

## Difficulties of Standard Arabic Phonemes Spoken by Non-Arab Primary School Children based on Formant Frequencies

N.A. Abdul-Kadir and R. Sudirman  
Faculty of Electrical Engineering, University Technology Malaysia,  
81310 UTM Johor Bharu, Malaysia

**Abstract: Problem statement:** The study of Malaysian Arabic phoneme is rarely found which make the references work difficult. Specific guideline on Malaysian subject is not found even though a lot of acoustic and phonetics research has been done on other languages such as English, French and Chinese. **Approach:** This study discussed about the correct and simplest way of Arabic phonemes pronunciation in Malay accent. The International Phonetic Alphabet of Arabic chart was considered as the reference of every recorded speech samples using Malaysian children for their sound localization (makhray point) of every alphabet. The recorded sound was analysed to determine the origin of each alphabet data by measuring its formant frequencies. The consonants of Standard Arabic (SA) phonemes were studied and the appropriate place of articulation of every phoneme was measured through its formant. **Results:** Only seven out of 25 consonants of SA phonemes of the children's samples did not give the appropriate formants value. The formants are /kof/, [ق], /zo/, [ظ], /kho/, [خ], /gheyn/, [غ], /ha/, [ح], /ain/, [ع] & /ha/, [ه] which consider as the difficult SA to utter among Malaysian children. **Conclusion/Recommendations:** The values obtained are used as the reference of the database for our recognition system.

**Key words:** FFT spectrogram, place articulation, Arabic phonemes, formant frequencies

### INTRODUCTION

Spectrogram is well-known of its representation of visual image of a moving signal. Through a spectrogram, the signal waveform can be interpreted according to its colour distribution, time and frequency. The spectrum of the signal is shown in every small vertical slice of the spectrogram. The robustness and convenience of the spectrogram in speech application is approved through reading (Laprie and Berger, 1996; Hatazaki *et al.*, 1989; Silverman and Lee, 1987; Hunt, 1987; Connolly and Edmonds, 1986; Zue and Lamel, 1986; Zue and Cole, 1979; Mporas *et al.*, 2007).

From the spectrogram, Awais *et al.* (2006) has described about the detection of the phonemes boundaries and identify it as pauses, vowels or consonants. The result for segmenting continuous Arabic speech using FFT spectrogram is excellent. Every phoneme of 10 speakers is detected as consonants or vowels with overall accuracy of 95.39%.

Therefore, this study is to use the robustness of the spectrogram of each of the signal of SA phonemes to detect the formant frequencies and to study the waveform patterns. The formant frequency,  $F_n$  of the

phoneme is acquired according to Eq. 1 (Fant, 1960; Johnson, 2003; Rabiner and Juang, 1993; Mohd *et al.*, 2009).

$$F_n = \frac{(2n-1)c}{4L} \quad (1)$$

where,  $L$  is the length of the vocal tract which is measured from the place of articulation (makhray point) of the sound production to the lips. Based on equation (1), since  $L$  is the length of the vocal tract, formant frequency,  $F$  will be lower as the constriction located into the mouth as  $n$  and  $c$  is fixed, where  $n$  is the formant;  $c$  is the speed of sound in warm and moist air ( $35000 \text{ cm sec}^{-1}$ ). For example, phoneme /ain/, [ع] (in pharynx location) will possess greater formant frequency compared to /tsa/, [ث] (placed at dental place of articulation) as reported by Ali *et al.* (2001).

All speech sounds have its unique speech waveform patterns. Therefore, to obtain the spectrographic image of speech sound, speech data must be compressed and undergo signal acquisition process or digital signal processing method (Smith, 1997; Mitra, 2006).

