

Identification of Diphthongs in Urdu and their Acoustic Properties

Rashida Bhatti and Benazir Mumtaz
Center for Language Engineering,
Al-Khawarizmi Institute of Computer Science,
University of Engineering and Technology, Lahore
firstname.lastname@kics.edu.pk

Abstract

The present research is carried out to finalize the list of diphthongs for the development of Urdu Phonetic Inventory. The corpus is specifically designed in carrier sentences to attest the existence of diphthongs in Urdu. For the identification of diphthongs in Urdu, two approaches are used in this paper i.e. perceptual approach and acoustic approach respectively. In perceptual approach, diphthongs are identified by ten native speakers using syllable identification technique. Diphthongs which passed the perceptual test were sent forward for acoustic testing. In acoustic approach, speech of six native speakers is analyzed using durational and formant cues both at stressed and unstressed forms on PRAAT. The combined analysis of perceptual and acoustic approaches indicates that Urdu has fifteen diphthongs.

1. Introduction

Urdu language has total 67 sounds in its phonetic inventory, i.e. 36 consonants, 15 aspirate consonants, 7 long vowels, 3 short vowels and 3 medial (majhul) vowels [1]. There are also nasalized forms of 7 long and 3 short vowels [1]. Having a large count of vowels in Urdu language is the triggering point to identify the diphthongs in Urdu and define a final list in phonetic inventory of Urdu.

To date, four studies have been carried out to identify the diphthongs in Urdu language and in results there are four different lists of diphthongs. Moreover, there are also 2 diphthongs which are claimed in present research after analyzing the ten hours of speech developed for Urdu TTS [1]. Due to the lack of

consistency among previous researches, this study is carried out. This work is based on a perceptual and an acoustic analysis of Urdu diphthongs hence a defined list of diphthongs can be added in Urdu phonetic inventory.

The results of this study are very important as identification of diphthongs is very essential for smooth annotation process of Urdu speech corpus hence the development of speech database of Urdu language. Moreover, identification of diphthongs will help to develop a robust pronunciation lexicon of Urdu language. Annotated speech corpus and pronunciation lexicon play important role in the development of Natural Language Processing (NLP) tools, i.e. Text to Speech Synthesizer (TTS) system, Automatic Speech Recognition (ASR) system [2] and Screen Readers. By adding these diphthongs in Urdu Phonetic Inventory, quality of the automatic speech would be improved. It would be more natural, audible and pleasant to hear.

The paper is organized in the following sections. The previous researches on the identification of diphthongs in the languages of world, Indo Aryan languages and specifically Urdu are presented in section 2. The methodology to study Urdu diphthongs is detailed in section 3. Data analysis description is presented in section 4, results and conclusions are discussed in section 5, while future work and recommendations are presented in section 6.

2. Literature review

A diphthong may be defined as a sequence of two perceptually different vowel sounds within one and the same syllable [3] or as a single vowel with continuously changing qualities [4]. There is no evidence of phonemic diphthongs in standard Urdu language [5] but present study shows that phonetic

diphthongs are observed in the speech of native Urdu speakers.

The reported researches in other languages have used different approaches to investigate the diphthongs, i.e. phonological marking of diphthongs through minimal pairs [6], perceptual identification of diphthongs using syllabification [7] and acoustic study of diphthongs using spectrum (formant) and duration analysis [6][8]. Acoustically, diphthong has some important information at three points, i.e. starting vowel, transition period and final vowel [9]. Okati, Helgason & Jahani [10] have used these points for acoustic analysis for the identification of diphthongs; they have analyzed formant values and duration of these three components. In another study on the learners of South African English, pitch contour is used for acoustic analysis along with formant analysis [11].

