❏   570

Objective and Subjective Evaluations of Adaptive Noise Cancellation Systems with Selectable Algorithms for Speech Intelligibility

Roshahliza M. Ramli[1], Salina Abdul Samad[2], Ali O. Abid Noor[3]
[1]Faculty of Engineering Technology, Universiti Malaysia Pahang, Gambang, Pahang, Malaysia
[2]Centre for Integrated Systems Engineering and Advanced Technologies (INTEGRA), Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia
[3]Department of Communication Engineering, University of Technology-Iraq, Baghdad, Iraq

Article Info

ABSTRACT

Adaptive Noise Cancellation (ANC) systems with selectable algorithms refer to ANC systems that are able to change the adaptation algorithm based on the eigenvalue spread of the noise. These systems can have dual inputs based on the conventional ANC structure or a single input based on the Adaptive Line Enhancer (ALE) structure. This paper presents a comparison of the performance of these two systems using objective and subjective measurements for speech intelligibility. The parameters used to objectively compare the systems are the Mean Square Error (MSE) and the output Signal to Noise Ratio (SNR). For subjective evaluation, listening tests were evaluated using the Mean Opinion Score (MOS) technique. The outcomes demonstrate that for both objective and subjection evaluations, the single input ALE with selectable algorithms (S-ALE) system outperforms that of the dual input ANC with selectable algorithm (S-ANC) in terms of better steady-state MSE by 10%, higher SNR values for most types of noise, higher scores in most of the questions in the MOS questionnaire and a higher acceptance rate for speech quality.

Corresponding Author:

Roshahliza M. Ramli,
Faculty of Engineering Technology,
Universiti Malaysia Pahang,
Gambang, Pahang, Malaysia.
Email: roshahliza@ump.edu.my

1. INTRODUCTION

In many speech applications, the presence of environmental noise poses a challenge and causes a degradation in the performance of the system if not taken into considerations. Environmental noise includes sounds from traffic, industry, construction, babble or any unwanted sounds. Speech are degraded by environmental noise additively resulting in noisy speech signals.

Numerous solutions have been proposed to improve quality and intelligibility of noisy speech signals using adaptive filtering [1–3] and other approaches [4, 5]. Most have reported their findings based on objective measurements such as the SNR of the noisy and cleaned speech as well as other parameters such as the MSE of the adaptive algorithms.

Occasionally, however, objective measurements may fail to adequately predict the quality of speech signals, whereby the speech signal is not inherently preserved due to either external or internal distortions. Speech intelligibility does not imply speech quality whereby a low quality speech signal may be completely comprehensible by listeners but is also judged to be unnatural, harsh, and unpleasant [5]. Therefore, subjective judgment of quality is often used instead, even though it is time-consuming and costly, especially

when a set of discriminating listeners within a consistent listening environment is required. Nevertheless, subjective evaluation provides a more accurate performance assessment since the degree of speech quality and intelligibility are determined by the human auditory system [6].

Most subjective evaluation focus on the MOS method. The MOS procedure is defined by the average score determined across subjects, which subjectively qualifies the perception level of an output speech signal obtained by specific systems [7]. In a typical MOS test, listeners hear sets of processed speech signals and are required to rate them based on a 5-point opinion scale. Moreover, the MOS procedure is utilised as the standard in telecommunication research conducted by the International Telecommunication Union (ITU). One of ITU standards is the ITU-T Recommendation P.85 that defines a subjective assessment method for the quality of speech output devices. This standard explains the practical procedures regarding the test method which include audio prompts, the questionnaire and instructions to respondents on how to interpret the set of words used in each query. Prior studies showed that this method has been used by researchers to assess and compare their proposed systems and obtain user perceptions before the systems are actually deployed for real world implementations. [8–10].

This paper presents the subjective and objective evaluations of two adaptive noise cancellation systems with selectable algorithms. A dual-input system based on conventional ANC structure is compared to a single-input system based on ALE structure. Both systems use the eigenvalue noise spread to determine how to select the adaptation algorithms, where the algorithms have been optimised for the respective structure and has been published separately [2, 11]. However, comparisons between the systems have not been reported for similar algorithms. Moreover, subjective tests performed on both systems as reported in this paper give a more meaningful insight on the performance of the algorithms in terms of of speech intelligibility and quality.

