International Journal of Engineering & Technology, 7 (4.15) (2018) 486-490

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

A Normalized Least Mean Square and Dynamic Time Warping (DTW) Algorithm for an Intelligent Quran Tutoring System

Mohammed Arif Mazumder, Rosalina Abdul Salam*

Faculty of Science and Technology, Universiti Sains Islam Malaysia (USIM), Negeri Sembilan, Malaysia
**Corresponding author E-mail: rosalina@usim.edu.my*

Abstract

Al-Quran is the most recited holy book in the Arabic language. Over 1.3-billion Muslim all over the world have an obligation to recite and learn Al-Quran. Learners from non-Arabic as well as from Arabic speaking communities face difficulties with Al-Quran recitation in the absence of a teacher (ustad) around. Advancement in speech recognition technology creates possible solutions to develop a system that has a capability to auricularly discern and validate the recitation. This paper investigates the speech recognition accuracy of template-based acoustic models and propose enhancement methods to improve the accuracy. A new scheme consists of enhancement of Normalized Least Mean Square (NLMS) and Dynamic Time Warping (DTW) algorithms have been proposed. The performance of the speech recognition accuracy was further improved by incorporating an adaptive optimal filtering with modified humming window for MFCC (Mel-frequency cepstral coefficients) using matching technique dynamic programming (DP), DTW (Dynamic Time Wrapping). The proposed scheme increases 5.5% of relative improvement in recognition accuracy achieved over conventional speech recognition process.

Keywords: *Dynamic Time Wrapping (DTW); Mel Frequency Cepstral Coefficient (MFCC); Normalize Least Mean Square (NLMS).*

1. Introduction

It is mandatory for all Muslims to be able to read and comprehend the commands of Allah in Al-Quran. So, Muslims all over the world, consider recitation of Al-Quran as a holy duty. In order to have a correct pronunciation in reciting the holy Quran, a presence of a teacher (ustad) is important. Consistent reading with face to face interaction is necessary. For a beginner, additional effort is also required. The proposed system aims to speed up the learning and pronunciation process. The speech recognition process proposed in this system allows a learner to record his recitation and automatically validate his reading with prominent teacher's (ustad) pronunciation. This will reduce the face-to-face time with the teacher and can improvise his/her pronunciation as well.

A new scheme that includes three proposed methods has been proposed. The first two methods are the pre-processing of voice signal to feature extraction. The first method reduces the sinusoidal noise from the input signal and also reduces the wideband noise. The second method extracts the speech features using spectral based parameter approach MFCC (Mel-frequency cepstral coefficients). The third approach called DTW (Dynamic Time Wrapping) is a template-based pattern matching approach for calculating the level of similarity between two time series in which any of them (test word or recorded word) may be warped in a non-linear fashion by shrinking and stretching the time axis.

Section II presents the review of available literature on the adaptive filtering method, MFCC and DTW. Section III proposes three enhanced methods to improve the accuracy of speech recognition. The performances of the proposed approaches are evaluated in Section IV and finally, the conclusions are drawn in Section V.

2. Related Research

According to related research and existing process of speech recognition system, it consists of three major stages -Pre-Processing, Feature Extraction, and Pattern Matching [1] as shown in Figure 1.

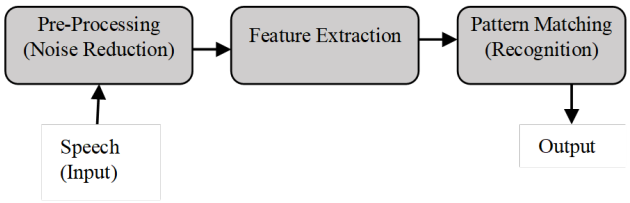

Fig. 1: Speech Recognition Process [1]

The following sub-sections define the methods used for the above stages. The relevant primary sources are summarized and presented in the next section.

2.1. Pre-Processing (Noise Reduction)

The main aim of pre-processing step of speech recognition is to organize the information and simplify the data (input speech signal) and further simplify the recognition.

When a noisy signal passes through a filter that inclines to suppress the noise while leaving the signal relatively unchanged, it is known as direct filtering. Direct filtering originates from the work of Wiener and was extended and enhanced by Kalman, Bucy and other researchers [1, 2]. Filters used for direct filtering can be either Fixed or Adaptive. If the signal is narrowband and noise broadband, a priori information is not needed. Otherwise, it will require a signal (desired replication) that is correlated with the signal to be estimated. Moreover, adaptive filters have the capability of adaptively tracking the signal under non-stationary conditions [1-3]. There are a number of different approaches that can be used in adaptive filtering as shown in Figure 2.

