly flexibility. Normally, it employs voice activity detection (VAD) in the form of statistical information of silence region. VAD performance degraded significantly at low signal to noise ration (SNR). However, difficulty emerged when background noise is nonstationary. Their shortcoming perceptually contains the remnant of unnatural noticeable to spectral artifacts known as musical noise in random frequencies. It correctly depends on precise

of noise estimation which is limited by the performance of speech/pause detectors. The algorithm's improvement of using spectral over-subtraction is to minimize the inevitable noise and distortion [5], [16]. Beside that, the algorithm assigned $(0 \leq \beta) << 1$ and $(\alpha \geq 1)$ in which to control the amount of power spectrum in noise subtraction from the power spectrum of the noisy speech in each speech frame [5]. Its spectral floor parameters are used to prevent the cause of spectrum floor from going to below the preset minimum level rather that setting to zero. This algorithm depends on a posteriori segmental SNR and over subtraction factor can be calculate from Equation (8) [5].

$$\left|\hat{S}(\omega)\right|^2 = \begin{cases} \left|Y(\omega)\right|^2 - \alpha\left|\hat{D}(\omega)\right|^2 & if \ \dfrac{\left|Y(\omega)\right|^2}{\left|\hat{D}(\omega)\right|^2} < \dfrac{1}{\alpha + \beta} \ ; \ \alpha = \alpha_0 + (SNR)\left(\dfrac{\alpha_{\min} - \alpha_0}{SNR_{\max}}\right) \\ \beta\left|\hat{D}(\omega)\right|^2 & else \end{cases} \qquad (7)$$

where $\alpha_{\min} = 1$, $\alpha_{\max} = 5$, $SNR_{\min} = -5$, $SNR_{\max} = 20$dB, α_0, $(\alpha_0 \approx 4)$, α at 0 dB *SNR*. In this technique will uniform the noise effects spectrum to the speech and predict the subtraction factors that was subtracted noisy by over-estimate of noise factor spectrum. Speech distortion and remnant musical noise is balanced using the various combinations of over-subtraction factor α and spectral floor parameter β. This parameter is to avoid the trade of the amount of remnant noise and the level of perceived musical noise. If parameter value β is large, it produces auditable noise due to a very little amount of remnant musical noise. If parameter value β is very small then the remnant noise greatly reduced but speech signal is quit annoying by the musical noise. Thus, the suitable design of its parameter value α is set following the Equation (7) and parameter β is set to 0.03. As such, the algorithm can reduce the level of perceived remnant musical noise while the remaining of the background noise is presented and distorted the enhanced speech signal. Many type of research also reported using other domain, e.g. signal subspace approach [1], [3]. It differs from the spectral subtraction by decomposing the noisy speech with Kahunen-Loeve-Transform (KLT) into subspace that occupied primarily by the clean speech vector space signal and noise vector space signal. This method used KLT instead of FFT which is proposed by spectral subtraction. It is then estimated the signal the signal of interest and noise subspace from a subspace of the noisy Euclidean space [1]. In [1] mentioned that there are several different types of the spectral subtraction algorithms family. Accordingly, this spectral subtraction type estimates the speech by subtracting noise estimation from the noise speech or by multiplying the noise spectrum with gain functions, and then combine it with the phase of noisy speech. Some of its examples, in briefly, are spectral over-subtraction, spectral subtraction based on perceptual properties, iterative spectral subtraction, multi-band spectral subtraction, Wiener filtering. Therefore, spectral subtraction types essentially were based on intuitive and heuristically based principles.

In Wiener filter type algorithms, the general idea is to minimize the mean square error criterion and to achieve the optimal filter as mention in [1], [18]. The typical formula of the Wiener filter with noncausal Wiener filter for which the frequency response [1], [18] and its formular can be expressed as follow

$$\Gamma_{A.Wiener}(\omega) = \frac{E\left[\left|S(\omega)\right|^2\right]}{E\left[\left|S(\omega)\right|^2\right] + E\left[\left|D(\omega)\right|^2\right]} = \frac{\left|\hat{S}(\omega)\right|^2}{\left|Y(\omega)\right|^2} \ \text{and} \ \Gamma_{Wiener}(\omega) = \left(\frac{P_s(\omega)}{P_s(\omega) + \alpha P_d(\omega)}\right)^\beta \qquad (8)$$

where $E[\bullet]$ is assigned as signal estimator and parameter α and β is assigned to some constant. These constant referred as parametric Wiener filters in which to obtain their characteristic for speech solution. In Equation (8) assign the α and β are equal to one. Thus, the enhanced speech estimation depends largely on the gain parameter's improvement. The enhanced speech estimation and its gain function is shown in Equation (9). This gain function is largely depend on the power spectrum density of the noise at a certain frequency that attenuates each frequency component.

