Indian Journal of Science and Technology, Vol 8(32), DOI: 10.17485/ijst/2015/v8i32/92116, November 2015
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Analysis on Quranic Accents Automatic Identification with Acoustic Echo Cancellation using Affine Projection and Probabilistic Principal Component Analysis

Noraziahtulhidayu Kamarudin[1*], S. A. R. Al-Haddad[1], Mohammad A. M. Abushariah[2], Shaiful Jahari Hashim[1], Asem Khmag[1] and Abd Rauf Bin Hassan[3]

[1]Department of Computer and Communication System Engineering, Faculty of Engineering, University Putra Malaysia, Selangor, Malaysia; hidayu.kamarudin@gmail.com, sar@upm.my, sjh@upm.edu.my, khmaj2002@gmail.com

[2]Department of Computer Information Systems, King Abdullah II School for Information Technology, The University of Jordan, Jordan; m.abushariah@ju.edu.jo

[3]Department of Foreign Language, Faculty of Modern Language and Communication, University Putra Malaysia, Selangor, Malaysia; raufh@upm.edu.my

Abstract

Background/Objectives: Audio recordings or live recitation is easy for noise and reverberation prone and contributes to the classification results. **Methods/Statistical Analysis:** Therefore, this paper proposed suitable usage of Affine Projection (AP) as adaptive filtering can be used to remove the echo and improve the percentage of results after the classification method takes place. **Findings:** The room impulse response is added for the signals in the study and AP algorithm is applies to remove the echo existing on each signal. Later, 161 signals then are converted to feature vector by Mel Frequency Cepstral Coefficients (MFCC) and the classification method using Probabilistic Principal Component Analysis (PPCA). This finding is to find the accuracies results based on audio on Quranic accents recitations after acoustic echo cancellation is done on the echoed signal. **Applications/Improvements:** The percentage of accuracies increases when acoustic echo cancellation is in used from the overall classification processes done to the segmented verses of the Quranic verse for Surah Ad-Duhaa. The implementation of AP as adaptive algorithm able to reduce the echo of Quranic accents and consistent output on the pattern classification and with 20dB of room impulse reverberation is applied and finally it can improve the results of classification to 96%.

Keywords: Acoustic Echo Cancellation, Adaptive Filtering, Affine Projection, Gaussian Mixture Model, Mel Frequency Cepstral Coefficents, Probabilistic Principal Component Analysis

1. Introduction

Echo and noise interruptions can happen from sound frameworks, response of frequency and masking with large peaks. The interruptions of noise[1,2] and music recordings, environment acoustics especially during system recordings will interrupts and degrading[3] the quality inside of speech or recording signals even to Quranic[4] recordings as well. The echoes too may influence the recordings procedure and it's viable for speaker, microphone and the transmission way. The speaker variable, voice signal is propagated and transmitted within different paths, and during the interruptions of elements from reverberation in room especially, it will decrease the voice signal quality and giving interference within the room space and dimensions. The signal from voice may contort and corrupt when the participant talks in front of the microphone during recordings or within live envi-

*Author for correspondence

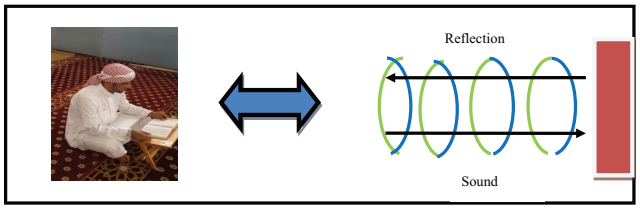

Figure 1. Reverberation condition.