Urdu is a member of Indo Aryan Languages [12]. Existence of diphthongs in the phonetic inventories of these languages is reported in different researches but those are less in numbers. In Bengali, there are 21 diphthongs [8]; two in Sanskrit while no diphthong in Sinhalese [13]. Six diphthongs are reported in Kashmiri and Pahari language [6]. Punjabi Language has 6 diphthongs which are comprised of short vowel and long vowels [2]. Such possibilities are also seen in Urdu. Samare [14] posits six diphthongs for the Persian language varieties spoken in Iran, but points out that they are only diphthongs from a phonetic point of view and can also be described as sequences of vowel and glide. Ganjavi et al. [15], Yaesoubi [16] and Hakimi [17] have reported diphthongs in different dialects of Persian, but many of those are either phonetic diphthongs or the combination of vowel and glides (i.e. /j/ and /w/). As Persian has much influence on Urdu, here is the possibility that Urdu language may have this type of diphthong combinations.

To date, four researches have been carried out for the identification of diphthongs in Urdu [7], [18], [19] and by CLE (Center for Language Engineering) researchers (available on: www.cle.org.pk). These studies present 4 lists and every list has different number of diphthongs. Almost, all the researches have used syllabification technique to identify the diphthongs and then acoustic cues are used to describe the characteristics of Urdu diphthongs.

Waqar and Waqar [7] have proposed total 13 diphthongs, i.e. /i̯u:/, /æe:/, /əi:/, /a:o:/, /a:i:/, /a:e:/, /e:a:/, /o:i:/, /a:ẽ:/, /o:e:/, /əĩ:/, /ia:/ and /a:ũ:/. Obtained results show that existence of diphthongs is speaker dependent and diphthong combination is in the result of deletion of phonemes, i.e. “ʔ”, “j” and “v”. Moreover, their analysis of duration shows that diphthongs are combination of one short and one long vowel and the duration of diphthongs is below 300

while duration of consecutive two vowels is round about 350ms. Phonemically the existence of diphthongs (using minimal pairs) cannot be proved, since they are formed as a result of deletion of either the consonant, or the timing slot.

Khurshid, Usman and Butt [18] have finalized 18 diphthongs in Urdu using syllabification and acoustic analysis. Their list includes /oi/, /oe/, /io/, /əi:/, /æe:/, /ua:/, /uə/, /a:ɪ/, /ao/, /aũ:/, /iũ:/, /io/, /ea/, /eo/, /ʊa/, /ui/ and /ue/ diphthongs. Another research conducted by Sarwar, Ahmad and Tarar [19] depicts that Urdu has 17 diphthongs, i.e. /ai/, /ae/, /ao/, /iũ/, /ia/, /aũ/, /oi/, /oe/, /oĩ/, /əi/, /aẽ/, /ea/, /əĩ/, /ʊa/, /ui/, /ue/ and /æe/. The obtained results show that diphthongs have had three parts, i.e. on glide, off glide and transition period. CLE has finalized a list of 7 diphthongs. This list reports unique combination of diphthong having medial (majhul) and long vowel qualities. The common thing in all these studies is that diphthongs are identified perceptually using native speaker intuition. Moreover, previous researches report that identification of syllabification and diphthong is speaker dependent. It is also noticed that identified diphthongs are mostly consisted of one short and one long or combination of two long vowels. [7] and [18] previous work claims that a diphthong will be considered a diphthong only, if 50% votes are in favor of a particular diphthong whereas [19] prefers 60% votes for the selection of diphthong.

There are two new diphthongs, which are not mentioned in the previous researches, i.e. /ea:/ and /a:e/. These diphthongs are studied during the annotation of 10 hours Urdu speech corpus; /ea:/ diphthong in words like اعادہ /ea:da:/ repeat and پیار /pea:r/ love etc and /a:e/ diphthong in words like جائزہ /dʒa:eza:/ overview, کائنات /ka:ena:t/ universe etc.

Hence, due to lack of inconsistency among lists there is no specified list of diphthongs which may add in phonetic inventory of Urdu language. Thus, the current study is built on the previous research efforts to develop a unified list of diphthongs.

The following section presents the methodology followed during the research.

3. Research Methodology

This study is carried out by combining the four previous lists of diphthongs along with 2 possible diphthongs proposed in present research. In combined list, total 26 possible diphthongs are selected for study.

