



2nd GLOBAL CONFERENCE on LINGUISTICS and FOREIGN LANGUAGE TEACHING,  
LINELT-2014, Dubai – United Arab Emirates, December 11 – 13, 2014

## Computer-Aided Training for Quranic Recitation

Huda M. Arfan Tabbaa<sup>a\*</sup>, Bassel Soudan<sup>a</sup>

<sup>a</sup>*Department of Electrical and Computer Engineering College of Engineering, University of Sharjah Sharjah, United Arab Emirates*

### Abstract

Computer Aided Language Learning (CALL) systems have gained popularity due to the flexibility they provide in empowering students to practice their language skills at their own pace. Detection/Correction of specific pronunciation error is an important component of an effective language learning system. Learning the correct rules of the Holy Quran recitation is important to every Muslim. In this work, we developed a Computer Aided Quranic Recitation Training system to detect errors in continuous recitation of Holy Quran and increase the accuracy of the error detection. We have integrated Automatic Speech Recognition (ASR) and classifier-based approach to detect recitation errors. Error detection is done in two successive stages: first, an HMM-based ASR recognizes the recitation, detects the insertion, deletion and substitution of phones and provides phonetic time alignments, and then classifier based approach is used to distinguish between confusing phones to achieve improved detection rate. In this implementation we implemented only two classifiers, one to discriminate between emphasized and non-emphasized utterances of the letter “R” in Arabic, and the other to distinguish between closely related, often confused letter pronunciations. The results show, that the system has achieved a 91.2% word-level accuracy.

© 2015 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of Academic World Research and Education Center.

*Keywords:* Computer-Aided Language Learning; Automatic Speech Recognition, Classifier-based approach

### 1. Introduction

Learning to recite the Quran properly is important for every Muslim. Quranic recitation is typically learned through extensive practice sessions with a teacher who listens to the learner’s recitation, identifies recitation errors, and instructs the trainee with the proper corrections. The process is repeated until the recitation rules become second nature to the trainee. This learning method requires a dedicated teacher for each learner or group of learners, and requires extensive training time until the trainee reaches the proper degree of proficiency.

\* Huda M. Arfan Tabbaa Tel: +3453643  
E-mail:[hudatabbaa@gmail.com](mailto:hudatabbaa@gmail.com)

If we analyze the training process, we find that the teacher spends a certain amount of time in the beginning educating the learner about the recitation rules (education phase), then the rest of the time is spent mainly with the learner reciting and the teacher correcting the errors (training phase). We believe that the education phase of the process cannot be automated and requires the explicit involvement of the teacher. On the other hand, the training phase extends over a very long period of time and as the skill of the learner improves, the need for the active involvement of the teacher diminishes.

This paper describes a multi-stage system for detection of errors in a learner's recitation. The system analyzes a learner's recitation, recognizes recitation errors, and provides informative feedback to assist the learner in correcting these errors. The system handles both common mispronunciation errors as well as violations of recitation rules.

The system's methodology can also be applied to language learning in general. The implemented system can follow the continuous readings of a learner, recognize errors and indicate corrections. The system's concept works whether the learner is reading through a story, a newspaper article, or any fixed known composition.

Technically, both segmental and supra-segmental features of the speech signal determine the accuracy of pronunciation. The segmental features are concerned with the distinguishable sound units of speech, i.e. phones. Whereas the supra-segmental features of speech comprise of intonation, pitch, rhythm and stress. In this work, we focus on segmental features, which provide learners with detailed corrective feedback (Witt & Young, 2000).

In the following section we will present previous work of related importance to this topic, then in section 3 we will discuss technical background information necessary for appreciating the problem, in section 4 we present the design of our proposed system, then in section 5 we discuss implementation details, in section 6 we discuss results and demonstrate the effectiveness of the proposed system, then in section 7 we presents conclusions and closing remarks.

