

Frontal Arabic Fricative Consonants Characteristics among Primary School Children: Spectral Density Function Approach

Nurul Ashikin Abdul-Kadir and Rubita Sudirman
Department of Electronics Engineering, Faculty of Electrical Engineering,
University Technology Malaysia, 81310 UTM Johor Bharu, Johor, Malaysia

Abstract: Problem statement: Arabic speech recognition system is un-developed until today. By using people of non-Arab nationality, the set of frontal fricative of Arabic consonants are study as the earlier step to develop the recognition system for Malaysian. As most of the developed Arabic speech recognition system for Malaysian are rarely found. **Approach:** The characteristic of each of the phonemes were study based on its spectral density function. The considerations were the peak and bandwidth of each consonant speech samples which appear in the spectrum. The total of each character was averaged to get the reliable reference value which will be the baseline for the selected database later. **Results:** The results showed that the bandwidth consideration of the first peak gave the increment value as the consonants distributed from outer part (labiodentals-3.6 kHz) to inner part of the mouth (alveolar-3.8 kHz). While the peak appearance is lowered as the group of the consonants moved further towards into the mouth. **Conclusion/Recommendations:** The values obtained were used as the reference of the database for our recognition system. Only suitable consonants samples were chosen to be stored as database to improve the system accuracy.

Key words: Arabic consonant, spectral density, place of articulation, International Phonetic Alphabet (IPA), Fast Fourier Transform (FFT), alveolar fricative, spectral peak

INTRODUCTION

In Arabic speech, fricative is the major group which composed of 13 consonants as followed the International Phonetic Alphabet (IPA) system (http://en.wikipedia.org/wiki/Wikipedia:IPA_for_Arab). Fricative, the act of producing sound by passing air through a narrow constriction that causes it flow turbulently with noisy sound, composed of eight places of articulation for thirteen consonants, but only seven consonants are taken into consideration in this study. Figure 1 shows the tabulated of places of articulation while Table 1 is the list of the consonants.

As seen in Fig. 1, dental, alveolar, postalveolar and palatal are located about the same range. Eventhough palatal used tongue as the medium to produce the correct sound, but the way of pronunciation is located in the middle of the oral tract which same as alveolar and postalveolar. Thus the correct pronunciation is just about the same which not easily heard as correct or not. Therefore, the speech signal are transform into the spectral in order to read the behavioral of each sample of different places of articulation especially the frontal consonants.

The common method to distinguish the speech samples are spectrogram reading which proved to be the powerful method to obtain the formants of each samples and had been used in long time ago in speech processing until now (Zue and Cole, 1979; Johannsen *et al.*, 1983; Hassan, 1982; Zue and Lamel, 1986; Debyeche *et al.*, 1998; Johnson, 2003; Awadalla *et al.*, 2005; Awais *et al.*, 2006; Steinberg and O'Shaughnessy, 2008; Iqbal *et al.*, 2008; Abdul-Kadir *et al.*, 2010, Mporas *et al.*, 2007, Sumithra *et al.*, 2011, Abdul-Kadir *et al.*, 2011a,b).

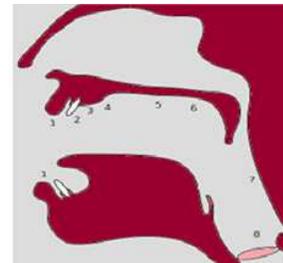


Fig. 1: The place of articulation: (1-2) Labiodental, (2) Dental, (3) Alveolar, (4) Post-alveolar, (5) Palatal, (6) Velar, (7) Pharyngeal, (8) Glottis

Corresponding Author: Rubita Sudirman, Department of Electronics Engineering, Faculty of Electrical Engineering, University Technology Malaysia, 81310 UTM Johor Bharu, Johor, Malaysia Tel: +6075535270
Fax: +6075566272

