# Urdu Vowel System and Perception of English Vowels by Punjabi-Urdu Speakers

Thesis submitted in partial fulfilment of requirements for the degree of

Doctor of Philosophy

to

Department of English Language and Linguistics (ELL)

School of European Culture and Languages (SECL)

University of Kent, Canterbury

*by*

*Ishrat Rehman*

2019

**Abstract**

A well-defined vocalic and consonantal system is a prerequisite when investigating the perception and production of a second language. The lack of a well-defined Urdu vowel system in the multilingual context of Pakistan motivated investigation of the acoustic and phonetic properties of Urdu vowels. Due to the significant influence of a number of first languages, the study focuses on the Urdu spoken in Punjab, Pakistan. A production experiment reports the acoustic properties of the monophthongs and six diphthongs in Urdu. The results showed that Urdu distinguishes between short and long vowels, and lacks an open-mid front and an open-mid back vowel. Since the central vowel is fairly open and retracted, it appears that the central vowel space is empty. This was reflected in the difficulty of perceiving the central vowels of Standard Southern British English (SSBE) by Punjabi Urdu speakers. The acoustic and phonetic evidence partially supports the phonetic existence of diphthongs in Urdu.

The acoustic investigation of the Urdu vowel system helped to predict the perceptual assimilation and classification patterns of SSBE vowels by Punjabi-Urdu speakers. A cross-language perceptual assimilation and a free classification experiment was conducted in three different consonantal contexts to test the predictions of three mainstream models of L2 perception: SLM, PAM and L2LP. The assimilation patterns in a cross-language and category goodness rating task varied according to familiarity with the target language. The patterns of perceptual assimilation failed to predict the perceptual similarity of the SSBE vowels in the auditory free classification task. Thus, the findings support the model predictions with regard to the role of L1; however acoustic similarities between L1 and L2 neither predict the patterns of cross-language perceptual assimilation nor perceptual similarity.

**Acknowledgements**

First of all, I am immensely grateful for the continuous support and guidance from my primary supervisor, Prof Amalia Arvaniti, throughout the course of my PhD. Prof Arvaniti spent numerous hours and sent feedback on multiple drafts up until the very last minute. This thesis would not have been possible without the continuous and informative input from Prof Arvaniti. I am forever grateful to Prof Arvaniti for her kindness, patience, continuous support, guidance and for such an amazing learning experience throughout the course of my PhD.

I am also very grateful to my secondary supervisor, Dr Gloria Chamorro, who gave me feedback on second language perception and advised me to prepare the questionnaires for participants via Google forms. I would also like to thank Dr Tamara Rathcke for her advice and guidance on acoustic data analysis and for introducing me to R. I am also very grateful to Dr Angelos Lengeris for his input and advice on the data analysis for my perception experiments.

I would also like to thank Jacqui Martlew for her continuous guidance over the years and help for research grant applications for conference presentations and travel for data collection. I am also very grateful to the Head of Department, Dr David Hornsby and other staff members in ELL and SECL for their kindness and support over the years.

This PhD would not have been possible without the partial funding from SECL (50th Anniversary Graduate Teaching Assistant (GTA) scholarship), and therefore I am very grateful for this once in a lifetime opportunity.

I would also like to thank the 101 participants without whose participation I would not have completed this thesis. I am very grateful to Dr Arif Mateen and Mrs Ferdos Arif, in Boston, and Prof Amra Raza, Head of Department at the Department of English Language and Literature, University of the Punjab, Lahore, Pakistan for facilitating my data collection and giving me permission and facilities to use for my data collection.

The continuous support and well-wishes from my family kept me going, and I am very grateful to my mother, sisters and brother for believing in me and supporting me throughout the course of my PhD. I cannot express in words how thankful I am to my husband, in particular for his patience, and because without his moral, technical and financial support I would not have been able to complete this thesis.

*For my mother*

[meːriː ɐmmĩː keː lɪjeː]

میری امی کے لئے

# Table of Contents

# List of Tables

## List of Figures

## List of Abbreviations

| | |
|---|---|
| AE | American English |
| ANOVA | Analysis of Variance |
| CP | Carrier Phrase |
| FS | Full Sentence |
| L2LP | Second Language Linguistic Perception Model |
| LEAP-Q | Language Experience and Proficiency Questionnaire |
| LMEMs | Linear Mixed Effects Models |
| MFC | Multiple Forced Choice |
| PAM | Perceptual Assimilation Model |
| PAM-L2 | Perceptual Assimilation Model of second language speech learning |
| PE | Pakistani English |
| RP | Received Pronunciation |
| SLM | Speech Learning Model |
| SSBE | Standard Southern British English |
| $TL_{roc}$ | Trajectory Length Rate of Change |
| $VSL_{roc}$ | Vowel Section Length Rate of Change |

**Chapter 1**

# Introduction

This thesis consists of two parts. Part I focuses on the investigation of the Urdu vowel system as spoken in Punjab, Pakistan. Part II focuses on the perception of Standard Southern British English (SSBE) vowels by Urdu speakers from Punjab, Pakistan. Each part consists of three chapters. The final chapter then gives the conclusions drawn from both parts and the implications for future research.

In this chapter, the background and status of Urdu in Pakistan is presented, followed by a brief overview of Pakistani English and perception of English vowels.

## 1.1 Background - Urdu in Pakistan

According to Ethnologue (Simons et al., 2017) and Hussain (2004), more than 100 million people speak Urdu around the world. Pakistan has the greatest number of Urdu speakers and India has the next greatest number of speakers. In India, Urdu is one of the 22 official languages and is spoken across six different states. In Pakistan, Urdu was declared in the constitution of 1956 as the *national language* (Javed et al., 2010), with some amendments in the constitution of 1973, where provincial governments were allowed to promote provincial languages as well as Urdu (Farooq, 2014:17). In this context, the term *national language* can be defined as a common language used by people from different linguistic backgrounds within the same country. Only 7.57 percent of the population of Pakistan speaks Urdu as their first language (Rahman, 2011; Rahman, 2008; Mansoor, 2004). The majority of the population speaks Urdu as their second or third language.

Urdu is close to Hindi and like Hindi it belongs to the New Indo-Aryan languages (Kachru, 1987). Although close, Urdu and Hindi differ from each other in their morphology, syntax, phonetics and phonology (see Kachru, 1987:53-72 for a detailed review). The word Urdu is derived from the Turkish word *ordu* which means “Camp or

Army with its followers" (Saleem et al., 2002:1). It is also considered an "offspring of Persian" and borrows its vocabulary mainly from Persian, Arabic, English, and Portuguese (Saleem et al., 2002; Khan and Alward, 2011). Due to the influence of local languages, various varieties/dialects of Urdu are spoken in Pakistan.

Urdu is mainly spoken by educated people and in urban areas: "...the middle and upper classes make more use of Urdu and English in Pakistan as compared to the rural classes that mainly use the regional languages" (Mansoor, 2004:336). Hence socio-economic background is a key indicator for the variety of Urdu spoken by the participants of any study.

Pakistan is a multilingual country where the vast majority of people speak at least two local languages. Urdu is taught as a compulsory subject throughout primary and secondary school education. According to a report from the British Council in Pakistan (Coleman, 2010) and Rahman (2008; 2011) six major and 58 minor languages are currently spoken in Pakistan. Major languages are shown in Table 1. 1 (see also Rahman, 2008; 2011) and Table 1. 2 shows the languages that are spoken as a first language by one million or more speakers (Coleman, 2010). From everyday experience, it is clear that Urdu is heavily influenced by these local languages. This was confirmed by the pilot study in Section 2.2.

Table 1. 1: Percentage of speakers of the major languages in Pakistan (Source: GoP 2001:107 cited in Rahman, 2011:56)

| Languages | Percentages of speakers |
|---|---|
| Punjabi | 44.15 |
| Pashto | 15.42 |
| Sindhi | 14.10 |
| Saraiki* | 10.53 |
| Urdu | 7.57 |
| Balochi | 3.57 |
| Others | 4.66 |

*Also sometimes spelled Seraiki / Siraiki

Table 1. 2: Individual languages with over 1,000,000 first language speakers in Pakistan (after Coleman, 2010:13)

| No | Language name | Speakers (millions) | Percentage of population |
|---|---|---|---|
| 1 | Punjabi, Western | 60.6 | 38.3 |
| 2 | Sindhi | 18.5 | 11.7 |
| 3 | Saraiki* | 13.8 | 8.7 |
| 4 | Urdu | 10.7 | 6.8 |
| 5 | Pashto Northern | 9.6 | 6.1 |
| 6 | Pashto Central | 7.9 | 5.0 |
| 7 | Balochi, Southern | 2.8 | 1.8 |
| 8 | Brahui | 2.0 | 1.3 |
| 9 | Hindko, Northern | 1.9 | 1.2 |
| 10 | Balochi, Eastern | 1.8 | 1.1 |
| 11 | Pashto, Southern | 1.4 | 0.9 |
| 12 | Balochi, Western | 1.1 | 0.7 |
| 13 | Farsi, Eastern | 1.0 | 0.6 |
| 14 | Punjabi, Mirpur | 1.0 | 0.6 |
| | **Sub-total** | **134.1** | **84.8** |
| | 58 other languages | 24.0 | 15.2 |
| | **Total** | **158.1** | **100.0** |

*Also sometimes spelled Seraiki / Siraiki

In India, Urdu is written in Devangari script, while in Pakistan, it is written in Perso-Arabic script in Nastalique style with an extended Arabic character set which includes diacritic marks (Ijaz and Hussain, 2007:1). The Nastalique style is quite complex as it is cursive (initially used by calligraphers), with no spaces between words, and context sensitive (Javed et al., 2010). Context sensitivity means that the shape of the characters changes depending on the syntactic and semantic context. The diacritics are used to represent the vocalic content (i.e. vowels); however most often diacritics are optional as the vocalic content can be deduced easily from the context. According to Ijaz and Hussain,

> "Urdu is normally written only with letters, diacritics being optional. However, the letters represent just the consonantal content of the string and in some cases (under-specified) vocalic content. The vocalic content may be optionally or completely specified by using diacritics with the letters... In certain cases, two different words (with different pronunciations) may have exactly the same form

if the diacritics are removed, but even in that case writing words without diacritics is permitted" (2007:2).

For example, the Urdu word بد can be pronounced as /bɐd̪/ "*bad*" or /bɪd̪/ "*new*" depending on the context. In a given context it is usually possible to correctly pronounce homographs that differ in pronunciation without recourse to diacritics marks; however, diacritics are required to pronounce such words and understand their meaning when they are written in isolation. In addition, there are certain words in Urdu that can only be written with diacritic marks, for instance /ɑːlɑː/ اعلیٰ "*superior*" (Ijaz and Hussain, 2007). The significance of diacritics is relevant for some of the results presented in Chapter 2.

Given the multilingual context of Pakistan and the status of Urdu in Pakistan, this thesis will mainly focus on the studies conducted in Pakistan, and the literature on Hindi-Urdu will not be explored further.

## 1.2 Urdu Sound Structure

Literature is sparse on Urdu linguistics and especially sparse on Urdu phonology. Most of the recent work in the past two decades has been done by Centre for Language Engineering, Al-Khawarizmi Institute of Computer Science, University of Engineering and Technology, Lahore, Pakistan.

The literature disagrees on the number of consonants and vowels in Urdu. For consonants, there are claims of 41 (Kachru, 1987); 36 (Bokhari, 1985; Bokhari, 1991; Hussain, 1997; Raza, 2009); and 44 (Saleem et al., 2002). The present study focuses on the investigation of Urdu vowel system, therefore the literature on Urdu consonants will not be discussed any further. A brief overview of the literature on the Urdu vowel system is as follows. The Urdu vowel system is presented using the slashes because the earlier studies consider these vowels to be phonemic, although not all the studies consider all the vowels to be phonemic.

### 1.2.1 Monophthongs

There are various claims on the number of Urdu vowels in the literature, as shown in Table 1.3. With the exception of Rauf (1997) and Bokhari (1985) who do not explicitly cite any experimental evidence for their analyses, most work is based on small studies conducted with 6 to 8 participants (sometimes only males) from Lahore, of unspecified socio-economic background.

Saleem et al. (2002) presented an Urdu consonantal and vocalic inventory and briefly reported a vowel /ɛ/ which has not been discussed in previous studies and suspected that it could be an allophonic realisation of /æ/. Saleem et al. (2002) did not provide any acoustic and phonetic analysis of the Urdu vocalic and consonantal sounds; however, they provided a comprehensive list of these sounds. The literature in Urdu phonology and phonetics not only disagrees on the number of vowels and consonants, but also reports inconsistent symbols/transcripts of the given sounds. The list of vowel symbols used across the literature is given in Table 1.4.

Table 1. 3: Number of Urdu vowels in a number of studies

| Oral long vowels | Oral short vowels | Nasal long vowels | Nasal short vowels | Total Vowels | Source |
|---|---|---|---|---|---|
| 7 | 7 | 7 | 7 | 28 | Bokhari, 1985 |
| 7 | 3 | 0 | 0 | 10 | Kachru, 1987; Hussain, 1997 |
| 5 | 5 | 5 | 5 | 20 | Bokhari, 1991 |
| 5 | 3 | 5 | 3 | 16 | Rauf, 1997 |
| 7 | 7 | 0 | 0 | 14 | Fatima and Aden, 2003 |
| 7 | 3 | 7 | 0 | 17 | Ali and Hussain, 2010 |
| 8 | 3 | 6 | 0 | 17 | Hussain et al., 2011 |

Table 1. 4: The vowel symbols used in the literature and the ones used in the present study

| Vowel Symbols used by different authors | | | | Used by the author in this thesis |
|---|---|---|---|---|
| Bokhari (1985) | Rauf (1997) | Hussain (1997) | Saleem et al. (2002) | |
| iː | iː | i | i | iː |
| i | I | ɪ | ɪ | ɪ |
| eː | eː | e | e | e |
| e | -- | -- | ɛ | ɛ |
| ɛː | -- | æ | æ | æ |
| ɛ | a | ə | ə | ə |
| a | -- | -- | -- | ʌ* |
| aː | aː | a | ɑ | ɑː |
| oː | oː | o | o | o** |
| o | -- | ɔ | ɔ | ɔː |
| u | U | ʊ | ʊ | ʊ |
| uː | uː | u | u | uː |

*None of the previous studies investigated the existence of the vowel /ʌ/ in Urdu vocalic inventory.

**The literature unanimously presents two open-mid back vowels /ɔ/ and /o/. In the present study, /o/ is used on the assumption that the difference is solely one of duration (i.e. /ɔː/ might be longer than /o/.

Most of the studies on Urdu vowels are outdated, especially in the diverse socio-cultural, economic and multilingual context of Pakistan. The studies are difficult to compare because it is not clear which variety of Urdu is investigated, and the symbols used are borrowed from the English phonetic tradition and are not always compatible with IPA usage, so it is not clear if authors used them to symbolise the same vowels.

Most of these studies have presented 7 long oral vowels, /iː/, /eː/, /æː /, /aː/, /ɔː/, /oː/, and /uː/, and 3 short vowels /ɪ/, /ə/, and /ʊ/ (Saleem et al., 2002); or 8 long vowels, /i/, /e/, /ɛ/, /æ/, /ɑ/, /ɔ/, /o/ and /u/, and 3 short vowels /ɪ/, /ə/ and /ʊ/ (Raza, 2009). Some studies have also claimed that all long oral vowels have a nasalised counterpart (Raza, 2009; Rauf, 1997; Bokhari, 1985). These studies vary not only in the number of vowels, but also the inconsistent symbols used for these vowels. In addition, most of the studies discussed here did not provide any experimental evidence based on acoustic, phonetic and/or phonological investigation. Given the disagreement across the literature and the prevalence of cross-language and regional influences, the present study focussed on the vowel system of Urdu as spoken in Punjab, Pakistan, and largely used the inventory proposed by Saleem et al. (2002), which to-date is the most comprehensive and legible, with slight modifications to address pedagogical goals of the study which are discussed in detail in Chapter 4 and Chapter 5.

### 1.2.2 Diphthongs

Most of the literature (Bokhari, 1991; Waqar and Waqar, 2002; Khurshid, Usman and Butt, 2003; Bhatti and Mumtaz, 2016) agrees that phonologically Urdu does not have diphthongs. However, this claim is disputed by the survey and perception experiments of Waqar and Waqar (2002), Sarwar, Ahmed and Tarar (2003) and Bhatti and Mumtaz (2016) who argue that, phonetically, diphthongs are found in Urdu.

With regard to Urdu phonological processes, Wali (2001) argues that the deletion of /h/ and /ʔ/ in word final position results in the elongation of the preceding short vowel. For example, /t̪əfrɪh/ "*break-time*"is pronounced as [t̪əfriː] and /bəʔd̪/ "*after*" is pronounced as [bɑːd̪] (Wali, 2001), and this deletion of /h/ and /ʔ/ in some cases may generate a diphthong, e.g. /məsɑʔɪl/ "*problems*" is pronounced as [məsɑɪl] (Wali, 2001:256). Waqar and Waqar (2002), Farooq and Mumtaz (2016) and Bhatti and Mumtaz (2016) report that deletion of one of the three consonants /ʔ/, /j/ and /v/ in a disyllabic word results in a diphthong, for example /zejɑːd̪ɑː/ "*excessive*" is pronounced as [zeɑːd̪ɑː] and /t̪ejɑːriː/ "*preparation"* is pronounced as [t̪eɑːriː]. Khurshid, Usman, and Butt (2003) reported the possibility of diphthongs and triphthongs in Urdu and claimed that out of a list of 37 possible diphthongs in Urdu, 18 were identified as diphthongs by 20 native speakers of Lahori Urdu. A brief survey of the literature on Urdu diphthongs is presented in Chapter 3.

## 1.3. Pakistani English

The second half of this thesis deals with second language (henceforth L2) perception, in particular the perception of Standard Southern British English (SSBE) vowels by Punjabi

Urdu speakers from Pakistan. Therefore, it is important to understand the status of English in Pakistan. A brief presentation is found below.

### 1.3.1 Background - Historical and Political

Although Urdu is the *national* language of Pakistan (see Section 1.1), English is the official language of Pakistan (Mahmood, 2004 cited in Farooq, 2014), as it was before the partition of India in 1947. "English is a lingua franca and medium of communication in affluent classes of the society, civil and military bureaucracy, official correspondence, and English is also the language of courts and the constitution of Pakistan" (Bughio, 2013:7). According to Mahboob (2017), due to *pragmatic* and *political* reasons English has remained the medium of instruction particularly in higher education and has played a significant role in educational context of Pakistan (see Mahboob, 2017:71-92 for a detailed review). The British colonial powers adopted an approach whereby the teaching of English was restricted to elites in India in order to maintain class and status divisions: "The current education system in Pakistan is a legacy of the British colonial powers. The British started two streams of education: English-medium and vernacular-medium, to serve their own political ends" (Rahman, 1996; Ramanathan, 2005 cited in Shamim, 2008:236). Due to the dual language education policies of the government (i.e. English for the elite and Urdu for the masses), it is very complicated to define the status of English in Pakistan.

Various Pakistani governments since independence in 1947 have tried to replace English with Urdu, but it has not been possible due to a lack of consistent education policies, training and resources. According to (Shamim, 2008:236), Urdu was promoted as a national language to unite the nation; however, it provoked some regional ethnic issues and as a result Bangladesh (then East Pakistan) seceded in 1972. General Zial-ul-Haque's government (1977-1988) tried to popularise Urdu as a national and official language and medium of instruction in schools; however, this enforcement left the country with a rather complicated situation, where English became the official language and enjoyed the status of "language of the elite", prestige, and modernity. In 1981 and 1982, the government recommended the use Urdu of as a medium of instruction and introduced either English or Arabic as an additional language from the age of 11.The language policies over the years, without any proper planning and implementation, have played havoc with the education system in government schools; whereas the private English medium schools and colleges have been insulated from these inconsistent policies (see Mahboob 2002:15-39 and 2017:71-92 for a detailed review).

The 'Urdu only' policy, which started in 1977, only exacerbated class divisions, since the privileged still had access to English unlike the poor:

> "Most high paying jobs in the private sector required English. Graduates from English-medium schools met this requirement and were hired, while graduates from non-English medium or Vernacular schools had to struggle to find decent jobs. Thus, there was (and is) discrimination in the higher circles of the society against people with a non-English medium education and against people with insufficient skills in English language" (Mahboob, 2002:16).

Many efforts have been made over the years to replace English with Urdu in Pakistan; however, an unstable and ever-changing political situation, has led to poor language policies and a big divide between the public and private sector education systems. Nowadays, English is considered a tool for social, economic, individual and national development. The pro-English stance of various Pakistani governments since 1989 seems to be in contrast with the majority of post-colonial states, where "small English-speaking elites have continued the same policies of the former colonizers, using access to English language education as a crucial distributor of social prestige and wealth" (Pennycook, 1995:40 cited in Shamim, 2008:244). “No English, no future” is the trend in vogue for the past few decades (Mahboob, 2002). In Pakistan, learning and speaking standard British English not only guarantees a better job, but also indicates a better social status (Mahboob, 2002). Therefore, Pakistanis strive to learn and speak standard British English, and this is the standard used in the curriculum of English medium schools. According to Rahman (2015:10),

> “Pakistani writers with international reputations … use British Standard English (BSE) with some indigenous lexical items and idiomatic turns of speech for artistic reasons. Pakistani writers do not manifest as much concern with the creation of a genuinely indigenous English”.

In his investigation of attitudes towards English in Pakistan, Baumgartner (1995) reported an emerging variety of English in Pakistan. However, Rahman (2015) did not like the use of the term “Pakistani English”, arguing it is “inadequate” and “incorrect” because his investigation of English in Pakistan was “... dealing with Punjabi English, Pashtun English, Sindhi English, Baloch English and Urduized English” (Rahman, 2015:10).

To summarise, English is the official language in Pakistan and is learned in schools. Private English medium schools are far better than public schools, where the medium of instruction has been changed from English to Urdu and Urdu to English by almost every government since 1947. However, the private English medium schools are autonomous, in the sense that, “Public sector schools follow the syllabus prescribed by the government whereas private schools follow their own choice for the selection of curriculum within the prescribed subjects” (Bughio, 2013:421). Therefore, unstable government policies do not affect the education system in the private sector. Since the majority of the population speaks at least two languages and each regional language has further dialects, the English language in Pakistan shows strong influence from the local languages. Although, some argue about the existence of a distinct Pakistani English (PE) dialect, which is hugely influenced by the phonological inventory of the native languages, e.g. Punjabi, Urdu, Sindhi, Saraiki, Balochi, Pashto and other regional languages, very little or no work has been done so far to investigate the influence of local languages on the perception and production of English in Pakistan. The limited available literature on the phonetics and phonology of Pakistani English is discussed in Section 4.1.

## 1.4 Second Language Perception

### 1.4.1 Background

As discussed above, English is the official language in Pakistan and is used as a lingua franca in higher education and formal communication in all aspects of life across Pakistan. English in Pakistan is learned in schools and is heavily influenced by the socio-economic, geographic, educational and linguistic background of the speakers (Mahboob and Ahmar, 2004). Mainstream second language learning models, for instance the Speech Learning Model (SLM; Flege, 1995, 2003) and Perceptual Assimilation Model (PAM; Best, 1995) have been tested for the last three decades on learners of English who received native English input and do not speak more than one mother tongue. The situation in Pakistan is quite different. Most people in Pakistan speak more than two languages and the English language input they receive is far from native. Keeping in view the status and role of English in Pakistan (as discussed in Sections 1.3 and 4.1), the present study investigated the perception of Standard Southern British English (SSBE) vowels by the Punjabi-Urdu speakers from Punjab; mainly from Lahore and its suburbs.

According to Scovel (1969) and Patkowski (1990), it is difficult to learn and speak a second language (henceforth L2) without a strong foreign accent especially after the "critical period" that is usually before the age of 15. It is believed that learners speak their L2 with a strong and identifiable foreign accent, possibly caused by inaccurate speech perception (Flege, 1992; Flege et al., 1995; Flege, 1995; Flege et al., 1999). Flege also proposes that L2 speech can be learned throughout one's lifetime, and learners can develop native-like accuracy if they are given sufficient native input/exposure to the target language (Flege, 1987; Flege, 1995; Flege, et al.,1995). In addition, according to Best and Strange (1992), the perception of vowels and consonants remains flexible even after the critical period. Some L2 perception studies report that linguistic experience not only influences the perception of L2 sounds but also the L1 sounds (Flege et al., 1997; Boomershine, 2013) and leads to the development of an interlanguage phonological space, which means bilinguals do not perform like monolinguals of either language. Research also shows that due to phonetic category assimilation or dissimilation, the vowels produced by bilinguals and monolinguals differ in their spectral and /or temporal quality (Flege, Schirru and MacKay, 2003).

Most of the research on L2 perception and production has been conducted on the predictions of L2 perception models; in particular the Speech Learning Model (SLM; Flege, 1995 and 2003) and Perceptual Assimilation Model (PAM; Best, 1995) and its extension to second language learning, i.e. PAM-L2 (Best and Tyler, 2007). Both of these models propose that both phonetics and phonology of L1 and L2 play a role in the perception and production of L2 sounds.

In addition to SLM, PAM and PAM-L2, Phonological Inference Model (PIM; Brown, 1998) has also been the focus of L2 research. Brown (1998) proposed a pure phonological model (i.e. Phonological Inference Model; PIM) which is based on distinctive features. According to Brown (1998), it is not the individual phonemes which are perceived, but the absence or presence of certain distinctive features in L1 shapes the perception and production of L2 sounds. For instance, the evidence he used to support this theory was the differences in the perception of English /ɹ/-/l/ by Chinese and Japanese learners of

English. As compared to Japanese, Chinese learners did not have any difficulty perceiving these two phonemes because the distinctive features, *retroflex* and *coronal*, are present in Chinese but absent in Japanese. However, this model was challenged by other L2 research, where its prediction did not bear out for the perception of French /i/-/y/ by native speakers of English (Colantoni, Steele and Escudero, 2015). In addition to the criticism as this model is purely based on distinctive features, it was not explored further for the present study.

In the early 2000s, the Second Language Linguistic perception (L2LP; Escudero, 2005) model was developed. The predictions proposed by L2LP are mainly based on the acoustic comparison of the L1 and L2 vowels. According to this model, the acoustic properties of L1 sounds have a strong influence on the perception of L2 sounds. Some of the predictions of this model are similar to PAM. Both PAM and L2LP predict that L2 contrasts that are not found in L1 are difficult to discriminate by L2 learners. L2 vowel contrasts that are assimilated to a single L1 vowel are labelled *single-category* (SC) assimilation in PAM and this assimilation type is treated as a *new scenario* in L2LP. L2 vowel contrasts that are assimilated to two L1 vowel categories are labelled *two-category* (TC) assimilation in PAM and this assimilation type is treated as a *similar scenario* in L2LP (Escudero, 2005; Escudero et al., 2014).

The present study focussed on some of the hypotheses proposed by SLM (Flege, 1995) and the predictions proposed by PAM and PAM-L2 (Best and Tyler, 2007) and L2LP (Escudero, 2005; Escudero et al., 2014). SLM and PAM-L2 overlap but are different in many ways. SLM mainly deals with individual phonemes whereas PAM makes predictions with regard to contrasting pairs of phonemes. According to Flege (1995), errors in production of speech segments are mainly based on errors in the perception of speech segments. SLM relies on the importance of experience, exposure and age of learning in L2 perception and production; however, PAM proposes that listeners focus on articulatory gestures of the L2 phonemes and try to assimilate those to similar L1 phonemes. Thus, the degree of similarity and/or differences of L2 phones to L1 phonemes indicate the difficulties in L2 learning.

Most of the previous studies explored the predictions of these models by second language learners of English who speak Spanish, Italian, French, German, Dutch or Japanese as their first language. Very little is known about how native Punjabi-Urdu speakers perceive and process Standard Southern British English (SSBE) vowels, and this study hopes to begin to shed light on this point. This will be the first study of its kind to investigate the perception of SSBE vowels by native speakers of Punjabi-Urdu. As discussed in Section 1.3, the linguistic context in Pakistan is quite different from previous studies in cross-linguistic speech perception.

L2 learning theories suggest that after the critical period it is difficult to achieve a native-like proficiency in L2 (Patkowski, 1989; DeKeyser, 2000). SLM emphasizes the importance of native input for a better perception and production of L2, and PAM predicts that the perceptual assimilation of L2 sounds to L1 categories can help to identify the difficulties faced by L2 learners in perception and production. Punjabi-Urdu speakers learn English in a non-native and multilingual context, and the current theories of L2 learning and perception do not account for such a scenario. Hence it is quite possible that existing theories are not applicable and cannot predict L2 learning difficulties by mere patterns of perceptual assimilation of SSBE vowels to Urdu vowel categories.

The experimental evidence from the literature and the predictions for the perception of SSBE vowels based on these models are discussed in detail in Section 4.2.

## 1.5 Aims and Objectives

The present study investigates the acoustic and phonetic properties of Urdu vowels (monophthongs and diphthongs) and the perceptual assimilation of SSBE vowels by Punjabi-Urdu speakers who learn English at school. According to L2 speech learning theories (SLM, PAM, PAM-L2 and L2LP), L2 learners face difficulties in the perception and production of L2 segments (i.e. vowels and consonants) that are either new (do not occur in their L1) or are very close (phonetically) to L1 segments. In contrast to previous studies (e.g. Strange et al., 2007; Gilichinskaya and Strange, 2010), the listeners were not naïve learners of SSBE; however, it must be noted that they learn English in Pakistan and as a result the input (especially speaking and listening) they receive has very little to do with Received Pronunciation (RP) or SSBE, except for some audio-visual materials used in classrooms, and TV, films and other media platforms in everyday life. The present study investigates if the predictions of PAM, L2LP and SLM are applicable to the perception of SSBE vowels when L2 (English) is learned from a very young age in a non-native context and used as lingua franca in everyday life; hence the L2 users are not inexperienced learners.

## 1.6 Overview of the Thesis

**Chapter 2** presents a brief survey of the literature on Urdu monophthongs followed by a production study on Urdu vowels. Based on the findings from this study, a proposal for the vowel system of Urdu as spoken by Punjabis is presented.

**Chapter 3** first presents a brief survey of the literature on Urdu diphthongs, followed by a production study on Urdu diphthongs. Based on the findings from this, the chapter discusses the status of diphthongs in Urdu.

**Chapter 4** has two main sections: the first section gives a brief rationale for the perception study, followed by description of Pakistani English, especially phonetics and phonology, including a brief comparison with Indian English. The second section presents survey of the literature on L2 perception, especially the perception of English vowels by speakers of other languages, followed by the main objectives and predictions for the perception study.

**Chapter 5** presents the experiment design and procedures for a perception experiment of Standard Southern British English (SSBE) vowels by Punjabi-Urdu speakers and reports on the findings.

**Chapter 6** presents the experiment design, procedures and findings for an auditory free classification experiment of Standard Southern British English (SSBE) vowels by Punjabi-Urdu speakers.

**Chapter 7** summarises the results from both parts of this thesis, proposes implications for the teaching and learning of English in Pakistan, and presents limitations of this work and directions for future research.

## Chapter 2

### Urdu Monophthongs

This chapter reports on a production experiment investigating the phonetic and acoustic properties of Urdu monophthongs by native speakers of Urdu from Punjab, Pakistan. Literature is sparse for Urdu linguistics, and what is available is often contradictory. There are many disagreements about the phoneme inventory of Urdu as shown in Chapter 1, and this is particularly evident with respect to the vowel inventory. However, there is no study in the literature that shows a comprehensive view of phonetic and acoustic properties of Urdu vowels. It was anticipated that defining a standard Urdu dialect would be difficult due the heavy influence of local languages (as confirmed in Section 2.2), and due to the small number of native (L1) speakers of Urdu (7.57%) relative to the overall Pakistani population of 207 million (Pakistan Bureau of Statistics, 2017). For this reason, the present study focuses on the investigation of Urdu as spoken by Pakistanis whose L1 is Punjabi and who reside in Punjab. Specifically, this study is an attempt at determining the phonetic and phonological system of Urdu vowels as spoken in Punjab, Pakistan. An understanding of the phonetic and in particular phonological system of Urdu (phonemic contrasts) can help to better understand the perception of SSBE vowels by Punjabi-Urdu speakers (see Section 1.4, Section 4.4 and Section 5.1 for further details). Before presenting the experimental design, analysis and results, an overview of the literature on Urdu vowels, expanding on the brief overview given in Section 1.2.1, follows.

### 2.1 Background - Urdu Vowels

Recall from Section 1.2.1 that most studies agree on seven long oral vowels: /iː/, /eː/, /æː/, /aː/, /ɔː/, /oː/, and /uː/; and three short vowels /ɪ/, /ə/, and /ʊ/ in Urdu (Saleem et al., 2002; Ali and Hussain, 2010). Ali and Hussain (2010) also reported seven long nasal vowels. However, according to Khan and Alward (2011), there are eight pure vowels: /i/, /iː/, /u/, /uː/, /ə/, /ɑ/, /e/, and /o/. Similarly, Raza (2009) reported 11 oral vowels, i.e. eight long vowels (/i/, /e/, /ɛ/, /æ/, /ɑ/, /ɔ/, /o/ and /u/), three short vowels (/ɪ/, /ə/ and /ʊ/) and seven nasal vowels (/ĩ/, /ẽ/, /æ̃/, /ɑ̃/, /ɔ̃/, /õ/ and /ũ/).

In an investigation of the effects of lexical stress on the phonetic properties of Urdu vowels and consonants, Hussain (1997) reported duration, fundamental frequency and mean F1 and F2 frequencies of 9 monophthongs /i/, /ɪ/, /e/, /æ/, /ə/, /a/, /o/, /ʊ/ and /u/. Hussain (1997:66) recorded data from seven speakers (three females and four males including the author) who were native speakers of Urdu from Pakistan. Three participants were visiting Chicago and the other four participants were students at the Northwestern University and had been living in the United States for at least six months. The main focus of this research was to investigate the influence of stress on the duration of vowels; hence he extracted spectral and temporal values from stressed and unstressed syllables in disyllabic words. Although he mentioned seven long vowels, he did not include /ɔ/ in his stimuli and analysis: "among the long vowels /ɔ/ shows a very limited usage" (Hussain, 1997:148). His results showed that stressed syllables exhibit longer duration for both short and long vowels as compared to unstressed syllables. His findings are shown in Table 2. 1.

Hussain (1997) did not give any justification for the selection of vowels except for that Kachru (1987) proposed seven long oral and three short oral vowels. In addition, Hussain (1997) disagreed with Kachru's (1987) front open-mid vowel /ɛ/ and considered it closer to /æ/ in quality.

In his book *Phonology of Delhi Urdu*, Rauf (1997:80) argues that Urdu has 8 orals vowels: 5 long (iː, eː, uː, oː, aː), and 3 short (I, U, a - interpreted here in IPA symbols by ɪ, ʊ, ə, respectively); and 8 nasal vowels: 5 long (ĩː, ũː, ẽː õː ãː), and 3 short (ĩ, Ũ, ã - in IPA symbols ɪ̃, ʊ̃, ə̃, respectively).

Table 2. 1: Duration of Urdu vowels in stressed and unstressed syllables by Hussain (1997)

| Vowel | Duration (ms) Stressed syllables | Duration (ms) Unstressed syllables |
|---|---|---|
| iː | 103 | 100 |
| ɪ | 55 | 48 |
| e | 104 | 92 |
| ɛ* | | |
| æ | 115 | 105 |
| ɑː | 114 | 110 |
| ɔː* | | |
| o | 98 | 84 |
| ʊ | 49 | 46 |
| uː | 87 | 80 |
| ə | 50 | 47 |
| ʌ* | | |

*/ɛ/, /ɔ/ and /ʌ/ were not included in this investigation

According to Bokhari (1985) every short vowel has a long counterpart and vice versa. Bokhari stated that there are 24 vowels in Urdu,

> "…there are, in all, 24 vowels in Urdu with two classifications, viz., (1) qualitative, further subdivided into (a) 14 plain and (b) 10 nasalised vowels and (2) quantitative, further subdivided into 12 short and 12 long vowels" … we have, in all 14 plain and 10 nasalised vowels because nasalisation is not permissible after /o/ /oː/ /ɛ/ and /ɛː/" (1985:5).

Wali (2005) states that long vowels in Urdu are nasalised if the following sound is an alveolar nasal /n/. However, his example shows the opposite, i.e. /mɑːn/ "*pride*" and /mɑ̃ː/ "*mother*" are two different words in Urdu with different meanings, and they differ in one sound, an oral vs. nasal vowel. In addition, the oral vowel in /mɑːn/ "*pride*" is followed by an alveolar nasal /n/, but it is not nasalised.

In one of the most recent studies, Farooq (2014) concluded that the Urdu phonetic inventory has 66 phonemes = 43 consonants + 13 Oral vowels (= 7 long vowels + 3 medial vowels + 3 short vowels) + 10 nasal vowels (= 7 long + 3 short). She investigated the effects of regional languages (five major provincial languages: Punjabi, Saraiki, Pashto, Balochi and Sindhi) on the pronunciation and production of four peripheral Urdu vowels /iː/, /æː/, /aː/ and /uː/. Different district/city names were used as corpus and data were collected over the telephone from 30 participants per language group. Thus, she analysed six accents of Urdu in 136 districts of Pakistan, and gave some acoustic and statistical analysis. However, her acoustic analysis shows only a range of formant frequencies for the four vowels: /i/, /æː/, /aː/, and /uː/, which are given as follows:

> "In Urdu language, /i:/ vowel has lower F1 and higher F2 values (200-2400 Hz) than /a:/ vowel; /æ:/ vowel has slightly lower F1 value and higher F2 value (700-1700 Hz) than /a:/ vowel, /a:/ vowel has upper F1 and lower F2 values (700-1100 Hz) than /i:/ vowel, /u:/ vowel has slightly upper F1 and lower F2 values (300-700 Hz) than /i:/ vowel" (Farooq, 2014:99-114).

Farooq (2014) reported that F1 and F2 frequencies showed significant effects of the mother tongue on the pronunciation of the Urdu vowels. As a result, each group differed in their pronunciation. Further, she reported that the formant frequencies of the Urdu vowels produced by native speakers of Punjabi were a close match to the ones produced by native speakers of Urdu. However, she did not provide any further details about the background of the native speakers of Urdu. It seems unlikely that all these speakers were native since only 7.5% of the population speaks Urdu as their first language and most of this population is assumed to reside in the federal capital, i.e. Islamabad (see Section 1.1). Farooq (2014:22) has listed Islamabad among 25 districts of Khyber-Pakhtunkhwa (NWFP) region. In this region, 35 minor languages are spoken and Pashto is the mother tongue of the majority of the population (Farooq, 2014; Rahman, 2002). In addition, Farooq (2014) did not provide any background information of speakers whose first language was not Urdu, such as age group, education level and whether they were born and raised in urban areas or rural areas.

Overall, some of the disagreement in the Urdu vowel inventory can be attributed to the different accents of Urdu due to the influence of local and regional languages. Another reason could be lack of experiments and acoustic analysis. For instance, in one study the data was collected over the phone (Farooq, 2014) and others (Bokhari, 1985; Saleem et al., 2001; Ali and Hussain, 2010) did not report any acoustic analysis. In all these studies, a number of Urdu words have been used regardless of whether they are monosyllabic, disyllabic or multisyllabic. For example, Farooq (2014) used district names which are very complex proper nouns and have complex linguistic and geographical origins, and therefore it is likely that speakers will pronounce those names differently. In most other cases the lack of information on speakers' linguistic, social and educational background, as well as incomplete or no information on the experiment design, stimuli, execution and analysis made these studies unreliable and difficult to compare with each other.

In addition to these disagreements on the number of vowels, there are some specific vowel contrasts which appear to be controversial in the literature and need further investigation. A brief overview of those vowel contrasts is given below.

There is major disagreement on the properties of /ɛ/ and /æ/. For instance, Saleem et al. (2002) states that the sounds /ɛ/ and /æ/ are allophones of the same phoneme, however Fatima and Aden (2003:74) state that /ɛ/ is a long and /æ/ is a short vowel or vice versa; Raza (2009) reported /ɛ/ and /æ/ as distinct vowels. However, neither Saleem et al. (2002) nor Fatima and Aden (2003) and Raza (2009) have given enough evidence to support their claims. Kachru (1990) used the symbol /ɛ/ for a front open-mid vowel in his impressionistic account of Hindi-Urdu vocalic inventory; however, Hussain (1997) used /æ/ in his description, claiming that the sound in Urdu is closer to vowel /æ/ than /ɛ/. However, none of these authors has given any acoustic and phonetic or phonological evidence to support their claims.

In addition, the status of /ɔ/ and /o/ appears to be unclear. The literature unanimously reports two back open and close-mid back vowels /ɔ/ and /o/ respectively (Saleem et al., 2002; Hussain, 1997; Kachru, 1987, Bokhari, 1985). However, none of these studies have provided any acoustic and phonetic or phonological evidence to support this claim. Hussain (1997) in his unpublished dissertation, listed these two vowels in the Urdu vocalic inventory; however, he excluded /ɔ/ vowel from his stimuli without any explanation.

Lastly, most of these studies reported /ə/ as the only central vowel, and none considered the presence of the vowel /ʌ/. The stimuli words used for the description of /ə/ are always disyllabic, where this vowel usually occurs in an unstressed syllable. Therefore, it was noted as worth investigating if the stressed monosyllabic words also have this sound or /ʌ/. Some of the main issues in these studies are discussed below.

Apart from Hussain (1997) and Farooq (2014), none of these studies provide any information on the speakers' socio-economic, linguistic background. They do not present a formal acoustic analysis with reference to the formant frequencies as used in other studies in other languages, such as English. For example, Wells (1962), Deterding (1997), Hawkins and Midgley (2005), and Ferragne, and Pellegrino (2010) presented acoustic, phonetic and statistical analysis of English vowels by embedding the target vowels in a hVd context. Although Hussain (1997) provided a detailed information about stimuli and speakers in his unpublished dissertation, the results are not comparable with the present study for the following reasons: (a) six speakers were reported as native speakers of Urdu from Lahore, , and one of the speakers was from Karachi and probably also spoke Sindhi or some other local language; (b) the set of Urdu vowels was recorded in disyllabic or trisyllabic words to investigate the stress patterns in Urdu. As a result, each vowel appeared in a stressed and unstressed syllable. Despite these fundamental differences, since there are no other studies available, the duration and F1 and F2 of the present study will be compared with those reported by Hussain (1997).

The studies on Urdu vowels discussed above are neither reliable nor comparable (especially with regard to experimental design and methods of analysis) with similar studies in other languages, such as English. Also, as discussed in Section 1.2.1, these studies are not reliable particularly because it is not clear which variety of Urdu is investigated and the symbols are borrowed from English rather than the IPA alphabet, so it is not clear if authors used them in the same way. In short, these studies vary not only in the number of vowels, but also the inconsistent symbols used for these vowels. The next section presents a brief review of the standard methods for acoustic and phonetic investigation of vowels.

### 2.1.1 Aims and Objectives

As mentioned in Chapter 1, given the disagreement across the literature and the prevalence of cross-language and regional influences, the present study focused on the vowel system of Urdu as spoken in Punjab, Pakistan. The inventory proposed by Saleem et al. (2002) (as discussed in Section 1.2.2.1) is used with slight modifications to address pedagogical goals of the study which are discussed in detail in Chapters 4 and 5.

The aims of the production study were to answer the following research questions:

a) Is there a consistent difference between short and long vowels in Urdu?
b) Do the vowels whose status has been disputed have different qualities; this applies in particular to front vowels /e/, /ɛ/ and /æ/, and back vowels /o/ and /ɔ/;

c) Does Urdu as spoken in Punjab have two central vowels, /ə/, /ʌ/?

d) Does Urdu as spoken in Punjab have diphthongs?

The analysis and results pertaining to questions (a)-(c) are presented in this chapter; the analysis and results pertaining to question (d) are presented in Chapter 3.

**2.2 Pilot Study**

Five participants (two females and three males) from varying backgrounds were recruited with the help of one of the administrators of the mosque at Canterbury, Kent for the pilot study. By providing evidence for different varieties of Urdu, the pilot study helped to refine the present study. For example, participants' linguistic, educational, social and regional background (as given in Table 2. 2) helped to highlight the regional and dialectal differences on the pronunciation and quality of Urdu vowels. For example, the word [peːʈ] "*stomach*" was pronounced as [piːʈ], homophonous with "*to beat*", by a Sindhi-Urdu speaker, and [peːʈ] by a Punjabi-Urdu speaker. Similarly, the Sindhi-Urdu and Punjabi-Urdu speaker's pronunciation varied for other words, e.g. [pʌ̃t̪ʰ] "*join*" was pronounced as [pĩt̪ʰ], and [sə̃d̪] "*certificate*" was pronounced as [sənəd̪].

The differences in vowel quality were evidence of regional dialects and influences of speakers' first language. The female speaker from Lahore was adamant that her first two languages were Urdu and English and that she rarely spoke Punjabi. On the other hand, the female speaker from Faisalabad had a very strong regional accent in all her utterances. In addition, she was the least qualified/educated among all five participants, which indicated that educational background also has an influence on the speakers' pronunciation of Urdu.

The Pashto-English speaker, who learned Urdu in a school in England, aspirated all the stops in Urdu words even those that should have been unaspirated. Urdu distinguishes aspirated stops from unaspirated stops hence the strong influence of English was very evident in the pronunciation of Pashto-English speaker. These variations are an example of one of the major criticisms of previous studies in the literature, as detailed in Sections 1.2 and 2.1, further justifying the choice of the present study to restrict to the Punjabi-Urdu dialect.

Table 2. 2: Participants' socio-linguistic and education background in the pilot study

| Speaker | Gender | Age range | Regional Background | First language | Education |
|---|---|---|---|---|---|
| S1 | Male | 45-55 | Faisalabad, Punjab | Punjabi | PhD |
| S2 | Male | 45-55 | Sukkur, Sindh | Sindhi | PhD |
| S3 | Female | 35-45 | Faisalabad, Punjab | Punjabi | B.A. |
| S4 | Female | 25-35 | Lahore, Punjab | Urdu | PhD student (second year) |
| S5 | Male | 18-25 | British born Pakistani, his parents emigrated from Peshawar, Khyber Pakhtunkhwa | Pashto and English | Undergraduate (stage 2) |

**2.3 Main Study**

**2.3.1 Methods**

**2.3.1.1 Speakers**

Twenty-six speakers, thirteen males and thirteen females from Punjab, Pakistan, took part in the experiment. All participants were aged between 18-84 with the median age range of 35-44. Twenty-one of the 26 participants were in the age range of 35-44 or 45-54. All participants were multilingual and spoke at least Punjabi, Urdu, and English. They were from different parts of Punjab and spoke different dialects of Punjabi. They all belonged to elite or upper middle class and were highly educated, except for one female and two male participants who had only secondary school level qualifications from Pakistan. All participants held British citizenship and had been living in England for 5 to 25 years at the time of the recording; however, they all used Urdu regularly. Most of the participants were couples and worked in the same profession (GPs and senior medical consultants) with a few exceptions: an Islamic scholar and his wife, two students from the University of Kent, and a housewife. The details are given in Table 2. 3.

Twenty-five out of twenty-six participants were recruited with the help of connections with the then-manager of the Boston, UK mosque. One participant volunteered in response to an advertisement placed in the Canterbury mosque. The participants were contacted in advance and the time and place of meetings were arranged at their convenience. For instance, some of the participants were able to go to the linguistics lab at the University of Kent, Canterbury; others went to a mosque in Boston, UK or were visited by the experimenter at their homes in Boston, UK.

Before the recording session started, participants filled in Language Experience and Proficiency Questionnaire (LEAP-Q) by Marian, Blumenfeld, and Kaushanskaya (2007) on paper (see a sample of this questionnaire in Appendix 2A), signed the consent forms and read the instructions and details of the experiment. All this information was written in English; however, the experimenter and the participants conversed in Urdu and all their questions were answered before the recordings began. None of the participants reported any history of speech and hearing or sight disorder. They were all familiar with Urdu script, having been taught as a compulsory subject from primary to secondary school. Their responses to the questions about linguistic background are given in the Table 2. 3.

#### 2.3.1.2 Materials

As noted, the inventory proposed by Saleem et al. (2002) has been used with slight modifications. This selection includes the controversial vowels, as discussed in Section 2.1. Thus, 25 vowels were selected for investigation and are given in Table 2. 4.

Table 2. 3: Participants' socio-linguistic and education background

| Speaker | Gender | Age range | Regional Background | First language | Second Language | Education |
|---|---|---|---|---|---|---|
| S1 | Male | 35-44 | Multan, Punjab | Punjabi | Urdu | PhD |
| S2 | Male | 35-44 | Lahore, Punjab | Punjabi | Urdu | Professional Training |
| S3 | Male | 45-54 | Multan, Punjab | Punjabi | Urdu | Doctoral Fellowship |
| S4 | Male | 45-54 | Multan, Punjab | Punjabi | Urdu | Professional Training |
| S5 | Male | 35-44 | Burewala, Punjab | Urdu | Punjabi | PhD |
| S6 | Male | 35-44 | Lahore, Punjab | Urdu | Punjabi | Professional Training |
| S7 | Male | 45-54 | Dera Ghazi Khan, Punjab | Urdu | Punjabi | Graduate |
| S8 | Male | 45-54 | Liaquat Pur, Punjab | Punjabi | Urdu | MBBS (FRCS) |
| S9 | Male | 25-34 | Lahore, Punjab | Punjabi | Urdu | Masters |
| S10 | Male | 25-34 | Lahore, Punjab | Urdu | Punjabi | Graduate |
| S11 | Male | 45-54 | Lahore, Punjab | Punjabi | Urdu | MBBS (FRCS) |
| S12 | Male | 45-54 | Lahore, Punjab | Punjabi | Urdu | MBBS (FRCS) |
| S13 | Female | 35-44 | Multan, Punjab | Punjabi | Urdu | Graduate |
| S14 | Female | 35-44 | Sahiwal, Punjab | Urdu | Punjabi | Masters |
| S15 | Female | 35-44 | Khushab, Punjab | Urdu | Punjabi | PhD |
| S16 | Female | 45-54 | Islamabad | Urdu | English | MBBS |
| S17 | Female | 45-54 | Rawalpindi, Punjab | Urdu | Punjabi | Professional Training |
| S18 | Female | 45-54 | Lahore, Punjab | Urdu | Punjabi | Other (MBBS) |
| S19 | Female | 55-64 | Lahore, Punjab | Punjabi | Urdu | High School |
| S20 | Female | 35-44 | Lahore, Punjab | Punjabi | Urdu | Doctor, MBBS |
| S21 | Female | 45-54 | Lahore, Punjab | Punjabi | Urdu | Masters |
| S22 | Female | 35-44 | Lahore, Punjab | Punjabi | Urdu | MBBS |
| S23 | Female | 18-24 | Lahore, Punjab | Urdu | Punjabi | College |
| S24 | Female | 75-84 | Lahore; however, born in Delhi (before partition) | Punjabi | Urdu | Masters |
| S25 | Female | 35-44 | Lahore, Punjab | Punjabi | Urdu | Masters |
| S26 | Female | 45-54 | Lahore, Punjab | Punjabi | Urdu | Masters |

These vowels are presented in square brackets because their phonemic status is unclear. The selected vowels were as follows: 12 oral vowels (7 long [iː], [eː], [æː], [ɑː], [ɔː], [o] and [uː] and 5 short [ɪ], [ɛ], [ʌ], [ə] and [ʊ]); 7 nasal vowels ([ĩ], [ẽː], [ə̃], [æ̃], [ɑ̃ː], [ɒ̃ː] and [ũː]); and 6 diphthongs ([ɑe], [oe], [ɑʊ], [ɪɐ], [eɐ] and [ʊɑ]). The six diphthongs were selected by the author of this thesis purely due to their expected resemblance to English diphthongs, which were expected to be helpful in the perception of SSBE vowels, discussed in Chapters 3 and 4.

Table 2. 4: Urdu vowels

| | | |
|---|---|---|
| Long vowels | [iː] [eː] [æː] [ɑː] [ɔː] [o] [uː] | 7 |
| Short vowels | [ɪ] [ɛ] [ʌ] [ə] [ʊ] | 5 |
| Nasal vowels* | [ĩ] [ẽː] [ə̃] [æ̃] [ɑ̃ː] [ɒ̃ː] [ũː] | 7 |
| Diphthongs | [ɑɪ] [ɔɪ] [ɑʊ] [ɪə] [eə] [ʊə] | 6 |
| Total vowels | | 25 |

*The nasal vowels were not analysed in this thesis.

The 25 vowels were embedded in 30 monosyllabic minimal or near minimal pairs. These test words (glossed in Table 2.5) were embedded in two types of sentences (as shown in Appendix 2B):

1. A standard carrier phrase (CP) of the form "I will say ___ once" (e.g., [mɛ̃ ɪsɛː b̲ʌ̲̃d̲̪ eːk bɑːr kəhõ giː]). This structure made it possible to have a vowel (/ɛ̃/ or /æ̃/) at the end of the preceding word and /eː/ at the beginning of the following word.
2. Longer and more varied full sentences (FS) in which these same words were likely to naturally occur; e.g., [d̪ərɪɑ keː kɪnɑːreː b̲ʌ̲̃d̲̪ bɑ̃ːd̪ dɪə geɑ] "*a wall was built on the river bank*".

The speaking mode in the CP structure can have a significant effect (Harris and Umeda, 1974), hence why the same vowels were recorded in natural full sentences (FS), so that the vowels in the two types of speech can be tested and compared for any differences in the quality, especially with regard to citation speech or connected speech. According to Deterding (1997) vowels from connected speech represent more natural data than the artificial data from the specially articulated citation speech.

Five additional words were used to resolve homographic ambiguity, as is explained in more detail below. In the pilot study it was observed that participants mispronounced and confused some words in the standard carrier phrases (CP), and some took long pauses (especially in their first recording of the five repetitions) to identify the correct pronunciation. This problem arose for two main reasons. Firstly, in the standard carrier phrases the words carrying the target sounds were out of context (e.g. "I will say ---------- once"). Secondly, in Urdu many words are homographs, so their pronunciation mainly depends on the context (see Section 1.1). Diacritics helped participants to pronounce the words correctly; however, the number of mispronounced words remains high (see Appendix 2C for the errors in production of the given stimuli).

To avoid this problem, the following additional words were used to avoid the ambiguity:

(1) [bʌd̪] "*bad*" was pronounced as [bɪd̪] "*new*" and vice versa so [bɪk] "*to sell*" was added to the list

(2) [puːt̪] "*son*" was pronounced as [pɔːt̪] "*son*" and vice versa so [pɔːd̪] "*descendant*" was added to the list; however, [pɔːd̪] was also mispronounced as [puːd̪] by some participants, therefore [suːt̪] "*yarn*" was added to the list.

(3) [suːt̪] "*yarn*" was also mispronounced by some participants as [sɔːt̪], so [kuːd̪] "*jump*" was added to the list as well.

(4) [sə̃d̪] "*certificate*" was not pronounced as a monosyllabic word, rather it was pronounced as a disyllabic word [sənəd̪]. Therefore, [pə̃kʰ] "*feather*" was added to the list.

(5) For long nasal vowel [ɑ̃ː] there are two words in the list [sɑ̃ːd] "*ox*" and [bɑ̃ːd̪ʰ] "*tie*/*build*".

Table 2. 5: Urdu words carrying 12 oral monophthongs

| No. | Transcription | *Glosses* | Urdu words |
|---|---|---|---|
| 1 | [biːt̪] | *"pass"* | بیت |
| 2 | [bɪd̪] | *"new"* | بدہ |
| 3 | [bɪk] | *"sell"* | بک |
| 4 | [peʈ] | *"stomach"* | پیٹ |
| 5 | [bɛd]/ | *"a willow tree"* | بید |
| 6 | [bʌd̪] | *"bad"* | بد |
| 7 | [bæt̪] | *"follow"* | بیعت |
| 8 | [bɑːd̪] | *"after"* | بعد |
| 9 | [pɔːd̪] | *"descendant/ offspring"* | پود |
| 10 | [puːt̪] | *"son"* | پوت |
| 11 | [bʊd̪ʰ] | *"Wednesday"* | بدھ |
| 12 | [buːd̪ʰ] | *"wisdom"* | بودھ |
| 13 | [kuːd̪] | *"jump"* | کود |
| 14 | [suːt̪] | *"yarn (cotton)"* | سوت |

In order to minimise the co-articulatory effects of the neighbouring consonantal sounds the syllable structure for oral and nasal monophthongs was $C_1VC_2$ (Hawkins and Midgley, 2005; Ladefoged, 2003; Johnson et al., 1993); for diphthongs the structure was CV ($C_1V$).

The target vowels were placed in the same environment within the word as much as possible. In the majority of words, the vowels were between a bilabial (/b/or/p/) and an alveolar or dental plosive (/t/, /d/, /t̪/, or /d̪/); For monophthongs, $C_1$ was always unaspirated and $C_2$ was aspirated only in two cases i.e. /bʊd̪ʰ/ “wisdom” and /bʊd̪ʰ/ “*Wednesday*”. For diphthongs, $C_1$ was bilabial plosive /p/ and /b/ in four words, dental plosive /d̪/ in one word and velar plosive /g/ in one word. The list of words for Urdu oral monophthongs is given in Table 2. 5.

The list of sentences for oral monophthongs is given in Appendix 2B. The list of words and sentences for six diphthongs are given in the Chapter 3, and the list of words and sentences for nasal vowels is given in Appendix 2B because nasal vowels were not analysed in this thesis.

## 2.4 Procedures

The majority of recordings took place in a quiet room of the Boston Mosque in Boston, Lincolnshire, UK. Four female participants were visited in their homes by the experimenter, and one participant was recorded in the linguistics lab in the Department of English Language and Linguistics at the University of Kent Canterbury, UK. The participants were given an information sheet, a consent form and the LEAP-Q questionnaire to fill in before the recordings began. Each participant read from a computer screen five sets of sentences in pseudo-randomised order, such that standard carrier phrases and natural sentences alternated. The Urdu sentences were written in Arabic script in the Nastalique style (as discussed in Section 1.1) and were presented on screen one at a time. To this end, Microsoft PowerPoint presentation software was used, with each slide containing a single sentence. The participants’ utterances of the test words were recorded in .wav format using a Zoom Handy Recorder H4N using the in-bulit microphone, and with sample settings of 44.1 kHz digitization and 16-bit quantization to obtain high quality acoustic data. The recorder was kept on the table away from the speaker's mouth.

The total number of tokens was 7800 (26 speakers × 30 words carrying the target vowels × 2 sentence types × 5 repetitions).

The speech rate increased with every repetition for both male and female speakers. The median duration for each repetition across all the speakers ranged from 151 seconds (first repetition) to 114 seconds (fifth repetition) excluding the pauses between the (CP and FS) sentences. As discussed above (Section 2.3.1.2), there were five sets of Urdu sentences, and each set contained 60 sentences (30 CP and 30 FS) which had 354 words. The median speech rate thus ranged from 2.3 (first repetition) to 3.1 (fifth repetition) words per second.

The repetition was not found to have an effect on the quality of the vowels, therefore the statistical model as discussed in Section 2.4.4 did not include repetition as an effect; however, the increased speech rate of later repetitions did result in mispronunciation. The participants were told to repeat the whole sentence if they mispronounce the target words in the carrier phrases (CP), and occasionally the experimenter (the author of this thesis) requested that a participant repeat the sentence if she thought the speaker had not pronounced it correctly. Even after additional repetitions some participants still mispronounced (or did not pronounce as expected) some words; hence the number of mispronounced words remains high (see Appendix 2C). Each set of data (one out of five repetitions) took 5 to 8 minutes, and the duration decreased as the participants read the last two sets.

In order to keep the data anonymous, the recordings were coded such that no personally identifiable information is given; for example, MS1 for Male Speaker 1 and FS1 for Female Speaker 1. The anonymized data have been stored on a personal laptop, as well as off site in private cloud storage (i.e. Dropbox and Mega) and in secure University of Kent server.

### 2.4.1 Segmentation

Measurements were made using Praat (Boersma and Weenink, 2016). Segmental intervals were measured manually, after visual inspection of the spectrograms and waveforms analysed using Praat's default settings. That is, the view range was set at 5000 Hz for males and 5500 Hz for females with dynamic range 70 dB and window length 0.005s. The maximum numbers of formants were set to five for male speakers and four for female speakers, with dynamic range 30 dB at window length 0.025s and dot size 1 mm. Using Praat TextGrids, the following *tiers* were created to annotate the intervals containing the tokens:

Tier 1 (named "Vowels") intervals were created containing the vowel sounds and labelled with the IPA symbols, i.e. for the oral vowels: /iː/, /ɪ/, /eː/, /ɛ/, / æː/, /ɑː/, /ʌ/, /ə/, /uː/, /ʊ/, /ɔː/ and / ɒː/, and diphthongs /ɑe/, /ɔe/, /ɪɐ/, /eɐ/, /ʊɑ/ and /ɑʊ/. The words carrying the target vowels were also inserted after the symbol. For example, /iː/ - biːt̪.

Tier 2 (named "sentences") labelled standard carrier phrases and natural full sentences. The codes for standard carrier phrase (CP) were LS1 /pɑɪ/, LS2 /bəd/, LS3 /pɪə/, …, LS30 /buːd/. The codes for full sentences (FS) were NS1 /pɑɪ/, NS2 /bəd/, NS3 /pɪə/, …, NS30 /buːd/. That is, LS for lab speech and NS for natural speech.

The following criteria were used to insert boundaries around the target vowels (as discussed by Ladefoged, 2011):

If the token word carrying the target sound began with a stop, such as /p/, /b/, /t/, /k/, /g/, and fricative /s/, the boundaries were inserted (as shown in the Figure 2. 1) where the voicing began, the formant structure for F1 and F2 became visible in the spectrogram and waveforms started to get complex. Following Ladefoged (2011), in order to be consistent across the word initial consonants, i.e. stops and fricatives, the burst and VOT of $C_1$ was not included in the vowel interval. In Urdu, plosives can be aspirated or unaspirated. In the chosen syllables, however, they were not aspirated.

As shown in Figure 2. 1 and Figure 2. 2, if the target sound was in a closed syllable and ended in a stop, i.e. /t̪/, /d̪/, /d/, /k/, or /ʈ/, the boundaries were inserted where the vowel started losing amplitude, closure and voicing of the following consonant began and the waveform started to get less complex. Though no disyllabic words were used in the stimuli, one word [sə̃d̪] was pronounced or mispronounced as a disyllabic [sənəd̪] word by all except one of the speakers. Therefore, the formant frequencies for the sound [ə] were extracted from the first and second syllable where the first consonant was /s/ and the final consonant was /d̪/, as shown in Figure 2. 3.

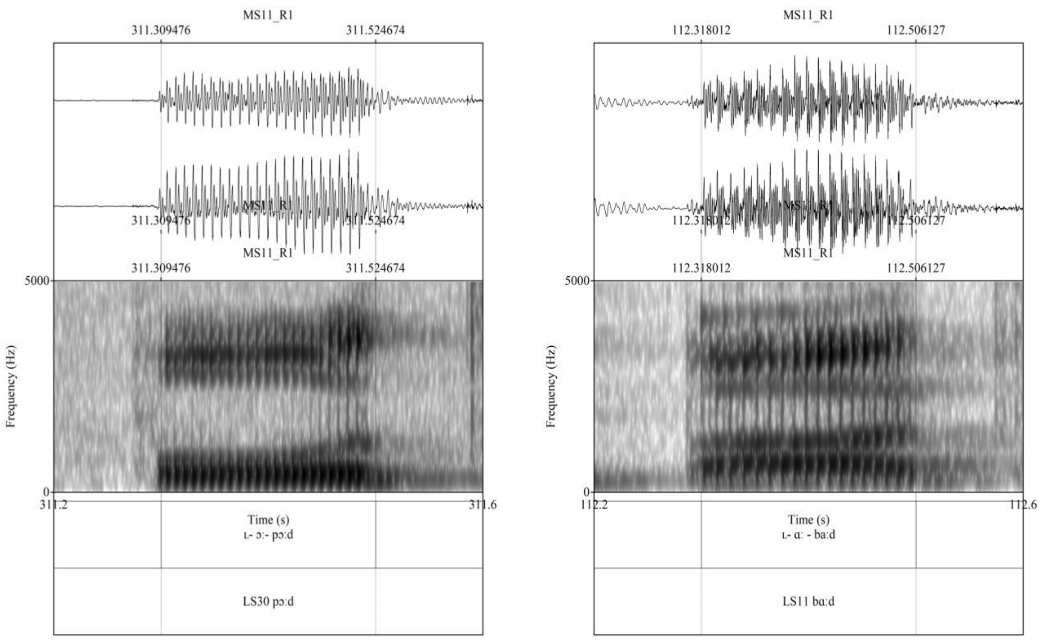



Figure 2. 1: Word level segmentation for vowels in [pɔːd̪] (C1 -voiced V C2 +voiced) [left] and [bɑːd̪] (C1 -voiced V C2 +voiced) [right]

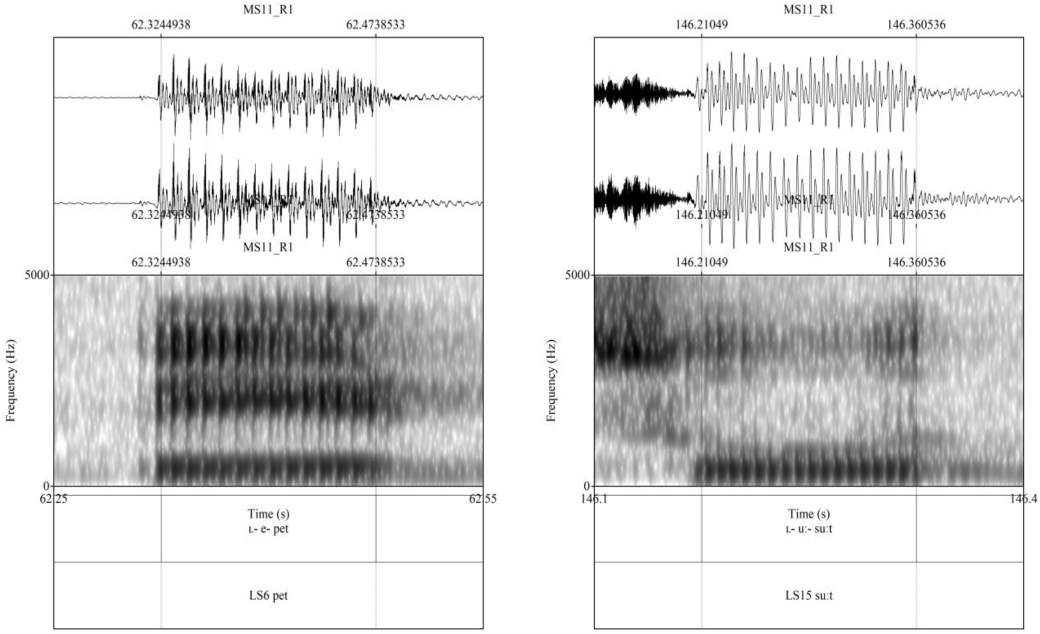



Figure 2. 2: Word level segmentation for vowels in [peʈ] ($C_1$ -voiced V $C_2$ +voiced) [left] and for [su:t̪] ($C_1$ -voiced V $C_2$ -voiced) [right]

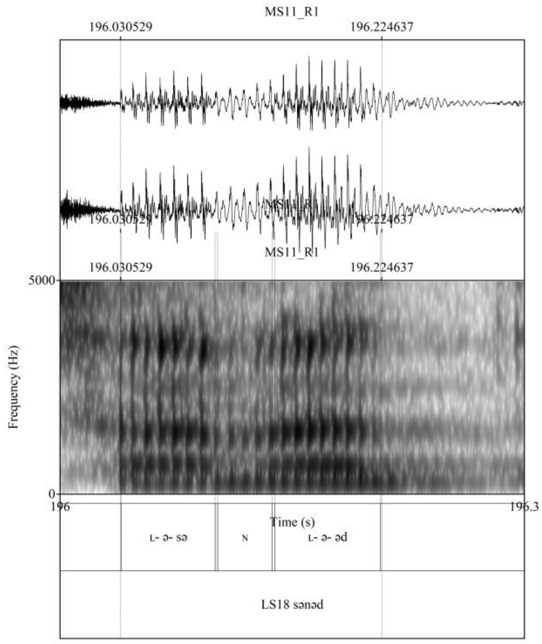



Figure 2. 3: Word level segmentation for vowels in [sənəd̪] ($C_1$ -voiced V nasal Consonant V $C_2$ +voiced) to extract the formant frequencies. There were only monosyllabic words in the stimuli; however, all but one of the speakers pronounced this word as disyllabic instead of monosyllabic.