2.   RESEARCH METHOD

ANC systems with selectable algorithms are able to change the adaptation algorithm based in on the eigenvalue spread of the noise. The dual-input system is based on conventional ANC structure [2, 3] while the single-input system is based on ALE structure [11].

2.1. Selectable ANC (S-ANC)

The S-ANC structure is as shown is Figure 1. This dual-input system based on conventional ANC [3]. However, a mechanism is introduced within the adaptive algorithm block in order to switch the adaptive algorithm according to the noise input.
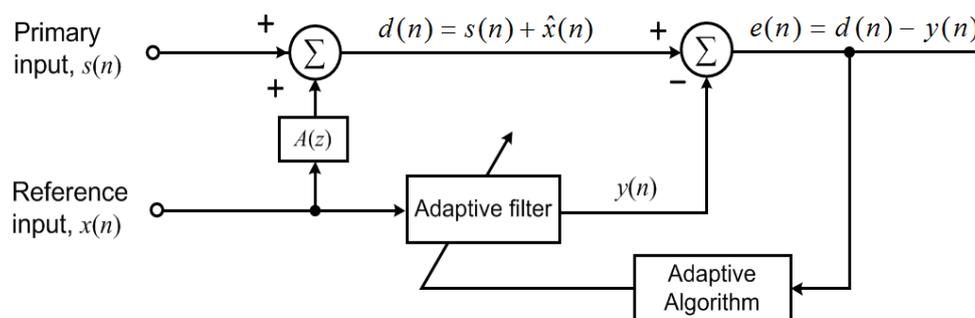
Figure 1. Basic block diagram of ANC

In Figure 1, the primary sensor supplies a signal and a noise uncorrelated with the signal $s(n) + \hat{x}(n)$ as the primary input $d(n)$ to the canceller. The other sensor, called as the reference sensor, receives a noise $x(n)$, which is uncorrelated to $s(n)$ but correlated in some unknown way with the noise $\hat{x}(n)$, to provide the reference input to the adaptive filter. $x(n)$ is transmitted over an unknown channel $A(z)$ and received by the primary sensor, then filtered by the adaptive filter to produce an output $y(n)$ closely resembling $\hat{x}(n)$. Ths signal $y(n)$ is subtracted from $d(n)$ to produce the system output known as error signal, or $e(n)=d(n)–y(n)$. Here, $e(n)$ provides the system control signal and updates the adaptive filter coefficients, which helps to minimize residual noise [12].

2.2. Selectable ALE (S-ALE)

The S-ALE structure is shown in Figure 2. It has a single input based on ALE but a mechanism is introduced in the adaptive algorithm block in order to switch the adaptive algorithm according to the noise input.
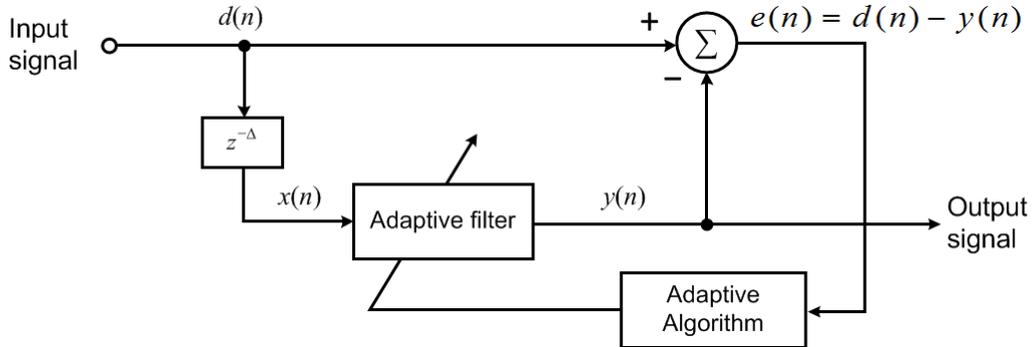
Figure 2. Basic block diagram of ALE