Table 1: The Consonants

		Cavity					
		Front			Back		
Labio	Dental	Alveolar	Post-alveolar	Palatal	Velar	Pharynx	Glottis
dental					geal		
[ف]	[ث]	[س]	[ش]	[ظ]/zo/	[خ]	[ح]	ه]
/fa/	/tsa/	/sin/	/shin/	/ð/	/kho/	/haa/	/ha/
/f/	/θ/	/s/	/ʃ/	/x/	/hh/	/h/	
-	[ذ]	[ذ]	-	-	/ [غ]	[ع]	-
	/zal/	/zai/		gheyn/	/ʕ/ /ain/		
	/ð*/	/z/			/x/		

As an alternative, power spectral density of each phoneme is computed and the graphical illustration of the spectrum is study. The peak, bandwidth and trough are determined and the average of each is set as the reference. The amplitude of spectral peak decrease as the place of articulation moves further forward into the mouth while the bandwidth becomes wider. In addition, the tilt of the spectrum at the beginning also helps the identification of the phoneme. Furthermore, Li (2009) found that English speaking children generally produced alveolar fricatives more accurately than post-alveolar, whereas the opposite was true for Japanese-speaking children. By using non-native children as the subject, this study study about the characteristics of frontal fricative consonants through power spectral density. Further observation is discussed in section 3.

Previous work: Shadle *et al.* (1995) also done the spectral analysis of [x] but using German speakers. The peaks are reported appeared at 2 and 3.8 kHz while trough is around 3 kHz. While Beautemps *et al.* (1995) had discover four peaks (600, 1211, 2180 and 3665 Hz) in the study of French speaker pronouncing [x]. Shadle and Mair (1996) had distinguished spectral parameters that are reliable to model a fricative corpus.

Phillips (2000) found the spectral peaks of fricative /s/ are 1, 1.4, 1.8 and 26 kHz in his study of characteristics of plosive, fricative and aspiration components in speech production.

Phillips (2001) had used the spectrum of plosive and obtains the spectral characteristics of the burst. The differentiation between voiced and unvoiced sound are discussed as stated that voiced sound composed of less aspiration.

Jesus and Shadle (2005) observed that the back place of articulation possess a high amplitude peaks. As in his study, [s] which is situated in the front cavity had a broad peak at 8 kHz, [ʃ] in the centered had peak at 3.5 kHz while [χ] with longer front cavity is having a series of peaks at 1.3 and 2.4 kHz. The peaks are found to be distributed along the range of 1-1.6 kHz, 2.1-2.8 kHz and 3.2-4 kHz. The obvious trough always appears along 1.8-2.5 kHz.

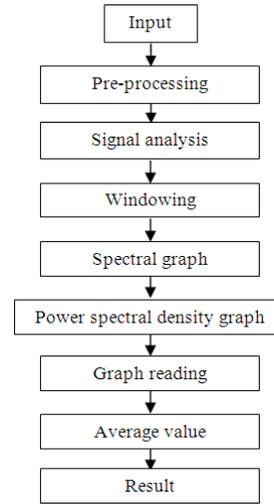


Fig. 2: The workflow to obtain the average value of spectral reading

Jesus and Jackson (2008) had track the value of F₀ of fricatives of British English and European Portuguese. Less than 300 Hz is the average value acquired.

Johnson (2009) had model the speech signal as he stated that each peak in the spectral form of each speech signal is the representation of the formant.

METERIALS AND METHODS

Experiment settings: This study involved children age 8-11 years old as the subject with a total number of 25. There are 15 girls and 10 boys of Malay native children (non-Arab nationality) who practiced Arabic in daily life only to recite Quran. Other specifications are as follows:

- Recording software: Easy Hi-Q Recorder with sampling rate of 44.1 kHz
- Analysis Software: Goldwave, Matlab 2008a
- Recording type and format: *.wav, 16-bit mono
- Recording device: External mic

Recording session is done in a room with normal environment. Each one of those subjects spoke each of the Arabic frontal fricative consonants in a record. With the specification obeyed, the total of all records of 175 phonemes signal are extracted and analyzed.