**Corresponding Author:** R. Sudirman, Faculty of Electrical Engineering, University Technology Malaysia, 81310 UTM Johor Bharu, Johor, Malaysia Tel: +6075535270 Fax: +6075566272

By using Fourier Transform technique, Equation (2) is visualized as an image with the convention that frequency is increasing from bottom to top as time is increases from left to right. While the pixel value at each point in the spectrogram is proportional to the magnitude of the spectrum at a certain frequency at some point in the time.

$$X_m(e^{j\omega}) = \sum_{n=-x}^{20} x(n)w(n-m)e^{-j\omega n} \quad (2)$$

Here:

- x(n) = The speech signal, and w(n) is a window of length L
- $X_m(e^{j\omega})$  = A collection of Fourier Transform of windowed segments of x(n). Where
- n = The number of sample while m is the number of windowed frame of the sample

An experiment done by Jackson (2001) on three English consonants for each voiced (/b, d, g/) and unvoiced (/p, t, k/) showed that formant  $F_4$  of English plosives is less reliable. While the value of formant  $F_2$  and  $F_3$  can distinguish the consonants according to the place of articulation. The inner the place of articulation, the results is higher in formant  $F_2$ . Iqbal *et al.* (2008) used formant frequencies to identify Arabic vowels and nasal formants /lam/, [ل], /mim/, [م] and /nun/, [ن]. Subjects were Quranic expertise which gave 90% average accuracy.

The following parts discussed the SA phonemes based on its manner of production. Each manner of productions is having the same characteristics of the spectrogram based on the way it is pronounced. The following parts describe the methods used, results and conclude the findings.

**Fricative phonemes:** Fricative is the largest set of SA because it contains thirteen phonemes in the set. Fricative sound is produced as air escaped a narrow constriction in the mouth (Fant, 1960; Johnson, 2003; Rabiner and Juang, 1993; Hassan, 1984). The classifications of fricative are as follows and its phonetic as pronounced by Malay subjects:

- Labiodental: /fa/, [ف]
- Dental: /zal/, [ذ] and /tsa/, [ث]
- Alveolar: /zai/, [ز] and /sin/, [س]
- Post-alveolar: /shin/, [ش]
- Coronal: /sod/, [ص]
- Palatal: /zo/, [ظ]
- Velar: /kho/, [خ] and /gheyn/, [غ]
- Pharyngeal: /haa/, [ح] and /ain/, [ع]

- Glottis: /ha/, [ه]

**Plosive phonemes:** It consists of nine phonemes in SA which is the second largest set of the phonemes after fricative. Plosive is the behaviour of sound produced by stopping the airflow in the mouth (Fant, 1960; Johnson, 2003; Rabiner and Juang, 1993; Hassan, 1984). Plosive is sometimes called stop or spirant. Those include bilabial, dental, velar, alveolar, interdental and uvular. The plosive classifications are as follows and their pronunciations are in Malay accent:

- Bilabial: /ba/, [ب]
- Dental: /dal/, [د] and /ta/, [ت]
- Interdental: /tho/, [ط]
- Alveolar: /dod/, [ض]
- Velar: /jim/, [ج] and /kaf/, [ك]
- Uvular: /kof/, [ق]
- Glottal: /alif/, [ا]

**Nasal phoneme:** In SA, only two phonemes are categorized in this group. Nasal is produced with a lowered velum in the mouth, allowing air to escape freely through the nose (Fant, 1960; Johnson, 2003; Rabiner and Juang, 1993; Hassan, 1984). The classifications of nasal phonemes are bilabial: /mim/, [م] and dental: /nun/, [ن].

**Approximant phonemes:** Approximant is the manner of articulation which composed of two phonemes of SA. Both of these approximant are semivowel, which produced by bringing one articulator (alveolar) in the vocal tract close to another (Fant, 1960; Johnson, 2003; Rabiner and Juang, 1993; Hassan, 1984). The approximant phonemes are palatal: /ya/, [ي] and velar: /wao/, [و].

**Trill phoneme:** Only a phoneme is categorized as trill, which is produced by vibration of the tongue against some other part of the mouth (alveolar). The trill phoneme is an alveolar: /ro/, [ر].

**Lateral phoneme:** Lateral also composed of a phoneme in the group of SA. The phoneme is produced by raising the tip of the tongue against the roof of the mouth so that the airstream flows past one or both sides of the tongue (Johnson, 2003; Rabiner and Juang, 1993). The lateral phoneme is /lam/, [ل].