The ALE uses a single sensor to detect input signal $s(n)$ with noise $x(n)$ or the desired signal $d(n)$, which can be expressed as

$$d(n) = s(n) + x(n) \tag{1}$$

The ALE is slightly different from the ANC where it consists of a single sensor and delay $z^{-\Delta}$. This produces a delayed version of input signal, denoted by $d(n-\Delta)$, which de-correlates the noise while leaving the target signal component correlated. The delayed input is then processed with an adaptive filter and subtracted from the $d(n)$ to produce error signal $e(n)$, as expressed by the following equation

$$e(n) = d(n) + y(n) = s(n) + x(n) - y(n) \tag{2}$$

Here, $y(n)$ is the output of the adaptive filter and can be determined based on the filter structure used as the adaptive filter.

2.3. Adaptive Algorithms

The selectable adaptation algorithms for both S-ANC and S-ALE systems are the Normalized Least Mean Squares (NLMS), Affine Projection (AP) and a modified version of AP algorithm called the Dynamic Set-Membership-AP (DSM-AP). The flowchart for algorithm selection is shown in Figure 3. The S-ANC and S-ALE select an adaptive algorithm intelligently based on a flag setting and apply an appropriate algorithm according to the characteristics of noise. In the selectable algorithm mechanism, predefined values are set to differentiate between low, medium and high eigenvalue spread. The predefined value 1 is set to 3 to represent low eigenvalue spread and the predefined value 2 is set to 10 to represent high eigenvalue spread.

If the eigenvalue spread is large, then the signal is considered as an ill-conditioned signal, for which a conventional Least Mean Squares (LMS) adaptive filter will not function properly [11–13]. A swichting between different algorithms according to the noise type is the basis of a selectable ANC system.

For both S-ANC and S-ALE systems, the NLMS algorithm is applied during intervals when the eigenvalue spread is quite low, which is considered to be the best case. The NLMS algorithm works similarly to AP algorithm when the projection order of AP is set low. Meanwhile, the DSM-AP is applied during intervals when the eigenvalue spread is very high, which is regarded as the worst case. Between the NLMS and the DSM-AP, the conventional AP is applied such that it would not cross some predefined projection order in order to keep the computational power as low as possible. The advantage of applying the AP is that the algorithm can efficiently reduce coloured noise with mild computational complexity, as long as the projection order is as low as possible.
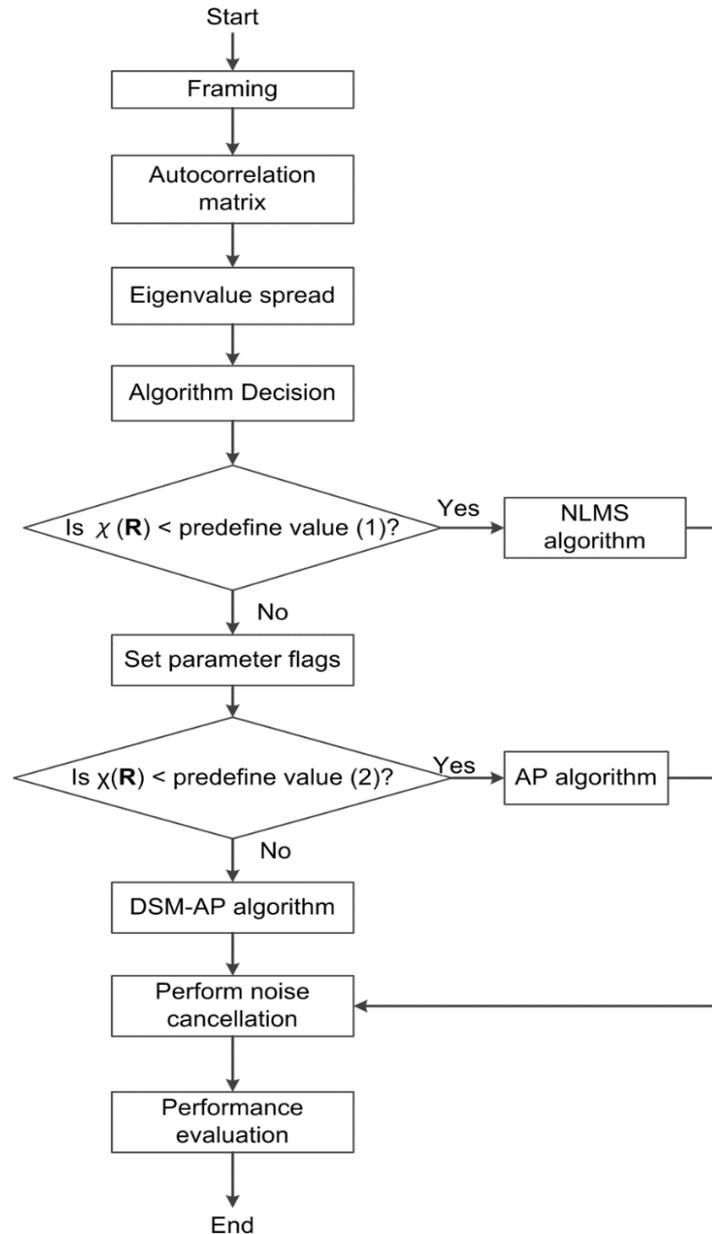
Figure 3. Flow chart for algorithm selection

