An Acoustic Investigation of Pakistani and American English Vowels

Abdul Malik Abbasi[1], Mansoor Ahmed Channa[2], Stephen John[3], Masood Akhter Memon[4] & Rabbia Anwar Ahmed[5]

[1] Department of Allied Subjects, Sindh Madressatul Islam University, Karachi, Pakistan

[2] Department of Basic Sciences, Quaid-e-Awam University, Nawab Shah Sindh, Pakistan

[3] Department of Education, Sindh Madressatul Islam University, Karachi, Pakistan

[4] Department of English, Cadet College Petaro, Hyderabad, Sindh

[5] Department of Computer Sciences, Sindh Madressatul Islam University, Karachi, Pakistan

Correspondence: Abdul Malik Abbasi, Department of Allied Subjects, Sindh Madressatul Islam University, Karachi, Pakistan. E-mail: amabbasi@smiu.edu.pk

Abstract

Acoustic analysis tests the hypothesis that the physical properties of Pakistani English (PaKE) vowels differ in terms of acoustic measurements of Native American English speakers. The present paper aims to document the physical behavior of English vowels produced by PaKE learners. The major goal of this paper is to measure the production of sound frequencies coupled with vowel duration. The primary aim of this paper is to explore the different frequencies and duration of the vowels involved in articulation of PaKE. English vowels selected for this purpose are: /æ/, /ɛ/, /ɪ/, /ɒ/ and /ə/. Total ten samplings were obtained from the department of computer science at Sindh Madressatul Islam University, Karachi. The study was based on the analysis of 500 (10×5×10=500) voice samples. Five vowel minimal pairs were selected and written in a carrier phrase [I say CVC now]. Ten speakers (5 male & five female) recorded their 500 voice samples using Praat speech processing tool and a high-quality microphone on laptop in a computer laboratory with no background sound. Three parameters were considered for the analysis of PaKE vowels i.e., duration of five vowels, fundamental frequency (F1 and F2). It was hypothesized that the properties of PaKE vowels are different from that of English native speakers. The hypothesis was accepted since the acoustic measurements of PaKE and English Native American speakers' physical properties of sounds were discovered differently.

Keywords: acoustic, analysis, Pakistani English, fundamental frequency and duration

1. Introduction

The study aims to analyze the sound properties of English vowels produced by English language learners. The present paper examines three basic parameters of human speech, the fundamental frequencies F1 and F2 and the duration of English vowels determined by the data. This experiment is based on 500 voice samples of PaKE learners with the hypothesis that PaKE differs in terms of its acoustic measurements of English vowels as compared to English native speakers' production. The primary aim of this paper is to explore the different frequencies and duration of the vowels involved in articulation of Pakistani English. English vowels were selected as follows: /æ/, / ɛ /, /ɪ/, /ɒ/ and /ə/. The study was based on the analysis of 500 (10×5×10=500) voice samples. Five vowel minimal pairs were selected and written in a carrier phrase. Ten speakers (5 male & 5 female) recorded their 500 voice samples using *Praat Speech Processing* tool and a high-quality microphone on laptop in a computer laboratory with no background sound. Three parameters were considered for the production analysis of English vowels i.e., duration, fundamental frequency (F1 and F2). Vowel duration is the length of the vowel whereas F1 shows the height of vowel and F2 shows the back-ness of vowel in mouth cavity.

2. Literature Review

Connor (1980) states that English language contains 24 consonantal sounds, 20 vowels (12 Mono-thongs and 8 diphthong). English is rich in its vocalic system, coupled with many variations across English language speaking countries. It is an international language in Pakistan where the undergraduate students in universities study most of their major courses in English. Since English is a second language of every Pakistani student, there will be a

