Study on Unique Pharyngeal and Uvular Consonants in Foreign Accented Arabic

Yousef A. Alotaibi, Khondaker Abdullah-Al-Mamun, and Ghulam Muhammad

Department of Computer Engineering, College of Computer and Information Sciences,
King Saud University, Riyadh, Saudi Arabia
{yalotaibi, mamun, ghulam}@ccis.ksu.edu.sa

Abstract

This paper investigates the unique pharyngeal and uvular consonants of Arabic from the automatic speech recognition (ASR) point of view. Comparisons of the recognition error rates for these phonemes are analyzed in five experiments that involve different combinations of native and non-native Arabic speakers. The most three confusing consonants for every investigated consonant are uncovered and discussed. Results confirm that these Arabic distinct consonants are a major source of difficulty for ASR. While the recognition rate for certain of these unique consonants such as /H/ can drop below 35% when uttered by non-native speakers, there are advantages to including non-native speakers in ASR. Regional differences in the pronunciation of Modern Standard Arabic by native Arabic speakers require attention of Arabic ASR research.

Index Terms: Arabic, pharyngeal, uvular, speech recognition, foreign accents

1. Introduction

Arabic is a Semitic language which has many differences when compared with other languages that include unique phonemes, phonetic features and complicated morphological word structure. Automatic Speech Recognition (ASR) system for Modern Standard Arabic (MSA) has major difficulties with distinctive characteristics of the Arabic sound system, namely, geminate, emphatic, uvular, and pharyngeal consonants, and vowel duration [1, 2]. Few researches have been done on Arabic ASR compared to other languages. Most efforts have concentrated on developing recognizers for MSA, which is the formal linguistic standard used throughout Arabic-speaking countries in the media, lectures, courtrooms [3]. This paper concentrates on the analysis and investigation of five Arabic unique pharyngeal and uvular sounds from an ASR perspective with effect of foreign-accented pronunciation.

Arabic is one of the oldest languages in the world. The estimated number of Arabic speakers is 250 million, of whom roughly 195 million are first-language speakers and 55 million are second-language speakers. Since it is also the language of religious instruction in Islam, many more speakers have at least a passive knowledge of the language. Arabic is an official language in more than 22 countries [4].

Compared to MSA, Classical Arabic is an older, literary form of language, exemplified by the type of Arabic used in the holy Quran. Spoken Arabic is a collection of regional and national varieties that are derived from Classical Arabic. Arabic dialects are primarily oral languages; written material is almost invariably in MSA. As a result, there is a serious lack of Language Model (LM) training material for dialectal speech. MSA is a version of Classical Arabic with a modernized vocabulary [5], and it is a formal standard common to all Arabic-speaking countries. It is the language used in the media (television, radio, press, etc.), in official speeches, in universities and schools, and, generally speaking, in any kind of formal communication situation [4].

Arabic texts are almost never fully diacritized and are thus potentially unsuitable for automatic speech recognition and synthesis [3]. Table 1 shows all Arabic alphabet letters and their correspondences to consonant and semivowel phonemes. This table also shows the phonetic description of each phoneme including the place of articulation. In this paper we use Language Data Consortium (LDC) WestPoint Modern MSA database phoneme set rather than International Phonetic Alphabet (IPA). This was decided because we are using WestPoint corpus in our research and they are using their custom Arabic phoneme list. Also IPA symbolization is hard to use in computer programs and files because symbols contain some special characters and special diacritizations.

Arabic has phonemes consisting of three short vowels (/i/, /a/, /u/), three long vowels (/i:/, /a:/, /u:/ which are the counterparts of the short vowels), and twenty-eight consonants [6]. Arabic has noticeably fewer vowels than English. In addition, vowel lengthening in Arabic is phonemic. Some Arabic dialects may have additional or fewer consonant phonemes. Arabic phonemes contain two distinctive classes that are named pharyngeal and emphatic phonemes.