### 2.4.2 Acoustic Measurements

The total number of tokens recorded was 7800 (= 26 speakers × 30 words carrying the target vowel sounds × 2 sentence types × 5 repetitions). Although data were collected and segmented for both oral and nasal vowels (as discussed above in Section 2.3 and 2.4.1), only the oral vowels are required for testing the perception SSBE vowels (as discussed in Section 1.4, Section 4.2 and Section 5.1). A full analysis of nasal vowels is thus left for future work.

Table 2. 6: The total number of tokens analysed per vowel (oral monophthongs) by 11 males (M) and 11 females (F)) in each context Carrier Phrases (CP) and Full Sentence (FS)

| Vowel | Tokens | Speakers | | Context | |
|---|---|---|---|---|---|
| | | F | M | CP | FS |
| iː | 182 | 93 | 89 | 80 | 102 |
| ɪ | 366 | 183 | 183 | 180 | 186 |
| e | 218 | 109 | 109 | 108 | 110 |
| ɛ | 213 | 106 | 107 | 103 | 110 |
| æ | 239 | 128 | 111 | 136 | 103 |
| ɑː | 217 | 108 | 109 | 110 | 107 |
| ɔː | 167 | 85 | 82 | 83 | 84 |
| o | 84 | 61 | 23 | 50 | 34 |
| ʊ | 222 | 113 | 109 | 109 | 113 |
| uː | 777 | 364 | 413 | 379 | 398 |
| ə | 419 | 211 | 208 | 211 | 208 |
| ʌ | 247 | 134 | 113 | 121 | 126 |
| **Total** | **3351** | **1695** | **1656** | **1670** | **1681** |

Data from two males and two female speakers were not included in the final analysis. One of the two male speakers was trying to speak Urdu with an English accent, and he aspirated all the stop consonants in Urdu words. The second male speaker read most of the sentences in each set very quickly and usually without taking a pause between

sentences. His speed increased with every set. As a result, segmentation was very difficult and unreliable. Of the two female speakers, one pronounced many words and sentences incorrectly - she reported no sight problems, however from her pronunciation/reading mistakes, it appeared that either she was unable to read written Urdu or she had some sight problem causing her to misread the words and sentences. The second female participant was breathing heavily during the recordings (due to old age) and therefore the recordings were not of high quality and comparable to those of other speakers. As a result, the data analysed was based on recordings from 22 participants (11 males and 11 females). Out of the initial 7800 tokens for 25 vowels, 6600 remained once the unsuitable participants were excluded, leaving 3351 tokens for analysis of 12 oral monophthongs.
The total number of tokens analysed per monophthong, by males (M) and females (F) and in Carrier Phrases (CP) and Full Sentence (FS) are given in Table 2. 6. Vowel tokens by individual speakers in each context are given in Appendix 2D.

**2.4.3 Automatic Formant Extraction**

Praat scripts (see Appendix 2E) were used to extract the frequencies of the first, second and third formant of monophthongs in three positions: at vowel onset (+10ms), at the middle of the vowel, and at vowel offset (-10ms). The Praat scripts use the same configuration as that used for manual segmentation, that is a maximum frequency of 5000 Hz, a window length of 0.025 seconds and dynamic range of 30 dB for male speakers; and 5500 Hz, 0.025 seconds and 30 dB for female speakers.

Previous studies have shown (Williams and Escudero, 2014; Hillenbrand, 2013; Morrison, 2013; Mayr and Davies, 2011; Fox and Jacewicz, 2009; Morrison, 2009; Watson and Harrington, 1999) that F1 and F2 alone in the vowel steady state (i.e. usually midpoint) is not sufficient to investigate the acoustic properties of vowels because the vowel quality changes. Therefore, the *rate of change* (ROC) approach, as employed by

Gay (1968), Deterding (2000) and Kent and Read (1992 cited in Deterding, 2000), was also used to measure the change in the quality as well as the spectral change in the monophthongs, because some monophthongs can show diphthongal patterns. For instance, Williams and Escudero (2014) observed that some English monophthongs such as /uː/ has diphthongal properties. In the ROC approach the difference between the value of F1 at the beginning and at the end is divided by the total duration. Positive and negative values indicate the direction of change. E.g. a negative value indicates that the vowel is changing from a more open position to a closing position, and a positive value indicates the reverse. In addition to formant frequencies, the duration of monophthongs was also measured between the boundaries of start and end. For monophthongs, temporal measurements are necessary to establish the difference between long and short vowels.

Raw Hertz formant frequencies are not considered reliable for comparing vowels on the same plot for different speakers (Watt et al., 2010 cited in Flynn, 2011:2). Therefore, formant frequencies for F1 and F2 were normalised by using Lobanov (z-score) normalisation (Thomas and Kendall, 2007):

$$F_{n[V]}^{N} = (F_{n[V]} - MEAN_{n})/S_{n}$$

> *Where $F_{n[V]}^{N}$ is the normalized value for $F_{n[V]}$ (i.e., for formant n of vowel V). $MEAN_n$ is the mean value for formant n for the speaker in question and $S_n$ is the standard deviation for the speaker's formant n.*

The Lobanov normalisation method is *vowel extrinsic*, where the frequency of each formant for a given vowel utterance and speaker is adjusted relative to the mean and standard deviation across *all* vowels for that formant and speaker. The Lobanov method was chosen due to its prevalence in the literature and the fact that it outperforms most other normalisation methods at “equalising and aligning the vowel space areas” of speakers (Flynn, 2011:13). This method was considered appropriate for the present study as there were five repetitions per vowel across two contexts by each speaker. This method normalises the possible variations in the productions of each speaker; hence allows for better comparison across speakers and contexts.

### 2.4.4 Statistical Analysis

For statistical analysis, *R* (R Core Team, 2016) was used to construct linear mixed effects models via the *lme4* package's *lmer* function (Bates et al., 2015), with the dependent variable being F1 (Lobanov normalised), F2 (Lobanov normalised) or duration (in milliseconds). Repeated measures ANOVAs followed by post-hoc Tukey tests were then conducted using the *lmerTest* package (Kuznetsova, Brockhoff and Christensen, 2017). Linear Mixed Effects Models (LMEMs) are particularly suitable for the structure of the data gathered by the present study. The data contained uneven cells in terms of the number of tokens per vowel, context and speaker (see Appendix 2D). LMEMs can be constructed with such data without discarding data points and statistical tests can be performed on them easily (see Politzer-Ahles and Piccinini, 2018; Cunnings, 2012; Jaeger, 2008; Krueger and Tian, 2004 for details). Further, LMEMs are considered more powerful because they can model both fixed and random effects (Muth et al. 2016; West, Welch and Galecki, 2014; Barr et al. 2013; Winter, 2013).

An initial trial linear mixed effects model included context, gender, speaker, and vowel as effects. The independent variables were Vowel, Context, Gender and Speaker. It was anticipated that data from more natural speech (FS) would be different from those obtained using standard carrier phrases (CP) which leads to slower, more careful speech. This could have led to differences in vowel quality, specifically leading to greater centralization for vowels in FS (Deterding, 1997). It was also anticipated that gender could affect the data even after Lobanov normalization; this would happen, e.g., if speakers use different gender-related variants (Gahl and Baayen, 2019; Clopper et al.

2018; Clopper, Mitsch and Tamati, 2017). Finally, speaker was included as a random effect, as anatomical and general stylistic differences could affect vowel quality (Vorperian and Kent, 2007; Clopper, Pisoni and De Jong, 2005; Hawkins and Midgley, 2005; Adank, Smits and Van Hout, 2004a; 2004b).

Hence, context, gender and vowel were modelled as fixed effects, along with all their possible interactions. Speaker was modelled as a random effect with random intercept. This configuration was reused with F1, F2 and duration as response variables. The formula as entered into R was thus

Context * Vowel * Gender + (1 | Speaker).

The model was then reduced with *lmerTest*'s function *step*, which helped to eliminate non-significant effects of the full linear mixed effects models for F1, F2 and duration. The ANOVA significance results reported by the *step* function, using significance levels of 0.01 for both fixed and random effects, are reported in the results section.

Using the final reduced models, post hoc differences of least square mean t-tests were performed on each factor of the model using the *lsmeans* (Lenth, 2016) package/function. Tukey corrections for multiple comparisons were used for the pairwise p-values.

For the visual representation of the data, F1/F2 vowel space plots, stacked-bar plots (for certain aspects of duration), and violin plots (also for duration) (Adler, 2005) have been used. Violin plots are preferred to box plots because, as well as showing the median, interquartile range and 95% confidence interval, they also show a kernel density estimation illustrating the distribution of the data. That is, a wider section of the kernel density represents a higher probability that members of the population will take on the given value.

## 2.5 Results

### 2.5.1 Formant Frequencies

The mean formant frequencies with standard deviation in parenthesis for the first three formants of the monophthongs at the midpoint in Hertz with standard deviation is given in Table 2.7 and Lobanov normalised in Table 2.8. Further, the mean formant frequencies with standard deviation for the first two formants of the monophthongs at seven equidistant points in time, i.e. 20%, 30%, …, 80%, in Hertz and Lobanov normalised are given in Appendix 2F. The mean formant frequencies and duration for each context and gender separately are given in Appendix 2G.

Table 2. 7: Mean frequencies of F1, F2 and F3 at vowel midpoint of 12 oral monophthongs in Hertz produced by 11 male and 11 female speakers. Standard deviations are given in parenthesis.

| Vowel | Gender | F1 (SD) | F2 (SD) | F3 (SD) |
|---|---|---|---|---|
| iː | F | 344 (94) | 2598 (214) | 3254 (372) |
| | M | 314 (63) | 2269 (182) | 2837 (292) |
| ɪ | F | 410 (91) | 2327 (236) | 2887 (267) |
| | M | 385 (41) | 1967 (184) | 2550 (259) |
| eː | F | 420 (110) | 2407 (110) | 2865 (386) |
| | M | 407 (58) | 1991 (144) | 2446 (235) |
| ɛ | F | 664 (83) | 2012 (240) | 2794 (265) |
| | M | 570 (41) | 1740 (132) | 2531 (234) |
| æ | F | 674 (48) | 2001 (287) | 2802 (228) |
| | M | 557 (31) | 1688 (221) | 2495 (288) |
| ɑː | F | 769 (43) | 1317 (307) | 2800 (227) |
| | M | 637 (25) | 1121 (203) | 2616 (198) |
| ɔː | F | 530 (104) | 958 (169) | 2921 (263) |
| | M | 500 (67) | 962 (363) | 2629 (243) |
| o | F | 513 (38) | 951 (368) | 2955 (311) |
| | M | 480 (33) | 838 (151) | 2662 (320) |
| ʊ | F | 419 (88) | 1193 (141) | 2920 (313) |
| | M | 393 (67) | 994 (205) | 2619 (177) |
| uː | F | 388 (44) | 1015 (130) | 2887 (197) |
| | M | 356 (31) | 836 (121) | 2587 (142) |
| ə | F | 610 (50) | 1797 (291) | 2959 (229) |
| | M | 544 (40) | 1477 (230) | 2607 (201) |
| ʌ | F | 647 (94) | 1584 (107) | 2860 (291) |
| | M | 542 (62) | 1306 (133) | 2611 (187) |

Table 2. 8 and Figure 2. 4, Figure 2. 5, and Figure 2. 6 show mean F1 and F2 frequencies of the 12 monophthongs on Lobanov normalised F1/F2 vowel space for each vowel produced by both males and females. The centroid of each vowel is marked by the appropriate vowel symbol, and the ellipses around the vowels represent +/- 1 standard deviation.

Table 2. 8: Mean frequencies of F1 and F2 at vowel midpoint of 12 oral monophthongs, Lobanov normalised. Standard deviations are given in parenthesis.

| Vowel | F1 (SD) | F2 (SD) |
|---|---|---|
| iː | -1.4 (0.3) | 1.7 (0.5) |
| ɪ | -0.8 (0.3) | 1.2 (0.3) |
| e | -0.6 (0.3) | 1.3 (0.4) |
| ɛ | 0.9 (0.5) | 0.7 (0.3) |
| æ | 0.9 (0.5) | 0.7 (0.3) |
| ɑː | 1.6 (0.5) | -0.6 (0.2) |
| ɔː | 0.0 (0.5) | -1.3 (0.2) |
| o | -0.1 (0.5) | -1.2 (0.4) |
| ʊ | -0.7 (0.3) | -0.9 (0.2) |
| uː | -1.0 (0.3) | -1.2 (0.5) |
| ə | 0.7 (0.6) | 0.2 (0.3) |
| ʌ | 0.9 (0.6) | -0.3 (0.2) |

Figure 2. 4, Figure 2. 5 and Figure 2. 6 show that Punjabi-Urdu speakers distinguished between 10 out of 12 candidate monophthongs. The results show full spectral overlap for [æ]-[ɛ] and [ɔː]-[o] (or most often /ɔ/ as reported in the studies on Urdu vowels), and partial spectral overlap can be seen for close and close-mid front vowels [ɪ]-[e] and open-mid central vowels [ə]-[ʌ]. The literature on the Urdu vowel system only reports /ə/ as a central vowel; however, the present study shows that there are two central vowels [ə] and [ʌ].

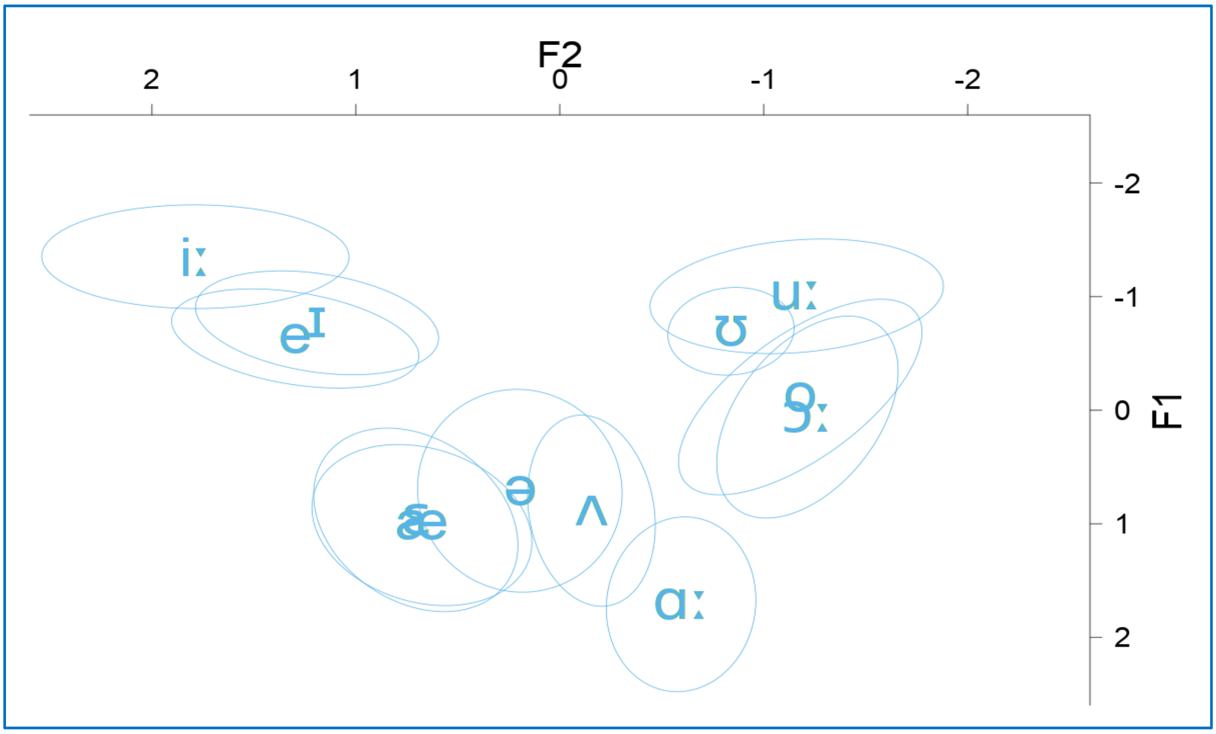



Figure 2. 4: Mean formant frequencies (Lobanov normalised) of the 12 monophthongs on F1/F2 vowel space by 22 speakers across both contexts. The ellipses around the vowels represent +/- 1 standard deviation.

In order to investigate the phonetic and acoustic differences between the monophthongs, the formant frequency values from each participant were analysed statistically.

For F1, the results revealed a significant fixed effect of Vowel ($F(11,3334.35) = 1612.41$, $p < 0.0001$), Context ($F(1,3328.99) = 45.51$, $p < 0.0001$), and interaction between Context × Vowel (Vowel: $F(11,3329.41) = 16.96$, $p < 0.0001$) and Gender × Vowel ($F(11,3334.38) = 6.94$, $p < 0.0001$) and Context × Gender ($F(1,3328.92) = 9.59$, $p < 0.001$). The random effect of Speaker was also significant ($\chi^2$ (1) = 48.12, $p < 0.0001$). The results further showed non-significant interaction effects between Context × Gender × Vowel ($F(11,3329.29) = 1.69$, $p = 0.06$) and non-significant effect of Gender ($F(1,24.66) = 1.39$, $p = 0.24$).

For F2 the results showed a significant fixed effect of Vowel ($F(11,3334.94) = 2296.70$, $p < 0.0001$), interactions between Context × Vowel ($F(11,3329.75) = 10.51$, $p < 0.0001$)

and Gender × Vowel ($F(11,3334.98) = 2.42$, $p < 0.01$), but non-significant effect of Gender ($F(1,25.74) = 0.97$, $p = 0.33$) and Context ($F(1,3329.21) = 0.95$, $p = 0.33$). The random effect of Speaker was also significant ($\chi^2$ (1) = 357.69, $p < 0.0001$). The results further showed non-significant interaction effects between Context × Gender ($F(1,3329.14) = 0.01$, $p = 0.91$), and non-significant three-way interaction between Context × Gender × Vowel ($F(11,3329.64) = 1.06$, $p = 0.38$).

In summary:

**For F1**, the combined interaction between gender, context and vowel was found to be non-significant. In terms of an R formula we have

*Context + Vowel + Gender + Context:Vowel + Context:Gender + Vowel:Gender + (1 | Speaker)*

**For F2**, the combined interaction between gender, context and vowel was found to be non-significant, as was any interaction between gender and context. In terms of an R formula we have

*Context + Vowel + Gender + Context:Vowel + Vowel:Gender + (1 | Speaker)*

Among front close vowels, and indeed overall, [iː] (329 Hz) and [ɪ] (398 Hz) have the lowest F1, and [ɔː] (893 Hz), [uː] (918 Hz) and [o] (941 Hz) have the lowest F2 values. A higher F1 for [ɑː] (702 Hz) than [ʌ] (603 Hz), with a difference of 101 Hz, suggests that [ʌ] is not as open as [ɑː]; however, the higher F2 for [ʌ] (1453 Hz) suggests that it is not as retracted and back as [ɑː] (1218 Hz). The F1 for [e] (413 Hz) is slightly higher than [ɪ] (399 Hz) but significantly lower than [ɛ] (608 Hz) and [æ] (621 Hz), and the higher F2 for [e] suggests that [e] is closer and more front than [ɛ] and [æ]. F1 and F2 for [ɔː] (F1=496 Hz; F2=893 Hz) is slightly lower than [o] (F1=501 Hz; F2=941 Hz); however, as the spectral overlap in Figures 2.4, 2.5 and 2.6 show, spectrally these two vowels are

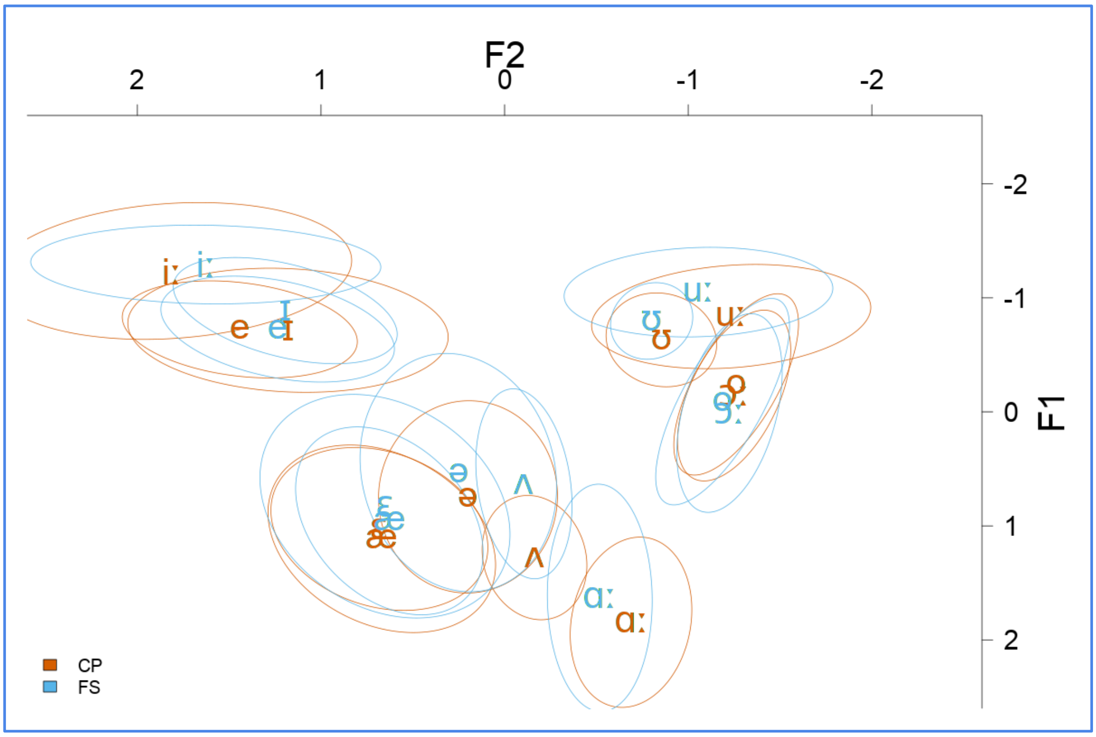



Figure 2. 5: Mean formant frequencies (Lobanov normalised) of the 12 monophthongs on F1/F2 vowel space by female speakers in both contexts. The ellipses around the vowels represent +/- 1 standard deviation.

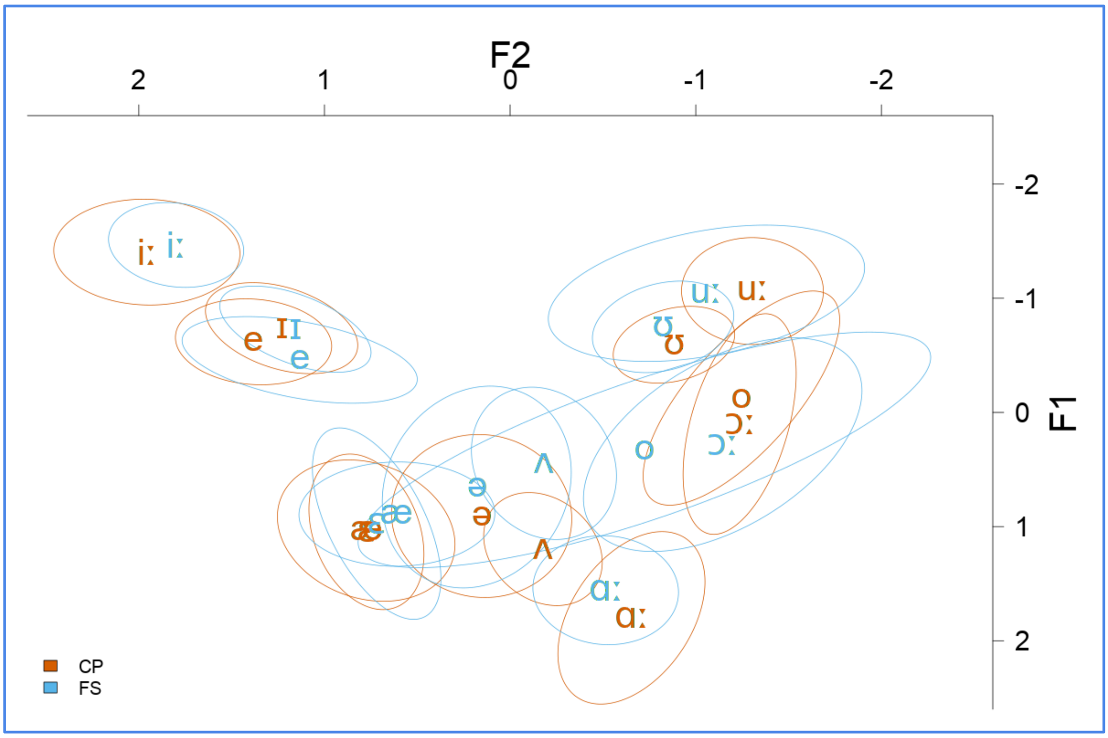



Figure 2. 6: Mean formant frequencies (Lobanov normalised) of the 12 monophthongs on F1/F2 vowel space by male speakers in both contexts. The ellipses around the vowels represent +/- 1 standard deviation.

similar. This finding is very interesting as these two vowels are treated in the Urdu literature as two distinct phonemes (Saleem et al., 2001; Hussain, 1997; Kachru, 1987), which, the present study does not support.

Spectral overlap can also be seen from Figure 2. 4, Figure 2. 5 and Figure 2. 6 for front close-mid vowels [ɪ] - [e]; front open-mid vowels [ɛ] - [æː]; central vowels [ə]- [ʌ]; back open-mid vowels [ɔː] - [o]; and back close-mid vowels [uː]-[ʊ].

**2.5.1.1 Front Vowels**

**[ɪ] - [e]**

As the Figure 2.7 (left) and (right) show spectral overlap for [ɪ] - [e] in all contexts and by both genders, the mean formant frequencies were analysed further to investigate if these vowels are spectrally distinct.

Post hoc pairwise comparisons of vowel LS means (*lsmeans)* using Tukey-adjusted significance level (Lenth, 2016) were performed separately for F1 and F2. The pairwise comparisons showed that that [ɪ] – [e] were distinct with regard to F1(t(3367.30) = 4.071, $p < 0.01$). However, these two vowels were not significantly distinct with regard to F2 (t (3367.13) = 3.253, $p = 0.05$). Despite visually showing large spectral overlap in Figure 2.7, we find that they are much more clearly distinct with respect to duration, as discussed below.

**[æ]- [ɛ]**

[æ] and [ɛ] are reported as distinct phonemes in Urdu; however, as can be seen in Figure 2. 8 (left) and (right), the present study shows that [ɛ] and [æ] do not differ spectrally. The pairwise comparisons showed that that these vowels were not significantly distinct with regard to F1(t(3368.68) = 0.898, $p = 0.99$) or F2 (t(3368.89) = -1.075, $p=0.99$).

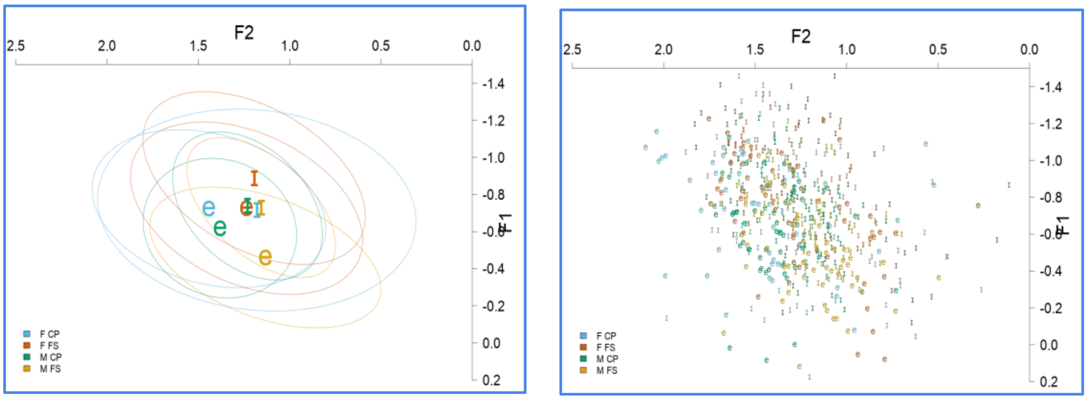



Figure 2.7: (left) mean formant frequencies (Lobanov normalised) of [e] and [ɪ] per speaker gender and context. The ellipses around the vowels represent +/- 1 standard deviation; (right) mean formant frequencies of /e/ and /ɪ/ for each individual speaker, colour coded by speaker gender and context.

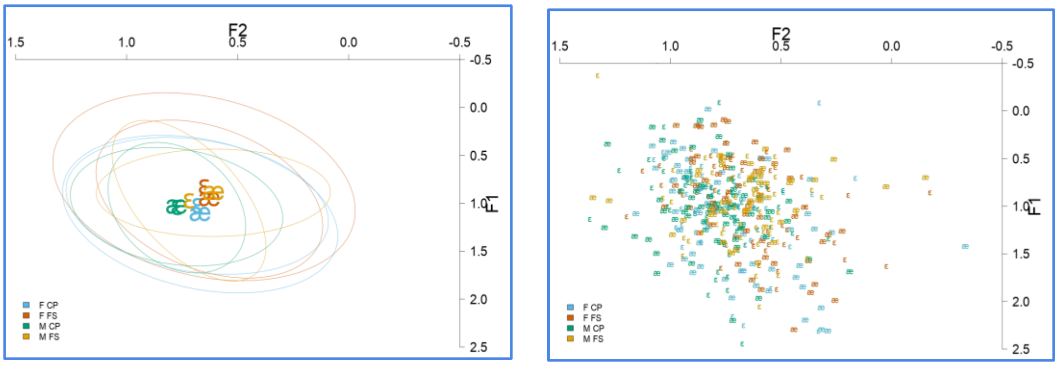



Figure 2. 8: (left) mean formant frequencies (Lobanov normalised) of [ɛ] and [æ] per speaker gender and context. The ellipses around the vowels represent +/- 1 standard deviation; (right) mean formant frequencies of /ɛ/ and /æ/ for each individual speaker, colour coded by speaker gender and context.

### 2.5.1.2 Central Vowels

**[ə]- [ʌ]**

Spectral overlap for [ə] and [ʌ] can be seen in the Figure 2. 9 (left) and (right). However, the post-hoc *lsmeans* pairwise comparisons showed that these vowels are distinct. F1 for [ʌ] is higher (by 23 Hz) than [ə] (F1(t(3369.16) =-5.374, p <0.0001), and F2 for [ʌ] is significantly lower (by 188 Hz) than [ə]  (F2(t(3369.49) =12.176, p <0.0001).

### 2.5.1.3 Back Vowels

**[ɔː] - [o]**

Spectral overlap as shown in the Figure 2. 10 (left) and (right) suggests that these vowels do not differ in quality. The post-hoc *lsmeans* pairwise tests showed that [ɔː] and [o] are not significantly distinct with regard to F1 (t (3386.96) = -0.861, p = 0.99) nor F2(t (3387.53) = 1.209, p = 0.98). This spectral similarity suggests that these vowels are the same and should be treated as one vowel.

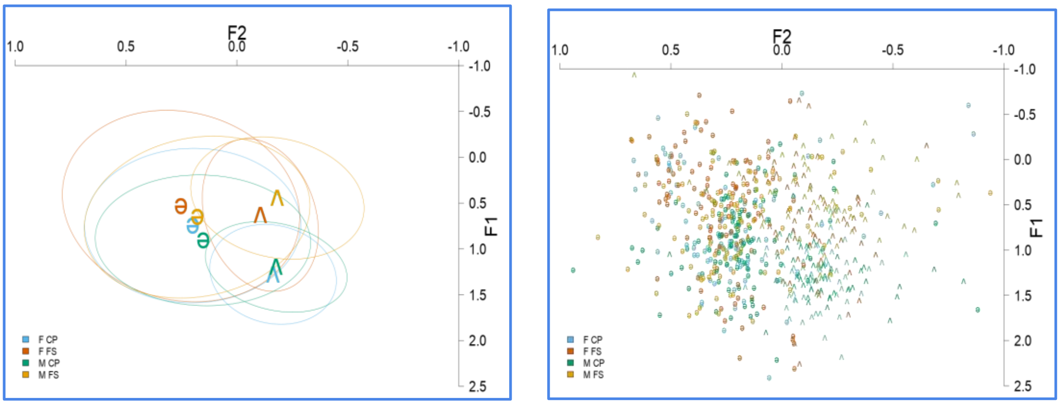



Figure 2. 9: (left) mean formant frequencies (Lobanov normalised) of [ə] and [ʌ] per speaker gender and context. The ellipses around the vowels represent +/- 1 standard deviation; (right) mean formant frequencies of [ə] and [ʌ] for each individual speaker, colour coded by speaker gender and context.

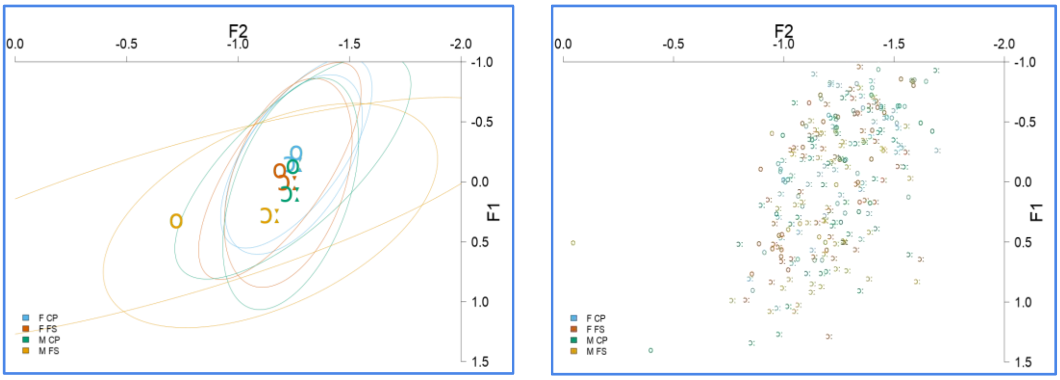



Figure 2. 10: (left) mean formant frequencies (Lobanov normalised) of [ɔː] and [o] per speaker gender and context. The ellipses around the vowels represent +/- 1 standard deviation; (right) mean formant frequencies of [ɔː] and [o] for each individual speaker, colour coded by speaker gender and context.

The rate of spectral change and diphthongization patterns for Urdu monophthongs are discussed in the next section.

### 2.5.2 Spectral Change

The F1 and F2 frequencies at seven equidistant points in time are given in Appendix 2F and shown in Figure 2.11. Formant trajectories of F1, F2 and F3 for diphthongnal monophthongs are shown in Figure 2.12. Visual inspection of the monophthongs shows that the vowel trajectories within words [bɑːd̪], [pɔːd̪], [bʊd̪], [kuːd̪] and [bʌd̪] display diphthongal movement as compared to other monophthongs. In order to investigate if these monophthongs are diphthongised, the F1 and F2 rate of change (ROC) was also calculated, as described in Section 2.4.3, and given in Table 2.9 for F1 and Table 2.10 for F2. As can be seen from the figures and tables, the F2 formant trajectory for the back vowels [ɑː], [ɔː], [ʊ], [uː], and central vowel [ʌ] show large movement.

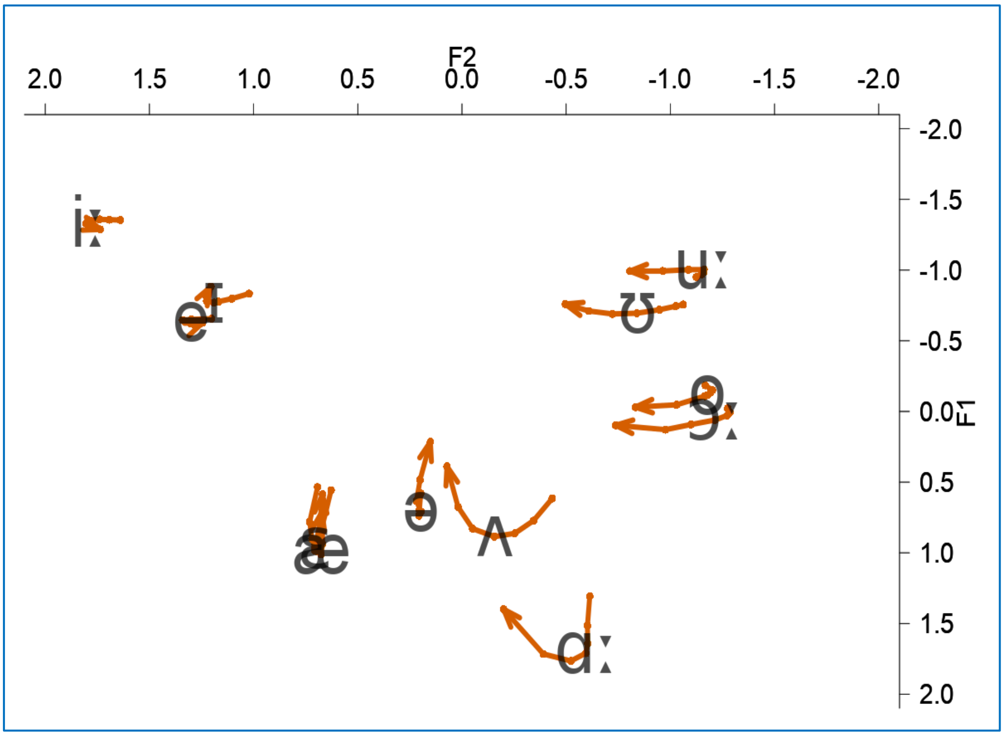



Figure 2. 11: Mean formant trajectories (Lobanov normalised) of 12 monophthongs for all speakers in both contexts. The dots on each trajectory show the formant measurements at seven equidistant points in time during each monophthong, i.e. 20%, 30%, …, 80%, and arrowheads show the direction and offset of each monophthong.

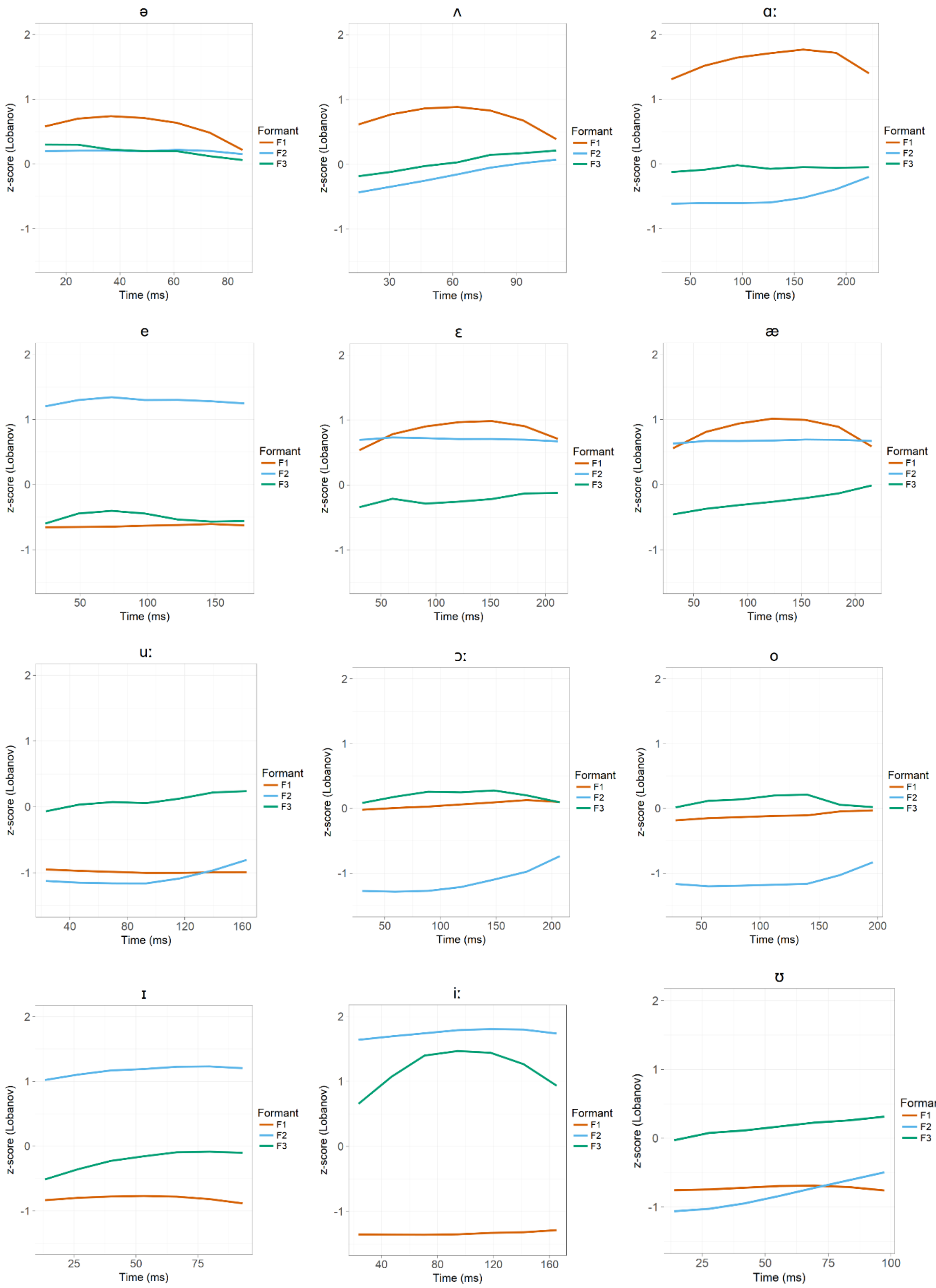



Figure 2. 12: Lobanov normalised (z-score) F1, F2 and F3 trajectories of 12 oral monophthongs pooled over male and female speakers and two contexts (CP and FS).

#### 2.5.2.1 F1 and F2 Rate of Change (ROC)

The F1 ROC was much lower for [ə] (F1 = -585 Hz/second) and [ʌ] (F1 = -316 Hz/second). These negative values suggest that F1 gradually decreases towards the offset of the vowels; hence the F1 formant structure of these vowels is not steady. These vowels showed even more movement in the F2 dimension, as can be seen in Figure 2. 12, and the F1 and F2 ROC values are given in Table 2.9 and Table 2.10 respectively. Similarly, the vowels in [bɑːd̪], [bʊd̪], [kuːd̪], [bʌd̪] and [pɔːd̪] show movement for F1 but more prominently for F2 as can be seen in Figure 2. 12. The substantial F2 movement could be explained by the dental consonant /d̪/ following the vowel. In addition to the observation of F1 and F2 movements, visual inspection of these trajectories also indicate that [e] and [ɛ] are distinct vowels; however, [ɔː] and [o], and [æ] and [ɛ] do not appear distinct and show similar patterns for F1 and F2 trajectories, as shown in Figure 2. 12.

Table 2. 9: Measurement of Rate of Change (ROC) in Hz/sec of F1 for 12 oral monophthongs

| Vowel | F1 start (Hz) | F1 End (Hz) | Change (Hz) | Duration (sec) | ROC (Hz/sec) |
|---|---|---|---|---|---|
| iː | 329 | 337 | 8 | 0.142 | 58 |
| ɪ | 390 | 384 | -5 | 0.080 | -65 |
| e | 411 | 415 | 4 | 0.147 | 24 |
| ɛ | 555 | 578 | 23 | 0.181 | 128 |
| æ | 561 | 571 | 10 | 0.184 | 56 |
| ɑː | 652 | 663 | 11 | 0.190 | 60 |
| ɔː | 486 | 503 | 17 | 0.178 | 93 |
| o | 491 | 512 | 21 | 0.168 | 123 |
| ʊ | 398 | 400 | 2 | 0.083 | 22 |
| uː | 377 | 372 | -5 | 0.140 | -32 |
| ə | 562 | 519 | -44 | 0.073 | -594 |
| ʌ | 569 | 543 | -26 | 0.093 | -281 |

Table 2. 10: Measurement of Rate of Change (ROC) in Hz/sec of F2 for 12 oral monophthongs

| Vowel | F2 start (Hz) | F2 End (Hz) | Change (Hz) | Duration (sec) | ROC (Hz/sec) |
|---|---|---|---|---|---|
| iː | 2372 | 2420 | 48 | 0.142 | 339 |
| ɪ | 2061 | 2155 | 94 | 0.080 | 1184 |
| e | 2151 | 2171 | 20 | 0.147 | 135 |
| ɛ | 1889 | 1868 | -20 | 0.181 | -113 |
| æ | 1856 | 1862 | 6 | 0.184 | 30 |
| ɑː | 1210 | 1419 | 209 | 0.190 | 1100 |
| ɔː | 861 | 1145 | 284 | 0.178 | 1597 |
| o | 949 | 1135 | 186 | 0.168 | 1108 |
| ʊ | 974 | 1268 | 294 | 0.083 | 3527 |
| uː | 938 | 1099 | 161 | 0.140 | 1150 |
| ə | 1636 | 1609 | -27 | 0.073 | -366 |
| ʌ | 1310 | 1570 | 260 | 0.093 | 2790 |

### 2.5.3 Duration

The mean duration for the 12 monophthongs is given in Table 2.11. The results show a clear distinction between long and short vowels, especially the four short vowels, i.e. [ɪ] (80ms), [ə] (73ms), [ʌ] (93ms), and [ʊ] (83ms), all having less than 100ms mean duration. However, [ɛ] (181ms) does not appear to be a short vowel as reported in the literature (Fatima and Aden 2003). Figure 2. 13 shows that speakers made a clear durational distinction between long and short monophthongs. That is, long monophthongs [iː], [e], [ɛ], [æ], [ɑː], [o] and [ɔː] are clearly longer than short monophthongs [ɪ], [ə], [ʌ] and [ʊ] in both contexts. [ʌ] appears to be the longest among the four short vowels, and [ə] appears to be the shortest among all the vowels. The results also showed that low vowels [æ] and [ɑː] are much longer than the high vowels [iː], [ɪ], [uː] and [ʊ], as shown in Figure 2. 13.

Table 2. 11: Mean duration in milliseconds (ms) of 12 oral monophthongs produced by 11 male and 11 female speakers. Standard deviations are given in parenthesis.

| Vowel | Gender | Duration (SD) |
|---|---|---|
| iː | F | 145 (34) |
| | M | 136 (35) |
| ɪ | F | 83 (24) |
| | M | 76 (14) |
| eː | F | 154 (37) |
| | M | 140 (40) |
| ɛ | F | 190 (42) |
| | M | 175 (43) |
| æ | F | 188 (37) |
| | M | 170 (30) |
| ɑː | F | 199 (45) |
| | M | 182 (47) |
| ɔː | F | 180 (38) |
| | M | 153 (38) |
| o | F | 183 (36) |
| | M | 172 (34) |
| ɔː | F | 180 (38) |
| | M | 153 (38) |
| ʊ | F | 86 (22) |
| | M | 79 (17) |
| uː | F | 145 (37) |
| | M | 135 (31) |
| ə | F | 78 (18) |
| | M | 65 (12) |
| ʌ | F | 94 (30) |
| | M | 82 (26) |

Statistical tests revealed a significant main effect of Vowel ($F(11,3329.58) = 974.18$, $p < 0.0001$), a significant main effect of Context ($F(1,3329.02) = 1438.95$, $p < 0.0001$), and a non-significant main effect of Gender ($F(1,21.99) = 3.86$, $p = 0.06$) for predicting duration. In addition, the results showed a significant Context × Vowel interaction

(F(11,3329.05) = 32.08, p < 0.0001), and a significant Gender × Context interaction (F(1,3329.01) = 10.92, p < 0.001). The random effect of Speaker was also significant ($\chi^2(1)$ = 1109.07, p < 0.0001). However, Gender × Vowel interaction (F(11,3329.72) = 1.46, p = 0.13); and a three-way Context × Gender × Vowel interaction (F(11,3328.95) = 0.79, p = 0.64) were not significant.

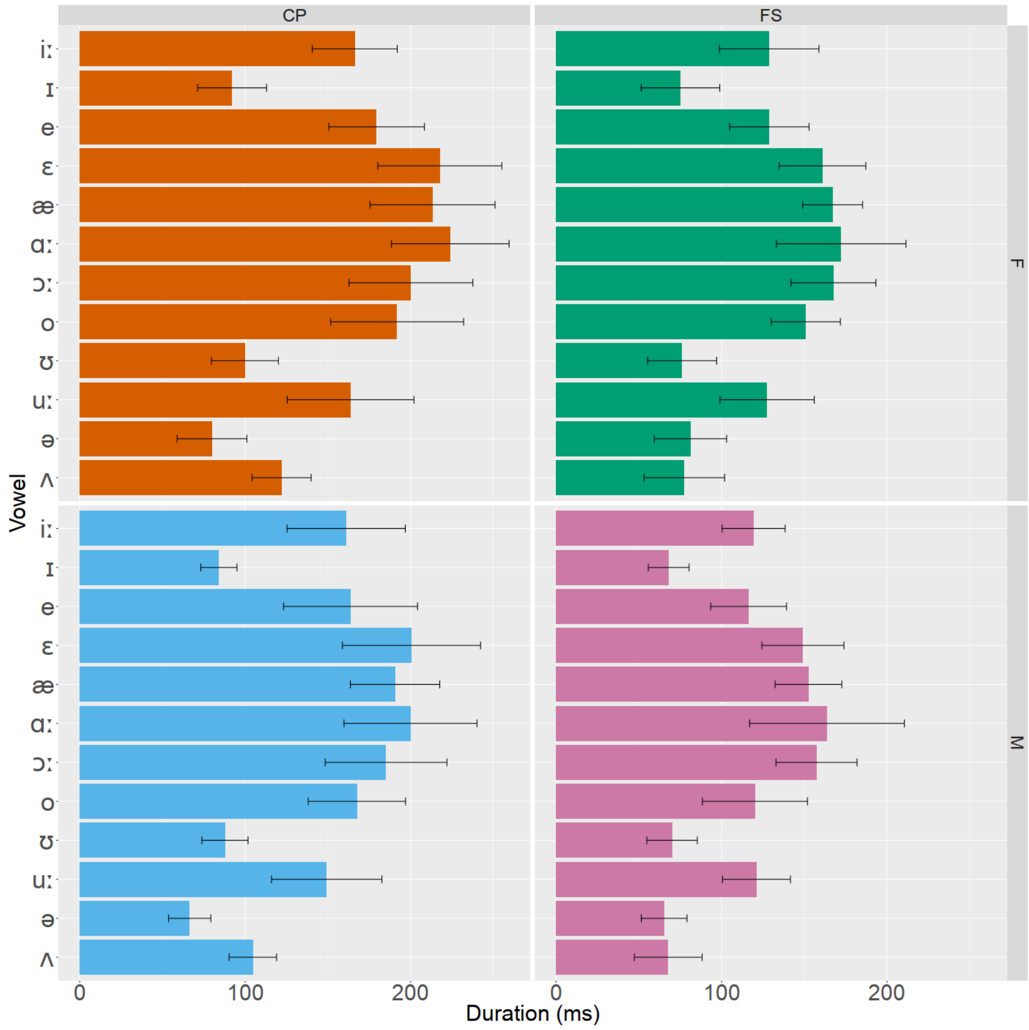



Figure 2. 13: Duration (ms) of 12 oral monophthongs for males and females in CP and FS context. Error bars denote +/-1 SD

Pairwise comparisons of the Gender × Context interaction using *lsmeans* (as discussed in Section 2.4.4) revealed that the effect of Gender on duration tends to counteract Context when moving from CP male to FS female (*FS,F – CP,M* in Table 2. 12). The other pairwise comparisons confirm a dominant effect of Context. The results of pairwise comparisons are shown in Table 2. 12.

Table 2. 12: Pairwise comparison of Gender × Context interaction using *lsmeans*

| Contrast | *t* test |
|---|---|
| CP,F - FS,F | t(4688.62) = 33.406, p < 0.0001) |
| CP,F - CP,M | t(24.34) = 1.858, p = 0.27) |
| CP,F - FS,M | t(25.03) = 5.608, p < 0.0001) |
| FS,F - CP,M | t(25.03) = -1.916, p = 0.25) |
| FS,F - FS,M | t(24.34) = 1.858, p = 0.27) |
| CP,M - FS,M | t(4688.62) = 33.406, p < 0.0001) |

In summary:

**For duration,** the combined interaction between gender, context and vowel was found to be non-significant, as was any interaction between gender and context. In terms of an R formula we have

*Context + Vowel + Gender + Context:Vowel + Vowel:Gender + (1 | Speaker)*

The results also showed that overall mean duration was longer for the vowels produced in CP context as compared to FS context. The vowels produced by female speakers were slightly longer than vowels produced by male speakers, but the difference was non-significant. The results for controversial vowel pairs are discussed below.

#### 2.5.3.1 Front Vowels

**[ɪ] and [e]**

As can be seen in Figure 2. 14, [ɪ] (80 ms) and [e] (147 ms) are distinct with regard to duration. The post-hoc *lsmeans* pairwise tests showed that the difference in duration between /ɪ/ and/e/ is significant (t(3353.40) = 33.928 , p < 0.0001).

**[æ] and [ɛ]**

The two mid-open vowels [ɛ] (181ms) and [æ] (184ms) appear to be similar in duration as shown in Figure 2. 15. The post-hoc pairwise tests showed that there is no significant difference in duration of these two vowels (t(3353.48) = -1.234, p = 0.98). These findings are in line with the impressionistic account of Kachru (1990 cited in Saleem et al., 2002:1) that "the front middle-low vowel [æ] exists as front middle low vowel [ɛ] in Urdu". However, they did not report any acoustic analysis, and therefore the findings from the current study cannot be compared quantitatively. In addition, Saleem et al. (2002) showed a spectrogram of [ɛ] in a disyllabic word [kɛhər]; however, there are no acoustic measurements to compare the quality of this vowel with other Urdu vowels.

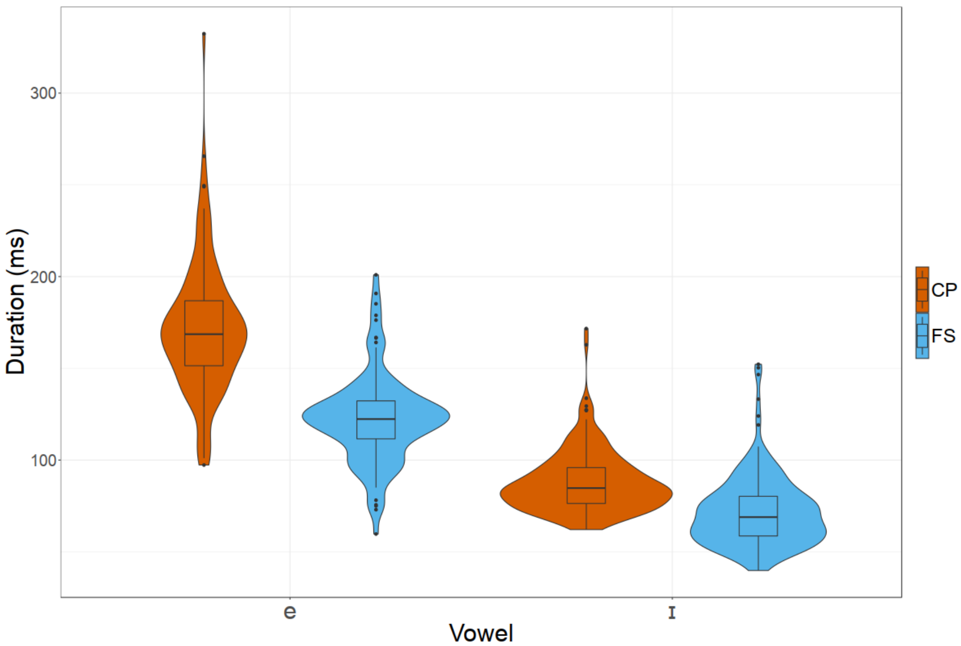



Figure 2. 14: Violin plot of the duration of [e] and [ɪ] in milliseconds, for each context.

### 2.5.3.2 Central Vowels

**[ə] and [ʌ]**

The mean duration values show that [ə] and [ʌ] are short vowels with [ə] being shorter (73 ms) than [ʌ] (93 ms) as shown in Figure 2. 16. The post-hoc pairwise tests showed that the difference in duration is significant (t(3353.53) = -10.015, p <0.0001).

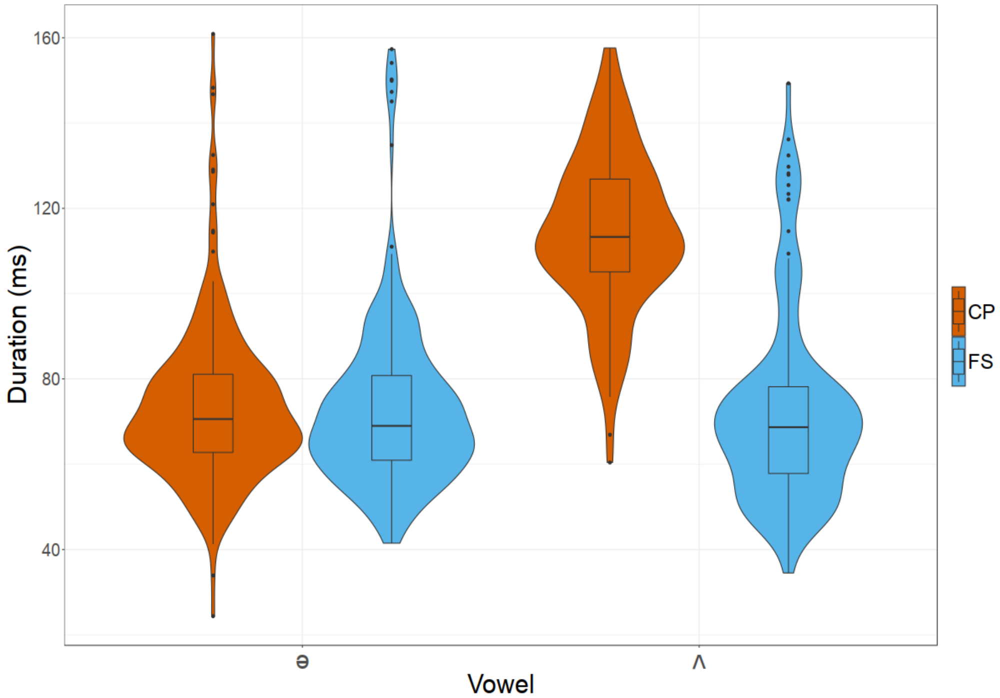



Figure 2. 15: Violin plot of the duration of [ə] and [ʌ] in milliseconds, for each context.

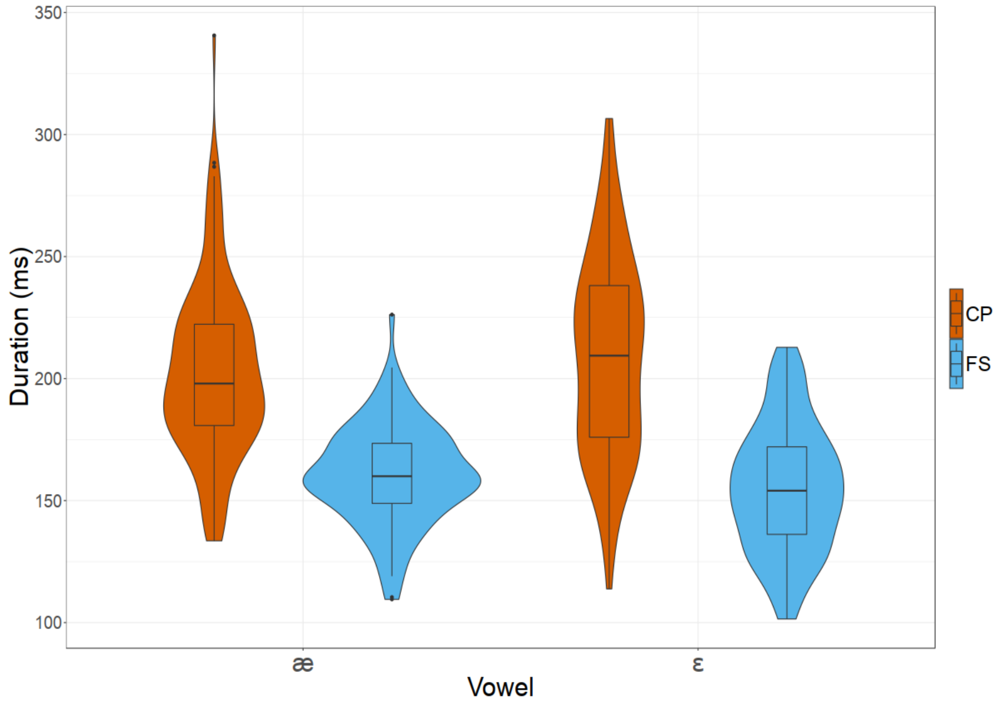



Figure 2. 16: Violin plot of the duration of [æ] and[ɛ] in milliseconds, for each context.

#### 2.5.3.3 Back Vowels

**[o] - [ɔː]**

As shown in Figure 2. 17, the two back vowels [ɔː] (178ms) and [o] (168ms) appear to be similar in duration; however the post-hoc pairwise tests showed that duration difference was significant ($t(3355.27) = -7.290$, $p < 0.001$). These two vowels have been reported as distinct in Urdu literature; however, there is no acoustic analysis available to compare these findings with. Hussain (1997) reported the mean duration of Urdu vowels produced by six speakers; however, he extracted this data from stressed and unstressed syllables and did not include [ɔː] in his investigation.

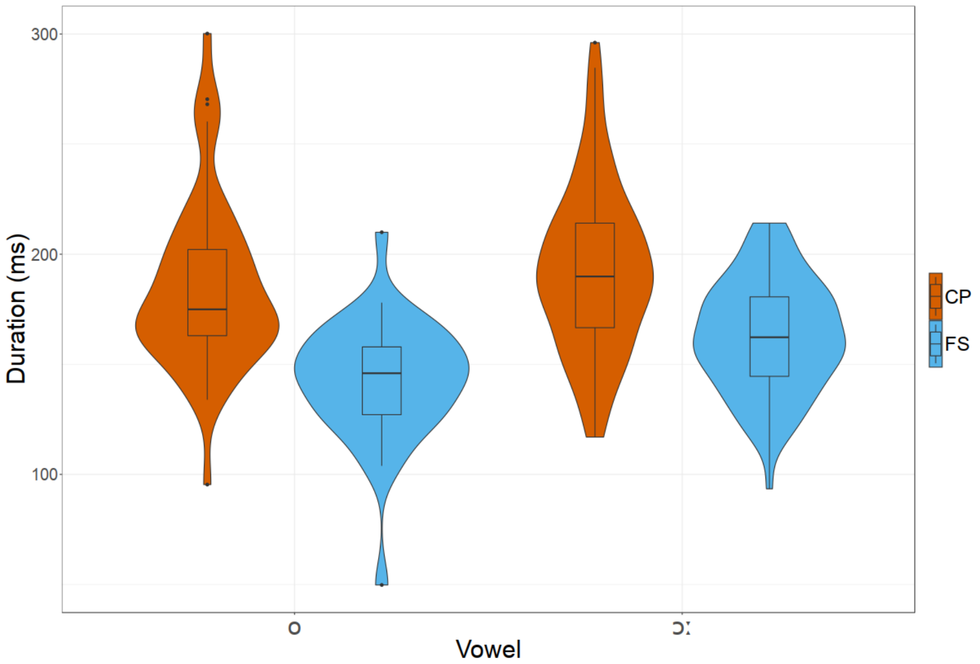



Figure 2. 17: Violin plot of duration of [o] and [ɔː] in milliseconds, for each context.