Figure 1 shows the distribution of the places of articulation in SA and Table 1 listed the SA phonemes with their manner and place of articulation (Awadalla *et al.*, 2005) under consideration.

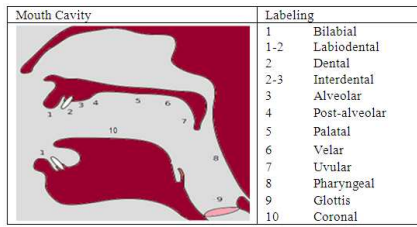


Fig. 1: Places of articulation

Table 1: The Standard Arabic phonemes in consideration

Phoneme	Manner of Production	Place of articulation
ب	Plosive	Bilabial
ت	Plosive	Dental
ث	Fricative	Dental
ج	Plosive	Velar
ح	Fricative	Pharyngeal
خ	Fricative	Velar
د	Plosive	Dental
ذ	Fricative	Dental
ر	Trill	Alveolar
ز	Fricative	Alveolar
س	Fricative	Alveolar
ش	Fricative	Post alveolar
ص	Fricative	Coronal
ض	Plosive	Alveolar
ط	Plosive	Interdental
ظ	Fricative	Palatal
ع	Fricative	Pharyngeal
ف	Fricative	Velar
ق	Fricative	Labiodental
ك	Plosive	Uvular
ل	Plosive	Velar
م	Lateral	Post alveolar
ن	Nasal	Bilabial
ة	Nasal	Dental
هـ	Fricative	Glottal

## MATERIALS AND METHODS

The recording of the speech signal is done on a room at a tuition centre with quiet environment. All of the subjects are Malaysian children age of seven to eleven years old. The children are ten boys and fifteen girls which are randomly selected. Their background in Arabic phoneme is unknown. The data sound is collected and analysed in order to determine the accuracy level of Malaysian children pronunciation of Arabic phoneme. Other requirements of the project are as follows:

- Recording Software: Easy Hi-Q Recorder with sampling rate of 44.1 kHz
- Analysis Software: Goldwave; and Speech Filing System (SFSWin) version 1.7 2008
- Recording type and format: \*.wav, 16-bit mono
- Recording device: External Mic

The experiment of data collection and data analysis is done separately. During the recording session, the children were asked to pronounce all of the 28 Arabic phonemes sequentially. The recording is started a few seconds before the children utter the first phoneme and continue until all of the phonemes pronounced. The recording is stopped a few seconds after the last phoneme is spoken for each individual. The speech sound collections of 25 children are gathered.

The first step in data analysis is to group every speech sound recorded according to its phoneme. By using Goldwave software, the recorded sound for each child is cut into one second length. Therefore, for each phoneme, a total of 25 samples are collected. After that, SFS software is used to analyse the data. The wide-band spectrograms for each data are produced and the formant frequencies ( $F_n$ ) are determined. The average data is calculated since the appropriate data for each phoneme is not single. All of the average measured formant frequencies are summarized in results and discussion. The average value is the average of 25 samples of each phoneme averaged using Eq. 3.

$$F_n = \frac{1}{25} \sum_{i=1}^{25} F_i \quad (3)$$

$F_n$  = the formant frequency where  $n = 1, 2, 3, 4$  in which  $F_1$  refer to the lowest formant and  $F_4$  = the highest formant measured in the study. Figure 2 is the flow summary of method in this study and Fig. 3 is the algorithm to obtain the spectrogram.

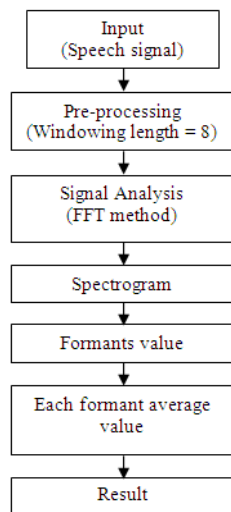


Fig. 2: The workflow to obtain the average value of formant frequencies

- 1- Short segmented the audio signal.
- 2- Used FFT method to find the frequency spectrum of each segment.
- 3- Converted these spectra into grayscale image (dark colour distribution shows the high amplitude while light colour shows the low amplitude).
- 4- Get the graphical image (spectrogram) which the frequency content of speech changes with time is provided.

