KACST Arabic Phonetics Database

Mansour M. Alghmadi
Computer & Electronics Research Institute,
King Abdulaziz City for Science and Technology,
Riyadh, Kingdom of Saudi Arabia
mghamdi@hotmail.com

ABSTRACT

Arabic sounds were the first to be fully described and analyzed. The place and manner of articulation of each Arabic sound were identified and documented in the eighth century AD by Sibawayah in his famous treatise *al-Kitaab*. Since then many Arabic scholars have quoted him, but not much has been added to what he wrote

Lately, King Abdulaziz City for Science and Technology (KACST) has published a detailed and comprehensive database called KACST Arabic Phonetics Database (KAPD) [1]. KAPD gives almost all the details of the articulatory mechanism of Arabic sounds. It contains more than 46000 files. They show the results of 9 experiments on 8 native Arabic subjects. The experiments include: airflow, air pressure, linguapalatal contact, nasality, perception, side and front facial images, and stroboscopic images of the glottis, pharyngeal cavity and velo-pharyngeal port. The data in KAPD are raw, i.e., it can be utilized for research and development such as speech therapy, speech perception, speech synthesis, speech recognition and speech modeling. KAPD is available on 3 CD's for researchers and students of Speech.

1. INTRODUCTION

Arabic sounds were identified and their places and manners of articulation were described and categorized 13 centuries ago by Sibawayh in his treatise *Al-Kitab* [2]. Many other Arab linguists came after him and wrote on Arabic sounds but *Al-Kitab* has remained their main reference. However, since the birth of Phonetics as an independent area of science in the 19th century and the use of modern equipment and research tools, the studies of Arabic sounds remain mostly individual efforts by Ph. D. students and a few interested phoneticians on certain aspects of Arabic sounds. Their results on articulatory phonetics are reported in references such as [3, 4, 5, 6, 7, 8, 9, 10 and 11]. Other results on Arabic acoustic phonetics are available in [6, 11, 12, 13, 14, 15 and 16], to mention some. A few researchers have worked on Arabic speech recognition [9 and 15].

Researchers usually collect data, analyze them, study them and write their findings on papers, as it is the case in almost all research fields. What other researchers would like to do is take the results and try to utilize them in other related areas of knowledge, but they do not have access to the original data. The original data are some times of an important value for certain researchers who would like to study them from different point of view or would like to utilize them in areas such as training a speech recognition engine or modeling a text-to-speech system.

King Abdulaziz City for Science and Technology (KACST) has realized the need for a database on Arabic phonetics which can be used by researchers of Arabic in particular and human speech in general. In 2000, KACST introduced its Arabic Phonetics Database (KAPD). The database includes more than 46000 files on Arabic sounds. The files are photos taken by a laryngoscope and video cameras (JPEG format), NSP file formats acquired by some of Kay's software and hardware systems, audio files in .wav format, and a table of the perceptual responses to all uttered tokens.

2. METHOD

The objective of KAPD was to provide researchers who are interested in Arabic sounds with as much detailed data as possible. Therefore, 9 experiments were designed to collect data for each sound: 1) Facial images, 2) velo-pharyngeal port images, 3) pharyngeal images, 4) laryngeal images, 5) aerodynamic data, 6) linguapalatal-contact data, 7) electroglottographic (EGG) data, 8) nasality data 9) perceptual results.

The tokens in experiments 2, 3, 4 and 5 are in the frame cvCvc/cvCCvc where c is /z/, v is /a/ and C is one of the Arabic 28 consonants. The target consonant is single in cvCvc and geminate in cvCCvc. The tokens in experiments 1, 6, 7 and 8 are in the frames: cvCvc/cvCCvc, Cvc and cvC where c is /z/, v is one of the vowels /a , i , u/ and C is one of the Arabic 28 consonants. The target consonant C is single in cvCvc and geminate in cvCCvc. The Arabic consonants which were used in the experiments are: the plain stops /b , t , d , k , q , ʔ /; the emphatic stops /tˤ , dˤ /; the plain fricatives /ɵ , ð , f , s , z , ʃ , ʒ , χ , ʁ , ħ , ʕ , h/; the emphatic fricatives /ðˤ , sˤ /; the nasals /m , n/; the glides /j , w/; the lateral /l/; and the trill /r/.