difference in pronunciation of native English speakers vs. PaKE speakers. Languages have different vowel systems; therefore, they differ in pronouncing the same words because of different vowel systems available their languages. The current study presents evidence that PaKE is different from native American English in terms of English vowels production. Additionally, another phonetic factor is the mother tongue influence, which is unconscious learning which influences the production of second language speakers. This study investigates the English vowels produced by Pakistani undergraduate students. This paper documents the frequencies and duration of the vowels produced by Pakistani undergraduate students. The purpose of the study was to determine the acoustic differences of English vowels produced by PaKE speakers. Pakistani speakers of English do not perceive the differences of vowels which cause misunderstanding amongst the native Americans. This study will benefit those speakers who make such errors on the production of English vowels. In this context, the study will add some knowledge to the existing literature. Hillenbrand et al. (1995) analyzed American native speech as illustrated in Table 1. Average duration, fundamental frequencies, and formant frequencies of vowels are produced by 45 men, 48 women, and 46 children. Averages are on a subset of the tokens that were well identified by listeners. The duration measurements were analyzed in (ms) all others were in Hz.

Table 1. Average duration, fundamental frequencies & formant frequencies of American vowels

		/i/	/ɪ/	/e/	/ɛ/	/æ/	/ɑ/	/ɔ/	/o/	/ʊ/	/u/	/ʌ/	/ɜ/
Dur	M	243	192	267	189	278	267	283	265	192	237	188	263
	W	306	237	320	254	332	323	353	326	249	303	226	321
	C	297	248	314	235	322	311	319	310	247	278	234	307
F0	M	138	135	129	127	123	123	121	129	133	143	133	130
	W	227	224	219	214	215	215	210	217	230	235	218	217
	C	246	241	237	230	228	229	225	236	243	249	236	237
F1	M	342	427	476	580	588	768	652	497	469	378	623	474
	W	437	483	536	731	669	936	781	555	519	459	753	523
	C	452	511	564	749	717	1002	803	597	568	494	749	586
F2	M	2322	2034	2089	1799	1952	1333	997	910	1122	997	1200	1379
	W	2761	2365	2530	2058	2349	1551	1136	1035	1225	1105	1426	1588
	C	3081	2552	2656	2267	2501	1688	1210	1137	1490	1345	1546	1719
F3	M	3000	2684	2691	2605	2601	2522	2538	2459	2434	2343	2550	1710
	W	3372	3053	3047	2979	2972	2815	2824	2828	2827	2735	2933	1929
	C	3702	3403	3323	3310	3289	2950	2982	2987	3072	2988	3145	2143
F4	M	3657	3618	3649	3677	3624	3687	3486	3384	3400	3357	3557	3334
	W	4352	4334	4319	4294	4290	4299	3923	3927	4052	4115	4092	3914
	C	4572	4575	4422	4671	4409	4307	3919	4167	4328	4276	4320	3788

Sheikh (2012) finds that PaKE is the closest to British RP English accent as compared to other English accents and further states that the PaKE is now considered an independent substance since it reconstructs its own English sound system. The basic articulator characteristics of F1 and F2 formants are: the displacement of the body of the tongue in the mouth (height and back) and the roundness of the lips (Ladefoged, 1993; Pfitzinger, 2003). First two formant frequencies are enough for the recognition of the speech (Parsons, 1987). For male adult speakers, the range of the frequencies is described for example, "F1 ranges between 200-800 Hz, F2 ranges between 600-2800 Hz, and F3 ranges between 1300- 3400 Hz" (Parsons, 1987). The mouth resonance frequency and throat vary due to different sizes of vocal tracts of different people from each other. The vocal tract is approximately 17 cm long for adult male speakers (Parsons, 1987).

The vowel sounds of PaKE i.e., /æ/ was analyzed as an unrounded below the half-open position front, vowel as in 'map' /mæp /. Short vowel / e /, is an unrounded front between half-close and half-open vowel. Short vowel /ɪ/ a centralized unrounded half-close front vowel e.g., 'sit' / sɪt /. / ɒ / is a rounded back vowel just above the open position. e.g., 'rod' /rɒd /. /ʌ/ was analyzed as an unrounded central vowel between open and half-open (Sabir & Al-Saeed, 2014). There are some vowels which are articulated exactly like the English vowels. These are /i/, /ɪ/, /u/, /ʊ/, /ɔ/, / ɒ/ and /ʌ/ in PaKE (Sheikh, 2012).