Fig. 3: Steps/Algorithm to spectrographic

## RESULTS AND DISCUSSION

The pattern for all of the phonemes is observed in the spectrogram. The waveform for each features are compared within the group of manner of articulation and the formant frequencies are observed. The average value of each formant is summarized as in Table 2, 4, 6, 7, and 9 according to its place of articulation. The waveform of the fricative sound is shown in Fig. 4. As the fricative sound is produced by the jet air flowing through a constriction in the vocal tract, the starting of the formant is hissing due to air hitting the front teeth and caused turbulence for all frequency, before the formant can be seen clearly at time 450 ms as circled in the spectrogram.

From the literature review,  $F_1$  and  $F_4$  are not reliable. It is proven when the frequencies are increased and decreased randomly not according to the place of articulation. As summarized in Table 2, the value of  $F_2$  is increasing from bilabial to post-alveolar phonemes. As the place of articulation move inner to the mouth, the value is increased and decreased inconsistently. While  $F_3$  shows the increment of the value from dental phonemes to velar phoneme (/kho/, [خ]) but decreased and increased inconsistently for velar phoneme (/gheyn/, [غ]) and inner place of articulation inside the mouth. Coronal phoneme (/sod/, [ص]) which pronounced as the blade of the tongue is raised and lips are rounded and the place of articulation is considered to be longer than /shin/, [ش]. Therefore, the value of  $F_2$  and  $F_3$  observed for phoneme /sod/, [ص] is greater than /shin/, [ش].

For fricative phonemes, only /fa/, [ف], /tsa/, [ث], /zal/, [ذ], /zai/, [ز], /sin/, [س], /shin/, [ش] and /sod/, [ص] give the increasing value of  $F_2$  respectively which distributed from bilabial to coronal place of articulation. It is found that, the difficult fricative consonants to utter are /zo/, [ظ], /kho/, [خ], /gheyn/, [غ], /ha/, [ح], /ain/, [ع] and /ha/, [ه].

The result is divided into two parts based on the place of articulation which are front and back. These

refer to the location of the place of articulation. Front consonants consists of seven phoneme which articulated at labio-dental, dental, alveolar, post-alveolar and coronal (/fa/, [ف], /tsa/, [ث], /zal/, [ذ], /zai/, [ز], /sin/, [س], /shin/, [ش] & /sod/, [ص]). While back consonants consists of six phoneme which articulated at palatal, velar and glottal (/zo/, [ظ], /kho/, [خ], /gheyn/, [غ], /ha/, [ح], /ain/, [ع] & /ha/, [ه]).

From the result, only seven fricative phonemes (front fricative consonant) gave the result appropriate to Jackson (2001). Other six fricative phonemes (back fricative consonant) did not obey the findings. The results show that  $F_2$  and  $F_3$  can distinguish the place of articulation for front fricative consonants as the formant increase from bilabial to post-alveolar while  $F_1$  and  $F_4$  is less reliable to be used in the same position as summarized in Table 3.

In Fig. 5, the waveform of the plosive consonant (/baa/, [ب]) is started with a burst before the waveform is easily seen on the spectrogram. This is because the plosive consonants are produced by stopping the air flow in vocal tract.

The average value of each phoneme is summarized as in Table 4. By neglecting the changes of  $F_1$  and  $F_4$ , the value of  $F_3$  is increased from 3.9 kHz to 4.4 kHz at dental phonemes to velar phonemes.

The formants  $F_2$  and  $F_3$  for uvular phoneme; /kof/, [ق] are less than velar phoneme /kaf/, [ك] and /jim/, [ج]. It is shown that only one phoneme (/kof/, [ق]) did not obey the result as concluded in above. This shows that Malaysian children mostly have difficulty in articulating the plosive uvular phoneme correctly as discussed by Abdul-Kadir *et al.* (2010). Comparison between previous findings with current is summarized in Table 5.