Processing step: The FFT method of windowing size of 8 is used as the processing method. The spectral of each of the samples are obtained in MatLAB environment. Figure 2 shows the overall workflows to get the power spectral density graph. While on the other hand, there are four steps to be concerned to get the reading of the spectral. First of all is to short-segmented the audio signal. Second, number of data points must be defined. Next is to define the length of window to use the FFT method in order to find the frequency spectrum of each segment. Last, power spectral density is plotted and analysed.

RESULTS AND DISCUSSION

The first part of this section discussed about voiced alveolar /s/ and voiced post-alveolar /ʃ/ while the second part discuss of other consonants.

Alveolar /s/ and Post-Alveolar /ʃ/: After all of the data had been collected, the spectrum representation of each speech samples is computed using Matlab software with specific algorithm. The power spectral density for every consonant are obtained as in Fig. 3 which shows a spectrum of one of /s/, [س]. The pattern of all spectrums is study. The peak(s), tilt and bandwidth are gathered and the reliable average values are listed in Table 2. As the representation of the spectrum is in Fast Fourier Transform (FFT) window, thus the results are then changed to Hertz form. The overall window is 2^7 , n for sampling frequency, F_s of 44.1 kHz. The Hertz value (f) is represented in Table 3 as computed using Eq. 1:

$$f = \frac{n-1}{129-1}(f_s) \tag{1}$$

In Fig. 3, the spectrum is rising from n = 0 until n = 3, the peak and drop drastically to n = 6. The bandwidth is measured from n = 1 to n = 6. No tilt since the peak only appeared once. Thus the peak appeared at 541.4 and bandwidth is 725.3 Hz.

As seen in Table 2 and 3, the spectral reading are then plotted as graph representation as in Fig. 4. Consonants /s/ and /ʃ/ which are located at alveolar and post-alveolar respectively composed of first peak at 541.4 and 725.3 Hz correspondingly.

Table 2: The spectrum reading from FFT window

Phoneme	1st Peak (n)	Bandwidth (n)
/s/ ,[س]	2.57	5.9
/ʃ/ ,[ش]	3.11	6.37

Table 3: The spectrum frequency of fricative consonants

Place of articulation	Phoneme	1st Peak (Hz)	Bandwidth (Hz)
Alveolar	/s/ ,[س]	541.4	1689.8
Post-alveolar	/ʃ/ ,[ش]	725.3	1849.6

While the bandwidth for both consonants are 1689.8 and 1849.6 Hz respectively. Bandwidth became wider as the place of articulation goes into the oral tract. For both consonants, the spectrum power is decreasing drastically after at nearly 2 kHz.

According to theory, post-alveolar consonant which is situated inner into the mouth than alveolar should possess peak appearance at frequency lower for the same peak number (first peak with first peak) While According to Li (2009), English speaker generally pronounced alveolar fricative accurately than post-alveolar while Japanese-speaking children pronounced post-alveolar better than alveolar. As can be seen in Fig. 4, the first peak of alveolar is less than post-alveolar. This might happen as the Malaysian-speaking children have difficulty in produced voiced post-alveolar sound as no fricative post-alveolar sound in the language. Both peak appeared at frequency lower than 1 kHz and the spectrum decrease drastically as in Fig. 3. The bandwidth is greater as the place of constriction moves from alveolar to post-alveolar.