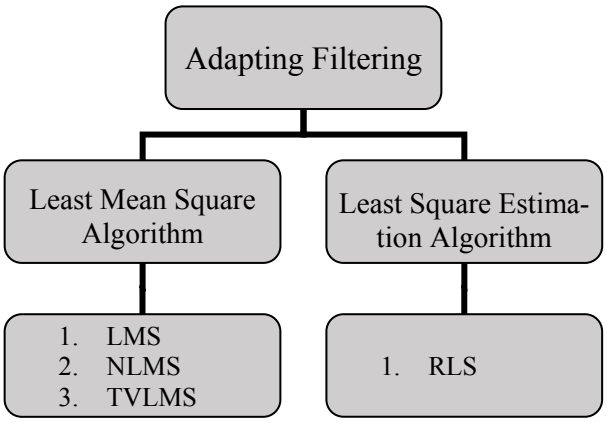

Fig. 2: Adaptive Filters Hierarchy [4, 5]

Least mean squares (LMS) algorithm [3, 6] is a class of adaptive filter used to mimic a desired filter by finding the filter coefficients that relate to producing the least mean squares of the error signal (difference between the desired and the actual signal) and will adapt based on the current error. Another variant of the LMS algorithm is Normalized Least Mean Squares (NLMS) filter [4, 5, 6] that solves the error signal problem by normalizing with the power of the input and achieve a slightly better accuracy than LMS. Time varying Least Mean Square (TVLMS) algorithm [3, 4, 7] is the new version of LMS, which is based on utilizing a time varying convergence parameter with general power for LMS algorithm. The core difference between LMS and TVLMS is that the LMS algorithm needs a large convergence parameter value. Whereas, in TVLMS, we can set relatively small convergence parameter for better accuracy [4, 6, 7]. Recursive least squares (RLS) [3, 4, 7] adaptive filter is an algorithm which recursively finds the filter coefficients that minimize a weighted linear least square cost function relating to the input signals. Table 1 compares and shows the performance of the RLS adaptive algorithm. The stability is higher as compared to other algorithms since the less mean-square error (MSE) is lower [3, 6].

Table 1 presents a review of major adaptive filtering algorithms. It shows that LMS is preferred over RLS algorithms for various noise cancellation purposes as RLS has increased computational complexity and stability compared to LMS-based algorithms which are robust and reliable. Complexity of LMS-based algorithms are less in comparison to RLS. Moreover, LMS provides a practical frame of reference for assessing any further improvement that may be attained through the use of more sophisticated adaptive filtering algorithms [3, 6, 7].

Table 1: Performance Comparison of Adaptive Algorithms

Algorithms	MSE	Complexity	Stability
LMS	$1.5*10^{-2}$	2N+1	Less Stable
NLMS	$9.0*10^{-3}$	3N+1	Stable
RLS	$6.2*10^{-3}$	$4N^2$	High Stable
*MSE: Mean Square Error			

2.2. Feature Extraction

The goal of feature extraction is to find the set of properties called parameter of utterances by processing the signal waveform of the utterances. It produces a meaningful representation of speech signals [4, 5, 7]. Feature extraction techniques such as linear predictive coding (LPC), mel frequency cepstral coefficients (MFCC), linear frequency cepstral coefficients (LFCC) and perceptual linear predictive analysis (PLP) are used for converting speech signals to digital format and then measures important characteristics of signal [4, 7].

The Linear Predictive Coefficient (LPC) method is a time domain approach, where LPC analysis is carried out by approximating each current sample as a linear combination of past samples speech samples [8, 9]. Though the computation speed of LPC is good and provides with accurate parameters of speech, but it generates residual error as output. This means that some amount of important speech gets left in the residue resulting in poor speech quality [9, 10]. The Linear Predictive Cepstral Coefficient (LPCC) is an extension of the LPC technique, where a cepstral analysis is executed after completing the LPC analysis in order to obtain the corresponding cepstral coefficients [11]. LPCC is implemented using the autocorrelation method for resolving the residual error. LPCC is highly sensitive to quantization noise. It is the main drawback of the LPCC. Another method is Perceptual Linear prediction (PLP), which is an analytical model perceptually motivated auditory spectrum by a low order pole function using the autocorrelation LP technique [8, 11, 12]. The PLP analysis provides similar results as with the LPC analysis, but the order of PLP model is half of the LP model therefore less computational storage [13, 14]. PLP sometimes has been slightly better than LPCC, when it comes to noisy environment. Among those techniques, the most widely used feature extraction methods is Mel frequency Cepstral Coefficient (MFCC) in the field of ASR [8, 14]. MFCC provides good discrimination [5] and low correlation between coefficients, but MFCC performance might be affected by the number of filters [10] and does not give accurate results if there are background noise [8].