## 2. Previous Work

Many different approaches have been used to detect pronunciation errors (Witt & Young, 2000) (Witt S. M., 2012). Comparison based approaches use dynamic programming algorithms to compare between the correct word pronunciation and the utterance under evaluation (Lee & Glass, 2012). Lee implemented a comparison-based framework to detect word-level mispronunciation. In this framework, the utterance of the learner and the teacher are compared through dynamic time warping (DTW). Mispronunciation detection is done by locating poorly matching alignment regions based on features extracted from spectral representation. The main drawbacks are the low accuracy and the restriction of having to pronounce the utterance in exactly the same manner as the reference. Any minor variation is flagged as a pronunciation error.

Confidence scoring approach depends on assessing the pronunciation quality of individual phones by calculating a confidence rating that a phone has been correctly recognized by the training software (Witt & Young, 2000). Kim et al (Kim, Franco, & Neumeyer, 1997) presented three scores that have been extensively used in research: a Hidden Markov Model (HMM) based log-likelihood score, an HMM-based log-posterior score, and a score based on segment duration. HMM-based log-posterior score is better than the others and has been extensively used (Witt & Young, 2000; Witt, 2012). Witt and Young (Witt & Young, 2000) proposed a goodness of pronunciation (GOP) score based on the log-likelihood scores. Including knowledge of the learner's native tongue leads to an improvement in the detection of mispronunciation errors (Harrison & M., 2009; Ito, 2007). Researchers have built an extended recognition network that includes both the correct pronunciation as well as common pronunciation errors (taking into consideration the particular learner's native language effect). Confidence scoring has the advantage of being very easy to compute. However, it relies on HMMs, which are not very successful in discriminating inter-class data, especially between classes that are spectrally similar (Strik H. , Truong, F, & Cucchiari, 2009).

Several studies have employed classifiers to discriminate specific error possibilities of specific phone pairs (Witt S. M., 2012). Specific acoustic features have been used in automatic error detection of frequent mispronunciation for learners of Dutch (Strik , Truong, & Cucchiari, 2009; Strik , Truong, de Wet, & Cucchiari, 2007). The acoustic properties of these pronunciation errors were examined to define a number of discriminative acoustic features to be used to train and test classifiers. The authors compared the scoring accuracy for four different classifiers (Decision Tree, Linear Discriminative Analysis (LDA) using specific acoustic phonetic features, LDA using Mel Frequency

Cepstrum Coefficients (MFCCs), and GOP confidence score). Their work demonstrated that LDA based classifiers can outperform confidence scoring. The authors also compared between MFCCs (Mel-Frequency Cepstral Coefficients) and acoustic phonetic features. They found that MFCCs with an LDA classifier is better than acoustic phonetic features if there is no mismatch between training and testing data (for example, trained on non-native speech and tested on non-native speech). On the other hand, when there is mismatch between the training and testing data, it is the other way around. Patil (Patil & Rao, 2012) also investigated the classifier-based approach in the context of mispronunciation of Hindi learners. He restricted his work to a specific type of error relating to an aspirated stop. He used the GMM (Gaussian Mixture Model) classifier to compare between MFCC features and acoustic phonetic features. He demonstrated that acoustic phonetic features provide better discriminability between correct and incorrect utterances. Overall, the classifier-based approach has been shown to outperform previous approaches, but its drawbacks are: common errors have to be known and a separate classifier is necessary for each error type (Witt S. M., 2012).

Abdou (Abdou, 2006) developed a system specifically for training on Quranic Recitation. The system uses confidence scoring to detect the most common recitation errors. The calculated confidence scores determine if each of the recognized phones should be accepted or rejected, or if the user is prompted to repeat the recitation. The system correctly identified 62.4% of the pronunciation errors, reported "Repeat Request" for 22.4% of the errors and falsely accepted 14.9% of total errors. Hammady (Hammady, 2008) noted that the main reason for the low performance was the weak confidence measure, which could not discriminate robustly between many phones that have very close acoustic models. Hammady tried to overcome the problem by implementing 5 HMM classifiers and merging their results into a single score. He used MFCCs as the input to the classifiers. He proposed a classifier for each articulation feature; where each features can have only two possible states: the feature and its opposite (voicing/unvoicing, elevation/lowering, plosive/fricative, fluency/desisting, adhesion/separation). Phone dependent thresholds were used to discard results from poor feature models. However, this approach still could not discriminate between phones that have same articulation features such as the Arabic "s" and "th".