In summary, all speakers produced long and short monophthongs with a clear duration distinction, except for two open-mid vowels, front [ɛ] and [æ].

### 2.6 Discussion and Conclusion

In this section, based on the results of the present study the vowels that are phonemically distinct are presented in slashes and the allophonic vowels are presented in brackets (note this is in contrast to the notation adopted in Section 2.3.1.2). Urdu has long and short vowels which differ both spectrally and temporally. /ɪ/ and /e/ show extensive spectral overlap. However, these are distinct phonemes especially with regard to F1 and duration. Therefore, it can be concluded that these two vowels differ temporally and spectrally. In addition, the duration of /e/ suggests that it is not a short vowel. Also, phonologically these are two distinct phonemes as they can be found in minimal pairs, for example /peːʈ/ "stomach", and /pɪʈ/ "got beaten up". In addition, both of these phonemes can be found in word initial and medial positions, for example /eːk/ "one", /ɪk/ "once", /peːʈ/ "stomach", and /pɪʈ/ "got beaten up". /e/ can be found in monosyllabic open syllables, such as /seː/ "from", and /keː/ "of". However, /ɪ/ cannot be found in monosyllabic open syllables as Urdu does not allow open light (mono-moraic) syllables (Hussain, 1997 as discussed further in detail in Section 3.1).

[ɛ] and [æ] are reported as distinct phonemes in Urdu literature (Saleem et al., 2001; Fatima and Aden 2003; Raza, 2009). However acoustic and phonetic analysis and findings from the present study show that [ɛ] and [æ] do not differ spectrally or temporally. Both these vowels appear in the open-mid and close-mid front region of the vowel quadrilateral. We conclude that they should be considered a single vowel, transcribed as /ɛ/ and not /æ/. Phonemically, /ɛ/ can be found in closed syllables, and in disyllabic or tri-syllabic words in stressed open syllables, for example /ˈbɛ.ʈʰɑː/ "*sat down (he)*" - the first syllable is open and has /ɛ/ at the end.

The inconsistent symbols for the Urdu vowels in the available literature cause difficulties when trying to compare the results of the present study. The most confused symbols are [e], [ɛ] and [æ]. These symbols have been used for distinct vowels or as allophones of the same vowel in different studies. Saleem et al. (2002) states that the sounds [ɛ] and [æ] are allophones of the same sound, however Fatima and Aden (2003:74) state that [ɛ] is a long and [æ] is a short vowel or vice versa; Raza (2009) reported [ɛ] and [æ] as distinct vowels. However, neither Saleem et al. (2002) nor Fatima and Aden (2003) and Raza (2009) have given enough evidence to support their claims. Kachru (1990) used symbol /ɛ/ for a front open-mid vowel in his impressionistic account of Hindi-Urdu vocalic inventory; however, Hussain (1997) argued that the sound in Urdu is closer to /æ/ than /ɛ/. In the present study, [e], [ɛ] and [æ] were used in near minimal pairs /peːʈ/ *"stomach"* /bɛd/ "*willow tree" /* bæt̪/ *"to follow"* and the results show that [ɛ] is not distinct from [æ]; however, the results show that speakers distinguished /e/ from /ɛ/.

The example Saleem et al. (2002) gave in their impressionistic account of /ɛ/ was a disyllabic word [kɛhər], and they gave a spectrogram of this sound. However, the position of F1 and F2 in their spectrogram suggest that it is actually /e/. There are no acoustic measurements to draw a quantitative comparison with the present study. The quality of this vowel might be different if it is used in a disyllabic word and so can be considered an allophone of /æ/ in certain contexts; however, this argument requires further acoustic investigation including disyllabic words. Therefore, based on the findings from the present study, it can be concluded that [ɛ] and [æ] are not distinct phonemes neither phonemically nor phonetically. In fact, the results show that the Urdu spoken in Punjab does not have an /æ/ vowel, and /ɛ/ is significantly higher than /æ/ and lower than /eː/ in the Urdu vowel space, as shown in Figure 2. 4, Figure 2. 5 and Figure 2. 6. However, it is quite possible that these two vowels are found in other dialects of Urdu as allophones or distinct phonemes.

Besides inconsistent symbols, one of the biggest problems in the available literature is the lack of detailed information to compare with. In particular, acoustic analysis, formant frequencies, specific accent information or information about the participants' linguistic background. As a result, there is no way to compare the results from the present study with the studies reported in the literature. The acoustic properties of Urdu vowels (duration and F1 and F2 frequencies) reported by Hussain (1997) are not comparable with the findings from the present study as those values were extracted from stressed and unstressed syllables from disyllabic words from varying consonantal contexts. He reported that stress changed the quality of the vowel, such that open vowels appeared to be more open and closed vowels appeared to be more closed, if the vowel was in stressed syllable.

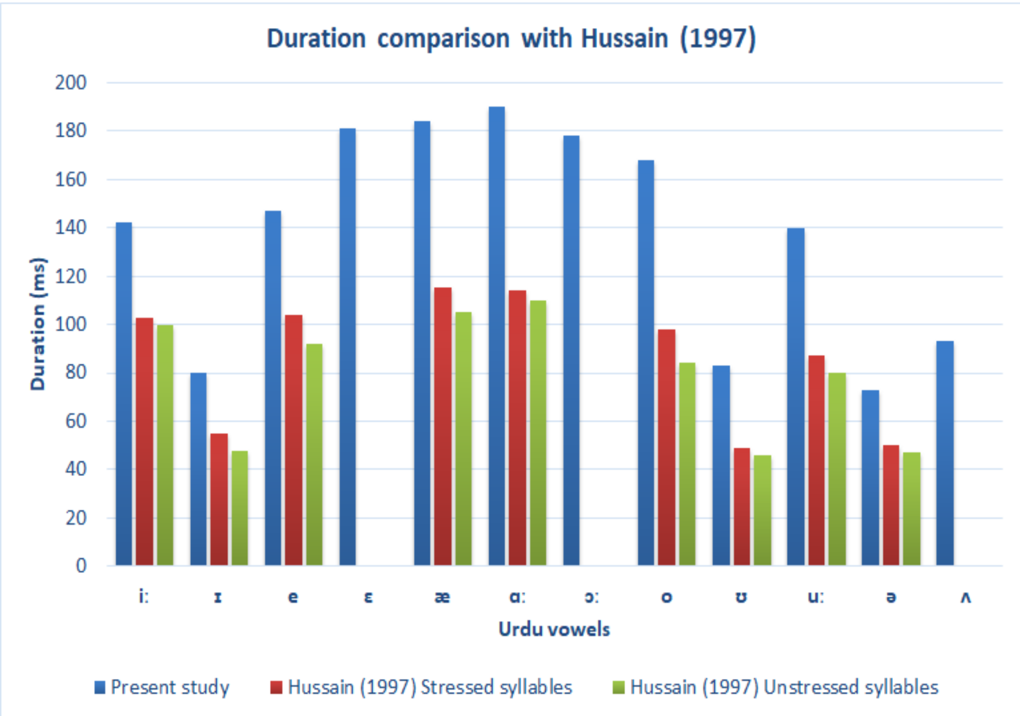



Figure 2. 18: A comparison of vowel duration (ms) from the present study with the vowel duration as reported by Hussain (1997)

For the comparison of vowels from the present study with Hussain (1997), all vowels show differences in mean F1 and F2 except for F1 for Urdu vowel /e/. These differences could be due to the context (present study) or limited number of speakers (Hussain, 1997). Hussain (1997) reported the duration of Urdu vowels from stressed and unstressed syllables, and his vowels in both stressed and unstressed syllables appear significantly shorter than the vowels in the present study. However, as shown in in Figure 2. 18, the distinction between long and short vowels can be seen across three contexts for those vowels that Hussain (1997) included in his study.

Whilst analysing six accents of Urdu in 136 districts of Pakistan, Farooq (2014) gave some acoustic and statistical analysis. However, her acoustic analysis showed only a range of formant frequencies for the four vowels, i.e. /i/ (200-2400 Hz), /æː/ (700-1700 Hz), /aː/ (700- 1100 Hz), and /uː/ (300-700 Hz). The formant frequencies analysed in the present study for [i], [æ], [ɛ], [ɑː] and [uː] fall in the ranges given by Farooq (2014), however those ranges are not very precise. For instance, the mean F1 for Urdu vowel [uː] as reported in the present study is 371 Hz. This value falls within the range given by Farooq (2014; [uː] (300-700 Hz); however, this is a rather broad range that could contain other distinct vowel formants.

Despite the fact that each member of the pair /æ/-/ɛ/ and /ɔː/-/o/ are treated in the literature (Fatima and Aden, 2003; Saleem et al., 2002) as distinct phonemes, in the present study these vowels show substantial spectral and temporal overlap, and statistical analysis confirms that these are not distinct phonetically. Saleem et al. (2002:3) contradicts this with an example of a disyllabic near minimal pair /sonɑ/ "*gold*" or *"to sleep"* /bɔnɑ/ "*to sow*"; however, the vowel sound in both of these words does not differ in quality at all, and does not change the meaning of the given words if pronounced with one or the other sound. In addition, there is no data given with which to compare the findings to the present study.

It is difficult to find minimal pairs in Urdu, which could mean that the sounds that occur at different positions in different words (initial, medial or final) are allophones, with slight variations, of the same sound. Therefore, further studies with minimal pairs can help to better establish the quality and existence of [ɔː] and [o]as distinct phonemes in the Urdu vocalic inventory.

It is quite possible that the duration of these [ɔː] and [o]/ may vary in disyllabic words, such as: /bɔːlɔː/ "*speak", /t̪ɔːlɔː/ "measure*", /kʰɔːlɔː/ "*open", /t̪ɔːɽɔː/ "break", /*ʤɔːɽɔː/ "*mend"*, /tʃʰɔːɽɔː/ "*leave*", /mɔːɽɔː/ "*bend"*. However, the change in sound from [o] to [ɔː], or vice versa, in the first or second syllable does not change the meaning. Therefore, these can be considered one phoneme in Urdu. The results of present study show that phonetically [ɔː] and [o] differ in duration; however, there are no strong arguments for treating them as distinct phonemes. It is likely that the studies reporting these two vowels as distinct phonemes obtained their data from groups of speakers with diverse first languages. Phonemically these two sounds do not contrast in Urdu spoken by Punjabi-Urdu speakers, therefore it can be concluded that [ɔː] and [o] are not distinct phonemes in Punjabi-Urdu.

Literature on Urdu vowels reports only /ə/ as a central vowel in Urdu; however, the present study shows that phonetically schwa [ə] and [ʌ] are distinct vowels and [ʌ] is a fairly open, central unrounded vowel, which has higher F1 and lower F2 than schwa [ə]. It is quite possible that this central vowel is pronounced differently by speakers of Urdu who speak different first languages. Although phonetically [ə] and [ʌ] vowels appear to be distinct with reference to duration, F1 and F2, phonologically these two sounds are not distinct. Based on the phonetic data analysis, it can be concluded that wedge /ʌ/ is used in closed monosyllabic syllabic words (CVC), for example, /bʌd̪/ "*bad*", /kʌb/ "*when*", /sʌb/ "*all*", /rʌb/ "*God*", /t̪ʌb/ "*then*", /dʒʌb/ "*when*", /ʃʌk/ "*doubt*" /rʌʃ/ "*busy*", /hʌt/ "*get aside*", /mʌt̪/ "*sanity*". /ə/ is used in disyllabic or tri syllabic words, for example, /sə.nəd̪/ "*certificate*", /ɣə.zəb/ "*wrath*", /rə.dʒəb/ "*7th month in Islamic calendar*", /sə.bəb/ "*cause*", /hə.məd̪/ "*hymn*", /və.dʒɑː/ "*reason*" and so on.

Recall that [ə] was extracted from the first and second syllable in a disyllabic word (see Section 2.4.1). Therefore, it can be considered an allophonic realisation of /ʌ/ in disyllabic words or in unstressed syllables. Based on the phonetic analysis from the present study and phonological information as discussed above, it can be concluded that [ə] and [ʌ] are allophonic in Urdu where [ʌ] is fairly open, central unrounded vowel with higher F1 and lower F2 than schwa [ə], and a better transcription for this vowel would be /ɐ/. In the present study, the results also show that context, speaker and gender had very minimal effect (except for the [ɔː] and [o] differences seen in the male data) - in most cases no effect, as shown in Figure 2. 4, Figure 2. 5, Figure 2. 6 and Figure 2. 13 in Section 2.4.

## 2.7 Summary – Urdu Vowel System

In summary, the analysis of formant frequencies and duration show that Urdu has nine distinct vowels: six long and three short vowels. Long vowels are significantly longer than short vowels and appear to be more peripheral in quality than short vowels. In addition, short vowels do not occur in open syllables (CV) in Urdu. Urdu long-short vowel pairs are given below and shown in Figure 2. 19 (F1/F2 vowel space) and Figure 2. 20 (duration):

| Long | | Short | |
|---|---|---|---|
| Vowel | gloss | vowel | gloss |
| biːt̪ | "*spent/pass*" | bɪd̪ | "*new*" |
| peːʈ | "stomach" | pɪʈ | "*get beaten up*" |
| bɛːd̪ | "*willow tree*" | bɪd̪ | "*new*" |
| bɑːd̪ | "*later*" | bɐd̪ | "*bad/evil*" |
| buːdʒʰ | "*guess*" | bʊdʒʰ | "*put out*" |
| /poːd̪/ | "*offspring*" | | |

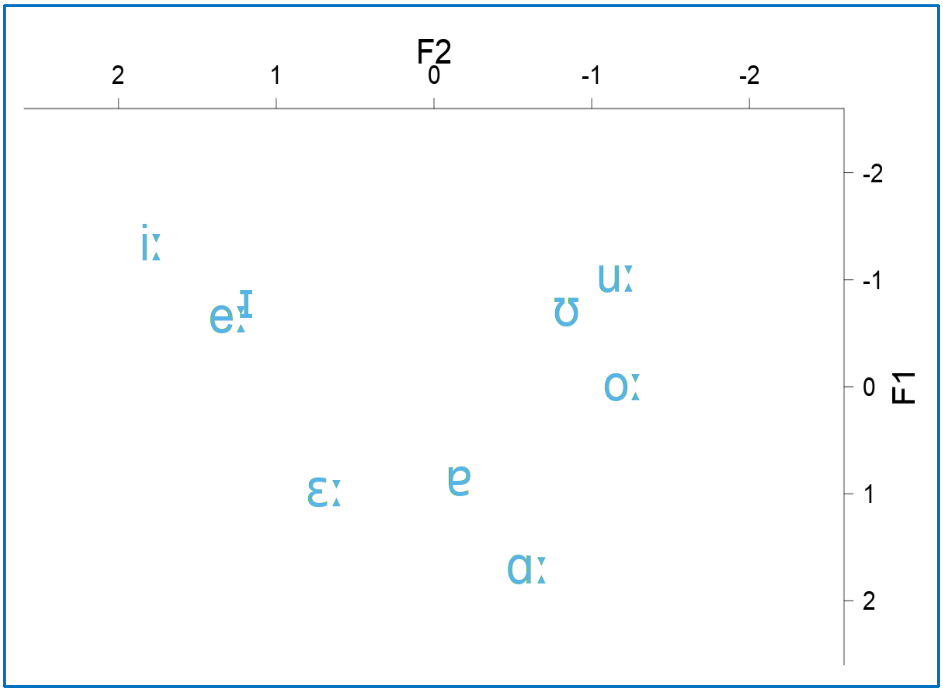



Figure 2. 19: Mean Lobanov normalised formant frequencies of the 9 monophthongs in F1/F2 vowel space, across all speakers and both contexts.

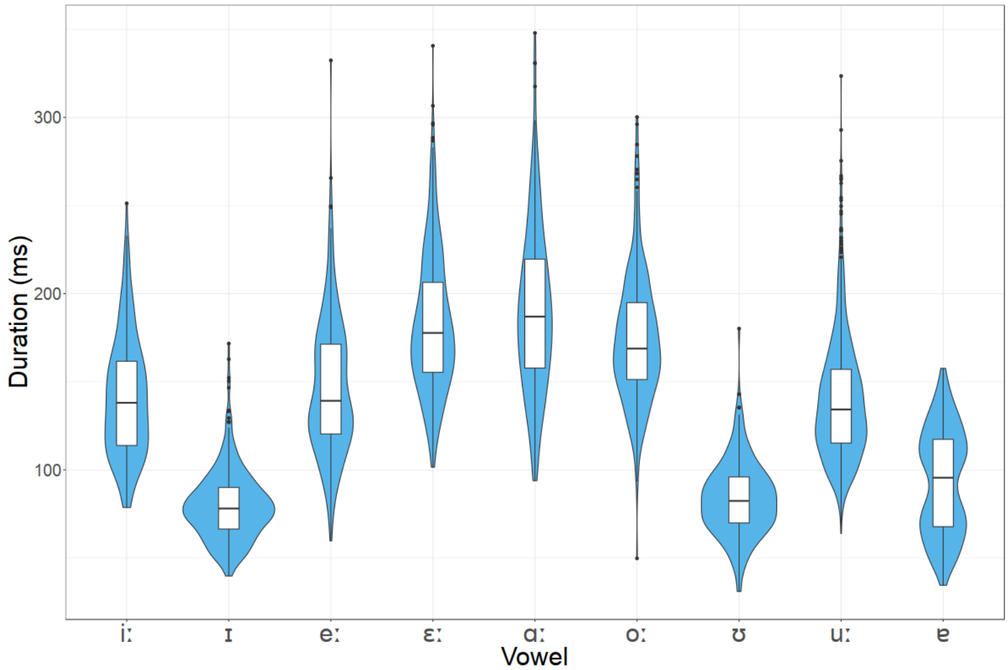



Figure 2. 20: Violin plots of the duration (ms) of 9 oral monophthongs across all speakers and both contexts

With regard to distinctive features, i.e. front, back, high and low, Urdu oral monophthongs can be placed in the vowel quadrilateral as shown in Table 2. 13.

Table 2. 13: Monophthongs of Punjabi-Urdu

| | Front | Centre | | Back |
|---|---|---|---|---|
| Close | iː | | | uː |
| | ɪ | | | ʊ |
| Close-mid | eː | | | oː(ɔː)* |
| Open-mid | ɛː | (ə)** | | |
| | | ɐ | | |
| Open | | | ɑː | |

* The vowels in parenthesis are not phonemic and share the acoustic properties with adjacent vowels outside the parentheses.
** This is an allophonic counterpart of the central mid open vowel /ɐ/ as it is shorter in duration and different with regard to F1 and F2, but not contrastive phonemically.

Following Wells (1982) lexical set for English vowels: an Urdu lexical set is proposed and given in Table 2. 14. These words cannot be mistaken for any other words, except /peːʈ/ if it is pronounced out of context. Just like Wells' lexical set, where possible these words end in a voiceless alveolar or dental consonant.

Table 2. 14: Lexical set of Urdu spoken in Punjab Pakistan

| Vowel | Lexical set Transcription | Gloss | Lexical set in Urdu script |
|---|---|---|---|
| iː | biːt̪ | *"success"* | جیت |
| ɪ | bɪk | *"to sell"* | بک |
| eː | peːʈ | *"stomach"* | پیٹ |
| ɛː | bɛːt̪ | *"to follow"* | بیعت |
| ɐ | bɐd̪ | *"bad"* | بد |
| ʊ | bʊd̪ʰ | *"Wednesday"* | بدھ |
| ɑː | bɑːd̪ | *"after"* | بعد |
| oː | boːl | *"speak"* | بول |
| uː | kuːd̪ | *"jump"* | کود |

This is the vowel system of Urdu as spoken in Punjab Pakistan. /ɔ/ and /æ/ are not found in this system, and [ə] is found as an allophonic variation of /ɐ/. The status of diphthongs in this Urdu vowel system is discussed in the following chapter (Chapter 3).

**Chapter 3**

## Urdu Diphthongs

This chapter reports on the production experiment investigating the phonological and phonetic properties of Urdu diphthongs. As discussed in Chapter 2, there are many disagreements about the phoneme inventory of Urdu and this is particularly evident with respect to the vowel inventory. Regarding diphthongs, most of the literature agrees that Urdu does not have them phonologically (i.e. as distinct phonemes) (Bokhari, 1991; Waqar and Waqar, 2002; Khurshid, Usman and Butt, 2003; Bhatti and Mumtaz, 2016).

Phonemically, the existence of diphthongs (using minimal pairs) cannot be proved, though "…their phonetic existence however remains undocumented" (Waqar and Waqar, 2002:16). The literature disagrees on how these diphthongs are formed: some claim that diphthongs are formed "as a result of deletion of either a consonant" or "a timing slot" if both vowels are long (Waqar and Waqar, 2002:20; Sarwar, Ahmed and Tarar, 2003), and others claim that short vowels are replaced by long vowels when word final consonant deletion results in the elongation of the preceding short vowel (Bhatti and Mumtaz, 2016; Waqar and Waqar, 2002; Wali, 2001).

Studies on second language perception report that perceptual similarities/dissimilarities between the phonetic and phonological system of L1 compared with the target language (Escudero, 2005; Best and Tyler, 2007), play a significant role in the perception and production of L2 phonemes. The phonetic investigation of the presence of diphthongs in Urdu (especially those similar to English diphthongs) can therefore help English language learners to perceive and produce the English diphthongs. This is one of the aims of the present study.

Before presenting the experimental design, analysis and results, a brief overview of the literature describing the status of diphthongs in Urdu is given below.

### 3.1 Background - Urdu Diphthongs

We first present an overview of literature on the syllable structure in Urdu, which is a prerequisite for most of the studies on Urdu diphthongs.

#### 3.1.1 Syllable Structure in Urdu

Hussain (1997:42) reported the following syllable templates in Urdu: CV (this open syllable with short vowel is not allowed in word final position), CVC, CVCC, CVV, CVVC, CVVCC. According to Hussain,

> "Long vowels are bi-moraic, short vowels are mono-moraic, consonants clusters in coda position are also bi-moraic. Therefore, open syllables with short vowels are monomoraic, closed syllables with short vowels and open syllables with long vowels are bimoraic and closed syllables with long vowels or with short vowels and a coda cluster are tri-moraic" (1997:44-45).

This means that Urdu phonology has a three-way quantity distinction and these syllables are labelled as light (L) monomoraic, heavy (H) bimoraic, and super heavy (S) tri-moraic. According to Hussain (1997), Urdu syllables have constraints on the coda and onset clusters. For example, Urdu does not allow more than two consonants in onset and coda position. The second consonant in the onset position is restricted to the glides /w/, /j/ and /h/. The second consonant in the coda position is limited to stops, and first consonant in the coda position is limited to a voiceless fricative /f/, /ʃ/ or /x/ or nasals /n/ or /m/ (Hussain, 1997:42). Hussain (1997) also reported that stress is not fixed in Urdu and stress assignment is sensitive to vowel length.

Ghazali (2002:190) presented 11 syllable templates in Urdu: CV, CVC, CVCC, CVV, CVVC, CVVCC, V, VC, VCC, VV, and VVC, and claimed that the first six templates are underlying and the remaining templates are derived (i.e. surface representations). According to Ghazali (2002) every onset-less syllable in Urdu has a glottal stop /ʔ/, which is not realised in pronunciation. Hence VC syllables are not common, but such syllables do exist. However, his examples for V and VC syllable templates do not show the presence of a glottal stop /ʔ/ before the vowel, for instance, from Ghazali (2002:200):

/ɪ.ləm/ (*knowledge*) /ɪn.sɑːf/ (*justice*) /saː.ət/ (*a moment*)

V. CVC VC.CVVC CVV.VC

(*long vowels are represented by VV*)

Ranjha (2012) agrees with Ghazali (2002) with regard to syllable templates. He additionally claimed that sometimes the syllable templates undergo changes due to the deletion of certain segments. For example, in the Urdu word /ʔəbr/ "cloud", the syllable structure is CVCC; however, as reported by Ghazali (2002), the word initial /ʔ/ is not realised in pronunciation; hence the syllable template surfaces as VCC. This syllable template contradicts with Hussain (1997), especially with regard to coda cluster.

Ranjha (2012) reported that the maximum number of syllables in Urdu is three. In addition, there are limitations on the number of consonants in onset and codas position in a syllable. He further reported that only two consonants are allowed in the coda, for example, /ərzmə̃d̪/ "applicant" in Urdu is syllabified as /ərz.mə̃d̪/. He added that this pattern of syllabification follows the sonority sequence principle (SSP), because in the coda of the first syllable (i.e. /-rz/), liquid /r/ is more sonorant than fricative /z/ (Ranjha, 2012:31). Ranjha (2012) reported that the most frequent coda clusters were /st/, /rd/ and /xt/. Although Ranjha (2012) presented arguments that Urdu syllable structure follows descending sonority order in coda clusters, other studies (e.g. Nazar, 2002) disagree as discussed below.

Nazar (2002:191-194) analysed 5,000 words from an Urdu dictionary *"Jaibi Feroze-ul-Lughat, Feroze Sons, Lahore"*, and reported that CV(V)(C) is the most common syllable template in Urdu, and the second most frequent is CV(V)(CC). He further reported that

V(V)(C) are the least frequent and onsetless syllables such as VVCC are prohibited in Urdu. He mentioned the constraints on the coda clusters especially with regard to SSP but did not explore this further, except to argue that the descending sonority order in coda clusters does not apply to all syllables in Urdu. Nazar (2002) also mentioned that Urdu does not allow complex onset clusters as reported by Hussain (1997), which was attributed to Hussain's biased approach due to his exposure to the English language or the re-syllabification of English loanwords in Urdu, which require extensive phonetic investigation.

Bokhari and Pervez (2003) reported syllabification and re-syllabification rules and patterns for Urdu words. For most patterns they agree with the literature discussed above and explain the rules for derived templates, where the basic syllable template changes due to insertion or deletion of a vowel segment. Their data consisted of 1000 Urdu words, and they asked three native speakers of Urdu to syllabify those words. They reported that Urdu has a very simple syllable structure as it allows only one consonant in onset position and a maximum of two consonants in coda position, and re-syllabification only occurs when a vowel is either deleted or inserted in any given word. They further reported that a glottal stop /ʔ/ gets deleted in the syllable initial and final position, and this deletion of the glottal stop /ʔ/ results in changes in syllable structure but does not affect the number of syllables. For example, /ʔo.rət̪/ "*woman*" has syllable structure CV.CVC, and after the deletion of /ʔ/ the syllable structure becomes V.CVC; the syllable structure of /bər.ʔəks/ "*opposite*" is CVC.CVCC, and after the deletion of /ʔ/ the syllable structure changes to CV.CVCC. They also reported that unlike English, the nuclei in syllables are only vowels, and diphthongs are also considered a nucleus. For example, in [koi] "any" is a monosyllabic word. Overall, this study helped to clarify the concept of underlying and derived syllable templates; however, most of the given examples are misprinted (i.e. IPA symbols are missing) and therefore the arguments are not clear.

The syllabication patterns in Hindi-Urdu as discussed by Kachru (1987:402) are similar to the ones discussed in the literature above; however there are some differences, for instance, /pɑː.jeː/ is syllabified as /pɑː.eː/ by Kachru (1987:402). Although very little is known about the stress patterns in Urdu, stress is not distinctive in Urdu except for some grammatical forms, such as past participle form of verb [ˈdʒəlɑː] "*burnt*" and infinitive form of verb [dʒəˈlɑː] "*to burn*" (Nyyar, 2002; Hussain, 1997). Stress can help to distinguish the status of the Urdu diphthongs. Hussain (1997) proposed an algorithm for stress patterns in Urdu; however, this is very different from the stress patterns in Hindi-Urdu discussed by Kachru (1987:402). In addition, the stress patterns reported by Hussain (1997) are also not considered very comprehensive and further investigation is required.

Based on all the arguments discussed above, we can conclude that that the syllable structure in Urdu is (C)VV(C)(C). As mentioned above, most studies on Urdu diphthongs refer to the syllable structure in Urdu. A review of the literature on Urdu diphthongs can now be presented.

### 3.1.2 Literature on Diphthongs in Urdu

Waqar and Waqar (2002) reported 13 diphthongs in Urdu. They reported that phonemically diphthongs do not exist in Urdu as there are no minimal pairs to show that contrast; however phonetically the deletion of any one of the three consonants /ʔ/, /j/and /v/ in a disyllabic word results in a diphthong. For example, deletion of /ʔ/ from /nə.ʔeː/ نئے *"new"* results in diphthong [nəeː]; deletion of /j/ from /ke.jaː/ کیا "*what"* results in diphthong [keaː]; and deletion of /v/ from [hʊ.viː] ہوئی "*happened"* results in diphthong [hʊiː].

The experimental design of Waqar and Waqar (2002) is not very clear. For the perception of diphthongs, they prepared a list of words containing monophthongs and possible diphthongs. For the identification task they first trained 25 native speakers of Urdu on how to identify the number of syllables in a word. They did not give any details of the speakers' backgrounds or the list of words carrying the possible diphthongs and monophthongs. The criteria for a vowel-vowel sequence to be considered a diphthong was whether more than half of listeners perceived a word carrying a vowel-vowel sequence as monosyllabic. After the perception task, the words that were perceived as monosyllabic (i.e. carrying diphthongs) were recorded by five male native speakers of Urdu in carrier phrases for further acoustic analysis. Out of 22 words carrying vowel-vowel sequences, listeners perceived 13 as monosyllabic. Their acoustic analysis of these 13 diphthongs is based on onglide (first vowel) + offglide (second vowel) duration and F1 and F2 at the onset and offset of these vowels. It is not clear how this data was measured as they did not provide any information. Their results show that Urdu has rising diphthongs, for instance the onglide duration is only 39% and offglide duration is 61%. They defined *rising* diphthongs as follows:

> "If the first vowel of the diphthong is prominent, it is called a falling diphthong, and if the second vowel of the diphthong is prominent, it is called a rising diphthong". (Waqar and Waqar, 2002:19)

According to Waqar and Waqar (2002:19), in Urdu a syllable cannot start with a vowel "except word initially"; hence the vowel-vowel sequences are treated as diphthongs instead of vowel sequences belonging to different syllables. Therefore, after the deletion of the consonant /ʔ/, the remaining word looks like [nə.eː], where /eː/ cannot stand alone as a syllable. Hence /eː/ gets merged with the preceding syllable and forms a diphthong, [nəeː], with a syllable structure CVVV, where the first vowel is shorter (mono-moraic) than the second vowel (bi-moraic). They conclude that the diphthong structure in Urdu is VVV, where the first component is short and the second component is long.

Waqar and Waqar (2002) insist on the deletion of a time slot and argue against a vowel sequence of two long vowels, VVVV. For example, they reported that in the disyllable Urdu word /ko.ʔi/ "*any*" the syllable structure is CVV. CVV; however, after the deletion of /ʔ/ the syllable structure will be like CVV.VV. Urdu does not allow onset-less syllables except word initially, so this word will go through re-syllabification and the preceding vowel will have to lose a time slot in order to form a legitimate syllable. Their definition of legitimate syllable structure contradicts with the literature (as discussed in Section 3.1.1). In a phonological study of Urdu, Wali (2001) argues that deletion of /h/ and /ʔ/ in word final position results in the elongation of a preceding short vowel, which contradicts Waqar and Waqar (2002), as they report a deletion in time slot. Wali (2001:256) further claimed that "if {ʔ} occurs in the middle of the word, it may sometimes generate diphthongs as in [məsɑʔɪl] "*problems"* [məsɑɪl]".

In a follow-up study, Sarwar, Ahmed and Tarar (2003) followed the same methods as employed by Waqar and Waqar (2002) except for the number of speakers (i.e. 3 males and 3 females) and listeners (i.e. 30). They reported that there are 17 diphthongs in Urdu and claimed that diphthongization results in the loss of a time slot. For example, they reported that in the Urdu word /dʒɑ.ʔo/ "go" with CVV.CVV structure, after the deletion of /ʔ/ the syllable structure becomes CVV.VV. As onset-less syllables are not allowed in Urdu, the second syllable merges with the preceding syllable and becomes CVV. This structure contradicts with the ones discussed by Waqar and Waqar (2002), i.e. CVVV. Further, they discussed the individual variation between speakers and reported that the diphthong in [dʒɑo] "go" was rising (i.e. the second vowel was longer than the first vowel) for male speaker A and falling for male speaker B (i.e. the first vowel was longer than the second vowel). Overall, their arguments lack clarity, exacerbated by numerous typographical mistakes. In addition, they did not provide any conclusive acoustic and statistical evidence to support their arguments about the deletion of a time-slot in the production of a diphthong.

Khurshid, Usman, and Butt (2003) reported the possibility of diphthongs and triphthongs in Urdu and claimed that out of a list of 37 possible diphthongs in Urdu, 18 were identified as diphthongs by 20 native speakers of *Lahori* Urdu (as in, Urdu as spoken in the city of Lahore, where speakers mostly have Punjabi as their first language). They trained the participants for syllable identification, then gave them a list of words carrying the possible diphthongs (i.e. vowel-vowel sequences) and asked them to syllabify those words. If more than 50% of the participants syllabified a word containing a vowel-vowel sequence as monosyllabic, they considered those vowel-vowel sequences as diphthongs. Based on the syllable identification task, they concluded that there were 18 diphthongs but no triphthongs. They also concluded that individual differences play an important role in syllabification and perception of a given set of words. For example, some speakers identified as few as 7 and others identified as many as 32 diphthongs out of a set of 37 words carrying vowel-vowel sequences.

After the identification task, the identified words were recorded by 3 males and 3 female speakers for acoustic analysis. Their acoustic analysis was based on the duration in milliseconds to identify if a sound was a diphthong or two separate vowels. The baseline for the verification of a diphthong was the maximum duration for a long monophthong, which they set at 350ms. Therefore, if a vowel-vowel sequence was pronounced within this duration, it was considered a diphthong, otherwise these were considered two separate long vowels. Khurshid et al. (2003) reported that in all cases the duration was below 350 ms for diphthongs and less than 150 ms if both vowels in the diphthong happen to be the short vowels. They summed the duration of two separately recorded monophthongs for each speaker and compared that sum with the duration of the corresponding diphthong. For example, [oe] (325 ms) was compared with the sum of /o/ (234 ms) and /e/ (236 ms), which turned out to be less than the average summed duration of two separate monophthongs.

Lastly, contrary to Waqar and Waqar (2002) and Sarwar et al. (2003), Khurshid et al. (2003:18) reported that Urdu has very few words containing two consecutive short vowels and "Majority of the diphthongs identified by the native speakers of Urdu contained two long vowels or one long and one short vowel". They further added that in their list of 37 words they had only two words which contained two consecutive short vowels and only one of those was identified as a diphthong by the participants. A particular limitation of this study is that they did not perform any spectral analysis, and based their results solely on temporal analysis. They did not present any phonological arguments and the syllable identification task was done on a paper where participants read the words written in Urdu script.

Bhatti and Mumtaz (2016) present a follow up of three previous studies (Waqar and Waqar, 2002; Sarwar et al. 2003; Khurshid et al., 2003) plus two additional diphthongs, [eaː] and [aːe]. For the possible 26 diphthongs they recorded 78 words: three words per diphthong, produced by three male and three female speakers in a carrier phrase *"I said…."*. Five male and five female speakers took part in the perceptual identification task. Based on acoustic analysis (F1, F2, F3 at three different points: on-glide, transition and off-glide) and a perceptual identification task, where listeners were asked to identify the number of syllables in each word (70% votes or more), they concluded that Urdu has 16 diphthongs and five of these diphthongs are nasalised. Their acoustic analysis was mainly based on F1 and F2 of the first and second segment and the total duration of each diphthong. They reported that diphthongs behave like monophthongs in stressed and unstressed syllables and the maximum duration of a diphthong in an unstressed syllable was 148 ms. Hence they rejected a nasalised diphthong (i.e. [ɑːĩː]) due to its longer duration.

Bhatti and Mumtaz (2016) disagreed with Waqar and Waqar (2002) and Sarwar et al. (2003) about the deletion of a time-slot to form a diphthong, since none of the diphthongs they found were formed via reduction of a time slot. Bhatti and Mumtaz (2016) reported a variety of combinations of long and short vowels in diphthong formation, such as: long-short, short-long, or long-long vowels. However, they did not give any temporal or spectral information to support this claim. Further, they claimed that in the formation of diphthong, /ə/ and /j/ were replaced by /æ/, and front vowels /ɪ/ and /j/ were replaced by /e/; however, they did not give any suitable examples and analysis to support this claim.

Farooq and Mumtaz (2016) investigated Urdu phonological processes in connected speech. They analysed 13,717 words for multiple pronunciation and reported that segment *alternation* (short vowel to long vowel), *deletion* (a consonant or vowel) and *insertion* (mainly vowels to break consonant clusters) occurs at multiple levels due to a number of factors, such as syllable structure and stress. The most interesting findings were: segment deletion never occurs at word initial position; consonantal deletion can result in the elongation of the preceding short vowel; /j/ deletion occurs word medially, usually to form a diphthong (e.g. /keːjaː/ "*what*" becomes [kæaː]); and /v/ deletion occurs intervocalically to form a diphthong and converts a disyllabic word into monosyllabic (e.g. /hʊ.viː/ "*happened*" becomes [huːiː]). They also reported that sometimes /v/ deletion occurs in unstressed syllables and does not result in the formation of diphthongs.

To summarise, despite contradictory arguments, most of these studies agree on the basic syllabic structure in Urdu; however, there are a number of disagreements on the onset and coda constraints as well as the re-syllabification (i.e. surface representations) of the underlying syllables. For instance, some studies (Ghazali (2002; Bokhari and Pervez,

2003) reported the deletion of glottal stop at word initial a position, while others (Farooq and Mumtaz, 2016) reported that segment deletion does not occur at word initial position. Most of these studies agree that intervocalic deletion of /j/, /v/ and /ʔ/ converts the disyllabic words into monosyllabic words, and hence results in diphthongs. According to Ren (1986 cited in Aguilar, 1999:72), the distinction between diphthong and vowel-vowel sequence (hiatus) should be reflected acoustically, because hiatuses are two vowels from two syllables whereas diphthongs are two vowels in the same syllable. None of the studies on Urdu diphthongs mentioned above analysed the acoustic data with regard to this distinction; however, they do insist on the deletion of a time-slot and re-syllabification of disyllabic words into monosyllabic words. In addition, vowel-vowel sequences (hiatus) is reported to have a quicker transition than a diphthong (Quilis, 1981 cited in Aguilar, 1999:72). This was another parameter which none of the studies on Urdu diphthongs mentioned above explored.

Keeping in view these contradicting arguments from phonetics and phonology in the literature, the present study was designed to investigate the acoustics of the Urdu vowel-vowel sequences claimed as diphthongs in the literature. To this end, six vowel sequences resulting from the deletion of /j/ and /w/ were selected as candidate diphthongs and will be referred to as diphthongs in this study for the ease of reference.

### 3.1.3 Aims and Objectives

Besides all the contradictions, the studies discussed above indicate that it is possible to investigate the phonetic existence of diphthongs in Urdu. The phonetic investigation of diphthongs is required to test the predictions proposed by SLM, PAM, and PAM-L2 for the perception of SSBE vowels as discussed in Chapter 4, Chapter 5 and Chapter 6. The literature on Urdu phonetics and phonology suggests that diphthongs arise through the deletion of /j/, /v/ and /ʔ/ and restructuring of the syllable (Sarwar et al., 2003; Waqar and

Waqar, 2002; Farooq and Mumtaz, 2016). This description conforms to the arguments presented by Ren (1986 cited in Aguilar, 1999:72):

> "…[T]he idea is that the syllable components are planned before the phonetic realisation; so, the acoustic result of hiatus (i.e. two vowels in two syllables) has to be necessarily different from the acoustic result of diphthongs (i.e. two vowels in a syllable), which requires a restructuring in time and frequency to adjust both vocalic segments to the syllable frame".

The present study analyses the acoustic properties of six vowel-vowel sequences (diphthongs) in order to answer the following questions:

a) What are the phonetics of Urdu *vowel-vowel sequences*?

b) Does Urdu have diphthongs or vowels in hiatus?

c) Does Urdu have rising and/or falling diphthongs? (The rising and falling diphthongs will be discussed with regard to the definition provided by Waqar and Waqar (2002:19) as discussed in Section 3.1.2)

d) Is the off-glide (second segment) always longer than the on-glide (first segment) in these vowel-vowel sequences (or diphthongs)?

### 3.2 Methods and Procedures

In this chapter, the experimental design for the production experiments mainly follows on from the previous chapter. The methods and procedures are the same as discussed in Chapter 2 on Urdu monophthongs. The materials for diphthongs are presented below.

#### 3.2.1 Speakers

The data reported here come from 22 speakers (11 males and 11 females), as explained in Section 2.3.

### 3.2.2 Materials

Based on the reasons and rationale discussed in Section 3.1.1, 3.1.2 and 3.1.3, six diphthongs (resulting from the deletion of /j/ and /w/) were selected for investigation: ([ɑe] from /ɑje/, [oe] from /oje/, [ɑʊ] from /ɑwo, [ɪɐ] from /ɪjɑ/, [eɐ] from /ejɑ/ and [ʊɑ] from /uwɑ/). The main motivation for this selection was as a prerequisite for the subsequent perception experiment of Standard Southern British English (SSBE) vowels by Punjabi-Urdu speakers (as discussed in Chapters 4, 5 and 6). In other words, putative diphthongs were selected that have similarity with established diphthongs of SSBE.

#### 3.2.2.1 Minimal and Near Minimal Pairs

These six diphthongs were embedded in six monosyllabic minimal or near minimal pairs. The syllable structure for diphthongs was CV ($C_1V$) where the $C_1$ was bilabial plosive /p/ and /b/ in four words, dental plosive /d̪/ in one word and velar plosive /g/ in one word. The list of the words is given in Table 3. 1.

Table 3. 1: Urdu words carrying possible diphthongs

| Phonetic transcription | *Gloss* | Urdu words |
|---|---|---|
| [pɑe] | *"goat's trotter"* | پا ئے |
| [boe] | *"sow"* | بوئے |
| [pɑʊ] | *"gain"* | پاؤ |
| [pɪɐ] | *"lover"* | پیا |
| [geɐ] | *"went"* | گیا |
| [dʊɑ] | *"prayer"* | دعا |

#### 3.2.2.2 Carrier Phrases and Sentences

As discussed in Chapter 2, the test words were embedded in two types of sentences: (1) a standard carrier phrase (CP) of the form "I will say ___ once" (e.g. [mɛ̃ ɪseː geɐ kəhõ giː]), and (2) longer and more varied full sentences (FS). The list of sentences is given below in Table 3. 2.

Table 3. 2: Urdu Diphthongs in standard carrier phrases and sentences

| Phonetic transcription | Gloss | Urdu sentence<br>Carrier Phrases (CP)<br>Full sentence (FS) |
|---|---|---|
| [mɛ̃ ɪse pɑe kəhõ gɪ] | *"I will say* pɑe *once"* | میں اسے پا ئے کہوں گی |
| [je bəkre ke pɑe hæ̃] | *"these are goat's trotters"* | یہ بکرے کے پا ئے ہیں |
| [kɪsɑn ne kʰetõ mẽ bɪdʒ bɔe] | *"farmer sowed seeds in the fields"* | کسان نے کھیتوں میں بیج بوئے |
| [kʰəbʰi səkũ: nəhĩ pɑʊ ge] | *"you will never find peace"* | کبھی سکون نہیں پاؤ گے |
| [dʒɔ pɪɐ mən bʰɑe] | *"the one who is loved by her lover"* | جو پیا من بھائے |
| [wɔ ʧəlɑ geɐ] | *"he went away"* | وہ چلا گیا |
| [je merɪ mɑ̃ kɪ dʊɑ hɛ] | *"this is my mother's prayer"* | یہ میری ماں کی دعا ہے |

### 3.2.3 Segmentation and Annotations

The methods and procedures for segmentations and acoustic measurements are discussed in Sections 2.3 and 2.4. Measurements were made using Praat (Boersma and Weenink, 2016).

For the diphthongs that are in open syllables, and the following word starts with a stop consonant /k/, it is easier to visually segment the vowel offset in the spectrogram than if the following word started with a vowel. In one case, [geɐ] "*went*" in the CP context proved difficult when inserting boundaries for the diphthong such that the target sounds were in open syllables (CV), since the following word started with a vowel /e/. In this case, along with the clear formant structure of F1 and F2, the waveforms were used to insert the boundaries, especially at the end of the vowel, when the waveform started to get less complex, just before the beginning of the following vowel.

In addition, in order to measure the formant transition duration, another Praat *tier* named “Transition” was used in order to manually insert interval boundaries around the diphthong transition period. The start point of the interval was inserted where the second formant started to change from its steady state and the end point of the interval was inserted when the second formant started to appear in a steady state (cf. Lindau, Norlin, and Svantesson, 1990). These criteria are shown in Figure 3. 1 and Figure 3. 2.

The number of analysed tokens was 1307. As discussed in Section 2.4.2, data from two males and two female speakers were not included in the final analysis. The total number of tokens analysed per vowel, by males (M) and females (F) and in Carrier Phrases (CP) and Full Sentence (FS) are given in Table 3. 3.

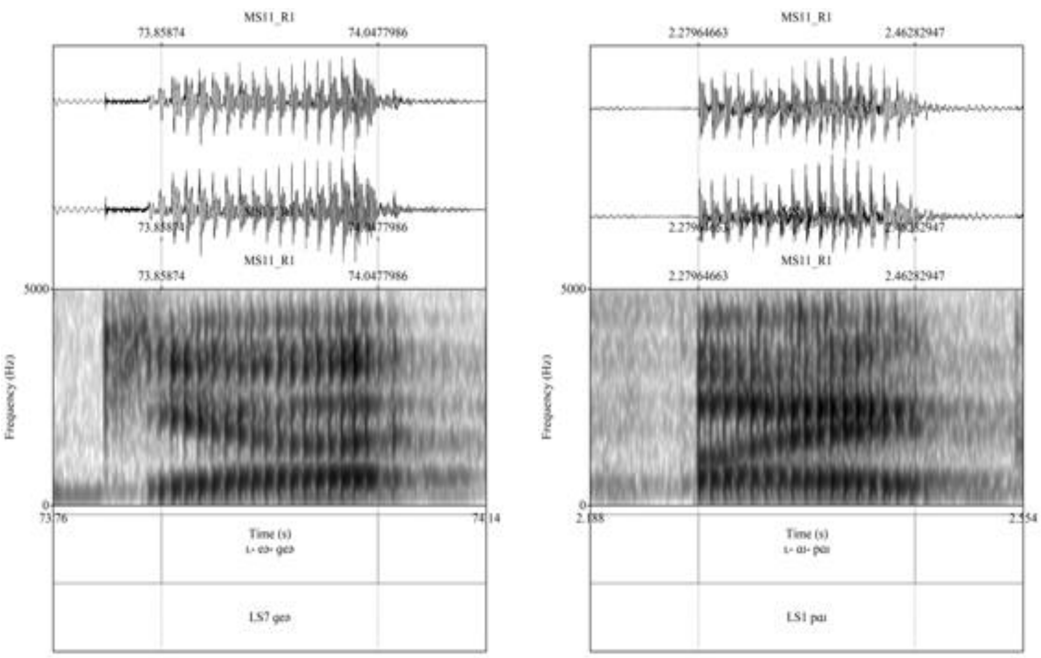



Figure 3. 1: word level segmentation for diphthong in [geɐ] (+voiced V, open syllable – the word following the target does not begin with /k/) [left] and for diphthong in [pɑe] (-voiced V, open syllable - the word following the target begins with /k/) [right]

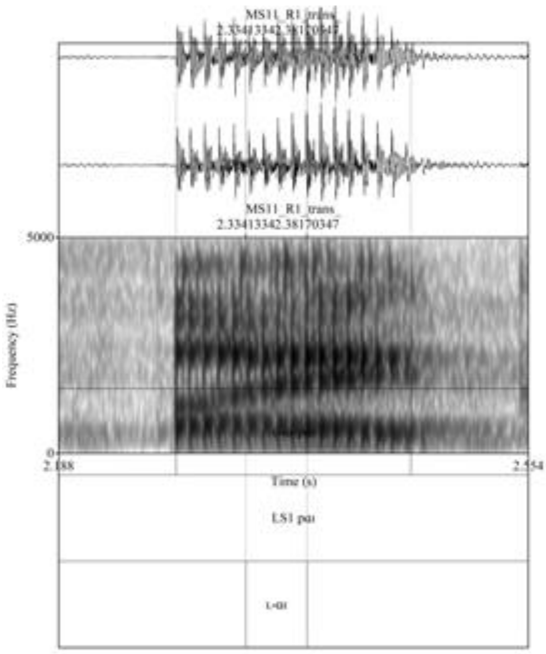



Figure 3. 2: word level segmentation for transition interval for diphthong in [pɑe]

Table 3. 3: total number of tokens analysed for each vowel, per gender (M and F) and per context (CP and FS).

| | | Speakers | | Context | |
|---|---|---|---|---|---|
| Vowel | Tokens | F | M | CP | FS |
| ɑe | 220 | 110 | 110 | 111 | 109 |
| ɑʊ | 218 | 109 | 109 | 108 | 110 |
| oe | 212 | 108 | 104 | 103 | 109 |
| ɪɐ | 220 | 110 | 110 | 109 | 111 |
| eɐ | 217 | 109 | 108 | 107 | 110 |
| ʊɑ | 220 | 110 | 110 | 110 | 110 |
| **Total** | **1307** | **656** | **651** | **648** | **659** |

### 3.2.4 Automatic Formant Extraction

Praat scripts (see Appendix 2E) were used to extract the frequencies of the first, second and third formant of monophthongs in two temporal positions, 20% and 80% (cf. Williams and Escudero, 2014; Hillenbrand, 2003), and the duration in milliseconds. Following measurements by Mayr and Davies (2011), Kirtley et al. (2016) and Williams and Escudero (2014) of diphthong trajectories, F1 and F2 movement, and *vowel inherent spectral change* (VISC), the formant frequencies were additionally measured at seven equidistant points for each formant, i.e. 20%, 30%, …, 80%.

As discussed in Section 2.1.2, F1 and F2 in the vowel steady state (usually midpoint) is not sufficient to investigate the acoustic properties of diphthongs, because the vowel quality changes resulting in a decrease or increase in F1 value, depending on whether the first segment is open (e.g. /ɑ/) or closed (e.g. /ɪ/). Therefore, the *rate of change* (ROC) approach, as employed by Gay (1968), Deterding (2000) and Kent and Read (1992 cited in Deterding, 2000), was also used to measure the change in the quality and spectral change in diphthongs. Further, following Lindau et al. (1990), the transition duration was measured for each diphthong, as detailed above. In addition to formant frequencies, the total duration of diphthongs was measured.

In order to compare the first target and the second target in the diphthongs with their monophthongal counterparts, the formant frequencies of the first two formants were extracted at the midpoint of the first and second target and were analysed acoustically and statistically. In addition, in order to compare the duration of the first and second target of the diphthongs beyond a visual inspection of the spectrogram (cf. Mayr and Davies, 2009), the duration before and after the transition period was measured for acoustic and statistical analysis. These measurements aided in determining whether the second vowel in Urdu diphthongs is always long (Waqar and Waqar, 2002), or both vowels are equally long (Bhatti and Mumtaz, 2016), or it merely depends on each individual vowel. These measurements also help with the IPA transcriptions of Urdu diphthongs.

Previous studies have reported diphthong duration to compare cross-dialect differences (cf. for Welsh: Mayr and Davies, 2011; for American English: Jacewicz and Fox, 2013; for Southern and Northern dialect of British English: Williams and Escudero, 2014). In the present study diphthong duration was compared with monophthong duration in order to determine if the two vowels in the target words have a combined duration comparable to a single vowel (i.e. total duration will be equal to or less than the long monophthongs in Urdu, as reported by Khurshid et al., 2003) or two separate vowels (i.e. the total duration of the two vowels in the diphthong will be less than the sum of the two corresponding monophthongs).

Following Mayr and Davies (2011) and Fox and Jacewicz (2009), the vowel section length (VSL) was calculated. In the present study six sections were calculated as opposed to the four sections calculated in previous studies to provide sufficient resolution for subsequent visual comparison with manual segmentation (see Section 3.2.3). That is, we calculate VSL for sections 20%-30%, 30%-40%, 40%-50%, 50%-60%, 60%-70%, and 70%-80% across each diphthong duration with the following Euclidean distance formula:

$$VSL_n = \sqrt{(F1_n - F1_{n+1})^2 + (F2_n - F2_{n+1})^2} \qquad \textbf{(1)}$$

where $VSL_n$ is the section length with section number $n$ (i.e. $n$=1 for 20%-30%, $n$=2 for 30%-40%, …, $n$=6 for 70%-80%) and $F1_n$/$F2_n$ are the format values at sample number $n$ (i.e. $n$=1 for 20%, $n$=2 for 30%, …, $n$=7 for 80%).

The trajectory length (TL) was then calculated for each diphthong.

$$TL = \sum_{n=1}^{6} VSL_n \qquad \textbf{(2)}$$

Trajectory length (TL) can be defined as the length of the diphthong's path through F1/F2 vowel space.

The overall rate-of-change of this trajectory is then the trajectory length divided by the portion of the overall duration that the trajectory covers (i.e. 60% of the duration)

$$TL_{roc} = \frac{TL}{0.60 \times V_{dur}} \quad \textbf{(3)}$$

This gives the values of trajectory length rate of change in Hz per millisecond.

Vowel section length (VSL) rate of change was calculated separately for each section of each diphthong with the following formula:

$$VSL_{roc_n} = \frac{VSL_n}{0.15 \times V_{dur}} \quad \textbf{(4)}$$

This means that the VSL of each section of diphthong (in Hz) was divided by the duration (in ms) of that section; this gave the values of spectral rate of change in Hz per milliseconds.

### 3.3 Statistical Analysis

The models for statistical analysis are identical to those discussed in Section 2.4.

### 3.4 Results

The results are discussed with reference to mean F1, F2 at onset and offset points, transition duration of F2 (Lindau et al., 1990), and TLroc and VSLroc (Mayr and Davies, 2011). In addition, the mean duration of first and second component in the diphthongs is presented.

In order to validate to the manual segmentation of F2 transition duration (Lindau et al., 1990) as discussed in Section 3.2.2, the spectral rate of change (Mayr and Davies, 2011) was also calculated and plotted. By comparing the peaks in spectral rate of change with the F2 transition segmentation, it can be seen that manual segmentation is not subject to author's subjective bias.

### 3.4.1 Formant Frequencies

The mean formant frequencies in Hz of six diphthongs with standard deviation in Hz are given in Table 3.4 and Lobanov normalised are given in Table 3.5. With regard to spectral change and direction of trajectories, as shown in Figure 3.3, all six diphthongs appear to be distinct. The mean and standard deviation of frequencies of the first two formants in Hertz and Lobanov normalised at seven equidistant points, i.e. 20%, 30%, …, 80% are given in Appendix 3A.

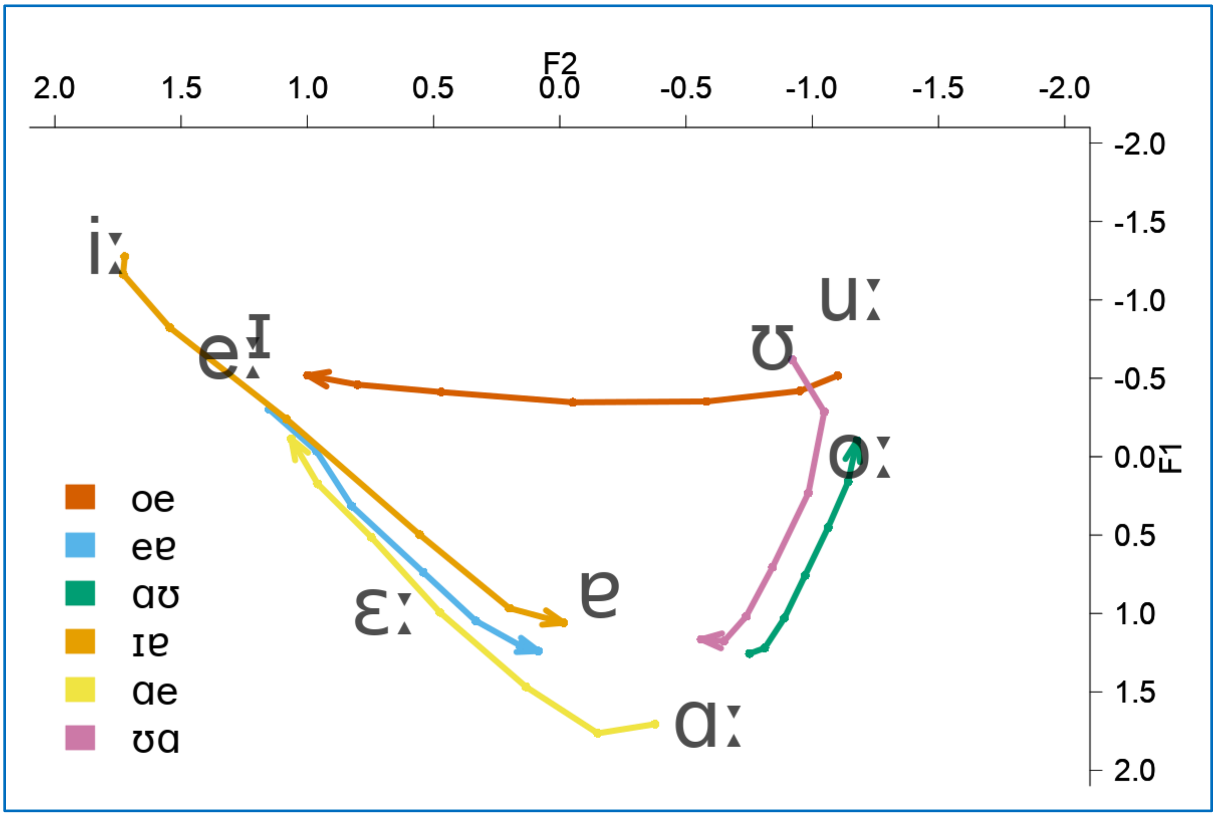



Figure 3. 3: Trajectories of mean formant frequencies of six diphthongs overlaid on monophthongs (steady state), Lobanov normalised (z-score). The dots on each trajectory show the formant measurements at seven equidistant points in time for each formant, i.e. 20%, 30%, …, 80%, and arrowheads show the direction and offset of each diphthong.

Table 3. 4: Formant frequencies of F1 and F2 (in Hz) at onset and offset of diphthongs produced by 11 male and 11 female speakers, pooled over contexts (standard deviation in parenthesis).

| Vowel | Gender | F1 20% | F1 80% | F2 20% | F2 80% |
|---|---|---|---|---|---|
| ɑe | F | 778 (125) | 481 (86) | 1440 (120) | 2292 (280) |
| | M | 630 (57) | 470 (44) | 1221 (167) | 1865 (255) |
| ɑʊ | F | 695 (107) | 488 (68) | 1226 (108) | 957 (184) |
| | M | 596 (48) | 463 (46) | 1049 (108) | 870 (125) |
| eɐ | F | 476 (75) | 715 (151) | 2276 (242) | 1607 (139) |
| | M | 439 (99) | 583 (73) | 1956 (216) | 1510 (186) |
| ɪɐ | F | 364 (39) | 699 (129) | 2608 (346) | 1581 (134) |
| | M | 317 (32) | 557 (64) | 2209 (204) | 1438 (155) |
| oe | F | 435 (55) | 433 (64) | 1002 (150) | 2248 (235) |
| | M | 421 (32) | 422 (30) | 906 (179) | 1836 (257) |
| ʊɑ | F | 417 (55) | 684 (133) | 1133 (199) | 1329 (118) |
| | M | 413 (35) | 590 (49) | 962 (136) | 1144 (147) |

Table 3. 5: Formant frequencies of F1 and F2 (Lobanov normalised) at onset and offset of diphthongs, pooled over speakers and contexts (standard deviation in parenthesis).

| Vowel | F1 20% | F1 80% | F2 20% | F2 80% |
|---|---|---|---|---|
| ɑe | 1.71 (0.6) | -0.11 (0.5) | -0.38 (0.3) | 1.06 (0.5) |
| ɑʊ | 1.26 (0.5) | -0.10 (0.4) | -0.75 (0.2) | -1.18 (0.3) |
| eɐ | -0.30 (0.7) | 1.24 (0.8) | 1.15 (0.4) | 0.09 (0.4) |
| ɪɐ | -1.28 (0.3) | 1.06 (0.6) | 1.72 (0.5) | -0.02 (0.3) |
| oe | -0.52 (0.3) | -0.52 (0.3) | -1.10 (0.3) | 1.00 (0.4) |
| ʊɑ | -0.62 (0.3) | 1.17 (0.6) | -0.92 (0.3) | -0.56 (0.2) |

Statistical tests with F1 and F2 as dependent variables were performed using the same methodology as in Section 2.4 using diphthong onset and offset, plus monophthongs, as input data.

For F1, the results revealed a significant main effect of Vowel ($F(23,5945.94) = 1038.90$, $p < 0.0001$), Context ($F(1,5943.59) = 101.32$, $p < 0.0001$), and significant interaction between Context × Vowel ( $F(23,5943.59) = 14.04$, $p < 0.0001$) and Gender × Vowel ($F(23,5945.94) = 8.13$, $p < 0.0001$) and Context × Gender ($F(1,5943.52) = 16.94$, $p < 0.0001$). The random effect of speaker was also significant ($\chi^2(1) = 43.88$, $p < 0.0001$). The results further showed non-significant effect of Gender ($F(1,23.68) = 1.32$, $p = 0.26$), and non-significant three-way interaction between context × gender × vowel ($F(23,5943.53) = 1.40$, $p = 0.09$).

For F2, the results showed a significant main effect of Vowel ($F(23,5945.82) = 1928.66$, $p < 0.0001$) and Context ($F(1,5943.80) = 17.81$, $p < 0.0001$), and significant interactions between Context × Vowel ($F(23,5943.66) = 7.48$, $p < 0.0001$) and Gender × Vowel ($F(23,5945.82) = 4.62$, $p < 0.0001$), and significant three-way interaction between Context × Gender × Vowel ($F(23,5943.66) = 2.09$, $p = 0.001$). The random effect of speaker was also significant ($\chi^2(1) = 15.58$, $p < 0.0001$). The results further showed a non-significant effect of Gender ($F(1,24.70) = 6.18$, $p = 0.02$) and non-significant interaction effects between Context × Gender ($F(1,5943.80) = 0.03$, $p = 0.85$).

In summary:

**For F1**, the combined interaction between gender, context and vowel was found to be non-significant. In terms of an R formula we have

*Context + Vowel + Gender + Context:Vowel + Context:Gender + Vowel:Gender + (1 | Speaker)*

**For F2** the combined interaction between gender, context and vowel was found to be non-significant, as was any interaction between gender and context. In terms of an R formula we have

*Context + Vowel + Gender + Context:Vowel + Vowel:Gender + (1 | Speaker)*

These F1 and F2 formulae were used to construct the model for post-hoc pairwise comparisons as discussed below.

The pairwise comparisons showed that all six diphthongs were distinct from each other at the onset and offset with regard to F1 and F2, and the nested vowels within each diphthong are distinct from the respective monophthongs.

These results suggest that the vowel at the end of [ɪɐ], [eɐ], and [ʊɑ] are closer to /ɑː/ than /ɐ/, but not the same; the final sound in [ɑʊ] is closer to /oː/ than /ʊ/ but not the same; and the first sounds in [oe] and [ʊɑ] are quite close to the monophthongs /oː/ and /ʊ/ respectively with regard to F1 and F2. However, the first sound in [ɪɐ] is different from /ɪ/ with reference to F1 ($p < 0.001$) and suggests that /ɪ/ in [ɪɐ] is higher and more fronted than the monophthong /ɪ/. The first sound in [eɐ] is different from /e/ with regard to F1 ($p < 0.001$); and the first sound in [ɑʊ] and [ɑe] is very close to /ɑː/ with regard to F1, but different with regard to F2. These results are in line with the description of Welsh diphthongs (Ball, 1983) and English diphthongs by Ladefoged and Johnson (2011:92) "…[t]he diphthong often do not begin and end with any of the sounds that occur in simple vowels". Comparisons of diphthong onset and offset are shown in Figure 3. 4.

The steady decrease in F1 in [ɑe], [ɑʊ] and [oe], and steady increase in F2 in [ɑe] and [oe] can be seen in the F1 and F2 movement plot in Figure 3. 4 (top row). A steady increase in F1 in [ʊɑ], [eɐ] and [ɪɐ] and a sharp decrease in F2 in [eɐ] and [ɪɐ] can be seen in the Figure 3. 4 (bottom row), which suggests that these vowels start from front close and close-mid region and are centring towards central open-mid region. The decrease in F2 in [ʊɑ] is not as sharp as compared to [eɐ] and [ɪɐ], and unlike these two diphthongs [ʊɑ] starts with a close-mid back vowel, which already has lower F2; hence movement towards the centre does not show a sharp change in F2.

With regard to trajectory direction and movements, Urdu diphthongs can now be divided into two groups: closing and centering (cf. Harrington, Cox, and Evans, 1997), which is discussed in the following subsection.