Fig. 6 shows the waveform of /mim/, [م] and its spectrogram. From the spectrogram, the starting region is less dark compared to middle region as for nasal pronunciation, the vocal tract is divided into nasal branch and oral branch. Therefore, more antiresonances are produced as the interference between these two branches. Antiresonances in the vocal tract will eliminate formants near it which appear to be weak when look at the spectrogram.

From Table 6, the values of  $F_1$  and  $F_3$  for /mim/, [م] are decreased from front to end of the phoneme. While  $F_2$  and  $F_4$  decreased from front to middle but increased to the end of the phoneme. Meanwhile, for phoneme /nun/, [ن], the value of  $F_1$  increasing from front to middle but decreased from middle to end, but the values of  $F_2$ ,  $F_3$  and  $F_4$  are decreased from front to middle and increased from middle to end of the phoneme. The difference between /mim/, [م] and /nun/, [ن] nasal consonants pronunciation is that the lip rounding when the phonemes pronounced.

Table 2: The average value for formants of fricative consonants

Place of Articulation	Phoneme	F1	F2	F3	F4
Labiodental	/fa/, [ف]	1.0	2.0	4.0	5.4
Dental	/tsa/, [ث]	1.0	2.1	3.8	5.3
	/zal/, [ذ]	0.4	2.2	3.8	5.1
Alveolar	/zai/, [ز]	1.0	2.2	4.0	5.5
	/sin/, [س]	0.5	3.1	4.0	5.2
Post-alveolar	/shin/, [ش]	1.6	3.5	4.3	5.1
Coronal	/sod/, [ص]	1.8	3.6	4.7	5.5
Palatal	/zo/, [ظ]	1.2	3.0	4.4	5.3
Velar	/kho/, [خ]	0.8	2.3	4.4	5.5
	/gheyn/, [غ]	0.6	2.0	3.7	4.9
Pharyngeal	/ha/, [ح]	0.6	1.2	3.9	5.3
	/ain/, [ع]	1.0	2.8	4.4	5.7
Glottis	/ha/, [ه]	1.0	1.9	3.5	4.5

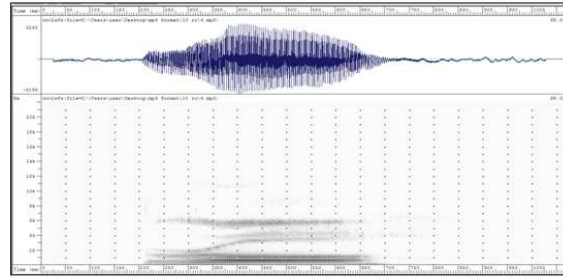


Fig. 8: Spectrogram of trill /ro/, [ر]

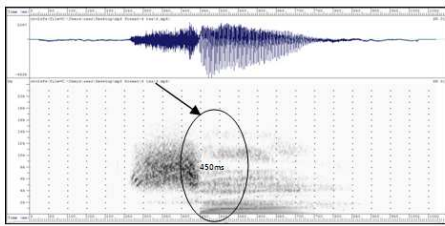


Fig. 4: Spectrogram of fricative /tsa/, [ث]

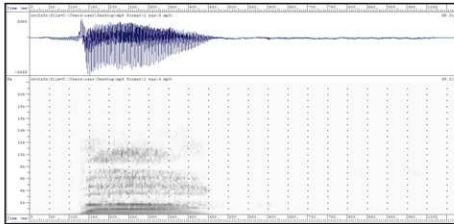


Fig. 5: Spectrogram of plosive /baa/, [ب]

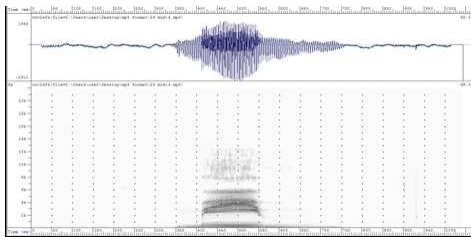


Fig. 6: Spectrogram of nasal /mim/, [م]

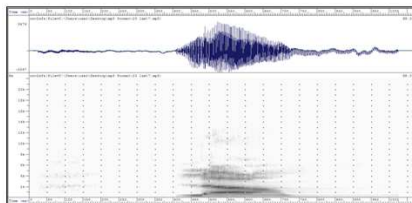


Fig. 7: Spectrogram of lateral /lam/, [ل]

Lateral /lam/, [ل] is produced when the air escape from both sides of the tongue while at that time, the tip of the tongue is placed on the alveolar edge (post-alveolar). The effect of the blocked oral tract in the centre and effect of side branches can be seen through the spectrogram in Fig. 7.