There are two pharyngeal phonemes in Arabic that are part of interested phoneme in this research: fricatives, Ain /C/ and H\aa /H/. In addition to this, there are three uvular phones that are also part of this research. These phonemes are one stop Qaaf /q/, two fricatives Ghain /G/ and Khaa /x/. Table 2 shows the two pharyngeal and three Arabic uvular sounds. This table displays the place of articulation of these five Arabic phonemes. More description about pharyngeal and uvular phonemes will be represented in the results section.

Newman [7] investigated the phonetic status of Arabic within the world's languages, with the concentration on the uniqueness of special Arabic phonemes. In this study he considered the framework of IPA as used with the UCLA Phonological Segment Inventory Database – commonly known as UPSID. In that database there are about of 317 languages with total of 58 phonetic features. By concentrating on only pharyngeal and uvular Arabic phonemes that are of interest in our paper, he concluded that the voiced pharyngeal fricative /C/ is limited only to eight languages (2.5%) – five of them are Afro-Asiatic, where the lengthened version is unique to Arabic. The voiced uvular fricative /G/ is reported in only fourteen languages (4.38%) while the long version of this phoneme is unique for Arabic. In addition to this, Newman found that the pharyngeal unvoiced phoneme, /H/, is accruing in only thirteen languages (4.1%), and its longer variant was found in Arabic and in another language. On the other hand, he attributed the uvular plosive /q/ as the least stable sound inasmuch as in many local varieties of Arabic. To make it clear, /q/ phoneme is realized either as a voiced velar plosive (e.g., in Egypt, Libya, and Tunisia), or as

Table 1. *Arabic consonants* [8].

			Bilabial	Labio-dental	Inter-dental	Alveo-dental	Alveolar	Palatal	Velar	Uvular	Pharyngeal	Glottal
Stop	Voiced	Emphatic				ض /D/						
		Non- Emphatic	ب /b/			د /d/		ج /j/				
	Unvoiced	Emphatic				ط /T/						
		Non- Emphatic				ت /t/			ك /k/	ق /q/		ء /Q/
Fricative	Voiced	Emphatic			ظ /Z/							
		Non- Emphatic			ذ /TH/	ز /z/				غ /G/	ع /C/	
	Unvoiced	Emphatic				ص /S/						
		Non- Emphatic		ف /f/	ث /th/	س /s/		ش /sh/		خ /x/	ح /H/	هـ /h/
Nasal	Voiced	Non- Emphatic	م /m/				ن /n/					
Liquid	Voiced	Non- Emphatic					ل ر /l/r/					
		Emphatic					ل /l/					
Semivowels	Voiced	Non- Emphatic	و /w/					ي /y/				

Table 2. *Investigated Arabic consonants.*

Arabic Alphabet Carrier		LDC Symbol	Place of Articulation Considered
Ain	ع	C	**Pharynx**
Ghain	غ	G	**Uvular**
H\aa	ح	H	**Pharynx**
Qaaf	ق	q	**Uvular**
Khaa	خ	x	**Uvular**

Table 3. *LDC WestPoint corpus summary*.

Number of Speakers			
	male	**female**	**total**
native	41	34	75
non-native	25	10	35
totals	66	44	110
Hours of Data			
	male	**female**	**total**
native	6	4.4	10.4
non-native	0.74	0.28	1.02
totals	6.74	4.68	11.42

glottal stop as the case in the Syro-Lebanese and Cairene. This phoneme was reported in only thirty-eight languages (11.9%) in UPSID. Finally in that study, the unvoiced uvular fricative /x/ occurs in only twenty-seven languages (8.5%) within the database (interestingly enough from nearly every continent), whereas the lengthened variety is limited to only Arabic and two more other languages.

2. Experimental framework

The system is designed to recognize Arabic phonemes. We analyze the performance of the system with respect to the pharyngeal consonants - /C/ and /H/ - and the uvular consonants /G/, /q/, and /x/. The investigation focuses on the effect of native and non-native speakers in both training and testing data.

Hidden Markov Models (HMMs) are a well-known and widely-used statistical method for characterizing the spectral features of speech frame. The assumption underlying HMMs is that the speech signal can be well characterized as a parametric random access, and the parameters of the stochastic process can be predicted in a precise, well-defined manner. The Hidden Markov Model Toolkit (HTK) [9] is a portable toolkit for building and manipulating HMMs; it is widely used for designing, testing, and implementing ASR systems and related research tasks. HTK is used in all the experiments reported here.