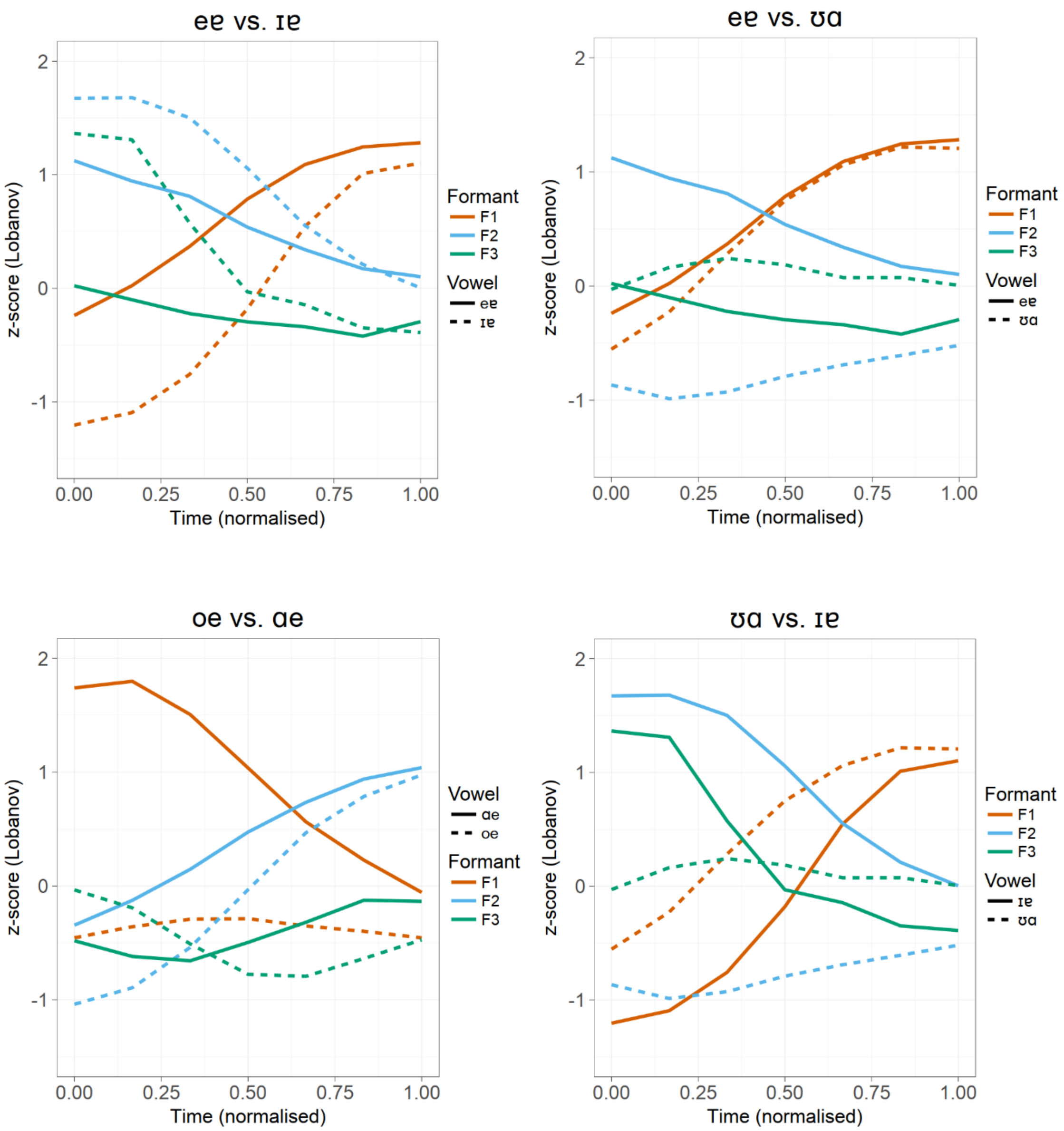



Figure 3. 4: : Lobanov normalised (z-score) F1, F2 and F3 trajectories of [ɪɐ] (dashed line) and [eɐ] (solid line) top left and [ʊɑ] (dashed line) and [eɐ] (solid line) top right; [oe] (dashed line) and [ɑe] (solid line) bottom left and [ʊɑ] (dashed line) and [ɪɐ] (solid line) bottom right pooled over male and female speakers and two contexts (CP and FS).

#### 3.4.1.1 Closing Diphthongs

Three diphthongs with offset points in the close front and close back region are labelled as closing diphthongs, i.e. [ɑe], [oe] and [ɑʊ]. The trajectory length, direction and movements of these diphthongs are discussed below.

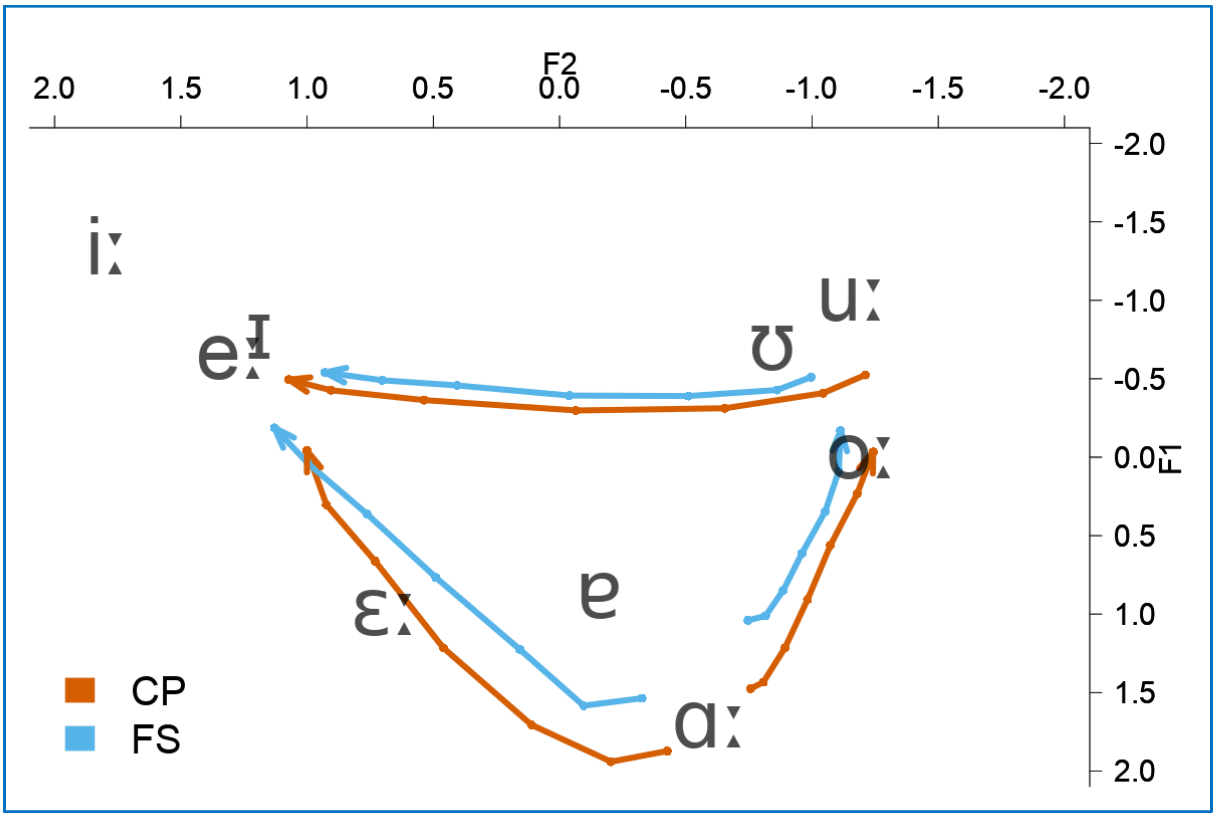



Figure 3. 5: Trajectories of mean formant frequencies of [ɑe], [ɑʊ] and [oe] diphthongs overlaid on monophthongs (steady state), Lobanov normalised (z-score), for each context. The dots on each trajectory show the formant measurements at seven equidistant points in time for each formant, i.e. 20%, 30%, …, 80%, and arrowheads show the direction and offset of the diphthong.

[ɑe] and [ɑʊ] have higher F1 values at onset (20%) and lower F1 values at offset (80%) which suggest that these diphthongs start in open or open-mid region and end in close region. However, F2 at the onset (20%) of [ɑe] is higher than the onset of [ɑʊ]. A lower F2 at the offset (80%) suggests that the second vowel in [ɑʊ] is further back and retracted as compared to the second vowel in [ɑe], which has a higher F2 value (2078 Hz) and so appears to be in the close-mid front region.

The diphthong in [boe] has similar values of F1 at onset (423 Hz) and offset (434 Hz); however, F2 at onset is lower than F2 at offset, which suggests that this vowel starts somewhere in the back close-mid region and ends in the front close region. F2 for [ɑe] starts with lower F2 (1288 Hz) and ends with higher F2 (2078 Hz); similarly, F2 for [oe] starts with lower F2 (931 Hz) and ends with higher F2 (2059 Hz). The higher F2 for [ɑe] at the onset suggests that it is in the open-mid or open region, and the lower F2 for [oe] at the onset suggests that this vowel is in the close or close-mid region. The F2 at offsets of these two diphthongs are similar, which suggests that these diphthongs share an offset point and end in the front close-mid region.

F2 at the onset of [ɑʊ] is higher (1139 Hz) than the F2 at offset (918 Hz), which suggests that this vowel starts somewhat in the centre and ends as a back vowel. This movement can be seen in the vowel trajectory in Figure 3. 5.

Post-hoc pairwise comparisons showed that the onset and offset points of diphthongs are significantly different from each other and also from individual monophthongs. For instance, /ɑ/ in the onset of [ɑe] and [ɑʊ] is significantly different with regard to F1 ($t(6014.83) = 7.235$, $p < 0.0001$), and F2 ($t(6038.43) = 8.336$, $p < 0.0001$). /ɑ/ in the onset of [ɑe] and [ɑʊ] is also significantly different from the monophthongs /ɐ/ and /ɑː/, with regard to both F1 and F2 as shown in Table 3. 6.

Table 3. 6: Pairwise comparison of diphthongs [ɑe] and [ɑʊ] onset with monophthongs /ɐ/ and /ɑː/

| Diphthongs vs monophthong | F1 | F2 |
|---|---|---|
| ɑe vs /ɐ/ | t (6016.87) = 14.805, p < 0.0001 | t (6041.85) = -8.931, p < 0.0001 |
| ɑe vs /ɑ/ | t (6014.90) = 3.588, p = 0.05 | t (6038.54) = -3.896, p = 0.02 |
| ɑʊ vs /ɐ/ | t (6016.820 = 7.345, p < 0.0001 | t (6041.76) = -17.461, p < 0.0001 |
| ɑʊ vs /ɑ/ | t (6014.81) = 10.790, p < 0.0001 | t (6038.39) =4.419, p < 0.01 |

The offset in [ɑe] was significantly different from monophthongs /eː/ with regard to both F1 (t(6014.83) = 9.667, p < 0.0001), and F2 (t(6038.43) = -6.742 , p < 0.0001), monophthong /iː/ with regard to F1(t(6016.84) = 24.338 , p < 0.0001), and F2 (t(6041.76) = -20.220, p < 0.0001); however it was distinct from monophthong /ɪ/ with regard to F1 (t(6018.52) = 14.368, p < 0.0001) only. The higher F1 from /ɪ/ in the offset position of [ɑe] suggests that the offset in the vowel is lower than /ɪ/ but in the same front region.

The offset in [ɑʊ] is significantly different from /uː/ with regard to F1 (t(6016.22) =23.020, p <0.0001), but not F2. The higher F1 for the offset in [ɑʊ] suggests that it was significantly lower than monophthong /uː/. However the offset in [ɑʊ] is distinct from /oː/ with regard to F1 only (t(6022.61) =-4.711, p <0.01).

The onset in [oe] is significantly different from the monophthong /oː/ with regard to F1 (t(6023.46) =-12.543, p <0.0001) but not F2. The lower F1 in the onset of [oe] suggests that it is significantly lower than /o/; however, the non-significant difference in F2 suggests that it is in the same back region and as retracted as /oː/.

The offset in [oe] is significantly different from the monophthong /iː/ with regard to both F1 (t(6017.31) = -17.906 , p <0.0001) and F2 (t(6042.54) = 21.227 , p <0.0001), and also from monophthong /ɪ/ with regard to both F1 (t(6019.76) = -6.883, p < 0.0001), and F2 (t(6046.33) = 5.497, p < 0.01). The higher F1 in the offset of [oe] suggests that it was significantly lower than monophthongs /iː/ and /ɪ/. The lower F2 suggests that it was retracted and was not in the same front region as /iː/ and /ɪ/; however, the offset is significantly different from monophthong /eː/ only with regard to F2 (t(6039.77) = 7.924, p < 0.0001). The higher F2 than /eː/ suggests that this vowel was more front than /eː/.

### 3.4.1.2 Centering Diphthongs

[eɐ], [ɪɐ], and [ʊɑ] have lower F1 at the onset and higher F1 at the offset. F2 at the offset of [eɐ], [ɪɐ] and [ʊɑ] varies. As can be seen from the trajectories in Figure 3. 6, [eɐ] seems to end further back and higher than [ɪɐ], and [ɪɐ] appears to be more front than [ʊɑ] and lower than /ɐ/. These offsets suggest that diphthongs end in the same region with substantial variations in F1 and F2. Further [ɪɐ] starts in the close front region and ends in the open-mid central region; however, [eɐ] starts in the close-mid region and ends further back in the open-mid region. [ʊɑ], which starts in the close-mid back region near /ʊ/, ends further back and retracted than [ɪɐ] and [eɐ], between /ɐ/ and /ɑː/. The exact offset point of [ʊɑ] appears to be lower and further retracted than /ɐ/ but higher than /ɑː/.

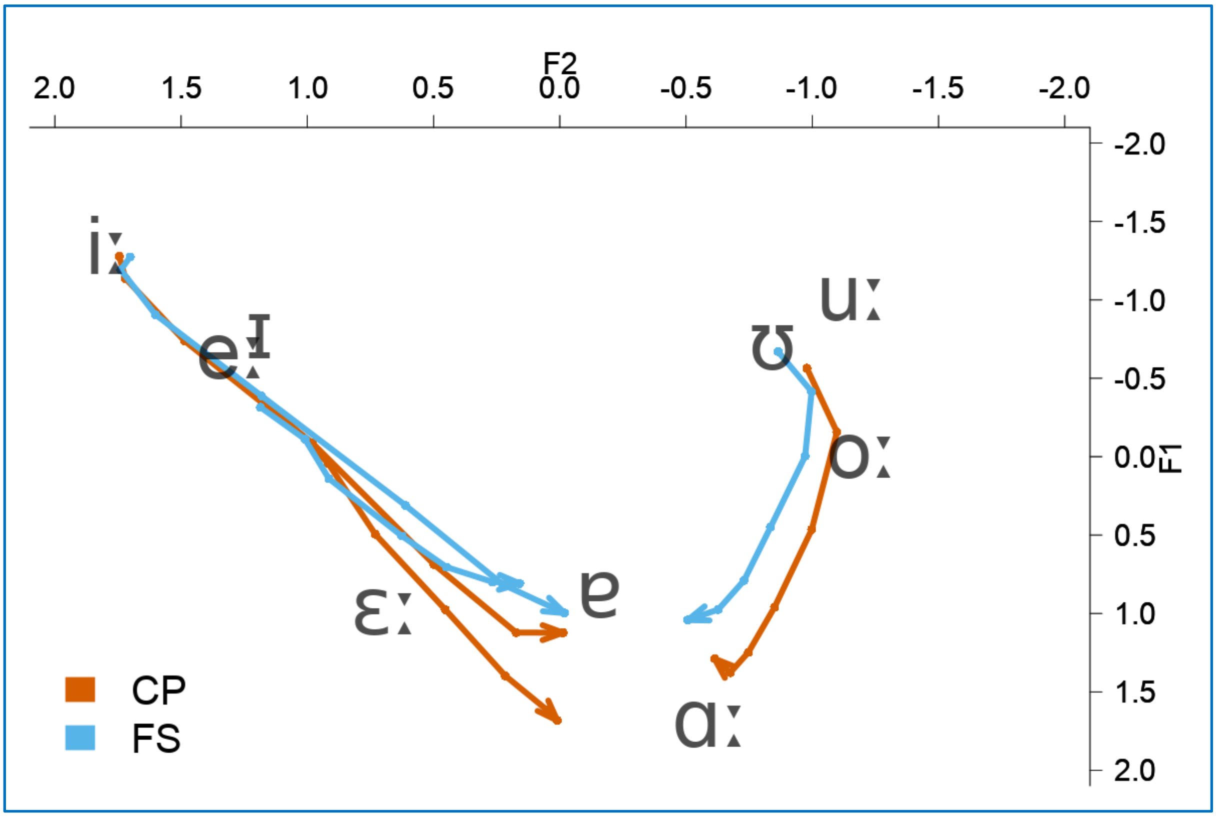



Figure 3. 6: Trajectories of mean formant frequencies of [ɪɐ], [eɐ] and [ʊɑ] diphthongs overlaid on monophthongs (steady state), Lobanov normalised (z-score), for each context. The dots on each trajectory show the formant measurements at seven equidistant points for each formant, i.e. 20%, 30%, …, 80%, and arrow heads show the direction and offset of the diphthong.

The pairwise comparison tests showed that these points are significantly different. For instance offset in [eɐ] and [ɪɐ] is significantly different from each other with regard to F1 ($t(6014.90) = 5.106$, $p < 0.0001$) and F2 ($t(6038.54) = 4.187$, $p < 0.01$). The results also showed that the offset of [eɐ] and [ʊɑ] was different from the central vowel /ɐ/, that is the offset in [eɐ] was significantly different from /ɐ/ with regard to both F1(t(6016.91) = 8.138, $p < 0.0001$) and F2 (t(6041.90) = 6.891, $p < 0.0001$); and the offset in [ʊɑ] was significantly different from /ɐ/ with regard to both F1(t(6016.87) = 5.955, $p < 0.0001$) and F2 (t(6041.84) = -11.637, $p < 0.0001$). However, the offset in [ɪɐ] was not significantly different from /ɐ/ with regard to either F1 (t(6016.87) = 2.910, p =0.35) and F2 (t(6041.83) = 2.605, p =0.59).

The results also showed that the onset in [eɐ] was significantly different from /ɪ/ with regard to F1 (t(6018.51) = 9.276, $p < 0.0001$) but not F2; however onset in [ɪɐ] was significantly different from monophthong /ɪ/ with regard to both F1 (t(6018.52) = 12.584, $p < 0.0001$) and F2 (t(6044.37) = -17.910, $p < 0.0001$). Onset in [ɪɐ] was not significantly different from monophthong /iː/ with regard to F1 and F2, which suggests that the onset in [ɪɐ] was in the same close front region as /iː/. The onset in [ʊɑ] was not significantly different from the monophthong /ʊ/.

### 3.4.1.3 Trajectory Length Rate of Change (TL roc)

The trajectory length rate of change (TLroc) is shown in Figure 3. 7. Statistical tests were once again performed using the same methodology as Section 2.4.4, with TLroc as the dependent variable and using diphthong input data augmented with the TLroc for each utterance.

The results revealed a significant main effect of Vowel (F(5,1285.17) = 135.09, $p < 0.0001$) and Gender (F(1,22.04) = 8.37, $p < 0.01$). In addition, the results showed a significant interaction between Context × Vowel (F(5,1285.19) = 5.93, $p < 0.0001$) and Gender × Vowel (F(5,1285.17) = 2.98, p = 0.01). The random effect of Speaker was also significant ($\chi^2(1) = 119.60$, $p < 0.0001$). However, Context × Gender interaction (F(1,1285.25) = 2.34, p = 0.12); and a three-way Context × Gender × Vowel interaction (F(5,1285.19) = 0.50, p = 0.77) were not significant. In summary:

**For TLroc**, the combined interaction between gender, context and vowel was found to be non-significant, as was any interaction between gender and context. In terms of an R formula we have

*Context + Vowel + Gender + Context:Vowel + (1 | Speaker)*

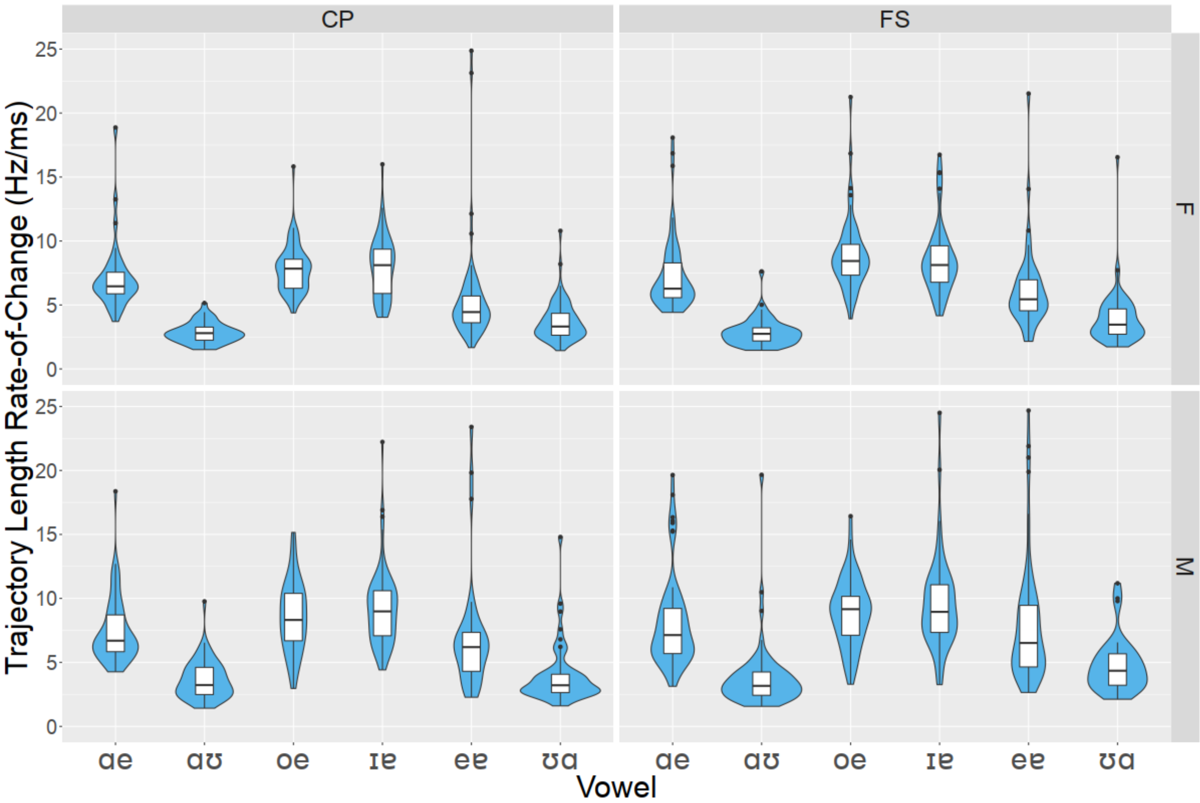



Figure 3. 7: Violin plots of mean trajectory length rate of change (in Hz/ms) of the six diphthongs in two contexts (CP and FS) by 11 male and 11 female speakers.

Pairwise comparisons showed non-significant differences in spectral rate of change for the following diphthongs; [ɑe] did not differ significantly in $TL_{roc}$ from [eɐ] ($t(1285.09) = 3.084$, $p = 0.02$). Also, [ɑe] did not differ significantly in $TL_{roc}$ from [oe] ($t(1285.41) = -3.136$, $p = 0.02$). [ɑʊ] was not significantly different from [ʊɑ] ($t(1285.06) = -1.812$, $p = 0.45$), and [ɪɐ] and [oe] also did not show significant difference in $TL_{roc}$ ($t(1285.41) = 1.681$, $p = 0.54$). $TL_{roc}$ is slower for [ʊɑ] and [ɑʊ], across both contexts and genders, than all other diphthongs. [eɐ] shows the fastest rate of change across both genders and contexts, except for male CP context. Diphthongs show consistent spectral rate of change by male and female speakers across both contexts; however, the results are inconclusive with regard to the differences in rate of change between each diphthong.

#### 3.4.1.4 Vowel Section Length Rate of Change (VSLroc)

VSLroc for each diphthong in each context by each gender is given in Figure 3. 8. The figure shows roughly similar patterns of spectral change for each diphthong across all contexts and genders. However, some diphthongs show differences in the peaks of spectral change across six sections. For instance, [ɑʊ] does not show extensive spectral change for F1 and F2 as compared to other diphthongs. [ɑe], [eɑ], and [oe] show constant increase in spectral change. However, the peaks, where this constant increase in spectral change culminates, are different for each of these diphthongs.

In CP context by female speakers, as shown in Figure 3. 8, [eɐ] shows the spectral peak in the 30%-40% section of the vowel; [ɑe], [ɑʊ] and [oe] show spectral peaks in the 50%-60% section of the vowel; and [ɪɐ] and [ʊɑ] show spectral peaks in the 40%-50% section of the vowel.

In diphthongs produced by female speakers in FS context, [eɐ], [ɑe] and [oe] shows spectral peaks in the 50%-60% section of the vowel. [ɑʊ], [ɪɐ] and [ʊɑ] show spectral peaks in the 40%-50% section of the vowel. In the diphthongs produced by male speakers in CP context, [oe], [eɐ], [ɪɐ] and [ʊɑ] show spectral peaks in the 40%-50% section of the vowel; [ɑʊ] shows a spectral peak in the 50%-60% section of the vowel; and [oe] shows spectral peak in 30%-40% section of the vowel. In the diphthongs produced by male speakers in FS context all diphthongs except [ɪɐ] show a spectral peak in the 40%-50% section of the vowel, and [ɪɐ] shows a spectral peak in the 50%-60% section of the vowel.

Statistical tests were once again performed using the same methodology as Section 2.4.4, with VSLroc as the dependent variable and using diphthong input data augmented with the VSLroc for each utterance.

The results revealed a significant main effect of Vowel ($F((5,7820.54) = 220.46$, $p < 0.0001$) and Section ($F(5,7820.02) = 79.89$, $p < 0.0001$). In addition, the results showed a significant interaction between Context × Vowel ($F(5,7820.60) = 9.68$, $p < 0.0001$) and Gender × Vowel ($F(5,7820.54) = 4.81$, $p = 0.001$), Context× Section ($F(55,7820.02) = 4.62$, $p = 0.0001$), Vowel× Section ($F(25,7820.02) = 18.25$, $p = 0.0001$), and a significant three-way interaction between Context ×Vowel × Section ($F(25,7820.02) = 2.57$, $p = 0.0001$), and Gender ×Vowel × Section ($F(25,7820.02) = 2.26$, $p = 0.001$).

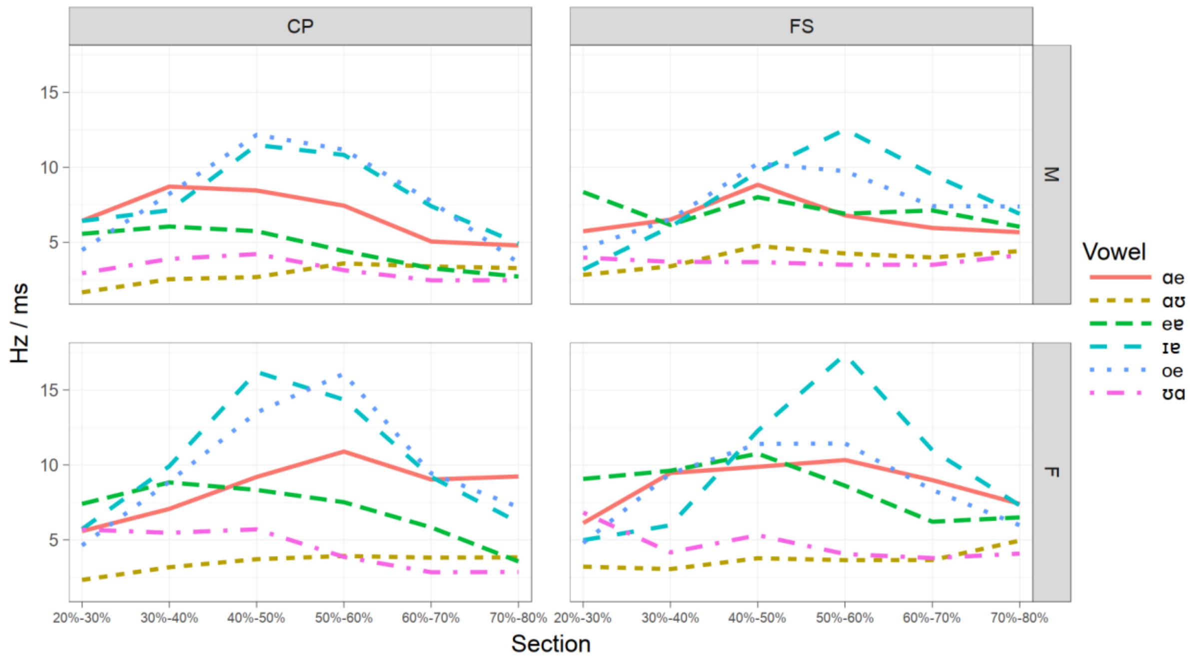



Figure 3. 8: Vowel section length rate of change for each diphthong in both contexts (CP, FS) by male and female speakers

The random effect of Speaker was also significant ($\chi^2(1) = 236.42$, $p < 0.0001$). However, Context × Gender interaction ($F(1,7820.82) = 3.76$, $p = 0.05$); and a three-way interaction between Context × Gender × Vowel ($F(5,7820.62) = 0.81$, $p = 0.54$) and Context × Gender

× Section (F(5,7820.01) = 1.61, p = 0.15) were not significant. In addition a four-way interaction between Context × Gender × Vowel × Section (F(25,7820.01) = 1.33, p = 0.12) was also not significant. In summary:

**For VSLroc**, the combined interaction between gender, context and vowel-section was found to be non-significant, as was any interaction between gender and context. In terms of an R formula we have

*Context + Gender + Vowel + Section + (1 | Speaker) + Context:Vowel + Gender:Vowel + Context:Section + Gender:Section + Vowel:Section + Context:Vowel:Section + Gender:Vowel:Section*

Pairwise comparisons showed significant differences in VSLroc for all diphthongs; however [ɪɐ] did not differ significantly in the pattern of spectral change from [oe] (t(7821.44) = 2.145, p = 0.26). [ɑʊ] also did not show a significant difference from [ʊɑ] in the pattern of spectral change (t(7820.11) = -2.313, p = 0.18). Thus, results for [ɪɐ], [oe] and [ɑʊ] are inconclusive and more data is required for further investigation. The other three diphthongs ([ɑe], [ʊɑ] and [eɐ]) showed significant differences in the patterns of spectral rate of change for each section of the diphthong.

### 3.4.2 Duration

The mean duration of six diphthongs with standard deviation is given in Table 3. 7. Figure 3.9 shows the mean duration in all contexts by the 22 speakers. Figure 3. 9 shows that all diphthongs have comparable duration; however, [ʊɑ] and [ɑʊ] are shorter than any other diphthongs.

Statistical tests were once again performed using the same methodology as Section 2.4, with duration as the dependent variable and using diphthong input data.

Table 3. 7: Mean duration in milliseconds of the diphthongs produced by 11 male and 11 female speakers pooled over contexts (standard deviation in parenthesis).

| Vowel | Gender | Duration (SD) |
|---|---|---|
| ɑe | F | 233 (41) |
| | M | 208 (37) |
| ɑʊ | F | 234 (52) |
| | M | 212 (37) |
| eɐ | F | 245 (65) |
| | M | 219 (55) |
| ɪɐ | F | 235 (40) |
| | M | 223 (47) |
| oe | F | 243 (55) |
| | M | 227 (51) |
| ʊɑ | F | 244 (48) |
| | M | 232 (39) |

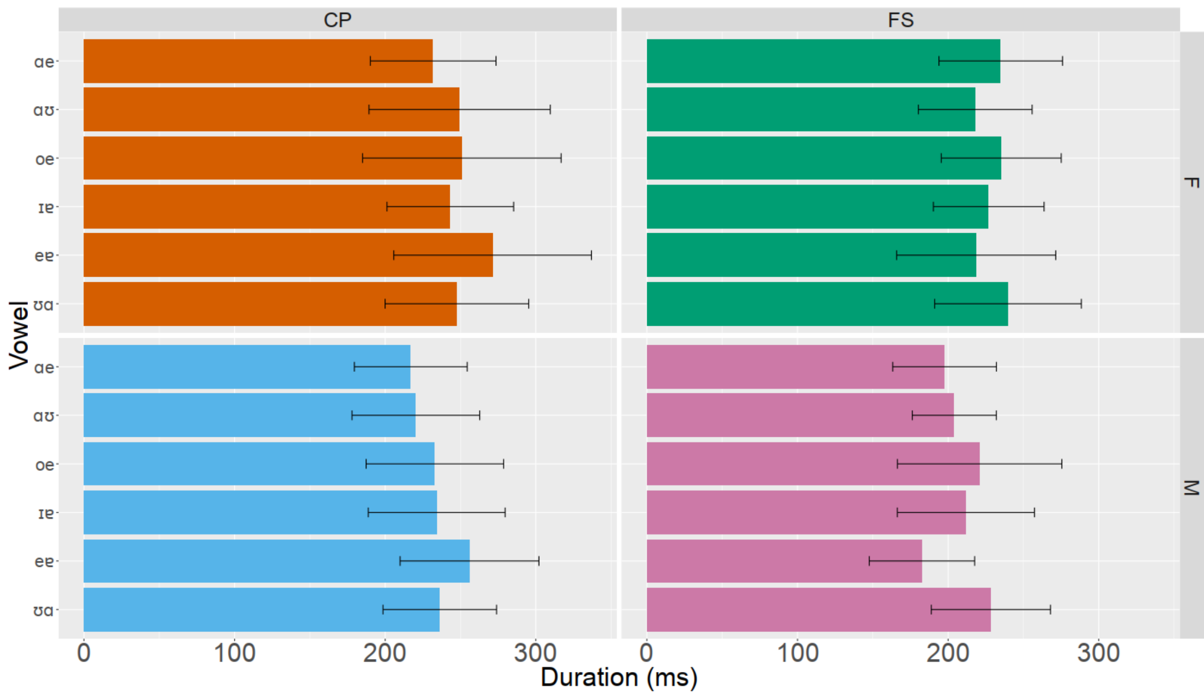



Figure 3. 9: Mean duration of the six Punjabi-Urdu diphthongs in each context (CP, FS) for each gender, across 22 speakers (11 males and 11 females). Error bars represent +/- SD.

The results revealed a significant main effect of Vowel ($F(5,1285.02) = 8.26$, $p < 0.0001$), a significant main effect of Context ($F(1,1285.04) = 134.23$, $p < 0.0001$), and a non-significant main effect of Gender ($F(1,22.01) = 2.34$, $p = 0.13$). In addition, the results showed a significant interaction between Context × Vowel ($F(5,1285.02) = 18.35$, $p < 0.0001$). The random effect of Speaker was also significant ($\chi^2(1) = 551.92$, $p < 0.0001$). However, Gender × Vowel interaction ($F(5,1285.15) = 1.90$, $p = 0.09$), Context × Gender interaction ($F(5,1285.02) = 18.35$, $p < 0.0001$), and a three-way Context × Gender × Vowel interaction ($F(5,1284.91) = 2.15$, $p = 0.05$) were not significant. Non-significant three-way interactions suggest that males and females responded to the changing context in the same manner.

The duration of diphthongs was found to be longer in CP than FS, which is backed up by the significant effect of context. Figure 3. 9 shows the mean duration of the six diphthongs.

The pairwise comparisons showed that in CP context the following diphthongs significantly differ in duration: [ɑe] is different than [eɐ] ($t(1296.12) =-8.131$, $p <0.01$), [oe] ($t(1296.25) =-3.847$, $p <0.01$), and [ʊɑ] ($t(1296.10) =-3.650$, $p <0.01$). [eɐ] is different than [ɪɐ] ($t(1296.13) =5.148$, $p <0.01$), [oe] ($t(1296.26) =4.160$, $p <0.01$), and [ʊɑ] ($t(1296.12) =4.497$, $p <0.01$).

The pairwise comparisons showed that in FS context the following diphthongs significantly differ in duration: [ɑʊ]/ is different than [ʊɑ] ($t(1296.10) =-4.758$, $p <0.01$). [eɐ] is shorter than [oe] ($t(1296.10) =-5.666$ , $p <0.01$), and [ʊɑ] ($t(1296.10) =-6.914$, $p <0.01$).

#### 3.4.2.1 F2 Transition Duration

In addition, F2 transition duration percentage was calculated from the ratio between the transition duration and total duration of each diphthong:

Transition duration percentage (TD%) = $\frac{Transition(end) - Transition(start)}{duration\ total} \times 100$

In order to calculate the transition duration, an interval was inserted in Praat, such that the transition began where the steady state of the first vowel started to change and ended where the steady state of the second vowel began.

The mean duration of the onset, transition and offset and transition duration percentages are given in Table 3. 8. These percentages suggest that F2 transition duration occupies almost 30% of the total duration of each diphthong. The results show that F2 transition duration covers 33% for [ɑe] and 31% for [ɑʊ], which is shorter than the F2 transition duration of [ɑe] (60%) and [ɑʊ] (73%) diphthongs in English (cf. Lindau, Norlin, and Svantesson, 1990).

As shown in Table 3. 8, the duration of the first component tends to be slightly shorter than the second component, and the transition duration tends to be shorter than the second vowel but comparable with the first vowel. Pairwise post-hoc Tukey test showed that the onset duration in [ʊɑ] was significantly different than offset ($t(2570.03) = 19.31$, $p < 0.0001$). For other diphthongs the difference was not significant. Therefore, contrary to Waqar and Waqar (2002), on-glide is not always shorter than off-glide for all six diphthongs.

Literature on Urdu vowels disagrees on the duration of first and second vowel in the diphthongs. The present study shows, however, that the difference in duration between first and second vowel is not significant.

Table 3. 8: Mean duration in milliseconds of first and second vowel and transition duration, and transition duration percentages for the six diphthongs across all speakers and both contexts (standard deviation in parenthesis)

| Vowel | Vowel 1 | Transition | Vowel 2 | Transition (%) |
|---|---|---|---|---|
| ɑe | 68 (15) | 72 (17) | 80 (22) | 33 |
| ɑʊ | 73 (16) | 68 (18) | 81 (23) | 31 |
| eɐ | 76 (24) | 70 (22) | 86 (31) | 30 |
| ɪɐ | 82 (18) | 67 (17) | 80 (19) | 29 |
| oe | 75 (19) | 79 (22) | 81 (27) | 33 |
| ʊɑ | 74 (17) | 73 (19) | 91 (23) | 31 |

### 3.4.2.2 Diphthong vs. Monophthong Duration Comparison

The mean duration of diphthongs versus the summed duration of the corresponding onset and offset monophthongs is given in Table 3. 9. The total duration of two separate monophthongs appears to be longer than the duration of diphthongs.

Statistical tests were once again performed using the same methodology as Section 2.4.4 with duration as the dependent variable, and diphthongs and summed monophthong pairs as input.

The formula as entered into R was thus

Duration ~ (Vowel/Mono.vs.diph) * Gender * Context + (1 | Speaker)

Stimulus Vowel (here Vowel refers to diphthongs as well as summed monophthong pairs), Gender and Context were fixed effects and Subject was a random factor. In contrast to the previous model, the new binary effect Mono.vs.diph was added to disambiguate the diphthongs and summed monophthong pairs.

As discussed in Section 2.4.4, the *lmerTest*'s function *step* was used to reduce the model. The results revealed a significant four-way interaction Vowel × Mono.vs.diph × Gender

× Context (F(6, 3682.0) = 4.3714, p <0.0001) meaning that the model could not be reduced.

Table 3. 9: Mean duration (ms) of diphthongs and sum of two monophthongs (onset+offset)

| Diphthong | Duration | Monophthong 1 | Duration | Monophthong 2 | Duration | Total |
|---|---|---|---|---|---|---|
| ɑe | **220** | ɑː | 190 | eː | 147 | **337** |
| ɑʊ | **223** | ɑː | 190 | oː | 174 | **365** |
| eɐ | **232** | eː | 147 | ɑː | 190 | **337** |
| ɪɐ | **229** | iː | 142 | ɑː | 190 | **332** |
| oe | **235** | oː | 174 | eː | 147 | **321** |
| ʊɑ | **238** | uː | 140 | ɑː | 190 | **330** |

The pairwise comparisons of each diphthong with the sum of two respective monophthongs showed that the total duration of each monophthong was significantly shorter than the total duration of two individual monophthongs. The output of the pairwise differences of contrast is given in Table 3.10.

Table 3. 10: Pairwise comparison of mean duration of diphthong with mean duration of two monophthongs.

| Diphthong vs. monophthongs duration | t-test |
|---|---|
| ɑe vs ɑː+eː | t(3728.58) = -30.295, p < 0.0001 |
| ɑʊ vs ɑː+oː | t(3728.58) = -39.485, p < 0.0001 |
| eɐ vs eː+ɑː | t(3728.58) = -26.587, p < 0.0001 |
| ɪɐ vs ɪː+ɑː | t(3728.58) = -25.085, p < 0.0001 |
| oe vs oː+eː | t(3728.58) = -21.075, p < 0.0001 |
| ʊɑ vs uː+ɑː | t(3728.58) = -43.708, p < 0.0001 |

The mean duration of diphthongs is below 250ms. These findings are partially in line with the results reported by Khurshid, Usman and Butt (2003). They reported that the duration of some diphthongs was below 150 ms, which is shorter than the duration of a long monophthong. The results from the present study show that the total duration of a diphthong is shorter than two monophthongs but longer than one long monophthong.

### 3.5 Discussion and Conclusion

Diphthongs are not distinct with regard to duration. The onset and offset points of diphthongs are distinct from the pure monophthongs. The steady decrease in F1 closing diphthongs and steady increase in F1 of centering diphthongs showed that Urdu diphthongs can be divided into closing and centering categories. The closing diphthongs have offset points in front close and back close regions; however, these points are distinct from the monophthongs in those regions. Centering diphthongs do not share the same offset point; as a result, [ɪɐ] and [eɐ] have offsets in the front central fairly open region, and the offset point of [ʊɑ] is further retracted and slightly more open. Trajectory length and vowel section length spectral rate of change showed the internal structure of each diphthong across both contexts and genders. The spectral peaks (as shown in Figure 3. 8) showed that the duration of the first vowel in the diphthong is not always shorter than the second vowel, since in some diphthongs the spectral peak was shown at 50% - 60% or 60%- 70% which suggests that the first component in the diphthong was longer than the second component. These findings contradict the claims of Waqar and Waqar (2002) who reported that the second element is longer.

Diphthong trajectories were shown graphically overlaid on the relevant monophthongs, and show a probable correlation. The mean F1 and F2 of these diphthongs before and after transition was extracted and compared with the mean F1 and F2 of the relevant monophthongs. Trajectory length ($TL_{roc}$) and vowel section length ($VSL_{roc}$) rate of change were also calculated and analysed statistically. The spectral rate of change showed that all diphthongs had spectral peaks in different sections of the vowels, which suggest that these diphthongs differ from each other with regard to transition behaviour.

Phonetically, however, these Urdu vocalic sequences do not behave like diphthongs. The results show that these diphthongs (vocalic-sequences) are shorter than two monophthongs together, but longer than a long monophthong, which suggests that a time slot is not deleted as claimed by Waqar and Waqar (2002). All six diphthongs are distinct from each other at the onset (20%) and offset with regard to F1 and F2, and first and second vocalic segments in these vowel-vowel sequences are also distinct from pure monophthongs in quality.

The time spent in the transition, and the duration at the onset and offset, are all of similar magnitude, each occupying approximately 1/3 of the total sequence, which suggests that Urdu does not have rising diphthongs as claimed by Waqar and Waqar (2002: 20): “Urdu has rising diphthongs (second vowel is of longer duration)”.

This result is markedly different from English where transition duration can be very high, e.g., 60% for [aɪ] and 73% for [aʊ] (Lindau et al., 1990). Lindau et al. (1990) also reported that in Chinese 40-50% of the diphthong duration is covered by transition, and in Arabic and Hausa for the same two vowels it is between 16-20%. They concluded that the transition duration for each diphthong is language specific and depends on the Euclidean distance between the first and second vowel in the diphthongs. The languages with larger vowel inventories (such as English) show larger transition duration as compared to languages with smaller vowel inventories (such as Arabic). Thus, according to Lindau et al. (1990) transition duration appears to be distinct across diphthongs as well as languages.

Overall, the results show that these diphthongs are pronounced as two vowels with almost equal duration by Punjabi-Urdu speakers. In addition, the shorter transition duration suggests that these are vowels in hiatus and not diphthongs (in line with (Quilis, 1981 cited in Aguilar, 1999:72, as discussed Section 3.1.2). The only possible reason to treat these vowel-vowel sequences as diphthongs could be that the total duration of each diphthong is less than the duration of two monophthongs. Literature on Urdu phonology unanimously reports that onsetless syllables, except for word initial position, are not allowed in Urdu. If we divide the words, e.g. [pɑe] and [boe] into two syllables such as /pɑ.e/ and /bo.e/, we are left with an onsetless syllable, which is not permitted in Urdu and violates the onsetless syllable structure rule (see Section 3.1.1). The stimuli used here were all considered monosyllabic words, therefore the effects of stress on the quality of these vowels cannot be accounted for. In future it will be interesting to see the qualities of these diphthongs, or vowels in hiatus, in disyllabic and multisyllabic words, with stressed and unstressed syllables. Future perception and syllable identification tests will also help to better establish the status of these diphthongs.

**Chapter 4**

## Second Language Perception

### Rationale

This chapter is focussed on the perception of SSBE vowels by Punjabi-Urdu speakers. The previous chapters investigated the acoustic and phonetic properties of Urdu vowels. This was necessary in order to formulate hypotheses and investigate the predictions of L2 perception models, such as Perceptual Assimilation Model (PAM; Best, 1995), PAM-L2 (Best and Tyler, 2007), Second Language Linguistic Perception model (L2LP) (L2LP; Escudero, 2005) and Speech Learning Model (SLM; Flege, 1995), as discussed in more detail in Section 4.2. Besides other factors, such as age of learning and exposure to L2, these models emphasize the relationship (e.g. similarities and differences) between the L1 and L2 phonetic and phonological inventories. Therefore, it was crucial to understand the Urdu vowel system in order to better understand the perception of English vowels by native speakers of Urdu.

As discussed briefly in Section 1.3, despite English being the official language in Pakistan, there is not a single study which has investigated the perception of Standard Southern British English (SSBE) vowels by these speakers. The context in Pakistan is very different from most of the studies reported in L2 perception literature for a number of reasons: Pakistan is a multilingual country where most of the population speaks more than two languages; English is a lingua franca in higher education and bureaucracy, and is even the language of the constitution, so it is not a foreign language; and the input learners receive is far from native except for some very prestigious English training institutions where BBC recordings are used for listening and speaking. Despite the multilingual context, Pakistanis strives to speak Standard Southern British English (SSBE). IELTS and other English language tests as a requirement for entry into national and international institutions are another reason to learn Standard British English. With regard to Received Pronunciation (RP) and General American models of standard English varieties, Kachru (1992:50) rightly stated that

> "Non-native speakers of English often aim at close approximation of these models (i.e. *RP* and *General American (GA)*) even at the risk of sounding affected. The works of Daniel Jones and John S. Kenyon encouraged such attempts. What Jones' outline of English Phonetics (1918) or English Pronouncing Dictionary (1956) did for RP, Kenyon's American Pronunciation did for GA…"

The limited available literature on Pakistani English is discussed in further detail in section 4.1. There is a gap in L2 perception literature for multilingual speakers with very little input of native L2, i.e. English in this case. The present study was designed to test the predictions and hypotheses proposed by L2 perception theories in a multilingual context, i.e. Pakistan. The main question to be addressed is whether L2 perception models are applicable to multilingual speakers in a multilingual context, and findings are discussed in Chapter 5 and Chapter 6. A brief overview of the L2 vowel perception literature is discussed in section 4.2 followed by the research questions and predictions for the present study. However, before that a brief overview of Pakistani English, and how it is distinct from Indian English, is presented.

### 4.1. Pakistani English (PE)

As discussed in Section 1.3, English is the official language in Pakistan, Urdu is the *national* language, and six major regional languages are spoken in the country. However due to the historically unstable political situation and dual education and language policies of the government, defining the status of English in Pakistan is not easy. Since the British colonial period (until 1947), English has been the official language, as it was considered a neutral language for multi-linguistic and multi-ethnic Pakistan and India.

English in Pakistan is considered a tool for social, economic, individual and national development. According to Rahman (2015:10), “English is very much in demand by Pakistani students and their parents and employers”. Mahboob (2002) conducted a survey to investigate student attitudes and beliefs about the English language and its role and status in Pakistan. Mahboob reported that most participants defined English as an international and global language, a medium of communication with foreign countries, a tool to enhance knowledge, learn new scientific discoveries, and a key to a bright and successful future career; one participant simply wrote, “No English, no future!” (2002:31).

Shamim (2008) reported the poor quality and outdated approaches to teaching English in government schools, where the focus is on grammar and translation into local languages or Urdu. English in Pakistan is heavily influenced by local languages. Some literature categorises English spoken in Pakistan as a distinct variety of English; however, there is very little literature on this point. Prior to Rahman (2015:24) the only available literature was from 1989: “...in 1989, the first printed version of the monograph was intended to fill the gap in knowledge about Pakistani English”.

Rahman (2015), in his revised monograph of Pakistani English (PE), reported that before 1984, the term Pakistani English was non-existent, and it was assumed that educated Pakistanis spoke standard British English (though there is no evidence to support this); any deviations in spoken or written English were considered mistakes and avoided at all cost. This prejudice against PE still continues. Rahman (2015:21-22) further reports that there are four sub-varieties of English spoken in Pakistan: the *anglicized* variety which is identical to RP except for some phonetic and phonological features; this variety is mainly spoken by highly educated and westernised families; the *acrolect*, the variety used by upper middle class who studied in elitist English-medium schools or had exposure to RP later; the *mesolect* used by most Pakistanis who studied in Urdu-medium schools and had no exposure to the standard British English; the *basilect* "... used by clerks, minor officials and typists etc who have not had much education. This kind of English is full of bureaucratic clichés and is the least intelligible variety for foreigners" (Rahman, 2015:22). He further argues that "...the ideal for teaching is RP, and all indigenous features of English are taken as deficiencies or errors" (Rahman, 2015:42). According to Rahman (2015), these sub-varieties of English in Pakistan reflect the class structure, where the influence of first languages is at a minimum in the Anglicized variety.

Rahman (2015:27-42) listed the phonological and phonetic features of the four sub-varieties of Pakistani-English and concluded that although English spoken in Pakistan is different from region to region (with regard to both the linguistic and social background of the speakers) as well as from Indian English, Pakistani Urdu and Punjabi speakers of English share some phonological features with the English spoken in North India.

In his description of Pakistani English, Rahman (2015) did not report any acoustic and phonetic analysis of these varieties and based his description on *personal observations* and written scripts from *"newspapers, magazines and other publications*". It is not clear how he identified the social class of the writer, and he did not provide any information about the newspapers and magazines. This information would have been helpful, since according to Baumgardner (1990:60) "There are 18 English language daily newspapers in Pakistan, 35 weekly publications, 33 fortnightlies, 152 monthlies and 111 quarterlies". A brief overview of the limited available literature on the phonetics and phonology of Pakistani English is given below.

### 4.1.1 Literature on Pakistani English (PE)

Kachru called Asian English "a transplanted colonial language" (Kachru, 1998:94) and divides the English language in three circles: inner circle (Australia, New Zealand) where

English is used as a first language, outer circle (Pakistan, India, Bangladesh, Malaysia, Singapore and Philippines), where English is used as an "institutionalised additional" language, and expanding circle (Bhutan, Brunei, China, Fiji, Hong Kong, Indonesia, Japan, Maldives, Myanmar, Nepal, Thailand, South Korea) where English is used as a foreign language.

> "...the input that English language learners received in South Asia was non-native and local. There was relatively little contact with native varieties of English in India, and after independence, this contact was further reduced. These factors have contributed to the institutionalization and evolution of South Asian English as a distinct variety." (Mahboob and Ahmar, 2004:1003)

According to Mahboob and Ahmar (2004), PE is considered heterogeneous for a number of reasons: socio-economic, geographic, educational and linguistic background of speakers. For example, Punjabi and Urdu speakers insert different vowels at different positions in English words that have consonant clusters that are not permitted in Urdu or/and Punjabi; for example, *start* is pronounced differently by native speakers of Punjabi and Urdu, i.e. [sətɑːrt] and [ɪstɑːrt], respectively.

Mahboob and Ahmar, (2004:1005) stated that speakers' linguistic background has effects on the production of consonant sounds. For example, Punjabi speakers replace English /ʒ/ with /j/ or /dʒ/. They further claimed that since Pakistan is a multilingual country, speakers with different linguistic backgrounds will probably speak English differently; hence the findings from their report cannot be generalised. Their claims are made without any acoustic investigation on the production of English consonants. For detailed review of corpus-based analysis of Pakistani English see Mahmood (2009).

According to Bolton (2008), Asian Englishes at a phonological level demonstrate a lack of distinction between vowel contrasts - such as long and short, high and low vowels -

and replacement of diphthongs with monophthongs. However, studies on PE show a distinction between short and long vowels (Sheikh, 2012; Bilal et al., 2011a and 2011b; Raza, 2008). Some diphthongs are pronounced as monophthongs (Mahboob and Ahmar, 2004; Rahman, 1991). Khan (2012) reviewed the limited available literature on Pakistani English and highlighted the fact that studies on Pakistani English are mainly based on its comparison with British and American English. Therefore, these studies reflect the English in Pakistan from an elitist perspective.

Mahboob and Ahmar (2004) collected data from six educated speakers (22-37 years old), four females and two males, from Karachi who spoke Urdu as their first language. They used the Sheffield set for recording and reported a phonological description of Pakistani English. They did not report any acoustic or statistical analysis. They classify the vowel realisation of PakE (Pakistani English – PE) by these native speakers of Urdu into three groups: (1a) vowels that are close to RP; (1b) vowels which are different from RP; (2a) vowels with no inter-speaker variation; (2b) vowels with inter-speaker variation. This classification of vowels is shown in Table 4.1 and Table 4.2 (after Mahboob and Ahmar, 2004:1003-1016).

They reported that most of the monophthongs and diphthongs are close to RP. However, in disyllabic words, where the second syllable is usually unstressed and its vowel is reduced to /ə/, the native speakers of Urdu failed to reduce and pronounced those vowels as full vowels.

For monophthongs, the variation is between tense and lax vowels, while for diphthongs the variation is between monophthongs and diphthongs. According to Rahman (1991), monophthongisation of RP diphthongs is a common characteristic of PE, especially the monophthongisation of /eɪ/ to /eː/ and of /əʊ/ to /ɔː/ or /oː/. The centring and closing diphthongs vary in their point of start and end; for example, NEAR: /ɪə/ is pronounced as

/eə/, SQUARE: /eə/ is pronounced as /əɪ/ or /ɑɪ/, and CURE: /jʊə/ is pronounced as /jeɔː/ or /eɔː/. Overall, their description of Pakistani English is in line with other studies on Pakistani English (Bilal et al., 2011a, 2011b and 2011c; Rahman, 1997). In addition, the monopthognisation of RP diphthongs could be because these diphthongs do not exist in the Urdu vowel inventory (see production experiments in Chapters 2 and 3 of this thesis for details).

Table 4. 1: List of English vowels similar to RP and showing no variation among Pakistani speakers (after Mahboob and Ahmar, 2004:1007-1008)

| Lexical item | PakE (PE) | RP |
|---|---|---|
| Monophthongs | | |
| KIT | ɪ | ɪ |
| HAPPY | ɪ | ɪ |
| THOUGHT | ɔː | ɔː |
| NORTH | ɔː | ɔː |
| FORCE | ɔː | ɔː |
| PALM | ɑː | ɑː |
| START | ɑː | ɑː |
| DRESS | e | e |
| TRAP | æ | æ |
| STRUT | ʌ | ʌ |
| FLEECE | iː | iː |
| GOOSE | uː | uː |
| Diphthongs | | |
| PRICE | aɪ | aɪ |
| CHOICE | ɔɪ | ɔɪ |
| MOUTH | aʊ | aʊ |

Table 4. 2: List of English vowels showing variation among Pakistani speakers (after Mahboob and Ahmar, 2004:1008-1009)

| Lexical item | PakE (PE) | RP |
|---|---|---|
| Monophthongs | | |
| FOOT | ʊ~uː | ʊ |
| BATH | ɑː~æ | ɑː |
| CLOTH | ɔ~ɔː~oː | ɔ |
| Diphthongs | | |
| FACE | eː~eɪ | eɪ |
| GOAT | oː~əʊ~ʊ | əʊ |
| GOAL | oː~əʊ | əʊ |
| NEAR | ɪə~eə | ɪə |
| SQUARE | eə~əɪ~ɑɪ | eə |
| CURE | jʊə~jeɔː~eɔː | jʊə |

Raza (2008) gave a description of Pakistani English based on his auditory observations of the utterances of 20 speakers, five speakers from four different L1 backgrounds: Urdu, Punjabi, Sindhi, and Pashto. He did not report any acoustic or phonetic analysis. He reported that RP /æ/ and /eɪ/ are pronounced as /e/ in PE, and that RP /ɪ/, /e/ and /æ/ are much longer in PE. RP /ɒ/, /ɔː/ and /əʊ/ are not pronounced as distinct vowels and are often pronounced as [o] or [a] depending on the social and educational background of the speakers. Further Raza also reported that in PE, RP /ɜː/ is realised as /əː/, and the RP central vowel /ʌ/ is retracted in PE. However, he did not provide any acoustic evidence to support this claim. He also reported that RP /ɔɪ/ is frequently replaced with a combination of a short vowel /ʊ/ and a diphthong /ae/. For instance, "toy is pronounced as [tʊae]" (Raza, 2008:107). This is an interesting observation that is not discussed in any other study. This pronunciation of "toy as [tʊae] could be due to influence of orthography; however acoustic and phonetic evidence is required to establish this claim.

In another impressionistic study, Sheikh (2012) collected data from 50 participants from 13 colleges and universities in Lahore to investigate PE vowels. Thus, unlike Raza (2008), her data were based on Punjabi speakers of English. Like many other studies, she only reported results based on a questionnaire and auditory observations of pronunciation of individual words from a list. She reported that /ɪ/, /i/, /æ/ and /u/ vowels in PE are the same as in RP. She further reported that participants made the following vowel substitutions: /e/ with [æ]; /ɒ/ and /ɔː/ with [ɑː], and /ʊ/ with [uː]. Some of the results reported by Sheikh (2012) are not clear. For instance, she reported that /ɪ/ in the word /wɪmen/ was pronounced as [ʊ]. This vowel alternation could be attributed to orthography, as participants were asked to read words from a list; however, Sheikh (2012) did not discuss the influence of orthography on the production of English vowels.

The above-mentioned studies discussed the whole vowel inventory of PE. However, the reported features of PE lack acoustic analysis in order to determine how similar the vowels of PE are to those of RP. Recently, studies on Pakistani English have been conducted with a focus on front, central and back vowels. These studies present some acoustic and phonetic analysis with a focus on comparison with British and American English (Bilal et al., 2011a, 2011b and 2011c; Hassan and Imtiaz, 2015; Farooq and Mahmood, 2017) and are discussed below in turn.

#### 4.1.1.1 Front Vowels in PE

Bilal et al. (2011a and 2011b) presented an acoustic analysis of the front four vowels /iː/, /ɪ/, /e/ and /æ/ of PE with reference to RP and AmE (American English) vowels. They collected data from 60 participants (30 males and 30 females) who were fluent in English and spoke Punjabi as their first language. They recorded these front vowels in three different contexts (i.e. hVd, sVt, pVt) and compared the acoustic properties with RP and American English vowels.

Bilal et al. (2011a and 2011b) reported the following: (1) PE speakers distinguished between /iː/ and /ɪ/; however, their vowels were less distinct than when spoken by RP speakers, (2) /e/ and /æ/ appeared less distinct when spoken by PE speakers as opposed to RP speakers, who produced /e/ as a close-mid front vowel and /æ/ as an open-mid front vowel, (3) there were fewer differences between PE and AmE, where /iː/ was produced as more raised and fronted by AmE speakers than by PE speakers, (4) there were significant differences in male and female productions, but these differences could be due to male speakers having more exposure to RP and AmE than female speakers. They concluded by saying that both RP and AmE, as well as local languages, have great influence on PE. However, they did not report any statistical tests. Their results are based on average formant frequencies which are given in Table 4. 3.

Table 4. 3: Formant frequencies of PE front vowels (after Bilal et al., 2011a:22-24)

| | Male | | Female | |
|---|---|---|---|---|
| All contexts: hVd, sVt, pVt | | | | |
| Vowel | F1 | F2 | F1 | F2 |
| [iː] | 374 | 2255 | 433 | 2724 |
| [ɪ] | 426 | 2031 | 509 | 2419 |
| [e] | 591 | 1790 | 716 | 2028 |
| [æ] | 632 | 1773 | 751 | 1870 |

Table 4. 3 shows that in PE /iː/ and /ɪ/ are distinct vowels, and /e/ and /æ/ are distinct only with regard to F1. This contrasts with other literature on Asian Englishes such as

Singaporean, Malaysiana, Indian English (Deterding, 2007; Kachru, 2005), which claims that speakers of Asian Englishes do not maintain the distinction between these tense and lax vowels. The formant frequencies reported by Bilal et al. (2011a and 2011b) are comparable with Hawkins and Midgley (2005); however, F1 for /iː/, /ɪ/ and /e/ is substantially higher, and F2 for /æ/ is substantially lower than the formant frequencies of RP monophthongs reported by Hawkins and Midgley (2005).

#### 4.1.1.2 Central Vowels in PE

Bilal et al. (2011c) investigated the status of three central RP vowels /ə/, /ʌ/ and /ɜː/ in PE. They collected data from 20 speakers (10 males and 10 females) who were undergraduate students at the University of Sargodha and spoke Punjabi as their first language. The stimuli were three words per vowel i.e. [ʌ] hunt, *stunt*, *punt*; [ə] *mother, oven, famous*; [ɜː] *skirt, spurt, hurt*. They reported that PE has two central vowels, [ə] and [ʌ], where these two phonemes are distinct with regard to F1, which is slightly higher for [ʌ] ($p < 0.01$). [ʌ] is slightly lowered and fronted whereas [ə] is slightly raised and retracted.

In addition, Bilal et al. (2011c) reported that Pakistani speakers of English do not differentiate between [ə] and [ɜː], and indeed their F1 and F2 formant frequency difference is non-significant. They concluded that Pakistani English does not distinguish between /ə/ and /ɜː/. They further claimed that due to the influence of Punjabi, the difference between [ə] and [ʌ] is only significant with regard to F1. They found that gender has non-significant effect. Their findings were in line with Mahmood et al. (2011), who investigated the phonological adaptation of English loanwords in Punjabi, and reported that RP [ɜː] is not present in Pakistani English spoken by Punjabi speakers. The average formant frequencies of PE central vowels, as reported by Bilal et al. (2011c), are given in Table 4. 4.

Table 4. 4: Formant frequencies of PE central vowels (after Bilal et al., 2011b:8-10)

| | Male | | Female | |
|---|---|---|---|---|
| Vowel | F1 | F2 | F1 | F2 |
| [ə] | 514 | 1365 | 654 | 1624 |
| [ɜː] | 537 | 1396 | 678 | 1610 |
| [ʌ] | 626 | 1473 | 805 | 1696 |

**4.1.1.3 Back Vowels in PE**

Hassan and Imtiaz (2015) investigated the presence of RP /ɒ/ and /ɔː/ in PE by collecting data in two contexts, hVd and kVd, spoken by 20 students (10 males and 10 females) from Government College University, Faisalabad who speak Punjabi as their first language. They reported that Pakistani speakers of English do not differentiate between RP /ɒ/ and /ɔː/ with regard to F1/F2. Hence in PE these two vowels are considered a single phoneme. They did not report any temporal differences between these two vowels. They also did not report any statistical analysis, and based their arguments on the average F1 and F2 values as shown in Table 4. 5. They concluded that for both /ɒ/ and /ɔː/, F1 and F2 are not different in either context, specifically hVd (hod, horde) and kVd (cod, cawed).

Despite the lack of statistical analysis, these results are in line with Raza (2008) who reported that English /ɒ/ and /ɔː/ are found in the speech of very few Pakistanis, specifically those who either have been educated or raised in the inner circle of English-speaking countries. Otherwise, most Pakistani speakers replace English /ɒ/ and /ɔː/ with [oː]. For example, the students from rural areas or suburbs of big cities, like Karachi, replace /ɒ/ and /ɔː/ with [ɑː] (Raza, 2008). Thus, PE has a lot of variety, especially influenced by a speaker’s socio-economic, linguistic, geographical, and educational background.

Table 4. 5: Average Formant frequencies of PE back vowels (after Hassan and Imtiaz, 2015:29)

| | Male | | Female | |
|---|---|---|---|---|
| Vowel (hVd) | F1 | F2 | F1 | F2 |
| [ɒ] | 521 | 1051 | 704 | 1221 |
| [ɔː] | 574 | 1034 | 646 | 1164 |
| Vowel (kVd) | | | | |
| [ɒ] | 557 | 976 | 682 | 1169 |
| [ɔː] | 586 | 1013 | 655 | 1170 |

**4.1.1.4 Diphthongs in PE**

Farooq and Mahmood (2017) investigated the acoustic behaviour of RP diphthongs in Pakistani English. They collected data from 30 educated Pakistanis (12 males and 18 females). Their results are based on perceptual identification of syllables and acoustic investigation of F1, F2 and duration. They showed that RP diphthongs /ʊə/ and /əʊ/ were pronounced as the monophthong [o], and RP /eɪ/ was pronounced as the monophthong /e/. However, RP /ɪə/ and /eə/ were both pronounced as [eə]. Interestingly, RP /aɪ/, /aʊ/, eə/ and /ɔɪ/ showed comparable features with RP only when these vowels occurred at word medial position. It is not clear if this applies only to diphthongs found in monosyllabic words consisting of a closed syllable, or also in multisyllabic words. They further reported that RP /ɔɪ/ had two different pronunciations at the word final position, i.e. [ɔe] or [ʊae]. Overall, their findings are in line with other studies (Rahman, 2015; Raza, 2008; Mahboob and Ahmar, 2004; Rahman, 1991). However, there are two main problems with this study: firstly, they did not provide any information on the linguistic background of the speakers, hence it is not clear if these findings can be generalised across speakers of Pakistani English with different L1s; secondly, their results are based on descriptive statistics of extracted formant frequencies and lack statistical significance tests to support their claims.

