JOURNAL OF APPLIED SCIENCES TECHNOLOGY AND COMPUTING
e-ISSN: 3036-0250

JASTEC

Vol. 1 No. 1 (2024) 39-51
https://publisher.uthm.edu.my/ojs/index.php/jastec

Acoustic Echo Cancellation using Adaptive Filter for Quranic Accent Signals

Noraziahtulhidayu Kamarudin[1*], Syed Abdul Rahman Al Haddad[2], Rauf Hassan Azhari[3], Azli Basiron[4]

[1] *Department of Multimedia, Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, MALAYSIA*
[2] *Department of Computer Communication Engineering, Faculty of Engineering, Universiti Putra Malaysia, MALAYSIA*
[3] *Department of Modern Language, Faculty of Moden Language and Communication, Universiti Putra Malaysia, MALAYSIA*
[4] *Department of Information Security and Web Technology, Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, MALAYSIA*

*Corresponding Author: noraziah@uthm.edu.my
DOI: https://doi.org/10.30880/jastec.2024.01.01.005

Article Info

Received: 23 November 2023
Accepted: 13 February 2024
Available online: 27 February 2024

Keywords

Adaptive filtering, acoustic echo cancellation, recursive least squares, least mean square, affine projection, accuracy rate

Abstract

Quranic recordings and echoed portions of the emphasis are susceptible to signal reverberation, particularly when being listened to in a conference room. Tajweed and Quranic verse rule identification are susceptible to additive noise, which could lower classification accuracy. In order to reflect the most correct rate following pattern categorization, this study suggested the appropriate use of three adaptive algorithms: Affine Projection (AP), Least Mean Square (LMS), and Recursive Least Squares (RLS). For feature extraction, Mel Frequency Cepstral Coefficient is used together with Probabilities Principal Component Analysis (PPCA), K-Neural Network (KNN) and Gaussian Mixture Model (GMM). AP indicates 93.9% for all of the classification algorithm in used, while for LMS and RLS the results are differed varies on different pattern classification algorithm stated whereby with LMS and PPCA classification, 96.9 % for accuracy and 84.8% accuracy for LMS and KNN. While for RLS and GMM, 96.9% was achieved and the results were reduced for both KNN and PPCA. The analysis has resulted for both on accuracies within different filtering algirithm and classification for accuracy and ERLE(dB).Towards this research it is hope will embark more understanding towards echo cancellation and quality of sound recordings that may affected even to the Quranic recordings.

1. Introduction

Unusually big peaks have been produced by frequency response, masking, and emerging peaks from sound systems. Noise can be introduced into music recordings from the surrounding environment, during the recording system when audio signals change, or during the indexing process. These noise signals can disrupt and lower the quality and efficiency of the recording. Quranic recordings may also be affected by echoes, which can interfere with the recording process. The speaker, microphone, and transmission path are among the variables that affect

echo in a room. Voice signals are transmitted and spread via various pathways in relation to the speaker factor. Due to numerous path interruptions [1], echoing elements in the room may cause the voice signal strength to decrease. The voice signal may distort and degrade when spoken by the participants in front of the microphone [2]. The characteristics of a sound signal are its volume, or amplitude, which is measured in decibels (dB), its pitch, which is expressed in frequencies and measured in hertz (Hz), the duration of time, which is expressed in seconds, and all of these combined into one dimension.

The psychoacoustic qualities of sound, or timbre, that are multidimensional in nature, are present in melodies of music. Those that convey distinctive characteristics of the audio file are referred to as audio descriptors. An audio descriptor can be a scalar or one-dimensional variable that provides a range of values for a feature vector [2]. The two primary categories of echo signals are telephone signal echo and acoustic signal echo. Basically, there are two main types of echo signals which are acoustic signal and telephone signal echo. On the other hand, in applications which involve music recognition, the artistic impacts can be easily heard include reverberation and propagation delay[3]. Telephone signal echo and acoustic signal echo are the two main types of echo signals. The two primary categories of echo signals are telephone signal echo and acoustic signal echo. Reverberation and propagation delay, on the other hand, are examples of creative effects that are audible in music recognition applications [3].

In addition, background noise may exist in any sound element, which can cause disturbances or distractions. Background noises may vary from those that are undetectable to ones that are extremely irritating. The source of noise can degrade the performance of the receiver. This does not come from just one direction, but it can be in every direction, which is known as surrounding noise [3]. Time delay is a condition whereby the original signal is reflected in the upcoming origins or known as echo as shown in Figure 1. It is known as the delayed and degraded version of the original signal, which travels back to its source after several reflections to a point delayed and degraded version of the original signal. Echoes may arise within reiteration of waveform upon reflection from a point where a signal gets through and propagates changes. The critical issue that arises for echo signal is finding an appropriate algorithm that is suitable to be used in acoustic echo cancellation [4]. The method can be applied in communication network and/or hands-free communication environment. Figure 1 shows the preprocessing of cancellation of acoustic echo in Quranic signals.

2. Background

The Acoustic Echo Canceller (AEC) is proper for use in audio/video conferencing, hands-free telephony, and speakerphones. The echo paths should be estimated based on the adaptive filter and subsequently reduced estimating and echo in the transmitted signals. Typical algorithms for filter update procedure in AEC include for instance, Normalized Least Mean Square (NLMS), Least Mean Squares (LMS), Affine Projection (AP), Recursive Least Squares (RLS), and Fast Recursive Least Squares (FRLS). The audio signals recorded with microphones are also exposed to echo from speakers, which are composed of the desired speech and background noise [5]. It is crucial in designing the adaptive filters as it may iteratively change its characteristics in order to achieve optimum desired output for d(n) and actual output known as y(n), and this is understood as a cost function [6]. The aim of acoustic echo cancellation is to cancel the desired input signal d(n) by making sure the error signal(e) is at its best minimum value possible. Therefore, it may reflect the best result, which may be acquired for the accuracy rate to enable signal classification to take place. Convergence rate helps to determine the rate where filters may converge to a resultant state. The faster convergence is the desired characteristic for an adaptive system and does not depend on any performance characteristics. The convergence rate factors would decrease and vice versa with performance, for example, stability would decrease if the convergence performance is increased, and the system would be more stable if the convergence rate is decreased (7,8). Table 2.1 explains best some useful comparisons for each of the adaptive algorithms.