From the spectrogram, the average value for /l/ F<sub>1</sub> is 0.4 kHz, F<sub>2</sub> is 2.3 kHz, F<sub>3</sub> is 3.8 kHz and F<sub>4</sub> is 4.9 kHz as summarized in Table 7. From front to middle (in order to produced /la/ sound), F<sub>1</sub> is increased from 0.4 kHz to 1.1 kHz, while F<sub>2</sub>, F<sub>3</sub> and F<sub>4</sub> decreased to 1.9 kHz, 3.5 kHz and 4.6 kHz respectively. To complete the /lam/, [ل] pronunciation, the formants changed from middle to end is reversed as from front to middle.

Both nasal and lateral formant F<sub>2</sub> is not in the range as result discussed by Iqbal (2008) as shown in Table 8. This might be the effect of the vocal tract length since the study used children age 7 to 11-year-old, while Iqbal used expertise in Quranic recitation age 15 to 30 years old.

Trill consonant is produced by vibrating the tongue against alveolar. From the waveform and spectrogram in Fig. 8, it can be observed that the formants are decreasing except at F<sub>4</sub>. This is due to the way to pronounce the phoneme which less vibration (friction) can be heard as the tip of the tongue curled back towards the mouth. From the spectrogram, F<sub>4</sub> is observed to rise drastically.

The average value for F<sub>1</sub>, F<sub>2</sub>, F<sub>3</sub>, and F<sub>4</sub> is 0.6, 1.6, 3.8, and 5.3 kHz respectively as summarized in Table 9.

Consonant /ro/, [ر] is observed through the spectrogram only. It is shown that the formant transition for F<sub>4</sub> is drastically increased compared to other formants in the same spectrogram as found by Ladefoged (2003) (refer Table 10).

Table 3: Comparison findings between current study and previous research

Position	Study	F1	F2	F3	F4
Front	Jackson, 2001	Not stated	Distinguish the place of articulation	Distinguish the place of articulation	Less reliable
	This Study	Randomly increased and decreased	Increasing from bilabial to post-alveolar and coronal	Increasing from bilabial to post-alveolar and coronal	Randomly increased and decreased
Back	Jackson, 2001	Not stated	Distinguish the place of articulation	Distinguish the place of articulation	Less reliable
	This Study	Randomly increased and decreased	Randomly increased and decreased	Randomly increased and decreased	Randomly increased and decreased

Table 4: The formants average value of plosive consonants

Place of Articulation	Phoneme	F1 (kHz)	F2	F3	F4
Bilabial	/ba/, [ب]	0.5	3.0	4.0	5.0
Dental	/ta/, [ت]	0.8	2.1	3.9	5.2
	/dal/, [د]	0.6	2.4	3.9	4.7
Interdental	/tho/, [ط]	0.6	2.0	3.9	5.0
Alveolar	/dod/, [ذ]	0.5	2	4.0	4.9
Velar	/kaf/, [ك]	1.3	2.4	4.0	5.1
	/jim/, [ج]	1.0	3.3	4.4	5.2
Uvular	/kof/, [ق]	0.9	1.5	4.1	5.0

Table 5: Comparison findings between current study and previous research

Study	F1	F2	F3	F4
Ali <i>et al.</i> , 2001	-Formant is significant when transitions involved are sharp and clear -Bilabial is weaker than alveolar and velar			
Jackson, 2001	Not stated	Distinguish the place of articulation	Distinguish the place of articulation	Less reliable
This Study	Randomly increased and decreased	Increasing from bilabial to velar	Increasing from bilabial to velar	Randomly increased and decreased