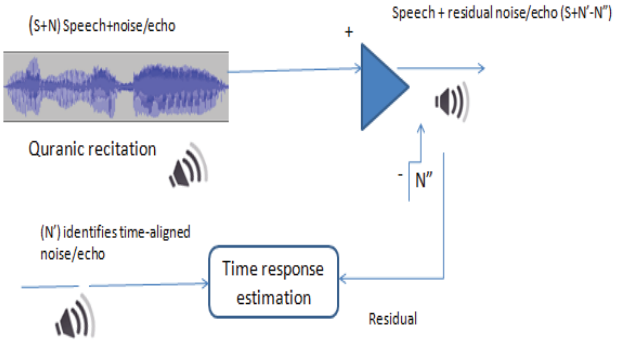

Figure 2. Acoustic echo cancellation process for removing echo in preprocessing stage.

ronment too as shown in Figure 1. Mainly the process of acoustic echo cancellation that takes place can be referred as in Figure 2.

2. Literature Review

Echo paths is normally measured based on adaptive filter[5] and echo reduction within transmitted signals too. Most of algorithms that in used for Acoustic Echo Cancellation (AEC) including Normalized Least Mean Square (NLMS)[6], Least Mean Squares (LMS), Affine Projection (AP), Recursive Least Squares (RLS) and Fast Recursive Least Squares (FRLS). Audio signals which recorded directly from speakers and comprises with desired speech and the environment and background noise too. Therefore, it is important for adaptive filters to continue change its characteristics for getting the maximum result for d (n) and actual output known as y (n); and for both of this is knows as cost function[8]. Acoustic Echo Cancellation (AEC)[7] is purposely to eliminate specific input signal d (n) by ensuring the e (n) will be at the most smallest value too. Finally the results will all be affected in terms of its accuracy especially whereby the signal classification to take place. In this matter of research too, convergence rate also important as it helps to determine the date where filters would converge to a resultant state. The desired characteristics for chosen and better convergence[9] are not depending for the performance of the adaptive system. In this case, the performance may act as vice versa example; stability decreased when convergence is increased while system will be much better and convergence[9] would decreases. For certain values too, the coefficients will change based on certain option to improve performance and focuses on the error signals. Different combination coefficients of digital adaptive filters are based from the error signals and updated accordingly. Techniques of Affine Projection (AP) has been used and managed to get 55dB for Echo Return Loss Enhancement (ERLE) and were getting better Signal to Noise Ratio (SNR) by implementing in noise cancellation for speech enhancement in their[10] research using AP for 24.07 dB compared to Least Mean Square 13.54 dB. Another techniques[6], the researchers used Affine Projection (AP) and Least Mean Square (LMS), for the analysis and managed to get higher Signal to Noise Ratio (SNR) and optimum response[10]. This algorithm has appreciable significance in speech processing as they managed to get 20.03 dB for AP while 13.59 dB for LMS.

3. Methodology

Figure 3 shows the current process for methodology and workflow which consists of three major phases provided and Quranic accents (Qiraat) speech signal that be used as main input of the whole processes. Most of the crucial process in this current studies, is preprocessing, whereby, this process will remove unnecessary echoes and noises from Quranic accents signal. These steps will involve adaptive algorithm AP within step 1 and 2. After the preprocessing stage takes place, the new clean signal will be collected again and be as input for second phase in step 3, and the feature extraction process in step 4 and 5. In this step, Mel Feature Cepstral Coefficients (MFCC), is the algorithm in used and converts that clean signal to feature vectors (step 6). Patten classification phase for the

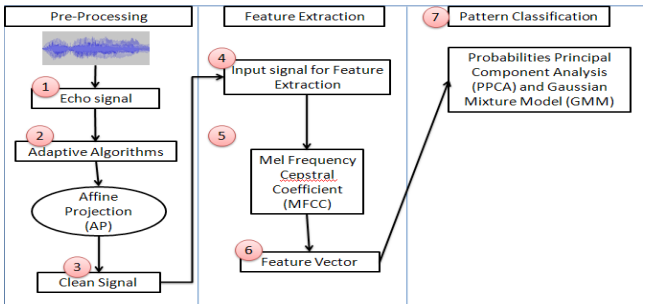

Figure 3. Research Workflow and Methodology.

third phase for the feature vectors. Within this steps, there are two method of classifiers in used; namely Gaussian Mixture Model (GMM) with Principal Component Analysis (PCA) and with GMM. The experimental results and accuracies are shown later in Figure 8.