Labiodental /f/, Dental /θ, ð/, Alveolar /z/ and Palatal /ðʰ/: This part (B) discussed about the remaining five frontal fricative consonants under study which are labiodental /f/, dental /θ, ð/, alveolar /z/ and palatal /ðʰ/. The spectrum of power spectral density is illustrated as in Fig. 5 and 6 of consonants /f/ and /θ/ respectively. Both have more than a peak in each spectrum representation of each samples of same length duration. In Fig. 5, two peaks appeared at n = 6 and n = 10 which are 1722.7 and 3100.8 Hz respectively. While in Fig. 6, three peaks appeared at n=3, 6 and 11 which correspond to 689.1, 1722.7 and 3445.3 Hz and the slope at n=5 which is 1378.1 Hz. The spectrums of all samples are obtained and the results are averaged as summarized in Table 4. Table 5 is the interpretation of those samples in frequencies as computed using Eq. 1.

Most of these five consonants have two peaks with /θ/ and /z/ has three. The third peak always appear at f = 3937.5 and 3962.1 Hz for both consonants. As the place of constriction moves from labio-dental to palatal, the first peak is decrease as in Fig. 7.

Table 4: The spectrum reading from FFT window

Phoneme	1st Peak (n)	2nd Peak (n)	3rd Peak (n)	Trough (n)	Bandwidth (n)
/f/ ,[ف]	5.94	9.67	-	7	11.5
/θ/ ,[ث]	4.06	8.44	12.43	6.22	11.35
/ð/ ,[ذ]	4.6	8.5	-	7.33	11.95
/z/ ,[ز]	4.29	10	12.5	9.5	12.14
/ʔð/ ,[ظ]	4.26	7.5	-	6.8	9.83

Table 5: The spectrum frequency of fricative consonants

Place of articulation	Phoneme	1st Peak (Hz)	2nd Peak (Hz)	3rd Peak (Hz)	Trough (Hz)	Bandwidth (Hz)
Labiodental	/f/ ,[ف]	1701.1	2985.9	-	2067.2	3617.6
Dental	/θ/ ,[ث]	1053.9	2562.5	3937.5	1799.2	3566.9
	/ð/ ,[ذ]	1240.3	2584	-	2182	3772.6
Alveolar	/z/ ,[ز]	1132	3100.8	3962.1	2928.5	3839.1
Palatal	/ʔð/ ,[ظ]	1123.5	2239.5	-	1998.3	3040.9

Table 6: Comparison of previous work

Study	Speakers	Peak (kHz)	Peak range (kHz)	Trough (kHz)
Shadle <i>et al.</i> (1995)	German	[x] at 2 and 3.8	-	3
Beautemps <i>et al.</i> (1995)	French	[x] at 0.6, 1.211, 2.18 and 3.665	-	-
Jackson (2000)	-	/s/ at 1, 1.4, 1.8 and 26	-	-
Jesus and Shadle (2005)	-	[s] at 8 [ʃ] at 3.5 [χ] at 1.3 and 2.4	1 - 1.6 2.1 - 2.8 3.2 - 4	1.8 - 2.5
Jesus and Jackson (2008)	British english and European portuguese	F ₀ of fricatives < 0.3	-	-
This study	Malaysian	/f/ at 1.7 and 3.0 /θ/ at 1.1, 2.6 and 3.9 /ð/ at 1.2 and 2.6 /s/ at 0.5 /z/ at 1.1, 3.1 and 4 /ʃ/ at 0.7 /ðʔ/ 1.1 and 2.2	-	/f/ at 3.6 /θ/ at 3.6 /ð/ at 3.8 /s/ at 1.7 /z/ at 3.8 /ʃ/ at 1.8 /ðʔ/ at 3