3.1. Corpus development and speech recordings

In order to study the proposed phenomenon, corpus was specifically designed for recording. Seventy eight (78) words containing diphthongs were selected and embedded in carrier phrase. For example:

- 1) میں نے آئینی کہا
/mē: ne: a:ɪni: kəhɑ:/
I said constitutional.

It was made sure that there was a valid coverage of all 26 possible Urdu diphthongs in the corpus (3 words for each possible diphthong (3*26=78)). Recordings were obtained from six native speakers of Urdu (3 males and 3 females) in an anechoic chamber who also use Punjabi in their daily routine. PRAAT software was used for recordings and analysis. The speech samples were recorded in mono form at the sampling rate of 48 KHz, and stored in wav file format. Six speakers were asked to read out the sentences in their natural style of speaking. Three instances of each speaker's voice samples were recorded (3*3*26=234) and stored in wav file format for subsequent offline processing. Afterwards segments were marked on phoneme, syllable, word and stress levels using process described in [1]. Moreover, ten sentences containing distracters were also recorded from the speakers. Distracters were the words imbedded in carrier phrases to attest the respondents' perception regarding syllabification, i.e. عادلانہ /ʔɑ:ɖɪla:nɑ:/ uprightly, اسطوخودوس /oʃtu:xoɖɖu:s/ lavender, اسلامیات /isla:mja:t/ Islamic studies, اشارت /ɪʃɑ:rəʃ/ insinuation, بلاغت /bəla:ɣət/ eloquence, دلیرانہ /dɪle:ra:nɑ:/ courageous, رؤیت /ru:jət/ visibility, رؤف /rəu:f/ rauf, ہوائی /həva:i:/ hawaii, لاجورد /la:dʒvərd/ Armenian stone and لاعلم /la:ʔilm unaware.

3.2. Perceptual experiment methodology

To verify the defined list of diphthongs, first of all, recorded sentences containing 26 possible diphthongs were listened and segmented using PRAAT. Mispronounced or having bad quality voice were not selected for perceptual analysis. Three utterances of each diphthong were selected at word initial, middle and final positions from the speech of six speakers for the perceptual experimentation (26*3*6=468). However, there are few diphthongs such as a:e, iā:, iō: and əi: which do not exist at word initial and middle position in Urdu. For these diphthongs, three instances are taken only at word final position. Thus, there were three waves of one diphthong, which consisted of utterances of six speakers. Moreover, to evaluate the native speaker perceptual understanding of diphthongs, utterances of six speakers containing one diphthong

and two sentences containing distracters were combined in one wave file.

Later on, 10 native speakers (5 males and 5 females whose age vary from 20 to 30) were asked to listen to these 78 wave files one by one using headphones to identify the diphthongs in Urdu. Syllable count is a good cue to identify the diphthongs; therefore, respondents were asked to count the syllables in recorded words. Respondents listened to all three files against each diphthong and wrote the syllable count in a given questionnaire. On the basis of their syllable count log sheet, 16 diphthongs are finalized by the respondents. Among these 16, 5 are the nasalized diphthongs (See Section 4 Table 1).

3.3. Acoustic Experiment methodology

To verify the proposed list of 16 diphthongs, durations of finalized diphthongs and formant frequencies are analyzed manually. Duration of diphthongs in both stressed and unstressed forms are calculated separately (3 unstressed+3 stressed*16 diphthongs*6 speakers=576). Average values of males and females are calculated and enlisted in Appendix A. Moreover, minimum duration in unstressed form and maximum value at stressed form is also mentioned in Appendix A. Only one perceptually selected diphthong /ɑ:ĩ:/ is rejected at this stage of experimentation.