In this work we will describe a system that detects pronunciation errors in two successive stages. First, HMM-based Automated Speech Recognition (ASR) engine will be used to detect the insertion, deletion and substitution of some phones as well as generating phonetic time alignments. Then, the classifier-based approach will be used to distinguish between confusing phones to improve the detection rate. The implemented system combines the speed and simplicity of the scoring approach with the improved accuracy and discriminability of the classifier-based approach.

### 3. Background

The system described in this paper shares numerous characteristics with systems for learning foreign languages. While Quranic Recitation cannot be considered a "foreign language", the process requires similar repetitious practice until the learner masters the proper articulation point for the different letters and the different pronunciation rules.

Computer-Aided Language Learning (CALL) systems encompass applications and approaches for teaching and learning foreign languages using a virtual learning environment (Tuncer, 2009). CALL systems provide an effective learning environment so that students can practice on their own pace. A major component of modern day CALL systems is the Computer Aided Pronunciation Training (CAPT) sub-system. While CALL systems are multifaceted and assist the learner with the numerous aspects of language learning, CAPT focuses specifically on detecting mispronunciation errors in the learner's speech (Neria, Micha, Gerosaa, & Giuliana, 2008). CAPT can be particularly beneficial for second language learning: not only does it provide a private, stress-free environment for self-paced practice; it can also provide individualized, instantaneous feedback (Neri, Cucchiarini, & Strik, 2002). It is the CAPT sub-system's concentration on mispronunciation detection that is of interest in regards to the proposed system. Most state-of-the-art CALL and CAPT systems utilize an ASR engine to convert the speech signal to a sequence of words for evaluation.

The speech recognition process can be divided into three stages:

1. Feature extraction stage: The speech signal is transformed into a sequence of acoustic feature vectors. Each vector represents the signal information in a small time window. The most common features used in ASR are the Mel Frequency Cepstral Coefficients (MFCCs), which approximate the sensitivity of the human ear (Jurafsky D. , 2008).
2. Training stage: In this stage, the parameters of a set of HMMs are estimated using training utterances and their associated transcriptions.
3. The Recognition stage: The recognizer uses the trained HMMs to recognize the speech. The recognizer requires three supporting modules in order to recognize speech (Acoustic Models, Language Model, and Lexicon).

### 3.1. HMM-based Speech Recognition

The recognizer uses probabilistic implementation to answer the question: “What is the most likely sentence out of all sentences in the language given a certain acoustic vector?”

An acoustic vector sequence is denoted by  $O = o_1, o_2 \dots o_t$ . Similarly, a sentence is treated as a sequence of words as  $W = (w_1, w_2 \dots w_t)$ . The decoder needs to find the sentence  $\hat{W}$  that satisfies the relationship in equation (1) below:

$$\hat{W} = \underset{W \in L}{\operatorname{argmax}} P(W|O) \quad (1)$$

Where  $P(W|O)$  is the probability of having the sequence of words  $W$  given the observation vectors  $O$ , and  $L$  is the set of all possible words in the given language. The optimal word sequence  $\hat{W}$  can be computed by utilizing Bayes' Formula to compute the probability of word sequence given the observation –  $P(W|O)$  – as shown in equations (2) below (Jurafsky D. , 2008):

$$\hat{W} = \underset{W \in L}{\operatorname{argmax}} \frac{P(O|W).P(W)}{P(O)} = \underset{W \in L}{\operatorname{argmax}} P(O|W).P(W) \quad (2)$$

The most probable sentence  $W$  given some acoustic sequence  $O$  can be computed by taking the product of two probabilities for each sentence, and choosing the sentence for which this product is greatest.  $P(W)$ , the prior probability, is computed by the language model. While  $P(O|W)$ , the observation likelihood, is computed by the acoustic model.  $P(O)$  is the probability of the acoustic observation sequence, which doesn't change for each sentence. Therefore, it can be ignored.

