A Complete Mispronunciation Detection System for Arabic Phonemes using SVM

Muazzam Maqsood*, Hafiz Adnan Habib, Tabassam Nawaz, Khurram Zeeshan Haider
University of Engineering and Technology Taxila, Pakistan

Summary
Computer Assisted Language Learning Systems have gained a lot of attention in recent decades. Mispronunciation detection is probably the most important feature of these systems. It helps user to find out their pronunciation mistakes and provide useful feedback related to that mistake. Mispronunciation detection systems can be categorized in two classes; Posterior Probability based and Classifier based systems. In this paper pronunciation assessment problem is formulated as a classification problem. This research paper explores the Acoustic Phonetic Features (APF) rather than traditional Confidence Measure based scores for mispronunciation detection. Support Vector Machines (SVM) is used as a classifier to detect pronunciation mistakes. As a test case five Arabic phoneme are tested for mispronunciation detection. APF based classifier produced excellent results and give average accuracy of 97.5%. The proposed system outperforms the existing systems that have been developed for Arabic phonemes.

Key words:
Computer Assisted Language Learning Systems, Mispronunciation Detection, SVM, Acoustic Phonetic Features

1. Introduction:

In this era, the world has become a global village. Modern technologies has made it possible for everyone to stay in contact with each other while staying in different regions of the world. This has led to an increase in demand to learn new languages. Arabic is the 5th largest language in terms of number of native speakers. There are over a billion of speaker who speaks Arabic. Arabic is also important for all the Muslims across the world because their religious book “The Holy QURAN” is in Arabic. So all the Muslims across the world use Arabic for their religious obligations. Computer Assisted Language Learning (CALL) systems have gained a lot of attention because of the advancement in Artificial Intelligence. Pronunciation training is an important feature of CALL systems. These systems detect pronunciation mistakes from user’s speech and provide them useful feedback [1-5].

Pronunciation mistakes can be grouped in two classes; complete mispronunciation and partial mispronunciation. In Complete mispronunciation, a phoneme is mispronounced completely different. While in partial mispronunciation detection, a phoneme is not completely mispronounced but there is only slight variations from actual pronunciation of the phonemes. In this research, a complete mispronunciation detection system is designed for Arabic phonemes. There are two types of mispronunciation detection systems used; posterior probability based and classifier based. When complete mispronunciation detections are considered, classifier based approaches can be more useful as compared to other methods. In this case, mispronunciation detection problem is considered as 2-class classification problem. [1-2]

There has been many systems developed by different researchers for mispronunciation detection. In posterior probability based methods, different approaches has been designed. Goodness of Pronunciation (GOP) has been proposed by Witt et al. [6] to check the pronunciation quality and it is considered as a benchmark method for posterior probability based methods. There are different variations of GOP proposed by researchers [1-12]. Al-hindi et al. [18] developed a GOP based mispronunciation detection system for 5 Arabic phonemes. In classifier based approaches, Troung et al. [4] has developed a system to detect mispronunciation for Dutch using decision trees and Linear Discriminant Analysis (LDA). Wei et al. [2] developed a system for Mandarin syllables using SVM classifier. Strik et al. [5] carried out a comparative analysis between four different techniques of mispronunciation detection. These approaches included GOP, LDA with MFCCs, decision trees and LDA with acoustic phonetic features (APF). The results show that LDA with APF outperformed GOP methods and all other methods. The advantage of classifier based approach is that more acoustic phonetic features can be tested and incorporated in mispronunciation detection systems. [5]

Manuscript received March 5, 2016
Manuscript revised March 20, 2016

Table 1: Details for dataset used for this experiment

	No. of Speakers			
	Adult Male	Adult Female	Children	total
Native	40	10	10	60
Non-Native	30	05	05	40
Total	70	15	15	100
	No. of Phonemes			
	Adult Male	Adult Female	Children	Total
Native	200	50	50	300
Non-Native	150	25	25	200
Total	350	75	75	500

In this paper, a classifier based approach is proposed for Arabic mispronunciation detection system. SVM is considered as a good binary classifier due to its general nature which is suited for 2-class classification problems. Different acoustic features have been used to train the classifiers. For a test case, five Arabic phonemes have been tested and results are very promising.

The rest of the paper has been organized as follows: section 2 explains the materials and methods section and section 3 covers results and discussion followed by conclusion and future work.