a. Normalized Least Mean Squares (NLMS)

The main drawback of the conventional LMS which is mainly used in adaptive filters is the difficulty in choosing a suitable value for the step size parameter that guarantees stability. The NLMS has been proposed to overcome this problem in controlling the convergence factor of LMS through modification using a time-varying step size parameter. The NLMS converges faster than the conventional LMS because it employs a variable step size parameter aimed at minimizing the instantaneous output error [1, 12].

The NLMS is defined as an extension of the LMS due to its step size parameter that is inversely proportional to the actual input signal energy. The weight update recursion of NLMS is stated as,

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \frac{\mu}{\gamma + \mathbf{x}^T(n)\mathbf{x}(n)} e(n)\mathbf{x}(n), \tag{3}$$

where $\mathbf{x}(n)$ is the input signal vector, $y(n)$ is the output of an adaptive transversal filter, $e(n)$ is the error signal and $\mathbf{w}(n)$ is the weight vector of the adaptive transversal filter. Meanwhile, μ is the step size parameter

controlling the convergence rate within its suitable range and the value of μ has to be set within 0 and 2. The step size value affects the convergence behaviour of the filter; a too low value of μ leads to an extremely long convergence duration, whereas a too high value causes the algorithm to diverge, thus degrading the error performance of the adaptive filter. A small value of γ is used to avoid possible division by zero. The NLMS has an advantage that exhibits potentially faster convergence speed than that of conventional LMS algorithm for both uncorrelated and correlated input data [14].

b. Affine Projection (AP)

The AP algorithm reuses past and current information in order to increase the speed of convergence in adaptive filter when the input signal is highly correlated [1]. This algorithm can be viewed as an extension or generalization of the NLMS algorithm where the coefficient update of NLMS is interpreted as a one-dimension affine projection. Meanwhile in AP, projections are made in multiple dimensions per coefficient update. With the step size μ as the convergence factor to control convergence, stability and final error, a diagonal matrix $\delta\mathbf{I}$ (δ is a small constant) is added to regularize the inverse matrix in the equation, the coefficient update for the conventional AP algorithm is given as follows

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu\mathbf{X}(n)\left[\mathbf{X}^T(n)\mathbf{X}(n) + \delta\mathbf{I}\right]^{-1}\mathbf{e}(n) \qquad (4)$$

c. Dynamic Set-Membership AP (DSM-AP)

The Dynamic Set-Membership Affine Projection (D-SM-AP) algorithm has been proposed to improve the adaptation process and could be utilized as alternative to high complexity algorithm [15, 16]. The modification of this algorithm is based on using a simplified version of the Set Membership AP (SM-AP) in order to obtain better performance in terms of the convergence rate [12]. The adaptation step-size μ is modified as follows

$$\mu = \frac{\bar{\mu}e(n)}{|x(n)|^2} \qquad (5)$$

Here, $\bar{\mu}$ is provides a dynamically variable step size which possesses a better tracking capability for variable environments where

$$\bar{\mu}(n) = \begin{cases} 1 - \dfrac{\gamma}{|e(n)|}, & |e(n)| > \gamma \\ 0 & \text{otherwise} \end{cases} \qquad (6)$$