Table 6: The average value for formants of nasal consonants

Vocal tract position		Front (kHz)				Middle (kHz)				End (kHz)			
Place of Articulation	Phoneme	F1	F2	F3	F4	F1	F2	F3	F4	F1	F2	F3	F4
Bilabial	/mim/, [م]	0.5	3.1	4	5.0	0.5	2.8	3.8	4.6	0.4	2	3.5	4.8
Dental	/nun/, [ن]	0.5	1.7	4	5.2	0.5	1.6	3.6	4.8	0.4	2.5	3.9	4.9

Table 7: The average value for formants of lateral consonants

Vocal tract position		Front				Middle				End			
Place of Articulation	Phoneme	F1	F2 (kHz)	F3	F4	F1	F2 (kHz)	F3	F4	F1	F2 (kHz)	F3	F4
Post-alveolar	/lam/, [ل]	0.4	2.3	3.8	4.9	1.1	1.9	3.5	4.6	0.4	1.9	3.7	5

Table 8: Comparison findings between current study and previous research

Position	Study	F1	F2	F3	F4
Nasal	Iqbal, 2008	Between 0.3- 0.5 kHz	1.25 – 1.65 kHz	Not stated	Not stated
	This Study	0.5 kHz	3.1([م]) and 1.7([ن]) kHz	4kHz	5 kHz
Lateral	Iqbal, 2008	Between 0.3- 0.5 kHz	1.25 – 1.65 kHz	Not stated	Not stated
	This Study	0.4 kHz	2.3 kHz	3.8 kHz	4.9 kHz

Table 9: The average value for formants of trill consonants

Place of Articulation	Phoneme	F1 (kHz)	F2	F3	F4
Alveolar	/ro/, [ر]	0.6	1.6	3.8	5.3

Table 10: Comparison findings between current study and previous research

Study	F1	F2	F3	F4
Ladefoged, 2004	Not stated	Not stated	Not stated	Raising
This Study	Not stated	Not stated	Not stated	Drastic changes (raising)

### CONCLUSION

In conclusion, the results are compared with previous research findings according to their manner of articulation. There are five types for all 25 consonants under consideration of this study. Most of manner of articulations show that F<sub>2</sub> and F<sub>3</sub> are reliable to know the exact place of articulation. The first and forth (F<sub>1</sub> and F<sub>4</sub>) formants cannot be relied on. From the experiment, only seven out of 25 consonants did not get the appropriate value. Those consonants are /kof/, [ك], /zo/, [ظ], /kho/, [خ], /gheyn/, [غ], /ha/, [ح], /ain/, [ع] & /ha/, [ه] which consider as the difficult consonants to utter with correct sound production.

### ACKNOWLEDGMENT

The researchers would like to express their gratitude to University Technology Malaysia and Ministry of Science, Technology and Innovation Malaysia (MOSTI) for supporting and funding this study under Vote 79368. The second gratitude goes to bio-Medical Instrumentation and Electronics Research Group (bMIE) for ideas and comments to this study.

### REFERENCES

Awais, M.M., W. Ahmad, S. Masud and S. Shamail, 2006. Continuous Arabic speech segmentation using FFT Spectrogram. *Innovations Inform. Technol. Dubai*, 19-21. 1-6. DOI: 10.1109/INNOVATIONS.2006.301939

Awadalla, M., F.E.Z Abou Chadi and H.H. Soliman, 2005. Development of an Arabic speech database. *Information and Communications Technology, 2005. Enabling Technologies for the New Knowledge Society: Proceedings of the ITI 3rd International Conference, Dec. 5-6, Cairo*, pp: 89-100. DOI: 10.1109/ITICT.2005.1609617

Abdul-Kadir, N.A., R. Sudirman and N.M. Safri, 2010. Modelling of the Arabic plosive consonants characteristics based on spectrogram. *Proceedings*

of the Fourth Asia International Conference on Mathematical/Analytical Modeling and Computer Simulation (AMS), Sabah, Malaysia, June 26-28, pp: 282-285. DOI: 10.1109/AMS.2010.63