## 4. Mispronunciation Detection in Quranic Recitation

This research describes the implementation of a system to detect the mispronunciation in continuous Quranic recitation. The system detects errors in phone pronunciation as well as violations of recitation rules.

The next subsection discusses the special issues associated with Quranic Recitation and the additional difficulties it poses to automated speech recognition systems.

#### 4.1. *Quranic Recitation*

The rules of “Tajweed” (the set of rules that dictate the proper pronunciation and melodic intonation of Quran recitation) introduce additional difficulties to a speech recognition system as compared with simply learning a new language. This section briefly discusses the additional requirements imposed by the rules of Tajweed.

Phonetics of Arabic language state that the standard Arabic language has 28 consonants plus 3 short vowels and their 3 long counterparts (Abdou, 2006). However, the rules Tajweed and its melodic rhythm require the handling of additional sounds. The following are examples of Tajweed rules that may affect the speech recognition system:

1. Prolongation – is the prolongation of the sound of a vowel.
2. Nasalization – a nasal sound that accompanies the vocalization of some letters.
3. Vibration – an echo noise that should be produced after any non-vowelized occurrence of five letters of the Arabic language.
4. Emphasis – a heaviness that enters the body of the letter, so that the mouth is filled with its reverberation.

Recitation rules that govern supra-segmental features of the Holy Qur'an are few and their effect is minor compared to rules governing segmental features. Therefore, previous work concentrated mainly on the segmental features of Quranic recitation.

Tabbal (Tabbal, El Falou, & Monla, 2006) focused on identifying the aspects of the rules of Tajweed that affect the recognition phase. They considered the prolongation as the repetition of the vowel n-corresponding times. The same consideration was used for the nasalization. Abdou (Abdou, 2006) added a new phone to identify vibration and used a new phoneme to identify the emphatic pronunciation of the letter R.

#### 4.2. *Proposed System*

Previously, researchers have mostly employed HMM based confidence scoring to detect recitation errors. HMMs have demonstrated reasonable success in modelling intra-class data. However, HMMs are not very successful in discriminating between classes, especially between classes that are spectrally similar. An investigation of the different mispronunciation detection approaches reported in the literature has yielded the following findings:

1. Comparison based approaches are not suitable for this application due to:
  - a. The relatively low accuracy rate of comparison-based systems (Lee & Glass, 2012).
  - b. The need for a reference for comparison, which is not practical for this application as Quranic recitation varies depending on where the reciters elect to stop.
  - c. Comparison-based approach detects errors at the word and syllable level. It is incapable of detection at the phone level as required for detection of pronunciation errors in Quranic recitation.
2. Confidence scoring approach based on HMM scores is not suitable for discriminating between spectrally similar phones such as “t” and “d” (Strik H. , Truong, F, & Cucchiari, 2009).
3. Classifier approach has the highest accuracy in discriminating between different sounds. However, it is prohibitively expensive to build a classifier to discriminate between each pair of phones.

To overcome the problems of the different individual approaches, our proposed system integrates the ASR and classifier approaches to build a multi-stage system for the detection of pronunciation errors in Quranic recitation.

Figure 1 below illustrates the “block diagram” of the proposed system.

The linguistic generation module is used for the automatic generation of grammar and pronunciation hypotheses (correct pronunciations and common pronunciation errors) in addition to the diagnosis of errors. The input of this module is an orthographic transcription of the recitation piece. It was broken at page boundary in this implementation for size considerations. The output of the linguistic generation module is in the form of two files: a grammar file and a lexicon. These files become the reference inputs to the speech recognition module. The linguistic generation process can be performed offline and the linguistic data may be stored on a storage medium, as both the lexicon and the grammar are not user-dependent.

An important problem in continuous speech recognition is the determination of speech periods within a given audio signal. The “Speech/Silence detection” module examines the audio signal and identifies speech periods from periods of silence. Then, speech features are extracted in the “Feature Extraction” module.