3.1 Preprocessing

Simulated Room Impulse Response (RIR) shown in Figure 4, is used as the artificial impulse response for the Quranic accents signal. RIR is considered well establishes concepts to be used for echo[11] cancellation as it includes the parameters setups of reverberation time, room dimensions, and the source array distance. Other prevalent features include; sound database, microphone positions and the reflection time. In this study, there are three types of parameters that involved which includes: 1. Far end speech, 2. Random Delay for Near Speech, and 3. A single microphone, and the adaptive algorithm which used in this study is; Affine Projection (AP)[12]. Error signal, *e(n)*, is fed back to the algorithmic changes and reduced the cost function by the implementation of the adaptive filter, while unnecessary echoed signal that is uniformed of the maximum output of the adaptive algorithm filter in used. The desired signal is similar with adaptive filter output, when error signals turn to 0. Echoed signal will be removed completely while far end user would not be affected from returned signals[11].

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

The multiple input and updated weight vectors were earlier found by[12] for AP algorithm. In this study too, coefficient changes is updated based on equation for each iteration *n* as follows[8,13] in equation (1).

3.2 Feature Extraction Using MFCC

Mel Frequency Cepstral Coefficients (MFCC) is considered popular techniques and widely in used for speech recognition areas which relates to frequency domain within Mel scale and based on the human ear scale. The coefficients are based on the frequency domain features that is better in precision rather time domain features[14,15]. MFCCs is considering more on human auditory that commonly in used for automatic speech recognition systems[15]. It extracts important characteristics and parameters which similar for human hearing speech signal and also deemphasized other non related information. While

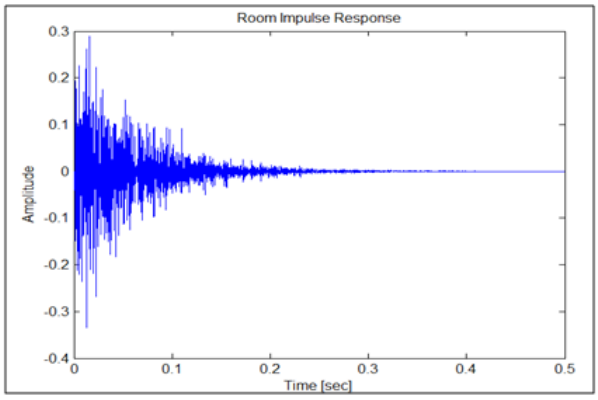

Figure 4. Simulated Room Impulse Response.

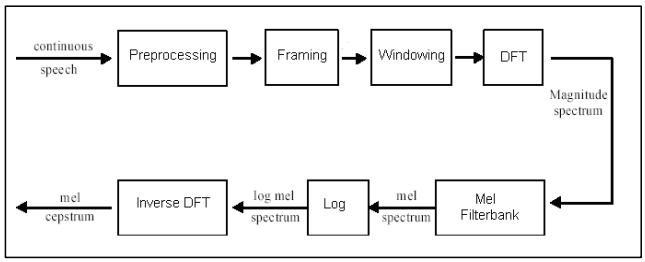

Figure 5. BlockDiagram of theComputation Processof MFCC[23].

for spectral feature, it provides complementary data of the MFCC feature during execution which contains the vocal source pitch and harmonic. MFCCs is beneficial for energy distribution, bandwidth of speech spectrum, voice and unvoiced excitation[16,17] and automatic speech recognition systems[20]. All the computational steps which is important for MFCC is presented in Figure 5. The whole process are included as the followings; preprocessing, framing, windowing using hamming window, performing Discrete Fourier Transform (DFT), the best spectrum perceived by human auditory system which performing the logarithm and taking the final inverse DFT for algorithm within magnitude spectrum[18,19].