Overall, the literature on Pakistani English tends to agree on the set of vowels. However due to a lack of comprehensive analysis and standard experimental designs, these studies are not easy to compare to each other. Some of these studies included speakers with different L1s (Rahman, 2015; Raza, 2008), but the majority of the studies are based on speakers from a single city, specifically Karachi (Mahboob and Ahmar, 2004), Lahore (Sheikh, 2012), Sargodha (Bilal et al., 2011a, 2011b and 2011c), and Faisalabad (Hassan and Imtiaz, 2015). Therefore, the findings from these studies cannot be conclusively generalised as Pakistani English. Most of these studies also report that the social and educational background of the speaker, along with linguistic background, has a strong influence on the production of English in Pakistan (Raza, 2008; Rahman, 1991).

The above-mentioned studies on Pakistani English claimed that the features of Pakistani English are different from the English spoken in India. However previous studies on Indian English and Asian Englishes do not draw this distinction (Kachru, 1998; Deterding, 2007; Bolton, 2008).. For a comparison of Indian English vowels with the above discussed vowels of Pakistani English vowels see Gargesh (2006 and 2004:994-997); Maxwell and Fletcher (2009 and 2010). Overall, the common features between Pakistani and Indian English are the confusion with central vowels and monophthongisation of some diphthongs. Tense and lax vowels are reported to be problematic in Indian English; however, literature on Pakistani English reported the contrary. Lastly, the extensive literature on Indian English shows that this variety of English is well researched and documented as compared to Pakistani English. Knowing how IE and PE varieties relate to one-another will help us in the following sections when we wish to interpret the perceptual assimilation patterns of SSBE vowels by Punjabi-Urdu speakers.

As discussed earlier, there is no study on the perception of Standard Southern British English (SSBE) vowels by Punjabi-Urdu speakers, however similar studies on different languages have been performed. A brief overview of the literature on the perception of second language vowels is presented in the next section.

**4.2 Second Language Perception**

As discussed in the previous section, English spoken in Pakistan shares some features with Indian English, but this variety is highly influenced by local languages. Studies on second language acquisition report that after the "critical period" second language learners speak L2 with a strong foreign accent (Scovel, 1969; Patkowski, 1990). According to the Critical Period Hypothesis (CPH) as proposed by Lenneberg (1967), maturation of neurobiological system hinders L2 acquisition like the first language (L1). However, the CPH has been challenged by other literature on L2 speech perception and production, especially with regard to the period of learning (i.e. the number of years) and the exposure to native input (Flege, Munro and MacKay, 1995; Flege, 2009). In addition, research has shown that length of residence in an L2 speaking country (if applicable), motivation for L2 learning, general language learning aptitude, and everyday use of L1, also contribute to the mastery of L2 (Piskey, MacKay and Flege, 2001).

The studies on multilingual and bilingual learners' acquisition of a third language (L3/L*n*) report that both L1 and L2 affect the acquisition of L3 (Hammarberg and Hammarberg, 2009; Hammarberg, 2014; Llama, et al., 2010; Lipińska, 2017). As L1 influences the perception and production of L2, the research on bilingual language acquisition proposed three different models: (a) *unitary* system (i.e. a single system develops into two separate systems), (b) *dual* system (i.e. bilingual children develop two separate systems which do not interact), and (c) *interdependence* system, which is further subdivided into *transfer*, *deceleration* and *acceleration*, and relates to the interaction between L1 and L2 (Paradise and Genesee, 1996 cited in Fabiano-Smith and Goldstein, 2010:160). The interaction between the two languages of bilingual learners enhances the learning process for certain aspects, e.g. acquisition of phonetically similar sounds, but slows the learning of certain linguistic features, e.g. phonology. Thus, the interaction between L1 and L2 results in errors in production as compared to monolinguals of a similar age group (see Fabiano-Smith and Goldstein, 2010 for review).

There is substantial research available on the perception and production of L2 (Fox, Flege and Munro, 1995; see Flege, 2003 for review); however very little is known how a target language is perceived by multilingual speakers. Early research on L2 speech focussed on production only; however recent research has also focussed on the perception of L2 speech (e.g. Colantoni, Steele and Escudero, 2015). The errors in L2 speech production are believed to be based on the errors in perception of those segments (Strange, 1995). A number of theoretical models have been proposed and tested to investigate how L2 learners perceive and produce speech sounds (i.e. vowels and consonants). A brief overview of the main competing models on second language perception is presented in the next section.

### 4.2.1 Second Language Perception Models

#### 4.2.1.1 Native Language Magnet

For instance, the Native Language Magnet model (Kuhl, 1993) focuses on the role and influence of L1 on the perception and production of L2. According to Kuhl (1991) "perceptual magnet effect" is a phenomenon that shows how experience of a language alters phonetic perception. This model suggests that exposure to a language at an early age (i.e. between 6 and 12 months) results in language specific perceptual sensitives

(Kuhl, 2000) and these sensitives alter the perception of subsequent sounds. In other words, the language specific perceptual sensitivity hinders the perceptual sensitivity to non-native phonemes, especially in adulthood (Iverson et al., 2003). This model has been challenged by research (e.g., Lively and Pisoni, 1997; Lotto, Kluender and Holt, 1998; Frieda et al., 1999), and as the focus of the present study is on L2 speech perception by adults, this model will not be explored further.

#### 4.2.1.2 Speech Learning Model (SLM)

The Speech Learning Model (SLM) (Flege, 1995; 2002) focusses on how L2 learners perceive and produce L2 segments. SLM proposes that L2 speech production is highly influenced by L2 speech perception. SLM also proposes that it is hard to formulate new categories for L2 sounds if they are very close to L1 sounds, and as a result these categories will be more difficult to acquire. Therefore, for dissimilar sounds new categories will be formed. However, the similar sounds will be subsumed under the relevant L1 category. As a result, this single category formation creates difficulty for L2 learners when they cannot detect the phonetic differences between two L2 phonemes.

#### 4.2.1.3 Perceptual Assimilation Model (PAM)

The Perceptual Assimilation Model (PAM) (Best, 1994, 1995) was designed to investigate how naïve listeners assimilate and/or discriminate unfamiliar non-native contrasts. Best and Tyler (2007) proposed an extension of PAM, PAM-L2, to investigate how L2 learners will discriminate non-native contrasts. This model has been divided into three sections (Best and Tyler, 2007): (1) the L2 contrasts, which are assimilated to the same L1 category (*single-category* pattern - SC), are predicted to be very difficult to differentiate; (2) L2 contrasts, which are assimilated to two different L1 segments (*two-category* pattern - TC), are predicted to be very easy to differentiate and assimilate; (3) L2 contrasts, which are assimilated to the same category, can be differentiated and assimilated very easily depending on the rate of *category-goodness* (CG). The higher CG rating will indicate which L2 phoneme was perceived as a better example of an L1 phoneme, whereas a lower CG rating will show the reverse. In addition, PAM predicts that given an L2 contrast pair, if one of the L2 phones is distinct from any L1 phoneme, then they are treated as *Uncategorised-Categorised* (UC); and if two L2 phones are perceived as noise and so not assimilated to any L1 segment, they are "uncategorisable" or "unassimilable" (*Uncategorised-Uncategorised* -UU) (Best, 1995; Strange 1998; Guchiliakaya, 2007; Tyler et al., 2014).

According to Best and Strange (1992), the perception of vowels and consonants remains flexible even after the critical period. PAM (Best, 1995; PAM-L2; Best and Tyler, 2007) and SLM (Flege, 1995) are two competing models in the literature on L2 perception. These models agree that the degree of difficulty in the perception and production of L2 segments varies depending on their resemblance with L1 phonemes (SLM) or how those L2 segments are assimilated to L1 phonemes.

#### 4.2.1.4 Second Language Linguistic Perception Model (L2LP)

More recently, the Second Language Linguistic Perception model (L2LP) has been proposed (L2LP: Escudero, 2005), and predicts the difficulties faced by L2 learners based on the acoustic similarities and differences between L1 and L2. According to L2LP, cross-language acoustic similarities between L1 and L2 vowels can help to predict the assimilation, categorisation and discrimination patterns of L2 vowels. L2LP was mainly "designed to account for individual variations in non-native speakers of varying proficiency" (Colantoni et al., 2015:44). Its predictions are similar to PAM, especially with regard to *single-category* (SC) and *two-category* (TC) assimilation patterns. In addition, like SLM, L2LP predicts cross-language categorisation based on acoustic properties of L1 and L2 vowels. As briefly discussed in Section 1.4.1, the *similar* and *new* *scenario* features of L2LP are explored in the present study under the umbrella of PAM (*two-category* and *single-category*) and SLM, since L2LP shares predictions with PAM and the role of L1 acoustic properties with SLM.

#### 4.2.1.5 SLM and PAM – Similarities and Differences

Although SLM was developed to investigate the perceived similarities and differences between individual L2 phones and L1 phonemes, and PAM was developed to investigate the perceptual assimilation and identification of L2 contrasts, SLM and PAM agree on the notion that accuracy in the production of L2 sounds is based on the perception of L2 sounds. In other words, L2 perceptual errors lead to inaccurate L2 productions. PAM and SLM also agree on the notion of that "perceptual learning process remains intact throughout life" (Best and Tyler, 2007:24). PAM and SLM also agree that L1 and L2 categories exist in the same phonological space, therefore, the two systems interact and as a result L2 learners cannot achieve monolingual-like performance (Colantoni, et al., 2015). However, these models differ in their predictions of perception of L2 segments.

The differences between PAM and SLM are given in Table 4.6 and the common features in PAM and SLM are given in Table 4.7

Table 4. 6: The differences between SLM and PAM

| PAM | SLM |
|---|---|
| PAM tests the perceptual assimilation patterns of pairs of L2 phones to L1 phonemes | SLM focuses on the individual phones |
| According to PAM the speakers' articulatory gestures (tongue position, vocal tract, manner of articulation, i.e. fricative, lateral) play an important role in the perception of L2 sounds | SLM proposes that phonetically relevant features (phonetic categories) are stored in the long-term memory and are used for cross language perception of L2 segments |
| PAM-L2 tests inexperienced L2 learners | SLM tests the effects of experience on L2 learning, so in practice SLM tests and compares both experienced and inexperienced L2 learners |
| PAM predicts L2 difficulty based on perceptual assimilation of L2 contrasts to native language phonemes | SLM predicts difficulty with regard to acoustic comparisons between L1 and L2 |

Table 4. 7: The common features in SLM and PAM

| PAM | SLM | Common features |
|---|---|---|
| *Single-Category* (SC) assimilation | *Perceptual equivalence* | If two L2 phonemes are heard as similar to one L1 phoneme, this will be SC, according to PAM; no new category will be formed according to SLM. |
| *Two-Category* (TC) assimilation | *L1-L2 phonetic similarity* | If two L2 phonemes are perceived as similar to two L1 phonemes (as opposed to a single L1 phoneme), this will result in TC according to PAM; listeners will be able to detect the phonetic differences between the L1 and L2 sounds, according to SLM. |
| *Uncategorised-Uncategorised* (UU) | *New category formation* | This means that these phonemes will be discriminated easily according to PAM, for this new phoneme a new category will be formed according to SLM, since it is absent from the L1. |

Previous studies on SLM and PAM mainly focussed on the perception of English vowels as a foreign or second language, with some exceptions, e.g. French vowels (Levy and Strange, 2008) and Norwegian, French and Thai vowels (Tyler et al., 2014). These models have proposed a number of hypotheses as discussed above, and a number of studies have been conducted to test these hypotheses for the perception of vowels. In the next section, a survey of these studies is presented for each model.

### 4.2.2 Studies Testing SLM

#### 4.2.2.1 New Category Formation and Position-Sensitive Allophones

Flege's SLM (1987; 1995) proposes that it is the difference (phonetic dissimilarity) between L1 and L2 segments that results in better perception and production of L2. With regard to the formation of new categories for L2 segments that are not present in L1, according to SLM it is easier for native English speakers learning French to perceive and produce the French vowel /y/ more accurately and easily than /u/, which can be easily confused (*equated*) with the English vowel /uː/ (Flege, 2003). In other words, new sounds from L2 are easier to perceive and produce than those that are phonetically close to the L1 sounds. "The greater the perceived phonetic dissimilarity between an L2 sound and the closest L1 sound, the more likely it is that phonetic differences between the sounds will be discerned" (Flege, 1995:239).

In order to support this hypothesis, Aoyama et al. (2004) conducted perception and production experiments to investigate Native Japanese speakers' perception of English /l/ and /ɹ/, where English /ɹ/ is considered perceptually more dissimilar from Japanese /r/ than English /l/. Their longitudinal study was conducted on both young and adult learners of English. Their findings supported the hypothesis that it was easier for Japanese learners to perceive and produce English /ɹ/, which is quite different from Japanese /r/ as compared to English /l/. Further, their findings showed that over time young learner's performance improved for English /ɹ/.

Previous research has also shown that due to phonetic category assimilation or dissimilation, the vowels produced by bilinguals and monolinguals differ in their spectral and/or temporal quality. For example, Flege, Schirru, and MacKay (2003) investigated the production of English /eɪ/ by four groups of Italian-English bilinguals. Their findings showed that early bilinguals, with low use of L1, established new categories for L2 segments; however late bilinguals, both with low and high use of L1, merged the L2 phoneme with an existing L1 phoneme. Thus English /eɪ/ was produced with less tongue movement by late bilingual Italian speakers of English than native English speakers, and early bilinguals produced English /eɪ/ with more tongue movement than native English speakers. This study showed that early Italian-English bilinguals formed a new category for English /eɪ/, which was not only different from Italian /e/ but also from English /eɪ/. These findings supported the hypothesis that L1 and L2 phonetic systems interact either by new category formation or merging two (L1 and L2) categories.

Further, SLM proposes that the position of an L2 segment in a word (*position-sensitive allophones*) plays an important role in the perception and production of that segment. For instance, Flege, Tagaki, and Mann (1996) found that Japanese-speakers of English perceive English /ɹ/ better than English /l/, especially in word initial position, which suggests that formation of a new or similar category for an L2 segment depends on the level of position-sensitive allophone. English /l/ is considered to be closer to Japanese /r/. Hence Japanese speakers distinguish English /ɹ/ as a new or different phoneme from their L1, and can establish a new category for this sound. According to Sheldon and Strange (1982), Japanese-speakers of English perceive English /ɹ/ and /l/ better in word final position than word initial position, which supports Flege's (1995) *position sensitive allophone* hypothesis.

**4.2.2.2 Effects of Experience and Exposure to L2**

Some L2 perception studies report that linguistic experience not only influences the perception of L2 sounds but also of L1 sounds (Flege et al., 1994; Boomershine, 2013) and leads to the development of an interlanguage phonological space, so that bilinguals do not perform like monolinguals of either language.

SLM also emphasises the effects and correlation between age of learning, exposure to L2 and the use of L1(see Flege, Bohn, and Jang, 1997; Flege et al. 1998; Flege, Schirru, and MacKay, 2003 for details). However, these factors are not relevant to the present study and cannot be tested due the following reasons: learners start learning L2 (English in this case) from a very early age; learners are multilingual (they speak at least three languages) and are fluent in all; and the input they receive is far from native except for British and American cinema and television.

Previous studies have also shown the effects of proficiency on the perception of L2 vowels and suggested that increased L2 proficiency causes the modifications in the vowel space. Fox et al. (1995) investigated the perception of seven English vowels, /i/, /ɪ/, /eɪ/, /ɛ/, /æ/, /ʌ/, and /ɑ/, and three Spanish vowels, /i/, /e/ and /a/, by English monolinguals and Spanish–English bilinguals. Their multidimensional analysis showed three-dimensional patterns of identification of the vowels by English monolinguals (vowel height correlated with duration, which highlighted the language-dependent sensitivity to this phonetic feature; front/back; and central/non-central) and two-dimensional identification patterns by Spanish-English bilinguals. Further, their results showed that the performance of Spanish-English bilinguals was similar to native speakers of English, as compared to the Spanish speakers who were not proficient in English and so confused the vowels. This led the authors to conclude that with the increase in L2 proficiency, identification of vowels is altered due to the modifications in the vowel space. However, it has not been tested if this sensitivity to the correlation between vowel height and duration can be seen in multilingual speakers of English.

With regard to L2 experience, Bohn and Flege (1990) investigated the identification and discrimination of four American English (AE) vowels, /i/, /ɪ/, /ɛ/ and /æ/, by native speakers of German. Based on perceptual similarities (which were drawn by comparing spectral and temporal properties) between AE and German vowels /i/, /ɪ/ and /ɛ/, and the absence of /æ/ from the German vowel inventory, the identification (labelling) task yielded the predicted results: AE /i/ and /ɪ/ were mapped to expected German vowel categories with a strong goodness of judgment ranking (on a scale of 1-5). However listeners were not very confident when mapping AE /ɛ/ and /æ/ to German vowel categories. According to Bohn and Felge (1990:310), this could be due to the fact that AE /ɛ/ and /æ/ "do not have clear counterparts in German". Further, based on the acoustic comparison between AE and German vowels, they concluded that spectral cues were significant in identification of /i/ and /ɪ/; however, duration cues were significant for the identification of /ɛ/, and the results for /æ/ were difficult to interpret as listeners did not have a corresponding vowel category in their L1. Therefore, AE vowel /æ/ was classed as new for native speakers of German. According to SLM a new phonetic category will be established for this vowel.

In the second part of this study, Bohn and Flege (1990) conducted a discrimination experiment of the continua *beat–bit* /i/-/ɪ/ and *bet-bat* /ɛ/-/æ/, differing in both spectral quality and duration. Their results showed that both experienced and inexperienced listeners' performance was comparable for the perceptually (acoustically) similar vowels (/i/-/ɪ/) in AE and German. Both experienced and inexperienced listeners relied on temporal cues, thus L2 experience was not significant. However, experienced learners performed better, and relied more on spectral than temporal cues, for the discrimination of the /ɛ/-/æ/ contrast. This suggested that experience with L2 can lead to native-like performance. They concluded that experience with L2 plays an important role for the perception of vowels that are absent from L1. This was an -informative study. However, it is not made clear why experience with L2 did not affect the perception of /i/-/ɪ/.

Bohn and Flege (1992), further investigated the effects of experience on the production of English vowels by native speakers of German with varying experience of English. Acoustic comparisons of AE vowels /i/, /ɪ/, /ɛ/ and /æ/ produced by German (both experienced and inexperienced) and English speakers confirmed their hypothesis that L2 experience affects the production of new vowels and does not affect the production of L2 vowels that are similar to L1 vowels. However, their perception tests for the new vowel /æ/ did not support this hypothesis.

#### 4.2.2.3 Role of L1 in L2 Perception

The L2 perception studies show that the perceived relation between L1 and L2 vowels, and experience and exposure to L2, plays an important role in the accurate production and perception of L2 vowels. Flege, Bohn, and Jang (1997) investigated the role of L1 and experience on the production and perception AE vowels by experienced and inexperienced non-native speakers from four different linguistic backgrounds: German, Korean, Mandarin and Spanish. Similar to Bohn and Flege (1990), the perception tests consisted of an identification task of synthetic continua of *beat–bit* /i/-/ɪ/ and *bet-bat* /ɛ/-/æ/, differing in both spectral quality and duration. Their predictions for perception of these contrasts were based on the acoustic analysis of the similarities and differences between L1 and L2 vowels, i.e. German, Korean, Mandarin, Spanish and English vowels. Their findings supported the hypothesis that the production of L2 vowel contrasts depends on the perception of those segments. However, they found that the production and perception of *beat–bit* /i/-/ɪ/ vowels did not differ between the experienced and inexperienced native Spanish speakers. In their production, there was very little spectral difference between the two vowels; however, in their perception they showed the identification and reliance on spectral cues, which showed that L2 learners' perception and production did not match. They concluded that the perceived relation between L1 and L2 vowels, and experience and exposure to L2, plays an important role in the accurate production and perception of L2 vowels. However, they did not explain all cases where the perception and production of AE vowels did not differ between experienced and inexperienced speakers of English.

#### 4.2.2.4 Effects of Age of Learning (Early vs Late)

According to Flege, MacKay and Meador (1999), early exposure to L2 results in better perception and production of L2 sounds. Flege et al (1999) investigated the categorical discrimination of English and Italian vowels in three different sets as follows: English vowel pairs: /i/-/ɪ/, /u/-/ʊ/, /æ/-/ʌ/, and /ɒ/-/ʌ/; one English and one Italian in vowel pairs: /æ/-/a/, /ʌ/-/a/, /u/-/o/ and /e/-/e/; Italian vowel pairs /u/-/o/, /e/-/a/ and /u/-/i/). Their participants were highly experienced Italian-English bilinguals and they were divided into four groups based on their Age of Arrival (AoA) in Canada. Their results showed that native Italian speakers who started to learn English early discriminated English vowels better than late learners, “early bilinguals who are highly experienced in their L2 may perceive L2 vowels in a nativelike fashion” (Flege et al., 1999:2981). These findings led them to conclude that early bilinguals might have established phonetic categories (Flege, 1995) for English vowels /ɪ/ and /ʊ/ which do not occur in Italian. Hence the discrimination patterns for English /i/-/ɪ/ and /u/-/ʊ/ were different between early/mid (AoA) bilinguals and late bilinguals. From the production experiment, Flege et al. (1999) found that late Italian learners of English neutralised the differences between low and mid-central English vowels such as /æ/, /ɒ/ and /ʌ/. Hence the categorical discrimination task for vowel contrasts /æ/-/a/, /ʌ/-/a/, /ʌ/-ɒ/ and /ʌ/-/æ/, the late bilingual group scored less than native speakers of English. Overall, this study supported two SLM hypotheses, i.e. category formation by early learners of L2, and accurate production of L2 depends on the accurate perception of L2. However, later studies (Flege and MacKay, 2004) have shown that native-like perception does not guarantee native-like production of L2 segments.

In an earlier study, Flege et al. (1998) found that three-year exposure to native English input affected the perception of English consonants by native Japanese speakers. Experienced Japanese speakers performed better than inexperienced speakers for the discrimination of consonant contrasts: /ɹ/-/l/, /s/-/θ/, /b/-/v/, and /ɹ/-/w/; however, there was no difference in their performance for the identification of vowel contrasts: /u/-/ʊ/, /ɑ/-/ʌ/, /eɪ/-/ɛ/, and /i/-/ɪ/. This suggests that perceived relations between English and Japanese consonants may change with experience and exposure, but perceived relations between English and Japanese vowels remains unaffected.

Previous studies have supported SLM hypothesis that children are good at category establishment, since they can detect the phonetic and acoustic differences between L1 and L2. Baker et al. (2002) investigated the identification of eight English vowels, /i/- /ɪ/, /u/-

/ʊ/, /æ/-/ɛ/, and /ɑ/-/ʌ/ by Korean speakers. According to Flege, Bohn and Jang (1997 cited in Baker et al., 2002:3), these pairs of vowels are considered to be difficult to discriminate and easily confused by Korean speakers in perception and production. Cross-language identification tests were carried out with adults (22-23 years old) and children (7-9 years old), all monolingual Korean speakers who have been living in the US for a year or less. The English vowels were embedded in three different contexts: /bVt/, /nVt/ and /hVd/. The listeners had to pick one of the 10 standard Korean vowels and rate on a 7-point scale how similar that sound was to a given Korean vowel. The results showed that assimilation patterns for English vowels were similar for both adults and children. For example: English /i/-/ɪ/ were matched to Korean vowel /i/; English /u/-/ʊ/ were matched to Korean /u/; English /æ/-/ɛ/ were matched to Korean /ɛ/ and /e/ (these were the most confusing vowels for them); and English /ɑ/-/ʌ/ were matched to two separate Korean vowels /a/ and /ʌ/ respectively. The only difference in perception between the two age groups (i.e. adults and children) was the goodness rating, where children's goodness rating was lower than adults, which suggests that they did not consider English vowels good examples of native Korean vowels. This finding supports the SLM hypothesis that children are good at category establishment, since they can detect the phonetic and acoustic differences between L1 and L2, and as a result perform better (more native-like) in perception and production of L2 vowels. Overall, this study suggests that adult/late learners are unlikely to establish new categories for L2 once their L1 phonological system is fully developed. Further, their results showed that L2 exposure and input improves an adult learner's perception and production of only those vowels that are non-confusing and similar to L1 vowels.

In order to answer the question of whether early L2 learners can perceive L2 vowels like native speakers of that L2, Flege and MacKay (2004) investigated the perception of six English vowels /ɒ/-/ʌ/, /ɛ/-æ/ and /i/-/ɪ/ by native speakers of Italian, grouped by: early (child) vs. late (adult) learners, age of arrival in Canada (AoA), length of residence (LoR) in Canada, and regular use of L1 (Italian). Their findings showed that though early learners were better in their discrimination and perception, some adult learners also perceived the contrasts accurately. Despite this, their findings broadly supported the SLM hypothesis that late learners' perception of L2 vowels is not as accurate as early learners. They showed that experience and exposure to L2 can help to establish new categories for L2 vowels, as this ability remains intact throughout life (Flege, 1995). Interestingly, some of their findings from early learners of L2, who used L1 more often than L2, were not as predicted and they concluded that learning an L2 at a young age does not guarantee native-like competence in L2. Similar to previous studies, this study did not answer why age (child vs adult) and experience (exposure to L2) affect some learners' perception of L2 vowels but not others.

Contrary to the above-mentioned studies, Jia et al. (2006) reported that experienced adult L2 learners perform better than young L2 learners. They reported that the amount of L2 exposure and age (young vs. adult learners) has a strong effect on the perception and production of L2 vowels. In particular, their findings suggested that in a non-native context with non-native input of L2, adult learners have an advantage in accurately perceiving and producing L2 sounds. In China, participants' lack of exposure to native input "…an older chronological age predicted a significantly higher discrimination accuracy of all vowel contrasts and higher production accuracy of two difficult vowels" (Jia et al., 2006:127). This contradicts the SLM hypothesis that children are better at perceiving and producing L2 vowels because their L1 phonetic system is not fully developed (Baker et al., 2002).

Individual differences in vowel perception were highlighted by Wang (2006) who investigated the perceptual assimilation patterns of Mandarin speakers living in Canada for six English vowels, /i/-/ɪ/, /u/-/ʊ/ and /æ/-/ɛ/, with temporal and spectral variations by synthetic manipulation. The results showed that Mandarin speakers used duration cues more so than spectral cues for the English vowel pair /i/-/ɪ/. Hence their assimilation patterns were not native-like. In addition, the results from individual listeners showed different strategies for discrimination/identification, and that most of the listeners did not manage to identify English vowels /æ/-/ɛ/ and /u/-/ʊ/ as two distinct categories. This is somewhat similar to how native speakers of American English integrated both spectral and duration cues to identify French /ɔ/-/o/ in synthetic stimuli, whereas native speakers of French used only spectral cues to identify these two vowels (Gottfried and Beddor, 1988).

Gottfried and Beddor, (1988) investigated the perception of French /o/-/ɔ/ vowels in /kot/-/kɔt/, with temporal and spectral variations using synthetic manipulation. Participants consisted of a group of native French speakers and two groups of American English speakers, one that studied French and another that did not study French. Their results showed that French native speakers did not pay attention to the temporal cues in both categorisation and category rating tasks. However native speakers of American English integrated spectral and temporal information in order to categorise the two vowels. This led them to conclude that

> "[P]erceptual integration of the acoustic properties relevant to a given vowel contrast does not simply follow from experience with that contrast. Rather, perceptual integration depends on the extent to which the acoustic properties correlate within the broader context of a phonological system" (Gottfried and Beddor, 1988:63).

Although experience and exposure to L2 (early versus late) is considered an important factor in the perception of L2, the relationship between temporal and spectral cues and the phonological system of L1 also plays a significant role in the perception of L2 sounds. This could explain why experience does not affect the perception of certain L2 sounds (Flege and MacKay, 2004).

All the above studies are based on monolingual L2 learners who receive native input at some point in their life. None of these studies considered L2 learners who start learning English at a very young age, but do not receive native input except for Jia et al. (2006).

In the past three decades, the research on SLM has focussed on the role of L1, experience, exposure to L2, and age of learning. The studies on SLM showed that accurate perception and proficiency do not guarantee accurate production. These studies lack investigation of individual variations in speech perception and production in order to answer why age, experience and exposure to L2 do not affect perception and production in some cases.

### 4.2.3 Studies Testing PAM and L2LP

Best's Perceptual Assimilation Model (PAM; Best et al., 1988; Best, 1995; Best and Tyler, 2007) predicts that an L2 listener's discrimination of L2 or non-native sounds depends on the perceived relation of these non-native sounds to those in their first language. This suggests that L1 greatly influences the perception and production of L2.

A number of studies have been conducted to test these PAM predictions. However, most of those studies investigated the perceptual assimilation and discrimination of L2 consonants. More recently, PAM predictions have been tested on the perceptual assimilation of L2 vowels (Faris et al., 2016; Tyler et al., 2014; Bundgaard-Nielsen et al., 2011; Escudero and Williams, 2011; Gilichinskaya and Strange, 2010; Levy and Strange, 2008; Strange et al., 2004; Strange et al., 2005).

Tyler et al. (2014) extended PAM predictions to test the categorisation and assimilation patterns of non-native vowel contrasts by native speakers of American English. The six vowel contrasts were chosen from three different languages, i.e. Norwegian, French and Thai, and listeners were not told about these languages. They employed two tasks, AXB discrimination and categorical assimilation followed by goodness rating judgement on a 1-5-point scale. The analysis of individual participant assimilation patterns revealed SC, TC, CG, UC and UU patterns. The results from the AXB discrimination task supported the PAM predictions that, for UC and TC contrasts, discrimination performance was excellent. They further concluded that assimilation patterns and types vary across individuals. Even though the UU assimilation pattern showed excellent discrimination, they did not analyse these patterns further.

Polka and Bohn's (1996) investigation of German and English adult listeners' discrimination and identification of two vowel contrasts /u/–/y/ showed category goodness assimilation as German /u/–/y/ were assimilated to English /u/. They found that discrimination patterns showed accurate discrimination, however assimilation patterns showed that English monolinguals matched the German /u/–/y/ to English /u/, where German /u/ got higher ratings (mean 3.89) than German /y/ (mean 2.8). According to PAM, this is an example of CG where one segment is a good exemplar of a native category and the other a poor exemplar; hence discrimination should be easy, as indeed was shown in the discrimination task.

Bundgaard-Nielsen et al. (2011) investigated the assimilation patterns for the whole vowel inventory (18 vowels) of Australian English in two contexts, (sentence and citation) to Japanese vowel categories (both mono-moraic and bimoraic) by 31 Japanese learners of Australian English. The listeners were divided into two groups based on their vocabulary size in L2 (high vs. low). After the mapping of each vowel they also rated its goodness on a 1-7-point scale. Overall, their results did not show any difference in the assimilation patterns in sentence and citation context, which meant the learners were sensitive to both spectral and temporal information in both contexts. However, the higher mean goodness in the sentence context suggested that Japanese listeners were sensitive to temporal information (vowel duration) as they are sensitive to duration in their L1. The assimilation patterns were also similar across the two (high vs. low) vocabulary groups, however learners differed in their assimilation patterns, especially in the consistency of mapping L2 vowels to L1 categories.

Their findings show that larger vocabulary size facilitates learners' perception of L2 vowels. Overall, these results showed the influence of L1 and vocabulary size. These findings supported the PAM-L2 hypothesis that a larger vocabulary size *rephonologises* the L1 system, as learners integrate L2 phones into the existing L1 phonological system (Bundgaard-Nielsen et al., 2011). The results from this study also showed that, apart from consistency, the vocabulary size did not have a significant effect on the assimilation patterns. Thus, it is not clear if this means vocabulary size affects the perception of L2 vowels, or just refines the perception because they are advanced learners and/or have more exposure to the L2. The assimilation of Australian English /eɪ/ and /eː/ to Japanese /eɪ/ and /eː/ indicates their phonetic and phonological similarities in both languages.

#### 4.2.3.1 Uncategorised or Multiple Category Assimilation Patterns

Escudero and Boersma (2002) showed that Dutch learners of Spanish show *multiple category assimilation* (MCA) patterns, which is affected by the learners' *perception mode* (Escudero and Boersma 2001). The same set of vowels (/i/-/ɪ/) was perceived differently when presented as Spanish or Dutch vowels, and the assimilation/identification patterns were different when they were perceived as Spanish or Dutch vowels. As they predicted, various token of Spanish /e/ and /i/ were perceived as Dutch /ɛ/, /ɪ/ and /i/ when they were told that they were listening to Dutch vowels; however, listeners did not perceive /ɪ/ when they were told that they were listening to Spanish. The experience of L2 affected the overall perception resulting in fewer errors. They also reported that other than SC and TC, MCA also poses some problems for L2 learning and accurate assimilation/categorisation of L2 segments. They further proposed that the learners' *perception mode* (Escudero and Boersma 2001) can be helpful to dispose of the L1 categories that do not exist in L2, and as a result improve their perception and production of L2. This study is based on the perception of an L2 with fewer vowels than L1. They did not consider if MCA patterns can be observed if L2 and L1 have a similar number of vowels.

In a more recent study, Faris, Best and Tyler (2016) investigated the uncategorised assimilation patterns for Australian English vowels by Egyptian Arabic speakers in Egypt. They found evidence for *dispersed* (L2 phone is assimilated to a large number of L1 phones), *focalised* (L2 phone is assimilated to one L1 phone but below (50%) the threshold for it to be categorised); and *clustered* (L2 phone is assimilated to a small number of L1 phones) patterns of uncategorised assimilation. They reported that focalised and clustered patterns of uncategorised assimilation suggest that listeners were sensitive to the phonetic details that are distinct in the L1 phonology. However, in dispersed patterns of uncategorised assimilation, listeners paid attention only to phonetic details. Faris et al. (2016) further reported that the degree of perceptual overlap (partial or full overlap) can determine the difficulty in discrimination of L2 vowels. They additionally predict that the *focalised-focalised* uncategorised contrasts will be relatively easy to discriminate, whereas *disbursed-dispersed* contrasts will be the most difficult. However, they did not test these contrasts in this study. This was the first study to analyse and define the nature of uncategorised assimilation patterns. However, it is not clear if these uncategorised patterns will result in the formation of new categories for L2 and/or result in the expansion of the L1 phonological and phonetic system.

#### 4.2.3.2 Effects of Acoustic and Phonetic Similarities between L1 and L2

A number of studies have been conducted to investigate the effects of acoustic and perceptual similarities of cross-linguistic phonemes, i.e. vowels and consonants (Elvin et al., 2014; Escudero et al., 2014; Escudero and Williams, 2011; Escudero and Chládková, 2010; Gilichinskaya and Strange, 2010; Escudero, Benders, and Lipski, 2009; Strange et al., 2004; Strange et al., 2005; Escudero and Boersma, 2002). Although, according to Strange et al. (2004), acoustic similarity does not predict perceptual assimilation patterns, these studies suggest that acoustic and phonetic similarities between L1 and L2 can predict the patterns of perceptual assimilation by L2 learners.

In a recent study Elvin et al. (2014) and Escudero et al. (2014) reported that acoustic similarities between L1 and L2 are better predictors of both the assimilation patterns and discrimination difficulties faced by L2 learners. Elvin et al. (2014) investigated if the vowel inventory size and acoustic properties (i.e. similarities and differences) between Australian English (12 monophthongs), Iberian Spanish (5 monophthongs) and Brazilian Portuguese (7 monophthongs) can help to predict the difficulties faced by Australian and Iberian Spanish learners of Brazilian Portuguese. Their results showed that Iberian Spanish learners of Brazilian Portuguese outperformed Australian English learners of Brazilian Portuguese, which suggested that vowel inventory size was non-significant; however, acoustic similarities between Iberian Spanish and Brazilian Portuguese were good predictors.

Escudero et al. (2014) investigated the assimilation patterns of Southern British English (SBE) vowels by 12 male speakers of Salento Italian (SI) who learned English as a foreign language at school. Escudero et al. (2014) compared these assimilation patterns with Peruvian Spanish learners of Southern British English. Despite the fact that both Salento Italian and Peruvian Spanish have a five-vowel system, their assimilation patterns of

Southern British English vowels were quite different. Therefore, Escudero et al. (2014) concluded that cross-language acoustic similarities between L1 and L2 can predict the assimilation patterns of L2 learners as well the difficulties faced by them. For example, they reported that identical F2 values for SBE /ɪ/ and SI /i/ successfully predicted that SBE /ɪ/ was assimilated to SI /i/, and that similar F1 values for SBE /ɔː/ and SI /o/ predicted that SBE /ɔː/ was assimilated to SI /o/.

Research shows that acoustic similarities between L1 and L2 can help to predict the patterns of assimilation. For instance, Escudero and Williams (2011) investigated the perceptual assimilation patterns of Dutch vowels by naïve Spanish speakers (20 males and 20 females), and they found that there were more single-category (SC) assimilation patterns than two-category (TC). They concluded that the acoustic similarity between Dutch and Spanish vowels helped to predict the possible assimilation patterns. Due to there being far fewer Spanish vowels (5 monophthongs) as compared to Dutch (12 monophthongs) it was predicted that there will be more SC patterns. They did not test the category goodness rating to see if these patterns were representative of good or bad exemplars of L1.

Similarly, in an earlier study, Escudero and Chládková (2010) investigated Spanish listeners' (20 males and 20 females) perception of American English and Standard Southern British English vowels in a synthetic stimulus of 9 English monophthongs. As they predicted, Spanish listeners showed perceptual assimilation patterns based on the acoustic/spectral similarity between the vowels of Spanish and the particular variety of English, and thus their assimilation patterns for the two varieties of English differed. This study supported L2LP predictions. However, they did not investigate the effects of temporal differences between vowels across languages.

English tense and lax vowels have proven quite difficult for L2 learners of other languages, as shown in a number of studies (Flege et al., 1998; Flege et al., 1999; Baker et al., 2002; Escudero and Boersma, 2002; Escudero, 2005; Escudero and Chládková, 2010; Escudero et al., 2014). Gilichinskaya and Strange (2010) investigated the perceptual assimilation of 8 American English (AE) vowels /iː/, /ɪ/, /ɛ/, /æː/, /ɑː/, /ʌ/, /ʊ/ and /uː/ in disyllables /Vpə/ by 19 Russian speakers. Their results showed that Russian speakers assimilated AE vowels based on their acoustic similarity with the respective Russian vowels. However, the rest of the vowels were not assimilated consistently. Their findings also showed that /ɪ/-/ɛ/ and /ɑː/-/ʌ/ were the most difficult for Russian speakers. AE /ɑː/-/æː/ were categorised-uncategorised, whereas /iː/-/ɪ/ and /uː/-/ʊ/ showed two-catego*ry* assimilation patterns. Based on these patterns they predicted that English tense and lax vowels will be easier for Russian speakers to perceive and produce. This shows that L1 plays a significant role in the perception of L2 sounds.

Previous studies have also shown that with regard to the native language, both monolingual and bilingual learners vary in their reliance on spectral and temporal cues for the perception of L2 vowels. For example, Escudero, Benders, and Lipski (2009) investigated the use of spectral and temporal cues for the categorization of the Dutch /ɑ/-/aː/ contrast. The three groups of listeners: monolinguals; L1-Dutch (31), and L1-German (31), bilinguals; L1-Spanish and L2-Dutch (38), were presented with synthetic stimuli of the Dutch vowel contrast in an XAB task. Their findings showed that bilingual (L1-Spanish and L2-Dutch) listeners rely more heavily on temporal than spectral cues. On the other hand, though L1 German listeners showed more reliance on spectral than on temporal cues, L1-German listeners were not accurate in their categorisation of the Dutch contrast. Overall, L1-Spanish and L2-Dutch, with high or intermediate proficiency of Dutch, categorised the vowels correctly as compared to L1 German, who did not have any experience/knowledge of Dutch. Hence experience of L2 affects categorisation; at the same time, the role of L1 cannot be neglected since it alters the perception patterns, i.e. paying more attention to temporal (Spanish listeners) or spectral (Dutch and German listeners) cues.

#### 4.2.3.3 Effects of Experience and Consonantal Context

Strange et al. (2001) investigated the effects of consonantal context on the perception of 11 AE vowels by 24 Japanese listeners. 11 AE vowels in six syllabic contexts /b-b, b-p, d-d, d-t, g-g, g-k/ were embedded in carrier sentences produced by four AE speakers. Japanese listeners mapped these to 18 Japanese categories and rated the category goodness on a seven-point scale. Strange et al. (2001) found that patterns of spectral and temporal assimilation were affected by the consonantal context, which further showed the difficulty to identify and discriminate AE vowels by Japanese speakers. For example, the preceding voiced and voiceless consonant affected the perceptual assimilation of AE /iː/ and /uː/. In addition, the effects of context and speaker were also observed in the assimilation patterns of AE vowels /ɪ/, /ɛ/, /æː/, /ʌ/ and /ʊ/.

Levy and Strange (2008) investigated the effects of experience and consonantal contexts on the perception of French vowels /u/, /y/, /i/ and /œ/ by 20 native speakers of American English who were divided into two groups based on their experience of French, such as experienced vs. inexperienced. Experienced listeners studied French from the age of 13 and inexperienced listeners did not study/learn French at all. The French vowels were presented in disyllables in two contexts: /rabVp/ and /rabVt/. Their results from an AXB task showed that experienced AE listeners performed better for /i/-/y/, /u/-/œ/ and /y/-/œ/ vowel contrasts than inexperienced listeners. However, both groups did not differ on the discrimination of /u/-/y/. Overall, consonantal context did not show any effect on the discrimination of experienced listeners; however inexperienced listeners performed better in the bilabial context. In addition, the effects of consonantal context were quite prominent. The inexperienced group confused /i/-/y/ in the bilabial context and /u/-/y/ in the alveolar context; however experienced listeners confused /u/-/y/ in both contexts. Although they concluded that naïve listeners show different perceptual patterns in different consonantal context, for segments of an unfamiliar language, they did not investigate how these discrimination patterns can be assimilated to L1 vowels.

In addition to acoustic similarities and differences, exposure to L2, experience and vocabulary size, a more recent study has found that L2 proficiency plays an important role in the perception and production of L2 vowels and consonants. Evans and Alshangiti (2018) investigated the perception and production of British English consonants and vowels by native speakers of Arabic with varying fluency in English. Their results showed that participants with higher proficiency levels showed less difficulty with British English vowels compared to the participants with low proficiency in English. However, in both groups' participants found /ɪ/, /ɒ/ and /eə/ difficult and performed poorly on accurate identification. In addition, all participants found the following contrasts most confusing: /ɪ/-/e/, /ɜ/-/eə/, /uː/-/ʊ/ and /ɒ/-/ʌ/-/əʊ/. Although the participants in this group started to learn English in Saudi Arabia, most of them had exposure to native British English for three years, on average.

**4.3 Summary**

Previous research has shown that that the L1 vowel inventory, age of learning, experience and exposure to L2, and the acoustic and phonetic similarities between L1 and L2, affect the perception and identification/discrimination of L2 vowels. A number of theoretical models have been proposed and tested to investigate how L2 learners perceive and produce speech sounds (i.e. vowels and consonants): the Native Language Magnet model (Kuhl, 1993; 2000), the Speech Learning Model (Flege, 1995; 2002), the Perceptual

Assimilation Model (PAM; Best, 1994, 1995; PAM-L2; Best and Tyler, 2007), and Second language Linguistic Perception model (L2LP: Escudero, 2005).

According to Best and Strange (1997) and Flege et al. (1987; 1995), the perceived similarities between L1 and L2 segments affect the perception of L2 segments as *similar*, *identical* or *new*. Most recent studies support the second language Linguistic Perception model (L2LP: Escudero, 2005), which states that cross-language acoustic similarities between L1 and L2 vowels can help to predict the assimilation, categorisation and discrimination patterns of L2 vowels. A number of the above-mentioned studies have been conducted on listeners whose native language has fewer vowels (i.e. Spanish, Italian, Arabic, Russian, French) as compared to L2, and in most cases L2 was American English, British English and Australian English.

Investigation of the perception of English vowels by German, Norwegian, Spanish and French speakers showed that a larger L1 inventory makes it easier to perceive L2 vowels accurately. Studies on L2 perception and production further showed that learners from different linguistic backgrounds demonstrated similar patterns of assimilation and used spectral and temporal cues in their identification and assimilation of L2 vowels. Some studies reported that L2 vowel inventories that are smaller than L1 result in multiple category assimilation patterns (Escudero and Boersma, 2004), and L2 inventories that are larger than L1 result in uncategorised-uncategorised assimilation patterns (Faris et al., 2016). However, none of the aforementioned models were designed to specifically address the performance of early fluent multilingual learners of English, who have no exposure to native input except for media (TV and films). In the present study, the tests are mainly based on the predictions of SLM, PAM and PAM-L2. With regard to acoustic similarities between Urdu and SSBE, the predictions of L2LP are also tested.

### 4.4 Present Study

The present study investigated the perceptual assimilation of SSBE vowels by Punjabi Urdu speakers who learn English at school. Therefore, as opposed to previous studies (Strange et al., 2007; Gilichinskaya and Strange, 2010; Escudero and Chládková, 2010), the listeners were not naïve or inexperienced learners of SSBE; however, the English language they learn is Pakistani English (PE) spoken in Punjab, Pakistan (as discussed in Section 4.1.1), which has very little to do with SSBE. According to L2 speech learning theories, L2 learners face difficulties in the perception and production of L2 segments (i.e. vowels and consonants) which are either new (do not occur in their L1) or are very close (phonetically) to L1 segments. This study intended to investigate if the predictions of PAM, L2LP and SLM are applicable to the perception of SSBE vowels where L2 (English) is learnt from a very young age in a non-native context and used as lingua franca in everyday life. Hence the L2 users are not inexperienced learners.

In summary, very little is known about how native Punjabi-Urdu speakers perceive and process Standard Southern British English (SSBE) vowels, and this study hopes to begin to shed light on this research question. This will be the first study of its kind to investigate the perception of SSBE vowels by native speakers of Punjabi-Urdu. In Pakistan, almost everyone is multilingual, and at least bilingual. Previous studies on multilingual and bilingual learners' acquisition of a third language (L3/L*n*) report that although both L1 and L2 affect the acquisition of L3, it is mainly the dominant language (L2) that interferes with the learning of L3 (Hammarberg and Hammarberg, 2009; Llama et al., 2010; Hammarberg, 2014; Lipińska, 2017). Therefore, the present study was designed to investigate the perception of SSBE vowels after the investigation of the Urdu vowel system as spoken in Punjab, Pakistan (i.e. the Urdu spoken by Punjabi L1 speakers). Based on the second language perception models and literature review, the predictions for the present study were as follows:

### 4.4.1 General Predictions: According to SLM

- Due to *equivalence classification* (Flege, 1995), Punjabi-Urdu speakers will not establish separate categories for English monophthongs that are found in the same phonological/acoustic space as a monophthong in Urdu, for example: /iː/, /ɪ/, /ɛ/, /ʌ/ and /ɒ/. As a result, the speakers' production of such monophthongs will not be very accurate because they will be collapsed with the similar L1 segments.
- Punjabi-Urdu listeners will be able to establish new phonetic categories for English monophthongs and diphthongs that do not have a counterpart in Urdu phonological and phonetic system, for example: /æ/, /ɜː/, /ɔː/, /əʊ/ and /eɪ/.
- Punjabi-Urdu listeners will be able to discern the phonetic differences between English and Urdu monophthongs and diphthongs, and will be able to establish new phonetic categories for them.

### 4.4.2 General Predictions: According to PAM and L2LP

Based on an acoustic comparison of SSBE and Punjabi-Urdu vowel system, we can predict the following patterns of assimilation:

- Given the number of Urdu vowels there will be more patterns for *two-category* (TC) assimilation of SSBE vowels.
- Due to the cross-language acoustic/perceived similarity (Escudero and Boersma, 2005; Flege, Munro and Fox, 1994) there will be some *single-category* (SC) assimilation patterns of SSBE /ɛ/-/æ/, /ɜː/-/ʌ/ and /ɔː/-/ɒ/, since there is only one vowel in the Urdu vowel inventory in the corresponding vowel space, i.e. open-mid front, central, and open-mid back, respectively.
- There will be fewer *category-goodness* (CG) assimilation patterns.
- There will be more *uncategorised-uncategorised* (UU) assimilation patterns, especially for English front vowels /ɛ/, /æ/, back vowels /ɔː/, /ɒ/ and diphthongs.

### 4.4.3 Broad Research Questions

- Which dimensions Punjabi-Urdu listeners use to classify the 19 SSBE vowels?
- How do Punjabi-Urdu listeners perceptually assimilate the 19 SSBE vowels to the Urdu vowels?
- How do Punjabi-Urdu listeners perceptually assimilate the long and short SSBE vowels to Urdu long and short vowels?
- To what extent do the perceptual assimilation patterns differ for the 19 SSBE vowels produced and presented in two different contexts, i.e. disyllabic *hVba* vs. monosyllabic *bVd*?
- To what extent do the perceptual assimilation and free classification patterns differ for the 19 SSBE vowels produced and presented in different contexts, i.e. disyllabic *hVba* vs. monosyllabic *bVd* and *hVd*?

In order to answer these questions, a cross-language perceptual assimilation experiment and a free classification experiment was conducted. Chapter 5 reports on the perceptual assimilation of SSBE vowels. Chapter 6 reports on the free classification of the same 19 SSBE vowels. The reasons for selecting two different types of experiments and three different contexts are further discussed in the following chapters (Chapters 5 and 6).

## Chapter 5

### Perceptual Assimilation of SSBE vowels

This chapter reports on an experiment investigating the perception of Standard Southern British English (SSBE) vowels by Punjabi-Urdu speakers from Punjab, Pakistan. As discussed in Section 1.3 and 4.1, Received Pronunciation (RP) is the standard variety of English taught in Pakistan. In the present study, RP vowel symbols are used as reported by Hawkins and Midgley (2005), however, the stimuli were recorded from SSBE speakers as discussed in Section 5.2.2. The Cross-language mapping and category goodness rating task is reported in the literature as the most reliable method of investigating the perceived relation between L1 and L2 vowels (Best, 1995; Flege, Bohn and Jang, 1997; Ingram and Park, 1997; Schmidt, 1996). In a cross-language mapping task, L2 learners who are phonetically untrained, assimilate multiple natural tokens of L2 vowels to the given L1 categories. After the mapping of L2 vowels to the given L1 categories, they rate the category goodness of the L2 vowels on how similar or dissimilar the L2 vowel is from those in L1. These goodness ratings (usually from 1-7, Strange et al., 1998) then show which L2 vowels were considered good or bad examples for L1 vowels. If multiple vowels are mapped to a certain L1 category, it is labelled as single-category assimilation, and category goodness ratings show which one of those L2 vowels were perceived as the best example of or closest to the L1 category. If two L2 vowels are mapped to two separate L1 categories, those are labelled as two-category assimilation (Flege, 1995; Best, 1995; Flege et al., 1997; Ingram and Park, 1997; Best and Tyler, 2007).

Given the status and role of English in Pakistan (as discussed Chapters 1 and 4), this study investigated if the predictions of SLM and PAM models (see Chapter 4) are applicable to the perception of SSBE vowels, where L2 (English) is learnt from a very young age in a non-native context and used as lingua franca in everyday life. Hence the L2 speakers are not inexperienced learners. Since there is no literature investigating the perception of SSBE vowels by Punjabi-Urdu speakers, this study employed perceptual assimilation and goodness rating task for the SSBE vowels (11 monophthongs and 8 diphthongs). Such a cross-language mapping task is ideal for examining the perceptual phonetic distance between L1 phonemes and L2 phones (Best, Faber and Levitt, 1996) and predicting the difficulties in the discrimination of L2 phones. The main objective of this experiment was to address the following research questions:

a) How do Punjabi-Urdu speakers perceptually assimilate the SSBE vowels; i.e. is it spectral or temporal information they are sensitive to, or both?
b) What are the most confusing English vowels for Punjabi-Urdu listeners, which lead to either mispronunciation or a strong accent while speaking English?
c) Which English vowels are perceptually assimilated with which Urdu vowels?
d) To what extent perceptual assimilation is influenced by a context *bVd* familiar to the participants as a word of English (e.g. *bud*) vs. an unfamiliar context *hVba* (e.g. *huba*)?

### 5.1 Predictions

According to SLM, PAM and L2LP, the phonetic resemblance between L1 and L2 plays an important role in the perception and production of L2 sounds. Therefore, the predictions for the present study are based on the acoustic comparison of Urdu vowels (as reported in Chapter 2 and Chapter 3) with SSBE vowels as reported by Hawkins and Midgley (2005). They reported formant frequencies of received pronunciation (RP) monophthongs by four different age groups. Figure 5. 1 shows the mean frequencies of the first two formants in F1/F2 vowel space for SSBE vowels (Hawkins and Midgley, 2005) and Urdu vowels.

### 5.1.1 Acoustic Similarities between RP and Urdu (Predictions according to L2LP)

According to L2LP (as discussed in Section 4.3), based on the acoustic similarities and differences between Urdu and SSBE, the following predictions can be made. From the visual inspection and acoustic measurements, it can be predicted that SSBE /ɪ/ may be confused with Urdu /ɪ/ and /eː/, and SSBE /ɛ/ will be assimilated to Urdu /ɛː/. F1 for SSBE /ɛ/ is higher than Urdu /ɛː/; however, F2 for SSBE /æ/ is lower than Urdu /ɛː/. The F2 of English /æ/ suggests that it may be assimilated to Urdu /ɑː/.

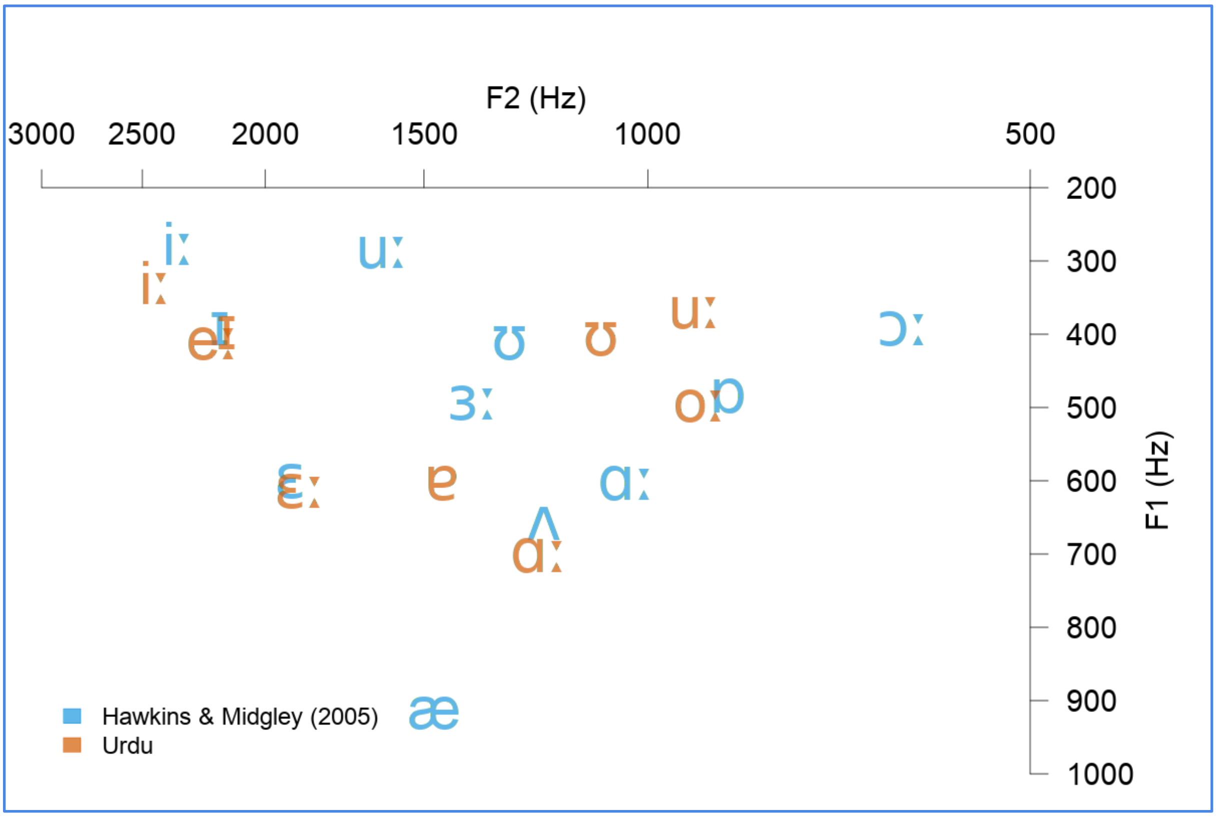



Figure 5. 1: Mean frequencies of the 11 RP monophthongs as reported by Hawkins and Midgley (2005) and 9 Urdu monophthongs on F1/F2 vowel space

Further, SSBE /ʌ/ appears to be acoustically similar to Urdu /ɑː/. Therefore, it can be predicted that SSBE /ʌ/ will be mapped to acoustically similar Urdu /ɑː/. SSBE /ɜː/

appears to be quite different from Urdu /ɐ/. F1 and F2 of SSBE /ɜː/ are lower than Urdu /ɐ/. The perceptual assimilation of this central vowel cannot be predicted based on the acoustic measurements. However, it seems that Punjabi-Urdu speakers will find this vowel difficult to assimilate to the Urdu vowel categories.

Urdu /oː/ lies somewhere in the middle of SSBE /ɔː/ and /ɒ/. Hence it can be predicted that those two back SSBE vowels will be assimilated to this one Urdu back vowel. SSBE /uː/ and /ʊ/ are quite front in the vowel space as compared to Urdu /uː/ and /ʊ/. With regard to F1 it can be predicted that English /uː/ and /ʊ/ will be assimilated to Urdu /uː/ and /ʊ/; however, F2 of English /uː/ is much higher than Urdu /uː/. Therefore, it is hard to predict whether Punjabi-Urdu speakers can identify these two vowels as distinct or not.

### 5.1.2 Predictions: According to PAM

As discussed in Section 4.4.2, according to PAM, it can be predicted that there will be some *Two-Category* (TC) and more *Single-Category* (SC) (Best, 1995; Best and Tyler, 2007) assimilation patterns for SSBE vowels by Punjabi-Urdu speakers. For example, as noted in Section 4.4.2, there will be *Single-Category* (SC) assimilation patterns, especially for SSBE /ɛ/-/æ/, /ɜː/-/ʌ/ and /ɔː/-/ɒ/, as there is only one vowel in the Urdu vowel inventory in the corresponding vowel space, i.e. open-mid front, central, and open-mid back. In addition, there will be fewer *Category-Goodness* (CG) assimilation patterns, and there will be more *Uncategorised-Uncategorised* (UU) assimilation patterns, especially for English front vowels /ɛ/ and /æ/, back vowels /ɔː/ and /ɒ/, and diphthongs.

### 5.1.3 Predictions: According to SLM

As discussed in Section 4.4.1, according to the SLM *equivalence classification* hypothesis, Punjabi-Urdu speakers will not establish separate categories for English monophthongs that are found in the same phonological/acoustic space in Urdu, i.e. /iː/,

/ɪ/, /ɛ/, /ʌ/, /ɑː/, /ɒ/ and /ʊ/. As a result, the speakers' production of those vowels will not be very accurate because the vowels will be collapsed with similar L1 phonemes. However, with experience and exposure, Punjabi-Urdu listeners will be able to discern the phonetic differences between English and Urdu monophthongs and diphthongs, and will be able to establish new phonetic categories for them. Lastly, Punjabi-Urdu listeners will be able to establish new phonetic categories for English monophthongs and diphthongs that do not have a counterpart in Urdu phonological and phonetic system, i.e. /æ/, /ɜː/, /ɔː/, /uː/ and diphthongs.

This chapter presents the perceptual assimilation experiment in two contexts, *bVd* and *hVba* (for an explanation, see below). Auditory free classification in a third context, *hVd*, is presented in Chapter 6. Previous research shows that the consonantal context has an effect on the quality of vowels (Hillenbrand et al., 2001), and variations in the phonetic context due to the neighbouring consonantal context also affects cross-language perceptual assimilation patterns (Strange et al., 2004; Bohn and Steinlen, 2003; Strange et al., 2001). Three different consonantal contexts were chosen for the present study.

Monosyllabic words in bVd and hVd contexts were chosen for two reasons. Firstly, the vowels in these two contexts are not affected by the neighbouring consonants, and as a result the coarticulation effects are factored out (Ferrag and Pellegrino, 2010; Hawkins and Midgley, 2005; Deterding, 1997; Wells, 1982). Secondly, to test if the familiarity with the target language affects listeners' perception. In addition, another reason to embed SSBE vowels in a bVd context was to make them appear very different than those in a hVba context, which aids in the cross-language assimilation task's conceit that vowels in the hVba context are from a foreign language, whereas bVd context vowels are from SSBE.

Nonce disyllabic words using hVba as a context were created to test if the (perceived) lack of familiarity with the target language affects listeners' perception, and if listening to vowels in an unfamiliar context results in better identification of phonological and phonetic details and hence better perception. Best et al. (1988) reported that American English monolinguals identified the isiZulu click consonants as nonspeech because clicks are not found in American English; however, they discriminated the minimal contrasts very well. Therefore, following Strange et al. (1998) the 19 SSBE vowels are embedded in nonsense disyllabic words so that listeners could attend to the vowels without any perceptual effects due to experience of English.

## 5.2 Experiment Design

### 5.2.1 Participants

In the perceptual assimilation and goodness rating experiments, 70 (24 male and 46 female) listeners from Lahore participated. Forty-six undergraduate students in their second and third year, and 24 MA students took part in the experiment. They were studying either in the department of English Language and Literature or Mass Communication at the University of the Punjab, Lahore, Pakistan. Their age ranged from 18 to 25. In order to obtain information about participants' linguistic and social background as well as proficiency in English language (independent of their perceptual abilities), two online surveys were conducted via Google forms (Rehman, 2016).

Prior to the listening tasks participants were sent emails containing links to two online forms that were prepared with Google forms. Firstly, participants were given a test to assess their level of English – the English Language Proficiency Test (ELP-T). This was adapted from *Cambridge Assessment English* (CAE, 2016) C1 advanced level *reading and use of English test*. Secondly, participants were given the Language Experience and Proficiency Questionnaire (LEAP-Q) adapted from Marian, Blumenfeld and

Kaushanskaya (2007) to get their essential linguistic and social background. Participants filled in the forms online in English. These tests were prepared to assess participants' proficiency in English language and educational, linguistic, social and regional background. The scores from ELP-T ranged from 11% to 87% with a median of 33% and a mean of 38%. These scores showed that participants varied in their ability, from lower intermediate to advanced level of competence in English.

LEAP-Q had four main sections: *Personal information*, *Linguistic background*, *English language acquisition*, and *Social background.* According to LEAP-Q, 56 (76%) of participants speak 3 languages. Further, 41 (55.4%) participants reported Urdu as their first language and 28 (37.8%) reported Punjabi as their first language. Among all the participants, 21.6% reported Punjabi as their second language, 37.8% reported Urdu as their second language, and 39.2% reported English as their second language. Most participants reported that they use Urdu (55.4%) and Punjabi (37.8%) at home and Urdu (74.3%) and English (20.3%) for communication with friends at school, college and university.

Most participants reported that they began to learn English from primary school (51.4%), some at the age of 4 (21.6%) and a few began to learn English in middle school (12.2%). The majority of the participants attended private schools (73%). In their respective primary schools, 47.3% reported the medium of instruction as English and 50% reported the medium of instruction as Urdu. With regard to their proficiency in speaking English (as shown in Figure 5. 2) and understanding spoken English, more than 70% of participants reported an intermediate to high level (5-7) on the scale, 16% participants reported a higher level (8-10), and less than 10% reported a low proficiency level.

More than 50% of participants reported an extensive exposure to English language and culture through TV, radio, social media and social interactions, readings and English language courses. Almost 60% of participants reported a higher percentage of foreign accent in their spoken English. More than 60% of participants reported that they grew up in urban (i.e. socially and economically developed) towns. In addition, none of the participants have been abroad except two, a male who had visited China for an arts festival, and a female who spent her summer holidays in the Middle East. Three participants reported hearing impairment, and data from those participants was not included in the final analysis. The overall responses to each section of LEAP-Q are given in Appendix 5A.