2.1. Database

We use WestPoint Arabic Speech Corpus provided by LDC [10]. A descriptive summary of this database is given in Table 3. The amount of data provided by the native speakers is significantly greater than non-native speakers and all non-native Arabic speakers are native English speakers. The corpus includes both male and female speakers.

2.2. System description and parameters

The parameters of the system are 22 kHz sampling rate with 16 bit sample resolution, 25 milliseconds Hamming window duration with a step size of 10 milliseconds, MFCC coefficients with 22 as the length of cepstral leftering and 26 filter bank channels, 12 as the number of MFCC coefficients, 12 first derivatives and 0.95 as the pre-emphasis coefficients.

Our baseline system is designed as a phoneme-level recognizer with 3-state, continuous, left-to-right, no skip HMM models. It considers all 37 MSA monophones as given in the LDC catalog [10]. We note that the WestPoint Corpus contains more monophones than the number of MSA phonemes mentioned in the linguistic literature [11, 12, 13]. The phoneme /g/ was used because some native and non-native speakers produced it in certain MSA words. Two extra diphthongs were added because of variations in the pronunciations of non-native speakers. We use the WestPoint Corpus phonemes, transcriptions and other settings without any modification.

Since most of the words consisted of more than two phonemes, context-dependent triphone models were created from the monophone models. In the training phase, the model was aligned and tied by using the decision tree method. The last step in the training phase is to re-estimate the HMM parameters using the Baum-Welch algorithm three times.

Five types of experiments were carried out in the experiments. These experiments differ only in the type of the training and testing data sets. These experiments are labeled as N/N, N/NN, NN/N, NN/NN, and M/M, where N, NN and M indicate native, non-native and mixture of native and non-native Arabic speakers respectively. For example, experiment N/N indicates that native Arabic speakers are used in both training and testing phases, while N/NN indicates native Arabic speakers are used in training data and non-native Arabic speakers are used in testing data. The other labels follow the similar interpretation.

3. Results and discussions

3.1. Pharyngeal consonants

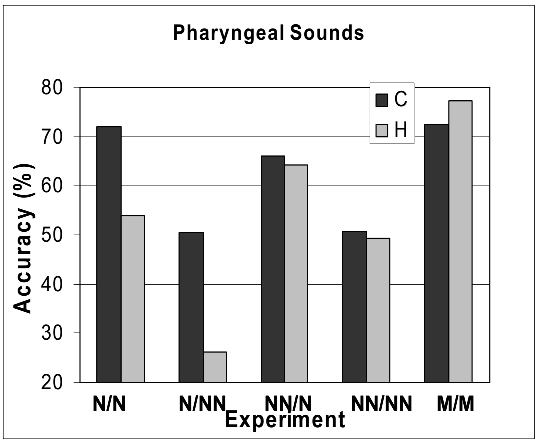

Figure 1: *Accuracy of Pharyngeal consonants.*

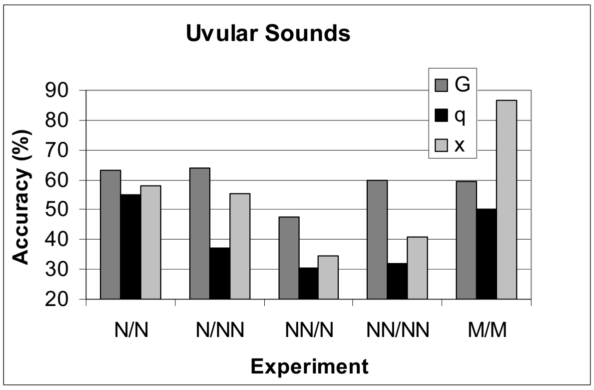

Figure 2: *Accuracy of Uvular consonants.*

Table 4. *Most three confusing consonants with the Pharyngeal and Uvular consonants.*