An adaptive filter changes automatically based on input signals for the given algorithm. The coefficients would change according to certain options, which typically focus on error signals to improve its performance. Adaptive filters work like digital filters with a combination of algorithms and the coefficients of the filter also vary accordingly. Least squares algorithms are able to minimize the sum of squares for the difference between desired signal and model of filter output [9]. For new samples of each incoming signal of each iteration, the solution is computed using Recursive Least Squares (RLS) algorithm. RLS algorithm is suitable for filtering the presence of speech in a background of noise [2].RLS algorithm also relies on the Least Squares (LS) and estimates the filter coefficients w(n-1) at iteration of (n-1). RLS is excellent in pursuing fast convergence when the eigenvalue spread of the input signal correlation matrix is large. The algorithm works well in time-varying environments, which is widely used in echo cancellation, channel equalization, speech enhancement and radar where the filter is able to be changed at a rapid rate. In addition, Least Mean Squares (LMS) uses gradient based methods, which are understood as an estimation of the gradient vector form of the available data and requires correlation function calculation or matrix inversion that makes it easier as compared to other algorithms [10].

J. of Applied Sciences Technology and Computing Vol. 1 No. 1 (2024) p. 39-51

Table 1 *Comparison of different adaptive algorithms*

Techniques Criteria	Recursive Least Squares (RLS)	Least Mean Squares (LMS)	Affine Projection (AP)
Implementation	Relies on the Least Squares (LS) and estimates the filter coefficients *w(n-1)* at iteration of *(n-1)*.	Estimation of gradient vector form of the available data, higher Echo Return Loss Enhancement (ERLE) (dB) for mu (step size).	In Affine Projection (AP), the adaptive tap weight vector is h=[ho,……..,hL-1] where, *hi.n* is the i^{th} tap at sample period *n*.
Advantages	Best convergence (Hadei, 2010) when the eigenvalue spread of the input signal correlation matric is large. Less time for iteration Better Signal to Noise Ratio.	Easy to implement and computationally inexpensive. Poor convergence rate but least computational complexity.	Affine projection gives similar performance in nonlinear and noisy environments. Faster convergence rate compared to LMS (Ramli et al., 2012).
Disadvantages	More complex compared to LMS.	Sensitive to the scaling of its input *x(n)*	Higher computationally compared to the conventional algorithms.

The minimization of mean square error is able to be achieved as its iterative procedure is able to perform successive corrections in the negative direction of the gradient vector. The LMS filter will have n+1 coefficient, which requires n+1 multiplication for n+1 addition for the filter coefficients and calculation of output in the adaptive filter. The Affine Projection (AP) algorithm is known as a generalization or extension of the NLMS algorithm, whereby it reuses both past and present information (data reusing), while NLMS may only be used for current information [11]. For this algorithm, its superior factor of convergence property may overcome other similar adaptive algorithms such as LMS, NLMS and even for RLS. Filter coefficient is updated each time to overcome the tradeoff between the convergence rate and computational complexity [8]. While the Echo Return Loss Enhancement (ERLE) is known as the calculation of ratio for send-in power (Pd) against the residual error signal power directly after cancellation (Pe) and calculated in decibel (dB). It also understands how much echo is attenuated in decibel (dB). It measures the amount of loss, which is introduced by the adaptive filter. If the ERLE gives a higher value, the better it means for the echo canceller and normally traditional ERLE is measured after removal of near end speech signals, therefore, only a clear amount of echo cancellation can be counted during the noisy time.

2.1 Acoustic Echo Cancellation

The echo level may be higher if the local talker speaks softly, or it might need louder playback levels than usual if it's intended for hearing-impaired people in a noisy environment. In contrast, if the microphone and loudspeaker are placed in the same close-by, easily accessible area, a louder echo is generated. It might take the shape of a consumer-grade gadget, such as a single-unit phone. From [14] finding that echoes can be produced in a reverberated room lends credence to this scenario. A group of reflected sounds from a surface environment, such a room, is known as reverberation [12]. It takes less than 0.1 seconds to measure.

Reverberation happens when speech signals that reach a microphone undergo multiple reflections. There are three types of sound components, namely:1) direct sound, 2) early reverberation and 3) late reverberation. Figure 2 shows the general block diagram of the adaptive filter where it represents the speech signal with the interruption of echo and noise, which is then filtered through the adaptive algorithm and adjustable filter. Results from analog signals are played out through speakers and produce echo in the microphone. Furthermore, echo from speakers and audio signals captured also comprise the desired speech and background echo. In implementing the RLS algorithm, for non-stationary signals, the filter may track time variations while in stationary signals, the convergence behavior is the same as the Wiener Filter.

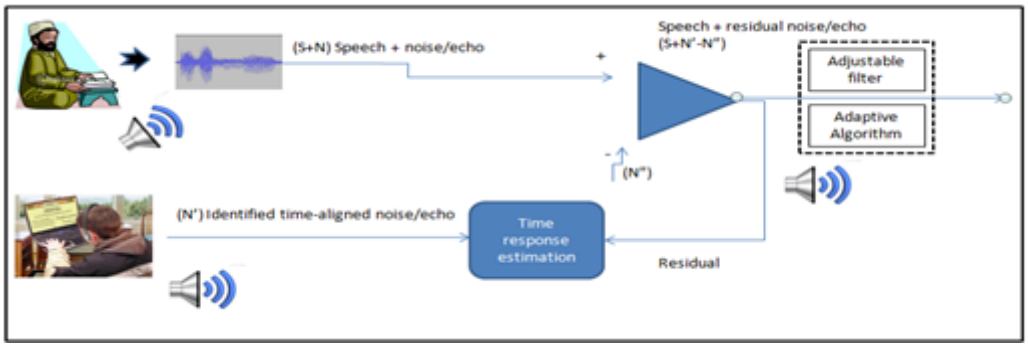

Fig. 1 *Block Diagram of Acoustic Echo Cancellation Component*

The main purpose of hands-free communication is to cancel the acoustic echo in providing a noise-free environment. Acoustic echo cancellers can solve the interference in teleconferencing and hands-free telecommunications [12]. In a comparison conducted [14] on both RLS and LMS algorithms based on a few significant studies, it was found that the convergence rate and weight factor may affect the results for echo cancellation. More iteration is needed for convergence and if *mu* is increased, less iteration is needed. In order to decrease the number of weights, a more ready state is needed for converging compared to RLS. While for RLS, it needs less iteration as compared to LMS to converge and reach a ready state, and more iteration is needed if the number of weights is decreased. Factors [13] of signal to noise ratio to make comparison on his/her research within two different adaptive algorithms, which were the RLS and the LMS. The results were also compared on the computational complexity of those algorithms. Three different step sizes [14], namely: **1)** new variable step size, **2)** fixed step size, and **3)** variable step size parameter. They managed to achieve significant results for the new variable step size (better convergence time and smaller mis-adjustment). More explanations can be viewed in Table 1 as stated previously.