In [1], [11] reviewed the statistical model based algorithm. Its method is justified by the statistics of speech and noise that are not available and there is no knowledge of the best distortion measure in the perception sense by modification of using Hidden Markov Model (HMM) based enhancement [1]. In general, this method adapted a composite source model by choosing a finite set of statistically independent Gaussian subsources model. This finite set is consider as switch that controlled by a Markov chain. The HMM-based

enhancement systems allow a separation between speech and noise beside that it introduced of a prior information about speech and a modeling of noise lead to an improvement over classical methods, especially at low SNRs and for speech corrupted by nonstationary noise. The limitation os the HMM-based system require a training phase to obtain the speech and noise models. It relatively increase the computational requirement. The evaluation of this stage followed clean speech estimation using Maximum A Posteriori estimation (MAP) [1], [11], Minimum Mean-Square Error (MMSE) estimation (it also known as Epharaim and Mallah's estimator) [19]. This method [19] focused on producing colorless residual noise by introducing the gain function's estimator as a function of *a posteriori SNR* and *a priori SNR*.

$$\left|\hat{S}(\omega)\right|^2 = \Gamma_{Wiener}(\omega)\left|Y(\omega)\right|^2 \quad \text{and} \quad \Gamma_{Wiener}(\omega) = \left[\frac{E\left[\left|S(\omega)\right|^2\right]}{E\left[\left|S(\omega)\right|^2\right] + \alpha E\left[\left|D(\omega)\right|^2\right]}\right]^\beta \quad (9)$$

Later, [18] proposed the modification of a priori signal to noise estimation that leads to the best subjective results and achieved the trade-off between noise reductions with low computational load for real time operations. Moreover, [1] adapted with a non-causal estimator for a priori SNR and a corresponding non-causal to enhance speech signal. This estimator technique produced a higher improvement in segmental SNR, lower log-spectral distortion, and better perceptual evaluation of speech quality assessment tests (PESQ scores based on ITU-T P.862 standard [10]). Besides that, other speech enhancement techniques [1], [4] also introduced. In [16] mentioned the modification of boosting techniques and its adaptation to temporal masking threshold of the human auditory system. This masking threshold depends on human auditory system that typically using in speech and audio coding to lower the bitrate requirement.The gain function was depended on the global forward masking threshold and forward masking threshold in each subband [16]. It acted as the filter operation that expressed in time domain in order to evaluate the noise effects to the speech signal in each subband.

3.   PROPOSED SPEECH ENHANCEMENT ALGORITHM
In this section, Figure 1 shows in the block diagram of the proposed algorithm. This speech enhancement algorithm is designed based on Wiener filter and compressive sensing (CS).

3.1. Noisy Spectrum and Update of Noise Estimate
As shown in Figure 1, the speech signal has been contaminated by noise and it is well-known as noisy speech. With this method, the noisy speech is separate into a frame of 20 milliseconds in which each frame is corresponded to 160 sample per frame by using the sampling rate of 8 kHz. Let noisy speech $y(n)$ as the input signal in term of time domain that consist of the clean speech $s(n)$ and additive noise $d(n)$ of independent source respectively. The equations are restated and simplified in order to make understandable. From Equation (1) and (2), noise estimate [20] with the hypothesis formula can be expressed in Equation (10). The noise estimation will calculate based on frame-by-frame noise estimation of Equation (11). The hypothesis of Equation (10) is update the noise estimate $\sigma_D^2(\omega)$. The rang $\alpha_d (0 \leq \alpha_d \leq 1)$ was assigned for smoothing factor. $H_0(\omega)$ and $H_1(\omega)$ denoted the speech absent and the speech present hypothesis respectively. Hence, the noise estimate $D(\omega)$ obtained from Equation (11) where $p'(\lambda,k) \underline{\Delta} P\left(H_1^{'}(\lambda,k)\,|\,Y(\lambda,k)\right)$ denoted as the speech presence probability of the noise variance that corrupted in high nonstationary noise environments.

