Comparative Analysis of Arabic Vowels using Formants and an Automatic Speech Recognition System

Yousef Ajami Alotaibi[1], Amir Hussain[2]

[1]*Computer Eng. Dept., College of Computer & Information Sciences, King Saud University*
[2]*Department of Computing Science, Stirling University*
yaalotaibi@ksu.edu.sa, ahu@cs.stir.ac.uk

Abstract

Arabic, the world's second most spoken language in terms of number of speakers, has not received much attention from the traditional speech processing research community. This study is specifically concerned with the analysis of vowels in modern standard Arabic dialect. The first and second formant values in these vowels are investigated and the differences and similarities between the vowels explored using consonant-vowels-consonant (CVC) utterances. For this purpose, a Hidden Markov Model (HMM) based recognizer is built to classify the vowels and the performance of the recognizer analyzed to help understand the similarities and dissimilarities between the phonetic features of vowels. The vowels are also analyzed in both time and frequency domains, and the consistent findings of the analysis are expected to enable future Arabic speech processing tasks such as vowel and speech recognition and classification.

Keywords: *MSA, Arabic, Vowels, Analysis, Speech, Recognition, Formants, HMM, ASR*

1. Background

1.1. Arabic language and research

Arabic is a Semitic language, and is one of the world's oldest languages. Currently it is the second most spoken language in terms of number of speakers. Modern Standard Arabic (MSA) has 36 phonemes, of which six are vowels, two diphthongs, and 28 are consonants. In addition to the two diphthongs, the six vowels are /a, i, u, a: , i:, u:/ where the first three ones are short vowels and the last three are their corresponding longer versions (that is, the three short vowels are /a, i, u /, and their three long counterparts are /a:, i:, u:/) [1], [2], [3]. As a result, vowel sound duration is phonemic in Arabic language.

Some researchers consider Arabic vowels to number eight in total by counting the two diphthongs as vowels and, this is normally considered to be the case for MSA [4]. Arabic phonemes comprise two distinctive classes, termed pharyngeal and emphatic phonemes. These two classes can be found only in Semitic languages such as Hebrew and Persian [1], [4], [5]. Arabic dialects may have different vowels - for instance, Levantine dialect has at least two extra types of diphthongs /aj/ and /aw/. Similarly, Egyptian dialect has other extra vowels [3].

The development of accurate Automatic Speech Recognition (ASR) systems is faced with two major issues. The first problem is related to diacritization where diacritic symbols refer to vowel phonemes in the designated words. Arabic texts are almost never fully diacritized:

implying that the short strokes placed above or below the consonant, indicating the vowel following this consonant, are usually absent. This limits the availability of Arabic ASR training material. The lack of this information leads to many similar word forms, and consequently, decreases predictability in the language model. The three short Arabic vowels are represented using diacric symbols in the written form. The second problem is related to the morphological complexity since Arabic has a rich potential of word forms which increases the out-vocabulary rate [4], [6].

Arabic language is comparatively much less researched compared to other languages such as English and Japanese. Most of the reported studies to-date have been conducted on Arabic language and speech digital processing in general, with only a few focusing on Arabic vowels specifically. A limited number of research studies have been carried out on MSA, classical and Quraanic (Islamic Holy Scripture based) versions of Arabic. More recently, Iqbal et al. [7] reported a new preliminary study on vowels segmentation and identification using formant transitions occurring in continuous recitation of Quraanic Arabic. The paper provided an analysis of cues to identify Arabic vowels. Their algorithm extracted the formants of pre-segmented recitation audio files and recognized the vowels on the basis of these extracted formants. The study was applied in the context of recitation principles of the Holy Quraan. The vowel identification system developed showed up to 90% average accuracy on continuous speech files comprising around 1000 vowels.