“Mispronunciation Detection” detects pronunciation errors in two successive stages:

- **The Recognition Stage:** an HMM speech recognition system will be used in this stage to recognize the spoken recitation. The inputs to this stage comprise of: the extracted features, the linguistic data of the selected Quranic page and acoustic models. This stage detects the insertion, deletion and substitution of phones and determines the phonetic time alignments.
- **The Pronunciation Verification Stage:** Discrimination classifiers are used in this stage to verify the pronunciation of some specific phones (such as emphasized and non-emphasized phones).

Finally, the “Feedback Generation” Module presents the result to the user.

## 5. Implementation Details

### 5.1. Speech Recognition

CMU Sphinx was used in this research to train a 3-state, tri-phone, continuous HMM as the first stage of the system (speech recognition).

The training/testing database is collected from telephone calls into a television program where a recitation scholar recites a page of the Quran then listens to students’ recitations and identifies errors in the recitation. The database contains recitations from 100 different female and 68 different male callers. The database contains more than 7.5 hours of recorded recitations. The database was divided into training and testing data, 6.5 hours of the database are used for training the HMM and the remaining one hour was used as the test database.

The speech recognition system achieved word level accuracy of about 97.6%. That is, the HMM-based speech recognizer was able to identify the spoken words correctly in 97.6% of the cases (without regard to whether there were any phone pronunciation errors).

### 5.2. Pronunciation Verification

In order to determine the phone-level accuracy, classifiers for a small number of specific scenarios were implemented. Mainly, two classifiers were needed:

1. An “R-Classifier” to discriminate between emphasized/non emphasized pronunciations of the letter R.
2. A “T-Classifier” that is applied to very closely sounding letters that emanate from the same articulation point and are often confused by student reciters.