2.3. Pattern Matching

In order to determine the identity of the unknown in pattern-comparison, a direct comparison is made between the unknown speeches (the speech to be recognized) with each possible pattern learned in the training stage [1, 5, 6]. Various techniques have been used for isolated and continuous speech recognition process, such as, dynamic programming algorithm, Dynamic Time Wrapping (DTW), Artificial Neural Network (ANN) and Hidden Markov Model (HMM). These are prominent speech classifiers that are used for achieving an optimal recognition rate of speech signal.

This paper focuses on DTW, which is used to find an optimal alignment between two given (time-dependent) sequences under certain restrictions intuitively. DTW feature classifier is useful for isolated word recognition in a limited dictionary, though it has O(n2v), high complexity for larger dictionary [10, 12, 15]. Hidden Markov Models (HMM) has been widely used in lot of speech recognition applications because of its high reliability [12]. It can be used for large amounts of data, but also increase higher computational complexity. Artificial Neural Networks (ANN) are another classifier of speech recognition with acceptable accuracy. ANN is a nonlinear model which is easier to use and understand than statistical methods, but ANN may give unpredictable output quality [14, 15].

From the above comparative studies of different methods for speech recognition, there are still issues that can be further improved, especially on the accuracy. This paper proposes an enhanced NLMS for noise cancellation with the collaboration of MFCC and DTW algorithm to gain a higher rate of accuracy in speech recognition.

3. Methodology

A new scheme is proposed in order to address the problem of noise removal and distortion from voice signal. Enhanced methods have been proposed to increase the accuracy of speech recognition in Quran recitation, which is shown in Figure 3.

The proposed method used in this paper in which input speech signal for the Quran recitation from the teacher (ustad) is extracted through segmented features. The teacher voice signal and its features are collected from an existing database. These recorded recitations are free from noise.

Voice input for the correct Quran recitation recorded from the verified teacher (ustad) will be stored as reference recitation. Then, the difference

between the segmented new voice from the user and the set of segmented voice from the database will be calculated.
The enhanced methods for the pre-processing stage, feature extraction and pattern matching are described in the following section. Finally, the results will be compared with the conventional techniques.

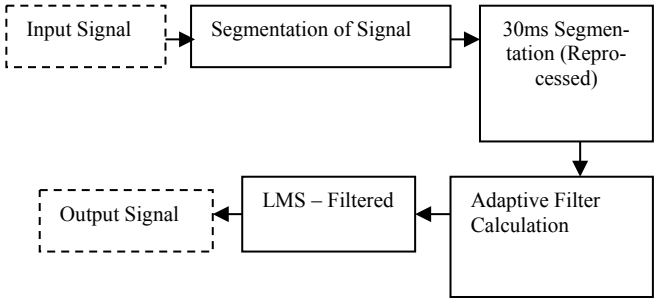

Fig. 3: Approach of Smart Quran Tutoring System

3.1. Normalize Least Mean Square (NLMS) with time segmentation

Speech signals are varying with time, and to get the efficacious noise reduction with LMS filtering method, the input signal must be segmented. The unprocessed voice signal needs to be segmented every 30ms for reprocessing as shown in Figure 4. Let $S_n = \{S_t: t = 1, 2,...,T\}$ be a noisy test signal with T frames and the frame at time t.