Arabic Alphabet Carrier		LDC Symbol	Experiment				
			N/N	N/NN	NN/N	NN/NN	M/M
Ain	ع	C	/Q, r, f/	/f, r, D/	/r, Q, TH/	/r, G, d/	/r, TH, G/
Ghain	غ	G	/Q, r, Z/	/C, r, f/	/r, Q, C/	/r, Z, sh/	/r, x, T/
H\aa	ح	H	/Q, G, h/	/f, r, D/	/h, Q, r/	/h, G, r/	/h, T, r/
Qaaf	ق	q	/Q, G, f/	/f, r, Q/	/k, Q, T/	/Q, k, h/	/T, Q, TH/
Khaa	خ	x	/Q, r, G/	/h, Q, r/	/h, Q, H/	/h, k, d/	/Z, T, r/

Figure 1 plots the accuracies of the two Arabic pharyngeal consonants for all the five experiments. In addition to this, Table 4 depicted the three most confusing Arabic consonants for every pharyngeal and uvular phoneme for each of the experiments. The brief description and experimental result about each consonant are discussed below.

3.1.1. Ain /C/

This phoneme is one of the unique Arabic consonants. All Arabic phoneticians agree about the place of articulation regarding this phoneme and they determine it as pharynx [8, 11, 14, 15, 16, 17, 18]. This phoneme can be described as fricative, voiced, non-emphatic, and pharyngeal sound. From Fig. 1, the second best accuracy for this phoneme is achieved from the experiment N/N and the accuracy is 71.4% while NN/NN encountered the lowest accuracy (50.6%).

In the experiment M/M, where mixed native and non-native Arabic speakers were used in both training and testing phases, the system achieved the highest accuracy (72.0%). Depending on the experimental result we can conclude that non-native Arabic speakers cannot pronounce this phoneme correctly. From Table 4, we can see /C/ is mostly confused with /r/ phoneme in all experiments. This phoneme /r/ is liquid, voiced, non-emphatic, and alveolar sound. /r/ phoneme also exists in English language. In addition to that /C/ phoneme is confused with /Q/, /TH/, and /f/ phonemes.

3.1.2. H\aa /H/

This phoneme is same as Ain /C/ except that it is unvoiced sound. It is investigated in all the five experiments and its accuracy reached its highest value (76.9%) with M/M experiment while the worst accuracy (26%) is encountered in N/NN. This is the worst accuracy in all experiment for Arabic pharyngeal and uvular sounds. From this result we can say that non-native Arabic speakers pronounce this sound in a completely wrong manner. They pronounce this like /h/ phoneme. /h/ phoneme exists in most languages including Arabic and English and it is the glottal counter part of the /H/ sound. From Table 4 we can see that /H/ phone is confused mostly by /h/ phoneme. This happened for all experiments but not N/NN experiment. This means that non-native Arabic speakers turn /H/ phoneme to its glottal counter part, /h/. Other phonemes that caused confusions for /H/, depending on Table 4, are /r/, /Q/, and /G/.

These findings are consistent with our other on going research [19, 20, 21].

3.2. Uvular consonants

Both the classical and MSA versions have three uvular consonants, namely *Ghain* /G/, *Qaaf*, /q/, and *Khaa* /x/. Figure 2 plots the accuracies of these three uvular consonants for all the five experiments. The brief description and experimental result about each consonant are discussed below.

3.2.1. Ghain /G/

We count it as uvular because it is vocalized from the uvular as its place of articulation, though there is confusion between researchers regarding this. The confusion occurs because the closeness in locations of the velar, palatal, and uvular in the human being vocal tract area. The phonemes /q/, /x/, and /G/ have the same place of articulation when we experience vocalization of them in same setting of words. We can define it as a fricative, voiced, non-empathic, and uvular Arabic phoneme.

The best accuracies are 63.3% and 64% occurred in case of N/N and N/NN, respectively, while the worst accuracy (47.5%) is encountered in NN/N. For all experiments, phoneme /r/ is mostly confused with phoneme /G/. Phonemes /C/ and /Q/ are also confused with phoneme /G/. Table 4 shows that there are big similarities in pronunciation of /C/, /G/ and /r/.