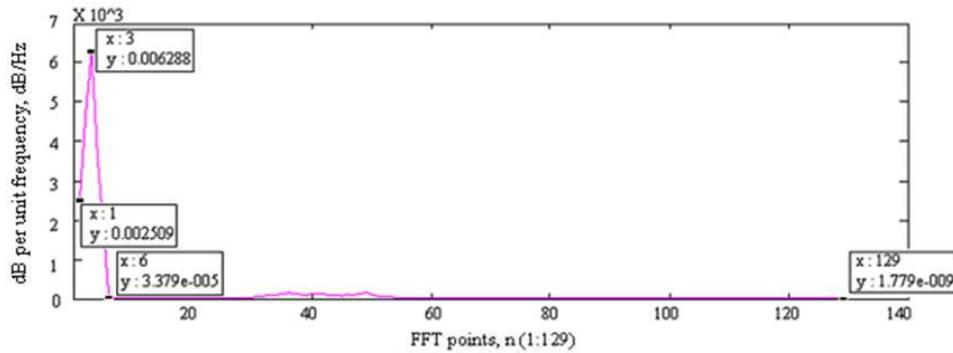


Fig. 3: Power spectral density for /s/ , [س]

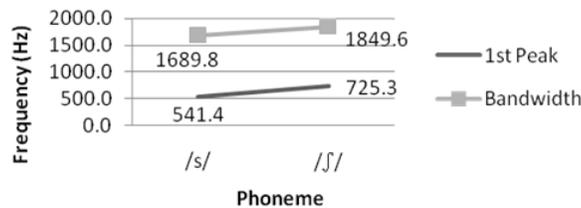


Fig. 4: Bandwidth and 1st peak value for alveolar /s/ and post-alveolar /ʃ/

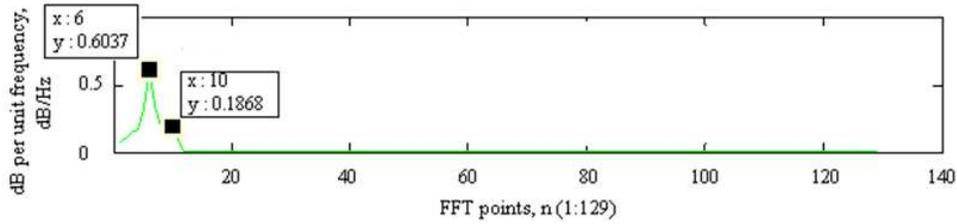


Fig. 5: Power spectral density for /f/, [ف]

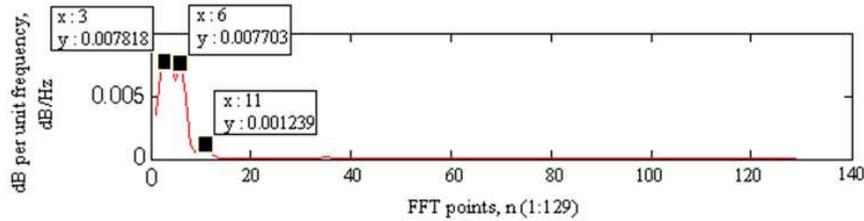


Fig. 6: Power spectral density for /θ/, [ث]

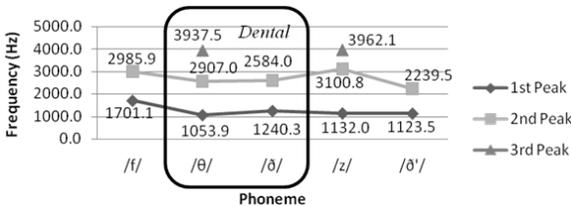


Fig. 7: Power spectral density for labiodental /f/, dental /θ,ð/, alveolar /z/ and palatal /ð'/

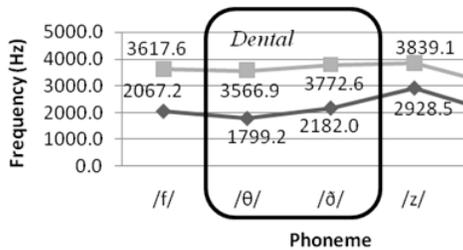


Fig. 8: Power spectral density for labiodental /f/, dental /θ,ð/, alveolar /z/ and palatal /ð'/

While the second and third peak of each consonants are greater than previous peak. Trough for [ف], /f/, [ث], /θ/, [ð], /ð/, [ز], /z/ and [ذ], /ð'/ appear at 2067.2, 1799.2, 2182.0, 2928.5 and 1998.3 Hz respectively while bandwidth is 3617.6, 3566.9, 3772.6, 3839.1 and 3040.9 Hz respectively as in Fig. 8. Bandwidth is increasing from labio-dental to alveolar but not increase as the place is at palatal. This may happen as proven by Li (2009) that children will have difficulty at producing the correct

sound of consonant of non-native language. Comparison with previous findings is summarized as in Table 6.