3.3 Pattern Classification

For pattern classification, Probabilistic Principal Component Analysis (PPCA) with GMM as presented as in Figure 3. The description as explain in the sub section below.

3.3.1 Gaussian Mixture Model (GMM)

Gaussian Mixture Model (GMM) is famously in used in terms for machine learning[20], data mining, time series classification, image texture detection and speaker iden-

tification which generated from pool of Gaussian model together with mixture weights. From the mixture models, the maximum likelihood is estimated, and predicts test data with the largest probabilities. Using the concept of GMM, the expectation of mismatch and non mismatch is predicted between clean and mean vector of noise speech within each frames. And by using this concept, it shows the significant approaches for accuracy[21]. Expectation Maximization (EM) is performed by iterative for both Expectation (E) step and Maximization (M) step. The number of components within the GMM[22] is interrelated to each other. Number for each cluster within data points will be segregated to wrap local variations[30]. For Quranic accents (Qiraat), the information is captured within the work involved and different of the training number of components varies for each GMM state. Significant results within both Minimum Mean-Squared Error (MMSE) and Maximum Likelihood Estimation (MLE) are also achieved as the usage is supported[24] when in used for GMM mapping with condition of low pass filtering.

Expectation Maximization for this GMM is derived for auxiliary function[21].

Initial guesses of the parameters:

$$Q(\theta, \phi), \text{ where } \Phi = \{wjk\}$$

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

Expectation (E) Step: Compute the responsibilities:

$$w_{jk}^{(t)} = P(y_j = k | x_j, \theta^{(t)}) = \frac{\alpha_k^{(t)} \, p_k\left(x_j \middle| \theta_k^{(t)}\right)}{\sum_{i=1}^{K} \alpha_i^{(t)} \, p_i\left(x_j \middle| \theta_i^{(t)}\right)} \qquad (2)$$

3.3.2 Probabilities Principal Component Analysis (PCA)

Principal Component Analysis (PCA)[23] established a technique for dimensionality reduction which explored numerous texts on different variety analysis and processes by iteratively steps until convergence is achieved. Orthogonal projection for each data within different dimension and for the lower ones can cause the variance for the projected data to be expanding and useful for visualization, exploratory data analysis, data compression, image processing, pattern recognition and predicting time series. PCA is able to employ data into some reduced-dimensionality representation and algebraic manipulation of Maximum-Likelihood Estimators (MLE), the obtained results are standard projection for principal axes if desired. While in Probabilistic PCA (PPCA), the principal axes may be found increasing[24]. For the PPCA and the GMM, the algorithms used are as follows:

Table 1. Sampling parameters of the wave files in used

Sampling rate	→: 8000 Kbps
Bit-Depth (Bits)	→: 16
Channels	→: 1 Channel (Mono)

$$\log p(X) = -\frac{Nd}{2}\log 2\pi - \frac{N}{2}\log|C| - \frac{N}{2}\mathrm{Tr}(C - 1S) \qquad (3)$$

S is for covariance matrix while N for number of data points.

4. Experimental Work

In this analysis, the clean signal and echoed signal are then used for the classification purpose; in order to find the comparison of results acquired especially after the echo cancellation process is done. There are 161 totals of files that in used for this case study, where 60% of them are used for training purpose while another 40% are used for the testing purpose. The segmented verse that involved 1.والضحى 2.سجى 3.قلى 4.فترضى 5.فآوى 6.فهدى 7.فأغنى And the Quranic accents that involved here cover 6 types of accents which comprises: 1. Ad-Duri, 2. Al-Kisaie, 3. Hafs an A'asem, 4. IbnWardan, and 5. Warsh and 6. Qaloon. For each verse, the attributes of parameters that involved are as follows;

5. Discussion

Based on the experiments, the results acquired are very significance, and showing that echo cancellation is considers really important during pre-processing stage for signal analysis and speech identification. The whole process of identification is done using MFCC and PPCA and GMM as the technique for feature extraction and pattern classification like in stage 5 and 7 in Figure 3. The sample of signal in used for classification can be viewed in Figure 6 and 7 between echo and clean signals of the segmented verse for Quranic accents IbnWardan. During classification, echo signal is in used and acquire 90% of accuracy rate while for clean signal they achieved 96% of the signal accuracy after pattern classification takes place as shown in Figure 8.