3.2.2. Qaaf /q/

This can be described as stop, unvoiced, non-emphatic, and uvular Arabic phoneme. We investigated /q/ phoneme and found that the best accuracy (55%) for in the experiment N/N while the worst performance (30.4%) is encountered in NN/N. Also the accuracy of the experiment NN/NN is 31.9%. We can conclude that the non-native speakers cannot pronounce this phoneme correctly. Depending on Table 4 this phoneme is confused with /Q/ phoneme in all experiment, which means that there is big similarity between /q/ and /Q/ phonemes. In fact /Q/ phoneme is the glottal counterpart of /q/ phoneme (i.e., /Q/ phoneme differ from /q/ phoneme only in the place of articulation where the former one is glottal and the latter one is uvular). /r/ and /T/ are among the confusing phonemes of /q/ phoneme.

3.2.3. Khaa /x/

If we practice vocalization of this sound, we have the confidence it is an uvular sound, hence we count it as uvular one. We say this because all /q/, /x/, and /G/ phonemes have the same place of articulation when we experience vocalization of them in same kind of words.

Hence /x/ phoneme can be described as fricative, unvoiced, non-emphatic, and uvular sound. In our experiments, it has been found that the best accuracy (86.7%) is encountered in experiment M/M while the worst accuracy (34.5%) is encountered with experiment NN/N. This accuracy is excellent compared to all other experiments and phoneme accuracies in this group of sounds. Similar low accuracy is encountered with experiment NN/NN. In case of the experiment N/N the accuracy is 58%. Depending on Table 4 we can see that this phoneme is confused mostly by /h/, /Q/, and /r/ phonemes.

General conclusion is that the accuracies in the experiments N/N and M/M are better than the other experiments namely N/NN, NN/N, and NN/NN. From this we can say that non-native Arabic speakers cannot pronounce these pharyngeal and uvular phonemes correctly. The most puzzling phonemes for pharyngeal and uvular are dependent on the mother tongue of the speakers used in training phase and/or testing phase.

4. Conclusions

An Arabic phoneme recognition system is designed and used to investigate five Arabic pharyngeal and uvular phonemes in Modern Standard Arabic (MSA). This speech recognition recognizes speech signal by using phoneme level without using any language model. The most three confusing phonemes that degraded accuracy of every phoneme in our set were presented and discussed. The most important outcome from all manipulation presented in this research can be summarized in that the non-native Arabic speakers contributed a big portion of miss-recognizing pharyngeal and uvular sounds. This problem is caused by the wrong pronunciation of non-native Arabic speakers to these five Arabic unique sounds. The reason for this problem is difficulty of correctly pronouncing these sounds by non-native Arabic speakers and similarity between these phonemes with other phonemes. In one experiment, phoneme /h/ which is the glottal counterpart of /H/ phoneme badly confused with the /H/ phoneme and thereby increasing the error rate. This investigation provides several significant directions for future research. Close aural inspection of the pronunciation by native Arabic speakers in the WestPoint Corpus found considerable regional variation, in other words a foreign-accented MSA. This suggests the need for research on Arabic ASR to control for regional and other social correlates of phonetic variation. Attention to processes such as the spread of the pharyngeal feature of the emphatic consonants to neighboring segments should also inform future work in this area.

5. References

[1] Selouani, S. and Caelen, J., "Arabic Phonetic Features Recognition using Modular Connectionist Architectures", Interactive Voice Technology for Communication, IVTTA'98, in Proceedings of IEEE 4th Workshop, 155-160, Trino, Italy, 29-30 Sep 1998.

[2] Economic and Social Commission for Western Asia-United Nations Report, "Harmonization of ICT Standards Related to Arabic Language Use in Information Society Applications", United Nations, New York, 2003.

[3] Kirchhoff, K., Bilmes, J., Das, S., Duta, N., Egan, M., Gang J., Feng H., Henderson, J., Daben L., Noamany, M., Schone, P., Schwartz, R., and Vergyri, D., "Novel approaches to Arabic speech recognition: Report from the 2002 Johns-Hopkins Summer Workshop", in Proceedings of ICASSP 2003, 1:344-347, Apr 2003.