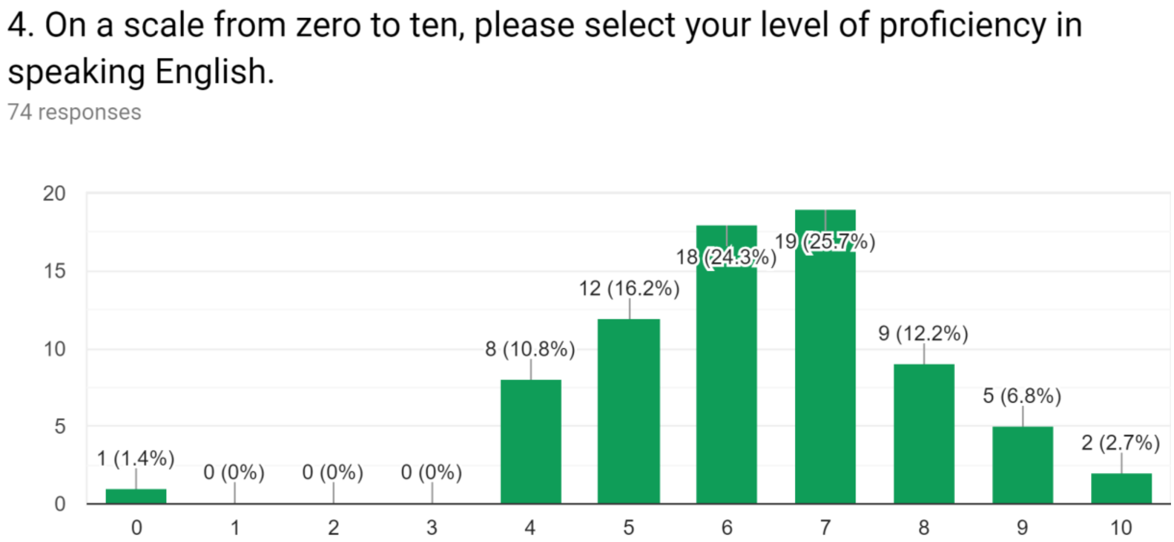



Figure 5. 2: Participants' self-reported level of proficiency in speaking English (from LEAP-Q)

As the participants took part in more than one experiment (see Chapter 6), they each attended two sessions, on two different days. Each participant was given a reward in cash (equivalent to 7.50 GBP) in Pakistani rupees for their participation. Experiments were conducted at the University of the Punjab, Lahore, Pakistan from 29th August 2016 to 20th September 2016.

### 5.2.2 Stimuli

The perceptual assimilation and category goodness experiment was based on 19 SSBE vowels, 11 monophthongs /iː/, /ɪ/, /ɛ/, /æ/, /ɜː/, /ʌ/, /ɑː/, /ɒ/, /ɔː/, /ʊ/ and /uː/ and 8 diphthongs /ɪə/, /eə/, /ʊə/, /eɪ/, /aɪ/, /ɔɪ/, /əʊ/ and /aʊ/, in the contexts bVd and hVba. That is, English minimal pairs were constructed in two contexts: bVd (19 vowels × 2 speakers × 3 repetitions = 114) + hVba (19 vowels × 2 speakers × 3 repetitions = 114). Table 5. 1 lists the English words and nonsense words that were used for the cross-language assimilation and goodness rating task.

Table 5. 1: Test words in bVd and hVba context

| **Test words in bVd context** | **Test words in hVba context** | **Lexical Set** |
|---|---|---|
| bead | heba | FLEECE |
| bid | hiba | KIT |
| bed | heba | DRESS |
| bad | haba | TRAP |
| bard | harba | START |
| bod(y) | hoba | LOT |
| bawd | horba | FORCE |
| budd(hist) | hooba | FOOT |
| booed | who’ba | GOOSE |
| bud | huba | STRUT |
| bird | hurba | NURSE |
| bayed | haba | FACE |
| bide | hiba | PRICE |
| boyd | hoiba | CHOICE |
| bode | hoeba | GOAT |
| bowed | howba | MOUTH |
| bared | hareba | SQUARE |
| beard | heerba | NEAR |
| boored | hureba | CURE |

Two talkers (a male in his late 20s and a female in her late 30s), who spoke Standard Southern British English as a native language, recorded the set of English and nonsense words. These words were shown on a PowerPoint slide one at a time, and talkers could see the lexical sets parallel to the target word for ease of pronunciation. The stimuli were produced in citation form, and the talkers were instructed to pronounce these words in a normal tone, and to stress both syllables in the disyllabic words of the hVba context (i.e. the final vowel was pronounced as /ɑ/ rather than unstressed schwa /ə/).

Following Strange et al. (1998), a CVCV structure for nonsense words, with the final syllable stressed, were used so that the utterance conformed to the Urdu CV phonotactic structure, where word final mono-moraic syllables are not allowed (see Section 3.1.1) for Urdu syllable structure and constraints). As a result, it was easier to tell participants that the final syllable was /bɑ/ in all the words, so they should focus on the vowel in the first syllable. In addition, having a full vowel at the end of each word ensured that participants did not suspect that these vowels were from English. Each speaker produced three randomized blocks of stimuli. They were asked to repeat an utterance immediately if they mispronounce it (in their own or the experimenter's judgment). The total number of tokens across two contexts was 228 (2 talkers × 19 vowels × 2 context types × 3 repetitions).

The test words were recorded in .wav format using a Zoom Handy Recorder H4N with sample settings of 45.1 kHz digitization and 16-bit quantization. In order to keep the data anonymous, the recordings were coded such that no personally identifiable information was given. Data were stored on a personal laptop, as well as off site in private cloud storage (Dropbox and Mega) and secure University of Kent server.

SSBE vowels in the present stimuli were compared with the RP monophthongs reported by Hawkins and Midgley (2005). This comparison was considered necessary to ensure that talkers in the present study produced these vowels in the standard dialect (SSBE). This standard dialect (as discussed in Section 4.1) was set as a requirement for the current perception experiments by the author.

### 5.2.3 Acoustic Analysis of Stimuli

Acoustic measurements were made using Praat. In order to insert the intervals, the following spectrogram settings were used:

- Female speaker: View range (Hz): 5500.0, Dynamic range (dB): 70.0, Maximum formant (Hz): 5500, Maximum number of formants: 4.
- Male speaker: View range (Hz): 5000.0, Dynamic range (dB): 70.0, Maximum formant (Hz): 5000, Maximum number of formants: 5.

Unless specified otherwise, all settings were left as the default for Praat version 6.0.15. Praat scripts (see Appendix 5B) were used to measure the duration of the monophthongs and the frequency of the first, second and third formant of the monophthongs in three positions: at vowel onset (+10ms), at the middle of the vowel, and at vowel offset (-10ms). For diphthongs, the mean frequencies of the first two formants were measured at the seven equidistant points in time i.e. 20%, 30%, … 80%.

Mean acoustic measurements across the six tokens of each monophthong in bVd and hVba contexts are given in Appendix 5C. Figure 5. 3 shows the mean frequencies of the first two formants in F1/F2 vowel space for SSBE vowels (produced by a male and female native speaker in a bVd and hVba context in citation form) and RP vowels produced by a group of male speakers (20-25 years old) as reported in Hawkins and Midgley’s (2005). Hawkins and Midgley (2005) reported formant frequencies of RP monophthongs by four different age groups, i.e. 20-25, 35-40, 50-55 and 65+ years. The vowels produced by the speakers used for the present study are most acoustically similar to their youngest group (i.e. 20-25). The mean frequencies of the first two formants of each vowel are presented in Figure 5. 3.

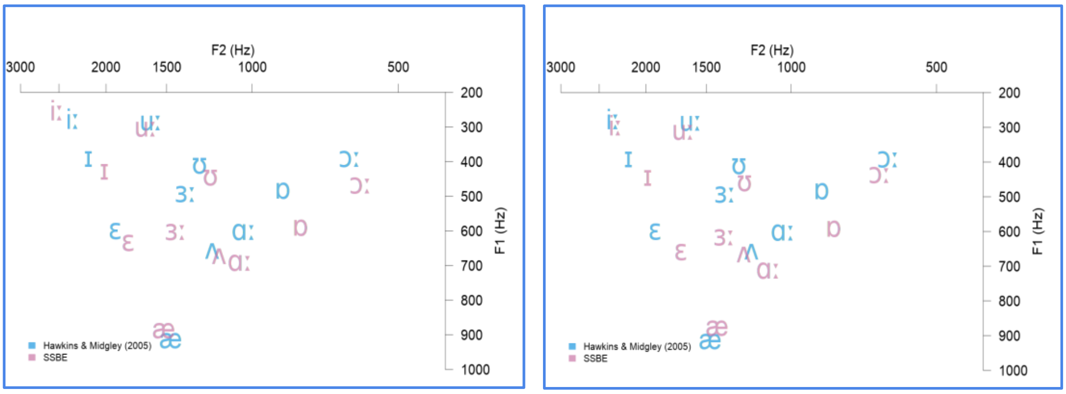



Figure 5. 3: Mean F1/F2 frequencies of the 11 SSBE vowels in bVd (left) and hVba (right) contexts, superimposed with RP monophthongs from Hawkins and Midgley's (2005) youngest age group.

The stimuli used for the cross-language assimilation experiment are very similar to the RP monophthongs reported by Hawkins and Midgley (2005) as shown in Figure 5.3. The stimuli in both contexts are comparable to other studies on SSBE vowels (Deterding, 2007; Ferragne and Pellegrino, 2010; Escudero et al., 2014).

## 5.3. Procedures

### 5.3.1 Listening task

Cross-language perceptual assimilation and goodness rating tasks were conducted, using the Multiple Forced Choice (MFC) experiment function in Praat (Boersma and Weenink, 2016) for stimulus presentation and response acquisition, on desktop computers. For each of the 19 SSBE vowels, 23 Urdu words were displayed on the computer screen as potential responses. These response alternatives were selected based on a preliminary acoustic and phonetic analysis of Urdu vowels (both monophthongs and diphthongs) as presented in Chapters 2 and 3. The 23 response alternatives were monosyllabic Urdu words, written in Perso-Arabic script in Nastalique style, containing the nine monophthongs and six diphthongs. The 23 Urdu words covered the 15 Urdu vowels. An open syllabic (CVV) word was used for the Urdu vowel category /oː/, /ɛː/, and all diphthongs. The remaining Urdu words were monosyllabic in closed syllables, CVC or CVVC. Keeping in view the issue of homographic ambiguity that was observed in the production experiment (as discussed on page 29 in Section 2.3.1.2), there were additional words for some Urdu vowel categories. There were three words, [bɐd̪] "*bad*", [bɐrf] "*ice"* and [kɐrz] "*debt",* for the Urdu vowel /ɐ/. There were two words for diphthong /oe/, [koe] "*any*" and [boe] "*to sow*"; two words for monophthong /uː/, [kuːd̪] "*jump*" and [puːt̪] "*offspring*"; and two words for /eː/ [peːʈ] "*stomach*" and [bʰeːd̪] "*secret*".

In addition to 23 response alternatives, "*none*" was also given as an option. This option was provided for determining whether any of the English vowels were judged to fall outside the bounds of the Urdu vowel system (cf. Butcher, 1976 cited in Flege, 1991:705). Butcher (1976) gave this option to French and English listeners to identify a set of cardinal vowels that were different from their L1 vowels. His findings showed that English speakers used the option "none" more often than French speakers, especially for front rounded and nasalised vowels that are not found in English. The list of response categories for 15 Urdu vowels is given in Table 5. 2. Note the vowels with more than one response category are in bold.

All participants were tested in a quiet, sound-attenuated room in the Department of English Language and Literature and in a computer lab in the Institute of Communication Studies at the University of the Punjab, Lahore, Pakistan. The participants were given access to a computer (laptop) and Sennheiser HD 650 headphones and were asked to listen to a set of English sounds, choose one of the given Urdu words that may contain that sound, and rate the Urdu word's category goodness (perceived goodness) on a 7-point scale, from 1, not Urdu like, to 7, Urdu like (Strange et al., 1998; Bundgaard-Nielsen, Best and Tyler, 2011; Tyler et al., 2014; Faris, Best and Tyler, 2016). Stimuli were presented at the participant’s self-selected comfortable listening level (which was then maintained throughout the experiment).

Table 5. 2: Urdu vowels and response categories for the perceptual assimilation

| Urdu vowel | Urdu Response categories and *glosses* | | | | | |
|---|---|---|---|---|---|---|
| iː | dʒiːt̪ جیت | *victory* | | | | |
| ɪ | zɪd̪ ضد | *stubborn* | | | | |
| **e**ː | bʰeːd بھید | *secret* | peːʈ پیٹ | *stomach* | | |
| **ɛ**ː | kɛːd̪ قید | *prison* | sɛː سہ | *tolerate* | | |
| ɑː | bɑːd̪ بعد | *after* | | | | |
| **o**ː | roːk روک | *stop* | d̪oː دو | *two* | poːd̪ پود | *plant* |
| ʊ | bʊd̪ʰ بدھ | *Wednesday* | | | | |
| **u**ː | kuːd̪ کود | *jump* | puːt̪ پوت | *offspring* | | |
| **ɐ** | bɐd̪ بد | *bad* | bɐrf برف | *ice* | kɐrz قرض | *debt* |
| ɪɐ | dʒɪɐ جیا | *lived* | | | | |
| eɐ | geɐ گیا | *went* | | | | |
| ʊɑ | d̪ʊɑ دعا | *prayer* | | | | |
| ɑe | pɑe پائے | *gained* | | | | |
| **oe** | boe بوئے | *sowed* | koe کوئی | *any* | | |
| ɑʊ | pɑʊ پاؤ | *gain* | | | | |
| None | کوئی بھی نہیں | *None* | | | | |

The participants were tested individually. A trial session to familiarise the participants with the task consisted of nine random stimuli in succession.  Before the trial session, the participants were asked to read all the Urdu key words aloud to make sure that they had the appropriate vowels in mind whilst completing the perceptual assimilation task. They heard each English vowel over headphones and were asked to judge to which Urdu vowel it was most similar by selecting one of the 23 response alternatives presented on a computer screen, or “none” if they thought the vowel they just heard did not match with any Urdu vowel, as shown in Figure 5. 4 and Figure 5. 5.

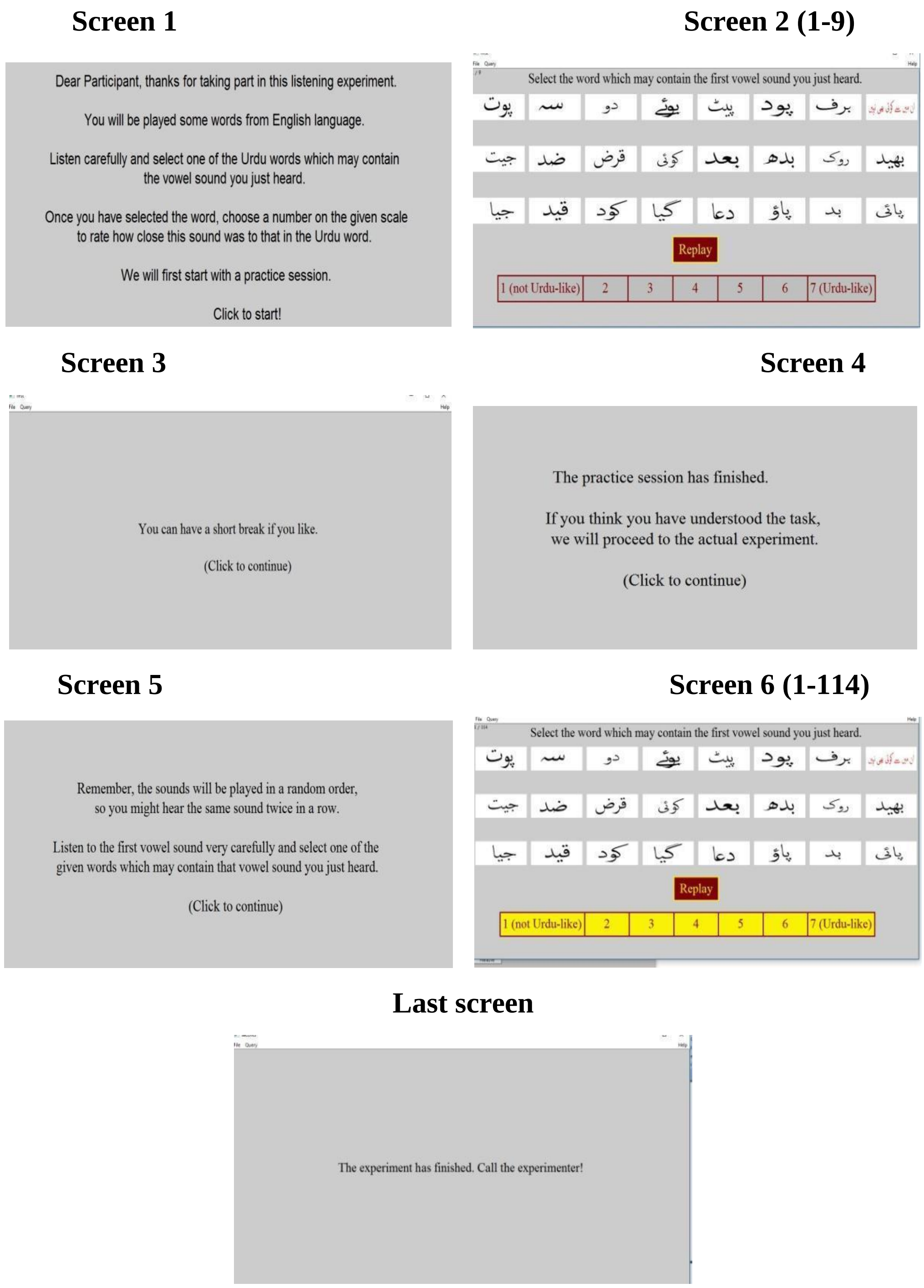



Figure 5. 4: Screenshots from the cross-language perceptual assimilation experiment (bVd)

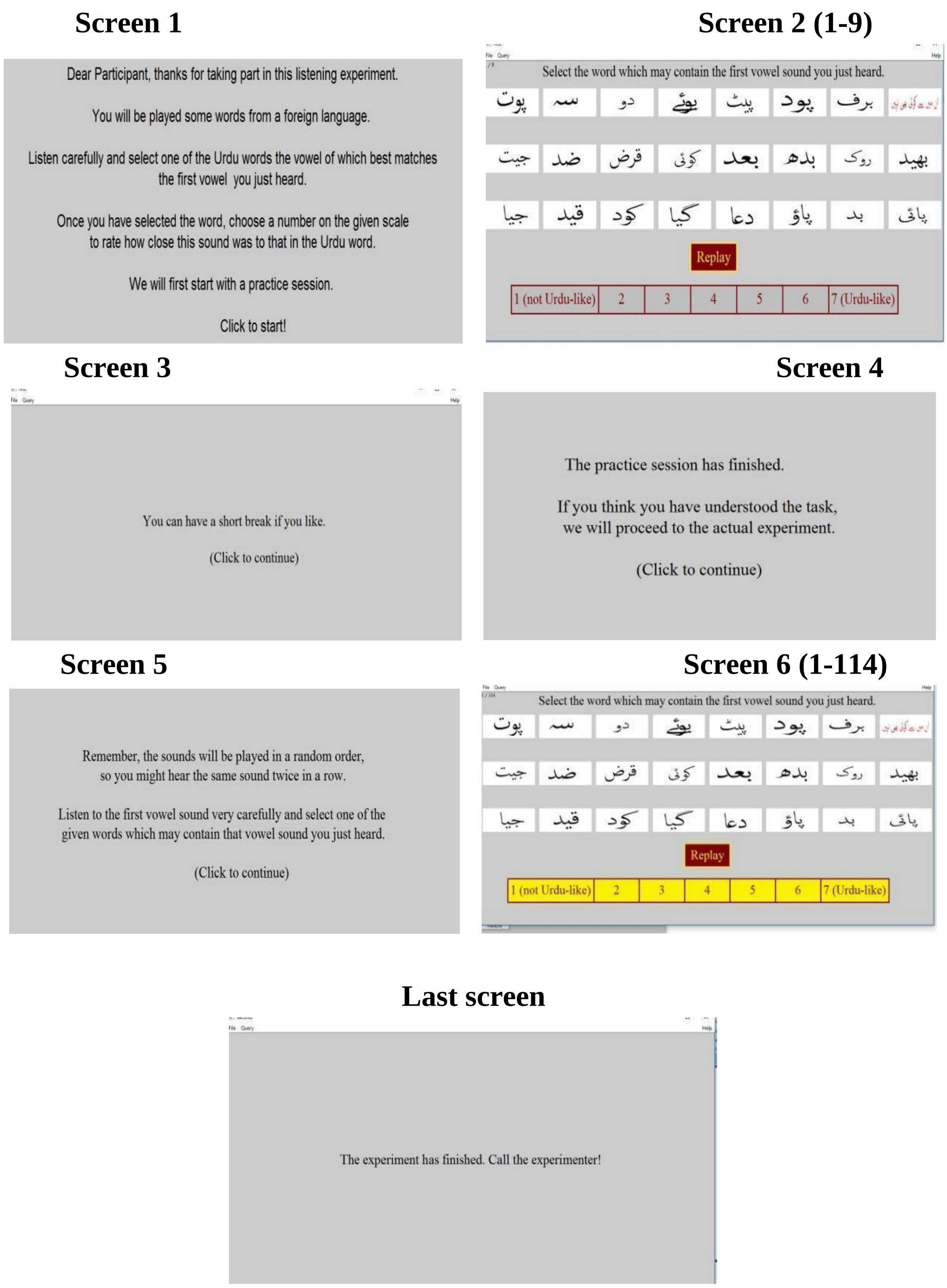



Figure 5. 5: Screenshots from the cross-language perceptual assimilation experiment (hVba)

Note that the instructions in these figures are slightly different (i.e. Screen 1) because in Figure 5.4 the stimuli are presented as the English sounds and in Figure 5.5 the stimuli are presented as words from an unknown foreign language.

Following the participant’s response, the participant then rated its ‘‘goodness’’ as an instance of the chosen response alternative on a scale from 1 to 7, and the endpoints were labelled as ‘‘Urdu-like’’ (7) and ‘‘not Urdu-like’’ (1) (Strange et al., 1998). Once a response was submitted for categorisation, they were not allowed to change it. The next stimulus was presented after the rating response was completed. Thus, all testing was participant-paced. Each stimulus was played after a 0.5 second silence followed by a 0.5-second-long beep followed by 0.5 second silence. Participants could replay the sound up to two more times if they wished. They were also prompted to take a short break after every 19 stimuli.

There were two testing sessions, one for each context, bVd and hVba. The order of presentation was counterbalanced across participants, and sessions were conducted on two separate days to minimize the risk of participants becoming aware that both contexts contain target vowels from the same language (i.e. English). In informal questioning after each experiment only one participant reported that the words in both contexts contained English vowels. Within a session the stimuli were fully-randomised, such that tokens from different vowels, speakers and repetitions were presented in random order.

For one session, bVd participants completed a total of six judgments on each vowel (2 talkers × 3 repetitions) for a total of 114 stimuli (2 talkers × 3 repetitions × 19 vowels). Similarly, for the other session, hVba participants completed a total of six judgments on each vowel (2 talkers × 3 repetitions) for a total of 114 stimuli (2 talkers × 3 repetitions × 19 vowels). Total testing time for each session ranged from 45 minutes to 2 hours, with short breaks, if desired.

## 5.4 Data Analysis

### 5.4.1 Assimilation Criterion for Categorisation

The fraction of opportunities where a given SSBE vowel is assimilated to a particular Urdu vowel used as the criterion for an SSBE vowel to be considered categorised was defined as 70% or above (cf. Tyler et al., 2014). In previous studies different criteria have been employed depending on the number of response categories or complexity of the task (Faris et al., 2016; Antoniou, Tyler and Best, 2012). A 70% criterion is neither a lenient 50%, nor stringent 90% that expects native like performance and in some cases might be too difficult even for native speakers (Bundgaard-Nielsen et al, 2011; Bundgaard-Nielsen et al., 2008). Given the nature of the task and response categories it was decided that a 70% criterion for bVd and a 50% criterion for hVba will be appropriate for the assimilation and categorisation of the 19 SSBE vowels (Faris et al., 2016).

The number of times each Urdu vowel was selected and the median ratings they were given was computed for each SSBE vowel and converted to percentages of total opportunities. The percentages of the first three (or five, for some of the stimuli in hVba context) most frequent responses in the bVd and hVba contexts are given in Table 5.6 and Table 5.7, respectively. The median goodness rating summed and averaged across all listeners for that stimulus is also given in parentheses.

Categorised responses were defined as those where participants selected a particular Urdu vowel in response to a given SSBE vowel more frequently than chance. Uncategorised responses were defined as those where participants selected multiple Urdu response vowels more often than chance. In order to differentiate categorised and uncategorized responses, t-tests were conducted comparing the mean percent categorisation of an SSBE vowel with each Urdu vowel response option against a chance score of 7%, a value that takes into account the 15 possible Urdu vowel categories.

### 5.4.2 Statistical Analysis

In order to investigate the effects of proficiency on the perception of SSBE vowels, the listeners were divided into two groups, i.e. high proficiency (advanced learners) and low proficiency (non-advanced learners).

The data were divided into two groups based on self-rating proficiencies (please see Appendix 5A), i.e. the median self-ratings ranged from 2.5 to 10 across participants, with an overall median rating of 7. Based on this median rating, participants were assigned to either the *advanced group*, if their median rating was higher than or equal to the median (N=34), or to the *intermediate group*, if their rating was lower than the median (N=32). Data from four of the 70 participants were not included in the analysis. Two participants did not complete the online questionnaire, and two participants' responses were lost due to hardware failure. Hence the results are based on the data collected from 66 participants.

In order to analyse the effects of context and proficiency on the assimilation of SSBE vowels by Punjabi-Urdu speakers, the assimilation frequencies and goodness ratings were combined and a fit index was calculated for each of the 19 SSBE vowels to the Urdu vowels by multiplying the proportion of the assimilation frequency and mean goodness ratings (Guion et al. 2000; Iverson and Evans 2007).

Further, the fit indexes calculated for the 19 SSBE vowels were analysed statistically. The methods for statistical analysis were similar to the one discussed in Section 2.4.4.

## 5.5 Results

For the mean percent categorization of a given SSBE vowel to an Urdu response vowel, a significant p-value ($p < 0.001$) indicates that a specific Urdu vowel was selected significantly more often than chance.

Table 5.3 and Table 5.4 show the Urdu vowels which were perceived to be the closest to the English vowels in bVd and hVba context respectively. The mean assimilation frequencies above 30% are presented in these tables.

As shown in Table 5.3, SSBE /iː/ and /æ/ were mapped to two Urdu vowels. However, the rest of the SSBE monophthongs and diphthongs were assimilated to one Urdu vowel. The proportion of assimilation to a particular Urdu vowel was weighted by the mean goodness rating for stimuli receiving that identification. For example, SSBE /iː/ was most frequently assimilated to two Urdu vowels, /iː/ and /eː/. Fit indexes were calculated for both Urdu response vowels. The proportion of /iː/ assimilation (0.52) was multiplied by the mean goodness rating for the /iː/ response (5.5). This yielded a fit index of 2.9. The proportion of /eː/ assimilation (0.39) was multiplied by the mean goodness rating for the /eː/ response (4.8). This yielded a fit index of 1.9. The fit index for SSBE /aʊ/ was obtained by multiplying the proportion of the highest frequency (0.73) by the goodness rating of that assimilation (5.4). This resulted in a fit index of 3.9 for SSBE /aʊ/ to Urdu /ɑʊ/.

As shown in Table 5.3, only three SSBE vowels (diphthongs: /ɪə/ /eɪ/ /aʊ/) received the highest proportion of assimilation (> 70%) as they were consistently assimilated to one Urdu vowel. SSBE /ɪə/ and /eɪ/ were assimilated to Urdu monophthong /eː/; however, SSBE /aʊ/ was assimilated to Urdu diphthong /ɑʊ/. Only two of the SSBE vowels /ɔː/ and /aɪ/ received the second highest proportion of assimilation (> 60%). Rest of the SSBE vowels were assimilated to particular Urdu vowels less than 60%, and SSBE /ɜː/ and /ʊə/ were not assimilated consistently to any Urdu vowel. Hence these are not included in Table 5.3.

In hVba context only one SSBE vowel /aʊ/ received the highest proportion (> 70%) of assimilation, and only two SSBE vowels, /ɒ/ and /ɔː/ received the second highest proportion (> 60%) of assimilation. Rest of the SSBE vowels were assimilated to a number of Urdu vowels and the most frequent responses received the proportion less than 50%.

Table 5. 3: Fit indexes calculated for SSBE vowels in terms of Urdu vowels in bVd context. Following Guion et al (2000) only identifications that were more than 30% are included.

| Stimulus vowel | Response vowel | Frequency | Goodness | Proportion | Fit Index |
|---|---|---|---|---|---|
| iː | iː | 196 | 5.5 | 0.52 | 2.9 |
| | eː | 148 | 4.8 | 0.39 | 1.9 |
| ɪ | eː | 204 | 5.3 | 0.54 | 2.9 |
| ɛ | eː | 200 | 4.8 | 0.53 | 2.6 |
| æ | ɛː | 118 | 5.3 | 0.31 | 1.7 |
| | eː | 120 | 4.4 | 0.32 | 1.4 |
| ɑː | ɑː | 120 | 5.5 | 0.32 | 1.8 |
| ɒ | oː | 170 | 5.2 | 0.45 | 2.3 |
| ɔː | oː | 240 | 5.2 | 0.64 | 3.3** |
| ʊ | ʊ | 149 | 5.5 | 0.39 | 2.2 |
| uː | uː | 150 | 5.4 | 0.40 | 2.2 |
| ʌ | ɐ | 164 | 5.4 | 0.44 | 2.4 |
| ɪə | eː | 271 | 5.2 | 0.72 | 3.7* |
| eə | ɛː | 150 | 5.4 | 0.40 | 2.2 |
| eə | eː | 172 | 4.6 | 0.46 | 2.1 |
| eɪ | eː | 286 | 5.2 | 0.76 | 4.0* |
| aɪ | ɑe | 261 | 5.4 | 0.69 | 3.8** |
| ɔɪ | oe | 202 | 5.2 | 0.53 | 2.8 |
| əʊ | oː | 134 | 4.8 | 0.36 | 1.7 |
| aʊ | ɑʊ | 277 | 5.4 | 0.73 | 3.9* |

Table 5. 4: Fit indexes calculated for SSBE vowels in terms of Urdu vowels in hVba context. Following Guion et al (2000) only identifications that were more than 30% are included.

| Stimulus vowel | Response vowel | Frequency | Goodness | Proportion | Fit Index |
|---|---|---|---|---|---|
| iː | iː | 178.00 | 5.3 | 0.48 | 2.5 |
| ɪ | eː | 139.00 | 5.2 | 0.37 | 1.9 |
| ɛ | eː | 148.00 | 4.9 | 0.40 | 1.9 |
| | ɛː | 111.00 | 5.1 | 0.30 | 1.5 |
| ɒ | oː | 224.00 | 5.2 | 0.60 | 3.1** |
| ɔː | oː | 237.00 | 5.2 | 0.64 | 3.3** |
| ʊ | oː | 118.00 | 5.0 | 0.32 | 1.6 |
| | uː | 112.00 | 5.1 | 0.30 | 1.5 |
| uː | uː | 126.00 | 4.8 | 0.34 | 1.6 |
| ɪə | eː | 139.00 | 4.9 | 0.37 | 1.8 |
| eə | eː | 133.00 | 4.9 | 0.36 | 1.8 |
| eɪ | eː | 172.00 | 4.9 | 0.46 | 2.3 |
| aɪ | ɑe | 209.00 | 5.5 | 0.56 | 3.1 |
| ɔɪ | oe | 221.00 | 5.3 | 0.59 | 3.2 |
| aʊ | ɑʊ | 261.00 | 5.6 | 0.70 | 3.9* |

SSBE /æ/ /ɑː/ /ʌ/ /ɜː/ /ʊə/ and /əʊ/ were not assimilated consistently to any Urdu vowel and their highest proportion of assimilation was less than 30%. Hence these are not included in Table 5.4.

As discussed above and shown in Table 5.4, the fit index for SSBE /iː/ was obtained by multiplying the proportion of the highest frequency (0.48) by the goodness rating of that assimilation (5.3). This resulted in a fit index of 2.5 for SSBE /iː/ to Urdu /iː/. Similarly, the fit index for SSBE /aʊ/ was obtained by multiplying the proportion of the highest frequency (0.70) by the goodness rating of that assimilation (5.6). This resulted in a fit index of 3.9 for SSBE /aʊ/ to Urdu /ɑʊ/.

The fit indexes in both contexts (bVd and hVba) spanned a wide range. For example, in bVd context, the fit indexes ranged from a low value of 1.4 (the fit of SSBE /æ/ to Urdu /eː/) to a high value of 4.0 (the fit of SSBE /eɪ/ to Urdu /eː/). In hVba context, the fit indexes ranged from a low value of 1.5 (the fit of SSBE /ɛ/ to Urdu /ɛː/) to a high value of 3.9 (the fit of SSBE /aʊ/ to Urdu /ɑʊ/). The higher fit index for the modal responses (the most frequently chosen response) suggests that those vowels were perceived as good examples or very similar to Urdu vowels, and the lower fit index suggests that those vowels were perceived as poor examples of Urdu vowels.

To further analyse the effects of context and proficiency on the assimilation patterns of SSBE vowels to Urdu vowels by Punjabi Urdu speakers, the fit indexes, calculated for the 19 SSBE vowels were analysed statistically. The methods for statistical analysis were similar to the one discussed in Section 2.4.4.

The formula as entered into R was thus

Fit index ~ Stimulus Vowel*Context*Proficiency*(1|Subject)

Stimulus Vowel, Context and Proficiency were fixed effects and Subject was a random factor.

The ANOVA significance tests showed a significant main effect of Vowel ($F(18, 1940.08)=19.8$, $p <0.001$) and Context ($F(1, 1961.23)=100.6$, $p <0.001$), and interaction between Vowel × Context ($F(18, 1940.08)=3.5$, $p <0.001$), but non-significant effect of Proficiency ($F(1, 55.04)=1.9$; $p = 0.17$), and non-significant interaction between Vowel × Proficiency ($F(18, 1940.08)=3.5$, $p = 0.21$), Context × Proficiency ($F(1, 1961.23)=2.5$, $p = 0.11$), and non-significant three-way interaction between Vowel × Context × Proficiency ($F(18, 1940.08)=0.9$, $p = 0.48$).

As proficiency was not found to have a significant effect on the assimilation patterns of SSBE vowels to Urdu vowels, the reduced model after eliminating the non-significant effects (i.e. Proficiency) was used for pair-wise comparisons to analyse the effects of context on the assimilation patterns for SSBE vowels in two contexts, bVd and hVba.

The formula as entered into R was thus

Fit index ~ Stimulus Vowel*Context*(1|Subject)

The results showed that five SSBE vowels differed in their fit to Urdu response vowels with regard to the context. As shown in Table 5.5, SSBE (/iː/ /ɪ/ /uː/ /ɪə/ /eɪ/) fitted better in bVd context with a higher fit index than hVba context.

For the rest of the SSBE vowels, context was not found to have a significant effect on their fit to the Urdu response vowels. Overall, the SSBE vowels fitted well to Urdu vowels in bVd context; however, in some cases the fit index was identical across two contexts. For example, SSBE /aʊ/ fitted equally well to Urdu /ɑʊ/ in bVd (fit index = 3.9) and hVba (fit index = 3.9). Another example, SSBE /ɔː/ fitted equally well to Urdu /oː/ in bVd (fit index = 3.3) and hVba (fit index = 3.3).

Table 5. 5: Pairwise comparisons of fit indexes for SSBE vowels to Urdu vowels in bVd and hVba context

| Stimulus vowel | Response vowel | Fit Index bVd | Fit Index hVba | *t* test |
|---|---|---|---|---|
| iː | iː | 2.9 | 2.5 | t(1978.35)= 3.836, p < 0.001 |
| ɪ | eː | 2.9 | 1.9 | t(1978.35)= 4.982, p < 0.001 |
| uː | uː | 2.2 | 1.6 | t(1978.35)= 4.408, p < 0.001 |
| ɪə | eː | 3.7 | 1.8 | t(1978.35)= 5.040, p < 0.001 |
| eɪ | eː | 4.0 | 2.3 | t(1978.35)= 5.066, p < 0.001 |

As the above discussed results showed that proficiency did not have any effect on the assimilation of Urdu vowels, the data were pooled over categories of proficiency and assimilation patterns from the bVd and hVba context are discussed in turn in further detail in the following sections.

**5.6 Results in bVd Context**

Considering the 70% or above criterion of vowel assimilation for categorisation, it can be seen in Table 5.6 that only three SSBE diphthongs (/ɪə/, /eɪ/ and /aɪ/) were assimilated to Urdu response vowels in 70% of opportunities, or above. SSBE /ɪə/ (71%) and /eɪ/ (75%) were assimilated to Urdu vowel /eː/, and SSBE /aʊ/ (73%) was assimilated to Urdu /ɑʊ/. The mean goodness rating for these vowels is quite high, i.e. /ɪə/ 5.2, /eɪ/ 5.2 and /aʊ/ 5.4, which suggests that listeners found these vowels good exemplars of the chosen Urdu vowels. The rest of the 16 SSBE vowels were assimilated to Urdu vowels with a percentage of 50% or below.

Considering the 50% or above criterion of vowel assimilation for categorisation (Faris et al., 2016), it can be seen in Table 5.6 that SSBE vowels /iː/, /ɪ/, /ɛ/, /ɔː/, /aɪ/ and /ɔɪ/ were assimilated to Urdu response vowels in 50% of trials, or above. For instance, /iː/ 52%, /ɪ/ 54%, /ɛ/ 51%, /ɔː/ 64%, /aɪ/ 69% and /ɔɪ/ 53%, with a goodness rating ranging from 4.8

to 5.5. SSBE /iː/ was assimilated to Urdu /iː/, SSBE /ɔː/, /aɪ/, and /ɔɪ/ were assimilated to Urdu vowels /oː/, /ɑɪ/ and /ɔɪ/, respectively, as expected; however, SSBE /ɪ/ and /ɛ/ were assimilated to the same Urdu vowel /eː/. The rest of the SSBE vowels were assimilated to Urdu vowels in below 50% of opportunities. Hence, they can be considered *uncategorised*. However, the goodness ratings show that listeners were aware of the similarities or differences of the stimulus with their chosen Urdu response vowel.

In order to see categorised/uncategorised patterns of perceptual assimilation of SSBE vowels to Urdu vowels, it is important to see which Urdu response vowels were chosen for each SSBE stimulus vowel. Thus, overall assimilation patterns for SSBE front, central and back vowels are based on the total frequencies for each SSBE vowel across all participants, i.e. 66 (participants) × 6 (opportunities for each SSBE vowel (3 tokens for female and 3 tokens for male speaker) = 396 and discussed below. The least well matched SSBE monophthongs and diphthongs are shown in Figure 5. 6 and Figure 5. 7.

### 5.6.1 Assimilation Patterns for SSBE Front Vowels /iː/, /ɪ/, /ɛ/ and /æ/

As shown in Figure 5. 6 and Table 5.6:

- SSBE /iː/ was assimilated to Urdu /iː/ (52%; 5.5) and /eː/ (39%; 4.8);
- SSBE /ɪ/ was assimilated to Urdu /eː/ (54%; 5.3), Urdu vowel /ɪ/ (18%; 4.9) and Urdu vowel /ɐ/ (10%; 4.7);
- SSBE /ɛ/ was assimilated to Urdu /eː/ (51%; 4.8), Urdu /**ɛ**ː/ (26%; 4.8); and Urdu /ɐ/ (10%; 4.8);
- SSBE /æ/ was assimilated to Urdu /eː/ (30%; 4.5), Urdu /**ɛ**ː/ (31%; 5.3), and Urdu /ɑː/ (18%; 5.3).

Overall, it can be seen in Figure 5.6 and Table 5.6, the SSBE front vowels /iː/, /ɪ/, /ɛ/ and /æ/ were most often assimilated to the Urdu vowel /eː/ with a goodness rating ranging from 4.5 to 5.3.

### 5.6.2 Assimilation Patterns for SSBE Central Vowels /ʌ/ and /ɜː/

As shown in Table 5.6:

- SSBE central mid-low vowel /ʌ/ was assimilated to Urdu /ɐ/ (44%: 5.4), Urdu /ɑː/ (17%: 5.0), and Urdu /oː/ (7%: 4.8), well below the 50% criterion level.
- SSBE central mid-high vowel /ɜː/ was assimilated to Urdu /ɑː/ (25%: 5.2), Urdu /ɐ/ (15%: 5.4), and Urdu /eː/ (14%: 4.6), well below the 50% criterion level.

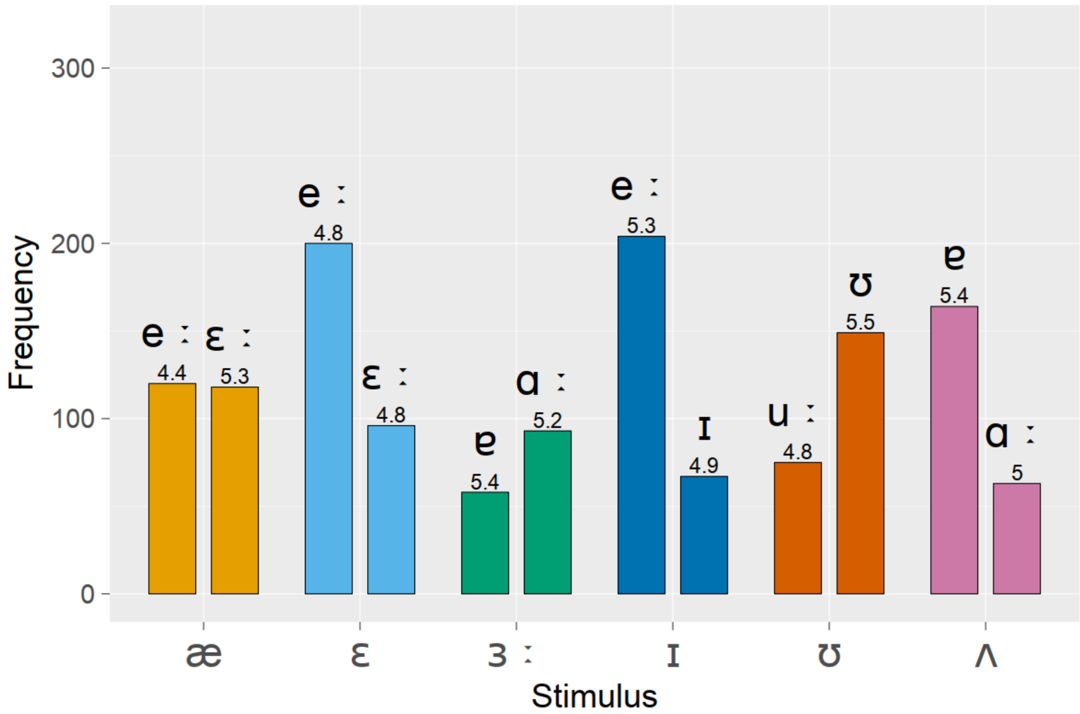



Figure 5. 6: The frequency and mean goodness rating for the two most frequent responses to each of the six least well-matched SSBE monophthongs in the bVd context. Maximum possible frequency N = 396.

### 5.6.3 Assimilation Patterns for SSBE Back Vowels /ɑː/, /ɒ/, /ɔː/, /ʊ/ and /uː/

As shown in Table 5.6:

- SSBE /ɑː/ was assimilated to Urdu /ɑː/ (32%; 5.5), Urdu /oː/ (19%; 4.9), and Urdu /ɑʊ/ (18%; 4.7).
- SSBE /ɒ/ was assimilated to Urdu /oː/ (40%; 5.2), Urdu /uː/ (24%; 4.8), and Urdu vowel /ʊ/ (10%; 5.5).

- SSBE /ɔː/ was assimilated to Urdu /oː/ (64%; 5.1), Urdu /uː/ (27%; 5.0), and Urdu /ʊ/ (4%; 4.7).
- SSBE /ʊ/ was assimilated to Urdu /ʊ/ (40%; 5.5), Urdu /oː/ (17%; 5.0), and Urdu /ɐ/ (13%; 5.4).
- SSBE vowel /uː/ was assimilated to Urdu /uː/ (40%; 5.5), Urdu /oː/ (26%; 5.1), and Urdu /ʊ/ (14%; 4.9).

### 5.6.4 Assimilation Patterns for SSBE Diphthongs /ɪə/, /eə/, /ʊə/ and /əʊ/

As shown in Figure 5. 7 and Table 5.6:

- SSBE /ɪə/ was assimilated to Urdu /eː/ (71%; 5.2), none (7%; 1.1) and Urdu /ɪɐ/ (3%; 3.3).
- SSBE /eə/ was assimilated to Urdu /eː/ (45%; 4.6), Urdu vowel /ɛː/ (40%; 5.4) and Urdu vowel /ɑː/ (7%; 5.4).
- SSBE /ʊə/ was assimilated to none (25%; 1.6), Urdu /eː/ (19%; 4.2), and Urdu /oː/ (9%; 4.6).
- SSBE diphthong /əʊ/ was assimilated to Urdu /oː/ (32%; 4.9), Urdu /ɑʊ/ (13%; 4.6) and Urdu /uː/ (11%; 4.5).

### 5.6.5 Assimilation Patterns for SSBE Diphthongs /eɪ/, /aɪ/, /aʊ/ and /ɔɪ/

A shown in Figure 5. 8 and Table 5.6:

- SSBE /eɪ/ was assimilated to Urdu /eː/ (75%; 5.2), Urdu /ɛː/ (6%; 5.4) and none (6%; 1.4).
- SSBE /aɪ/ was assimilated to Urdu vowel /ɑɪ/ (69%; 5.4), Urdu vowel /eː/ (12%; 4.1) and Urdu vowel /ɑʊ/ (5%; 4.5).
- SSBE /ɔɪ/ was assimilated to Urdu /oe/ (53%; 5.2), Urdu /oː/ (16%; 4.8) and none (6%; 1.4).
- SSBE /aʊ/ was assimilated to Urdu /ɑʊ/ (73%; 5.4), Urdu /uː/ (9%; 3.8) and Urdu /oː/ (8%; 4.4).

Table 5. 6: Mean percent categorization, and goodness ratings in parenthesis, of SSBE vowels in a bVd context by Punjabi-Urdu speakers with Punjabi-Urdu vowel response categories.

| | Stimuli bVd | Punjabi-Urdu vowels (three most frequent percentages) | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | iː | ɪ | eː | ɛː | ɑː | ɔː | ʊ | uː | ɐ | ɪɐ | eɐ | ʊɑ | ɑe | oe | ɑʊ | None |
| | iː | **52 (5.5)** ** | | 39 (4.8) | | | | | | | | | | | | | 3 (1.6) |
| | ɪ | | 18(4.9) | **54 (5.3)** | | | | | | 10 (4.7) | | | | | | | |
| | ɛ | | | **51 (4.8)** | 26 (4.8) | | | | | 10 (4.8) | | | | | | | |
| Categorised | ɔː | | | | | | **64 (5.1)** | 4 (4.7) | 27 (5.0) | | | | | | | | |
| | ɪə | | | **71 (5.2)** * | | | | | | | 3 (3.3) | | | | | | 7 (1.1) |
| | eɪ | | | **75 (5.2)** | 6(5.4) | | | | | | | | | | | | 6 (1.4) |
| | aɪ | | | 12 (4.1) | | | | | | | | | | **69 (5.4)** | | 5 (5.5) | 5 (1.2) |
| | ɔɪ | | | | | | 16 (5.8) | | | | | | | | **53 (5.2)** | | 6 (1.4) |
| | aʊ | | | | | | 8 (4.4) | | 9 (3.8) | | | | | | | 73 (5.4) | |
| | ɒ | | | | | | **40 (5.2)** | 10 (5.5) | 24 (4.8) | | | | | | | | |
| Focalised | ʊ | | | | | | 17 (5.0) | **40 (5.5)** | | 13 (5.4) | | | | | | | |
| | uː | | | | | | 26 (5.1) | 14 (4.9) | **40 (5.5)** | | | | | | | | |
| | ʌ | | | | | 17 (5.0) | 7 (4.8) | | | **44 (5.4)** | | | | | | | |
| | æ | | | 30 (4.5) | 31 (5.3) | 18 (5.3) | | | | | | | | | | | |
| | ɑː | | | | | 32 (5.5) | 19 (4.9) | | | | | | | | | 18 (4.7) | |
| | ɜː | | | 14 (4.6) | | 25 (5.2) | | | | 15 (5.4) | | | | | | | |
| Clustered | eə | | | 45 (4.6) | 40(5.4) | 7 (5.4) | | | | | | | | | | | |
| | əʊ | | | | | | 32 (4.9) | | 11 (4.5) | | | | | | | 13 (4.6) | |
| Dispersed | ʊə | | | 19 (4.2) | | | 9 (4.6) | | | | | | | | | | 25 (1.6) |

* Numbers in boldface present the mean percentages and goodness ratings for 70% criterion of categorised assimilation.

** Numbers in boldface and italics present the mean percentages and goodness ratings for 50% criterion of categorised assimilation.

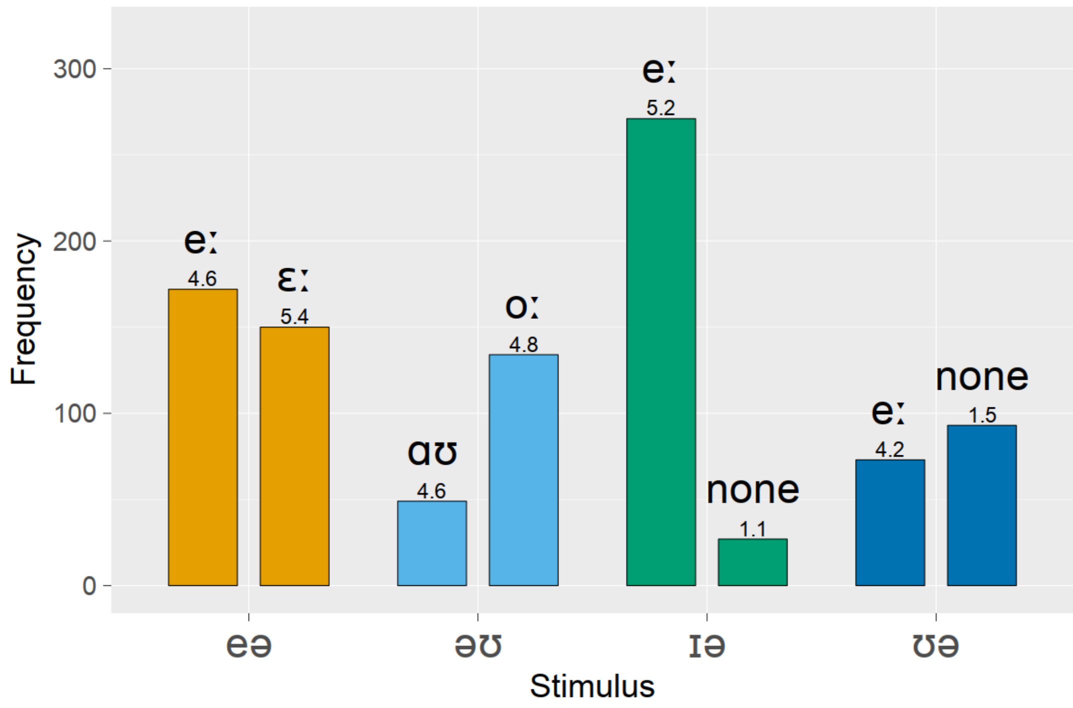



Figure 5. 7: The frequency and mean goodness rating for the two most frequent responses to each of the four least well-matched SSBE diphthongs in the bVd context. Maximum possible frequency N = 396.

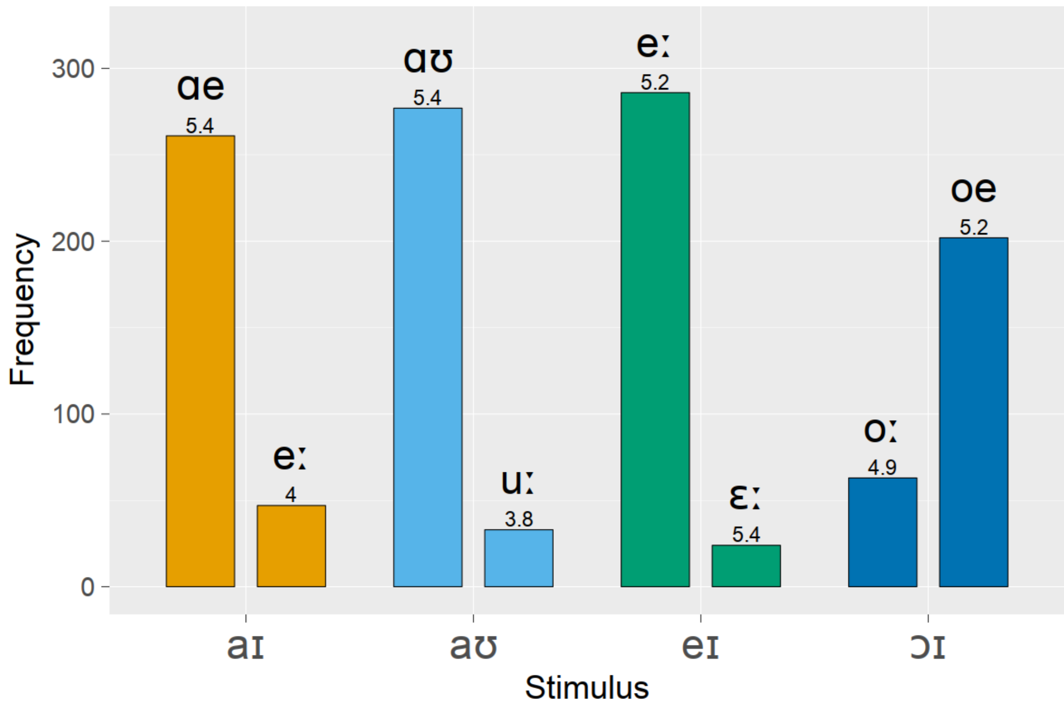



Figure 5. 8: The frequency and mean goodness rating for the two most frequent responses to each of the four most well-matched SSBE diphthongs in the bVd context. Maximum possible frequency N = 396.

Since 10 out of 19 of the SSBE vowels in the bVd context were assimilated to multiple Urdu vowels, this type of assimilation can be considered *Multiple Category Assimilation* (*MCA*; Escudero and Boersma, 2002) or *Uncategorised* assimilation (Faris et al., 2016), and discussed further in the next section.

### 5.6.6 Uncategorised Assimilation Patterns in bVd

As discussed in Section 4.2.3.1, according to Faris et al. (2016:3), "…uncategorized phones are operationally defined as those that are not consistently assigned to a single L1 category above a predefined threshold (e.g. 50%)". Therefore, the assimilation patterns in the present study can be considered *uncategorised*. These are further separated into three types, as discussed by Faris et al. (2016:3): *focalised*-uncategorised, where a non-native phoneme is usually mapped to a single L1 phone, but at a frequency below the categorisation threshold (i.e. 50% in this case); *clustered*-uncategorised, where the non-native phone is assimilated to a small set of L1 categories; and *dispersed*-uncategorised, where a non-native phoneme is assimilated to multiple L1 categories randomly because none of the L1 categories are similar (i.e. share phonetic or phonological similarity) to the non-native phoneme.

In the present study a number of SSBE vowels were assimilated to multiple Urdu vowels. The *focalised*, *clustered* and *dispersed* assimilation patterns in the bVd context can be seen in Table 5.6. However, the SSBE /iː/, /ɪ/, /ɛ/, /ɔː/, /ɪə/, /eɪ/, /aɪ/, /ɔɪ/ and /aʊ/ were categorised as they were consistently assimilated to an Urdu vowel at predefined threshold 50% or above. SSBE /ɪ/, /ɛ/, /ɪə/ and /eɪ/ were assimilated to Urdu /eː/. However, SSBE /iː/, /ɔː/, /aɪ/, /ɔɪ/ and /aʊ/ were assimilated to Urdu /iː/, /oː/, /ɑe/, /oe/ and /ɑʊ/ respectively. SSBE /ɒ/, /ʊ/, /uː/ and /ʌ/ were *focalised* as these were mostly assimilated to Urdu vowels /oː/, /ʊ/, /uː/ and /ɐ/, respectively, but below the 50% threshold.

SSBE /æ/, /ɑː/, /ɜː/ and /eə/ were *clustered* as they were mapped to a number of Urdu vowels. SSBE /əʊ/ and /ʊə/ were *dispersed* as they appear to be assimilated to multiple L1 categories randomly.

### 5.6.7 Summary and Discussion of bVd Results

According to the Perceptual Assimilation Model (PAM: Best, 1995), *Single-Category* (SC) assimilation patterns can be seen here. For instance, SSBE vowels /iː/, /ɪ/, /ɛ/, /æ/, /ɜː/, /ɪə/, /eə/, /ʊə/, /eɪ/ and /aɪ/ were assimilated to one Urdu vowel /eː/, with varying frequency and median goodness ratings. In addition, SSBE vowels /ɑː/, /ɒ/, /ɔː/, /ʊ/, /uː/, /ʌ/, /ʊə/, /ɔɪ/, /əʊ/ and /aʊ/ were assimilated to one Urdu vowel /oː/, with varying frequency and median ratings.

*Two-Category* (TC) assimilation patterns can be seen for the following: SSBE /iː/ and /ɪ/ were assimilated to Urdu /iː/ (52%; 5.5) and /eː/ (54%; 5.3) respectively. SSBE /uː/ and /ʊ/ were assimilated to Urdu /uː/ (40%; 5.5) and /ʊ/ (40%; 5.5) respectively. SSBE /ʌ/ and /ɜː/ were assimilated to Urdu /ɐ/ (44%; 5.4) and /ɑː/ (25%; 5.2) respectively. Clearly, some of these vowels were assimilated less frequently than the 50% criterion for a vowel to be considered as categorised. Therefore, these findings were further analysed as *uncategorised* assimilations and discussed in Section 5.6.7.

In the next section, predictions are discussed based on acoustic/phonetic and phonological similarities and differences between SSBE and Punjabi-Urdu vowels.

#### 5.6.7.1 Comparison with predictions of PAM, L2LP and SLM

**Monophthongs**

Based on acoustic comparison (see Sections 5.1.1, 5.1.2 and 5.1.3)., it was predicted that there will be some *Two-Category* (TC) and more *Single-Category* (SC) (Best, 1995; Best and Tyler, 2007) assimilation patterns for SSBE vowels by Punjabi-Urdu speakers. These predictions were supported to some extent. For example, from the visual inspection and acoustic measurements, it was predicted that SSBE /ɪ/ will be confused with Urdu /ɪ/ and /eː/. SSBE /ɛ/ will be assimilated to Urdu /**ɛ**ː/. Urdu /oː/ lies somewhere in the middle of SSBE /ɔː/ and /ɒ/; hence it was predicted that these two back SSBE vowels will be assimilated to this Urdu back vowel. The results showed these patterns. Hence these predictions were correct.

Based on the acoustic measurement (F1 for SSBE /æ/ is higher than Urdu /**ɛ**ː/; F2 for SSBE /æ/ is lower than Urdu /**ɛ**ː/) it was predicted that SSBE /æ/ may be assimilated to Urdu /ɑː/. However, the results show that /æ/ was mapped to three different Urdu vowel categories with mean goodness rating of 5.0, i.e. /eː/ (30%; 4.5), /**ɛ**ː/ (31%; 5.3), and /ɑː/ (18%; 5.3). It was predicted that SSBE /ʌ/ appears to be acoustically similar to Urdu /ɑː/, but the assimilation patterns show that SSBE /ʌ/ was mapped to Urdu /ɐ/ (44%; 5.4) with a higher percentage. However, SSBE /ɜː/ was often mapped to Urdu /ɑː/ with a mean goodness rating of 5.2 (25%). These findings suggest that acoustic measurements cannot always predict the perceptual assimilation patterns (cf. Strange et al., 2004; Nishi et al., 2008; Escudero, Simon and Mitterer, 2012).

Since F1 of SSBE /uː/ and /ʊ/ is quite front in the vowel space as compared to Urdu /uː/ and /ʊ/, it was predicted that SSBE /uː/ and /ʊ/ will be assimilated to Urdu /uː/ and /ʊ/. However, F2 of SSBE /uː/ is much higher than Urdu /uː/, and so it was difficult to predict whether Punjabi-Urdu speakers/listeners can identify these vowels as distinct or not. The results show that Punjabi-Urdu speakers/listeners identified these two vowels as distinct and assimilated SSBE /uː/ and /ʊ/ to Urdu /uː/ (40%; 5.5) and /ʊ/ (40%; 5.5) respectively. However, there was a slight overlap (14%, see Section 5.8) for the perceptual assimilation of SSBE /uː/ and /ʊ/, since they were both assimilated to Urdu /ʊ/. The classification overlap scores (cf. Flege and MacKay, 2004) are discussed in Section 5.8.

**Diphthongs**

Based on the acoustic measurements of SSBE and Urdu vowels, it was predicted that for SSBE diphthongs /eɪ/ and /əʊ/ participants will choose “none”, since these two diphthongs are not found in the Urdu vowel inventory. Contrary to this prediction, Punjabi-Urdu speakers/listeners assimilated SSBE /eɪ/ / to Urdu vowel /eː/ (75%; 5.2) and SSBE /əʊ/ was assimilated to Urdu /oː/ (32%; 4.9) and to a number of other Urdu vowels. This assimilation patterns shows that Punjabi-Urdu speakers/listeners focussed on the first element in the /eɪ/ diphthong and second element in the /əʊ/ diphthong and mapped these to the respective Urdu counterparts /eː/ and /oː/, respectively. However, as predicted Punjabi-Urdu speakers/listeners did not show much confusion for SSBE diphthongs /aɪ/, /ɔɪ/, and /aʊ/. Hence, they mapped these three diphthongs to the expected Urdu diphthongs /ɑe/ (69%; 5.4), /oe/ (53%; 5.2), and /ɑʊ/ (73; 5.4) respectively.

The assimilation patterns for SSBE diphthongs /ɪə/, /eə/ and /ʊə/ are not as expected. It was expected that these vowels will be assimilated to respective Urdu diphthongs, i.e. /ɪɐ/, /eɐ/ and /ʊɑ/. However, SSBE /ɪə/ was assimilated to Urdu /eː/ (71%; 5.2); SSBE /eə/ was assimilated to Urdu /eː/ (45%; 4.6) and /æ/ (40%; 5.4); and SSBE /ʊə/ was assimilated to “none” (25%; 1.6).

SSBE /eə/ was assimilated to both Urdu /eː/ (45%; 4.6) and /**ɛ**ː/ (40%; 5.4), which suggests a *Category-Goodness* (CG) assimilation pattern. With regard to mean goodness rating, /**ɛ**ː/ (5.4) is a better Urdu exemplar for SSBE /eə/ than Urdu /eː/ (4.6). Future production and/or discrimination experiments can establish if these predictions will bear out.

## 5.7 Results in hVba Context

Considering 50% or above as the criterion of vowel assimilation for categorisation, it can be seen in Table 5.7 that only five SSBE vowels, three of which are diphthongs (/ɒ/, /ɔː/, /aɪ/, /ɔɪ/, /aʊ/), were assimilated to Urdu response categories. SSBE /ɒ/ (60%; 5.3) and

/ɔː/ (64%; 5.1) were assimilated to Urdu /oː/ as expected, SSBE /aɪ/ was assimilated to Urdu /ɑe/ (56%; 5.6), SSBE /ɔɪ/ (59%; 5.3) was assimilated to Urdu /oe/, and SSBE /aʊ/ (70%; 5.6) was assimilated to Urdu /ɑʊ/. The mean goodness rating for these vowels is quite high, i.e. /ɒ/ 5.3, /ɔː/ 5.1, /aɪ/ 5.6, /ɔɪ/ 5.3, and /aʊ/ 5.6 which suggests that listeners found these vowels good exemplars of the chosen Urdu vowels. The remaining 14 SSBE vowels were assimilated to multiple Urdu vowels with a percentage below 50%; hence can be considered *uncategorised*. However, the goodness ratings show that listeners were aware of the similarities or differences of the stimulus with their chosen Urdu response vowel.

As discussed in Section 5.5, statistical tests showed a significance effect of context on the patterns of perceptual assimilation. Once again, in order to see categorised/uncategorised patterns of perceptual assimilation of SSBE vowels to Urdu vowels, it is important to see which Urdu response vowels were chosen for each SSBE stimulus vowel. Thus, overall assimilation patterns (as discussed above in Section 5.6) for SSBE front, central and back vowels are discussed below. The least well matched SSBE monophthongs and diphthongs are shown in Figure 5. 9 and Figure 5. 10 respectively.

### 5.7.1 Assimilation Patterns for SSBE Front Vowels /iː/, /ɪ/, /ɛ/ and /æ/

As shown in Figure 5. 9 and Table 5.7:

- SSBE /iː/ was assimilated to Urdu /iː/ (48%; 5.3) and /eː/ (15%; 5.1);
- SSBE /ɪ/ was assimilated to Urdu /eː/ (34%; 5.1), Urdu /ɪ/ (6%; 5.4), Urdu /ɪ/ (12%; 4.7), and Urdu /ɪə/ (11%; 4.1).
- SSBE /ɛ/ was assimilated to Urdu /eː/ (40%; 4.9) and Urdu /**ɛ**ː/ (30%; 5.1).
- SSBE /æ/ was assimilated to multiple Urdu vowels; however, the two most frequently chosen Urdu response categories were /ɑʊ/ (23%; 5.3) and /ɑː/ (16%; 5.0). These small percentages show that this vowel was confused with a number

of Urdu vowels. The five most frequently chosen Urdu vowel categories are shown in Table 5.7.