Abbasi & Hussain (2015) argue that 'overall mean of English vowels duration was 75 milliseconds for stressed and 66 milliseconds for unstressed. The mean difference between stressed and unstressed was 40 milliseconds for long vowels and 9 milliseconds for English vowels in Sindhi'. Abbasi (2017) further contends that the mean F1 and F2 values were higher for stressed long vowels and English vowels, while F1 and F2 values were lower for unstressed long and English vowels. Overall average of F1 and F2 values were statistically significant

difference within the group and between the groups'. Abbasi & Hussain (2015) note that 'the evidence presented argues for analyzing Indo-Aryan language in which intonation contours appear to be independent of stress. This finding is interesting since it suggests that stress is completely orthogonal to F0 contours unlike in most stress languages in which pitch accents dock on stressed syllables. Sindhi is an Indo-Aryan language whose pitch accent rises from the first syllable in disyllable words, irrespective of syllable weight, and the rise is followed by a fall at end of the word'. In addition, the study concludes that Sindhi behaves like a stress accent language as discussed in Beckman (1986).

3. Methods and Procedures

3.1 Speakers

Participating speakers were recruited from Sindh Madressatul Islam University, Karachi. The speakers were Urdu native speakers; their ages were between 18-25 years old. A total ten of Urdu speakers recorded their voice samples. They were five male and five female speakers. Entire samplings were taken from computer science department from the population of Sindh Madressatul Islam University, Karachi. There was no speech related problem with the subjects as was self-reported.

3.2 Recordings

Five hundred voice samples of 10 ESL (English as Second Language) learners were recorded and analyzed on *Praat Speech Processing Tool*. Three parameters were measured i.e., F1, F2 and the duration of English vowels. For recording process, the drop down, menu 'Record mono Sound' of 'New' option from 'object' window of Praat was used. For this purpose, a list of words was read three times each in a carrier sentence, to get the perfect recording. The recorded file through Praat was used 'save as WAV file' option. Later the Edit & View option was clicked where spectrograph of the token sound appeared. Token sound was selected and went to the menu of first formant frequency where the F1 option pops up which was entered where F1 frequencies appeared. In this way, F2 frequency was also examined. Duration of token vowel was taken manually by selecting horizontal length from right to left or left to right of the token vowel viewing dark bands of the energy carefully.

3.3 Procedure of Data Collection

The experiment recorded 500 voice samples of non-nasal /æ/, /e/, /ɪ/, /ɒ/ and /ə/. Monosyllabic word tokens were selected for the recording purpose. The words were CVC, a pattern that precedes vowel and followed by consonant (Consonant-Vowel-Consonant). Fifty words were taken for the sampling process, which contains ten words for each of five vowels. Later two words were removed from the study since the token words did not match the required phonetic parameter. These tokens were selected through the misleading orthography of English. The speakers were recorded in a silent zone with no background noise. They were seated six inches away from a microphone that was head-mounted. They were provided a list of monosyllabic English words that they were asked to read. The list included the CVC pattern which contains all the target English vowels. These words were placed in a sentence. For recording purpose, the carrier sentence 'I say [CVC] now' was used. The carrier sentence allowed the speakers to say carrier phrase with a constant rate. In each recording, the same sentence was repeated by the speaker three times for each English CVC vowel pattern. For such work, to record the possible best production, the speaker was highly encouraged. If there were any possible errors then the whole procedure would be performed again, to record the best sound. The list of CVC pattern can be found in appendix A at the end. The samples were recorded on 'Praat' software. The words were selected mono-syllable because mono-syllable words are easy and understandable to the ESL learners. The words were recorded in isolation in the same position in a sentence to avoid any variations in frequencies.

Table 2. Stimuli for the production of English words by ESL learners

Token words	Token words	Token words	Token words	Token words
æ	ɛ	ɪ	ɒ	ə
Bat	Bet	Big	Cot	Cut
Cat	Bed	Bit	Dog	Hug
Fat	Get	Fit	Got	Hut
Fan	Jet	Hit	Mom	Jug
Gas	Led	Kid	Lot	Nut
Map	Leg	Lip	Pot	Bus
Ran	Let	Sid	Rod	Bud
Sad	Pen	Tin	~~Son~~	But
Van	Set	Tip	Top	Cup
Hat	Pet	Win	~~Won~~	Fun

3.4 Data Analysis

For Data Analysis process the option 'View and Edit' is used from object window. The window pops up on a screen. On this, the recorded sound is shown as a *triangular wave form*. By using 'tab' button (i.e., from key board), it allows the user to play the recorded sound. Once the entire phrase is repeated, then select that part and use short key Ctrl + N to zoom in the selected recording whereas Ctrl + O key is used for zooming out. Repeat the same procedure until the whole three repetitions of the token sounds are done. Praat software figures out the spectrograms of duration and formant frequency etc.