$$\sigma_D^2(\omega) = E\left[|D(\omega)|\right]^2 \text{ and } \begin{array}{l} H_0^{'}(\omega): \; \hat{\sigma}_D^2(\omega+1) = \alpha_d\hat{\sigma}_D^2(\omega) + (1-\alpha_d)|Y(\omega)|^2 \\ H_1^{'}(\omega): \; \hat{\sigma}_D^2(\omega+1) = \hat{\sigma}_D^2(\omega) \end{array} \quad (10)$$

$$D(\omega) = \alpha_s(\omega)D(\omega-1) + (1-\alpha_s(\omega))|Y(\omega)|^2 \text{ and } \alpha_s(\omega) \underline{\underline{\Delta}} \alpha_d + (1-\alpha_d)p'(\omega) \quad (11)$$

3.2. SNR Estimator and Wiener Filter
The SNR estimator is represented by observing local *a posteriori SNR* and *a priori SNR* in Equation (12) respectively. This estimator was adapted by using [19] in order to produce colorless residual

noise and to improve the gain function of the Wiener filter. The $\left|\hat{S}_{n-1}(\omega)\right|$ is the previous frame estimation speech, where $SNR_{post} - 1$ is interpreted as instantaneous SNR (SNR_{inst}) while $\lambda = 0.98$ and $P(y) = y$ if $y \geq 0$ and $P(y) = 0\ otherwise$. This Wiener technique in Equation (13) was modified based on [18] to obtain the high amplitude spectrum weight estimate when applying Equation (12) to non-linear optimal gain function of Equation (13) and produced the enhanced speech signal. This modified technique will reduce the mismatch weight of the interested signal. Then, the inverse FFT transformed is synthesis. It also derived under assumption that of a key parameter in the reduction of the noise and improving the speech distortion where the technique given a decision-directed method as low computational load for real time operation.

$$SNR_{post} = \frac{|Y(\omega)|^2}{|D(\omega)|^2} \text{ and } SNR_{prio} = \frac{E\left[|S(\omega)|^2\right]}{|D(\omega)|^2} = \lambda \frac{\left|\hat{S}_{\bar{n}-1}(\omega)\right|^2}{|D(\omega)|^2} + (1-\lambda)\left(P(SNR_{post} - 1)\right) \quad (12)$$

$$\hat{y}(n) = IFFT\left(\Gamma_{Wiener}(\omega) \bullet |Y(\omega)| e^{j\phi(X(\omega))}\right) \text{ and } \Gamma_{Wiener}(\omega) = \sqrt{\frac{SNR_{prio}}{1 + SNR_{prio}}} \quad (13)$$
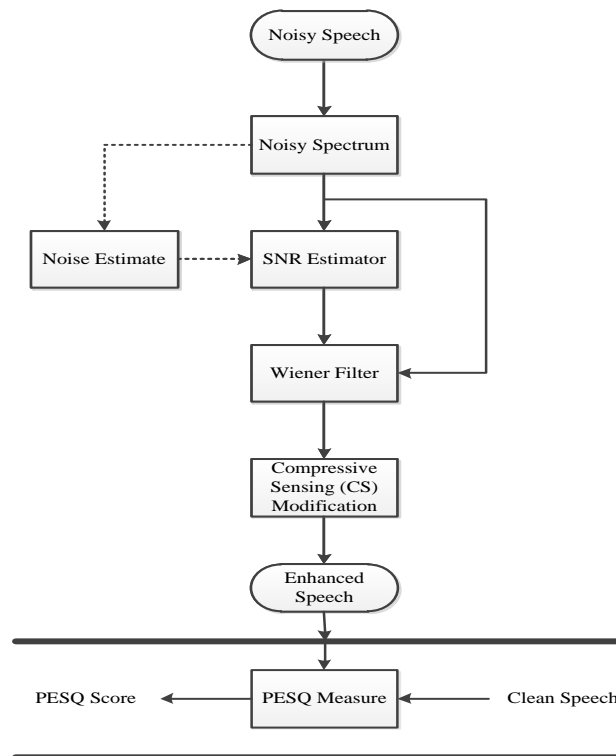

Figure 1. The proposed algorithm based on Wiener filter and compressive sensing technique