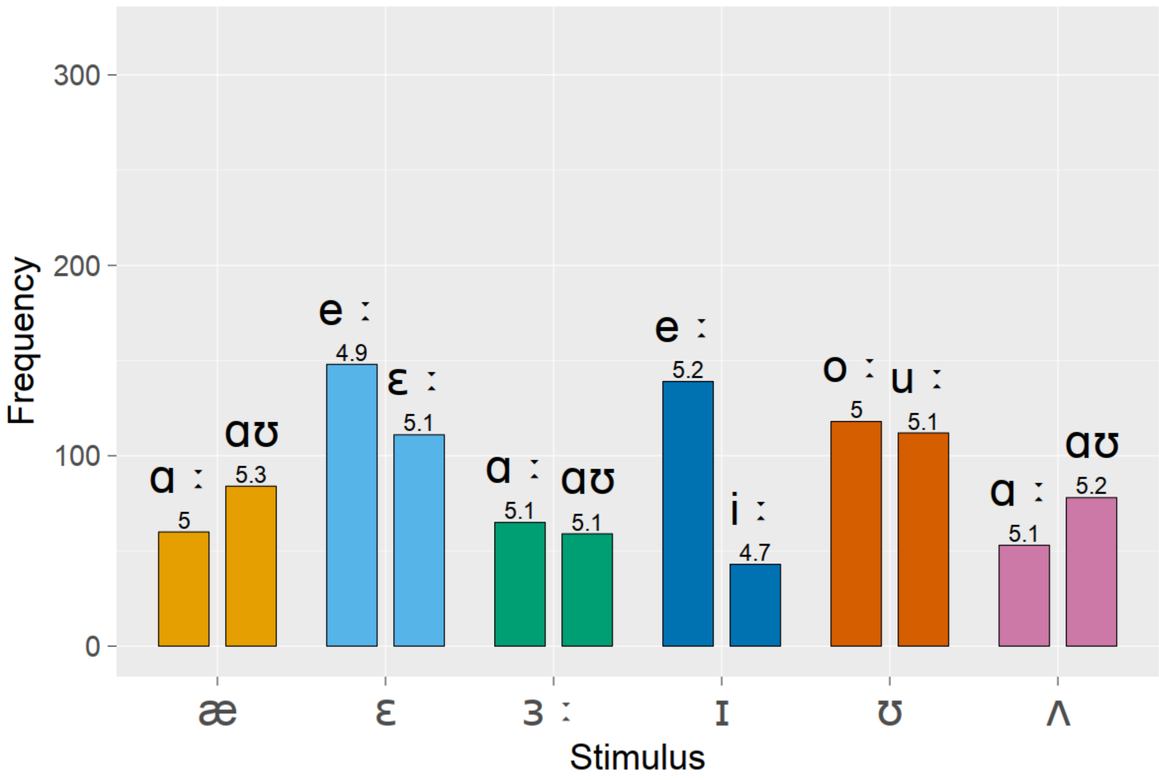



Figure 5. 9: The frequency and mean goodness rating for the two most frequent responses to each of the six least well-matched SSBE monophthongs in the hVba context. Maximum possible frequency N = 396.

Overall, SSBE front vowels /iː/, /ɪ/ and /ɛ/ were most often assimilated to Urdu vowel /eː/ with a goodness rating ranging from 4.9 to 5.1 SSBE /æ/ was assimilated to multiple Urdu vowels, with the most frequently chosen category as Urdu vowel-vowel sequence /ɑʊ/. It can be concluded that since the stimuli were nonsense words in disyllabic form, and participants were told untruthfully that these words were from a different language and not English, they did not use the same strategies to categorise these vowels that they used in the bVd context.

### 5.7.2 Assimilation Patterns for SSBE Central Vowels /ʌ/ and /ɜː/

As shown in Table 5.7:

- SSBE central mid-low vowel /ʌ/ was assimilated to multiple Urdu vowels, but never with Urdu /ɐ/. As shown in Table 5.7, SSBE /ʌ/ was assimilated to Urdu vowel-vowel sequence /ɑʊ/ (21%: 5.2), Urdu vowel /ɑː/ (14%: 5.1), and multiple other vowels.
- SSBE central mid-high vowel /ɜː/ was assimilated to Urdu /ɑː/ (17%: 5.1), Urdu vowel-vowel sequence /ɑʊ/ (16%: 5.1) and multiple other Urdu vowels.

These results show that in hVba context Punjabi-Urdu speakers/listeners found these two central vowels very difficult as compared to the ones they heard in bVd context, where they categorised /ʌ/ with Urdu /ɐ/.

### 5.7.3 Assimilation Patterns for SSBE Back Vowels /ɑː/, /ɒ/, /ɔː/, /ʊ/ and /uː/

As shown in Table 5.7:

- SSBE /ɑː/ was assimilated to Urdu /ɑʊ/ (28%; 5.2); Urdu /ɑː/ (13%; 4.9) and multiple other vowels.
- SSBE /ɒ/ was assimilated to Urdu /oː/ (60%; 5.3), Urdu /uː/ (8%; 4.7).
- SSBE /ɔː/ was assimilated to Urdu /oː/ (64%; 5.1), Urdu /uː/ (11%; 4.6), and Urdu /oe/ (9%; 5.3).
- SSBE /ʊ/ was assimilated to Urdu /uː/ (30%; 5.0); Urdu /ʊ/ (29%; 5.0), and Urdu /ʊɑ/ (8%; 5.0).
- SSBE /uː/ was assimilated to Urdu /uː/ (34%; 4.8), Urdu /oː/ (19%; 4.8); Urdu /oe/ (9%; 4.9), and Urdu /ʊɑ/ (7%; 4.7).

These results show that in the hVba context Punjabi-Urdu speakers/listeners did not categorise SSBE /ʊ/ and /uː/ as distinct vowels, even though they have counterparts in Urdu that were matched in the bVd context.

### 5.7.4 Assimilation Patterns for SSBE Diphthongs /ɪə/, /eə/, /ʊə/ and /əʊ/

As shown in Figure 5. 10 and Table 5.7:

- SSBE /ɪə/ was assimilated to Urdu /eː/ (35%; 4.8), and Urdu /eɐ/ (17%; 4.7) and some other Urdu vowels.
- SSBE /eə/ was assimilated to Urdu vowel /eː/ (31%; 4.8), Urdu /ɛː/ (24%; 5.1) and Urdu /eɐ / (9%; 3.9).
- SSBE /ʊə/ was assimilated to none (21%; 1.5), Urdu /eː/ (13%; 4.9), and Urdu /uː/ (11%; 4.4).

### 5.7.5 Assimilation Patterns for SSBE Diphthongs /eɪ/, /aɪ/, /ɔɪ/ and /aʊ/

As shown Figure 5. 11 and Table 5.7:

- SSBE /eɪ/ was assimilated to Urdu /eː/ (45%; 4.9), Urdu /eɐ/ (8%; 4.4); Urdu /oe/ (8%; 4.8), and none (9%; 1.3).
- SSBE /aɪ/ was assimilated to Urdu /ɑe/ (56%; 5.6), Urdu /oe/ (9%; 4.9) and Urdu /ɑʊ/ (9%; 5.0).
- SSBE /ɔɪ/ was assimilated to Urdu /oe/ (59%; 5.3), and Urdu /oː/ (10%; 5.0).
- SSBE /əʊ/ was assimilated to Urdu /oː/ (21%; 4.8), Urdu /oe/ (15%; 4.8) and Urdu /ɑʊ/ (15%; 4.5).
- SSBE /aʊ/ was assimilated to Urdu /ɑʊ/ (70%; 5.6), Urdu /ɑe/ (5%; 5.2) and Urdu /oː/ (5%; 3.7).

Since most of the SSBE vowels were assimilated to multiple Urdu vowels, this type of assimilation can be considered *Multiple Category Assimilation* (*MCA*; Escudero and Boersma, 2002) or *uncategorised* assimilation (Faris et al., 2016).

### 5.7.6 Uncategorised Assimilation Patterns in hVba Context

With regard to the uncategorised assimilation patterns as discussed in section 5.6, it can be seen in Table 5.6 that SSBE /iː/, /ɒ/, /ɔː/, /aɪ/, /ɔɪ/, /aʊ/ were categorised and assimilated to Urdu vowels /iː/, /oː/, /oː/, /ɑe/, /oe/, /ɑʊ/, respectively, and consistently above the 50% threshold.

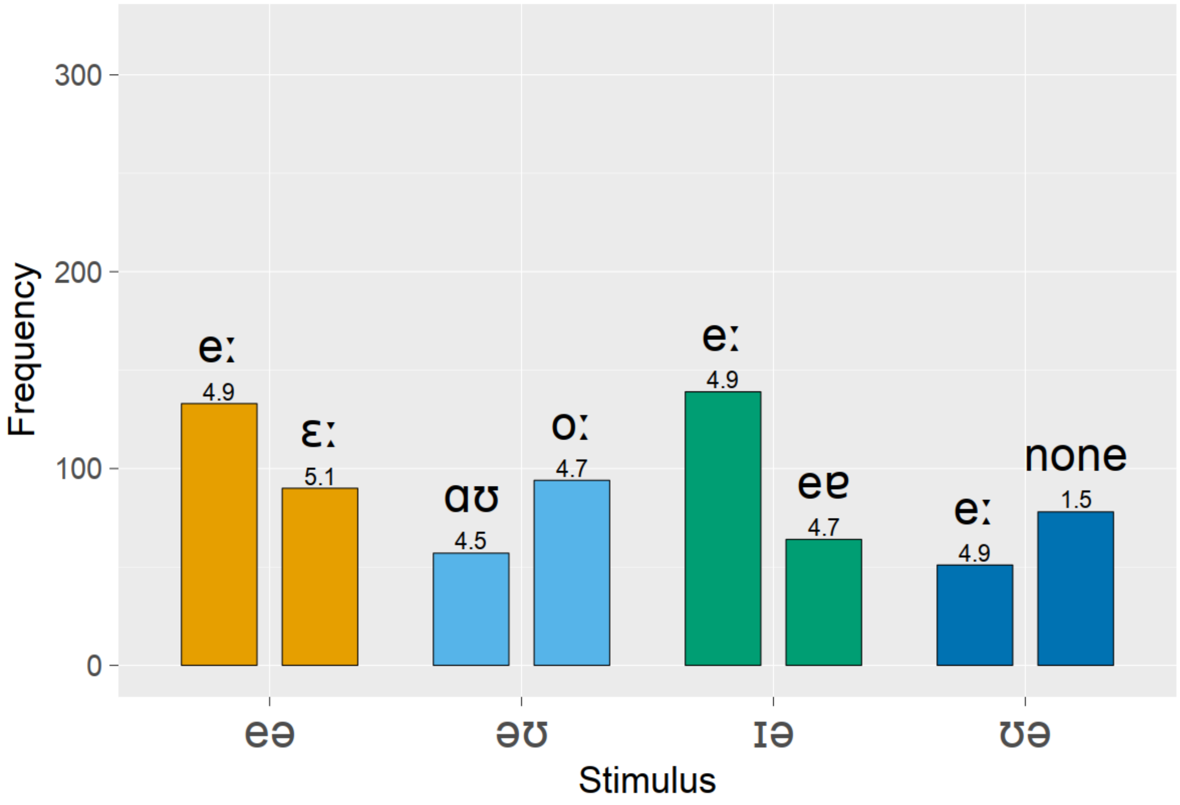



Figure 5. 10: The frequency and mean goodness rating for the two most frequent responses to each of the four least well-matched SSBE diphthongs in the hVba context. Maximum possible frequency N = 396.

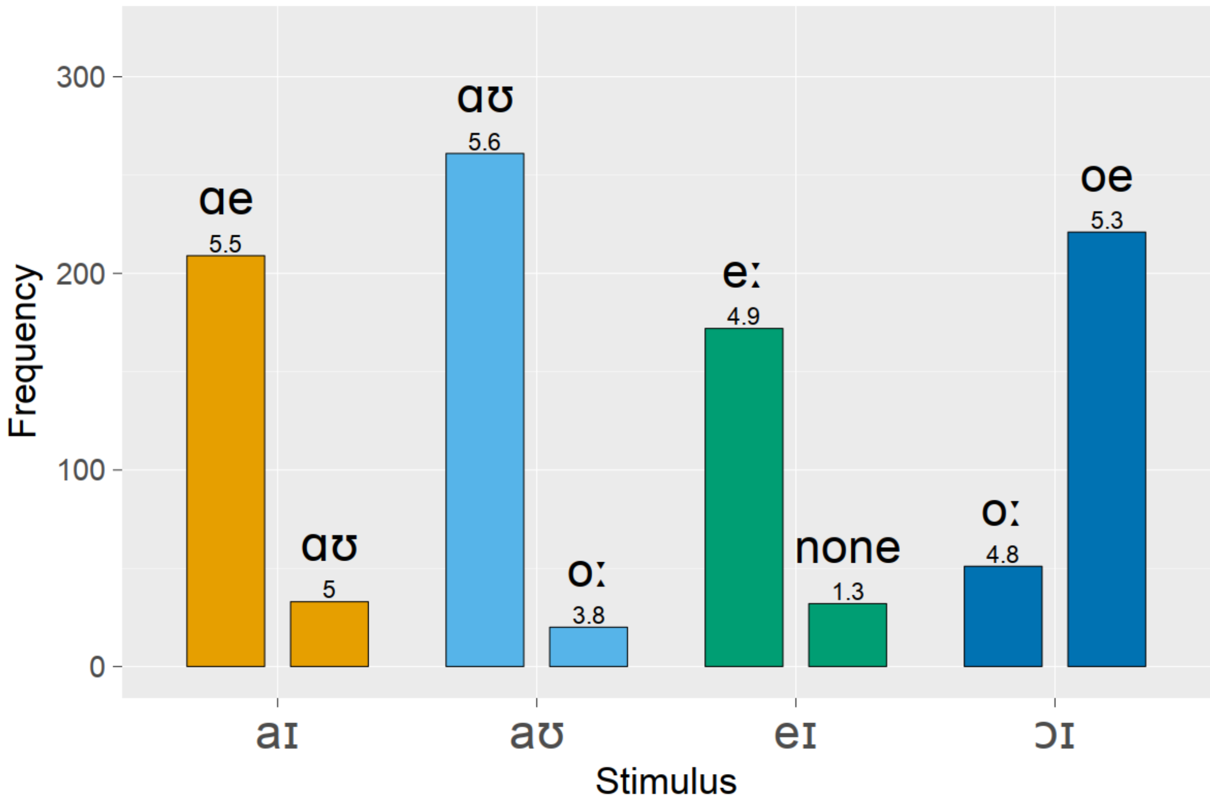



Figure 5. 11: The frequency and mean goodness rating for the two most frequent responses to each of the four most well-matched SSBE diphthongs in the hVba context. Maximum possible frequency N = 396.

Table 5. 7: Mean percent categorization, with goodness ratings in parenthesis, of SSBE vowels in a hVba context by Punjabi-Urdu speakers with Punjabi-Urdu vowel response categories.

| | hVba | | Punjabi-Urdu vowels (three most frequent percentages) | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | iː | ɪ | eː | ɛː | ɑː | ɔː | ʊ | uː | ɐ | ɪɐ | eɐ | ʊɑ | ɑe | oe | ɑʊ | None |
| | iː | ***48 (5.3)*** | | 15 (5.1) | | | | | | | 9 (5.7) | | | | | | |
| | ɒ | | | | | | ***60 (5.3)***[**] | | 8 (4.7) | | | | | | 7 (5.2) | 7 (4.7) | |
| Categorised | ɔː | | | | | | ***64 (5.1)*** | | 11 (4.6) | | | | | | 9 (5.3) | | |
| | aɪ | | | | | | | | | | | | | ***56 (5.6)*** | 9 (4.9) | 9 (5.0) | 5 (1.0) |
| | ɔɪ | | | | | | 10 (5.0) | | 5 (4.5) | | | | | | ***59 (5.3)*** | | |
| | aʊ | | | | | | 5 (3.7) | | | | | | | 5 (5.2) | | ***70 (5.6)*** | |
| | ɪ* | 12 (4.7) | 6 (5.4) | ***34 (5.1)*** | | | | | | | 11 (4.1) | | | | | | 8 (1.9) |
| | ɛ | | | ***40 (4.9)*** | 30 (5.1) | | | | | | | | | | | | 5 (1.2) |
| Focalised | eɪ | | | ***45 (4.9)*** | | | | | | | | 8 (4.4) | | | 8 (4.8) | | 9 (1.3) |
| | ʊ | | | | | | ***29 (5.0)*** | | ***30 (5.0)*** | | | | 8 (5.0) | | | | |
| | uː | | | | | | 19 (4.8) | | ***34 (4.8)*** | | | | 7 (4.7) | | 9 (4.9) | | |
| | ɪə | | | ***35 (4.8)*** | | | | | | | 8 (4.1) | 17 (4.7) | | | | | 11 (2.0) |
| Clustered | eə | | | ***31 (4.8)*** | 24 (5.1) | | | | | | | 9 (3.9) | | | | | 8 (1.6) |
| | æ* | | | | | 16 (5.0) | | | | | | 11 (4.1) | 9 (5.0) | 10 (5.4) | | 23 (5.3) | |
| | ɑː* | | | | | 13 (4.9) | | | | | | | 10 (4.7) | 10 (5.1) | | 28 (5.2) | 7 (1.1) |
| | ʌ* | | | | | 14 (5.1) | | | | | | 11 (3.8) | 10 (4.8) | 7 (5.2) | | 21 (5.2) | |
| | ɜː | | | | | 17 (5.1) | | | | | | 11 (4.3) | | | | 16 (5.1) | |
| Dispersed | əʊ | | | | | | 21 (4.8) | | 11 (4.6) | | | | | | 15 (4.8) | 15 (4.5) | |
| | ʊə | | | 13 (4.9) | | | | | 11 (4.4) | | | | | | | | 21 (1.5) |

[*] Five largest percentages shown, since the distribution reveals these stimuli were assimilated to multiple Urdu vowel categories.
[**] Numbers in boldface and italics present the mean percentages and goodness ratings for 50% criterion of categorised assimilation.

SSBE /ɪ/, /ɛ/ and /eɪ/ were *focalised* as these were consistently assimilated to Urdu /eː/, but below the predefined 50% threshold. SSBE /ʊ/, /uː/, /ɪə/ and /eə/ were *clustered* as they were mapped to a number of Urdu vowels. SSBE /æ/, /ɑː/, /ʌ/, /ɜː/, /əʊ/ and ʊə/ were *dispersed* as they appear to be assimilated to multiple Urdu categories randomly.

### 5.7.7 Summary and Discussion of hVba Context

In agreement with the Perceptual Assimilation Model (PAM: Best, 1995), *Two-Category* (TC) assimilation patterns can be seen for SSBE diphthongs /ɑɪ/ (56%), /ɔɪ/ (59%) and /ɑʊ/ (70%). *Single-Category* (SC) assimilation can be seen for SSBE /iː/, /ɪ/, /ɛ/, /ɪə/, /eə/, /ʊə/ and /eɪ/, as these were assimilated to the same Urdu vowel /eː/. Similarly, SSBE /ɒ/, /ɔː/, /ʊ/, /uː/, /ɔɪ/, /əʊ/ and /aʊ/ were assimilated to the same Urdu vowel /oː/.

In addition, *Two-Category* (TC) assimilation patterns can be seen here for the following:

SSBE /iː/ and /ɪ/ were assimilated to Urdu /iː/ (48%; 5.3, 12%; 4.7) and /eː/ (15%; 5.1: 34%; 5.1) respectively; and SSBE /uː/ and /ʊ/ were assimilated to Urdu /uː/ (34%; 4.8) and /ʊ/ (30%; 5.0) respectively.

*Single-Category* (SC) assimilation patterns can be seen here for the following: SSBE /ɪ/ and /ɛ/ were assimilated to the same Urdu vowel /eː/ (34%; 5.1) and (40%; 4.9), respectively, and SSBE /ʌ/ and /ɜː/ were assimilated to the same Urdu vowel /ɑː/ (14%; 5.1) and (17%; 5.1) respectively.

#### 5.7.7.1 Comparison with Predictions of PAM, L2LP and SLM

Based on acoustic comparison of SSBE and Urdu vowels (see Sections 5.1.1, 5.1.2 and 5.1.3), it was predicted that there will be some *Two-Category* (TC) and more *Single Category* (SC) (Best, 1995; Best and Tyler, 2007) assimilation patterns for SSBE vowels by Punjabi-Urdu speakers/listeners. These predictions were supported to some extent and patterns of assimilation were similar to those in the bVd context, with a few exceptions.

For example, it was predicted that SSBE /ʌ/ will be assimilated to Urdu /ɑː/ since it is acoustically similar. The assimilation patterns showed that SSBE /ʌ/ and /ɜː/ were both mapped to Urdu /ɑː/ with a mean goodness rating of 5.1. However, in the bVd context the SSBE /ʌ/ was not assimilated to the acoustically similar Urdu /ɑː/ but rather to Urdu /ɐ/. These findings suggest that acoustic measurements can help to predict the perceptual assimilation patterns in some contexts (i.e. hVba - when listener is unaware of the language) but not others (i.e. bVd - when listener is aware of the language) (cf. Strange et al., 2004; Nishi et al., 2008; Escudero et al., 2012).

In addition, unlike the assimilation patterns observed in the bVd context, Punjabi-Urdu speakers/listeners did not identify SSBE /uː/ and /ʊ/ as distinct, and assimilated SSBE /uː/ and /ʊ/ to Urdu /uː/ (34%; 4.8) and (30%; 5.0), respectively. However, there was a slight overlap for the perceptual assimilation of SSBE /uː/ and /ʊ/ with Urdu vowel /oː/ (19%; 4.8) and (29%; 5.0), as they were both also assimilated to Urdu /oː/.

For SSBE diphthongs, the assimilation patterns were similar to those in the bVd context, though had lower assimilation percentage, since SSBE vowels were more difficult to perceive in a hVba context and so confused with multiple vowels. However, SSBE /aɪ/, /ɔɪ/ and /aʊ/ were mapped to the expected Urdu response categories /ɑe/ (56%; 5.6), /oe/ (59%; 5.3), and /ɑʊ/ (70; 5.6), respectively. Surprisingly the percentage assimilation of the SSBE diphthongs /aɪ/, /ɔɪ/, and /aʊ/ are significantly higher than the rest of the vowels in the hVba context and are comparable to those in the bVd context.

In addition, the assimilation patterns of SSBE /eə/ with Urdu /eː/ (31%; 4.8) and /ɛː/ (24%; 5.1) can be considered *category-goodness* (CG) assimilation, where /ɛː/, with higher mean goodness rating (5.1), is probably a better Urdu exemplar for SSBE /eə/ than Urdu /eː/, with slightly lower mean goodness rating (4.8). Further, this CG pattern for SSBE /eə/ was also observed in the bVd context. However, the mean goodness ratings are higher for Urdu /eː/ (4.8) and lower for Urdu /ɛː/ (5.1) in the hVba than in the bVd context (/eː/ 4.6 and /ɛː/ 5.4). Future production and/or discrimination experiments can establish if these predictions will bear out. The classification overlap scores are discussed below.

### 5.8 Classification Overlap Scores in bVd and hVba

In order to see which SSBE vowels were the most confusing and/or difficult to perceive for Punjabi-Urdu speakers/listeners, the scores for classification overlap are computed (cf. Flege and MacKay, 2004) and a comparison of the classification overlap scores across both contexts (i.e. bVd and hVba) is given in Table 5.8.

Table 5. 8: The classification overlap scores percentages for SSBE vowels in bVd and hVba context (cf. Flege and MacKay, 2004)

| | bVd | | | hVba | |
|---|---|---|---|---|---|
| Stimuli | Response | Overlap | Stimuli | Response | Overlap |
| /ɪ/-/ɛ/ | /eː/ | 51% | /iː/-/ɪ/ | /iː/ | 12% |
| /ɛ/-/æ/ | /eː/ | 30% | /ɪ/-/ɛ/ | /eː/ | 34% |
| /ɪ/-/eɪ/ | /eː/ | 54% | /æ/-/ɑː/ | /ɑː/ | 13% |
| /ɛ/-/eɪ/ | /eː/ | 51% | /ɪ/-/eɪ/ | /eː/ | 34% |
| /ɪ/-/ɪə/ | /eː/ | 54% | /ɛ/-/eɪ/ | /eː/ | 40% |
| /ɛ/-/ɪə/ | /eː/ | 51% | /ɪ/-/ɪə/ | /eː/ | 34% |
| /æ/-/eɪ/ | /eː/ | 30% | /ɛ/-/ɪə/ | /eː/ | 35% |
| /ɛ/-/eə/ | /eː/ | 45% | /ɪ/-/eə/ | /eː/ | 31% |
| /æ/-/ɪə/ | /eː/ | 30% | /ɛ/-/eə/ | /eː/ | 31% |
| /ɪ/-/eə/ | /eː/ | 45% | /ʌ/-/ɜː/ | /ɑː/ | 14% |
| /æ/-/eə/ | /eː/ | 30% | /ʌ/-/ɜː/ | /ɑʊ/ | 16% |
| /ɜː/-/ɑː/ | /ɑː/ | 25% | /ʌ/-/ɜː/ | /eə/ | 11% |
| /ʌ/-/ɜː/ | /ɐ/ | 15% | /ɒ/-/ɔː/ | /oː/ | 60% |
| /ɒ/-/ɔː/ | /oː/ | 40% | /ʊ/-/uː/ | /uː/ | 30% |
| /ʊ/-/uː/ | /oː/ | 17% | /ʊ/-/uː/ | /oː/ | 19% |
| /ʊ/-/uː/ | /ʊ/ | 14% | | | |

An overlap below 25% suggests that discrimination of these contrasts will be relatively easy for Punjabi-Urdu speakers/listeners. On the other hand, a high classification overlap suggests that discrimination of those contrasts will be poor (Best, 1995; Faris et al., 2016). In bVd context, very little classification overlap was seen for SSBE contrasts /uː/-/ʊ/ (14%) and /ʌ/-/ɜː/ (15%). However, no classification overlap was seen for the SSBE contrast /iː/-/ɪ/.  In the hVba context, as can be seen in Table 5.8, SSBE /ʌ/-/ɜː/ were matched to three different Urdu vowels with different frequencies, but never with Urdu /ɐ/. This shows that Punjabi-Urdu speakers/listeners found these vowels the most confusing.

The selection of categories (response vowels) in a widely *dispersed* manner, with little overlap but a lot of categories in hVba, suggest that listeners can establish new categories for these vowels (Flege, 1995). However, the current study is limited as there are no discrimination or production data to test these predictions. These *dispersed* categorisation results are in line with PAM-L2 (Best and Tyler, 2007) and SLM (Flege, 1995) predictions that new phonological categories will be established for these highly *confused*/*dispersed* assimilations because there will be no interference from L1 attunement. In addition, *dispersed* assimilation patterns suggest that listeners failed "to detect clear higher-order phonological category invariants" (Faris et al., 2016:5).

The effect of context was found to be significant as discussed in Section 5.5.  As shown in Figure 5. 12, the mean duration for SSBE monophthongs in monosyllabic bVd vs. disyllabic hVba appears to be significantly different and could be the reason for variations in the patterns of perceptual assimilation. However, the differences in mean duration for SSBE diphthongs (as shown in Figure 5. 13), in monosyllabic bVd vs. disyllabic hVba, did not affect the patterns of perceptual assimilation, in particular for SSBE diphthongs /aɪ/, /aʊ/ and /ɔɪ/ (see Sections 5.6.5 and 5.7.5).

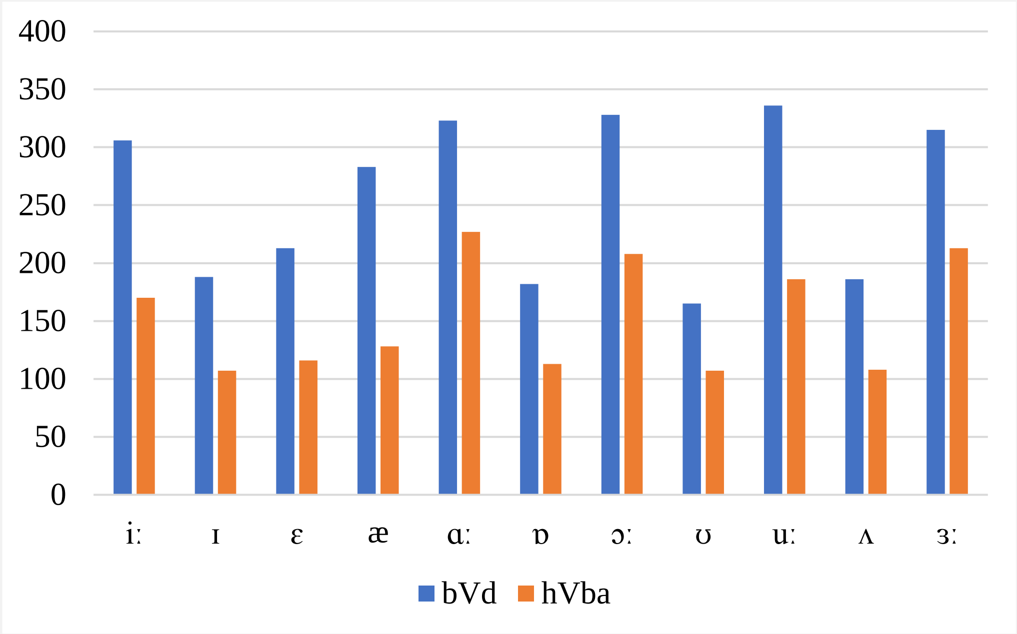



Figure 5. 12: Mean duration (ms) for SSBE monophthongs in a bVd and hVba context

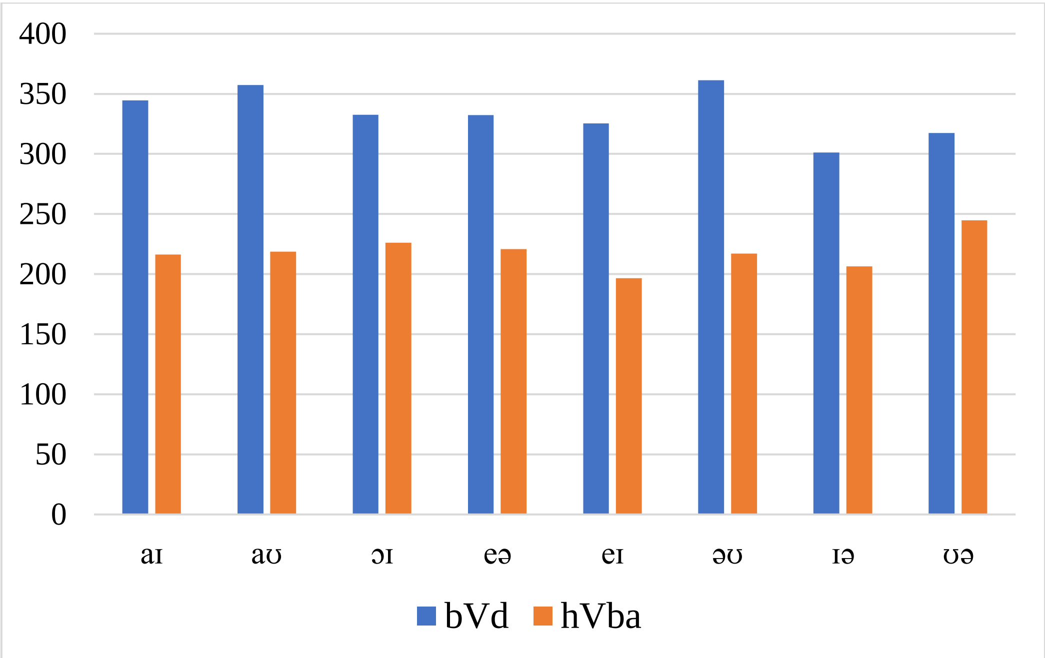



Figure 5. 13: Mean duration (ms) for SSBE diphthongs in a bVd and hVba context

The above discussed results are partially in line with the impressionistic description of Pakistani English (PE; as discussed in Section 4.1). According to the literature, in Pakistani English speakers make a distinction between both front and back tense and lax vowels (i.e. /iː/-/ɪ/ and /uː/-/ʊ/) (Sheikh, 2012; Bilal et al., 2011a and 2011b; Mahboob and Ahmar, 2004; Raza, 2008; Rahman 1997 and 2015). However, the results from the present study only show this in the bVd context. With regard to SSBE back and central vowels and diphthongs, the results are comparable to previous impressionistic and acoustic descriptions of PE (Hassan and Imtiaz, 2015; Bilal et al, 2011c; Raza, 2008; Rahman, 1997 and 2015). In the hVba context, for SSBE tense and lax vowels (/iː/-/ɪ/ and /uː/-/ʊ/), and for the remaining SSBE vowels in both contexts, the results are in line with the description of Asian Englishes (Bolton, 2008; Deterding, 2007; Kachru, 2005) and Indian English (Maxwell and Fletcher, 2009, 2010; Gargesh, 2004).

### 5.9 Conclusion

The results show that in the hVba context listeners were very confused - they confused all the SSBE vowels with multiple Urdu vowels. The option "none" was used more often in this context than in the bVd context. There are two possible reasons for that. Punjabi-Urdu speakers/listeners were told that these words/sounds are not from English but from a different European language that they do not know. This was done intentionally to see if their perceptual assimilation patterns vary if they can or cannot recognise the vowels as English.

Another possible explanation is due to the SSBE vowels being embedded in disyllabic words. Even though participants were given clear instructions and a practice session to emphasise that they should attend to the first vowel in these words, it is possible that when listening to the 114 tokens their focus shifted. In the hVba context they even confused SSBE monophthongs with Urdu vowel-vowel sequences, such as SSBE /æ/ and /ɑː/ with

Urdu /ɑʊ/ (23%; 5.3) and (28%; 5.2), respectively, and SSBE /ʌ/ and /ɜː/ with Urdu /ɑʊ/ (21%; 5.2) and (16%; 5.1), respectively.

Assimilation of SSBE vowels to multiple Urdu vowel categories shows that listeners were not sensitive to the phonetic similarity of these SSBE vowels to the phonological categories of Urdu (i.e. vowels). Therefore, it can be predicted that this insensitivity to the phonetic similarity of SSBE vowels to Urdu phonological categories will result in poor discrimination of these vowels by Punjabi-Urdu speakers/listeners (Faris et al., 2016). These experiments focussed on the perceptual assimilation patterns for SSBE vowels individually and not in contrasting pairs and so testing these predictions is left for future work.

The two-category assimilation patterns here suggest that listeners were sensitive to phonetic details with regard to L1 phonological categories. However, the multiple-category assimilation patterns show that listeners paid attention only to the phonetic details. In future it would be better to do the same experiments with nonsense words monosyllabic, to eliminate the effect of the syllabic context as a potential source of confusion. In addition, it would also be better to run the same tests with the response categories from Punjabi in order to see how two languages (L1s) affect learners' perception of an L2/foreign language.

Punjabi-Urdu speakers' superior perception of SSBE diphthongs /ɑɪ/, /ɔɪ/ and /ɑʊ/ in both contexts, suggests that these diphthongs are found in Urdu. They were each matched to the phonetic and phonological properties of Urdu phonemes. However, their high percentage of assimilation of SSBE /eɪ/ to Urdu /eː/ suggests that they only focused on the first part of this vowel. This diphthong does not exist in Urdu, so they did not have a counterpart to match with and confused this with a monophthong. Some of the patterns suggest that Punjabi-Urdu speakers were sensitive to duration, for example they identified

/i:/ correctly, but some of the findings (e.g. the assimilation patterns of SSBE central vowels) suggest that they were sensitive to both F1 and F2 - especially F2.

To conclude, overall the results show quite similar patterns of perceptual assimilation across both contexts; however, the biggest difference is in the percentage categorisation. In the bVd context, listeners chose fewer response categories with a higher percentage. In the hVba context listeners chose multiple response categories which resulted in overall lower percentages (see Table 5.6 and Table 5.7). The different assimilation patterns were found for the following vowels in both contexts:

The assimilation patterns for SSBE /æ/ were significantly different across the two contexts.

In the bVd context, SSBE /æ/ was assimilated to Urdu /eː/ (30%; 4.5), Urdu /ɛː/ (31%; 5.3), and Urdu vowel /ɑː/ (18%; 5.3). In the hVba context, SSBE /æ/ was assimilated to multiple Urdu vowels; however, the two most frequently chosen Urdu vowels were /ɑʊ/ (23%; 5.3) and /ɑː/ (16%; 5.0). These small percentages show that this vowel was confused with a number of Urdu vowels, and the five most frequently chosen Urdu vowel categories are shown in Table 5.7.

Assimilation patterns for SSBE central vowels /ʌ/ and /ɜː/ were also quite different in both contexts: In the bVd context, SSBE central mid-low vowel /ʌ/ was assimilated to Urdu /ɐ/ (44%: 5.4), Urdu /ɑː/ (17%: 5.0), and some other vowels; and SSBE central mid-high vowel /ɜː/ was assimilated to Urdu /ɑː/ (25%: 5.2), Urdu /ɐ/ (15%: 5.4), and Urdu /eː/ (14%: 4.6). On the other hand, in the hVba context, SSBE central mid-low vowel /ʌ/ and central mid-high vowel /ɜː/ were confused with multiple Urdu vowels, but never mapped to Urdu /ɐ/. These results show that in the hVba context Punjabi-Urdu listeners found these two central vowels very difficult as compared to the ones they heard in the bVd context, where they categorised /ʌ/ with Urdu /ɐ/ a high percentage of the time.

Among SSBE back vowels, in bVd context, SSBE /ɑː/ was assimilated to Urdu /ɑː/ (32%; 5.5); however, in the hVba context SSBE /ɑː/ was assimilated to Urdu /ɑʊ/ (28%; 5.2); Urdu /ɑː/ (13%; 4.9) and multiple other vowels. In addition, in the bVd context, SSBE /ʊ/ and /uː/ were mapped to Urdu /ʊ/ (40%; 5.5) and /uː/ (40%; 5.5), respectively. However, in the hVba context, both SSBE /ʊ/ (30%; 5.0) and /uː/ (34%; 5.8) were mapped to Urdu /uː/ respectively. These results show that inclusion of nonce word /bʊd/ in SSBE stimuli did not affect the perception of this vowel in bVd context.

These results show that in the hVba context Punjabi-Urdu speakers found SSBE /ɑː/ most confusing and mapped it to an Urdu vowel-vowel sequence. Similarly, listeners did not categorise SSBE /ʊ/ and /uː/ as distinct vowels in the hVba context, even though they have their counterparts in Urdu, though listeners did detect this distinction in the bVd context. The assimilation patterns for other monophthongs and diphthongs were quite similar across both contexts, as shown in Table 5.6 and Table 5.7. Lastly, the percentage classification overlap (see Section 5.8) suggests that discrimination will be easier for the vowels in the hVba context, and new categories can be established for the dispersed-uncategorised SSBE vowels (Faris et al., 2016; Flege, 1995); however only future discrimination and production experiments can verify these predictions. That is, it would be interesting to investigate whether the relative discrimination of the SSBE contrasts is consistent with PAM's predictions, and according to SLM, whether new categories can be formed for the SSBE vowels that are different phonetically from the closest Urdu vowels.

As discussed in Section 4.4.3, an auditory free classification experiment was also conducted to investigate the perceptual similarity spaces for 19 SSBE vowels in a hVd context. The experiment design and results are discussed in the following chapter (Chapter 6).

# Chapter 6

## Auditory Free Classification

This chapter reports on the auditory free classification of SSBE vowels by Punjabi-Urdu speakers. This study was used in order to mitigate the effects of association of oral stimuli with the written (orthographic) words (as can be seen in cross-language perceptual assimilation tasks) and also to investigate the perceptual similarity of SSBE vowels to each other. Free classification provides listeners with the freedom to classify the stimuli without the limitations of experimenter-defined category labels or specific dimensions of contrasts such as short vs. long. Contrary to traditional perceptual assimilation, categorization and discrimination experiments, auditory free classification is more flexible and effective as it allows the participants to categorise the stimuli (here, vowels) in high-dimensional space (Clopper, 2008). As a result, the experimenter can explore these categorisation strategies to better understand the perceptual dimensions and similarity across stimuli.

> "Free classification allows the experimenter to explore the complex interaction of perceptual cues to linguistic and indexical categories without requiring any priori judgments of what the relevant acoustic-phonetic cues or their weightings might be" (Clopper, 2008:575).

The free classification paradigm has proved popular in psychology and, more recently, in socio-phonetics; e.g. Clopper and Pisoni (2007) employed it for the perceptual classification of regional dialects; Bradlow et al. (2010) employed it for the perceptual classification of different languages based on their phonetic similarity or dissimilarity; Atagi and Bent (2013) employed it to investigate the perception of non-native speech. Clopper and Pisoni (2007) and Atagi and Bent (2013) used sentences as stimuli. Bradlow et al. (2010) used a sample of recording of "North Wind and the Sun" in four languages:

Dutch, Korean, Mandarin and Turkish, available from the website of the International Phonetic Association (IPA).

In the present study, 19 SSBE vowels were used in /hVd/ context, so the participants heard words in isolation. The main objective of this study was to investigate the following research questions:

a) Is the perceptual distance between SSBE vowels comparable with and predictable from cross-language perceptual assimilation patterns as observed in Chapter 5?

b) What acoustic dimensions do listeners pay most attention to: F1, F2, formant movement (diphthongization), or duration?

c) What is the role of the Urdu vowel system in determining the perceptual similarity of SSBE vowels?

## 6.1 Predictions

As discussed above, the specific patterns for auditory free classification (i.e. the clustering patterns for the SSBE vowels) cannot be predicted or predefined. Based on the general predictions as discussed in Section 4.5 and findings from the multiple forced choice cross-language assimilation and goodness rating task in a bVd context, reported in Section 5.6, the following broader predictions can be made. Firstly, participants will be able to differentiate between short and long vowels. Secondly, participants will be able to differentiate the central vowels from front and back vowels. Thirdly, participants will be able to differentiate monophthongs from diphthongs.

## 6.2 Experiment Design

### 6.2.1. Participants

The participants who participated in the perceptual assimilation task, as discussed in Section 5.2, also participated in the free classification task. The order of the tasks was counterbalanced across all participants and sessions were conducted on two separate days. This means that some of the participants took part in the free classification task before the perceptual assimilation and category goodness rating task and vice versa. Due to hardware failure, data from two participants could not be retrieved for the free classification task. Hence the results are based on the data collected from 68 participants.

### 6.2.2. Stimuli

The talkers were the same who recorded stimuli for the perceptual assimilation task (Chapter 5), i.e. two talkers (a male and a female, age range, 25-35 and 35-45) who spoke Standard Southern British English as a native language. Each talker produced three randomized blocks of the 19 SSBE vowels, i.e. 11 monophthongs /iː/, /ɪ/, /ɛ/, /æ/, /ɜː/, /ʌ/, /ɑː/, /ɒ/, /ɔː/, /ʊ/, and /uː/ and 8 diphthongs /ɪə/, /eə/, /ʊə/, /eɪ/, /aɪ/, /ɔɪ/, /əʊ/, and /aʊ/, in a hVd context. The recording procedures were identical to the ones discussed in Section 5.2.2. One token was selected for each vowel from each talker, i.e. one from the male talker and one from the female talker, from three repetitions of each SSBE vowel in a hVd context. The test words are given in Table 6. 1.

## 6.3 Acoustic Analysis of Stimuli

The methods for the formant extraction, and acoustic comparisons of SSBE vowels with RP vowels reported by Hawkins and Midgley (2005), were the same as discussed in Section 5.2.3. The mean formant frequencies of the first two formants are plotted in F1/F2 vowel space in Figure 6. 1, for both SSBE vowels in a hVd context and RP vowels produced by a group of male speakers (20-25 years old) as reported in Hawkins and Midgley (2005). The stimuli used for the free classification experiment in a hVd context appear to be more similar to the RP vowels (as reported in Hawkins and Midgley (2005)) than those in the bVd and hVba context (see Section 5.2.3). Hawkins and Midgley (2005) reported the formant frequencies of the 11 RP monophthongs in a hVd context, which explains why the stimuli in the hVd context are a closer match. The mean F1 and F2 frequencies and duration in the hVd context are given in Appendix 6A.

Table 6. 1: Test words (after Ferragne and Pellegrino, 2010)

| Test word | Lexical set |
|---|---|
| heed | FLEECE |
| hid | KIT |
| head | DRESS |
| had | TRAP |
| hard | START |
| hod | LOT |
| hoard | FORCE |
| hood | FOOT |
| who'd | GOOSE |
| Hudd | STRUT |
| heard | NURSE |
| hade | FACE |
| hide | PRICE |
| hoid | CHOICE |
| hoed | GOAT |
| howd | MOUTH |
| hared | SQUARE |
| heered | NEAR |
| hured | CURE |

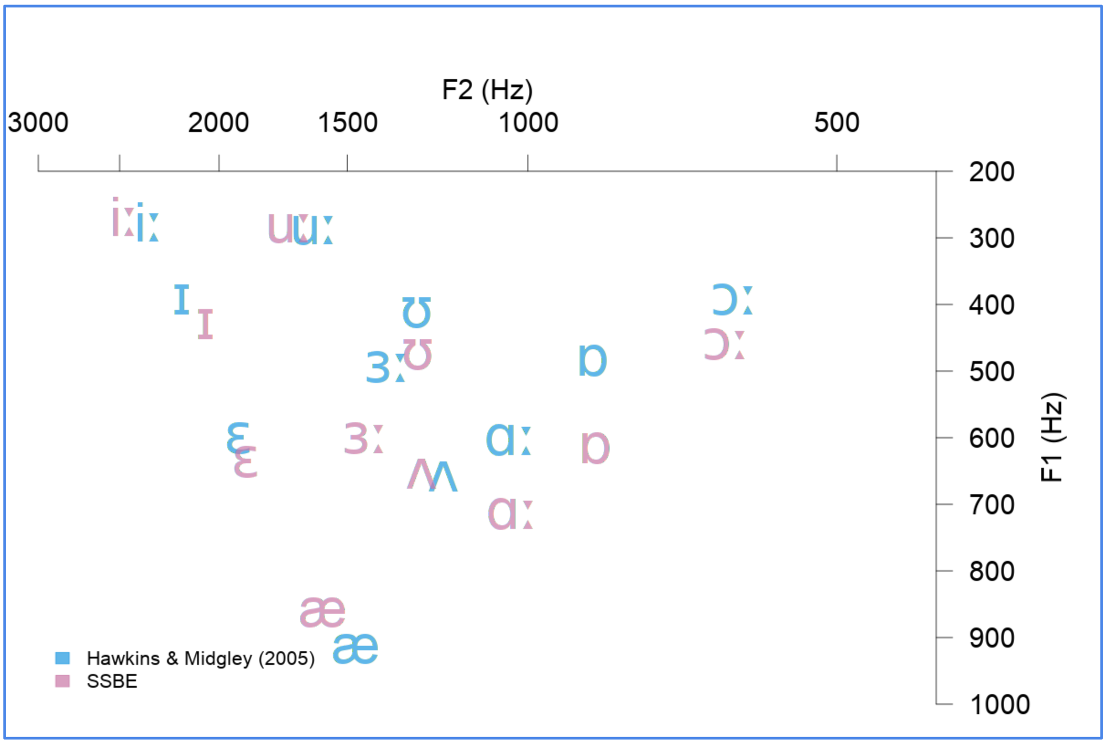



Figure 6. 1: Mean frequencies of the 11 SSBE monophthongs in a hVd context along with RP from Hawkins and Midgley's (2005) youngest age group in F1/F2 vowel space

### 6.4. Procedures

### 6.4.1 Listening Task - Auditory Free Classification

For the experiment, the stimuli were presented in PowerPoint with the interactive Drag and Drop AddIn (PPTAlchemy, 2016). All participants were tested in a quiet, sound-attenuated room in the Department of English Language and Literature and in a computer lab in the Institute of Communication Studies at the University of the Punjab, Lahore, Pakistan. The participants were given access to a computer (laptop) and Sennheiser HD 650 headphones to listen to a set of English sounds in a hVd context. The computer screen initially shows the instructions for the free classification task, as shown in Figure 6. 2. On clicking to move to the next page, the participants were presented with 38 light-blue squares with coded labels, each representing a sound, arranged in three columns on the left, plus a 10 x 10 cell grid in grey on the right, as shown in Figure 6. 3.

**Screen 1**

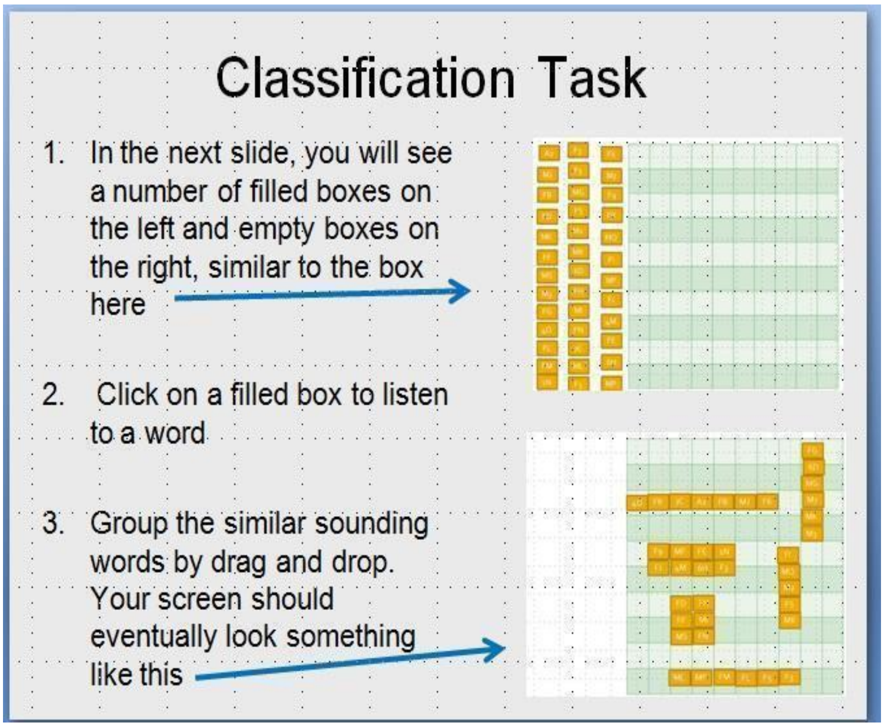



Figure 6. 2: Screenshot from the auditory free classification

**Screen 2**

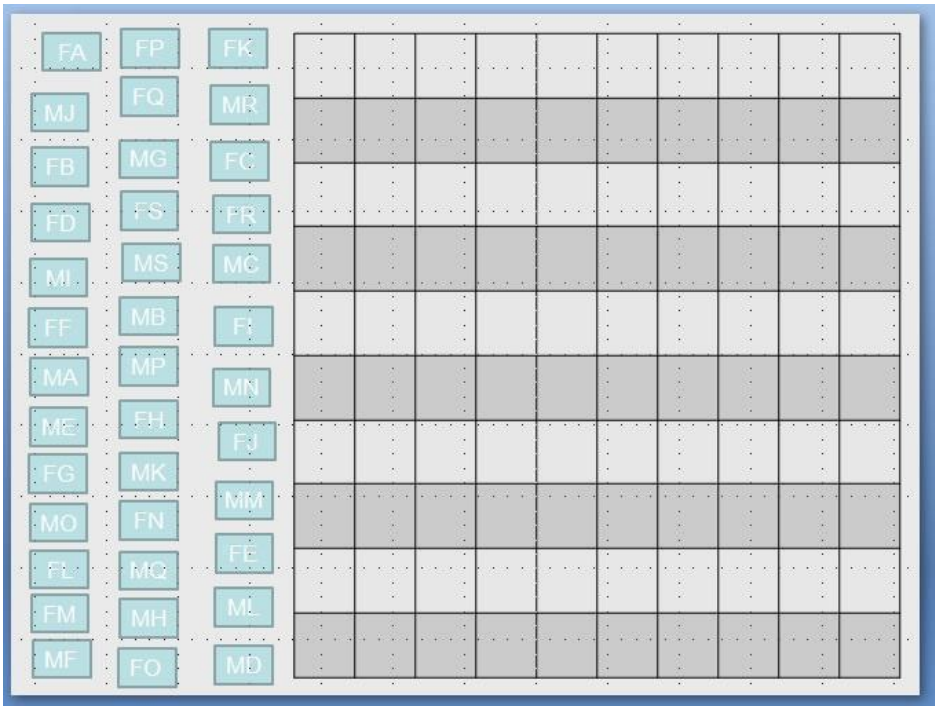



Figure 6. 3: Screenshot from the auditory free classification

The participants could listen to each sound file by clicking on the blue square once with the right mouse button. They could move and drag the light-blue squares around the screen once clicked. The participants were asked to group the similar sounding words by drag and drop. They were reminded to focus on the vowels only, as the initial and final consonants are the same across all 38 tokens, i.e. hVd. The listeners were allowed to make as many groups as they wished with as many words as they thought contained similar vowels. They were not required to put the same number of words in each group and they could listen and move the words around as many times as they wished. There was no time-limit set for this procedure. The participants were asked to inform the experimenter when they were finished. Typically, participants spent 15-20 minutes to complete the task; however, a couple of the participants took up to 60 minutes on the task.

## 6.5 Analysis

The first step was to arrange the data in square similarity *or proximity* matrices (Wickelmaier, 2003). The proximity matrices encode the dissimilarity between any two points. To this end, the groups created by each participant were converted into (symmetric) 38 by 38 matrices (19 SSBE vowels × 2 talkers, a male and female = 38), where each row/column represents a vowel, and each cell encodes the proximity of the row vowel to the column vowel. As shown in Table 6. 2, for each participant a 1 was placed in each cell that corresponds to a pair of vowel tokens that were grouped together. All other cells were set to zero (see Appendix 6B for an example). The overall proximity matrix is then the sum of all the individual participants' matrices and is given in Appendix 6C.

Statistical analysis is based on this aggregate similarity matrix using Hierarchical Cluster Analysis (HCA) and Multidimensional Scaling (MDS) analysis, which was conducted using R (R Core team, 2018) with *isoMDS* function from the *MASS* package (Venables and Ripley, 2002). Hierarchical clustering employs an iterative pairwise distance calculation, joining the two most similar objects (i.e. sounds in this case) at each iteration. The clusters form a tree pattern, highlighting the most similar objects (Bradlow et al., 2010). Following Clopper and Bradlow (2009), the relationship between acoustic properties of the stimulus materials and perceptual structure (i.e. clusters) was explored.

MDS is a multivariate data analysis approach that builds a map from distances, projecting points into a space in two, three or more dimensions. This representation of distances between objects can help to identify underlying dimensions, such as the perceptual similarity space in the present study. In three or fewer dimensions, MDS also has the advantage of providing a visual representation, highlighting perceptual clusters. The number or nature of perceptual dimensions are not predefined prior to the MDS analysis (Fox, 1983:27).

Table 6. 2: Matrix representation of clustered pairs for one participant. Empty cells represent zeros.

| | FA | FB | FC | FD | FE | FF | FG | FH | FI | FJ | FK | FL | FM | FN | FO | FP | FQ | FR | FS | MA | MB | MC | MD | ME | MF | MG | MH | MI | MJ | MK | ML | M | MN | MO | MP | MQ | MR | MS |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| FA | 1 | 1 | | 1 | | | | | | | | | | | | 1 | | | | | | | 1 | | | | | | | | | | | | 1 | | | 1 |
| FB | 1 | 1 | | 1 | | | | | | | | | | | | 1 | | | | | | | 1 | | | | | | | | | | | | 1 | | | 1 |
| FC | | | 1 | | | 1 | | | | | 1 | | | | | | | | | | | | | | | | 1 | 1 | | | | | 1 | | | 1 | | |
| FD | 1 | 1 | | 1 | | | | | | | | | | | | 1 | | | | | | | 1 | | | | | | | | | | | | 1 | | | 1 |
| FE | | | | | 1 | | 1 | | | | | | | | | | | | | | | | | 1 | | 1 | | | | | | | | | | | | |
| FF | | | 1 | | | 1 | | | | | 1 | | | | | | | | | | | | | | | | 1 | 1 | | | | | 1 | | | 1 | | |
| FG | | | | | 1 | | 1 | | | | | | | | | | | | | | | | | 1 | | 1 | | | | | | | | | | | | |
| FH | | | | | | | | 1 | | | | | | | | | | | | | | | | | | | | | | | 1 | | | | | | 1 | |
| FI | | | | | | | | | 1 | | | | | | 1 | | | | | | | | | | | | | | | 1 | | | | 1 | | | | |
| FJ | | | | | | | | | | 1 | | | 1 | | | | | | | | | | | | | | | | 1 | | | 1 | | | | | | |
| FK | | | 1 | | | 1 | | | | | 1 | | | | | | | | | | | | | | | | 1 | 1 | | | | | 1 | | | 1 | | |
| FL | | | | | | | | | | | | 1 | | 1 | | | | 1 | | | 1 | | | | 1 | | | | | | | | | | | | | |
| FM | | | | | | | | | | 1 | | | 1 | | | | | | | | | | | | | | | | 1 | | | 1 | | | | | | |
| FN | | | | | | | | | | | | 1 | | 1 | | | | 1 | | | 1 | | | | 1 | | | | | | | | | | | | | |
| FO | | | | | | | | | 1 | | | | | | 1 | | | | | | | | | | | | | | | 1 | | | | 1 | | | | |
| FP | 1 | 1 | | 1 | | | | | | | | | | | | 1 | | | | | | | 1 | | | | | | | | | | | | 1 | | | 1 |
| FQ | | | | | | | | | | | | | | | | | 1 | | | | | 1 | | | | | | | | | | | | | | | | |
| FR | | | | | | | | | | | | 1 | | 1 | | | | 1 | | | 1 | | | | 1 | | | | | | | | | | | | | |
| FS | | | | | | | | | | | | | | | | | | | 1 | 1 | | | | | | | | | | | | | | | | | | |
| MA | | | | | | | | | | | | | | | | | | | 1 | 1 | | | | | | | | | | | | | | | | | | |
| MB | | | | | | | | | | | | 1 | | 1 | | | | 1 | | | 1 | | | | 1 | | | | | | | | | | | | | |
| MC | | | | | | | | | | | | | | | | | 1 | | | | | 1 | | | | | | | | | | | | | | | | |
| MD | 1 | 1 | | 1 | | | | | | | | | | | | 1 | | | | | | | 1 | | | | | | | | | | | | 1 | | | 1 |
| ME | | | | | 1 | | 1 | | | | | | | | | | | | | | | | | 1 | | 1 | | | | | | | | | | | | |
| MF | | | | | | | | | | | | 1 | | 1 | | | | 1 | | | 1 | | | | 1 | | | | | | | | | | | | | |
| MG | | | | | 1 | | 1 | | | | | | | | | | | | | | | | | 1 | | 1 | | | | | | | | | | | | |
| MH | | | 1 | | | 1 | | | | | 1 | | | | | | | | | | | | | | | | 1 | 1 | | | | | 1 | | | 1 | | |
| MI | | | 1 | | | 1 | | | | | 1 | | | | | | | | | | | | | | | | 1 | 1 | | | | | 1 | | | 1 | | |
| MJ | | | | | | | | | | 1 | | | 1 | | | | | | | | | | | | | | | | 1 | | | 1 | | | | | | |
| MK | | | | | | | | | 1 | | | | | | 1 | | | | | | | | | | | | | | | 1 | | | | 1 | | | | |
| ML | | | | | | | | 1 | | | | | | | | | | | | | | | | | | | | | | | 1 | | | | | | 1 | |
| MM | | | | | | | | | | 1 | | | 1 | | | | | | | | | | | | | | | | 1 | | | 1 | | | | | | |
| MN | | | 1 | | | 1 | | | | | 1 | | | | | | | | | | | | | | | | 1 | 1 | | | | | 1 | | | 1 | | |
| MO | | | | | | | | | 1 | | | | | | 1 | | | | | | | | | | | | | | | 1 | | | | 1 | | | | |
| MP | 1 | 1 | | 1 | | | | | | | | | | | | 1 | | | | | | | 1 | | | | | | | | | | | | 1 | | | 1 |
| MQ | | | 1 | | | 1 | | | | | 1 | | | | | | | | | | | | | | | | 1 | 1 | | | | | 1 | | | 1 | | |
| MR | | | | | | | | 1 | | | | | | | | | | | | | | | | | | | | | | | 1 | | | | | | 1 | |
| MS | 1 | 1 | | 1 | | | | | | | | | | | | 1 | | | | | | | 1 | | | | | | | | | | | | 1 | | | 1 |

Fox (1983) reported four-dimensional scaling for the perceptual structure of 15 American English vowels; however, the data were not collected via a free classification task. Listeners had to listen to pairs of vowels in a hVd context and rate their similarity on a nine-point scale. The present study is different from Fox (1983) in both design and stimuli. The tokens were not played in pairs to make judgments on a scale. Fox (1983) reported the findings from volunteers who spoke the Midwestern dialect of American English. Despite the differences in experiment paradigms, Fox (1983) presented an interesting analysis of perceptual correlations and acoustic structure using MDS analysis, which inspired the analysis in the present study (see Section 6.6.2).

## 6.6 Results

### 6.6.1 Hierarchical Clustering Analysis

The number of clusters across participants ranged from 4 to 22 with a median 11 and mean 12. The sizes of clusters (the number of vowels in each cluster) per group ranged from 1 to 15 per group with a median of 2 and mean 3 as shown in Table 6. 3.

Table 6. 3: Summary of number of clusters and sizes of clusters

| | Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max |
|---|---|---|---|---|---|---|
| Number of clusters | 4 | 8 | 11 | 12 | 16 | 22 |
| Sizes of clusters | 1 | 2 | 2 | 3 | 4 | 15 |

The similarity matrix was submitted to the R function *hclust* using *complete linkage* (R Core Team, 2018) clustering criteria, where "...this method defines the distance between two groups as the distance between their two farthest-apart members. This method usually yields clusters that are well separated and compact" (NCSS Statistical Software, 2017:445-3). The dendrogram produced from this analysis is shown in Figure 6.4. In a dendrogram each branching point (horizontal line) is called a clade and the final end points (individual vowels, in this case) are called leaves. Each clade can have zero or multiple branches, though when plotting the results of hierarchical clustering as employed here, each clade will only ever have two branches. The arrangement of the leaves shows those that are most similar to each other. The height of the clades indicates how similar the leaves or groups that branch off from the clade are from each other: the greater the height, the greater the difference (Wheaton College Massachusetts, 2016).

The *Complete Linkage* method was used because it has the largest *cophenetic correlation coefficient* (goodness of fit), i.e. 0.85 for monophthongs and diphthongs together and 0.92 for only monophthongs. Other methods had lower than 0.75 *cophenetic correlation coefficient*. This goodness of fit measure is based on "... the correlation between the original distances and those that result from the cluster configuration" (NCSS Statistical Software, 2017:445-4), that is, it is a measure of how well the dendrogrammatic distance correlates with the source similarity matrix The *cophenetic correlation coefficient* was calculated using the *cophenetic* function of the *stats* package in R (R Core Team, 2018).

The clustering solutions were obtained for monophthongs and diphthongs combined, as well as separately, as shown in Figure 6. 4, Figure 6. 5 and Figure 6. 6. In the dendrogram in Figure 6. 4, the perceptual similarity is indexed by the least number of vertical branches connecting any two nodes. The horizontal distance is not relevant. The clustering analysis revealed two main clusters, corresponding to *front* and *back* vowels. These two main clusters are further subdivided into three sub-clusters, which can correspond to high front, high-mid front, and back vowels. The third sub-cluster is further subdivided into two sub-clusters that correspond to high back vowels (/uː/ and /ʊ/) and low back vowels.

Interestingly, some of the diphthongs were grouped together: /aʊ/, /aɪ/ and /ɔɪ/; and /ɪə/ and /ʊə/. However, /eɪ/, /eə/ and /əʊ/ were grouped with monophthongs. In addition, it can be seen from the dendrogram in Figure 6. 4 that /aʊ/, /aɪ/, /ɔɪ/ and /əʊ/ are part of the cluster interpreted as back, and the other diphthongs /eɪ/, /ɪə/, /ʊə/ and /eə/ are part of the main cluster interpreted as front. A more detailed analysis for monophthongs, followed by diphthongs, is given in the following sections.

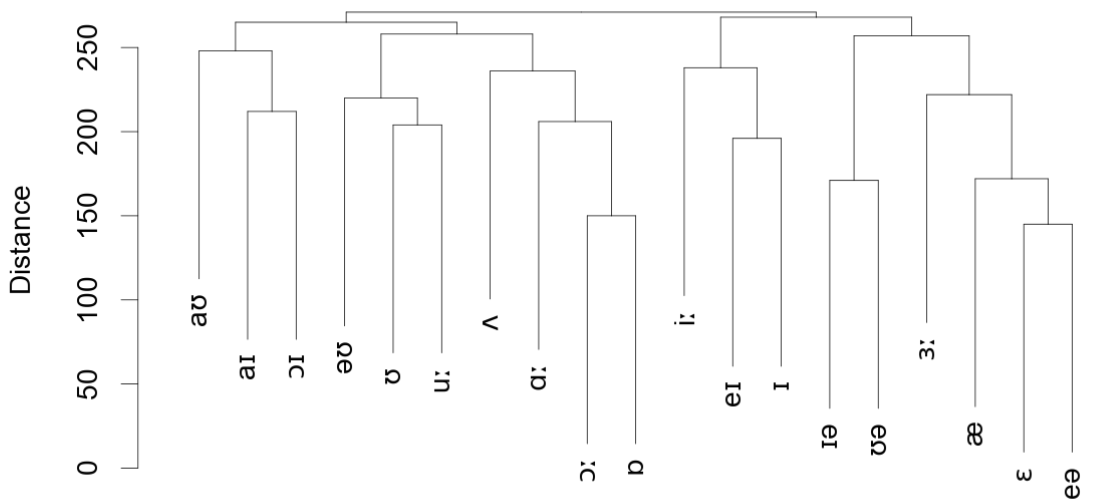



Figure 6. 4: Dendrogram of the hierarchical clustering of monophthongs and diphthongs

### 6.6.1.1 Monophthongs

The clusters shown in Figure 6. 4 and Figure 6. 5 can be interpreted as follows:

SSBE monophthongs /i:/ and /ɪ/ were grouped together. Similarly, /u:/ and /ʊ/ were grouped together. This suggests that listeners paid attention to spectral cues, in particular F2, since these front and back vowels are grouped in two different clusters. However, participants did not respond to temporal cues; hence long and short vowels are grouped together. SSBE monophthongs /ɛ/ and /æ/ were grouped together and perceived as much more similar to each other than they are to /ɜ:/. The dendrogram in Figure 6. 4 and Figure 6. 5 also shows that /ʌ/, /ɑ:/, /ɔ:/ and /ɒ/ are part of the third main cluster, where /ɔ:/ and /ɒ/ are perceptually closest, and /ʌ/ and /ɑ:/ were added to the cluster later.