In other related recent works, Razak et. al. [8] have investigated Quraanic verse recitation feature extraction using the Mel-Frequency Cepstral Coefficient (MFCC) approach. Their paper explored the viability of the MFCC technique to extract features from Quranic verse recitation. Features extraction is crucial to prepare data for the classification process. The authors were able to recognize and differentiate the Quranic Arabic utterance and pronunciation based on the extracted features vectors. Tolba et al. [9] have also reported a new method for Arabic consonant/vowel segmentation using the wavelet transform. In their paper, a new algorithm was presented for Arabic speech consonant and vowel segmentation without linguistic information. The method was based on the wavelet transform and spectral analysis and focused on searching the transient between the consonant and vowel parts in certain levels from the wavelet packet decomposition. The accuracy rate was about 88.3% for consonant/vowel segmentation and the rate remained fixed at both low and high signal to noise ratios (SNR). Previously, Newman et al. [10] worked on a frequency analysis of Arabic vowels in connected Speech. Their findings do not confirm the existence of a high classical style as an acoustically ‘purer’ variety of modern standard Arabic.

In another study, Alghamdi [11] carried out an interesting spectrographic analysis of Arabic vowels based on a cross-dialect study. He investigated whether Arabic vowels are the same at the phonetic level when spoken by speakers of different Arabic dialects, including Saudi, Sudanese, and Egyptian dialects. The author found that the phonetic implementation of the standard Arabic vowel system differs according to dialects.

Previously, Al-Otaibi [12] also developed an automatic Arabic vowel recognition system. Isolated Arabic vowels and isolated Arabic word recognition systems were implemented. The work investigated the syllabic nature of the Arabic language in terms of syllable types, syllable structures, and primary stress rules. Next, ASR based Hidden Markov models are briefly described and formant readings reviewed. The goals of this study are then outlined.

1.2. Hidden Markov models

ASR systems based on the Hidden Markov Model (HMM) started to gain popularity in the mid 1980's [13]. HMM is a well-known and widely used statistical method for characterizing the spectral features of speech frame. The underlying assumption of the HMM is that the speech signal can be well characterized as a parametric random process, and the parameters of the stochastic process can be predicted in a precise, well-defined manner. The HMM method provides a natural and highly reliable way of recognizing speech for a wide range of applications [14], [15].

In the main recognition module the feature vectors are matched with reference patterns, which are called acoustic models. The reference patterns are usually Hidden HMM models trained for whole words or, more often, for phones as linguistic units. HMMs cope with temporal variation, which is important since the duration of individual phones may differ between the reference speech signal and the speech signal to be recognized. Unfortunately this is not practical in short and long Arabic vowels, where the duration is very important and dictates the word meaning. A linear normalization of the time axis is not sufficient here, since not all allophones are expanded or compressed over time in the same way. For instance, stop consonants ("d", "t", "g", "k", "b", and "p") do not change their length much, whereas the length of vowels strongly depends on the overall speaking rate [15].

The recently developed Hidden Markov Model Toolkit (HTK) [16] is a portable toolkit for building and manipulating HMM models. It is mainly used for designing, testing, and implementing ASR and its related research tasks. The HTK [16] is a general toolkit for the HMM model that is mainly geared towards speech recognition, but can also be used for other tasks. HTK includes a large number of tools for training and manipulating HMMs, working with pronunciation dictionaries, n-gram and finite-state language models, recording and transcribing speech, etc.

1.3. Formant readings

By changing the vocal tract shape, different forms of a perfect tube are produced, which in turn, can be used to change the desired frequencies of vibration. Each of the preferred resonating frequencies of the vocal tract (corresponding to the relevant bump in the frequency response curve) is known as a formant. These are usually referred to as F1 indicating the first formant, F2 indicating the second formant, F3 indicating the third formant, etc. That is, by moving around the tongue body and the lips, the position of the formants can be changed [2].