2. Materials and Methods:

2.1 Dataset:

Arabic is the 5th largest language in terms of native speakers but still there hasn't been much of the work done for pronunciation training. This is the reason that there is no standard dataset available for Arabic. Most of the researchers make their own datasets from extracting the required information from recorded Arabic sentences available on internet. In this research a dataset has been recorded from the Pakistani speakers who have been learning Arabic as their second language. The dataset was recorded from 100 speakers, these speakers include from very proficient speakers to beginners. These speakers were asked to read Arabic phonemes in office environment. Details of dataset is presented in Table-1.

Dataset labeling is a time consuming yet important step in classification process. In order to train the classifier, labelled dataset is required. 3 different Arabic language experts have been asked to label the data. These language expert has extensive knowledge of Tajweed, they rate the pronounced phonemes as correct or incorrect and a phoneme is labelled as correct or incorrect if at-least 2 language experts agree on the same label. Details of Arabic Phonemes used in this research are given in Table-2.

Table 2: Details of all Arabic Phonemes used in this Research

Letter	**Name**		**IPA Symbol**
ث	*thaa'<*	ثَاْءٌ	[θ]
ح	*haa'<*	حَاْءٌ	[ħ]
ص	*saad*	صَاْدٌ	[sˤ]
ض	*daad*	ضَاْدٌ	[d̪ˤ]
ظ	*zaa'<*	ظَاْءٌ	[ðˤ]

2.2 SVM based Mispronunciation Detection System

CALL systems are usually based on Posterior Probabilities or Acoustic Phonetic Features. Most of the work done in this field are based on Posterior Probabilities which are calculated using Confidence Measure (CM). These CM are originally designed for Automatic Speech Recognition (ASR) toolkits. Very little emphasis has been given to APF based mispronunciation detection classifiers. The reason behind such little work is the fact that suitability of acoustic features for mispronunciation detection has not been researched properly. In this research, more APF features have been used to train classifier. In this way suitability of APF have also been analyzed. The APF used here are Root Mean Square Energy (RMSE), MFCCs along with its first and second derivative, Low Energy, Spectral features, zero-cross rate, Pitch and statistical features that includes mean, standard derivation, slope, periodic frequency, periodic amplitude, periodic entropy [13-17]. Then global statistical features have also been calculated for each of these feature. The statistical features used here are different from the statistical features used in ASR based mispronunciation detection systems.