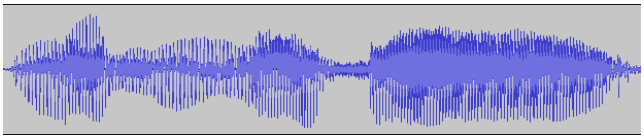

Figure 6. Signal of ' وَالضُّحَىٰ.. _qiraat Ibn Wardan' with echo.

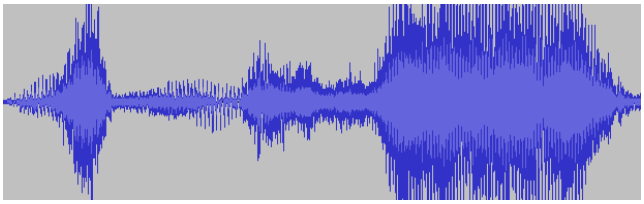

Figure 7. Signal of ' و الضحيٰ.. _qiraat Ibn Wardan' with echo.

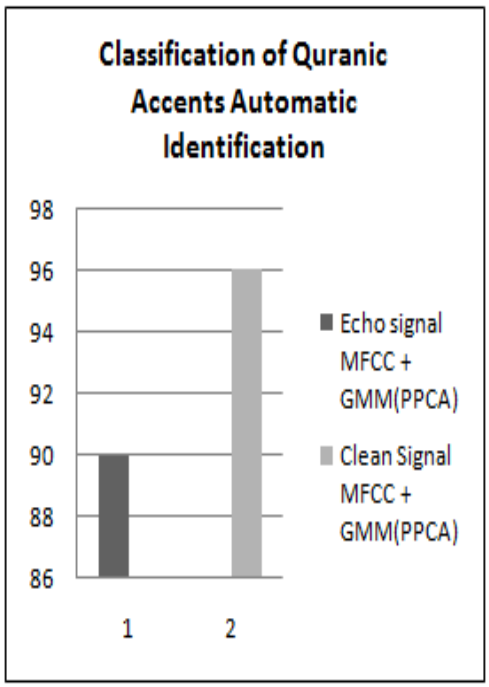

Figure 8. Results based on echo and clean signal after classification takes place.

6. Conclusion

Acoustic Echo Cancellation is considered a very important pre-processing stage as it improves the accuracies for Quranic accents identification. Therefore, for each Quranic verse recording, it shall be made sure that the steps shall include in the pre-processing stage with other steps like framing, windowing or noise cancellation. While, automatic identification for Quranic accents also plays a major role in improving understanding for the Muslims worldwide, as it improves their understanding while reciting or hearing Al Quran with different types of Quranic accents (Qiraat) recitation. Therefore, the whole implementation of identification of the Quranic accents, are valuable to justify and assured that only specific accents are classify based on the pattern of Quranic accents its belong to.