3.3. Compressive Sensing Modification

The compressive sensing (CS) technique is also modified. This novel CS approach is fundamentally different from the well-known Shannon sampling theorem [6]. This technique used sampling theory that of selecting the interested signal and recover with almost exact signal reconstruction from noiseless observations [6], [9]. The major advantage of the CS is the recovery predictions of the signals from incomplete measurements (information) that was applied in various applications. Moreover, the CS technique relies on the key efficiency of the empirical observation with well sparse approximations in suitable basis by

only a small amount of nonzero coefficients [6], [9]. The CS method used gradient projection for sparse reconstruction (GPSR) to experimentally investigate the interactive effects of the corrupted noise and obtain better improvement to the listener with noiseless reduction [21]. This method applied based on the weight adaptation ($\hat{y}(n) = Ax + w$) of inverse fast Fourier transform in Equation (14) to achieve high quality noise reduction and enhance speech signal $\hat{s}(n) = Ax$ where the nature of a matric is defined by measurement matrix $A \in R^{m \times n}$. The estimated coefficient $x \in R^n$, and model mismatch $w \in R^m$ is under assumption that $m < n$. To recover the ill-posed condition of signal with sufficient sparse x of unconstrained problem used the GPSR [36] technique, where the spurious components $w \in R^m$ are reduced noiseless distortions. This technique can be expressed as in Equation (14).

$$\min_{x} \frac{1}{2} \|y - Ax\|_2^2 + \tau \|x\|_1 \tag{14}$$

Let the sample y is input weigh signal correlation to predetermined the element of weigh adaptation $\hat{y}(n)$. The determination to exact solution of the sparse recovery y is utilized to regulate the recovery of the estimated coefficient in the predicted signal $\hat{x}$ of x and achieve the improvement of speech quality with noise reduction. This CS modification technique relies on the key efficiency of the empirical observation with well sparse approximation in suitable basis by only small amount of nonzero coefficients [6], [9].

4. EXPERIMENTAL RESULTS AND DISCUSSIONS

PESQ objective assessment test and its percentage improvement in δ was investigate in which to evaluate the enhancement of the speech signal and then compare with the clean speech signal that of a particular assessment signal [1], [24], [29]. The PESQ score has almost correlated with subjective assessment test of a 93.5% correlation while other objective test such as Itakura-saito distortion algorithm, Articulation index, segment SNR, and SNR have correlation assessment test of 59%, 67%, 77%, and 24% respectively [16]. In [16] also introduced the new speech quality assessment test in term of percentage PESQ improvement δ. This percentage improvement can be expressed as shown in Equation (15).

$$\delta = \frac{PESQ_{proc} - PESQ_{ref}}{PESQ_{ref}} \times 100\% \tag{15}$$

Equation (15) mentioned on $PESQ_{proc}$ and $PESQ_{ref}$, it denoted the objective PESQ assessment score of the enhanced speech compared with the clean speech signal while in $PESQ_{ref}$ refers to PESQ score of testing noisy speech performance quality compared with the clean speech respectively.