7 subjects participated in the experiments 1 to 8 (Table 1), while 10 other subjects participated in the perception experiment 9. All the subjects are native speakers of Arabic and carry the Saudi accent of Arabic. The subjects in the perception experiment are undergraduate students at

Umalqura University and none of them is one of the 7 subjects in the other experiment.

	Average	SD
Age	32.71 years	3.73
Height	167.71 cm	4.11
Weight	77.57 kg	11.66

Table 1. *The average and standard deviation (SD) of the subjects' age height and weight.*

In experiment 1, a color video camera (Sony Model: DXC-637P) was used for side view and another one (Sony Model: BVP-7P) was used for the front view. The recording was made by a video cassette recorder (Model: BVW-75P BETACAM SP). In experiment 2, 3 and 4, a Rhino-Laryngeal stroboscope (Model: RLS 9100) was used (Figure 1). In experiment 5, an aerophone (Model: Aerophone-II 6800) was used to collect and record the data (Figure 2). In experiment 6, Palatometer (Model: 6300) was used (Figure 3). In experiment 7, ElectroGlottoGraph was used. In experiment 8, Nasometer (Model: 6200-3) was used. In all the above experiments CSL (Computerized Speech Laboratory, Model: 4300B) was used to record the acoustic signal of the subjects' utterances. Most of the equipment used in these experiments carries the name of Kaytelemetrics. However, some of them were modified to cover the need of the experiments and the Arabic sounds (for example, Figure 3). In experiment 9, a computer program was written to give the subjects the choice of selecting the target sound that they have heard, its position in the token and its adjacent vowel.

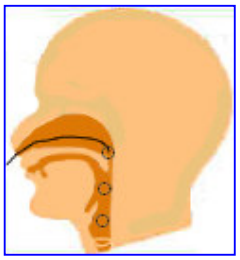

Figure 1. *The three circles show the position of the distal tip in experiments 2, 3 and 4.*

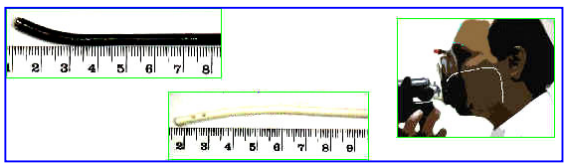

Figure 2. *The distal tip of the laryngoscope, the air-presser tube (the measurement is in centimeters), and one of the subjects wearing the air flew mask while the air-presser tube is inserted into the pharyngeal cavity via the nasal cavity.*

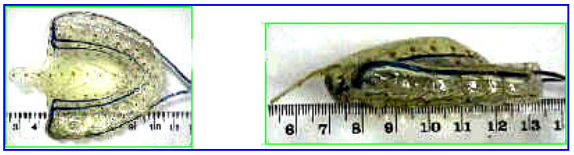

Figure 3. *The front and side views of the modified artificial palate with an extension to cover the uvula area.*

3. RESULTS

The results of experiment 1 are under directory C in KAPD. They include facial images of the 7 subjects during the production of Arabic sounds (Figure 4), NSP files with tags showing the target sound onset and offset, and WAV files of the same tokens. The NSP files contain the acoustic data plus other data such as linguapalatal contact and EGG. NSP files can be opened by a CSL or Multi-Speech program, both produced by Kaytelemetrics. WAV files can be opened by many programs including Windows and CSL. The results of experiment 2 are under the directory V. They include images of the velopharyngeal port (Figure 5), NSP and WAV files. The results of experiment 3 are under the directory E. The results of this experiment include images of the pharyngeal cavity (Figure 5), NSP and WAV files. The results of experiment 4 are under the directory X. They include images of the glottis (Figure 5), NSP and WAV files. The results of experiment 5 are under Directory A. They include NSP and WAV files. The NSP files have the aerodynamic data (Figure 6). The results of Experiment 6 are under Directory P. They include NSP (Figure 7) and WAV files. The results of experiment 7 are under Directory G. They include NSP (Figure 8) and WAV files. The results of experiment 8 are under Directory N. They include NSP (Figure 9) and WAV files. The results of experiment 9 are in the File DB1. They are the responses of the 10 informants who listened to the tokens produced by the subjects in the first 8 experiments and made their judgments on the target sounds they had heard.