Here, the parameter γ is chosen as less than $\sqrt{3\sigma_n^2}$, where σ_n^2 is the variance of the noise. The noise variance is obtained as

$$\sigma^2 = \frac{1}{N}\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2} \quad \text{where} \quad \bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i . \qquad (7)$$

The modified adaptation step size is substituted as the coefficient update for the AP algorithm.

3. EIGENVALUE SPREAD

The mechanism for algorithm selection is based on the calculation of the eigenvalue spread in order to select an adaptive algorithm intelligently for eliminating regular and irregular types of noise from noisy signals. The target application here is speech communications, where the useful signal is corrupted with irregular types of noise that are hard to eliminate using conventional methods. The eigenvalues spread is determined from the autocorrelation matrix $\mathbf{R}$, expressed as:

$$\mathbf{R} = E\left[\mathbf{X}(n)\mathbf{X}^H(n)\right] = \begin{bmatrix} E\left[|x_0(n)|^2\right] & E\left[x_0(n)x_1^*(n)\right] & \cdots & E\left[x_0(n)x_N^*(n)\right] \\ E\left[x_1(n)x_0^*(n)\right] & E\left[|x_1(n)|^2\right] & \cdots & E\left[x_1(n)x_N^*(n)\right] \\ \vdots & \vdots & \ddots & \vdots \\ E\left[x_N(n)x_0^*(n)\right] & E\left[x_N(n)x_1^*(n)\right] & \cdots & E\left[|x_N(n)|^2\right] \end{bmatrix} \tag{8}$$

Here, $\mathbf{X}^H(n)$ is the Hermitian transposition of the input signal vector $\mathbf{X}(n)$ in the time domain $x(n)$, and N is the filter length. The eigenvalues spread is determined from the ratio of the maximum to the minimum eigenvalues of the matrix $\mathbf{R}$. The eigenvalues are denoted by λ_j. We first establish the characteristic equation of $\mathbf{R}$ as follows.

$$\det(\mathbf{R} - \lambda_j\mathbf{I}) = 0 \tag{9}$$

where $\mathbf{I}$ is the identity matrix, and λ_j is given by following diagonal matrix,

$$\lambda_j = \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_M \end{bmatrix} \tag{10}$$

where, λ_1, λ_2, ..., λ_M are the eigenvalues of $\mathbf{R}$, and all of which may not be distinct from each other. Then, the eigenvalue spread of $\mathbf{R}$ is calculated as follows,

$$\chi(\mathbf{R}) = \frac{\max(\lambda_j)}{\min(\lambda_j)} \tag{11}$$

Here, $\max(\lambda_j)$ and $\min(\lambda_j)$ are the maximum and minimum eigenvalues of the autocorrelation matrix, respectively. Using the measurement of $\chi(\mathbf{R})$, the selection mechanism is set for cancelling different types of noise in the target noisy speech signal received by the primary sensor of the noise canceller. The eigenvalue spread is calculated on a frame by frame basis. The algorithm is updated with a new eigenvalue spread value every time a new frame of data is processed by the adaptive algorithm.

4. OBJECTIVE EVALUATIONS

Objective assessment methods are based on mathematical comparisons between the original and processed speech signals. These assessments can give a quick glance on performance of a speech enhancement algorithm or system. For this study, the objective assessment parameters are the MSE and output SNR.