The four different real artificial added form the noisy speech corpus (NOIZEUS) IEEE standard 1996 [1, 22] These noisy data set used the American English language, where the speech originally sampled at 25 kHz and down-sampled to 8 kHz. The traditional algorithms include Spsub [23], Ssrdc [24], Pklt [25], WnrWt [26], Mmask [27], and mmse [19] respectively. The PESQ assessment test was used to evaluate the main analysis and its significant diferent between the proposed SpEnCS and the other algorithms at various noise type SNRs. Figure 2 clearly indicated the improvement of the proposed algorithm in the waveforms and spectrogram results when compare with traditional algorithms, noisy speech. In Figure 3, the PESQ score in the proposed SpEnCS algorithm outperforms the speech quality compared to overall score with other algorithms of all noise types, i.e. 0, 5, 10, 15 dB SNR.
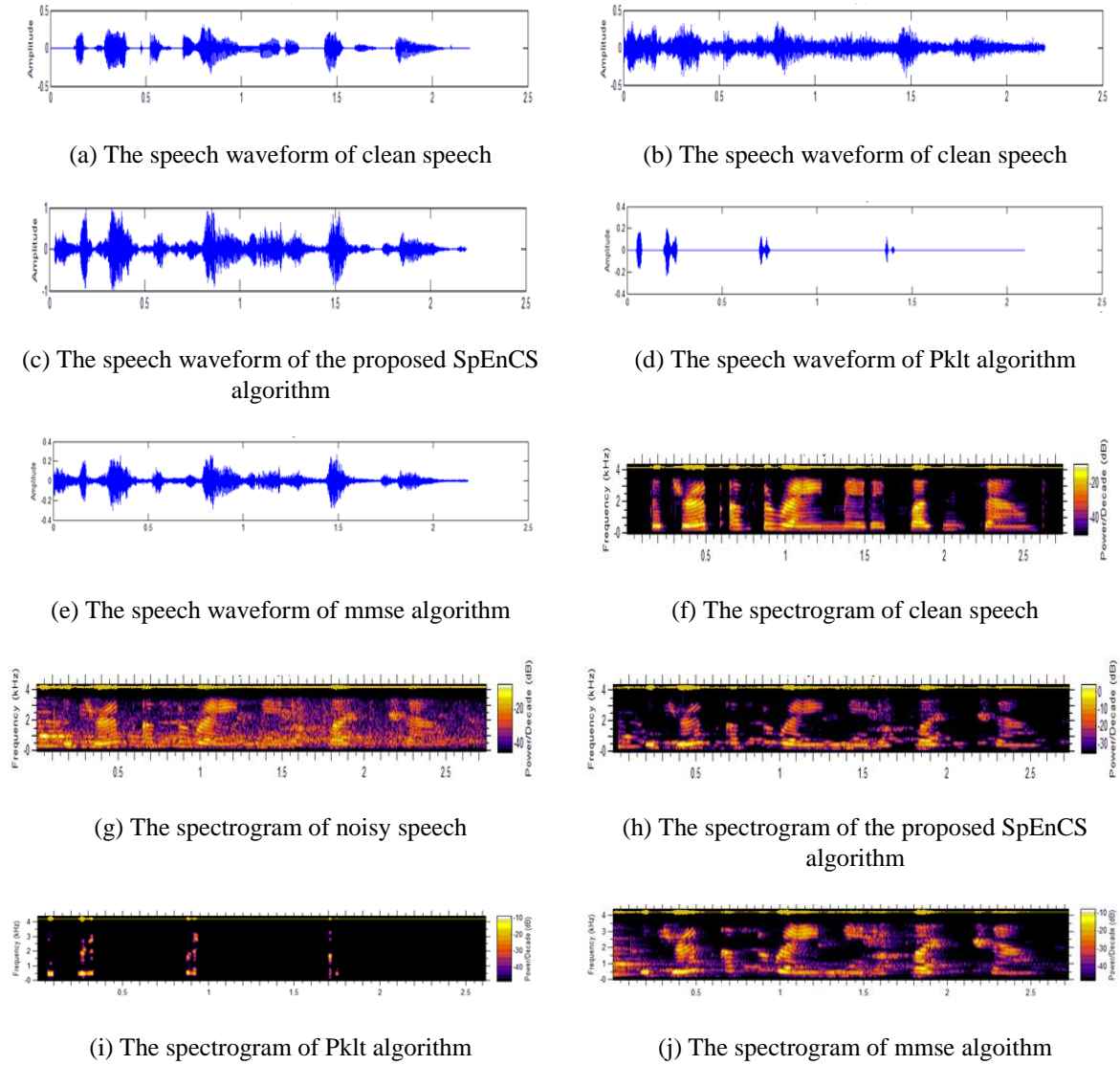
(a) The speech waveform of clean speech

(b) The speech waveform of clean speech

(c) The speech waveform of the proposed SpEnCS algorithm

(d) The speech waveform of Pklt algorithm

(e) The speech waveform of mmse algorithm

(f) The spectrogram of clean speech

(g) The spectrogram of noisy speech

(h) The spectrogram of the proposed SpEnCS algorithm

(i) The spectrogram of Pklt algorithm

(j) The spectrogram of mmse algoithm

Figure 2. The comparison of speech waveform (i.e. a-e) and its spectrogram (i.e. f-j) of the proposed SpEnCS algorithmthat of airport noise at"sp12.wav" at 0 dB SNR