3.5 Duration Measurement

All measurements of duration of vowels were taken manually by visual inspection of wideband spectrographic display on computer screen. Start and end points of the target vowels were measured in milliseconds on the spectrographic displays. Both beginning and ending points were measured using the speech processing tool Praat (Boersma & Weenink, 2017).

3.6 Formants Measurements

All formant measures for the lowest two formants, F1 and F2, were taken manually based on formant tracks at the visually located mid-point of the target vowel in stressed and unstressed tokens. Whenever a mismatch between the tracks and the visually apparent formant band in the spectrogram was detected, the formants were checked by visual inspection of wideband spectrographic display on computer screen. The pitch contours were manually extracted using the Praat autocorrelation method. Measurements were taken at the visually located midpoint of each target vowel in stressed and unstressed tokens.
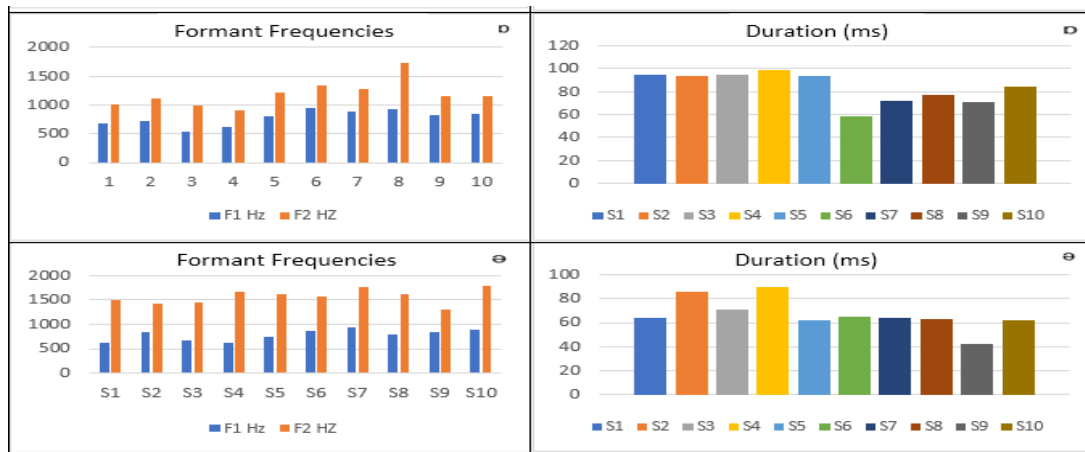
4. Results
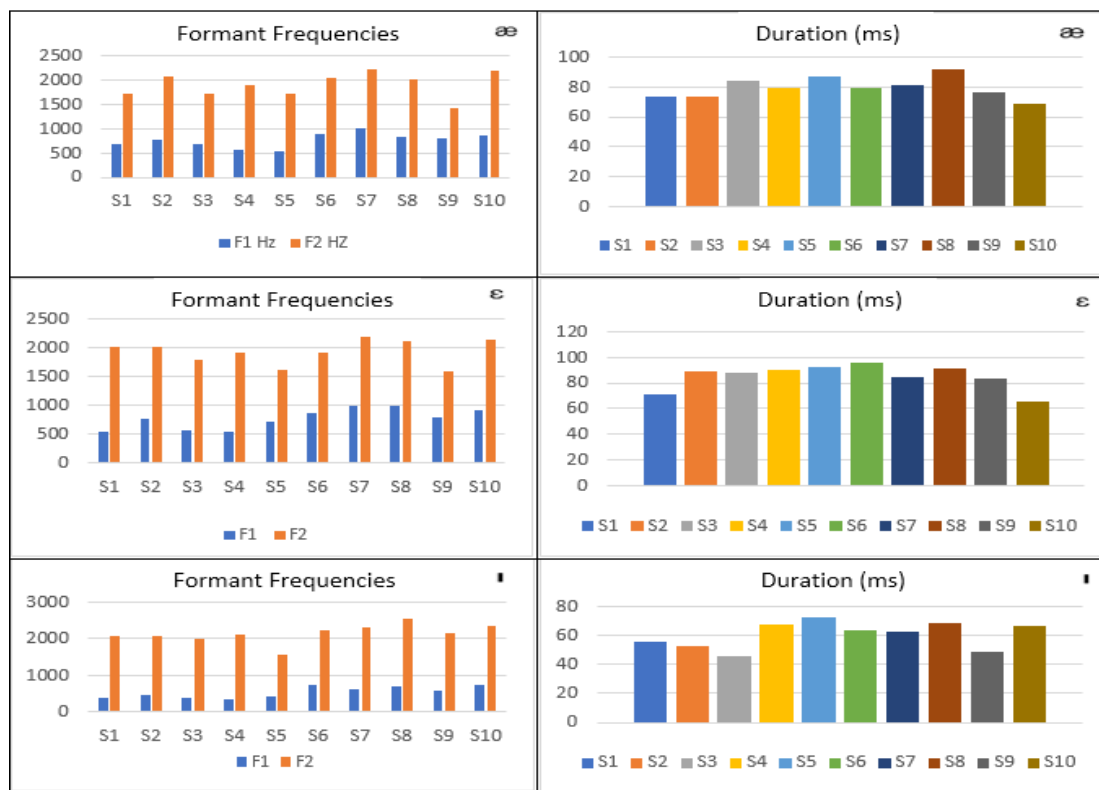
Figure 1. 10 graphs of bar charts of F1, F2, F0 and duration of five English vowels across speakers