CONCLUSION

The second and the third peaks always appeared greater than the first peak for each consonant of each place of articulation. The first peak always appeared with high amplitude peak of all peaks composed in each sample and drops drastically to produce slope, which is known as trough and increase again to produce the second peak and repeated until the last peak appeared with decrement in the amplitude as can be seen through the spectrum representation discuss previously. From the power spectral density representation, the characteristic of fricatives is easily to be known than spectrogram representation as the fricatives sound compose of hissing and aspiration more which make the formants not easily be seen and read from normal point of view. Furthermore, the spectral density function of each samples are readable of formants which represent by each peak.

ACKNOWLEDGMENT

The researchers are so indebted and would like to express thankfulness to the University Technology Malaysia and Ministry of Science and Innovation Malaysia for supporting and funding this study under Vote 79368. Our appreciation also goes to the Biomedical Electronics Research Group for their ideas and comments to this study.

REFERENCES

- Abdul-Kadir, N.A., R. Sudirman and N.M. Safri, 2010. Modelling of the Arabic plosive consonants characteristics based on spectrogram. Proceeding of the 4th Asia International Conference on Mathematical/Analytical Modelling and Computer Simulation, May 26-28, IEEE Xplore Press, Kota Kinabalu, Malaysia, pp: 282-285. DOI: 10.1109/AMS.2010.63
- Abdul-Kadir, N.A., R. Sudirman, 2011a. Vowel effects towards dental arabic consonants based on spectrogram. Proceeding of 2nd International Conference on Intelligent Systems Modelling and Simulation, Jan. 25-27, IEEE Xplore Press, Kuala Lumpur, pp: 183-188. DOI: 10.1109/ISMS.2011.38
- Abdul-Kadir, N.A., R. Sudirman, 2011b. Difficulties of standard arabic phonemes spoken by non-arab primary school children based on formant frequencies. *J. Comput. Sci.*, 7: 1003-1010. DOI: 10.3844/jcssp.2011.1003.1010
- Awadalla, M., F.E.Z. Abou Chadi and H.H. Soliman, 2005. Development of an arabic speech database. Proceeding of the ITI 3rd International Conference on Information and Communications Technology Enabling Technologies for the new Knowledge Society, Dec. 5-6, IEEE Xplore Press, Cairo, pp: 89-100. DOI: 10.1109/ITICT.2005.1609617
- Awais, M.M., W. Ahmad, S. Masud and S. Shamail, 2006. Continuous Arabic speech segmentation using FFT spectrogram. Proceeding of the Innovations in Information Technology, Nov. 19-26, IEEE Xplore Press, Dubai, pp: 1-6. DOI: 10.1109/INNOVATIONS.2006.301939
- Beautemps, D., P. Badin and R. Laboissire, 1995. Deriving vocal-tract area functions from midsagittal profiles and formant frequencies: A new model for vowels and fricative consonants based on experimental data. *Speech Commun.*, 16: 27-47. DOI: 10.1016/0167-6393(94)00045-C
- Debyeche, M., A. Houacine and J.P. Haton, 1998. A knowledge-based approach for arabic emphatic consonant identification based on speech spectrogram reading. Proceeding of the 13th Southeastern Symposium System Theory, Mar. 8-10, IEEE Xplore Press, Morgantown, WV, USA, pp: 325-328. DOI: 10.1109/SSST.1998.660085