(a) The PESQ assessment score of Airport Noise

(b) The PESQ assessment score of Babble Noise

(c) The PESQ assessment score of Car Noise

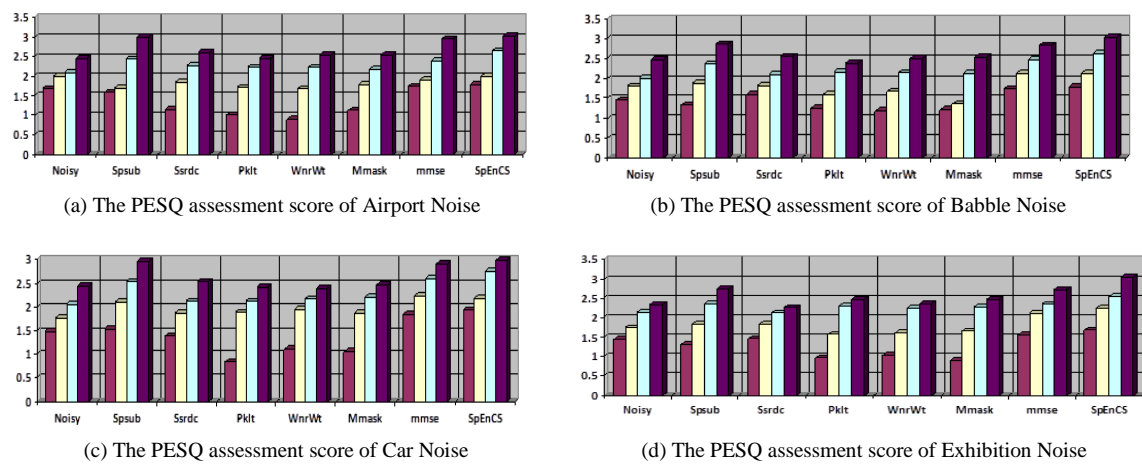(d) The PESQ assessment score of Exhibition Noise

Figure 3. Comparison of PESQ assessment test of the proposed SpEnCS algorithm with other conventional algorithms at 0 dB, 5 dB, 10 dB and 15 dB respectively

Table 1. The PESQ improvement in percentage (%) of the proposed SpEnCS compares with other algorithms

Speech	SNR	PESQ Improvement in Percentage (𝒫%)						
		Spsub	Ssrdc	Pklt	WnrWt	Mmask	mmse	SpEnCS
Airport Noise	0 dB	-0.95	-28.10	-37.30	-38.82	-42.97	10.50	13.55
	5 dB	-6.01	2.20	-5.46	-6.83	-0.61	5.68	10.74
	10 dB	11.21	3.74	1.37	1.78	-0.05	13.17	20.09
	15 dB	18.14	2.88	-3.00	0.83	0.59	16.75	19.36

Speech	SNR	PESQ Improvement in Percentage (𝒫%)						
		Spsub	Ssrdc	Pklt	WnrWt	Mmask	mmse	SpEnCS
Babble Noise	0 dB	-12.49	0.82	-22.32	-26.00	-33.91	10.91	13.57
	5 dB	0.37	-2.72	-16.28	-10.41	-28.07	14.25	14.25
	10 dB	10.86	-1.31	0.89	0.52	-0.05	16.00	22.74
	15 dB	13.40	-0.04	-5.27	-0.67	0.79	12.17	20.25

Speech	SNR	PESQ Improvement in Percentage (𝒫%)						
		Spsub	Ssrdc	Pklt	WnrWt	Mmask	mmse	SpEnCS
Car Noise	0 dB	2.48	-8.37	-44.31	-34.29	-30.19	23.29	29.65
	5 dB	18.93	5.18	6.20	10.14	5.80	25.63	27.65
	10 dB	23.80	3.76	3.76	6.16	7.82	26.74	34.31
	15 dB	21.15	3.60	-1.11	-2.41	1.23	18.74	21.93

Speech	SNR	PESQ Improvement in Percentage (𝒫%)						
		Spsub	Ssrdc	Pklt	WnrWt	Mmask	mmse	SpEnCS
Exhibition Noise	0 dB	-8.91	1.18	-8.48	-28.53	-38.07	8.14	17.82
	5 dB	5.79	5.67	-10.49	-8.83	-4.64	20.63	29.53
	10 dB	10.13	-0.56	7.65	5.09	6.16	9.47	18.81
	15 dB	17.58	-3.33	5.55	1.02	5.21	15.18	29.57

Table 1 shown that the worst case appear with 0dB at all type of noise conditions. Most of PESQ percentage improvement results of the traditional algorithms were below 10% and its improvement remain unconsistency. It was only at mmse algorithm produced comparable results with the proposed SpEnCS algorithm. The overall average of the improvement in the proposed SpEnCS is around 20% to all noisy assessment tests but other algorithm produced less than the proposed algorithms.

5.   CONCLUSIONS AND FUTURE WORKS

A new speech enhancement approach by using Wiener filter and compressive sensing was proposed for enhancing speech degraded by additive noise. The noise estimation is adapted in which to track noise update estimation continuously. The proposed approach is based on the Wiener filter and compressive sensing. The Wiener filter is modified for reducing colorless residual noise before Wiener filter is calculated. Wiener filter is then produced the optimal gain with increasing amplitude spectrum weight estimate and reducing mismatch signal estimate. The compressive sensing later is modified to predict the interested signals from incomplete measurements (signals) and recover with almost signal reconstruction from noiseless observations. Our investigation and evaluation of the proposed algorithms outperforms the other conventional algorithms at various noise types.