4.1 Duration

The duration of English vowels for both male and female speakers is around 60 to 86, measured in milliseconds. The averages are illustrated in Tables 3-4. The duration of English vowels, however, shows variations. The duration of English vowels is between 58 to 95 milliseconds as shown in Table 3, for males and 55 to 85 for females as shown in Table 4. As illustrated in Table 1 and Table 3 all durational values of vowels are found to be similar. The durational values of the PaKE vowels / æ/, /ɜ/ and / ə / are found to be approximately the same; about 79 to 86 milliseconds and the vowels / ɪ / and / ɒ / English vowels have too small duration approximately 60 to 67 millisecond vowels. This difference in length introduces a major quantitative difference. This visible difference (shown in Tables 1-3) is clear in vowels /æ/, /ɛ /, /ɪ/, /ɒ/ and /ə/. Another factor affecting the duration of vowels is the deliverance of speech rate. If the speed of production was fast, the time given to the articulation to reach the targets, necessary to correctly transmit the phonemes would be less necessary. Thus, the durational values of vowels and consonants would decrease (but this decrease is more evident for vowels) (Pickett, 1999).

4.2 Formant frequencies (F1 and F2)

PaKE vowels measurements are illustrated in Table 3 and 4 from the data analysis for five English vowels /æ/, / ɛ /, /ɪ/, /ɒ/ and /ə/. It is clear from the tables that there is a measurable difference between the frequencies of vowels and between F1 and F2 frequencies. Per observation the F1 is highest for /æ/, /ɛ /, /ɪ/, /ɒ/ and /ə/ lowest for /ɪ/ for all participants. However, the average of frequency F1 of /ə/ is highest, /ɒ/ and /æ/ is higher, high for /ɛ/ and low for /ɪ/ in most of the cases as shown in Table 1 and Table 2. The formant frequencies F2 average of all the English vowels are highest for /ɪ/ and higher for /æ/ and /ɛ/, high for /ə/ and low for /ɒ/. The vowels frequency values are illustrated in Table 1 and Table 2 (See Appendix A for more detail).