In vowels, F1 can vary from 300 Hz to 1000 Hz. The lower it is, the closer the tongue is to the roof of the mouth. F2 can vary from 850 Hz to 2500 Hz. The value of F2 is proportional to the frontness or backness of the highest part of the tongue during the production of the vowel. In addition, lips' rounding causes a lower F2 than with unrounded lips. F3 is also important in determining the phonemic quality of a given speech sound, and the higher formants such as F4 and F5 are thought to be significant in determining voice quality.

1.4. Goals

The goals of this study are to investigate MSA vowels using both speech recognition and time and frequency domain information especially formants. Firstly, a recognition system will be built to classify the vowels and determine the similarities and dissimilarities among the 8 different MSA vowels under investigation. Secondly, we carry out a formant based analysis of the six Arabic vowels as used in MSA. The outcomes of these two investigative methodologies will be crosschecked to conclude the final contributions of this research regarding the MSA vowels.

The rest of this paper is organized as follows: Section II introduces the experimental framework for both methods employed in this study. The results are described and discussed in Section III. Paper conclusions and some remarks and suggestions are given in Section IV.

2. Experimental Framework

The allowed syllables in Arabic language are: consonant-vowel (CV), consonant-vowel-consonant (CVC), and consonant-vowel-consonant-consonant (CVCC), where V indicates a (long or short) vowel while C indicates a consonant. Arabic utterances can only start with a consonant [1]. Table 1 shows the eight Arabic vowels along with their names, examples, and IPA symbols. In this paper the formants of Arabic vowels will be analyzed to determine their values. These are expected to prove helpful in subsequent speech processing tasks such as vowel and speech recognition and classification. In carrying out the analysis, Arabic vowels have been viewed as if they are patterns on papers. Specifically, the vowels were plotted on paper or computer screen in the form of their time waveform, spectrograms, formants, and LPC spectrums.

2.1. Recognition system overview

An ASR based on HMM was developed. The system was partitioned into three modules according to their functionality. First is the training module, whose function is to create the knowledge about the speech and language to be used in the system. Second is the HMM models bank, whose function is to store and organize the system knowledge gained by the first module. Finally, there is the recognition module whose function is to figure out the meaning of the speech input in the testing phase. This was done with the aid of the HMM models mentioned above.

The parameters of the system were 10 KHz sampling rate with 16 bit sample resolution, 25 millisecond Hamming window duration with step size of 10 milliseconds, MFCC coefficients with 22 set as the length of cepstral filtering and 26 filter bank channels, 12 as the number of MFCC coefficients, and 0.95 as the pre-emphasis coefficients.

Table 1. MSA Arabic vowels

Vowel	The name of Arabic vowels	Example	IPA code
V01	Short fatha	جَد	a
V02	Long fatha	جاد	a:
V03	Short dummah	جُد	u
V04	Long dummah	جود	u:
V05	Shot kasrah	جِد	i
V06	Long kasrah	جِيد	i:
V07	Fatha dummah	جَوَد	aw
V08	Fatha kasrah	جَيد	ay

A second speech recognition system was also implemented using the HMM technique with the help of HTK tools. The speech ASR was designed initially as a phoneme level recognizer with 3-states: continuous, left-to-right, no skip HMM models. The system was designed by considering all 21 MSA monophones by considering a subset of the thirty-four MSA monophones as given by the KACST labeling scheme given in [17]. The silenc`e (sil) model was also included in the model set. In a later step, the short pause (sp) was created from and tied to the silence model. Since most digits consisted of more than two phonemes, context-dependent triphone models were created from the monophone models mentioned above. Prior to this, the monophones models were initialized and trained using the training data. This was done using more than one iteration and repeated again for triphones models. The pre-final training phase step was to align and tie the model by using a decision tree method. The last step in the training phase involved re-estimating the HMM parameters using the Baum-Welch algorithm [14] three times.

2.2. Database

An in-house database was built to help in investigating Arabic vowels depending on good selected and fixed phonemes. The utterances of ten male Arabic speakers, all aged between 23 to 25 years with the exception of one child, were recorded. Nine of the speakers were from different regions in Saudi Arabia and the remaining one from Egypt. Each of the ten speakers participated in five different trials for every carrier word in the data set used along with all the eight intended Arabic phonemes. Some of the speakers recorded the words in one session and others in two or three sessions.