4.1. Mean Square Error (MSE)

MSE is used to show the convergence of the investigated algorithm based on the mean square error function $E[|e^2(n)|]$. Figure 4 shows the comparisons of convergence performance S-ANC and S-ALE with other single adaptation algorithms for the variable noise, where the noise type changes with time. From this result, for the single algorithms, the error decays between 20dB to 60 dB, meanwhile for both S-ALE and S-ANC, the error decays about 100 dB initially, but then increases up to 30% in the middle due to amplitude changes in the noise signal. However, the error of S-ALE decays 10% more and shows that it has a lower steady-state error compared to S-ANC.
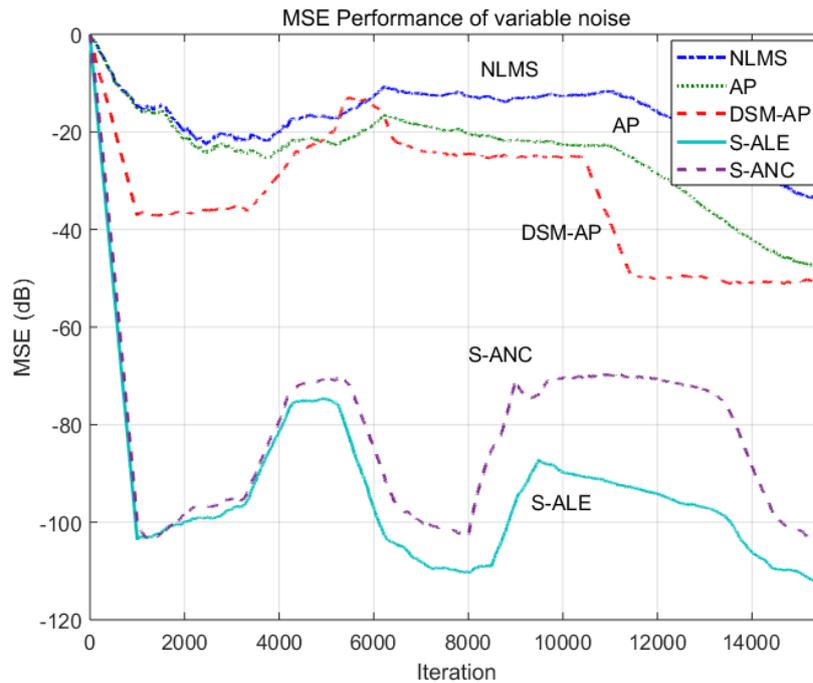
Figure 4. MSE comparisons between S-ANC and S-ALE

4.2. Signal to Noise Ratio (SNR)

Another objective measurement used in this study is the SNR which takes the ratio of signal energy to noise energy expressed in decibel (dB). This measure compares the output signal with the reference or clean signal. A high SNR value indicates a good perceptual quality of the speech. The SNR is given by

$$SNR = 10 \log\left(\frac{\text{Power of processed speech}}{\text{Power of processed speech - power of clean speech}} \right) \tag{12}$$

Table 1 shows the SNR comparisons between S-ANC and S-ALE with the other single algorithms. Although both systems show better SNRs with all noise types compared with the single algorithms, much higher SNR values are obtained for S-ALE compared to S-ANC for all noise types except for babble which show comparable values.

Table 1. SNR Comparisons between S-ANC and S-ALE

Noise	SNR, dB				
Type	NLMS	AP	D-SMAP	S-ANC	S-ALE
White	11.445	12.537	20.041	52.669	66.861
Car	31.712	31.576	29.333	33.884	56.627
Babble	22.825	24.485	29.123	58.369	55.157
Variable	19.427	21.616	40.626	45.522	65.281

5.   SUBJECTIVE EVALUATIONS

Quality and intelligibility of processed speech signals need to be evaluated before implementing the proposed schemes. Subjective listening tests are used to evaluate both S-ANC and S-ALE in order to guage the perception of listeners on the quality of the processed speech signals obtained at the output of the systems. Subjective listening tests are conducted using a standard measure recommended by ITU targeting communication applications, the ITU-T Recommendation P.85 MOS [7]. Wording of questions and scaling of grades were referred to this document and modified based on Viswanathan's study [17].

5.1. Mean Opinion Score (MOS)

Speech messages produced or processed by machines may suffer from certain impairments such as lowered intelligibility of processed speech and reduced pleasantness among listeners. MOS method can be applied to compare several systems by speech output devices using a 5-point scale questionnaire with the rating listed in Table 2. In this work, the subjective listening test using MOS method consisted of questions that assessed overall sound quality, listening effort, comprehension problems, articulation, pronunciation, speaking rate, and pleasantness.