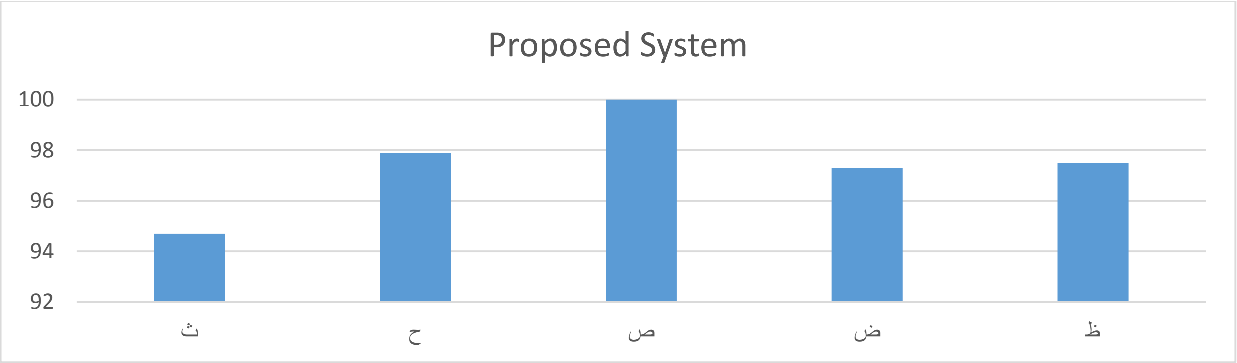

Fig. 1: Accuracies for all Arabic Phonemes

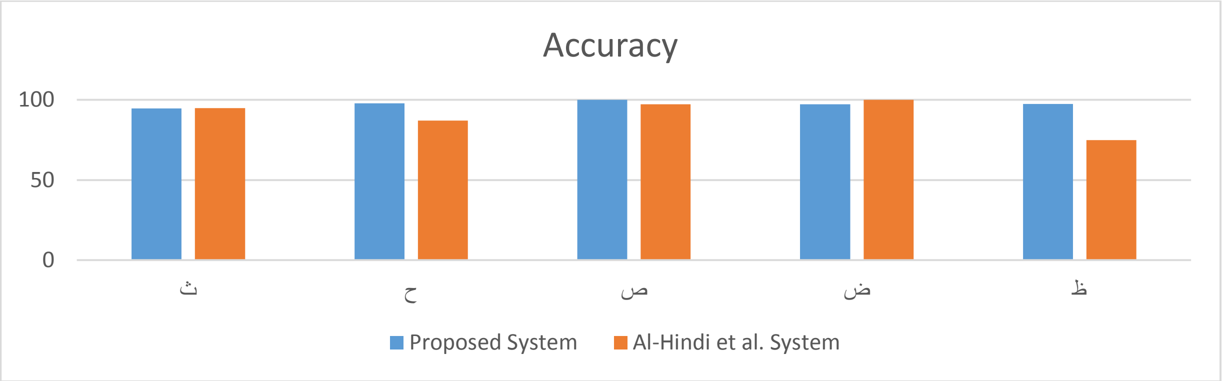

Fig. 2: A comparison with an existing System

When mispronunciation detection is considered as a classification problem, different classifiers can be used for mispronunciation detection. Artificial Neural Networks, Random Forest, decision trees etc. Support Vector Machines (SVM), is a very good binary classifier. SVM has been selected due to its generalization ability and its suitability to 2-class classification problem. In this research a separate classifier for each Arabic phone has been designed because each phone represents a different pronunciation mistakes. The extracted features for each phone are used to train each SVM classifier. Then this trained classifier can detect the pronunciation mistakes.

1. Results and Discussion:

This section presents the results for the proposed system. Here, mispronunciation detection problem is defined as a 2-class classification problem. All the correctly pronounced phonemes are categorized in class 1 and all the mispronunciations have been categorized in class 2.

Support vector machines (SVM) has been used for classification. To evaluate the results accuracy has been used. Accuracy of the system has been defined as:

$$Accuracy = \frac{Corectly\ Classified\ Phonemes}{Total\ No.of\ Phonemes}\ X\ 100 \qquad (1)$$

To evaluate the effectiveness of the proposed APF based classifier for mispronunciation detection system, as a test case 5 Arabic phonemes have been considered. These five Arabic phonemes are mostly mispronounced by the Pakistani and Indian Speakers while learning Arabic as their second language. A separate SVM classifier has been trained for each phoneme because this research is based on phone level mispronunciation detection.

Table 3: Comparison of our proposed technique with existing Arabic CAPT systems

Mispronunciation Detection Systems for Arabic						
Techniques	Proposed Acoustic Feature Selection based Technique	Metwalli *et al.* System	Strik *et al.* System	Alhindi *et al.* system	Witt. System	Cucchiarini *et al.* System
Avg. Accuracy	97.5%	52.2%	81-88%	92.95%	80-92%	86%

When these system was tested, it shows excellent result. The accuracy of the system is presented in Fig 1.The results shows that the proposed system gives the accuracy of 94.7%, 97.89%, 100%, 97.3% and 97.5% respectively for the phonemes. These excellent accuracies shows that when mispronunciation detection has been treated as a 2-class problem, it can be solved relatively easily. The proposed system beat the other existing systems developed for Arabic mispronunciation detection. The system developed by Al-Hindi et al. also detect mispronunciations for these 5 Arabic phonemes. A detailed comparison is presented in fig 2.

The proposed system continuously beat the system developed by Al-Hindi et al. [18]. A comparison between the weighted average accuracies is shown in Table 3. The average accuracy of the proposed system is better than the existing systems. This strengthen this argument that when mispronunciation detection problem is formulated using APF using a good classifier, it can outperform the traditional ASR based mispronunciation detection systems.

Conclusion:

In this paper, mispronunciation detection problem is defined as 2-Class classification problem rather than an ASR based system. The proposed system were trained using acoustic phonetic features. The feature set includes Pitch, Zero-Cross rate, MFCCs along with first and second derivative, RMSE, global statistical features and low energy. The proposed system produced excellent results with an average accuracy of 97.5% for five Arabic phonemes. There is still a need to develop a system which can automatically extract acoustic features based on the pronunciation error.

References

[1] Strik, H., Truong, K., de Wet, F., &Cucchiarini, C. , Comparing different approaches for automatic pronunciation error detection. Speech Communication, 2009; 51(10), 845-852.