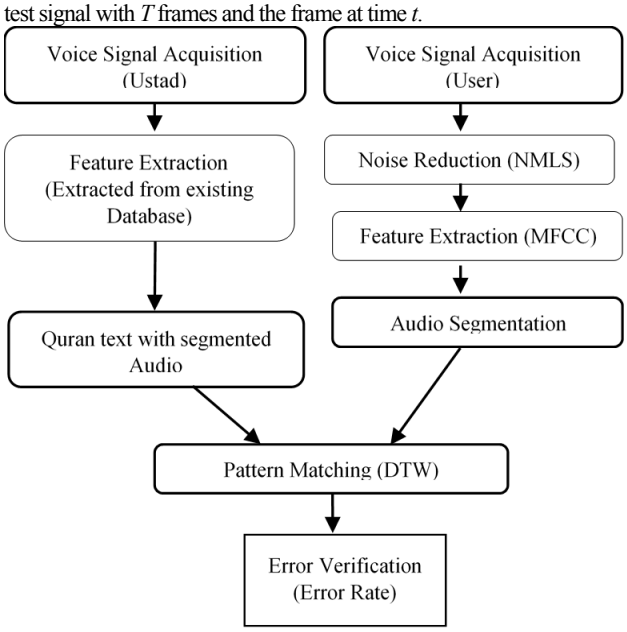

Fig. 4: Adaptive LMS Filtering with Time Segmentation

Signals are segmented between consecutive speech signal frames. Normalized Least Mean Square (NLMS) algorithm can be calculated to find weight vector wt. Total weight for adaption equation for time varying NLMS is as follows:

$$\boldsymbol{w(n+1)} = \boldsymbol{w(n)} + \frac{\breve{\mu}}{x^T(n)x(n)}\boldsymbol{e(n)x(n)} \quad (1)$$

where $\breve{\mu}$ is the input parameter, $\frac{\breve{\mu}}{x^T(n)x(n)}$ is actual segmentation size and $x(n)$ is the input value.
A more reliable implementation of the NLMS algorithm in practice requires the assistance of a regularization parameter ε, and it results in the ε -NLMS algorithm with the improved step size:

$$\left(\frac{\breve{\mu}}{x^T(n)x(n)}\right)/t(p) \quad (2)$$

Here $t(p)$ = 30ms which segment the weight vector in a small fraction for better noise cancellation. It shows that normalize LMS can have better performance than LMS algorithm. This proposed method pays attention to the non-stationary nature of voice signals and improves the performance quality of noisy voice signal.

3.2. Mel Frequency Cepstral Coefficients (MFCC) with enhanced features

Block diagram of the standard MFCC that includes fundamental steps to derive MFCC from an original input speech shown in Figure 5.

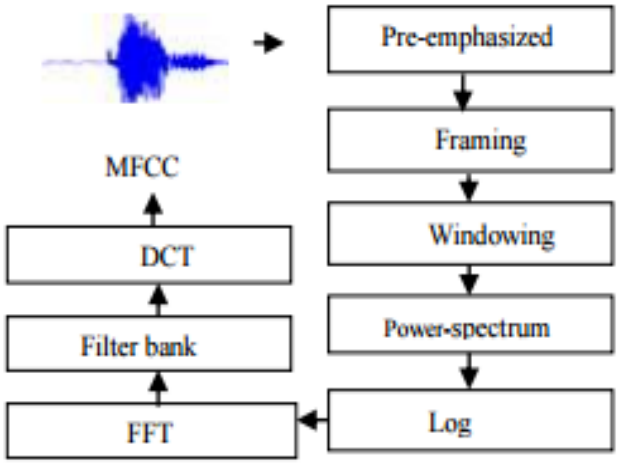

Fig. 5: Standard MFCC Extraction Algorithm [9]

Figure 6 shows ameliorate MFCC algorithm with deference to integrating complementary blocks and modification in the standard block.

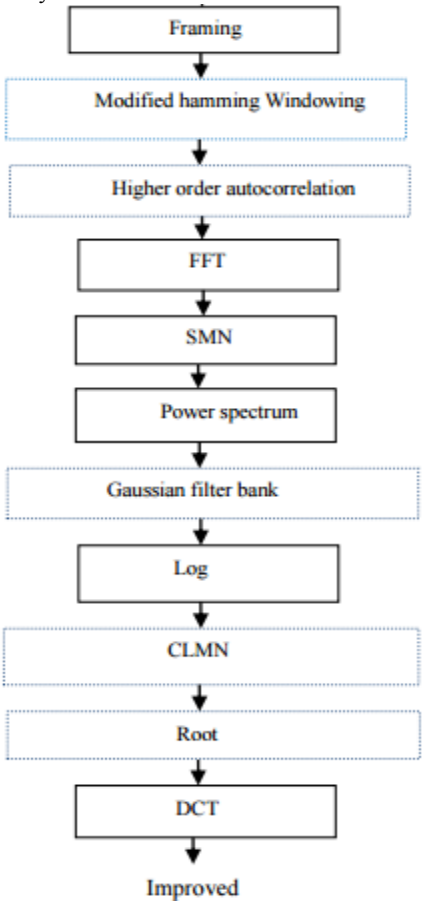

Fig. 6: Improved MFCC Extraction Algorithm

Though in antecedent step (Noise Reduction), two of the rudimentary steps already consummated for MFCC. Therefore, in an enhanced MFCC, it will include modified hamming window to minimize the hamming window size by half. If w(n) be a simple hamming window, then the new hamming window size will be as follows:

$$w_{new}(n) = w(n) / 2 \quad (3)$$

One-sided autocorrelation sequences of the framed signal passed from the modified hamming window, which are obtained and the lower lags of the autocorrelation sequences are removed. It can further suppress the noise.