Formant frequencies of finalized 15 diphthongs are measured from the recorded speech. To measure the formant frequency of first (F1), second (F2) and third formant (F3), diphthongs were divided into three components, i.e. on glide (1), transition (2) and off glide (3). F1, F2 and F3 are measured manually from the middle of component 1 and 3. Window of PRAAT was assured to be 20 ms to take formant values of each component. Three instances of every diphthong from the recorded speech of six speakers (3*15*6=270) are considered for formant values. Average formant frequencies of finalized diphthongs are reported in Appendix B.

4. Results

Selection of diphthongs was done on the basis of frequency of the responses. In this research, a diphthong is considered a diphthong only, if 70% votes are in favor of a particular diphthong. Only one diphthong /ɑ:ĩ:/ having 60% votes was selected for further testing and highlighted in green color in table 1. The list of perceptually selected diphthongs is given in Table 1.

Table 1: Results of perceptual analysis

Sr No .	Diphthongs	Perceptual Agreed Diphthongs	Perceptual Disagreed Diphthongs
1	a:e:	90 %	10 %
2	a:e	80 %	20 %
3	a:ẽ:	70 %	30 %
4	a:I	90 %	10 %
5	a:i:	80 %	20 %
6	a:ĩ:	60 %	40 %
7	a:o:	80 %	20 %
8	a:u:	40 %	60 %
9	æe	100 %	0 %
10	æa:	90 %	10 %
11	əi	100 %	0 %
12	əĩ:	80 %	20 %
13	ea: ¹	70 %	30 %
14	e:o:	20 %	80 %
15	ia:	10 %	90 %
16	iã:	30 %	70 %
17	io:	20 %	80 %
18	iõ:	80 %	20 %
19	iu:	0 %	100 %
20	iũ:	90 %	10 %
21	o:e:	50 %	50 %

¹ Red colored are proposed diphthongs in present research

22	o:i:	80 %	20 %
23	o:ĩ:	50 %	50 %
24	ua:	30 %	70 %
25	ue:	20 %	80 %
26	ui:	80 %	20 %

□ Grey highlighted are diphthongs which are rejected in perceptual analysis.

5. Discussion and Data Analysis

The obtained results from perceptual analysis show that there are 16 diphthongs in Urdu which are identified by ten native speakers of Urdu. Respondents rejected ten proposed diphthongs out of 26, i.e. /iã:/, /u:e:/, /a:u:/, /e:o:/, /o:e:/, /o:ĩ:/, /ia:/, /io:/, /iu:/ and /ua:/ in perceptual testing. Sixty percent respondents accepted the diphthong /a:ĩ:/ but we have had delimited the votes in favor to 70 %. However, we accepted this diphthong for further experimentation. Four speakers out of six have pronounced diphthong /ia:/, differently. Either the speakers pronounced it with 'j' or without 'j'. Respondents did not recognize it as diphthong (See Figure 1: J sound in /lija:/). They counted it bi syllabic word. Similarly /ua:/ and /u:e:/ are also not pronounced as diphthongs by speakers. Only one speaker pronounced it as diphthong and only one listener recognized it as diphthong, but the rate of speech was comparatively speedy during this particular utterance.

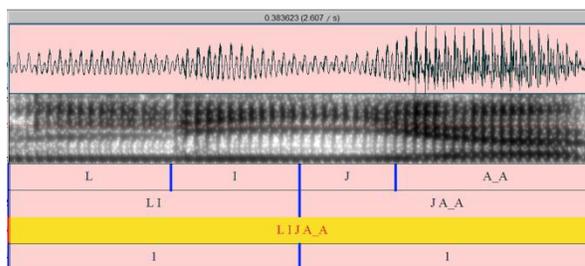


Figure 1: J sound in /lija:/

Two speakers have pronounced /v/ in second syllable **ہوئے** /huve:/ happened. Either those are spoken with /v/ or without /v/; native speakers did not recognize those as diphthongs (See **Figure 2: V sound in /huv/**). /iu:/ is not accepted by the respondents as diphthong, although it is a controversially accepted diphthong in American English but Urdu speakers did not speak it as diphthong.