The carrier words were chosen to represent different consonants before and after the intended vowel. These carrier words are displayed in Table 2 using the second vowel /a:/. The sampling rate used in recording these words was 16 kHz and 16-bit resolution mono. Total of the recorded audio tokens was 4000 (i.e., eight phonemes times ten speakers times ten carrier words times five trials for each speaker). These audio tokens were used for analyzing the intended phonemes in frequency, during the training phase of the recognition system, and in its testing phase.

The database was partitioned into two subsets one for training phase (60% of the total database) and the other for testing phase (40% of the whole database). For the first subset, for each vowel, speaker, and carrier word the first three trials were considered as training data for the recognition system. This gave a total of 2,400 (i.e., eight phonemes times ten speakers times ten carrier words times three trials) speech tokens. On the other hand, the testing subset consisted of the last two trials for all speakers, carrier words, and vowels. It contained a total of 1,600 (i.e., eight phonemes times ten speakers times ten carrier words times last two trials) speech tokens.

Table 2. Carrier words used (with 2nd vowel)

Word (CVC)	In Arabic	Start Consonant		End Consonant	
		IPA	Info.	IPA	Info.
CW01	سار	/s/	unvoiced fricative	/r/	voiced lateral
CW02	صار	/sˤ/	unvoiced fricative emphatic	/r/	voiced lateral
CW03	بات	/b/	voiced stop	/t/	unvoiced stop
CW04	عاد	/ʕ/	voiced fricative glottal	/d/	voiced stop
CW05	طار	/tˤ/	unvoiced stop emphatic	/r/	voiced lateral
CW06	زاد	/z/	voiced fricative	/d/	voiced nasal
CW07	صام	/sˤ/	unvoiced fricative emphatic	/m/	voiced nasal
CW08	فاز	/f/	unvoiced fricative	/z/	voiced fricative
CW09	نام	/n/	voiced nasal	/m/	voiced nasal
CW10	جاع	/3/	voiced affricative	/?/	voiced fricatives glottal

3. Results

3.1. Speech recognition analysis

An HMM based speech recognition system was designed, tested, and used for recognizing Arabic vowels using the given database. The overall system performance (recognition success rate (%) with respect to Arabic vowels was 91.6% as depicted in Table 3. This table shows the confusion matrix that was generated by the recognition system. The last three columns in the table show, respectively, individual system accuracies for each vowel separately, total number of speech tokens for each individual vowel, and missed tokens per vowel. In addition to this, the table depicted the number of inserted and deleted tokens for each vowel – all of which were with zero deletion and insertion in this experiment.

It can be seen from Table 3, that the totals of missed tokens were 135 out of 1599. The best vowel accuracy was encountered with Vowel 8 which is the diphthong “Fatha Kasrah” vowel. The accuracy for this vowel is 99% and only two tokens of Vowel 8 were missed by the system. On the other hand, the worst accuracy was encountered for the case of Vowel 1 which is the “Short Fatha”. The system accuracy for this vowel was 77.5% with a total of 45 missed tokens. Five of the vowels achieved over 90% accuracy but the remaining three did not.