- Hassan, A., 1982. *Linguistik Am Untuk Guru Bahasa Malaysia*. Wikipedia: IPA for Arabic.
- Iqbal, H.R., M.M. Awais, S. Masud and S. Shamail, 2008. New challenges in applied intelligence technology. *Studies Comput. Intell.*, 134: 1-401. DOI: 10.1007/978-3-540-79355-7
- Jesus, L.M.T. and C.H. Shadle, 2005. Acoustic analysis of European Portuguese uvular [ʔ] and voiceless tapped alveolar fricatives. *J. Int. Phonetic Assoc.*, 35: 27-44. DOI: 10.1017/S0025100305001866
- Jesus, L.M.T. and P.J.B. Jackson, 2008. Frication and voicing classification. computational processing of the portuguese language. *Lecture Notes Comp. Sci.*, 5190: 11-20. DOI: 10.1007/978-3-540-85980-2_2
- Johannsen, J., J. MacAllister, T. Michalek and S. Ross, 1983. A speech spectrogram expert. 8: 746-749. DOI: 10.1109/ICASSP.1983.1172057
- Johnson, D., 2009. *Modeling the Speech Signal*. Connexions.
- Johnson, K., 2003. *Acoustic and Auditory Phonetics*. 3rd Edn., Blackwell Publishing, United Kingdom, ISBN1444343076, pp: 232.
- Li, F., 2009. Contrast and covert contrast: The phonetic development of voiceless sibilant fricative in English and Japanese Toddlers. *J. Phonetics*, 37: 111-124. DOI: 10.1016/j.wocn.2008.10.001
- Mporas, I., T. Ganchev, M. Siafarikas and N. Fakotakis, 2007. Comparison of speech features on the speech recognition task. *J. Comput. Sci.*, 3: 608-616. DOI: 10.3844/jcssp.2007.608.616
- Phillips, J.B.J., 2000. Characterization of plosive, fricative and aspiration components in speech production. Ph.D. thesis, University of Southampton.
- Phillips, J.B.J., 2001. Acoustic cues of voiced and voiceless plosives for determining place of articulation. University of Birmingham.
- Shadle, C.H. and S.J. Mair, 1996. Quantifying spectral characteristics of fricatives. Proceeding of the 4th International Conference on Spoken Language, Oct. 3-6, IEEE Xplore Press, Philadelphia, PA, USA, pp: 1517-1520. DOI: 10.1109/ICSLP.1996.607906
- Shadle, C.H., S.J. Mair, J.N. Carter and N. Millner, 1995. The effect of vowel context on acoustic characteristics of [x]. University of Southampton.
- Steinberg, R. and D. O'Shaughnessy, 2008. Segmentation of a speech spectrogram using mathematical morphology. *Acoustics*. Proceeding of the IEEE International Conference Acoustics Speech and Signal Processing, Mar. 31-4, IEEE Xplore Press, Las Vegas, NV, pp: 1637-1640. DOI: 10.1109/ICASSP.2008.4517940
- Sumithra, M.G., K. Thanuskodi and A.H.J. Archana, 2011. A new speaker recognition system with combined feature extraction techniques. *J. Comput. Sci.*, 7: 459-465. DOI: 10.3844/jcssp.2011.459.465

Zue, V.W. and L.F. Lamel, 1986. An expert spectrogram reader: a knowledge-based approach to speech recognition. Proceeding of the IEEE International Conference on Acoustics, Speech, and Signal Processing, (ICASSP '86), IEEE Xplore Press, USA, pp: 1197-1200. DOI: 10.1109/ICASSP.1986.1168798.

Zue, V.W. and R.A. Cole, 1979. Experiments on spectrogram reading. Proceeding of the IEEE International Conference on Acoustics, Speech, and Signal Processing, (ICASSP '79), IEEE Xplore Press, USA, pp: 116-119. DOI: 10.1109/ICASSP.1979.1170735