3.3. Dynamic Time Wrapping for Voice Pattern Matching by adding additional Dynamic Programming (DP)

DTW is a nonlinear reformed technology that is coalesced with the time warping and distance measure calculation. DTW algorithm predicated on the minimum distance between the test and the reference patterns along the aligned path, which is obtained utilizing DP (Dynamic Programming). In the conventional DP-matching technique, the plane of grids shown in Figure 7 is generally utilized. The figure also shows an example of the alignment path [14].

Figure 8 shows the proposed DTW Time Alignment flow. In the proposed algorithm, the supplemental DP utilized for the proposed system endeavours to find the best-matched transition points. By introducing additional DP, the recognizer can be more robust against the variability in the verbalization pattern and possible errors in speech/silence, voiced/unvoiced decisions.

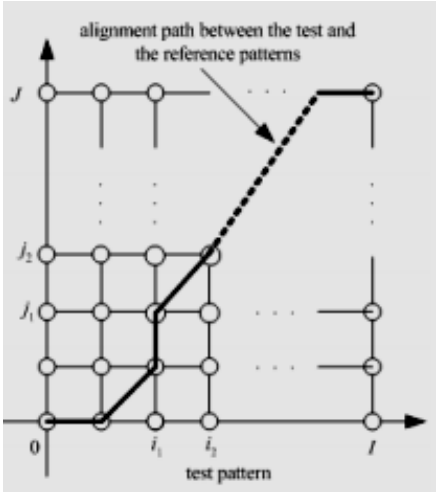

Fig. 7: Conventional DTW Time Alignment [14]

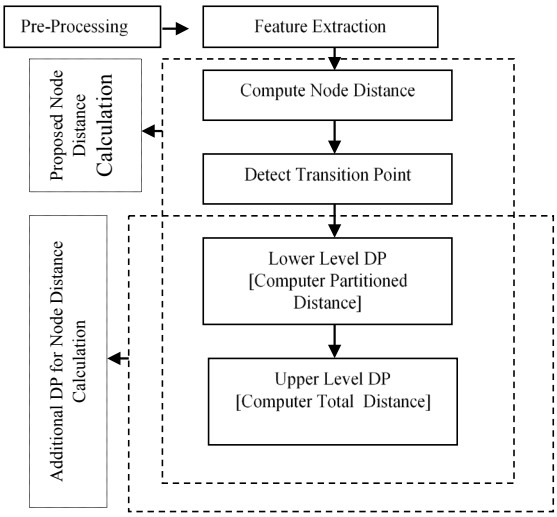

Fig. 8: Proposed DTW Time Alignment

4. Results and Discussion

This section shows the comparison between the performance of the LMS and time varying NLMS algorithms as noise canceller as shown in Figure 9. The noise abrogated signal has a time domain waveform [7, 14] that is relatively identically equal to pristine.

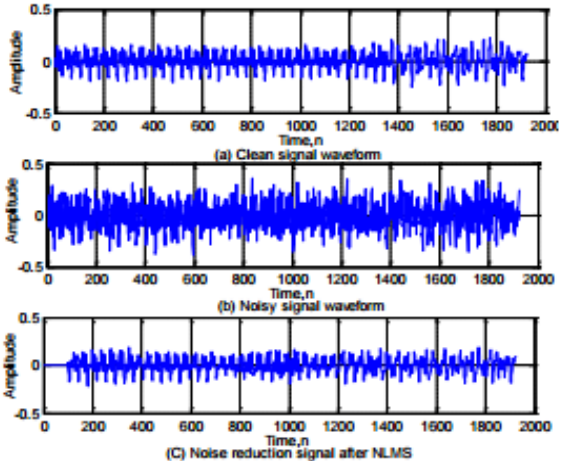

Fig. 9: Proposed noise reduction algorithm for audio signal

To compare the proposed DTW algorithm with a conventional DTW-predicated recognizer, the proposed algorithm is tested on 20 Quranic Sentences (Ayah') verbalized by 3 speakers. The system recorded the test speech samples at 10 kHz sampling rate utilizing a mobile microphone. Test results of applying improved DTW algorithm with MFCC feature extraction method is shown in Table 4.