ACKNOWLEDGEMENTS

REFERENCES

[1]   P. C. Loizou, "Speech Enhancement: Theory and Practice," CRC Press, 2013.
[2]   R. Sudirga, "A Speech Enhancement System Based on Statistical and Acoustic-Phonetic Knowledge," 2009.

[3]   N. Upadhyay, A. Karmakar, "Speech Enhancement using Spectral Subtraction-type Algorithms: A Comparison and Simulation Study", *Procedia Computer Science*, vol. 54, pp. 574_584, 2015.

[4]   S. V. Vaseghi, "Advanced Digital Signal Processing and noise Reduction," John Wiley & Sons, 2008.

[5]   N. Upadhyay, A. Karmakar, "Single-Channel Speech Enhancement using Critical-Band Rate Scale Based Improved Multi-Band Spectral Subtraction," *Journal of Signal and Information Processing*, vol. 4, no.3, pp. 314-326, Jul. 2013.

[6]   M. Fornasier, H. Rauhut, "Compressive sensing. In Handbook of Mathematical Methods in Imaging", pp. 187-228, Springer New York, 2011.

[7]   M. Unser, *"Sampling-50 years after Shannon",* Proceedings of the IEEE, Vol. 88, no. 4, pp. 569_587, Apr. 2000.

[8]   M. Unser, *"Sampling-50 years after Shannon",* Proceedings of the IEEE, Vol. 88, no. 4, pp. 569_587, Apr. 2000.

[9]   R. G. Baraniuk, "Compressive Sensing", *IEEE Signal Processing Magazine*, vol. 24, no. 4, Jul. 2007.

[10] ITU-T, R.P. 862, "Perceptual Evaluation of Speech Quality (PESQ): An Objective Method for end-to-end Speech Quality Assessment of narrow-band Telephone Networks and Speech Codecs," I*nternational Telecommunication Union-Telecommunication standardization Sector*, 2001.

[11] J. Benesty, S. Makino, J. Chen, "Speech Enhancement," Springer Science & Business Media, Mar. 2005.

[12] S. Boll, "Suppression of Acoustic noise in Speech using Spectral Subtraction," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 27, no. 2, pp. 113-120, Apr. 1979.

[13] D. Wang, J. Lim, "The Unimportance of Phase in Speech Enhancement," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 30, no. 4, Aug. 1982.

[14] M. Berouti, R. Schwartz, J. Makhoul, *"Enhancement of Speech Corrupted by Acoustic noise",* InAcoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'79, vol. 4, pp. 208-211, Apr. 1979.

[15] J. S. Lim, A. V. Oppenheim, *"Enhancement and bandwidth Compression of noisy Speech",* Proceedings of the IEEE, Vol. 67, no. 12, pp. 1586-1604, Dec. 1979.

[16] T. S. Gunawan, "Audio compression and speech enhancement using temporal masking models", Doctoral dissertation, The University of New South Wales, 2007.

[17] N. Virag, "Single Channel Speech Enhancement based on Masking Properties of the Human Auditory System," *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 2, pp.126-137, Mar. 1999.

[18] P. Scalart, *"Speech Enhancement based on a Priori Signal to noise Estimation",* ICASSP-96, IEEE International Conference on, 1996.

[19] Y. Ephraim, D. Malah, "Speech Enhancement using a Minimum-Mean Square Error Short-Time Spectral Amplitude Estimator", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 32, no.6, pp. 1109-1121, Dec. 1984.

[20] S. Rangachari, P. C. Loizou, "A Noise-Estimation Algorithm for Highly Non-stationary Environments", *Speech Communication*, vol. 48, no. 2, pp. 220-231, Feb. 2006.

[21] M. A. Figueiredo, R. D. Nowak, S. J. Wright, "Gradient Projection for Sparse Reconstruction: Application to Compressed Sensing and other Inverse Problems", *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, no. 586-597, Dec. 2007.

[22] E. Rothauser, W. Chapman, N. Guttman, K. Nordby, H. Silbiger, G. Urbanek, M. Weinstock, "IEEE Recommended Practice for Speech Quality Measurements", *IEEE Trans. Audio Electroacoust*, vol. 17, no. 3, pp. 225-246, Sep. 1996.

[23] M. Berouti, R. Schwartz, J. Makhoul, *"Enhancement of Speech Corrupted by Acoustic noise",* InAcoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'79, Vol. 4, pp. 208-211, IEEE, Apr. 1979.