Ali, A.M.A., J. Van der Spiegel and P. Mueller 2001. Acoustic-phonetic features for the automatic classification of stop consonants. *IEEE Trans. Speech Audio Process*, 9: 833-841. DOI: 10.1109/89.966086

Connolly, J.H. and E.A. Edmonds, 1986. Automatic speech recognition based on spectrogram reading. *Int. J. Man-Machine Stud.*, 24: 611-621. DOI: 10.1016/S0020-7373(86)80012-8

Fant, G., 1960. Acoustic theory of speech production with calculations based on X-ray Studies of Russian Articulations. Netherlands: Mouton and Co, The Hague. Chapter 2 and 3. ISBN: 9027916004.

Hassan, A., 1984. *Linguistik am Untuk Guru Bahasa Malaysia*. 5th Edn., Fajar Bakti, Selangor, Malaysia. ISBN: 019581990X.

Hatazaki, K., K. Yasuhiro and K. Takeshi, 1989. Phoneme segmentation using spectrogram reading knowledge. *Acoustics, Speech, and Signal Process.*, ICASSP-89. May 23-26, Glasgow, Scotland, UK, pp: 393-396. DOI: 10.1109/ICASSP.1989.266448

Hunt, M.J., 1987. Delayed decisions in speech Recognition-The case of formants. *Patt. Recog. Lett.*, 6: 121-137. DOI: 10.1016/0167-8655(87)90093-6

Iqbal, H.R., M. Awais, S. Masud and S. Shamail, 2008. On vowels segmentation and identification using formant transitions in continuous recitation of Quranic Arabic. *New Challenges in Applied Intelligence Technologies*. 134/2008. Springer Berlin/Heidelberg, pp: 155-162. DOI: 10.1007/978-3-540-79355-7\_15

Jackson, P.J.B., 2001. Acoustic cues of voiced and voiceless plosives for determining place of articulation. *Proceeding of the Workshop on Consistent and Reliable Acoustic Cues for sound analysis, CRAC. Aalborg, Denmark*, pp: 19-22. DOI: 10.1.1.124.6942

Johnson, K., 2003. *Acoustic and Auditory Phonetics*. 2nd Edn., United Kingdom: Blackwell Publishing. Chapter 7, 8, & 9. ISBN: 978-1-4051-0122-6.

Laprie, Y. and M.-O. Berger, 1996. Cooperation of regularization and speech heuristics to control automatic formant tracking. *Speech Commun.*, 19: 255-269. DOI: 10.1016/S0167-6393(96)00036-2

Ladefoged, P., 2003. *Phonetic data analysis: An Introduction to Fieldwork and Instrumental Techniques*. 2nd Edn., United Kingdom: Blackwell Publishing. ISBN: 0-631-23269-9

- Mitra, S.K., 2006. Digital Signal Processing: A Computer-based Approach. 3rd Edn., Mc Graw Hill, Singapore.
- Rabiner, L. and B.H. Juang, 1993. Fundamentals of Speech Recognition. Prentice Hall, Englewood Cliffs, New Jersey. ISBN: 0-13-015157-2.
- Smith, S.W., 1997. The scientist and engineer's guide to digital signal processing. California California Technical Publishing. Chapter 22.
- Silverman, H.F. and Y.-T. Lee, 1987. On the spectrographic representation of rapidly time-varying speech. *Comput. Speech Language*. 2: 63-86. DOI: 10.1016/0885-2308(87)90001-5
- Zue, V.W. and L.F. Lamel, 1986. An expert spectrogram reader: A knowledge-based approach to speech recognition. acoustic, speech, and signal processing. Proceedings of the IEEE International Conference on ICASSP'86. Tokyo, Japan, Massachusetts Institute of Technology, Cambridge, MA, pp: 1197-1200. DOI: 10.1109/ICASSP.1986.1168798
- Zue, V.W. and R.A. Cole, 1979. Experiments on spectrogram reading. acoustic, speech, and signal processing. Proceeding of the IEEE International Conference on (ICASSP'79), pp: 116-119. DOI: 10.1109/ICASSP.1979.1170735