7. References

1. Ramli RM, Noor AOA, Samad SA. Adaptive line enhancer using affine projection algorithm for noise cancelling in speech. International Conference on Engineering and Built Environment; 2012. p. 1–6.
2. Ganesan V, Manoharan S. Surround noise cancellation and speech enhancement using sub band filtering and spectral subtraction. Indian Journal of Science and Technology. 2015; 8(33):77774.
3. Adapa NS, Bollu S. Performance analysis of different adaptive algorithms based on acoustic echo cancellation [Master Thesis]. Karlskrona, Sweden: Blekinge Institute of Technology; 2012.
4. Razak Z, Ibrahim NJ, Yamani M, Idris I, Tamil EM, Yakub M. Quranic verse recitation recognition module for support in j-QAF learning: A review. IJCSNS. 2008; 8:207–16.
5. Affandi A, Dobaie AM, Husain M. Digital filters design using Matlab with Graphical User Interface (GUI). Life Science Journal. 2014; 11(5):336–48.
6. Deepika M. Noise cancellation in speech signal processing using adaptive algorithm. International Journal on Recent and Innovation Trends in Computing and Communication. 1 Sep 2013; 1(9):743–6.
7. Stokes JW, Malvar HS, Way OM. Acoustic echo cancellation with arbitrary playback sampling rate. 2004 Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'04); 2004. p. 1–4.
8. Haykin S. Adaptive Filter Theory. 4th ed. Prentice Hall; 2002.
9. Sudhir VV, Murthy ASN, Rani DE. Acoustic echo cancellation using adaptive algorithms. International Journal of Advances in Computer Science and Technology. 2014; 3:248–52.
10. Hadei SA. A family of adaptive filter algorithms in noise cancellation for speech enhancement. International Journal of Computer and Electrical Engineering. 2010; 2(2):1793–802.
11. Liu KR, Hsieh SF, Yao K. Systolic block householder transformation for RLS algorithm with two-level pipelined implementation. IEEE Transactions on Signal Processing.1992; 40:946–58.

12. Ozeki K, Umeda T. An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties. Electronics and Communications in Japan. 1984; 67-A(5):19–27.
13. Diniz PSR. Adaptive filtering: Algorithms and Practical Implementations. 3rd ed. Boston, MA: Springer; 2008. ISBN: 978-0-387-31274-3.
14. Abushariah MAM. A vector quantization approach to isolated-word automatic speech recognition [Master Dissertation]. Malaysia: University of Malaya; 2006.
15. Jacobsen F, Juhl PM. Fundamentals of General Linear Acoustics. United Kingdom: John Wiley and Sons Ltd; 2013. p. 284.
16. Hosseinzadeh D, Krishnan S. On the use of complementary spectral features for speaker recognition. EURASIP Journal on Advances in Signal Processing. 2008; (1):258184.
17. Kamarudin N, Al-Haddad SAR, Rauf A, Hassan B, Hashim SJ, Nematollahi MA. Feature extraction using spectral centroid and mel frequency cepstral coefficient for quranic accent automatic identification. IEEE Student Conference on Research and Development (SCOReD); 2014. p. 1–6.
18. Khalifa O, Khan S, Islam MR, Faizal M, Dol D. Text Independent Automatic Speaker Recognition. 3rd International Conference on Electrical and Computer Engineering; Dhaka, Bangladesh. 2004. p. 561–4.
19. Chetouani M, Gas B, Zarader JL, Chavy C. Neural predictive coding for speech discriminant feature extraction: The DFE-NPC. ESANN'2002 Proceedings of European Symposium on Artificial Neural Networks; Bruges, Belgium. 2002. p. 275–80.
20. Rao KS, Koolagudi SG. Robust emotion recognition using spectral and prosodic features. Springer Briefs in Electrical and Computer Engineering; 2013. p. 17–46.
21. Sena E, De Antonello N, Moonen M. On the modeling of rectangular geometries in room acoustic simulations. IEEE/ACM Transactions on Audio, Speech and Language Processing. 2015; 23(4):774–86.
22. Ari C, Aksoy S, Arıkan O. Maximum likelihood estimation of Gaussian mixture models using stochastic search. Journal Pattern Recognition. 2012; 45:2804–16.
23. Tipping ME, Bishop CM. Mixtures of probabilistic principal component analyzers. Neural Computation Microsoft Research. 1999; 11(2):443–82.
24. Kuttruff H. Room Acoustics. 4th ed. Abingdon: Taylor and Francis Group U.K.: SPON; 2000.