Figure 4. *The side and front views of the face of one of the subjects.*

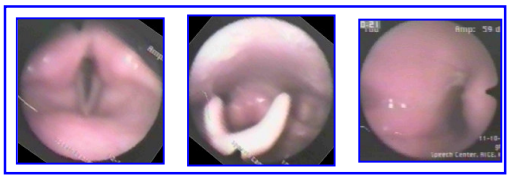

Figure 5. *Images of, from left to right, the vocal folds (experiment 2), epiglottis (experiment 3) and the velopharyngeal port (experiment 4).*

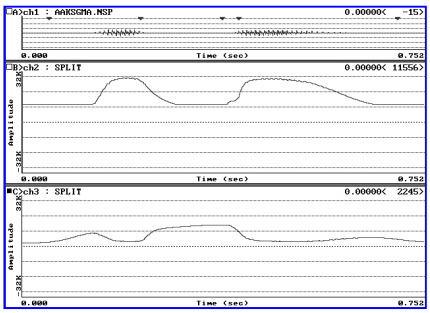

Figure 6. *Window 1 shows the waveform of /zakkaz/, window 2 shows the airflow and window 3 shows the air pressure of the same token (experiment 5).*

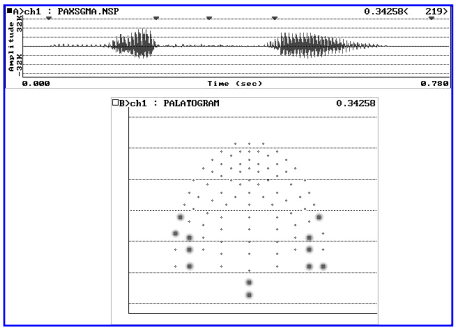

Figure 7. *Window 1 shows the waveform of /zaXXaz/, window 2 shows the linguapalatal contact during the production of the target sound /X/ (experiment 6).*

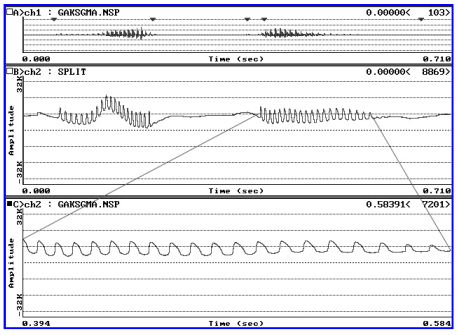

Figure 8. *Window 1 shows the waveform of /zakkaz/, window 2 shows the laryngeal activities and window 3 magnifies a portion of window 2 (experiment 7).*

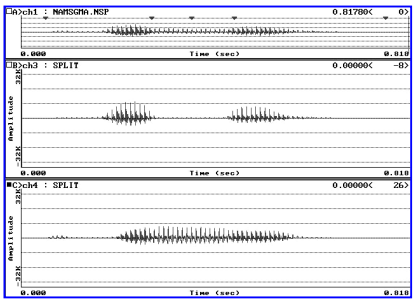

Figure 9. *Window 1 shows the waveform of /zammaz/, window 2 shows the oral air pressure and window 3 shows the nasal air pressure (experiment 8).*

The name of each file in the database has 7 letters. The first letter from left represents the experiment, the second represents the subject, the third and forth represent the target sound, the fifth represents the gemination, the sixth represents the position of the target sound in the token, and the seventh represents the adjacent vowel to the target sound. For example, the file PMBSSIU.NSP means that this file is about experiment 6 (linguapalatal contact), M is the subject's initial, BS represent the /b/ sound, S means the target sound is single not geminate, I means the target sound is in the initial position of the token, U means the adjacent vowel is /u/. The extension NSP is the file format.