Table 2. Scale of MOS

Rating	Speech quality
5	Excellent
4	Good
3	Fair
2	Poor
1	Unsatisfactory

Respondents were queried on perceived quality of the speech produced at the output of the systems, such as overall impression, listening effort, pronunciation of the speaker, speaking rate, pleasantness in listening, and speech comprehension.

5.2. Experimental Setup

Subjective listening tests were carried out on the validate speech outputs of by S-ANC and S-ALE. A total of 30 respondents aged between 22 to 36 years without any hearing loss participated. Before commencing this assessment, respondents took pre-tests to determine wether they can listen clearly. In the pre-tests, respondents were given three audio files consisting of two monophonic (monaural) and one binaural output sounds. They had to determine from which side of the headphone the sound was emitted. These pre-tests were conducted to verify that respondents have normal hearing ability, i.e., without any major hearing impairments. Only those who passed the pre-test were selected for the main listening test.

During the main subjective listening test assessments, two audio files of speech signals, each obtained from the output of S-ANC and S-ALE, were presented aurally using a computer. The files were played in random order to the respondents. Respondents were instructed to listen to the audio files using a portable stereo headphone without any noise cancelling function connected to a computer. While listening to the audio files, respondents were asked to answer a set of questions via an online form on a tablet. The subjective listening tests were held for approximately 15 minutes in a quiet area. Answers were automatically saved into a spreadsheet file for analysis. The actual scenario of this experiment is depicted in Figure 5.

Figure 5. A respondent taking the subjective listening tests

5.3. Subjective Evaluation Outcomes

Table 3 shows the average score and standard deviation for every question in each audio answered by the respondents. The average scores for S-ALE are higher in all items with smaller standard deviations

compared to S-ANC. This indicates that the responses are widely scattered for S-ANC. This also indicates that the speech processed by S-ALE is more desirable compared to S-ANC.

Table 3. The Analysis of MOS Taken from Respondents

Item	S-ANC		S-ALE	
	Average	Standard deviation	Average	Standard deviation
Overall impression	4.40	0.69	4.44	0.50
Listening effort	4.28	1.22	4.60	0.75
Pronunciation	4.60	0.75	4.56	0.70
Speaking rate	4.32	0.97	4.76	0.59
Pleasantness	4.08	1.02	4.16	0.78
Comprehension	4.20	1.02	4.68	0.55
Articulation	4.32	0.84	4.68	0.55

The findings showed in Table 3 were further confirmed by asking the respondents to state whether they find the quality of the processed speech is acceptable if it were the output of a communications application. This question was answered either 'yes' or 'no', and the results of the survey are illustrated in Figure 6. Most respondents agreed that the speech processed by S-ALE is more acceptable than S-ANC with 92% acceptance rate compared with 80%. It can be concluded from the subjective tests conducted that the S-ALE processed signals are perceived as being better to the listeners compared to those processed by S-ANC.
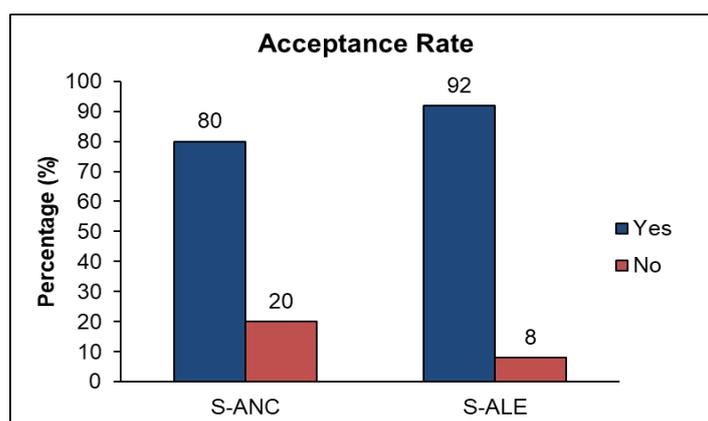
Figure 6. The MOS acceptance rate of perceived speech quality