[4] Kirchhoff, K., Bilmes, J., Das, S., Duta, N., Egan, M., Gang J., Feng H., Henderson, J., Daben L., Noamany, M., Schone, P., Schwartz, R., and Vergyri, D., "Novel Speech Recognition Models for Arabic: Johns-Hopkins University Summer Research Workshop 2002," Final Report, http://www.clsp.jhu.edu/ws02.html, accessed on 19 Mar 2008.

[5] Yousif A. El-Imam, "An Unrestricted Vocabulary Arabic Speech Synthesis System," IEEE Trans. on Acoustic, Speech, and Signal Proc., 37(12): 1829-1845, Dec 1989.

[6] Mansour Alghamdi, "Analysis, Synthesis and Perception of Voicing in Arabic," Al-Toubah Bookshop, Riyadh 2004.

[7] Daneil L. Newman, "The Phonetics Status of Arabic Within World's Languages: The uniqueness of the Dhaad Language", Antwerp Papers in Linguistics, 100:65-75, 2002, Online: http://www.dur.ac.uk/daniel.newman/Apil2.pdf, accessed on 19 Mar 2008.

[8] Mansour Alghamdi, "Arabic Phonetics," Al-Toubah Bookshop, Riyadh 2001. (In Arabic).

[9] Young, S., Evermann, G., Gales, M., Hain, T., Kershaw, D., Moore, G., Odell, J., Ollason, D., Povey, D., Valtchev, V., and Woodland, P., "The HTK Book (for HTK Version. 3.3)", Cambridge University Engineering Department, 2005. Online: http:///htk.eng.cam.ac.uk/prot-doc/ktkbook.pdf, accessed on 19 Mar 2008.

[10] Linguistic Data Consortium (LDC) Catalog Number LDC2002S02, Online: http://www.ldc.upenn.edu/, accessed on 19 Mar 2008.

[11] Ahmed Omar, "Derasat Alswaut Aloghawi," Aalam Alkutob, Eygpt 1991. (in Arabic)

[12] Muhammad Alkhouli, "Alaswaat Alaghawaiyah," Daar Alfalah, Jordan 1990 (in Arabic)

[13] M. Elshafei, "Toward an Arabic Text-to -Speech System," The Arabian Journal for Science and Engineering, 16(4B):565-83, Oct 1991.

[14] Mohammed A. Mahmoud, "Phonetics Science," Dar Eshbella, Riyadh 2003. (In Arabic).

[15] Essam Nour-Aldeen, "Sciences of Linguistic Phonology," Dar Alfikr Allubnani, Beirut 1992 (In Arabic).

[16] Kamal M. Bisher, "Arabic Phonetics," Shabab Liberary, Cairo, 1990. (In Arabic).

[17] Gloria J. Borden, Katherine S. Harris, "Speech Science Primer," Dar Al-Sharq Al-Arabi (Translated and expanded to Arabic Language by M. A. Homeidi), Beirut 1990.

[18] Muhammad Alkhouli, "Alaswaat Alaghawaiyah," Daar Alfalah, Jordan 1990 (in Arabic).

[19] Yousef Alotaibi and Sid-Ahmed Selouani, "Experiments on Adaptation to Non-Native Arabic Accent in Automatic Speech Recognition", Journal of Saudi Computer Society, Applied Computing and Information, Accepted on Feb.27, 2008.

[20] Yousef Ajami Alotaibi, Sid-Ahmed Selouani, and Douglas O'Shaughnessy, "Experiments on Automatic Recognition of Nonnative Arabic Speech," EURASIP Journal on Audio, Speech, and Music Processing, vol. 2008, Article ID 679831, 9 pages, 2008. doi:10.1155/2008/679831

[21] Sid-Ahmed Selouani, Yousef Ajami Alotaibi, "Investigating Automatic Recognition of Non-Native Arabic Speech", IEEE conference on Innovations in Information Technology, pp. 451-455, Dubai, UAE, 2007.