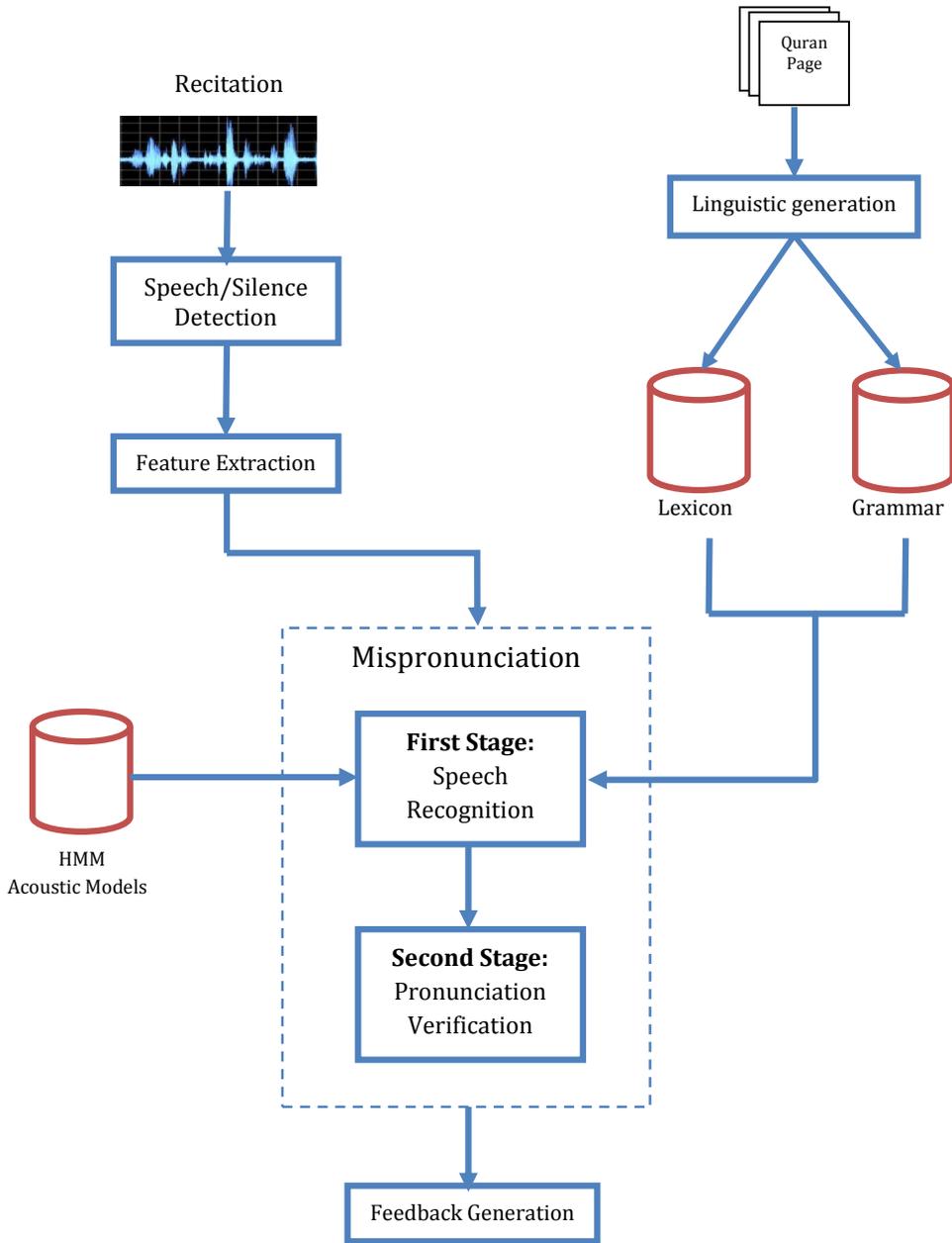


Figure 1 The block diagram of the proposed system.

**WEKA** (Waikato Environment for Knowledge Analysis) was used to build the classifiers. Four different machine-learning algorithms were tested and the most successful was chosen for each situation. The Algorithms that were evaluated are:

1. Support Vector Machine (SVM)
2. Neural Network Multilayer Perceptron (MLP)
3. Bagging
4. Random Committee

Wrong pronunciations of words were added to the lexicon in order to compare the accuracy of the trained HMM and other machine-learning algorithms. The wrong pronunciations were generated by changing a letter from the emphasized state to a non-emphasized state (and vice-versa).

### 5.2.1. R-Classifier

The training database contains 530 instances of the letter “R”, 298 are emphasized and the others are not. These were used to evaluate the different machine-learning algorithms available to determine the most appropriate algorithm to use for this classifier.

Figure 2 below shows the accuracy of the different machine-learning algorithms in differentiating between emphasized and non-emphasized letters. The results show that the HMM was able to discriminate between emphasized/non-emphasized utterances of the Arabic letter “R” in only 67% of the instances. All machine-learning algorithms fared much better. This corroborates the premise that HMM alone would not be sufficient to identify difficult pronunciation errors and that specific classifiers would be needed for these scenarios to achieve the sought error detection accuracy.

Based on the results of the different machine-learning algorithms, it was decided to use SVM for this classifier as it had the highest accuracy and F-measure values.