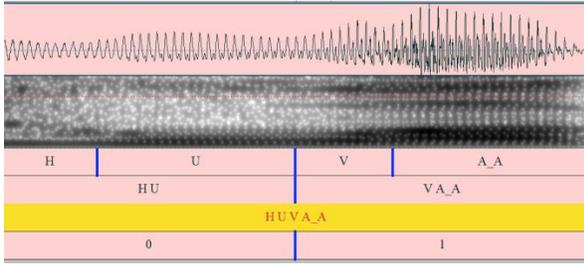


Figure 2: V sound in /huva:/

Three speakers pronounced /jā:/ instead of proposed diphthong /iā:/ (As in ناکامیاں /nakamīā:/ failures, لڑکیاں /laṛkiā:/ girls and کھڑکیاں /khiṛkiā:/ windows) while the other three pronounced it as diphthong. However, the 70% respondents did not recognize it as diphthong. Five speakers out of six could not speak the proposed diphthong /o:ī:/ rather they pronounced it as /o:i:/.

Data analysis of selected diphthongs suggests that diphthongs in Urdu can occur in 5 types of combinations.

- short and long oral vowels
 - /a:ɪ/, /æe:/ and /əi:/
- long and long oral vowels
 - /a:e:/, /a:i:/, /a:o:/, /o:i:/ and /u:i:/
- medial and long oral vowels
 - /a:e/, /æa:/ and /ea:/
- Short and long nasalized vowels
 - /əĩ:/, /iō:/ and /iū:/
- Long and long nasalized vowels
 - /a:ẽ:/

Combination of long-long vowel and long-medial vowel is unique property of Urdu language. /æa:/ is considered as diphthong, 5 speakers out of six pronounced it as diphthong while all respondents recognized it as one syllable hence diphthong.

Waqar and Waqar [7] had the conclusion that diphthong is made in the result of the deletion of any phoneme but the formant analysis shows that Urdu speaker alternate the schwa and J with medial vowel /æ/. The formants of medial vowel /æ/ in /æa:/ diphthong and individually has almost similar values (See Appendix B).

In the case of diphthong /ea:/ (e.g. in زیادہ /ze:ɖa:/ excessive, تیاری /te:ri:/ preparation and فلکیات /falkea:t/ universe), Urdu speakers alternate the sound /ɪ/ and /j/ with medial /e/. This medial vowel blends with the following vowel /a:/ and makes the diphthong /ea:/ as shown in Figure 3. All the speakers pronounced it as diphthong and 70% respondents recognized it as diphthong. Rest 30% respondents had

the confusion to identify it as diphthong in word /falkea:t/. They counted it as tri syllabic word.

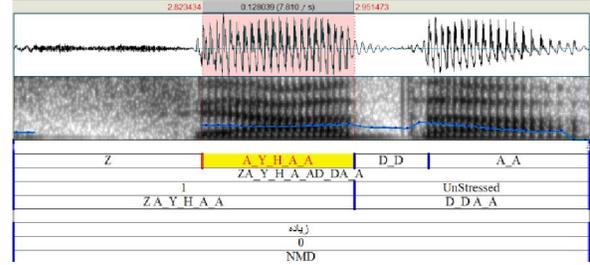


Figure 3: e sound in Diphthong /ea:/

Duration analysis shows that diphthongs show the qualities as an entity like long vowels. The duration of diphthongs increase in the state of stressed syllable like long vowels. On average the maximum durations of unstressed diphthongs is. 148 ms (See Appendix A). Therefore, on the basis of durations, /a:ī:/ diphthong was rejected at acoustic experiment stage. Obtained results show that there is no significance difference in average duration values of diphthongs on the bases of gender. Almost the duration is similar in the speech of males and females. During annotation and perceptual analysis of diphthongs, it is also observed that in diphthongs both vowels blends in such a way that listeners cannot separate them as shown in Figure 4.