Analyzing the confusion matrix in Table 3, we can notice that Vowel 1, Vowel 2, and Vowel 6 are the most confusing vowels for the recognition system (i.e. they are mistakenly picked up by the system in case of missing the correct one). Also, we can conclude from the confusion matrix that there are significant similarities between the following vowel pairs (listed in order of degree of highest to lowest similarity): (Vowel 1, Vowel 2), (Vowel 5, Vowel 6), (Vowel 3, Vowel 4), (Vowel 1, Vowel 3), and (Vowel 1, Vowel 5). The first, second and third pairs are similar but they differ in the length. For example Vowel 1 is the Short Fatha but Vowel 2 is the "Long Fatha". It is important to emphasize here that the duration of vowels in Arabic language is phonemic which is otherwise, as in English language. For the first, second and third pairs' similarities we can conclude that the system could not recognize the difference in duration and hence could not distinguish between Vowels 1 and 2 and between Vowels 5 and 6 and Vowel 3 and 4. This may be an important point to try to re-model the duration in HMM states and force the system to distinguish between short and long vowel in Arabic vowels since it is phonemic. For the fourth and fifth pairs we found that Vowel 2 (a:) was confused with Vowel 3 (u). This is considered a strange outcome due to the big difference between these vowels. The same thing can be said for the last pair. Also we can conclude from the recognition system outcomes that the vowel "Short Fatha" is the trouble source for the system where it made 3 major confusions to other vowels. Next, we will crosscheck the recognition system performance with the analysis carried out using the formant based approach.

Table 3. Confusion matrix and other statistics from the vowel recognition system

Vowel (IPA)	VWL1	VWL2	VWL3	VWL4	VWL5	VWL6	VWL7	VWL8	Del	Acc. (%)	Tokens	Missed
VWL1 (a)	155	33	5	0	6	0	1	0	0	77.5	200	45
VWL2 (a:)	4	195	0	0	0	0	1	0	0	97.5	200	5
VWL3 (u)	11	3	173	7	3	0	3	0	0	86.5	200	27
VWL4 (u:)	0	0	6	190	0	0	4	0	0	95	200	10
VWL5 (i)	8	5	5	0	166	15	0	0	0	83.4	199	33
VWL6 (i:)	0	0	0	1	0	194	0	5	0	97	200	6
VWL7 (ay)	1	0	1	5	0	0	193	0	0	96.5	200	7
VWL8 (ay)	0	0	0	0	0	2	0	198	0	99	200	2
Ins	0	0	0	0	0	0	0	0				
Total									0	91.6	1599	135

3.2. Formant based analysis

The aim of this part of the experiments was to evaluate values of the first and second formants, namely F1 and F2, in all considered Arabic vowels. This study considered frames in the middle of each vowel to minimize the co-articulation effects. Based on Figure 2, we can estimate the Arabic vowel triangle's location as (400, 800), (700, 1100), and (400, 2100) where the first value corresponds to F1 and the second value to F2.

Figure 2 shows a plot for all short vowels for one of the speakers for three different trials. It can be seen from Figure 2 that the F1 value is relatively high for /a/, medium for /u/, and minimum for /i/. But in the case of F2, it is medium for /a/, minimum for /u/ and high for /i/. For the long vowels it is clear that the same situation is true as observed for their short counterparts. The long vowels are peripheral while their short counterparts are close to center when the frequencies of the first two formants are plotted on the formant chart.

From the obtained results, it was found that F1 of /a/ is smaller than F1 of /a:/, while F2 of /a/ is smaller than F2 of /a:/ (and the values of F3 are close for both of the durational counterparts). Also F1 of /u/ is larger than F1 of /u:/ except for the child, whereas F2 of /u/ is smaller than F2 of /u:/ for all speakers and the values of F3 are close for both of them. Similarly, it has been found that F1 of /i/ is larger than F1 of /i:/ and F2 of /i/ is smaller than F2 of /i:/. To conclude these specific outcomes, it can be said that F1 in a given short vowel is larger than F1 in its long counterpart except for /a/ and /a:/; and F2 in a given short vowel is larger than F1 in its long counterpart except for /a/ and /a:/. These findings confirm those reported in previous studies [11].

The first formant, F1, can be used to classify between /a/ and /u/. F2 can be used to classify /i/ and /u/. The vowel /a/ has the largest value of F1 and /i/ has the largest value of F2. The vowel /i/ has the smallest value of F1 and /u/ has the smallest value of F2. In addition, F1 can be used to classify /a:/ and /u:/ whereas F2 can be used to classify /i:/ and /u:/. The vowel /a:/ has the largest value of F1 and /i:/ has the largest value of F2, while the vowel /i:/ has the smallest value of F1 and /u:/ has the smallest value of F2.