Table 3. Average of F1, F2 and duration

Five English Vowels Produced by Female Speakers			
Vowel	F1 Hz	F2 Hz	Duration
æ	881.235	1981.757	79.603
ɛ	907.092	1976.217	86.287
ɪ	667.426	2313.993	61.899
ɒ	885.791	1334.832	72.583
ə	867.237	1605.653	59.275

Table 4. Average of F1, F2 and duration

Five English Vowels Produced by Male Speakers			
Vowel	F1 Hz	F2 Hz	Duration
æ	648.813	1825.258	79.669
ɛ	620.49	1872.919	86.339
ɪ	403.177	1955.355	58.777
ɒ	670.995	1047.091	94.974
ə	696.825	1528.503	74.505

5. Discussion

Majority of PaKE language learners develops their own identity on the production of PaKE vowels. PaKE stands on its own as a non-native standard variety. The English vowel sounds produced by 10 speakers of the sample were measured and found the variations in characteristics of Pakistani English, representing as a PaKE speech. The current study discussed about the short vowel production by undergraduate students. The study analyzed five English vowels which are given in Table 2. Table 3 shows the average of the observations of the male speaker values and Table 4 shows the average values of the female speakers. The average durational values of each spoken vowel, the frequencies of formants F1 and F2 are shown in the tables for male speakers and female speakers (See Appendix A for more detail).

This present paper measures the variations in PaKE vowels produced by 10 university ESL learners in Pakistan. The results are based on the evidence analysed and compared. This study was conducted at Sindh Madressatul Islam University, Karachi and collected voice samples from undergraduate students. For this purpose, Praat software was used to analyse the sound recordings. The study discovered variations by the analysis of English vowel frequencies, i.e., F1 and F2 and the durational values of vowel sounds produced by ESL learners of English. Additionally, the study found that five English vowel frequencies of female is higher than males as their pitch was higher. The overall frequency F1 is highest for /ə/ than / ɒ / and /æ/ and lowest for / ɛ / and / ɪ / and for F2 is highest for / ɪ / than / ɛ / and / æ / and lowest for /ə/ and / ɒ / and duration is highest for / ɛ / than / ɒ / and /æ/ and lowest for /ə/ and / ɪ /. It has been noted that F1 for / ɛ / and / ɪ / is almost similar and F2 is almost the same for / ɛ / and /æ/ and duration is approximately the same for / ɛ / and / ɒ /. In most of the words, vowels were not produced as long vocalic sounds as compared to five native English vowels. It has also been noted that English vowels /æ/ sound as 'ah', / ɛ / sound as 'eh', / ɪ / sound as 'ih', / ɒ / sound as 'awe' and /ə/ sound as 'uh'. This study aims to make a significant contribution to the body of work done on five PaKE vowels production by the university students. The research suggests that the teachers should develop awareness about PaKE sounds and try their best to teach their students to improve their pronunciation in English to speak English naturally and fluently (See Appendix A for more detail).

6. Conclusion

The present study concludes that there is a strong evidence of speech variations between PaKE and American in terms of analysizing three acoustic parameters i.e., F1, F2 formant frequencies and duration of five PaKE vowels between American and PaKE. Speakers tend to move vowels higher and closer to the front positions. Inso far as teaching of English pronunciation is concerned to Pakistani ESL learners context, English language teachers should focus on correct pronunciation of the students because it is their second language they need to focus on English pronunciation in order to improve them. The study discovered different acoustic properties through measurements i.e., formant frequencies and duration across speakers between American and Pakistani English. Thus, this study concludes that there are explicit acoustic variations in American and PaKE vowels frequencies.

References

Abbasi, A. M., Memon, M. A., Channa, M. A., & John, S. (2018). Awareness of L2 American English word stress: Implications for teaching speakers of Indo-Aryan languages. *International Journal of English Linguistics, 8*(3), 101-107. https://doi.org/10.5539/ijel.v8n3p101

Abbasi, A. M. (2017). *The stress pattern of Sindhi and English*. Munich. Lincom Europa. Germany.

Abbasi, A. M., Channa, M. A., Kakepoto, I., Ali, R., & Mehmood, M. (2017). A perceptual study of phonological variations in Pakistani English. *International Journal of English Linguistics, 8*(2), 91-100. http://doi.org/10.5539/ijel.v8n2p92

Abbasi, A. M., & Hussain, S. (2015). Phonetic analysis of lexical stress in Sindhi. *Sindh University Research Journal-SURJ, 47*(4), 749-756.