6. CONCLUSION

Objective and subjective evaluations of single and dual input adaptive noise cancellation systems with selectable algorithms have been presented. The selectable adaptation algorithms for the systems are the NLMS, AP and DSM-AP. The switching between algorithms is triggered by the value of the noise eigenvalue spread. Comparisons between the systems using objective calculated metrics and subjective listening tests with MOS show that, overall, S-ALE outperform S-ANC. In objective evaluations, the S-ALE has better steady-state error and a much higher SNR values for most of the noise types tested compared to S-ANC. In subjective evaluations, S-ALE also obtained higher scores for most items asked in the MOS questionnaire and a higher acceptance rate of perceived speech quality compared to S-ANC. This suggests that for the purpose of implementation and deployment in communication applications where environment noise is prevalent and changing, S-ALE is a good candidate to adaptively cancel out the noise.

REFERENCES

[1] Diniz PSR. Adaptive Filtering: Algorithms and Practical Implementation. 3rd ed. New York: Springer; 2008.

[2] Ramli RM, Noor AOA, Abdul Samad S. Noise cancellation using selectable adaptive algorithm for speech in variable noise environment. Int J Speech Technol. 2017;20(3):535–42.

[3] Sayed AH. Adaptive Filters. Adaptive Filters. 2008.

[4] Devi TM, Kasthuri N, Natarajan AM. Environmental noise reduction system using fuzzy neural network and adaptive fuzzy algorithms. Int J Electron. 2013;100(2):205–26.

[5] Nordholm S, Kellermann W, Doclo S, Välimäki V, Makino S, Hershey JR. Signal Processing Techniques for Assisted Listening. IEEE Signal Process Mag. 2015;32(2):16–7.

[6] Krishnamoorthy P. An Overview of Subjective and Objective Quality Measures for Noisy Speech Enhancement Algorithms. IETE Tech Rev. 2011;28(4):292.

[7] ITU-T. Recommendation P.85 - Telephone Transmission Quality Subjective Opinion Tests. A Method for Subjective Performance Assessment of the Quality of Speech Voice Output Devices. 1994.

[8] Zhenyu C, Kitawaki N, Yamada T, Makino S. Comparison of MOS evaluation characteristics for Chinese, Japanese, and English in IP telephony. In: 4th International Universal Communication Symposium (IUCS), 2010. 2010. p. 112–5.

[9] Waidyanatha N, Wilfred T, Perera K, Silva M. Mean Opinion Score performance in classifying voice-enabled emergency communication systems. In: International Conference on Computer & Information Science (ICCIS), 2012. 2012. p. 676–82.

[10] Bendoumia R, Djendi M. Variable step-size subband backward BSS algorithms for speech quality enhancement. Appl Acoust. 2014;86(0):25–41.

[11] Ramli RM, Noor AOA, Samad SA. Adaptive line enhancer with selectable algorithms based on noise eigenvalue spread. 2016 Int Conf Adv Electr Electron Syst Eng ICAEES 2016. 2016;(ICAEESE):356–61.

[12] Haykin S. Adaptive Filter Theory. Automatica. New Jersey: Prentice Hall; 2002. 920 p.

[13] Manolakis DG, Ingle VK, Kogon SM. Statistical and Adaptive Signal Processing: Spectral Estimation, Signal Modeling, Adaptive Filtering, and Array Processing. Massachusetts : Artech House, Incorporated; 2000.

[14] Douglas SC, Meng TH-Y. Normalized data nonlinearities for LMS adaptation. IEEE Trans Signal Process. 1994;42:1352–65.

[15] Kim SE, Lee JW, Song WJ. A Theory on the Convergence Behavior of the Affine Projection Algorithm. IEEE Trans Signal Process. 2011;59(12):6233–9.

[16] Paul TK, Ogunfunmi T. On the Convergence Behavior of the Affine Projection Algorithm for Adaptive Filters. IEEE Trans Circuits Syst I Regul Pap. 2011;58(8):1813–26.

[17] Viswanathan M, Viswanathan M. Measuring speech quality for text-to-speech systems: development and assessment of a modified mean opinion score (MOS) scale. Comput Speech Lang. 2005;19(1):55–83.