In Arabic vowels, as mentioned earlier, F1 can vary from 300 Hz to 1000 Hz, and F2 can vary from 850 Hz to 2500 Hz. F3 is also important in determining the phonemic quality of a given speech sound, and the higher formants such as F4 and F5 can also be significant in determining voice quality [18]. In /a/ and /a:/ the whole tongue goes down so the vocal tract becomes wider than in producing other Arabic vowels. In /u/ and /u:/ the end of the tongue comes near to the palate while the other parts of the tongue are in the regular position. In /i/ and /i:/ the front of the tongue comes near to the palate whereas other parts remain in their regular position. Lips are more rounded for /u/ and /u:/ than for /i/ and /i:/. Also, the formants' usual patterns were noticed from their spectrograms as concluded in [18]. In addition, the similarities of the first (and final) consonant in all words can be clearly noticed since the same first and final consonants are used in all the plots (with just the vowel being varied) as can be inferred from Table 2.

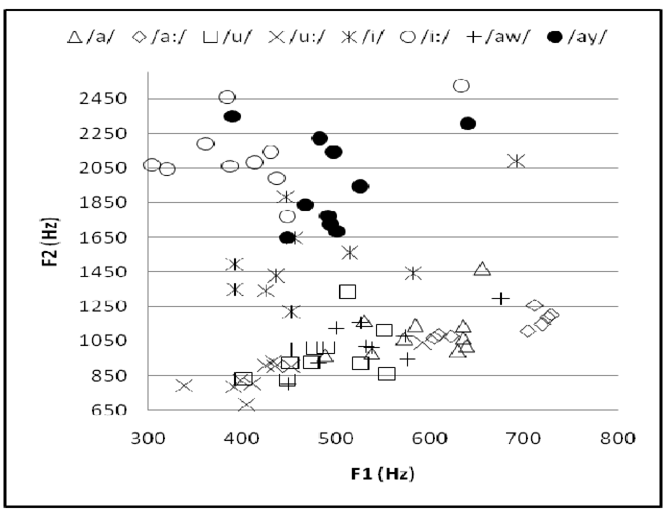

Figure 1. Vowels distribution depending on F1 and F2 for all speakers

3.3. Discussion

In summary, the vowel pairs (Vowel 1, Vowel 2), (Vowel 5, Vowel 6), (Vowel 3, Vowel 4), (Vowel 1, Vowel 3), and (Vowel 1, Vowel 5) that have major similarities depending on the speech recognition output, are mostly durational counterparts. This means that Vowel 1 is the short counterpart of Vowel 2 and so on. These pairs of vowels have major overlap in formant plot shown in Figure 1. Formants can thus be seen to be very effective in classifying vowels correctly and can be used in a future speech recognition system. Formants of the vowels can be included explicitly in the feature extraction module of the recognizer. If such a system is able to recognize the different vowels then this will tremendously assist in the Arabic speech recognition process. The reason behind this is that every word and syllable in Arabic language must contain at least one vowel; hence vowel recognition will play a key role in identifying the spoken utterance. Agreement between the recognition outputs and formant analysis can be inferred from these results. In addition to this, we can say that the distinguishing acoustic features of vowels are formants, in particular formants 1 and 2. Also, for Arabic vowels we can learn that the temporal duration is very important but this is missed by both the investigative methods used in this study and we need to find a way to incorporate this feature in future speech recognition and formant-based analysis approaches.