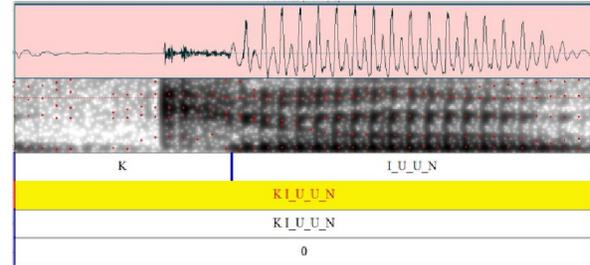


Figure 4: Diphthong I_U_U_N

Formant analysis of finalized 15 diphthongs shows that F1, F2 and F3 of vowel components in different diphthong combinations do not have much difference in values. One vowel shows almost similar values in different combinations of diphthongs. For example /a:/ has not much difference in different combinations of diphthongs like in /a:e:/, /a:e/, /a:ẽ:/, /a:i:/, /a:ɪ/ and /a:o:/ (See Appendix B). Obtained results show that although vowel maintains their qualities and formant frequency but they blend with other component to become a diphthong. Moreover, analysis of Urdu diphthong highlights that similar to other languages, Urdu diphthongs have three components, i.e. first vowel, transition period and second vowel.

6. Future work and recommendations

Finalized list of diphthongs is presented in this research using perceptual and acoustic approaches. The selected diphthongs can be added in Urdu phonetic inventory. It will be helpful in maintaining the accuracy and consistency during the annotation of speech corpus. By marking diphthongs, syllables and stress tier annotation can also be done more smoothly and accurately. Moreover, pronunciation lexicon can become more robust using list of diphthongs. Hence, Urdu speech database would be more accurate and will represent the quality speech of native speakers. Moreover, this study reports that the sounds /ə/ and /j/ replace with /æ/ and the sounds /ɪ/ and /j/ replace with medial /e/ to form diphthongs. These alternation results are based on the speech of six native speakers. This phenomenon needs to be studied on a large sample to confirm the trend of vowel shifting or alternation among the native Urdu speakers.

7. Acknowledgement

This work has been conducted through the project, Enabling Information Access for Mobile based Urdu Dialogue Systems and Screen Readers supported through a research grant from ICTRnD Fund, Pakistan.

8. References

- [1] B. Mumtaz, A. Hussain, S. Hussain, A. Mehmood., R. Bhatti, M. Farooq, & S. Rauf. "Multitier Annotation of Urdu Speech Corpus", Conference on Language and Technology (CLT14), Karachi, Pakistan, 2014.
- [2] L. Swaran, "Challenges for Design of Pronunciation Lexicon Specification (PLS) for Punjabi Language", Department of Information Technology, Govt of India, 2011.
- [3] J. C. Catford, "Fundamental Problems in Phonetics", Edinburgh: Edinburgh University Press, 1977, pp 215.
- [4] P. Ladefoged,. "A Course in Phonetics", Los Angeles: University of California, ed 2, 1982, pp 171.
- [5] M. A. Khan, "Urdu Ka Sauti Nizam", National Language Authority, Pakistan, 1997.
- [6] A. Q. Khan, "A Preliminary Study of Pahari and Its Sound System", The criterion An International Journal in English Vol IV, Issue IV, 2013, pp 1-20.
- [7] A. Waqar, & S. Waqar, "Identification of Diphthongs in Urdu and Their Acoustic Properties", 2002.
- [8] F. Alam, S. M. Habib & M. Khan, "Acoustic analysis of Bangla vowel inventory" BRAC University, 2008.
- [9] R. D. Kent, C. Read, & R. D. Kent, "The acoustic analysis of speech", (Vol. 58). San Diego: Singular Publishing Group, 1992.
- [10] F. Okati, P. Helgason, & C. Jahani, "Diphthongization in Five Iranian Balochi Dialects", *Orientalia Suecana*, Vol 61, 2013, pp 107-119.
- [11] C. P. Prinsloo, "A comparative acoustic analysis of the long vowels and diphthongs of Afrikaans and South African English", 2006.
- [12] B. Comrie, "Languages of the world. *The Handbook of Linguistics*", 2001, pp 522 (online book: cited on May 09,2016).
- [13] C. P. Masica, "*The Indo-Aryan Languages*", Cambridge University Press, 1993.
- [14] Samare, Yadollah (Samare, Yadollāh). "*Āvāšenāsi-ye zabān-e fārsi, āvāhā va sāxt-eāvāyi-ye hejā*", Tehran: Markaz-e Našr-e Dānešgāhi, 1368 [1989-90].
- [15] S. Ganjavi, P. G. Georgiou, & S. Narayanan, "ASCII based transcription systems for languages with the Arabic script: The case of Persian", 2003, In *Automatic Speech Recognition and Understanding*, 2003. ASRU'03. 2003 IEEE Workshop on IEEE, pp 595-600.
- [16] M. Yaesoubi, "*Machine Transcription Conversion between Perso-Arabic and Romanized Writing Systems*", Master thesis, Linköpings University, 2010. Online at: <http://liu.divaportal.org/smash/record.jsf?pid=diva2:360341> [Retrieved May 30, 2016].
- [17] A. Hakimi, "Comparative Phonetic Study of Frequently Used Words in Iranian Farsi versus Tajik Farsi", *Journal of American Science* 8(4), 2012, pp 6-16.
- [18] K. Khurshid, S.A. Usman, & N. J. Butt, "Possibility of existence and identification of dipht hongsand triphthongs in Urdu language". *Akhbar-e-Urdu*, 2003, pp 16-21.
- [19] A. Sarwar, S. Ahmed, & A. A. Tarar, "Diphthongs in Urdu language and analysis of their acoustic properties", *Annual Student Report*, Center for Research in Urdu Language Processing (CRULP), 2003, pp 9-15.