Room acoustic measurements are used to measure amplitude decay, while time-domain signals are used to distribute Room Impulse Response (RIR) for frequency distribution. Three different elements, related to room conditions, often influence the RIR: 1) source position; 2) reflection coefficient; and 3) absorption coefficient. The four parts of RIR are as follows: 1) direct sound, 2) propagation delay, 3) early reflections, and 4) reverberation tail section. Echo issues can arise with local talkers for a variety of reasons. Echo is typically employed in radar exploration and sonar detection. Echo comes in two flavors: 1) telephone line echo and 2) acoustic echo. Other artistic elements audible during music recognition include delay and reverberation, notable echo effects typically utilized in music dubbing and potentially complicating the automatic Room acoustic measurements are used to measure amplitude decay, while time-domain signals are used to distribute Room Impulse Response (RIR) for frequency distribution. Three different elements, related to room conditions, often influence the RIR: 1) source position; 2) reflection coefficient; and 3) absorption coefficient. The four parts of RIR are as follows: 1) direct sound, 2) propagation delay, 3) early reflections, and 4) reverberation tail section. Echo issues can arise with local talkers for a variety of reasons. Echo is typically employed in radar exploration and sonar detection. Echo comes in two flavors: 1) telephone line echo and 2) acoustic echo.

Other artistic elements audible during music recognition include delay and reverberation, notable echo effects typically utilized in music dubbing and potentially complicating the automatic Room acoustic measurements are used to measure amplitude decay, while time-domain signals are used to distribute Room Impulse Response (RIR) for frequency distribution. Three different elements, related to room conditions, often influence the RIR: 1) source position; 2) reflection coefficient; and 3) absorption coefficient. The four parts of RIR are as follows: 1) direct sound, 2) propagation delay, 3) early reflections, and 4) reverberation tail section. Echo issues can arise with local talkers for a variety of reasons. Echo is typically employed in radar exploration and sonar detection. Echo comes in two flavors: 1) telephone line echo and 2) acoustic echo. Other artistic elements audible during music recognition include delay and reverberation, notable echo effects typically utilized in music dubbing and potentially complicating the automatic recognition process [2]. Those two echoes may arise in teleconferences and hearing aid systems, and they are undesirable and can be of annoyance too[15]. They may also cause signal interference and reduced quality of transmission. Some of the major problems that exist in sound applications include background noise, reverberation and acoustic echo or acoustic feedback.

J. of Applied Sciences Technology and Computing Vol. 1 No. 1 (2024) p. 39-51

Table 2 *Performance comparison of acoustic echo cancellation research efforts using adaptive algorithms*

Source	Main Task	Techniques	Data	Results and Evaluation	
				SNR (dB)	ERLE (dB)
[14]	Speech signal enhancement based on adaptive noise cancellation	Least Mean Squares(LMS), Recursive Least Squares(RLS)	English and French speech spoken by male and female speakers	LMS a) English Male:19.1, Female:24.4 b) French Male:19.6 Female:21.6 RLS a) English Male:25.6 Female:38.1 b) French Male:27.1 Female:26.1	N/A
[16]	Noise cancellation for speech enhancement	Least Mean Squares (LMS), Recursive Least Squares(RLS) and Affine Projection (AP) Algorithm	Simulated signals	LMS :13.6 AP :20.03 RLS :29.74	N/A
[17]	Acoustic echo cancellation for speech processing	Least Mean Square (LMS) with different step size (μ)	Simulated signals	N/A	Mu(0.001):20 Mu(0.007):15 Mu(0.03):9
[13]	Acoustic echo cancellation inside a conference room	Recursive Least Square (RLS) and Least Mean Square(LMS)	Simulated signals	N/A	RLS:26 LMS:23.5
[18]	Performance analysis of acoustic echo cancellation	Affine Projection (AP)Algorithman d Recursive Least Square (RLS)	Speech files for male and female speakers	N/A	RLS: 65 AP:55

Table 2 provides a detailed performance comparison of research efforts on acoustic echo cancellation using different adaptive algorithms. Based on Table 2, the adaptive algorithms involved used different kinds of signals data either by human samples or simulated signals, whereby each of them gives different results. It is clearly shown that the RLS [14] has achieved better SNR compared to LMS, whereby higher SNR is achieved. The same conclusion is found [1][2][3] although they measured the ERLE factor that calculates the amount of signal loss applied by the echo canceller. For Tyagi (2012), the ERLE has different results based on different parameters used for Mu. In this research work, the justification of the best adaptive algorithm is based on accuracies after the pattern classification takes place, and the results are shown later in Section 4.0.

3. Methodology

The methodology and workflow of the whole process are shown in Figure 1 which consists of three major phases provided that Quranic accent (Qiraat) data are available and treated as input to the process. Preprocessing is the first phase that covers echoed Quranic accent (Qiraat) signals as main inputs. The echoed signals undergo the acoustic echo cancellation process using different adaptive algorithm including RLS, LMS and AP in steps 1 and 2. This process takes place in preprocessing stage of the current whole workflow. The results of cleaned signal as shown in step 3 will be considered as the input signals for the second phase, which is the feature extraction as shown in steps 4 and 5. Mel Frequency Cepstral Coefficient (MFCC) algorithm is currently used as the main feature

extraction algorithm, which converts clean signals into feature vectors as shown in step 6. These feature vectors will then be used during the third phase, which is the pattern classification. The pattern classification phase in step 7 uses three major classifiers namely: K-Nearest Neighbor (KNN), Gaussian Mixture Model (GMM), and Principal Component Analysis (PCA) with GMM. The experimental results and accuracies are shown later in Table 6.