Table 4: Recognition Accuracy when Dynamic Features are included with MFCC feature Extraction

Feature Type	Conventional Algorithm	Proposed Algorithm
2nd Order MFCC	0.65	0.71
12th order MFCC with differential Coefficients	0.76	0.81
12th order MFCC with differential and acceleration coefficient	0.64	0.83

5. Conclusion

The paper proposes a new scheme with an enhanced method for noise cancellation using a time variant LMS as pre-processing of the acoustic signal. It is then further improved by a modified hamming window size that was reduced by half using MFCC for finding accurate parameters or features of the speech signal. Lastly, an enhanced technique for feature recognition using supplement DP for robust pattern matching technique was introduced. The proposed scheme provides higher accuracy for Quran tutoring system. The future work will be to further improve the recognition speed without compromising its recognition accuracy and utilize the proposed methods in mobile apps for Smart Quran Tutoring.

Acknowledgement

The authors would like to express their gratitude to Universiti Sains Islam Malaysia (USIM) for the supports and facilities provided. This research study is sponsored by Universiti Sains Islam Malaysia (USIM) under USIM Competitive Grant [PPP/UTG-0114/FST/30/11414].

References

[1] Dhiman, J., Ahmad, S., & Gulia, K. (2013). Comparison between Adaptive Filter Algorithms (LMS, NLMS and RLS). International Journal of Science, Engineering and Technology Research, 2(5), 1100-1103.

[2] Mangamma, V., & Saravanan, V. (2014). Noise cancellation of speech signal by using adaptive filtering with averaging algorithm. international Conference on Innovations in Engineering and Technology, 3(3), 1917–1920.

[3] Hadei, S. & M. lotfizad (2011). A family of Adaptive Filter Algorithms in noise cancellation for speech enhancement. International Journal of Computer and Electrical Engineering, 2(2), 1793–8163.
[4] Anusuya, M. A., & Katti, S. K. (2010). Speech recognition by machine: A review. International Journal of Computer Science and Information Security, 6(3), 181-205.
[5] Ibrahim, N. J., Razak, Z., Yusoff, Z. M., Idris, M. Y. I., Tamil, E. M., Noor, N. M., ... & Naemah, N. (2008). Quranic verse recitation recognition module for support in j-QAF learning: A review. International Journal of Computer Science and Network Security, 8(8), 207-216.
[6] Ghule, K. R., & Deshmukh, R. R. (2015). Feature extraction techniques for speech recognition: A review. International Journal of Scientific and Engineering Research, 6(5), 2229-5518.
[7] Liu, Y., Xiao, M., & Tie, Y. (2013). A noise reduction method based on LMS adaptive filter of audio signals. Proceedings of the 3rd International Conference on Multimedia Technology, pp. 1001-1008.
[8] Darabian, D., Marvi, H., & Sharif Noughabi, M. (2015). Improving the performance of MFCC for Persian robust speech recognition. Journal of AI and Data Mining, 3(2), 149-156.
[9] Kim, C., & Seo, K. D. (2005). Robust DTW-based recognition algorithm for hand-held consumer devices. IEEE Transactions on Consumer Electronics, 51(2), 699-709.
[10] Afroz, F., Huq, A., & Sandrasegaran, K. (2015). Performance analysis of adaptive noise canceller employing NLMS algorithm. International Journal of Wireless and Mobile Networks, 7(2), 45-58.
[11] Ahmed, A. H., & Abdo, S. M. (2017). Verification system for Quran recitation recordings. International Journal of Computer Applications, 163(4), 6-11.
[12] Arora, S. J., & Singh, R. P. (2012). Automatic speech recognition: A review. International Journal of Computer Applications, 60(9), 34-44.
[13] Karpagavalli, S., & Chandra, E. (2016). A review on automatic speech recognition architecture and approaches. International Journal of Signal Processing, Image Processing and Pattern Recognition, 9(4), 393-404.
[14] Mansour, A. H., Salh, G. Z. A., & Mohammed, K. A. (2015). Voice recognition using dynamic time warping and mel-frequency cepstral coefficients algorithms. International Journal of Computer Applications, 116(2), 34-41.
[15] Mohammed, J. R., Shafi, M. S., Imtiaz, S., Ansari, R. I., & Khan, M. (2012). An efficient adaptive noise cancellation scheme using ALE and NLMS filters. International Journal of Electrical and Computer Engineering, 2(3), 325-332.