4. Conclusions

This paper has presented automatic speech recognition and formant based analysis of Arabic vowels using a spectrogram technique [18]. For the formant based technique, the Arabic vowels were studied as if they were patterns shown on screen or paper. Modern standard Arabic has six basic vowels and two diphthongs which are considered by some linguists as vowels rather than diphthongs. Thus the number of vowels in this study was considered to be eight (including the two diphthongs). All these eight Arabic phonemes were included for constricting the created database deployed in the speech recognition based investigation which has shown that the formants are very effective in classifying vowels correctly. The overall accuracy of the speech recognition system was 91.6%. Both methods of investigations agreed on the high degree of similarity between short Arabic vowels and their long counterparts. The temporal duration of the vowel which is phonemic in Arabic language could not be modeled in the recognition system. Also in the formant plots, there are big overlaps between pairs of vowels that also confused with each other in the recognition system. We see an agreement between the conclusions drawn from speech recognition results and formant analysis regarding these vowels.

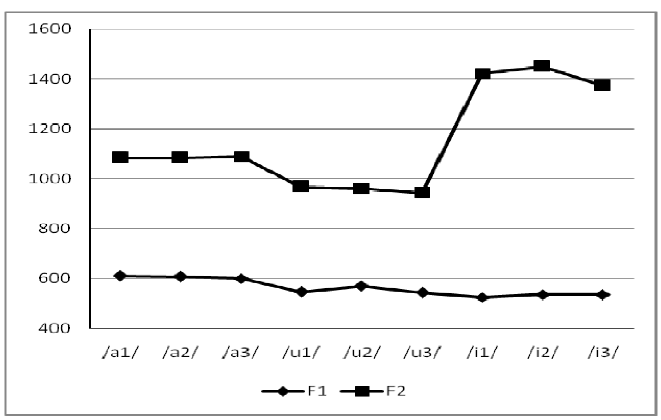

Figure 2. Values of F1 and F2 for short vowels for Speaker 6 for three trials

5. Acknowledgments

The authors would like to acknowledge the British Council (in Riyadh, Saudi Arabia) for funding this collaborative research between King Saud University and the University of Stirling (in Scotland).

6. References

[1] Muhammad Alkhouli, “Alaswaat Alaghawaiyah,” Daar Alfalah, Jordan 1990 (in Arabic)

[2] J. Deller, J.Proakis, and J. H.Hansen, “Discrete-Time Processing of Speech Signal” Macmillan, 1993.

[3] Mansour Alghamdi “Arabic Phonetics”, Al-Toubah Bookshop, Riyadh 2001. (In Arabic).

[4] Ahmed Omar, “Study of Linguistic phonetics,” Aalam Alkutob, Eygpt 1991. (in Arabic)

[5] M. Elshafei, “Toward an Arabic Text-to -Speech System,” The Arabian Journal for Scince and Engineering, Vol.16, No. 4B, pp.565-83, Oct. 1991.

[6] Y. A. El-Imam, “An Unrestricted Vocabulary Arabic Speech Synthesis System”, IEEE Transactions on Acoustic, Speech, and Signal Processing, Vol. 37,No. 12, Dec. 1989, pp.1829-45.

[7] Hafiz Rizwan Iqbal, Mian M. Awais, Shahid Masud, and Shafay Shamail, “On Vowels Segmentation and Identification Using Formant Transitions in Continuous Recitation of Quranic Arabic”, Chapter of “New Challenges in Applied Intelligence Technologies”, Springer Berlin, 2008: pp. 155-162.

[8] Zaidi Razak, Noor Jamaliah Ibrahim, Emran Mohd Tamil, Mohd Yamani Idna Idris, Mohd. Yakub and Zulkifli Bin Mohd Yusoff, "Quranic Verse Recitation Feature Extraction Using Mel-Frequency Cepstral Coefficient (MFCC)", In Proceedings of the 4th IEEE International Colloquium on Signal Processing and its Application (CSPA) 2008, 7-9 March 2008, Kuala Lumpur, MALAYSIA.

[9] M. F. Tolba, T. Nazmy, A. A. Abdelhamid, and M. E. GadallahA, “A Novel Method for Arabic Consonant/Vowel Segmentation using Wavelet Transform”, International Journal on Intelligent Cooperative Information Systems, IJICIS Vol 5, No. 1, pp 353- 364, July 2005.