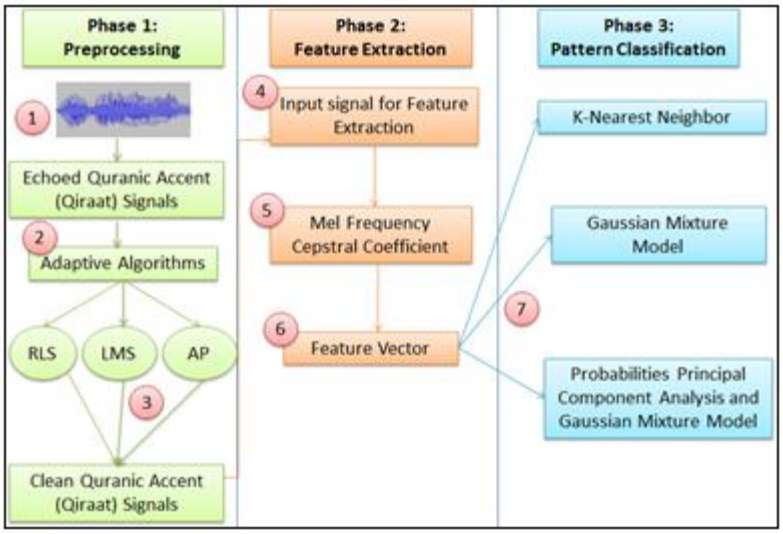

Fig. 1 *Research Workflow and Methodology*

In order to verify our methodology, audio files have been collected from "islamway.net" website (Islamway, 2014) for Surat Ad-Duhaa for five different Quranic accents (Qiraat), namely: 1) Ad-Duri, 2) Al-Kisaie, 3)Hafs an A'asem, 4)IbnWardan, and 5)Warsh. Figure 4 shows Surat Ad-Duhaa with its 11 verses, whereby segments that show differences between Qiraat types are underlined and marked in red color. Fig. 2 Surat Ad-Duhaa from The Holy Quran

وَالضُّحَى ﴿١﴾ وَاللَّيْلِ إِذَا سَجَى ﴿٢﴾ مَا وَدَّعَكَ رَبُّكَ وَمَا قَلَى ﴿٣﴾ وَلَلْآخِرَةُ خَيْرٌ لَّكَ مِنَ الْأُولَى ﴿٤﴾ وَلَسَوْفَ
يُعْطِيكَ رَبُّكَ فَتَرْضَى ﴿٥﴾ أَلَمْ يَجِدْكَ يَتِيمًا فَآوَى ﴿٦﴾ وَوَجَدَكَ ضَالًّا فَهَدَى ﴿٧﴾ وَوَجَدَكَ عَائِلًا
فَأَغْنَى ﴿٨﴾ فَأَمَّا الْيَتِيمَ فَلَا تَقْهَرْ ﴿٩﴾ وَأَمَّا السَّائِلَ فَلَا تَنْهَرْ ﴿١٠﴾ وَأَمَّا بِنِعْمَةِ رَبِّكَ فَحَدِّثْ ﴿١١﴾

Fig. 2 *Surat Ad-Duhaa from The Holy Quran*

3.1 Quranic Accent (Qiraat) Data

Details about the 24 audio files that are used for training purposes are shown in Table 3, whereas Table 4 provides details about the 33 testing audio files.

Table 3 *Details of the Training Audio Files*

Quranic Accent (Qiraat) Type	No. of Files	File Segments
Ad-Duri	**6**	**1.وَالضُّحَى 2.سَجَى 3.فَتَرْضَى 4.فَآوَى 5.فَهَدَى 6. فَأَغْنَى**
Al-Kisaie	**5**	**1.سَجَى 2.فَتَرْضَى 3.فَآوَى 4.فَهَدَى 5.فَأَغْنَى**
Hafs an A'asem	**4**	**1.سَجَى 2.فَآوَى 3. فَهَدَى 4. فَأَغْنَى**
IbnWardan	**5**	**1. قَلَى 2.3فَتَرْضَ 4.فَآوَى 5.فَهَدَى 6.فَأَغْنَى**
Warsh	**4**	**1.2.قَلَى سَجَى 3.فَآوَى 4.فَأَغْنَى**

Table 4 *Details of the Testing Audio Files*

Quranic Accent (Qiraat) Type	No. of Files	File Segments
Ad-Duri	**7**	**1.وَالضُّحَى 2.سَجَى 3.قَلَى 4. فَتَرْضَى 5.فَآوَى 6.فَهَدَى 7.فَأَغْنَى**

Al-Kisaie	**7**	1.وَالضُّحَى 2.سَجَى 3.قَلَى 4.فَتَرْضَى 5.فَآوَى 6.فَهَدَى 7.فَأَغْنَى
Hafs an A'asem	**6**	1.وَالضُّحَى 2.سَجَى 3.قَلَى 4.فَتَرْضَى 5.فَآوَى 6.فَهَدَى فَأَغْنَى
IbnWardan	**7**	1.وَالضُّحَى 2.سَجَى 3.قَلَى 4.فَتَرْضَى 5.فَآوَى 6.فَهَدَى 7.فَأَغْنَى
Warsh	**6**	1.وَالضُّحَى 2.سَجَى 3.قَلَى 4.فَتَرْضَى 5.فَآوَى 6.فَأَغْنَى

Table 5 shows the default attributes of the speech audio and the converted attributes that are used further in this research work. The difference is mainly with the sampling rate that was originally 44.1KHz and it is then converted to 8KHz.

Table 5 *Default and Converted/Used Attributes of the Speech Audio*

Attribute	Default	Converted/Used
Sampling Rate (KHz)	**44,1**	**8**
Bit-Depth (Bits)	**16**	**16**
Channels	**1 Channel (Mono)**	**1 Channel (Mono)**

3.2 Preprocessing

In this study, simulated Room Impulse Response (RIR) is used for impulse response on the signals used. It was proven that simulated response can provide comprehensive testing for acoustic signal processing algorithms which include controlling parameters such as reverberation time, room dimensions and source array distance. It is important to takeinto account sound source, microphone positions, and the reflection time(Jarrett et al., 2012).The estimation of the rectangular rooms with static and rigid wall with monochromatic sound pressure with sharp cornered frequency[19,20,21].From the algorithm, [*B,A]= cheby2(N,Astop,[fc1,fc2*]) where *N* is the filter order, *fc1* and *fc2* is the passband and stopband frequencies, respectively, for the bandpass or stop band filters. The value given here for *fc1*=0.1, *fc2*=0.7,*N*=4 and *Astop*=20dB. The algorithms for Chebyshev type II filter design (stopband ripple) used in this study is as follows (Equation1):

$$H(z) = \frac{B(z)}{A(z)} = \frac{b(1)+ b(2)z^{-1}+b(n+1)z^{-n}}{1+a(2)z^{-1}+\cdots+a(n+1)z^{-n}} \quad (1)$$