[24] H. Gustafsson, S. E. Nordholm, I. Claesson, "Spectral Subtraction using reduced delay Convolution and Adaptive Averaging", *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 8, pp. 799-807, Nov. 2001.

[25] F. Jabloun, B. Champagne, "Incorporating the Human Hearing Properties in the Signal Subspace Approach for Speech Enhancement", IEEE Transactions on Speech and Audio Processing, vol. 11, no. 6, pp. 700-708, Nov. 2003.

[26] Y. Hu Y, P. C. Loizou, "Speech Enhancement based on Wavelet Thresholding the Multitaper Spectrum", *IEEE Transactions on Speech and Audio Processing*, vol. 12, no. 1, pp. 59-67, Jan. 2004.

[27] Y. Hu Y, P. C. Loizou, "Incorporating a Psychoacoustical Model in Frequency Domain Speech Enhancement", *IEEE Signal Processing Letters*, vol. 11, no. 2, pp. 270-273, Feb 2004.

[28] S. Singh, M. Tripathy, R.S. Anand, "Subjective and Objective Analysis of Speech Enhancement Algorithms for Single Channel Speech Patterns of Indian and English Languages", *IETE Technical Review*, vo. 31, no. 1, pp. 34-46, Jan. 2014.

[29] P. Krishnamoorthy, "An Overview of Subjective and Objective Quality Measures for Noisy Speech Enhancement Algorithms," IETE Technical Review, vol. 28, no. 4, pp. 292-301, Jul. 2011.

BIOGRAPHIES OF AUTHORS

Amart Sulong received his B.S. in computer engineering from Sripatum University (SPU), at Bangkok, in 2000 and his M.S. degrees in Information and Computer Engineering in 2010 and his PhD degree in Engineering in 2017 from International Islamic University Malaysia (IIUM), at Gombak, Selangor. Currently, he is a postdoctoral fellow working on speech signal processing at IIUM. His research interests are the areas of signal processing, speech processing, artificial intelligent, and machine learning.

Teddy Surya Gunawan received his BEng degree in Electrical Engineering with cum laude award from Institut Teknologi Bandung (ITB), Indonesia in 1998. He obtained his M.Eng degree in 2001 from the School of Computer Engineering at Nanyang Technological University, Singapore, and PhD degree in 2007 from the School of Electrical Engineering and Telecommunications, The University of New South Wales, Australia. His research interests are in speech and audio processing, biomedical signal processing and instrumentation, image and video processing, and parallel computing. He is currently an IEEE Senior Member (since 2012), was chairman of IEEE Instrumentation and Measurement Society – Malaysia Section (2013 and 2014), Associate Professor (since 2012), Head of Department (2015-2016) at Department of Electrical and Computer Engineering, and Head of Programme Accreditation and Quality Assurance for Faculty of Engineering (since 2017), International Islamic University Malaysia. He is Chartered Engineer (IET, UK) and Insinyur Profesional Madya (PII, Indonesia) since 2016.

Othman Omran Khalifa received his Bachelor's degree in Electronic Engineering from the Garyounis University, Libya in 1986. He obtained his Master degree in Electronics Science Engineering and PhD in Digital Image Processing from Newcastle University, UK in 1996 and 2000 respectively. He worked in industrial for eight years and he is currently a Professor at Electrical and Computer Engineering Department, International Islamic University Malaysia. His area of research interest is Communication Systems, Information theory and Coding, Digital image / video processing, coding and Compression, Wavelets, Fractal and Pattern Recognition. He published more than 150 papers in international journals and Conferences. He is SIEEE member, IEEE computer, Image processing and Communication Society member.

Mira Kartiwi completed her studies at the University of Wollongong, Australia resulting in the following degrees being conferred: Bachelor of Commerce in Business Information Systems, Master in Information Systems in 2001 and her Doctor of Philosophy in 2009. She is currently an Associate Professor in Department of Information Systems, Kuliyyah of Information and Communication Technology, International Islamic University Malaysia. Her research interests include electronic commerce, data mining, e-health and mobile applications development.

Hassan Dao received the Ph.D. in engineering, 2013, the M.S. Computer and Information engineering in 2007 from International Islamic university Malaysia. He received Bachelor of Computer engineering from Sripatum university, Thailand in 2002. He is currently a senior lecturer within Computer engineering section at Universiti Kuala Lumpur, Malaysia. His research interests include signal processing communication, image processing, radio link design, RF propagation measurement, and wireless communication.