[10] Daniel L. Newman, and Jo Verhoeven, “Frequency Analysis of Arabic Vowels in Connected Speech”, pp. 77-87.

[11] M. M. Alghamdi, “A spectrographic analysis of Arabic vowels: A cross-dialect study” Journal of King Saud University, Vol.10, Arts(1), pp. 3-24, 1998.

[12] A. Al-Otaibi, “Speech Processing,” The British Library in Association with UMI, 1988.

[13] P. C. Loizou, and A. S. Spanias, “High-Performance Alphabet Recognition”, IEEE Trans. on Speech and Audio Processing, Vol. 4, No. 6, Nov. 1996, pp.430-445.

[14] L. R. Rabiner, “A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition”, Proceedings of the IEEE, Vol. 77, No. 2, pp.257-286, Feb 1989.

[15] B. Juang, and L. Rabiner, “Hidden Markov Models for Speech Recognition”, Technometrics, Vol. 33, No. 3, August 1991, pp.251-272.

[16] S. Young, G. Evermann, M. Gales, T. Hain, D. Kershaw, G. Moore J. Odell, D. Ollason, D. Povey, V. Valtchev, and P. Woodland, The HTK Book (for HTK Version. 3.4), Cambridge University Engineering [x] Department, 2006. http://htk.eng.cam.ac.uk/prot-doc/ktkbook.pdf.

[17] M. M. Alghamdi, Y. El Hadj, and M. Alkanhal, “A Manual System to Segment and Transcribe Arabic Speech”, IEEE International Conference on Signal Processing and Communication (ICSPC07). Dubai, UAE: 24–27 November 2007.

[18] Yousef Ajami Alotaibi and Amir Hussain, “Formant Based Analysis of Spoken Arabic Vowels”, BioID_MultiComm ’09: Joint COST 2101 & 2102 International Conference on Biometric ID Management and Multimodal Communication September 16-18, 2009, Madrid, Spain (Accepted)

Authors

Dr. Yousef Ajami Alotaibi received his B.Sc. degree in computer engineering from King Saud University, Riyadh, Saudi Arabia in 1988. He received his M.Sc. and Ph.D. degrees in computer engineering from the Florida Institute of Technology, Florida, USA, in 1994 and 1997, respectively. From 1988-1992 and 1998-1999 he worked as a research engineer at Al-ELM Research and Development Corporation, The National Information Center (NIC) Riyadh, Saudi Arabia, where he designed and implemented encryption and communication systems. In 1999 he joined the Computer Engineering Department, College of Computer and Information Sciences, of King Saud University in Riyadh, Saudi Arabia, as an Assistant Professor (1999-2008) and is currently an Associate Professor (since 2008). His research interests are centred round developing novel digital speech processing systems with a particular focus on speech recognition and Arabic language and speech processing. He has published more than 30 conference and journal papers to-date. He is also working on Arabic speech recognition projects with King Abdulaziz City for Science and Technology (KACST), Saudi Arabia.

Dr. Amir Hussain obtained his BEng (with the highest 1st Class Honours) and PhD, both in Electronic and Electrical Engineering from the University of Strathclyde in Glasgow, Scotland, in 1992 and 1997 respectively. He is currently a Reader in Computing Science at the University of Stirling in Scotland. His research interests are mainly inter-disciplinary and include speech signal processing and multi-modal audio-visual processing for next generation communication systems. He holds one international patent (in neural computing), is Editor of five Books and has published over 130 papers in journals and refereed international Conference proceedings. He is Editor-in-Chief of Springer's Cognitive Computation journal, Associate Editor for the IEEE Transactions on Neural Networks and serves on the Editorial Board of a number of other journals. He is a Senior Member of the IEEE, Chapter Chair of the UK & RI IEEE Industry Applications Society and a Fellow of the UK Higher Education Academy.