The acoustic echo cancellation for the Quranic Accent Signals was simulated in MATLAB for three types of parameters, namely: **1)** a single microphone, **2)** Random Delay for Near Speech, and **3)** Far End Speech, whereby they are used by three different adaptive algorithms, namely: **1)** Adaptive Recursive Least Square (RLS) Algorithm, **2)** Adaptive Least Mean Square (LMS) Algorithm, and **3)** Adaptive Affine Projection (AP) Algorithm. By applying different adaptive filters, the desired signal and adaptive filter output, *e(n)* differed. The error signal, *e(n)* that was fed back to the adaptive filter may have flowed relatively with algorithmic changes and reduced the function difference known as cost function, while the unwanted echoed signal was like the optimum output from the adaptive algorithm filter used. When the error signal turns to 0, the desired signal is equal to the adaptive filter output. During this condition, the echoed signal would be completely cancelled, and the far end user would not be interrupted to listen to anything from the original speech when the signals return [22,23]. While the RLS algorithm would minimize the cost function as in the following equation:

$$\xi(n) = \sum_{k=1}^{n} \lambda^{n-k} e_n^2(k) \quad (2)$$

The k remains as 1, in this equation (2), in which the RLS algorithm commences and λ is a small positive constant very close to, but of a valueless than 1. For values of $\lambda < 1$, the most recent error can estimate recent input samples, focusing more on observed dataand to forget the past. The advantage of RLS is more emphasis on recent samples and removal of the old scheme (Sudhir et al., 2014). For Echo Return Loss Enhancement (ERLE), it is defined by the following algorithm where E{} is the expected value for time sample[27, 28].

$$ERLE(n) = 10log_{10}\left(\frac{E\{y^2(n)\}}{E\{z^2(n)\}}\right) \quad (3)$$

In adapting RLS, the filter output was derived from previous iteration and current input vector using filter tap weights as shown in the following equation [23].

$$\bar{y}_{n-1}(n) = \bar{w}^T(n-1)x(n) \quad (4)$$

Where $\psi(n)$; is a matrix that can be rearranged into a recursive form. The convergence rate is 36dB after calculating ERLE using RLS and decreased until 9dB until the end of the convergence rate which can be seen in Figure4.While for the LMS implementation of each of the iteration of the LMS algorithm, the filter tap weights for LMS are updated according to the following equation [24]:

$$\hat{h}(n+1) = \hat{h}(n) + 2\mu x(n)e(n) \tag{5}$$

Here *x(n)* is the input vector of time delayed input values, *x(n) = [x(n)x(n-1)x(n-2) .... x(n-N+1)*^{T}. While, the vector *h(n)=[h₀(n)h₁(n)h₂(n) ... ...h$_{N-1}$(n)]*^{T} is the coefficient of the adaptive FIR filter tap weight based on *(n)* for time. For variable μ, it is related to the step size parameter and may influence the updating factor. For small values, the time taken for LMS to converge an optimal solution would be long but if $\mu$was large, the adaptive filter would be unstable and the output would diverge[24].The AP algorithm was initially investigated[25] as they utilized the update weight vectors and multiple input vectors[26].For coefficient update, the updated equation for each iteration *n*is as follows [29, 30]:

$$w(n+1) = w(n) - \mu x(n)t(n) \tag{6}$$

where

$$t(n) = [x^{T}(n)x(n) + \delta I]^{-1}e(n) \tag{7}$$

$$e(n) = d(n) - x^{T}(n)w(n) \tag{8}$$

Superscript *T* relates to matrix transpose, δ parameter regulation and *I* is the identity matrix. The input signal vector *x(n)* and desired signal vector *d(n)*are shown as follows:

$$d(n) = \begin{bmatrix} d(n) \\ d(n-1) \\ \vdots \\ d(n-P+1) \end{bmatrix}, x(n) = \begin{bmatrix} x(n) \\ x(n-1) \\ \vdots \\ x(n-P+1) \end{bmatrix} \tag{9}$$

Like LMS implementation, μis transposed to the step size parameter for convergence rate, estimation error of the algorithm, and the convergence rate, while P is the projection order for AP.

Each of the results acquired from the adaptive echo cancellation algorithm is used for the feature extraction and pattern classification, to find the accuracy pattern. Different kinds of pattern classification algorithm were used to check on the wave restored after weight update coefficient in order to get the accuracy rate for each Quranic accent signal based on the different adaptive algorithms used. The pattern classification algorithm used in this study includes K-Nearest Neighbor (KNN), Gaussian Mixture Model (GMM) and Probabilistic Principal Component Analysis (PPCA). The cleaned dataset files varied in terms of the speech signals accent. The clean Quranic accent (qiraat)signals were treated as inputs to the feature extraction phase using Mel Frequency Cepstral Coefficient (MFCC), which is described in the next section.

3.3 Feature Extraction Using MFCC

The computational process of MFCC is shown in Fig. 2, whereby the speech signal is divided into several frames followed by a windowing function within fixed intervals. Normally 20ms of the stationary signal is windowed with Hamming window to remove edge effects. Cepstral feature vector is then generated for each frame. MFCC is widely used for speech recognition as it fulfills the compact representation of spectral envelope, and the signal energy is mostly concentrated for the first coefficients. Mel Cepstral utilizes audio characteristics attributes of the human ear, predominantly utilizing the sound-related front channel bank model, separated talk a particular identity and personalized parameters. This element has made one of the characteristics portrayed in the use of voice-related assignments among the most successful. MFCC is generally utilized for speech recognition as it satisfies the minimal representation of spectral envelope, and the signal energy are generally thought for the first coefficients. Spectral feature is relied upon to enhance execution of MFCC feature as it can catch complementary data that identified with vocal source e.g., pitch, harmonic structure, energy distribution, bandwidth of the speech spectrum and even voiced or unvoiced excitation [31,32].

J. of Applied Sciences Technology and Computing Vol. 1 No. 1 (2024) p. 39-51

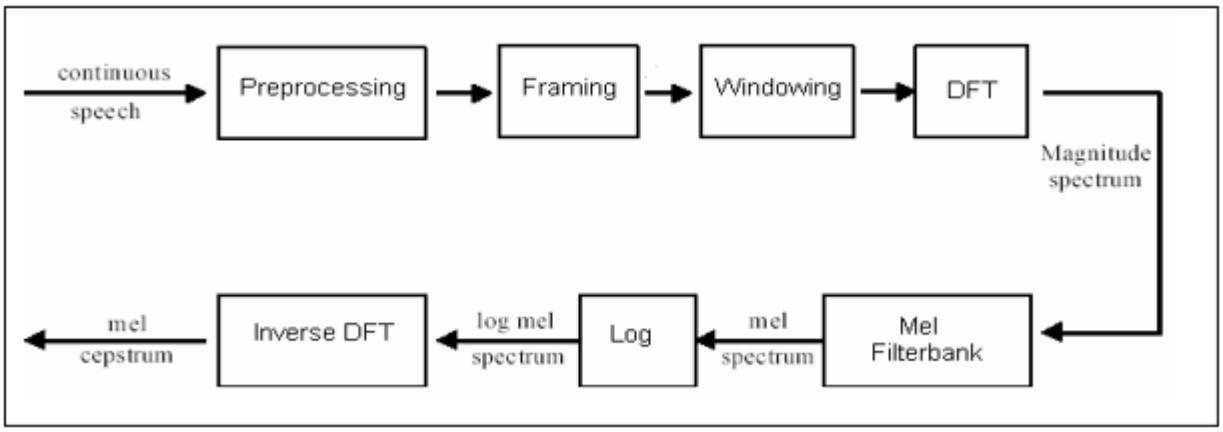

Fig. 2 *Block Diagram of the Computation Steps of MFCC[33].*

4. Pattern Classification

For pattern classification, the algorithm that in used involved K-Nearest Neighbour, Gaussian Mixture Model and Probabilistic Principal Component Analysis.