[2] Wei, S., Hu, G., Hu, Y., & Wang, R. H., A new method for mispronunciation detection using support vector machine based on pronunciation space models. Speech Communication, 2009; 51(10), 896-905.

[3] Franco, H., Neumeyer, L., Digalakis, V., Ronen, O., Combination of machine scores for automatic grading of pronunciation quality. Speech Comm., 2000; 121–130.

[4] Truong, K., Automatic pronunciation error detection in Dutch as a second language: an acoustic–phonetic approach. Master Thesis, Utrecht University, the Netherlands; 2004.

[5] Strik, H., Truong, K., Wet, F.D., Cucchiarini, C., Comparing classifiers for pronunciation error detection. In: Proc. Eur. Conf. on Speech communication and Technology, 2007; pp. 1837–1840.

[6] Witt, S.M., Young, S.J., Phone-level pronunciation scoring and assessment for interactive language learning. Speech Comm., 2000; 95–108.

[7] Kanters, S., Cucchiarini, C., &Strik, H., The Goodness of Pronunciation algorithm: a detailed performance study. Proceedings of SLATE; 2009.

[8] Neumeyer, L., Franco, H., Digalakis, V., Weintraub, M., Automatic scoring of pronunciation quality. Speech Comm., 1999; 83–93.

[9] Weigelt, L.F., Sadoff, S.J., Miller, J.D., Plosive/fricative distinction: the voiceless case. J. Acoust. Soc. Amer. 87, 1990; 2729–2737.

[10] Neri, A., Cucchiarini, C., Strik, H., ASR corrective feedback on pronunciation: does it really work? In: Proc. Interspeech, 2006a; pp. 1982–1985.

[11] Neri, A., Cucchiarini, C., Strik, H., Selecting segmental errors inL2 Dutch for optimal pronunciation training.IRAL – Internat.Rev.Appl. Linguist. Lang. Teaching 44, 2006b; 357–404.

[12] Cucchiarini, C., Strik, H., Boves, L., Quantitative assessment of second language learners' fluency by means of automatic speech recognition technology. J. Acoust. Soc. Amer. 107, 2000; 989–999.

[13] John Eulenberg, AbariesFarhad, Fundamental Frequency and the Glottal Pulse. Available at:

[14] https://www.msu.edu/course/asc/232/study_guides/F0_and_Glottal_Pulse_Period.html Sidney Wood, What are formants. Available at: http://person2.sol.lu.se/SidneyWood/praate/whatform.html

[15] Shete, D. S., and S. B. Patil., Zero crossing rate and Energy of the Speech Signal of Devanagari Script.

[16] Al-Shoshan, Abdullah I., Speech and music classification and separation: a review. Journal of King Saud University; 2006.

[17] Molau, Sirko, et al., Computing mel-frequency cepstral coefficients on the power spectrum. Acoustics, Speech, and Signal Processing, IEEE International Conference on. Vol. 1. IEEE, 2001.

[18] Al Hindi, Afnan, et al. "Automatic pronunciation error detection of nonnative Arabic Speech." Computer Systems and Applications (AICCSA), 2014 IEEE/ACS 11th International Conference on. IEEE, 2014.

Author Biography:

Muazzam Maqsood is currently doing his Ph.D. in Software Engineering from University of Engineering and Technology, Taxila. He has completed his MS degree in 2013 from University of Engineering and Technology, Taxila. His Major areas of interests are Machine Learning, Speech Processing, Recommender System and Image Processing.

Hafiz Adnan Habib completed his MS (Electrical Engineering) in 2004 and PhD (Electrical Engineering) in 2007 from University of Engineering and Technology, Taxila, Pakistan. He is currently serving as Head of Department of Computer Science in UET Taxila Pakisan. His research interests include Speech Processing, Image and Video Processing, Software Development, Artificial Intelligence and Artificial Neural Networks.

Tabassam Nawaz received his MS Computer Engineering in 2005 from CASE (Center for Advanced Studies in Engineering), Islamabad, Pakistan and subsequently he completed his Ph.D in 2008. He is currently serving as a Head of Department of Software Engineering. His research interest include Image and video processing, Software development, Artificial Intelligence and web development.

Khurram Zeeshan Haider completed his MS (Software Engineering) in 2011 from University of Engineering and Technology, Taxila, Pakistan. His research areas include Digital Image Processing, Machine Vision, Contextual Image Classification, video based Behavior Classification, Machine Learning, Software Development and Software Quality Engineering. He is PhD scholar at University of engineering and Technology, Taxila.