Abbasi, A. M., & Hussain, S. (2015). The role of pitch between stress and intonation in Sindhi. *Annual Research Journal of English Language Forum, 17*, 09-24.

Abbasi, A. M., & Hussain, S. (2012). Syllable structure and syllabification in Sindhi-English loanwords. *International Journal of Researchers, 1*(4), 120-134.

Beckman, M. (1986). *Stress and non-stress accent.* Dordrecht, Netherlands: Foris Publications. https://doi.org/10.1515/9783110874020

Boersma, P., & Weenink, D. (2017). Praat: Doing phonetics by computer (Version 5.3.35) [Computer software]. Retrieved on from http://www.praat.org

Hillenbrand, J., Getty, L. A., Clark, M. J., & Kimberlee, W. (1995). Acoustic characteristics of American English vowels. Acoustical Society of America. https://doi.org/10.1121/1.411872

Ladefoged, P. (1993). *Phonetics*. Harcourt Brace Jovanovich College Publishers.

Parsons, T. W. (1987). *Voice and speech processing*. New York: McGraw-Hill.

Pfitzinger, H. R. (2003). Acoustic Correlates of the IPA Vowel Diagram. *In The 15th International Congress of Phonetic Sciences,* 1441-1444. Barcelona, Spain.

Sabir & Al-Saeed. (2014). A brief description of consonants in modern standard Arabic. *Linguistics and Literature Studies, 2*(7), 185-189.

Appendix A

Short Vowel / æ / Male Speakers			Short Vowel / æ / Female Speakers		
F1 Hz	F2 HZ	millisecond	F1 Hz	F2 HZ	millisecond
688.183	1714.785	73.91	890.369	2053.679	79.093
761.527	2071.767	73.389	998.966	2228.724	81.742
677.995	1723.700	84.249	839.559	2017.011	91.812
572.635	1905.338	79.794	802.692	1411.476	76.653
543.726	1710.703	87.006	874.586	2197.892	68.713

Short Vowel / ɛ / Male Speakers			Short Vowel / ɛ / Female Speakers		
F1 Hz	F2 HZ	millisecond	F1 Hz	F2 HZ	millisecond
540.426	2010.609	71.673	863.075	1927.974	95.535
756.203	2031.132	89.379	990.393	2189.469	84.483
559.870	1796.995	87.856	994.592	2129.168	90.887
534.471	1907.32	90.284	786.692	1587.214	83.217
711.482	1618.541	92.502	909.420	2141.128	65.236

Short Vowel / ɪ / Male Speakers			Short Vowel / ɪ / Female Speakers		
F1 Hz	F2 HZ	millisecond	F1 Hz	F2 HZ	millisecond
380.351	2076.915	55.88	717.481	2220.787	63.337
479.142	2060.194	52.727	618.851	2324.084	62.738
373.547	1986.584	46.034	694.8917	2554.507	68.65
344.948	2092.784	67.256	564.582	2135.899	48.128
437.897	1560.297	71.987	741.324	2334.687	66.644

Short Vowel / ɒ / Male Speakers			Short Vowel / ɒ / Female Speakers		
F1 Hz	F2 HZ	millisecond	F1 Hz	F2 HZ	millisecond
676.515	1005.165	94.661	942.289	1349.647	58.11
727.722	1103.627	93.278	895.426	1273.006	71.543
541.316	998.896	94.82	922.296	1729.418	77.529
611.671	905.159	98.955	821.731	1159.177	71.104
797.752	1222.609	93.156	847.212	1162.913	84.631

Short Vowel / ə / Male Speakers			Short Vowel / ə / Female Speakers		
F1 Hz	F2 HZ	millisecond	F1 Hz	F2 HZ	millisecond
623.681	1501.452	63.92	860.959	1559.467	64.935
844.045	1417.272	85.858	945.226	1773.664	63.929
660.066	1436.912	71.014	793.949	1616.535	62.764
619.854	1674.663	89.798	837.533	1293.203	42.502
736.478	1612.216	61.936	898.519	1785.394	62.246