4.1 K-Nearest Neighbour

kNN is an uncomplicated characterization show that utilizes lazy learning [34]. It is a regulated learning algorithm by classifying the new instances query in light of larger part of kNN classification. Least separation between query instance and each of the training set is ascertained to classify using the kNN.Every query instance (test speech signal) will be comparedagainst each of preparing occurrence (preparing discourse signal). The kNN expectation of the question example is resolved considering the larger part voting of the nearest neighbor classification. Since question case (test speech signal) will analyze against all training speech signal, kNN experiences high response time [34].

4.2 Principal Component Analysis (PCA)

PCA classification is widely used for dimensional reduction and improving the feature vectors with other algorithms like Linear Discriminant Analysis (LDA). It is a widely used dimension reduction technique that performs a linear trans- formation on a set of data using the so-called principal components that best depict the variance within the data set. Other PCA usage derived for feature vector to have mutually uncorrelated elements so that no information about any elementcan be inferred from the remaining ones in a linear way, which present kind of reduction. After removing any linear relation (redundancy) it is vital to select those elements that carry most of the information which is related to their dispersion via entropy[35]. The selected features have enough information within it to identify each speaker class uniquely, and is used for feature space then mapped into eigenspace forclassification and identification [36].

4.3 Gaussian Mixture Model (GMM)

Gaussian Mixture Model (GMM) is widely used for data mining and machine learning. Some of Gaussian Mixture models have been used for time series classification, image texture detection and speaker identification. Gaussian Mixture Model normally using data points from specific object or class such as speaker identification which then generated from a pool of Gaussian model with mixture weights it estimates mixture models from the training data using a maximum likelihood method; it predicts test data with the classes that generate the test data with the largest probabilities.A Gaussian Mixture Model (GMM) based speech estimates the expectation of the mismatch factor between clean speech and noisy speech at each frame by using GMM of clean speech and mean vector of noise. This approach shows a significant improvement in recognition accuracy [30].

MFCC is generally utilized for speech recognition as it satisfies the minimal representation of spectral envelope, and the signal energy are generally thought for the first coefficients. Spectral feature is relied upon to enhance execution of MFCC feature as it can catch complementary data that identified with vocal source e.g.; pitch, harmonic structure, energy distribution, bandwidth of the speech spectrum and even voiced or unvoiced excitation [31,32].

K-Nearest Neighbor (KNN) classifies the instance from learning algorithm and based on 'distance' within training samples. The closer the test sample is to the reference; it conveys more probability to the sample in a group. By having more neighbors, the susceptibility to error due to environmental noise may be reduced as the training samples increase. The derivation of KNN is based on the Euclidian distance:

$$d(x_i, x_j) = \sqrt{For\ all\ attributes\ a\ \sum(x_{i,a} - x_{J,a})^2} \quad (10)$$

Gaussian Mixture Model (GMM) is largely known as an algorithm used for machine learning and data mining. It is also used widely in detection, time series classification, and speaker identification. Mixture weights estimate mixture model from the training data with a maximum likelihood method using data points collected from a pool of Gaussian model. It generates test data with the largest probabilities. The Gaussian Mixture Model (GMM) is able to estimate the mismatch factor between clean and noisy speech within each frame by using the GMM of clean speech and mean vector of noise. The normal methodology is to run the Expected Maximization(EM) algorithm ordinarily from diverse starting setups and to utilize the outcome relating to the most astounding log-probability value. On the other hand, even with a few heuristics that have been proposed to control the initialization, this methodology is as a rule a long way from giving an acceptable arrangement particularly with expanding measurements of the information space. Moreover, utilizing the results of different algorithms, for example, k-means for introduction is likewise frequently not tasteful on the grounds that there are no components that can quantify variability these numerous initializations are from one another. In addition, this is an extremely indirect methodology as numerous EM techniques that are instated with apparently diverse qualities may in any case focalize to comparable nearby maxima. Therefore, this methodology may not investigate the arrangement space successfully utilizing multiple independent runs.

Expected Maximization for this GMM [8] is derived for auxiliary function:

Initial guesses of the parameters: $Q(\odot, \phi)$, where $\Phi = \{wjk\}$

Gaussian components for j = 1, . . ., N and k = 1, . . ., K

Expectation Step: Compute the responsibilities

$$w_{jk}^{(t)} = P\left(y_j = k \middle| x_j, \odot^{(t)}\right) = \frac{\alpha_k^{(t)} p_k(x_j|\theta_k^{(t)})}{\sum_{i=1}^{K} \alpha_i^{(t)} p_i(x_i|\theta_i^{(t)})} \quad (11)$$

Maximization Step: Compute the weighted means and variances:

$$\hat{\mu}_1 = \frac{\sum_{i=1}^{N}(1-\hat{\gamma}_i)y_i}{\sum_{i=1}^{N}(1-\hat{\gamma}_i)},$$

$$\widehat{\mu_k^{(t+1)}} = \frac{\sum_{j=1}^{N} w_{jk}^{(t)} x_j}{\sum_{i=1}^{N} w_{jk}^{(t)}}, \hat{\Sigma}_k^{(t+1)} = \frac{\sum_{i=1}^{N} w_{jk}^{(t)} \left(x_j - \hat{\mu}_k^{(t+1)}\right)(x_j - \hat{\mu}_k^{(t+1)})^T}{\sum_{i=1}^{N} w_{jk}^{(t)}} \quad (12)$$

where *t* indicates the iteration number.