Appendix A

Sr No.	Diphthong	Average Duration Males (ms)	Average Duration Females (ms)	Minimum Duration males in Unstressed (ms)	Minimum Duration females in Unstressed (ms)	maximum Duration males in stressed (ms)	maximum Duration females in stressed (ms)
1	a:e:	254	207	127	147	296	267
2	a:e	231	212	186	169	280	280
3	a:ē:	282	212	200	170	277	276
4	a:ɪ	212	175	182	134	288	228
5	a:i:	233	205	205	150	261	262
6	a:o:	205	221	127	179	268	286
7	æe:	236	186	177	153	272	226
8	æɑ:	321	215	162	148	313	342
9	əi:	234	186	176	153	302	226
10	əĩ:	235	206	201	147	289	258
11	ea:	207	195	142	142	241	271
12	iõ:	203	152	179	146	225	165
13	iũ:	227	126	160	147	273	209
14	o:i:	244	192	229	96	267	293
15	ui:	201	210	167	161	231	327
16	a.ĩ: ²	300	285	350	336	322	354

² Diphthong rejected on the basis of durations.

Appendix B

Sr No.	Diphthong	First Component			Second Component		
		Average F1 Hz	Average F2 Hz	Average F3 Hz	Average F1 Hz	Average F2 Hz	Average F3 Hz
1	ɑ:e:	813	1707	2990	501	2407	2977
2	ɑ:e	797	1376	2700	542	2035	2970
3	ɑ:ē:	890	1534	3144	716	2221	3126
4	ɑ:ɪ	807	1527	2832	531	2154	2928
5	ɑ:i:	788	1694	2792	330	2379	3017
6	ɑ:o:	712	1431	2884	612	1184	2956
7	æe:	454	1830	2631	422	2187	2752
8	æɑ:	525	2048	2759	746	1541	2622
9	əi:	517	1835	2682	319	2477	3094
10	əĩ:	418	2228	3039	332	2118	3046
11	ea:	419	2103	2763	683	1492	2582
12	ɪð:	351	2104	2893	553	1412	2827
13	ĩũ:	331	1637	2596	347	1146	2732
14	o:i:	431	1212	2782	332	2360	2889
15	ui:	313	1735	2708	295	2464	2974