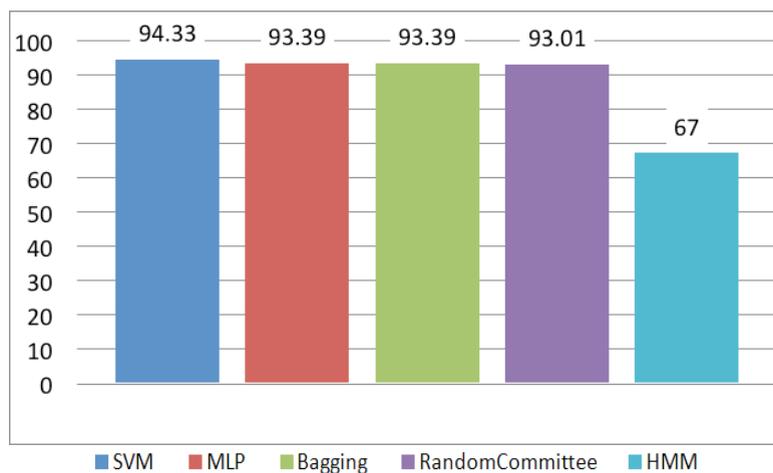


Figure 2. Accuracy of machine-learning algorithms when used in R-Classifier

### 5.2.2. T-Classifier

There are letters in the Arabic language whose articulation points are close and whose pronunciation is commonly confused by student reciters (such as the pronunciation of the phones “Sa” and “Su”). The main difference in the utterance of these similar sounding letters is in the amount of heaviness attached to the pronunciation. The T-classifier was built and added to the system to discriminate between the trainees’ utterances of these letters and ensure that the proper letter sound was produced.

The training database contained 1738 instances of these letters and Figure 3 below shows the results of using the different machine-learning algorithms to discriminating between these confusing letters. Again, an HMM-only approach fared quite badly compared to the machine-learning algorithm based classifiers.

While the machine-learning algorithms produced closely similar results in the previous case, their results varied quite a bit in this scenario. Based on the results in Figure 3, it was decided to use Random Committee for the T-classifier as it exhibited the highest accuracy and F-measure values.

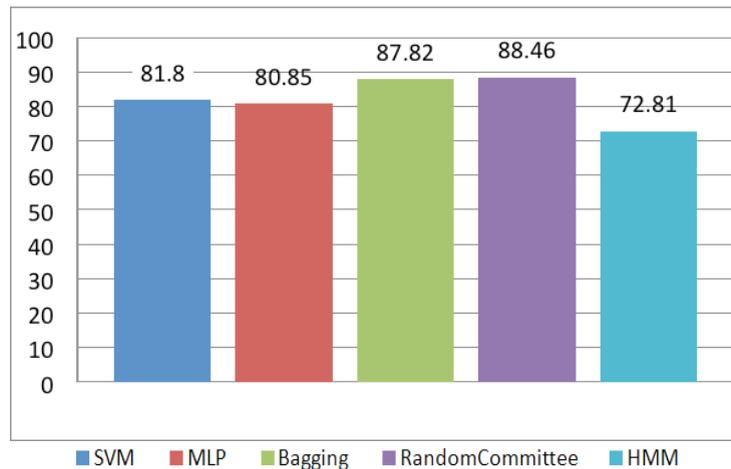


Figure 3. Accuracy of machine-learning algorithms when used in T-Classifer

## 6. Final Results

The system has been tested on 60 minutes of continuous recitation from 18 different female and 14 different male student reciters. The test database contains 2,689 words comprising a total of 23,198 phones.

It was mentioned earlier that the HMM-based speech recognition system was able to achieve a word-level accuracy of 97.6%. However, if we consider that a correctly recognized word containing a phone-level pronunciation error is actually a mispronounced word, then the real word-level accuracy of the HMM-based system comes out to be only 84 %.

After the inclusion of the two classifiers, the number of phone-level false-positive and false-negative cases was reduced. The real word-level accuracy has improved to 91.2%.

## 7. Conclusion

This work described an implementation of a Computer Aided Quranic Recitation Training system to detect errors in continuous recitation of the Holy Quran. The system integrates Automatic Speech Recognition (ASR) and classifiers to improve the detection rate. Error detection is done in two successive stages: first, an HMM-based ASR recognizes the recitation, detects the insertion, deletion and substitution of phones and provides phonetic time alignments, and then classifiers are used to distinguish between confusing phones to achieve improved detection rate.

This implementation describes the use of only 2 classifiers, one to discriminate between emphasized and non-emphasized utterances of the Arabic letter “R”, and the other to distinguish between closely related, often confused letter pronunciations. The results show that an HMM-only approach was able to achieve an accuracy of only 67% in discriminating between emphasized/non-emphasized utterances of the Arabic letter “R” and an accuracy of only 72% in discriminating between the utterances of highly confusing sounds combinations such as “Sa” and “Su”.