The steps are iterated until convergence is achieved. Principal Component Analysis (PCA)[37] established a technique for dimensionality reduction which explored numerous texts on multivariate analysis. The orthogonal projection of the data onto a lower dimensional vector space which causes the variance of the projected data to be maximized [38] and it covers applications including data compression, image processing, visualization, exploratory data analysis, pattern recognition and time series prediction[39]. PCA is able to manipulate data into some reduced-dimensionality representation and including algebraic manipulation of maximum-likelihood estimators (WML), the obtained results are standard projection for principal axes if desired. While in Probabilistic PCA (PPCA), the principal axes may be found incrementally [37]. For the Probability Principal Component Analysis and the Gaussian Mixture Model, the algorithms used are as follows:

$$log\ p(\boldsymbol{X}) = -\frac{Nd}{2} log\ 2\pi - \frac{N}{2} log|C| - \frac{N}{2} Tr(C - 1S) \quad (13)$$

Where N is the number of data points and S represents the covariance matrix.

5. Results and Discussion

There are two signals which include: (i)near end speech signal, s(n) and (ii) the far end echoed speech signal, v(n). The far end echoed signal is then cancelled from the microphone signal with the near end speech signal distributed. For the microphone signal, it may pick up the far-end and near-end signals, therefore the acoustic echo canceller will remove the far-end signal, and only the near end speech signal is heard by the listener from afar. The room impulse rate used for the three adaptive algorithms is 70dB. For each of the ERLE(dB) as in Table 5.1, the higher the value of ERLE(dB) acquired, the better the adaptive algorithm used in Acoustic Echo Cancellation(AEC)[26] but it also depends on the stability and performance of the signals which relate to the convergence rate factor[13,46].

This is the distinction in sign quality between the original far-end signal and the reverberation of that signal transmitted as the output of the near end, communicated in decibels. As such, ERLE is the measure of amount of signal loss connected to the first far-end signal that returns as echo.

Table 6 *Results of Erle(dB) for different adaptive algorithms.*

Adaptive Algorithm	Results in ERLE (dB)
Recursive Least Square (RLS)	16
Least Mean Square (LMS)	18
Affine Projection (AP)	17

Table 7 *Classification results based on different algorithms.*

Pattern Classification Adaptive Algorithm	Probability Principal Component (PPCA) + Gaussian Mixture Model (GMM) (%)	K-Nearest Neighbour (KNN) (%)	Gaussian Mixture Model(GMM) (%)
Recursive Least Square (RLS)	90.9	78.8	90.9
Least Mean Square (LMS)	96.9	84.8	96.9
Affine Projection (AP)	93.9	93.9	93.9

6. Conclusion

With reference to Table 6 it clearly shows that LMS adaptive filter algorithm provides better results on convergence rate based on ERLE (dB) as compared to the RLS and AP algorithms. However, based on the classification accuracy, cleaned Quranic signals with AP adaptive algorithm provided stable accuracy performance as compared to the other adaptive algorithms between different classification methods for acoustic echo cancellation based on the signals involved. The results provide another view of each of the adaptive algorithms involved in this acoustic echo cancellation. This shows that AP provides better ERLE (dB) and a consistent accuracy rate on the cleaned signals as in Table 6. The effectiveness of each algorithm can be justified too with its accuracy rate based on the classification results after the Acoustic Echo Cancellation has taken place for each Quranic signal.

Acknowledgement

The author(s) would like to convey much gratitude for Islamway[41], Faculty Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia and Faculty Engineering, Universiti Putra Malaysia for allowing this research be conducted successfully.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

[1] H. Zhang, S. Kandadai, H. Rao, M. Kim, T. Pruthi and T. Kristjansson, "Deep Adaptive Aec: Hybrid of Deep Learning and Adaptive Acoustic Echo Cancellation," ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Singapore, Singapore, 2022, pp. 756-760, doi: 10.1109/ICASSP43922.2022.9746039.

[2] Kamarudin, N., Al-Haddad, S. A. R., Khmag, A., bin Hassan, A. R., & Hashim, S. J. (2016). Analysis on Mel frequency cepstral coefficients and linear predictive cepstral coefficients as feature extraction on automatic accents identification. International Journal of Applied Engineering Research, 11(11), 7301–7307.

[3] Kamarudin, N., Al-Haddad, S. A. R., Khmag, A., Hashim, S. J., & Hassan, A. R. (2017). Sequential parameterizing affine projection (SPAP) windowing length for acoustic echo cancellation on speech accents identification. 2017 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting, EBBT 2017.

[4] Ren, Y., Zhi, Y. & Zhang, J. Geometric-algebra affine projection adaptive filter. EURASIP J. Adv. Signal Process. 2021, 82 (2021). https://doi.org/10.1186/s13634-021-00790-y