4. CONCLUSIONS

This paper introduces the reader to a database on Arabic sounds produced by Saudi subjects. The database includes articulatory, acoustic and perceptual information. The articulatory data consist of images of the face, velopharyngeal port, pharyngeal cavity and glottis. They also include linguapalatal contacts and aerodynamic data. The acoustic data consist of the audio recording of the tokens made by the subjects in all experiments. The perceptual data show the responses of 10 informants to the Arabic sounds in the acoustic data.

KAPD can be utilized in various areas. To mention some, in speech-to-text experiments, the acoustic part of the database can be used to train and test the system where the recording of 7 native Arabic speakers producing Arabic sounds 8 times in different token positions. KAPD can be used in text-to-speech systems especially when using the concatenation method. It can also be used in vocal tract modeling and facial animation during speech. It would be of value for researchers in phonetics, speech therapy and forensic phonetics.

5. ACKNOWLEDGEMENT

This paper and the database project are sponsored by King Abdulaziz City for Science and Technology. The database team consists of: Mansour Alghamdi, Ashraf Alkhairi, Abdullah Azzamil, Muhammad Alfarraj, Muhammad Aljabr, Muhammad Basalamah, Sayed Hussain, Surraya

Addawsari, and Waqayyan Addawsari.

REFERENCES

[1] (2000) *KACST Arabic Phonetics Database*, Computer and Electronics Research Institute, King Abdulaziz City for Science and Technology, Riyadh.

[2] Nassir, Abdulmun'im (1993) *Sibawayh the Phonologist*: *A Critical Study of the Phonetic and Phonological Theory of Sibawayh As Presented in His Treatise Al-Kitab*, Kegan Paul Intl.

[3] Kuriyagawa, F., M. Sawashima, et al. (1988). "Electromyographic study of emphatic consonants in standard Jordanian Arabic", *Folia Phoniatrica: International Journal of Phoniatrics, Speech Therapy and Communication Pathology* 40(3): 117-122.

[4] Giannini, A. (1982). *The Emphatic Consonants in Arabic*. Naples, Istituto Univ. Orientale, Seminario di Studi dell'Europa Orientale, Laboratorio di Fonetica Sperimentale.

[5] Gegov, K. (1989). "The Bulgarian velar consonant /x/ and three Arabic fricative consonants." *Contrastive Linguistics* 14(2): 21-24

[6] Butcher, A. and K. Ahmad (1987). "Some acoustic and aerodynamic characteristics of pharyngeal consonants in Iraqi Arabic." *Phonetica* 44(3): 156-172.

[7] Boff, M.-C. (1983). "A Contribution to the experimental study of back consonants in classical Arabic (Moroccan speakers)." *Travaux de l'Institut de Phonetique de Strasbourg* 15: 1-363.

[8] Ghali, M. M. (1983). "Pharyngeal articulation (Especially in Arabic and English)." *Bulletin of the School of Oriental and African Studies-University of London* 46(3): 432.

[9] Alwan, A. (1989). "Perceptual cues for place of articulation for the voiced pharyngeal and uvular consonants." *Journal of the Acoustical Society of America* 86(2): 549-556.

[10] Ghazali, S.; (1977); Back consonants and backing coarticulation in Arabic; Ph. D. dissertation; University of Texas at Austin.

[11] Al-Ani, S. (1970) *Arabic Phonology*. The Hague.

[12] El-Halees, Y. A. (1987). "F2 and back articulations in Arabic." *Journal of the College of Arts, King Saud University* 14(1): 21-44.

[13] El-Halees, Y. (1985). "The role of F1 in the place-of-articulation distinction in Arabic." *Journal of Phonetics* 13(3): 287-298.

[14] Flege, J., and Port, R. (1981) Cross-language phonetic interference: Arabic to English, *Language and Speech* 24, 125-146.

[15] Alghamdi, M. (1990) Analysis, synthesis and perception of Arabic voicing; Ph. D. dissertation; University of Reading.

[16] Alghamdi, M. (1998) A spectrographic analysis of Arabic vowels: a cross-dialect study, *J. King Saud Univ*. 10, 3-24.