Including machine-learning based discriminating classifiers for these specific scenarios improved the system’s word-level accuracy from 84% to 91.2%. Additional classifiers are being considered to further improve the accuracy of pronunciation and Tajweed error detection in continuous Quranic recitation.

## References

- Abdou, S. M. (2006). Computer aided pronunciation learning system using speech recognition techniques. *INTERSPEECH*.  
 Abdou. (2006). Computer aided pronunciation learning system using speech recognition techniques. *INTERSPEECH*.

- Hammady, H. e. (2008). An HMM system for recognizing articulation features for Arabic phones. *Computer Engineering & Systems ICCES 2008. International Conference on IEEE* .
- Harrison, & M., A. (2009). Implementation of an extended recognition network for mispronunciation detection and diagnosis in computer-assisted pronunciation training . *Proc. of SLATE* .
- Ito, A. (2007). Pronunciation error detection for computer-assisted language learning system based on error rule clustering using a decision tree . *Acoustical science and technology* .
- Jurafsky, D. (2008). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. Upper Saddle River: Prentice Hall.
- Jurafsky, D. e. (2008). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. Upper Saddle River: Prentice Hall.
- Kim, Y., Franco, H., & Neumeyer, L. (1997). Automatic pronunciation scoring of specific phone segments for language instruction . *In Eurospeech* .
- Langlais, P., Öster, A. M., & Granström, B. (1998). Phonetic-level mispronunciation detection in non-native Swedish speech. *In ICSLP*.
- Lee, A., & Glass, J. (2012). A comparison-based approach to mispronunciation detection . *Spoken Language Technology Workshop (SLT), IEEE* .
- Neri, A., Cucchiari, C., & Strik, H. (2002). Feedback in Computer Assisted Pronunciation Training: Technology Push or Demand Pull. *Proceedings of the International Conference on Spoken Language Processing (ICSLP)* .
- Neria, A., Micha, O., Gerosaa, M., & Giuliana, D. (2008). The Effectiveness of Computer Assisted Pronunciation Training (CAPT) for Foreign Language Learning by Children. *Computer Assisted Language Learning* .
- Patil, V., & Rao, P. (2012). Automatic pronunciation assessment for language learners with acoustic-phonetic features . *24th International Conference on Computational Linguistics* .
- Strik, H., Truong, K. P., de Wet, F., & Cucchiari, C. (2007). Comparing classifiers for pronunciation error detection . *In Interspeech* .
- Strik, H., Truong, K., de Wet, F., & Cucchiari, C. (2009). Comparing different approaches for automatic pronunciation error detection. *Speech Communication* .
- Strik, H., Truong, K., F, d. W., & Cucchiari, C. (2009). Comparing different approaches for automatic pronunciation error detection. *Speech Communication* .
- Tabbal, H., E. W., & B., M. (2006). "Analysis and implementation of a" Quranic" verses delimitation system in audio files using speech recognition techniques. *Information and Communication Technologies, 2006. ICTTA'06. 2nd. Vol. 2. IEEE* .
- Tabbal, H., El Falou, W., & Monla, B. (2006). Analysis and implementation of a" Quranic" verses delimitation system in audio files using speech recognition techniques. *Information and Communication Technologies, 2006. ICTTA'06. 2nd. Vol. 2. IEEE* .
- Truong, K., Neri, A., & Cucchiari, C. (2004). Strik, H. Automatic pronunciation error detection: an acoustic-phonetic approach . *In InSTIL/ICALL Symposium* .
- Tuncer, C. A. (2009). Learning and teaching languages online: a constructivist approach. *Novitas-Royal 3.1 Novitas-Royal 3.1* .
- Witt, S. M. (2012). Automatic error detection in pronunciation training: Where we are and where we need to go. *In International Symposium on Automatic Detection of Errors in Pronunciation Training, Stockholm, Sweden* .
- Witt, S. M., & Young, S. J. (2000). Phone-level pronunciation scoring and assessment for interactive language learning . *Speech communication* .