[5] C. Tourain et al., "Benefits of the Adaptive Algorithm for Retracking Altimeter Nadir Echoes: Results From Simulations and CFOSAT/SWIM Observations," in IEEE Transactions on Geoscience and Remote Sensing, vol. 59, no. 12, pp. 9927-9940, Dec. 2021, doi: 10.1109/TGRS.2021.3064236.
[6] T. O'Malley, A. Narayanan, Q. Wang, A. Park, J. Walker and N. Howard, "A Conformer-Based ASR Frontend for Joint Acoustic Echo Cancellation, Speech Enhancement and Speech Separation," 2021 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), Cartagena, Colombia, 2021, pp. 304-311, doi: 10.1109/ASRU51503.2021.9687942.
[7] Rafizah Mohd Hanifa a b, Khalid Isa a, Shamsul Mohamad a, A review on speaker recognition: Technology and challenges, Computers and Electrical Engineering,2021
[8] Caglar Ari, Selim Aksoy, O. A. (2012). Maximum Likelihood Estimation of Gaussian Mixture Models Using Stochastic Search. Journal Pattern Recognition, Volume 45, Pages 2804–2816.
[9] Chia Ai O, Hariharan M, Yaacob S, Sin Chee L (2012) Classification of speech dysfluencies with MFCC and LPCC features. Expert Systems with Applications 39:2157–2165. doi: 10.1016/j.eswa.2011.07.065
[10] Deepika M (2013). Noise Cancellation In Speech Signal Processing Using Adaptive Algorithm. International Journal on Recent and Innovation Trends in Computing and Communication 1:743–746.
[11] K. Kumar, R. Pandey, S. S. Bhattacharjee and N. V. George, "Exponential Hyperbolic Cosine Robust Adaptive Filters for Audio Signal Processing," in IEEE Signal Processing Letters, vol. 28, pp. 1410-1414, 2021, doi: 10.1109/LSP.2021.3093862.
[12] B. Chen, L. Xing, H. Zhao, N. Zheng and J. C. Príncipe, "Generalized correntropy for robust adaptive filtering", IEEE Trans. Signal Process., vol. 64, no. 13, pp. 3376-3387, Jul. 2016.
[13] Ramli RM, Noor AOA, Samad SA (2012).A Review of Adaptive Line Enhancers for Noise Cancellation. Australian Journal of Basic and Applied Sciences 6:337–352.
[14] Parvin S, Park JS (2007).An Efficient Music Retrieval Using Noise Cancellation. Future Generation Communication and Networking (FGCN 2007) 541–546. doi: 10.1109/FGCN.2007.59
[15] Gannamaneni G (2012). Acoustic Echo cancellation inside a Conference Room using Adaptive Algorithms Master Thesis Electrical Engineering June 2012. Blekinge Institute of Technology, Karlskrona, Sweden
[16] Mousa A, Qados M, and Bader S (2012).Speech Signal Enhancement Using Adaptive Noise Cancellation Techniques. Canadian Journal on Electrical and Electronics Engineering 3:375–383.
[17] Hutson M (2003). Acoustic Echo Cancellation Using Digital Signal Processing. Bachelor Engineering Thesis The University of Queensland
[18] Hadei SA (2010). A Family of Adaptive Filter Algorithms in Noise Cancellation for Speech Enhancement. International Journal of Computer and Electrical Engineering 2:10.
[19] Tyagi, R., and Agrawal, D. (2012).Analysis the Results of Acoustic Echo Cancellation for Speech Processing using LMS Adaptive Filtering Algorithm.International Journal of Computer Applications, 56(15), 7–11.
[20] Adapa N.S., and Bollu S., (2012). Performance Analysis of different Adaptive Algorithms based on Acoustic Echo Cancellation. Master Thesis, Blekinge Institute of Technology, 371 79 Karlskrona Sweden.
[21] Sena, E. De, Antonello, N., &Moonen, M. (2015).On the Modeling of Rectangular Geometries in Room Acoustic Simulations, IEEE/ACM Transactions on Audio,Speech, and Language Processing 23(4), 774–786.
[22] Jacobsen, Finn; Juhl, Peter Møller. (2013) Fundamentals of General Linear Acoustics. United Kingdom : John Wiley & Sons Ltd, 2013. 284 p.
[23] Heinrich Kuttruff (2000). Room Acoustics, 4th ed. Abingdon, Taylor and Francis Group U.K.: SPON.
[24] Liu KR, Hsieh SF, Yao K (1992).Systolic block householder transformation for RLS algorithm with two-level pipelined implementation. IEEE Transactions on Signal Processing 40:946–958. doi: 10.1109/78.127965
[25] Munjal A, Aggarwal V, Singh G (2008). Acoustic Echo Cancellation Using RLS Algorithm. Proceedings of 2nd National Conference on Challenges & Opportunities in Information Technology (COIT-2008) RIMT-IET, Mandi Gobindgarh. March 29, 2008. pp 299–303
[26] Sudhir V V, Murthy ASN, and Rani DE (2014). Acoustic Echo Cancellation using Adaptive Algorithms. International Journal of Advances in Computer Science and Technology 3:248–252.
[27] Ozeki, K. and Umeda, T. (1984). An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties. Electronics and Communications in Japan Pt. I, 67: 19–27. doi: 10.1002/ecja.4400670503
[28] Ramli RM, Noor AOA, Samad SA (2012)A Review of Adaptive Line Enhancers for Noise Cancellation. Australian Journal of Basic and Applied Sciences 6:337–352.
[29] Stokes JW, Malvar HS, and Way OM (2004).Acoustic Echo Cancellation with Arbitrary Playback Sampling Rate. IEEE International Conference on Acoustics, Speech, and Signal Processing, 2004.Proceedings. (ICASSP '04).(Volume:4). pp 1–4.
[30] R López-Cózar, A De la Torre, J.C Segura, A.J Rubio, V Sanchez Sánchez(2002),Testing dialogue systems by means of automatic generation of conversations, Interacting with Computers, Volume 14, Issue 5, October 2002, Pages 521–546

[31] Diniz P.S.R(2008). Adaptive Filtering: Algorithms and Practical Implementations Springer. Boston, MA, Third Edition,. ISBN: 978-0-387-31274-3.
[32] Haykin, S., (2002). Adaptive Filter Theory. 4th Edition, Prentice Hall.
[33] Hosseinzadeh, D., & Krishnan, S. (2008). On the Use of Complementary Spectral Features for Speaker Recognition. EURASIP Journal on Advances in Signal Processing, 2008(1), 258184. doi:10.1155/2008/258184
[34] Kamarudin, N., Al-Haddad, S. A. ., Rauf, A., Hassan, B., Hashim, S. J., & Nematollahi, M. A. (2014). Feature Extraction Using Spectral Centroid and Mel Frequency Cepstral Coefficient for Quranic Accent Automatic Identification. IEEE Student Conference on Research & Development SCOReD 2014 (pp. 1–6).
[35] Khalifa, O., Khan, S., Islam, M.R., Faizal, M., and Dol, D. (2004). Text Independent Automatic Speaker Recognition. 3rd International Conference on Electrical & Computer Engineering, Dhaka, Bangladesh, pp. 561-564.
[36] Pallabi, P., &Bhavani, T. (2006). Face Recognition Using Multiple Classifiers. In International conference on 18th IEEE tools with artificial intelligence, 2006. ICTAI'06.
[37] JurajKacur, RadoslavVargic, P. M. (2011). Speaker identification by K-nearest neighbors. In 18th International Conference on Systems, Signals and Image Processing (IWSSIP), 2011 (p. 4).
[38] Md. Rashedullslam , Md. Shafiul Azam, S. A. (2009). Speaker Identification System Using PCA & Eigenface. In 12th International Conference on Computer and Information Technology (ICCIT 2009) 21-23 December, 2009, Dhaka, Bangladesh Speaker (pp. 21–23).
[39] Tipping, M.E. and Bishop, C.M. (1999).Mixtures of Probabilistic Principal Component Analyzers .Neural Computation, Microsoft Research 11(2), 443-482.
[40] Lee C-H, Chou C-H, Lien C-C, Fang J-C (2011). Music genre classification using modulation spectral features and multiple prototype vectors representation. 2011 4th International Congress on Image and Signal Processing 2762–2766.doi: 10.1109/CISP.2011.6100759
[41] BabuM.,RaoK. M. L., and KhapardeA., (2011). Risk Priority Estimation for Major Corridors by using PCA Application of Traffic Mobility in Urban Areas : A Case Study of Gaddiannaram Municipality Area Hyderabad , A . P ., India. International Journal of Applied Engineering Research., vol. 6, no. 5, pp. 615–625.
[42] Kourav A, Soni BK (2011).RLS Algorithm for Adaptive Echo Cancellation. International Journal on Emerging Technologies 2:35–38.
[43] Islamway, (2014).http://en.islamway.net/ Retrieved May 2014.