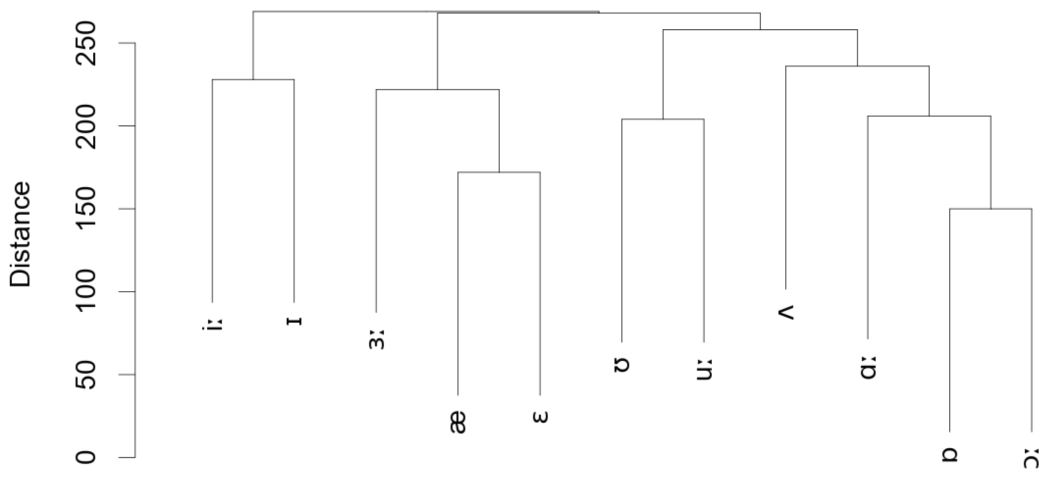



Figure 6. 5: Dendrogram of the hierarchical clustering of monophthongs

#### 6.6.1.2 Diphthongs

The clustering solution with all 19 SSBE vowels (i.e. including diphthongs), as shown in Figure 6. 4, shows similar patterns of perceptual similarity. In order to test if the presence or absence of diphthongs affect the perceptual space, future experiments can be designed to run separate tests for monophthongs and diphthongs. The present work suggests that the presence and absence of diphthongs does not influence the perceptual similarity/dissimilarity structure of monophthongs, as can be seen from Figure 6. 4 and Figure 6. 5. However, Figure 6. 6 shows a different clustering pattern for diphthongs. It must be noted that the in the present study, both monophthongs and diphthongs were presented together to the listeners, and so probably influenced the clustering choices.

As shown in Figure 6. 6, SSBE diphthong /aɪ/, /ɔɪ/, /əʊ/ and /aʊ/ were grouped in the same cluster. However, /aɪ/ and /ɔɪ/ are more similar to each other than /əʊ/ and /aʊ/, which were added to the cluster later. Although /ʊə/, /ɪə/, /eɪ/ and /eə/ form the second main cluster, /ʊə/ and /ɪə/ form a separate sub-cluster and appear to be more similar to each other than /eɪ/ is to /eə/, which form the second sub-cluster, as shown in Figure 6. 6.

Thus, Figure 6. 6 shows a different clustering solution for diphthongs alone than for the diphthongs and monophthongs together, shown in Figure 6. 4. SSBE diphthong /aɪ/ and /ɔɪ/ were grouped in the same cluster and /aʊ/ was added later on. /ʊə/ and /ɪə/ were part of that same cluster; however, they were more similar to each other than any other member of this group. /iː/, /ɪ/ and /eɪ/ were grouped with each other. /ɪ/ and /eɪ/ were more similar to each other than /iː/, which was added to this cluster later on. /əʊ/ was grouped with monophthongs /uː/ and /ʊ/ but appeared to be perceived differently from these two monophthongs. /eə/ was grouped with monophthongs /ɛ/ and /æ/ and appeared to be perceived as more similar to /ɛ/ than /æ/, which was added to the cluster later on.

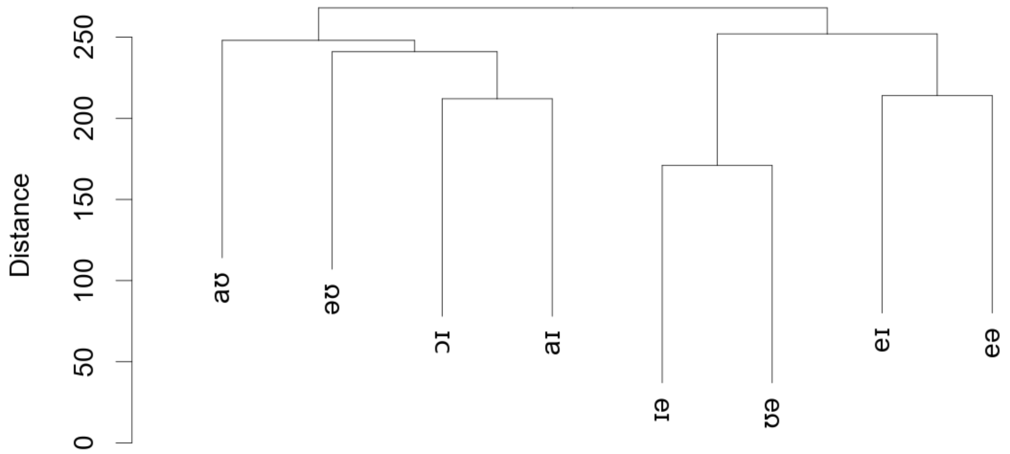



Figure 6. 6: Dendrogram of the hierarchical clustering of diphthongs

In order to explore the relationship between acoustic properties (F1, F2 and duration) of stimuli and the perceptual clusters, a *Generalised Linear Model* (GLM) model was fitted and analysed using the *glm* function of the *stats* package in R (R Core Team, 2018).

For each cluster, the three acoustic measures F1, F2, Duration, and talker Gender were entered as independent variables, and the cluster membership (Front or Back) for each vowel as the dependent variable (1 for members and 0 for non-members). In terms of an R formula we have

Front ~ F1 + F2 + Duration + Gender

The model was thus configured with a binomial error distribution and a logit link function in R (R Core Team, 2018).

Type II ANOVA tests were performed on the GLM model in R (R Core Team, 2018). The results revealed that F2 was a significant predictor of membership for each vowel in both front and back clusters (($\chi^2$ (1) = 14.7424, $p$ < 0.0001). The other two acoustic measures (i.e. F1 and Duration) and Gender were not significant. Hence these results suggest that listeners were more sensitive to F2 than F1 and duration to make their classification judgments.

### 6.6.2 Multidimensional Scaling Analysis

MDS analysis was also carried out to investigate the perceptual similarity of the SSBE vowels. A 19 by 19 matrix for all listeners was submitted to *isoMDS* from the MASS package of R (Venables and Ripley, 2002) for MDS analysis, as discussed in Section 6.5. A two-dimensional analysis (eventually interpreted as front/back and high/low) was selected for monophthongs. There were two reasons for the selection of a two-dimensional space: the two dimensions were highly interpretable, and the *stress* was significantly reduced from the one- to two-dimensional solution, but the reduction in stress from two- to three-dimensional solutions was relatively small, and smaller still as we increase the number of dimensions. This stress value gives a goodness-of-fit statistic for MDS analysis (Kruskal, 1964), which is based on the difference between actual distances and their predicted values.

A *scree* plot shows the stress values (*eigenvalues*: Smith, 2002) on the y-axis and the number of dimensions on the x-axis. The point where the slope of the curve starts levelling off (also known as "*the elbow*") (Clopper, 2008) indicates the number of dimensions that should be used to analyse the given data set. This stress value in the present study was calculated for the monophthongs and diphthongs combined as well as separately. A *scree* plot of the *stress* values obtained for each of the four MDS solutions for monophthongs and diphthongs combined is shown in Figure 6. 7, and for monophthongs and diphthongs separately is shown in Figure 6. 8 and Figure 6. 9 respectively.

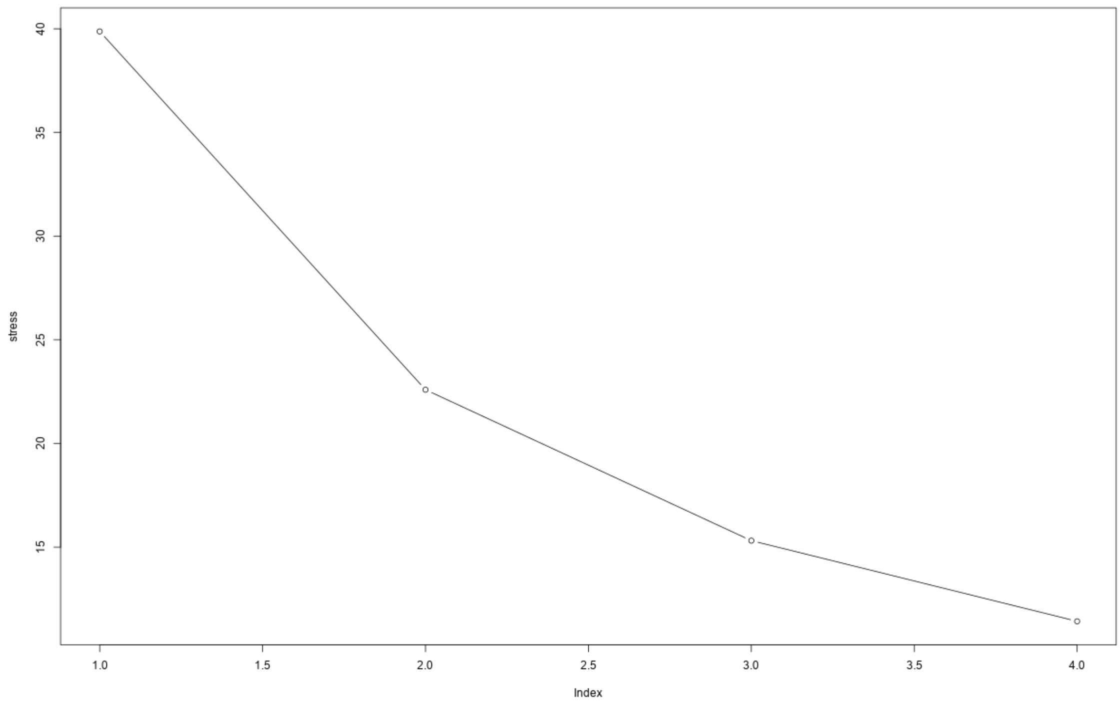



Figure 6. 7: Scree plot showing the stress for each MDS solution from one to four dimensions for monophthongs and diphthongs combined

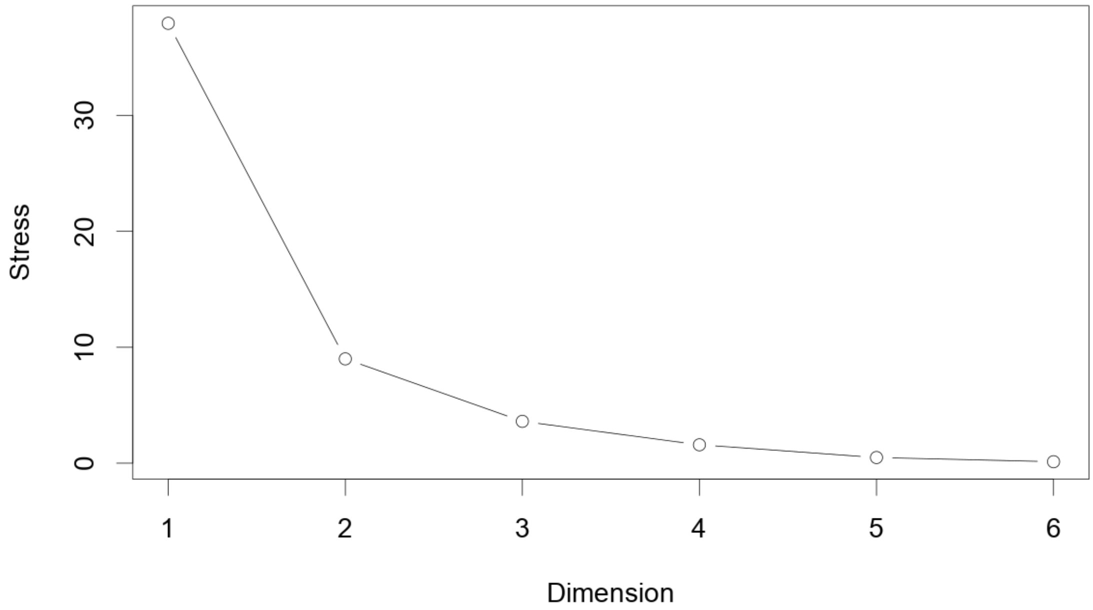



Figure 6. 8: Scree plot showing the stress for each MDS solution from one to six dimensions for monophthongs

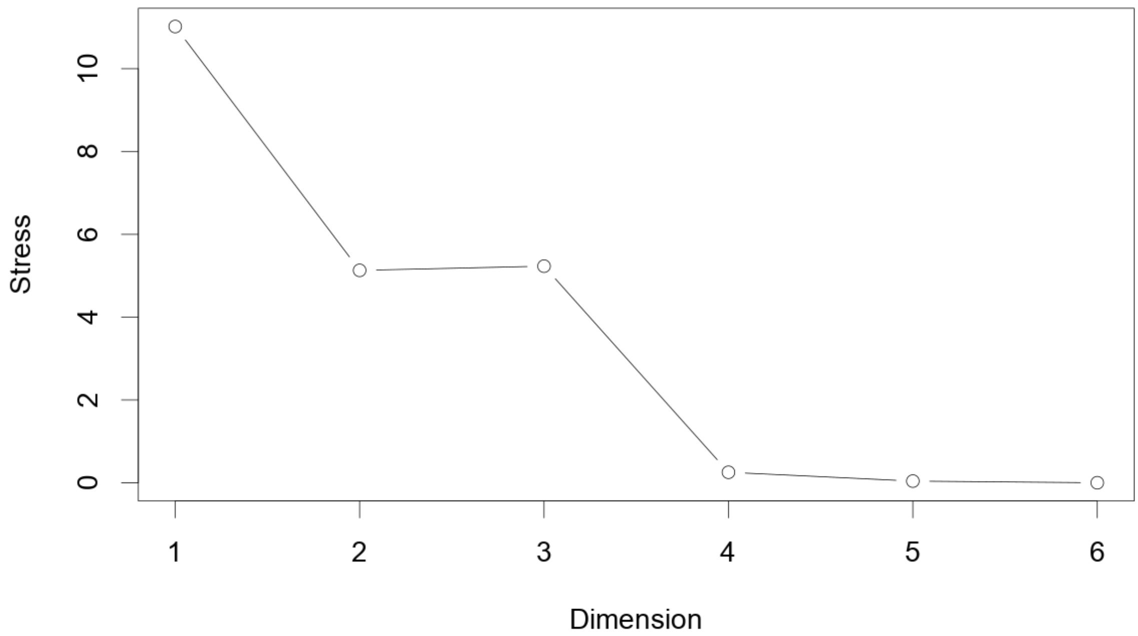



Figure 6. 9: Scree plot showing the stress for each MDS solution from one to four dimensions for diphthongs

The scree plot for diphthongs suggest that a four-dimensional solution should be chosen to interpret the underlying factors for perceptual classification of diphthongs. However, a four-dimensional solution was not obtained in the present study because dimension four was uninterpretable with regard to any standard linguistic interpretation (cf. Fox, 1983).

As shown in Figure 6. 10, Figure 6. 11 and Figure 6. 12, the first dimension of the two-dimensional solution separates the *high-front* from *non-high front* vowels. The second dimension separates the *high-back* vowels from *low-back* vowels. The two central vowels are also separated and grouped in these two dimensions: /ɜː/ is grouped with the non-high front vowels; /ɛ/ and /æ/; and /ʌ/ is grouped with the non-high back vowels; /ɒ/, /ɔː/ and /ɑː/. In this two-dimensional space (Figure 6. 10), the diphthongs are also grouped with front and back monophthongs as shown in hierarchical clustering (Figure 6. 4). As suggested by the *scree* plot, a three-dimensional solution for diphthongs is presented in Figure 6. 13.

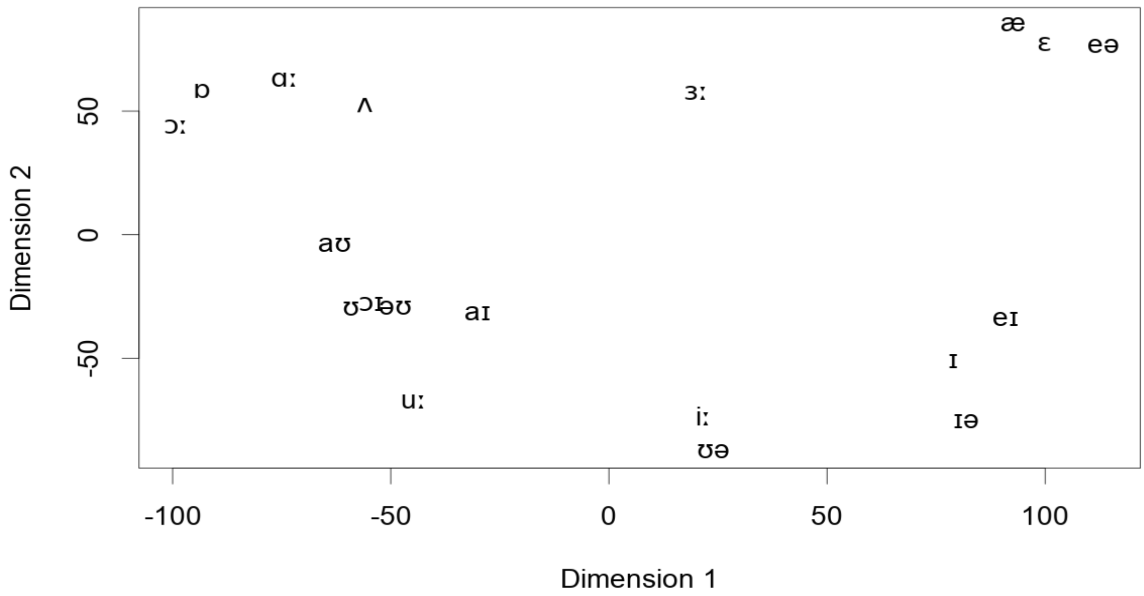



Figure 6. 10: Two-dimensional MDS of perceptual similarity for monophthongs and diphthongs

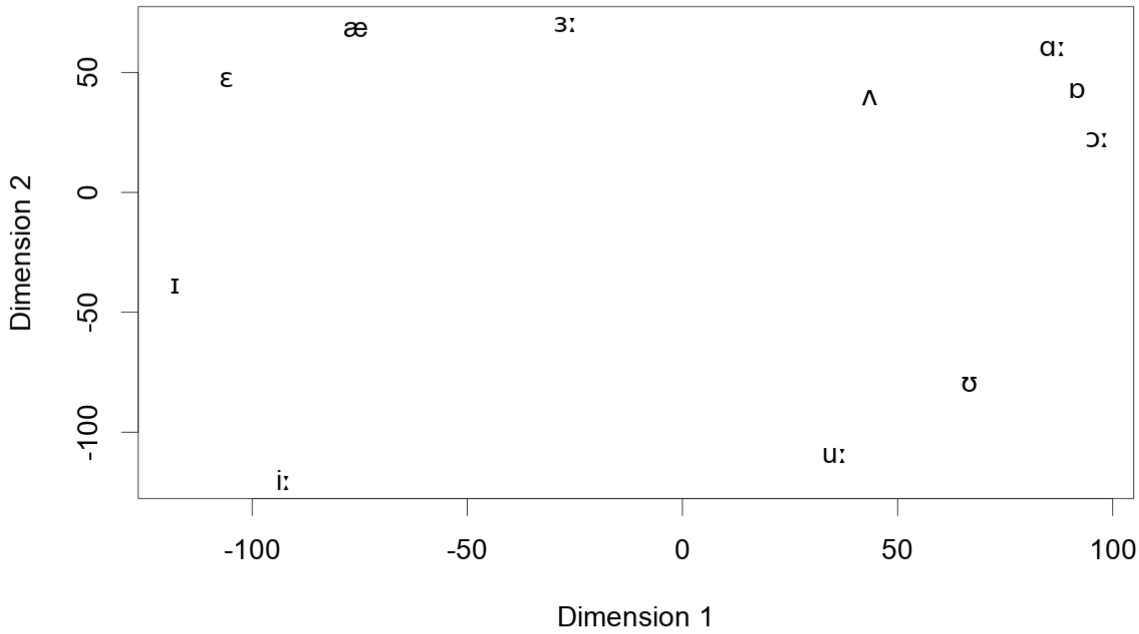



Figure 6. 11: Two-dimensional MDS of perceptual similarity for monophthongs

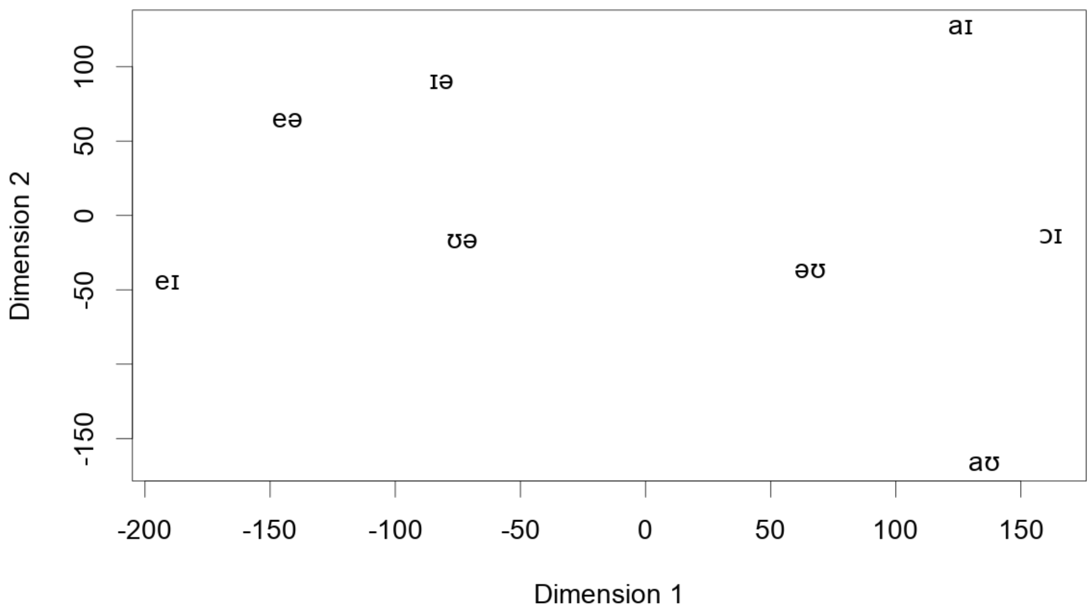



Figure 6. 12: Two-dimensional MDS of perceptual similarity for diphthongs

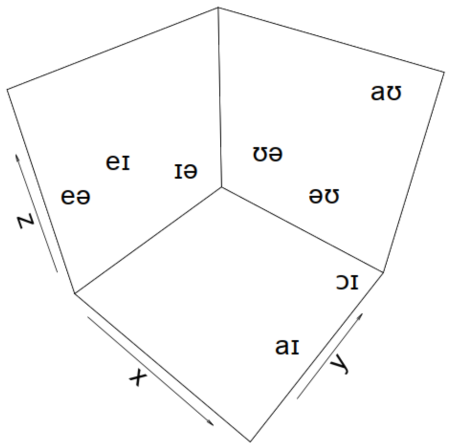



Figure 6. 13: Three-dimensional MDS of perceptual similarity for diphthongs

A three-dimensional solution for diphthongs, as shown in Figure 6. 13, can be interpreted as follows:

Dimension 1 in z-y plane, have diphthongs with close-mid front vowels as onset and central vowels as offset: /eə/, /eɪ/, and /ɪə/. Dimension 2 in x-y plane, have diphthongs with close-mid front vowels as offset and back vowels as onset: /aɪ/ and /ɔɪ/; and dimension 3 in x-z plane, have diphthongs with back and central vowels as onset and offset: /ʊə/, /əʊ/ and /aʊ/. Therefore, the first dimension can be interpreted as front/central, second dimension can be interpreted as back/low, and third dimension can be interpreted as back/high or back/central. Third dimension can also be interpreted as an extension of second dimension because the diphthongs in these two dimensions have low and high back vowels as well as front and central vowels as offset and onset respectively.

Although /ɪə/ and /ʊə/ are found in different dimensions, they appear to be quite distinct from the other members in their respective dimensions. Therefore, it was necessary to compute the MDS coordinates for the SSBE diphthongs in three-dimensional space to calculate the exact distances between the diphthongs in each dimension. These MDS coordinates are shown in Figure 6. 14 and the list of coordinates for each diphthong is given in Appendix 6D. As shown in Figure 6. 14, according to MDS coordinates in three-dimensional space, the following diphthongs can be grouped as they have coordinates in the same quadrant; hence can be considered more similar to each other than other diphthongs:

- First quadrant: /eə/ and /eɪ/ are found in the same quadrant;
- Second quadrant: /ɪə/ and /ʊə/ are found in the same quadrant but they differ with regard to y-axis coordinates;
- Third quadrant: /əʊ/ and /aʊ/ are found in the same quadrant; however, the x, y, z coordinates for /aʊ/ are significantly higher than /əʊ/;
- Fourth quadrant: /aɪ/ and /ɔɪ/ are approximately in the same quadrant with regard to x-axis coordinates; however, the y-axis and z-axis coordinates are very different, i.e. y-axis and z-axis are both negative and close to zero for /ɔɪ/, but y-axis is positive and z-axis is negative for /aɪ/.

These differences in coordinates suggest that their representation in three-dimensional space (as shown in Figure 6. 13) is only partially correct. For example, /ɪə/ and /ʊə/ were found in two different dimensions in three-dimensional space (Figure 6. 13). However according to their x, y, z coordinates (as shown in Figure 6. 14), /ɪə/ and /ʊə/ are found in the same quadrant and appear to be (perceived) as more similar to each other than any other diphthongs. Further, these findings can be interpreted that the diphthongs in these dimensions (quadrants) suggest that listeners paid more attention to the onset vowel in the first quadrant; however, they focused on the offset vowel in the second, third and fourth quadrant.

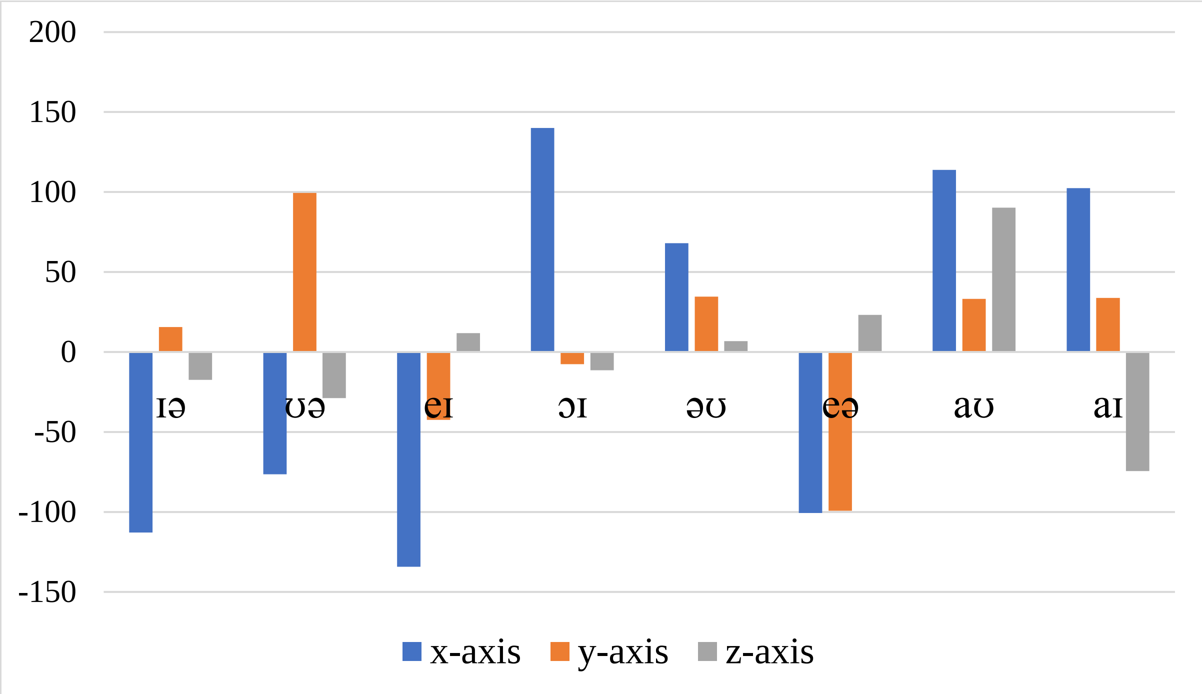



Figure 6. 14: MDS coordinates for each SSBE diphthong in three-dimensional space

Quantitatively, acoustic properties can be compared to the output of the MDS analysis, to establish a link between perceptual classification and physical attributes of the vowels. Pearson's r test for correlation was performed on F1, F2 and duration of the stimulus vowels against each dimension of the MDS analysis. The significance level of the correlations was then determined by calculating the t-value from Pearson's r-value. This was performed using the *cor.test* function in the *stats* package of R (R Core Team, 2018).

The monophthongs examined in isolation reveal a significant correlation between Dimension 1 and F2 ($r=-0.872$, $t(9)=-5.346$, $p < 0.001$), and Dimension 2 and F1 ($r=0.876$, $t(9)=5.438$, $p < 0.001$). F3 was not found to be significantly correlated with any dimension. The effect of duration was also not significant. This establishes a clear quantitative link between subjective classification by listeners and the objective acoustic properties of the vowels. We conclude that the listeners grouped vowels with similar F1 and F2 or F2 only properties, with no regard for duration. The Pearson's correlations for 2-dimensional MDS are given in Table 6. 4.

Table 6. 4: Pearson's correlations for 2-dimensional MDS for monophthongs

| | F1 | F2 | F3 | Duration |
|---|---|---|---|---|
| Dimension 1 | 0.03 | -0.87 | -0.23 | 0.13 |
| Dimension 2 | 0.88 | -0.54 | 0.01 | -0.22 |

The diphthongs examined in isolation with regard to three-dimensions reveal a significant correlation between Dimension 1 and F2 onset (r=-0.954, t(6)=-7.8057, p < 0.001), and a very weak correlation between Dimension 2 and F2 offset (r=-0.489, t(9)=-1.3726, p = 0.2), and Dimension 3 and F2 offset (r=-0.620, t(9)=-1.9384, p = 0.1). No other significant correlations were found. In particular, the effect of duration was not significant. These results suggest that listeners were more sensitive to F2 at onset and offset and perceived diphthongs similar to each other with regard to F2 onset and F2 offset. In addition, F3 was not found to be significantly correlated with any dimension. The Pearson's correlations for 3-dimensional MDS are given in Table 6. 5.

Table 6. 5: Pearson's correlations for 3-dimensional MDS for diphthongs

| | F1 onset | F1 offset | F2 onset | F2 offset | F3 onset | F3 offset | Duration |
|---|---|---|---|---|---|---|---|
| Dimension 1 | 0.54 | -0.25 | -0.95 | -0.16 | 0.14 | -0.02 | 0.69 |
| Dimension 2 | -0.39 | -0.22 | 0.01 | -0.49 | -0.48 | -0.53 | -0.16 |
| Dimension 3 | 0.43 | 0.40 | 0.07 | -0.62 | -0.31 | -0.15 | -0.29 |

### 6.7 Summary and Discussion

In this experiment, the listeners exhibited a range of free classification strategies. Some listeners created very few groups, while other created as many as 19 groups. The number of groups ranged from 4 to 22 with a median 11 and mean 12, and the sizes of groups ranged from 1 to 15 vowels per group with a median of 2 and mean of 6. These results show that by using a free classification task listener can make a relatively large number of fine distinctions based on the given set of words carrying the target vowel sounds.

Despite the fine-grained classifications made by individual participants, the clustering analysis of the aggregate data revealed just four overriding categories in the identification and grouping of the vowels: front, back, high and non-high. However, the listeners' central vowel space is empty. The classification of SSBE vowels into front and back clusters is also reported by Evans and Alshangiti (2018) for Arabic learners of English with high proficiency in English. However, the further sub-clusters they reported are very different from the ones found in the present study.

These four broad perceptual categories correspond directly to the perceptual categories revealed by MDS analysis. Overall, the perceived distribution of SSBE monophthongs showed by the MDS solution is similar to their distribution in acoustic space, except for the two central vowels: /ɜː/ was perceived as more similar to open-mid front vowels; and /ʌ/ was perceived as more similar to open-mid back vowels.

The MDS analysis also revealed several novel findings. When the perceptual distances between the SSBE vowels were plotted in a two-dimensional space for monophthongs, the dimensions that emerged corresponded to the vowel space in four categories: front, back, high, low. The two-dimensional MDS also reveals that SSBE /ɜː/ was grouped with the front non-high vowels, i.e. /ɛ/ and /æ/; and /ʌ/ was grouped with the back non-high vowels, i.e. /ɒ/, /ɔː/ and /ɑː/. The MDS solution results for diphthongs, using three-dimensions, are partially non-conclusive as dimension 3 did not show any relation to the given acoustic measures of diphthongs; except that the back dimension was split into two, which was interpreted as low back and high back. However, the interpretation of three-dimensional space with regard to x, y, z coordinates clarified the ambiguity and showed that Punjabi-Urdu speakers/listeners focused on the onset vowel in first quadrant and offset vowel in the second, third and fourth quadrant. These findings were backed up by Pearson's r test for correlation which was performed on F1, F2, F3 and duration of the stimulus vowels against each dimension of the MDS analysis as discussed in Section 6.6.2.

The results obtained from this free classification experiment look promising because they are consistent with previous research on the production of Pakistani English (PE), in particular vowels (see Section 4.1.1). That is, these grouping patterns show some similarities to those found in the PE production experiments presented by Mahboob and Ahmar (2004), Bilal et al. (2011a, 2011b and 2011c), Raza, (2008) and Rahman (1991). The results for front and back tense and lax vowels (i.e. /iː/-/ɪ/ and /uː/-/ʊ/) are partially in line with the impressionistic description of PE (as discussed in Chapter 4).

According to the literature, in Pakistani English speakers make a distinction between both front and back tense and lax vowels (i.e. /iː/-/ɪ/ and /uː/-/ʊ/) (Sheikh, 2012; Bilal et al., 2011a; Mahboob and Ahmar, 2004; Raza, 2008; Rahman 1997 and 2015); however, the results from the free classification task do not show this distinction. With regard to SSBE back and central vowels, as well as diphthongs, the results are comparable to the impressionistic and acoustic description of PE (Hassan and Imtiaz, 2015; Bilal et al., 2011c; Raza, 2008; Rahman, 1991, 1997 and 2015). Overall, the results from the free classification task are in line with the description of Asian Englishes (Bolton, 2008; Deterding, 2007; Kachru, 2005) and Indian English (Maxwell and Fletcher, 2009 and 2010; Gargesh, 2004), with slight variations, in particular for central vowels.

These grouping patterns suggest that listeners' distinctions between SSBE vowels are influenced by their L1s vowel inventory, i.e. Punjabi and Urdu in this case, even when specific labels are not imposed on the task a priori by the experimenter (as in the multiple forced choice experiments). In addition, the multidimensional scaling analysis of the classification data provided some insights into the relevant perceptual dimensions of distinction between SSBE vowels by the listeners.

### 6.7.1 Comparison of the Results from two Tasks: Free Classification and Cross-Language Perceptual Assimilation

As discussed in Chapter 5, a cross-language assimilation task was designed to explore the perceptual assimilation patterns of SSBE vowels in a bVd context by Punjabi-Urdu listeners. It was predicted that the results from the cross-language assimilation task will be similar to that of the free classification task discussed in this chapter; however, the findings from both contexts show some differences. A comparison of the findings from these two experiments is discussed below.

#### 6.7.1.1 Monophthongs

Firstly, in the cross-language assimilation task, Punjabi-Urdu listeners showed sensitivity to both spectral and temporal cues. However, in the free classification task, listeners were not sensitive to temporal cues.

The results showed that in the cross-language assimilation task, SSBE front vowels /iː/, /ɪ/, /ɛ/ and /æ/ were most often assimilated to the Urdu vowel /eː/ with a goodness rating ranging from 4.5 to 5.3. The clustering and MDS solutions for the free classification task showed similar results to the cross-language assimilation task for the front vowels. In the cross-language assimilation task, SSBE /æ/ was assimilated to Urdu /eː/ (30%; 4.5) and Urdu /**ɛ**ː/ (31%; 5.3). As can be seen in the dendrogram (Figure 6. 4 and Figure 6. 5), in the free classification task, SSBE /**ɛ**/ and /æ/ were grouped together and appeared to be more similar to each other than /ɜː/ which was added to the cluster later on.

In the cross-language assimilation task, the SSBE central close-mid vowel /ɜː/ was assimilated to Urdu /ɑː/ (25%: 5.2), but SSBE /ʌ/ was assimilated to Urdu /ɐ/ (44%: 5.4). However, in the free classification task, SSBE central open-mid vowel /ʌ/ and SSBE central close-mid vowel /ɜː/ were grouped with back and front vowels respectively. Thus, the findings for the central vowels are quite different across both experiments. Assimilation patterns for the back vowels can be considered broadly similar across both experiments. One contrasting result was that in the cross-language assimilation task listeners identified SSBE /ʊ/ and /uː/ as distinct vowels and then mapped to their corresponding Urdu vowels, whereas they failed to detect this temporal distinction in the free classification task.

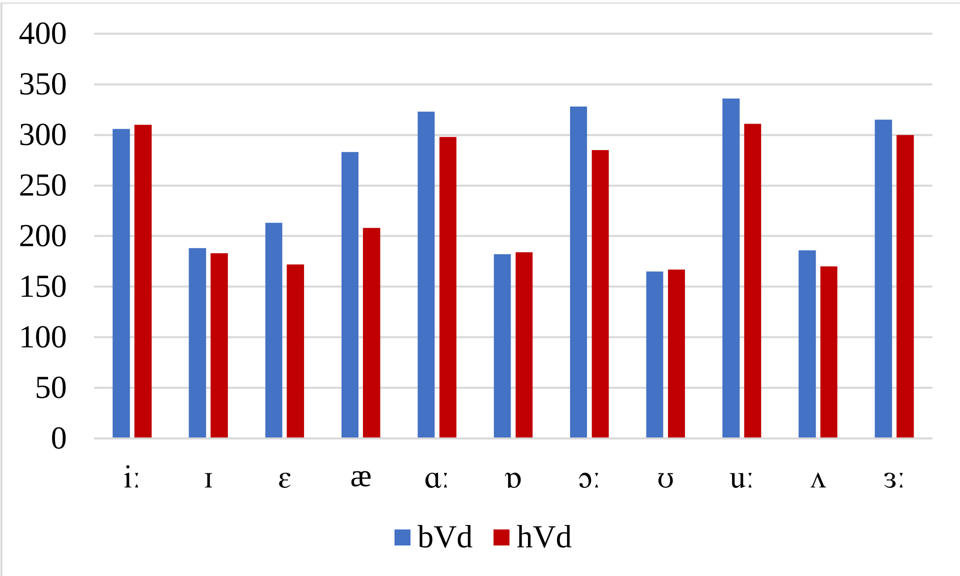



Figure 6. 15: Mean duration (ms) for SSBE monophthongs in a bVd and hVd context

Overall, for monophthongs the results from free classification and cross-language assimilation tasks are comparable, except for tense and lax vowels: /iː/, /ɪ/, /ʊ/ and /uː/ and central vowels: /ʌ/ and /ɜː/. We can eliminate differences in duration as a potential cause for this. As can be seen in Figure 6. 15 and Figure 6. 16, the difference in mean duration across both contexts (and hence, both experiments) was not sufficient enough to account for variations between the contexts.

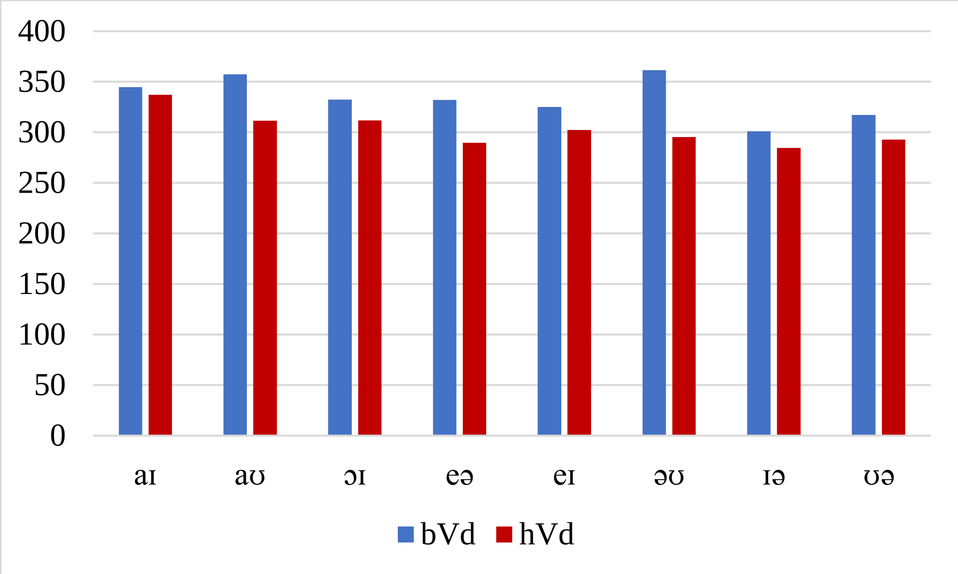



Figure 6. 16: Mean duration (ms) for SSBE diphthongs in a bVd and hVd context

#### 6.7.1.2 Diphthongs

For diphthongs, the assimilation patterns are similar to some extent across both experiments. In the cross-language assimilation task, SSBE /ɪə/, /eə/ and /eɪ/ were mapped to Urdu /eː/; SSBE /ʊə/ was most frequently mapped to "none", but second-most to Urdu /eː/. In the free classification task these diphthongs were clustered with the front monophthongs. /eɪ/ was grouped with /iː/ and /ɪ/. In the dendrogram (as shown in Figure 6. 4), /ɪ/ and /eɪ/ were more similar to each other than /iː/; /ʊə/ and /ɪə/ were part of the same cluster; and /eə/ was grouped with /ɛ/ and /æ/ where /eə/ was more similar to /ɛ/ than /æ/. These patterns show that listeners paid attention to the second element in these diphthongs.

In the cross-language assimilation task, SSBE /əʊ/ was mapped to Urdu vowel /oː/; however, in the free classification task, this diphthong was grouped with SSBE monophthongs /ʊ/ and /uː/. In the free classification task, the SSBE diphthongs /aɪ/, /ɔɪ/ and /aʊ/ were clustered together. For these vowels it seems that listeners paid more attention to the second element. Since these diphthongs were grouped with the back vowels, the pattern suggests that listeners were more sensitive to F2 than F1.

The assimilation patterns for diphthongs were quite different in both experiments. The free classification of diphthongs appears to be slightly more complex. These diphthongs can be clustered in two dimensions, i.e. front and back. The "back" dimension is further split in two dimensions: high back and low back, as shown in Figure 6. 13 and the accurate position of these diphthongs in three-dimensional space is best represented in four quadrants as shown in Figure 6. 14. In addition, the sub-clusters in the dendrograms (Figure 6. 4 and Figure 6. 6) show that listeners focussed on the second element more often. However, as discussed above (see Section 6.6.2), the MDS *stress* plot suggested a four-dimensional solution for diphthongs, but the fourth dimension was not interpretable with regard to any linguistic and acoustic properties (cf. Fox, 1983).

**6.8 Conclusion**

The results showed that listeners clustered long and short vowels together. The cluster distances suggest that the listeners found long and short vowel pairs very similar to each other, but very distinct from other vowels. Secondly, it was predicted that listeners will be able to differentiate the central vowels from front and back vowels. The clustering and MDS solution showed that listeners confused the central vowels /ɜː/ and /ʌ/ with front and back vowels, respectively. Thirdly, it was predicted that listeners will be able to differentiate monophthongs from diphthongs. The clustering solution showed that listeners indeed grouped /aʊ/, /aɪ/ and /ɔɪ/ together, however, the other diphthongs were grouped with monophthongs.

We can conclude that cross-language perceptual assimilation patterns cannot predict the perceptual similarity between L2 vowels as perceived by L2 learners. For example, the cross-language perceptual assimilation of SSBE central vowels /ʌ/ and /ɜː/ suggests that listeners found these vowels quite similar; however, in the free classification task hierarchical clustering and MDS analysis showed that listeners found these vowels very different, and they grouped SSBE central open-mid vowel /ʌ/ and SSBE central close-mid vowel /ɜː/ with back and front vowels, respectively.

Similarly, in the cross-language perceptual assimilation task, SSBE /ɪ/ and /ɛ/ were assimilated to Urdu /eː/. However, hierarchical clustering and MDS analysis showed that /ɪ/ and /eɪ/ were perceived more similar to each other than /iː/, and /ɛ/ was perceived as very different from /ɪ/. In the free classification task /ɛ/ appeared to be more similar to /eə/ than /æ/ and /ɜː/, which were added to the cluster later on. Similarly, in a two-dimensional space, /ɪ/ was found to be front-high, and /ɛ/ was found to be front-low.

The patterns for individual perceptual similarity are not tested here and can be the focus of future research. Variability in the individual perceptual similarity and assimilation patterns can be due to the individual listeners' knowledge and experience of the target language and influence of their L1s. In addition, in order to test if the presence or absence of diphthongs affect perceptual identification and classification, future experiments can be designed to run separate tests for monophthongs and diphthongs.

## Chapter 7

### Discussion and Conclusion

The aim of this thesis was to investigate the acoustic and phonetic properties of Urdu vowels as spoken in Punjab Pakistan and the perceptual assimilation of SSBE vowels by Punjabi-Urdu speakers. Chapter 1 provided an overview of the thesis structure, layout and research aims and objectives. In the first section of this thesis, the acoustic and phonetic properties of Urdu monophthongs and diphthongs were investigated. This acoustic and phonetic investigation followed by phonological interpretation of Urdu aided understanding of the cross-language perceptual assimilation and free classification of SSBE vowels in the second section.

In this chapter, firstly a summary of the methods and results of the production experiments are discussed. Secondly, the similarities and differences between cross-language assimilation patterns of SSBE vowels across two different tasks are discussed, followed by the comparison of findings with the impressionistic descriptions of Pakistani English. Lastly, the implications of these findings with regard to how patterns of perceptual assimilation and free classification may predict perceptual difficulties by SSBE learners of English who speak Punjabi and Urdu as their first and second language, respectively. Lastly, limitations of the present study and future work is discussed.

### 7.1 Urdu Monophthongs and Diphthongs

As discussed in Chapter 1, keeping in view the effects of local languages, it was decided to investigate the Urdu dialect as spoken in Punjab, Pakistan, in particular to establish the following:

a) whether there is a consistent difference between short and long vowels in Urdu;

b) whether vowels, whose status has been disputed, differ in quality or length; specifically, the front vowels /e/, /ɛ/ and /æ/, and back vowels /o/ and /ɔ/.

c) Whether Urdu as spoken in Punjab, Pakistan has two central vowels, /ə/ and /ʌ/.

d) Whether Urdu as spoken in Punjab, Pakistan has diphthongs.

In order to answer these questions, the acoustic data were collected from 22 speakers (11 males + 11 females) who speak Punjabi as their first language. The results, as discussed in Chapter 2, were based on F1 and F2 at steady state points, and F1 and F2 at seven equidistant points, and duration. The results showed that Urdu has long and short vowels, which differ both spectrally and temporally. The contradictory vowels in the literature were further analysed, and the results showed that /ɪ/ and /e/ have substantial spectral overlap; however, these are distinct phonemes, especially with regard to F1 and duration. Therefore, it was concluded that these two vowels differ temporally and spectrally. In addition, the duration of /e/ suggests that it is not a short vowel. Phonologically, these are two distinct phonemes, since they can be found in minimal pairs, for example /peːʈ/ "*stomach*" and /pɪʈ/ "*got beaten up*". In addition, both of these phonemes can be found in word initial and medial positions, for example /eːk/ "*one*", /ɪk/ "once", /peːʈ/ "*stomach*", and /pɪʈ/ "*got beaten up*". In addition, /eː/ can be found in monosyllabic open syllables, such as /seː/ "from", /keː/ "of"; however, /ɪ/ cannot be found in monosyllabic open syllables. As discussed in Section 3.1.1, Urdu does not allow open light (mono-moraic) syllables (Hussain, 1997). This suggests that Urdu has long and short vowels, where short vowels are mono-moraic and long vowels are bi-moraic.

/ɛ/ and /æ/ are reported as distinct phonemes in the Urdu literature (Fatima and Aden, 2003; Saleem et al., 2002); however, the acoustic and phonetic analysis from the present study show that [ɛ] and [æ] do not differ spectrally or temporally. These vowels appear in the open-mid and close-mid region of the vowel quadrilateral; hence it would be more appropriate to transcribe this vowel as /ɛ/ rather than /æ/. Phonemically, /ɛ/ can be found in closed syllables, and in disyllabic or tri-syllabic words in stressed open syllables, for example /ˈbɛ.t̪ʰɑː/ "*sat down (he)*", where the first syllable is open and has /ɛ/ at the end.

Despite the fact that [ɔ]-[o] are discussed in the literature as distinct phonemes, in the present study these vowels showed substantial spectral and temporal overlap and statistical analysis confirmed that these are not distinct phonemes in Urdu. Saleem et al. (2002:3) has given a disyllabic near minimal pair [sonɑ] "*gold*" or *"to sleep"* and [bɔnɑ] "*to sow*"; however, the vowel sound in both of these words does not differ in quality, and does not change the meaning of the given words if pronounced with one or the other sound. In addition, no data is offered to compare with the findings from the present study. It is hard to find minimal pairs in Urdu, which could essentially mean that the sounds that occur at different positions in different words (initial, medial or final) are allophones with slight variations of the same sound. Phonemically, these two sounds do not contrast in Urdu spoken by Punjabi-Urdu speakers, therefore it can be concluded that [ɔ] and [o] are not distinct phonemes in Urdu phonetically or phonemically. It is quite possible that the studies that report these two vowels as distinct phonemes took their data from speakers of Urdu with non-Punjabi L1. Therefore, further experiments with minimal pairs and using speakers with different L1s (e.g. Sindhi, Balochi, Saraiki) can help to better establish the quality and status of these two vowels in the Urdu vocalic inventory.

With regard to the central vowel, the results from the present study show that [ə] is an allophonic realisation of the fairly open, central unrounded vowel [ʌ], which has higher F1 and lower F2 than [ə]. It is quite possible that this central vowel is pronounced differently by speakers of Urdu with a non-Punjabi L1. Although [ə] and [ʌ] vowels appear to be distinct phonetically with reference to duration, F1 and F2, phonologically these two sounds are not distinct (see Section 2.6 for further details). Therefore, it can be concluded that [ə] and [ʌ] are allophonic in Urdu where [ʌ] is a fairly open, central unrounded vowel with higher F1 and lower F2 than [ɘ], and a more appropriate transcription for this vowel would be /ɐ/.

For the front open-mid vowels, [e], [ɛ] and [æ] have been used for distinct vowels or as allophones of the same vowel in different studies. For example: Saleem et al. (2002) states that the sounds [ɛ] and [æ] are allophones of the same sound; Fatima and Aden (2003:74) state that [ɛ] is a long and [æ] is a short vowel or vice versa; Raza (2009) reported [ɛ] and [æ] as distinct vowels. However, neither Saleem et al. (2002) nor Fatima and Aden (2003) and Raza (2009) have given enough evidence to support their claims. Kachru (1990) used the symbol [ɛ] for a front open-mid vowel in his impressionistic account of the Hindi-Urdu vocalic inventory; however, Hussain (1997) argued that the sound in Urdu is closer to vowel /æ/ than /ɛ/. In the present study, [ɛ], [e] and [æ] were used in near minimal pairs [peːʈ] *"stomach",* [bɛd̪] "*willow tree"* [bæt̪] *"to follow"* and the results show that [ɛ] is not distinct from [æ].

Another issue with the available literature was the lack of reliable information to compare the results, i.e. acoustic analysis and information about the participants' linguistic background. As a result, there is no way to compare the results from the present study with the studies reported in the literature. The acoustic properties of Urdu vowels (duration and F1 and F2 frequencies) reported by Hussain (1997) were compared with the findings from the present study. As Hussain (1997) extracted those values from stressed and unstressed syllables from disyllabic words from varying consonantal contexts, the comparison is not very comprehensive. However, the results showed that the vowel length contrasts and F1 and F2 values were partially comparable. This comparison confirmed the distinctions in both vowel spectral and temporal qualities that are found in the present study.

In summary, the analysis of formant frequencies and duration shows that Urdu has long and short pairs. Long vowels are significantly longer than short vowels and appear to be more peripheral in quality than short vowels. In addition, short vowels do not occur in open syllables (CV) in Urdu. Although literature reports on tense and lax vowels, we do not have articulatory evidence to support those claims here.

The research questions were answered as follows: (a) there is a consistent difference between short and long vowels in Urdu; (b) the front vowels /e/ and /ɛ/ are distinct in quality and length and Urdu does not have /æ/; (c) back vowels [o] and [ɔ] are not distinct in quality, in particular spectrally and temporal difference does not show phonological distinction; (d) it seems that Urdu as spoken in Punjab, Pakistan has two central vowels, [ə] and [ʌ], but [ə] is only an allophonic realisation of /ʌ/ in unstressed multisyllabic words, and a more appropriate transcription for this vowel would be /ɐ/. Nasalised vowels were not analysed in this study, despite the data being collected. Therefore, future studies will focus on the analysis of nasalised vowels and the allophonic status of [ə] in stressed and unstressed multisyllabic words. Thus, Urdu as spoken in Punjab Pakistan has six long vowels: /iː/, /eː/, /ɛː/, /ɑː/, /oː/ and /uː/; and 3 short vowels: /ɪ/, /ɐ/ and /ʊ/.

Chapter 3 provided an acoustic and phonetic investigation of six Urdu diphthongs. Although the phonological status of diphthongs remains unclear, the results showed that these diphthongs are not distinct with regard to duration. The acoustic analysis showed that phonetically these Urdu vocalic sequences partially (e.g. temporally) behave like diphthongs. The results also showed that these vocalic-sequences are shorter than two monophthongs together, but longer than a long monophthong which suggests that a time slot is not deleted as claimed by Waqar and Waqar (2002). All six diphthongs were distinct from each other at the onset (20%) and offset (80%) with regard to F1 and F2, and the first and second components were also distinct from pure monophthongs in quality.

Contrary to the assertion by Waqar and Waqar (2002:20) that “Urdu has rising diphthongs (second vowel is of longer duration)”, the results from the present study showed that the duration of the onset, transition and offset are all similar, which suggests that Urdu does not have rising diphthongs. The F2 transition occupies approximately 30% of the total duration of each diphthong. This is markedly different from English, where the transition duration can be very high, e.g., 60% for /aɪ/ and 73% for /aʊ/ (Lindau et al., 1990). According to Lindau et al. (1990), transition duration appears to be distinct across diphthongs as well as languages, for example, the transition duration of /aɪ/ and /aʊ/ occupies 16-20% in Arabic and Hausa, and 40-50% in Chinese.

Due to lack of enough phonetic evidence, we can conclude that the diphthongs discussed above were pronounced as two vowels (vowels in hiatus) by Punjabi-Urdu speakers. The stimuli used here were all monosyllabic words in open syllables, therefore the effects of stress (as discussed in Spanish; Aguilar, 1999) on the quality of these vowels cannot be accounted for.  Future work is therefore required to investigate the effect of stress.

The most compelling reason to treat these vowel-vowel sequences as diphthongs, other than their duration, is that the literature on Urdu phonology unanimously reported that onsetless syllables are not allowed in Urdu, except for word initial position. Therefore, if we divide the words, e.g. /pɑe/ and /bɔe/ in two syllables, such as /pɑ.e/ and /bɔ.e/, we are left with an onsetless syllable which is not permitted in Urdu (see Section 3.1.1). In future, it will be interesting to see the qualities of these diphthongs in disyllabic and multisyllabic words with stressed and unstressed syllables.

Another argument against these vocalic sequences being considered diphthongs regards language games (Ohala, 1986). The author of this thesis, a native speaker of Urdu, considered that if /ɑe/ can be broken down into two syllables, /pɑ.e/, a language game can be played where the penultimate syllable is swapped by the antepenultimate syllable, so /pɑ.e/ will become /e.pɑ/. However, this is not always possible. If the vowel sequence /ɑe/ is found in closed syllables, such as /məsɑel/ "*problems*", the syllabification will be /mə.sɑel/. If the last syllable in /mə.sɑel/ carrying /ɑe/ is broken down to give /mə.sɑ.el/, this would violate the onsetless syllable structure rule (see Section 3.1.1). Such a syllable structure would also sound incorrect to a native speaker of Urdu. This suggests that /ɑe/ is a diphthong. Therefore, future perception and syllable identification tests will help to better establish the status of these diphthongs.

In the present study, six diphthongs (as discussed in Section 3.2.2), which were considered close to SSBE diphthongs by the author, were investigated. They helped to further understand the perceptual assimilation and free classification patterns of SSBE vowels as discussed in Chapters 5 and 6. The IPA symbols for these diphthongs were chosen based on analysis of the acoustic and phonetic properties of Urdu vowels, as discussed in Chapters 2 and 3. Therefore, these symbols do not match the diphthongs reported in the Urdu literature.

The acoustic and phonetic investigation of Urdu vowels facilitated the predictions of the perceptual assimilation and free classification of the SSBE vowels as in Chapter 4, Chapter 5 and Chapter 6.

## 7.2 Perceptual Assimilation of SSBE Vowels

Cross-language perceptual assimilation and free classification of 19 SSBE vowels were tested in three different contexts: bVd, hVd, and hVba. These three different contexts were used for different reasons. bVd and hVba contexts were used for cross-language mapping and goodness rating tasks, where listeners had to match an SSBE vowels with one of the given Urdu vowel categories (Urdu words carrying the Urdu vowels) and give a goodness rating on a scale from 1 (unlike to Urdu) to 7 (like Urdu) (Strange et al., 1998; Faris et al., 2016). In the bVd context, listeners were told that these are English words; however, in the hVba context listeners were told that these words are from a different European language, and they can learn more about that after they finish the experiment. The use of nonsense words in a hVba context made it easier for listeners to believe that this was not English. The rationale behind this was to test if listeners' perceptual assimilation patterns are affected with regard to knowledge and familiarity of the target language. As well as the expected effects of context and syllable structure on the quality of vowels, the results did show a number of differences in the perceptual assimilation patterns across both contexts. As discussed in chapter 6, SSBE vowels in a hVd context were used for a free classification task to test how learners perceive and group these vowels when there is no categorisation or mapping to L1 vowels and no predefined categories.

The results showed quite inconsistent patterns of perceptual assimilation across bVd and hVba contexts. In the bVd context, listeners chose fewer response categories, which means those vowels were considered categorised. In the hVba context, listeners chose multiple response categories, which means most vowels were considered uncategorised. The assimilation patterns for SSBE /æ/ were significantly different, inconsistent and very complex across the two contexts: In the bVd context, SSBE /æ/ was assimilated to Urdu vowels /eː/, /ɛː/ and /ɑː/. In the hVba context, SSBE /æ/ was assimilated to multiple Urdu vowels, with the two most frequently chosen Urdu vowels being: /ɑʊ/ and /ɑː/. Listeners' sensitivity to temporal similarities and differences could be the reason for this perceptual assimilation in the bVd context. However, in the hVba context, listeners appeared to be sensitive to F1, since F2 and duration of /æ/ in a hVba context are significantly different from Urdu /ɑː/. The categorised patterns in the bVd context and uncategorised patterns in the hVba context suggest that familiarity with the target language made listeners sensitive to both phonetic and phonological contrasts, especially with regard to L1. However, unfamiliarity with L2 made listeners pay attention only to the phonetic details of the target language. These findings are neither supported by Perceptual Assimilation Model (PAM-L2; Best and Tyler, 2007) nor by Speech learning Model (SLM; Flege, 1995).

Assimilation patterns for SSBE central vowels /ʌ/ and /ɜː/ were also quite different in both contexts. In the bVd context, SSBE central open-mid vowel /ʌ/ was assimilated to Urdu /ɐ/, /ɑː/ and some other vowels; and SSBE central close-mid vowel /ɜː/ was assimilated to Urdu /ɑː/, /ɐ/ and /eː/. On the other hand, in the hVba context, SSBE central open-mid vowel /ʌ/ and central close-mid vowel /ɜː/ were confused with multiple Urdu vowels, but never mapped to Urdu /ɐ/. These results showed that in the hVba context, Punjabi-Urdu listeners found these two central vowels very difficult as compared to the bVd context, where they categorised /ʌ/ with Urdu /ɐ/ a higher percentage of the time. These differences in the patterns of perceptual assimilation of SSBE central vowels failed to support L2LP predictions (see Section 5.1.1 for the predictions and Sections 5.6.2 and 5.7.2 for the findings from the bVd and hVba contexts respectively).

Faris et al. (2016) reported that Egyptian Arabic learners of Australian English found the central vowel /ɜː/ difficult to perceive and map to any of Arabic vowels. Evans and Alshangiti (2018) reported that Saudi Arabian Arabic learners of SBBE with a higher proficiency in English confused the central vowel /ɜː/ and diphthong /eə/ with each other. They further reported that listeners with higher proficiency levels were able to detect some of the differences of these two vowels to other vowels. In the present study, in the bVd context listeners assimilated SSBE /ɜː/ to Urdu vowels /ɑː/, /ɐ/ and /eː/. However, in the hVba context they confused this vowel with multiple Urdu vowels. This could be due to the fact that Urdu does not have this central vowel, /ɜː/. In addition, according to Bilal et al. (2011c), Pakistani English does not have /ɜː/ vowel. Therefore, the difficulty in the perception of this vowel could be due to the fact that Punjabi-Urdu listeners did not have a good match for this vowel available and so confused it with multiple other vowels, especially in the hVba context. These results support both SLM and PAM predictions that it will be easier to form a new category for SSBE /ɜː/ (SLM), and it will be easier to discriminate SSBE /ɜː/ (PAM) because this vowel does not have a counterpart in Urdu.

As discussed in Section 5.9 assimilation patterns for SSBE back vowels were quite different in both contexts, i.e. in the bVd context /ɑː/ was assimilated to Urdu /ɑː/; however, in the hVba context /ɑː/ was assimilated to Urdu diphthong /ɑʊ/ and multiple other vowels. Further, in the bVd context, SSBE /ʊ/ and /uː/ were mapped to Urdu /ʊ/ and /uː/ respectively. However, in the hVba context, both SSBE /ʊ/ and /uː/ were mapped to Urdu /uː/, as well as other vowels. These results showed that in a hVba context Punjabi-Urdu speakers found SSBE /ɑː/ most confusing and mapped it to a diphthong. Similarly, participants did not categorise SSBE /ʊ/ and /uː/ as distinct vowels, even though they have their counterparts in Urdu. Although participants did detect this distinction in the bVd context, they failed to detect /ʊ/ and /uː/ as distinct vowels in the hVba context. As discussed above and in Section 5.9, these assimilation patterns in the bVd and hVba context suggest that familiarity with the target language played an important role in the perception of these vowels. The assimilation patterns for other monophthongs, and all diphthongs, were quite similar across both contexts.

In L2 perception literature, SSBE /iː/-/ɪ/ and /uː/-/ʊ/ contrasts have been reported as troublesome for second language learners (Evans and Alshangiti, 2018; Escudero and Chládková, 2010; Lengeris, 2009; Escudero, 2005). In addition, Evans and Alshangiti (2018), in an investigation of perception and production of British English vowels and consonants by Arabic learners of English with varying proficiency levels in English, reported that /ɪ/-/e/ contrast was found to be most confusing by Arabic learners of English.

Spanish is reported to lack /iː/-/ɪ/ and /uː/-/ʊ/ contrasts; hence these SSBE contrasts are perceived as the single Spanish vowel /i/ and /u/ respectively (Chládková, 2010; Escudero, 2005). According to Holes (2004, cited in Evans and Alshangiti, 2018: 17), as opposed to Urdu Modern Standard Arabic does not have an /e/ vowel and has three tense-lax pairs: /iː/-/i/, /aː/-/a/, and /uː/-/u/. Evans and Alshangiti (2018) reported SSBE contrasts that were not found in Arabic were the most difficult for Arabic learners of English, such as /ɪ/-/ɛ/. However, Punjabi-Urdu listeners in the present study do have /ɪ/-/ɛː/ contrast and long and short vowels (see Section 7.1) in their L1, but they failed to detect these differences in the perception of SSBE vowels, especially in the hVba context. Thus, the results from hVba context are in contrast with previous studies, which show that L2 contrasts that are not found in L1 appear to be the most difficult to perceive by L2 learners (Evans and Alshangiti, 2018; Escudero and Chládková, 2010; Escudero, 2005; Strange et al., 1998, 2001 and 2003).

Some of PAM's and L2LP's predictions were supported. The results showed some two-category (TC) (i.e. /iː/-/ɪ/ /iː/-/ɛ/) and single-category (SC) (i.e. /ɪ-e/ and /ɔː-ɒ/) assimilation patterns (mainly in the bVd context), but many more uncategorized-uncategorized (UU) assimilation patterns (mainly in the hVba context). The uncategorized assimilation patterns were further broken down into three-way uncategorized assimilation patterns: *focalized*, *clustered* and *dispersed* (Faris et al., 2016). According to second language perception theories, listeners use both fine-grained phonetic details as well as abstract phonological categories (Best, 1995) in speech perception. Unlike *dispersed* assimilation patterns, where listeners are only sensitive to phonetic information, the *focalized* and *clustered* assimilation patterns suggest that listeners were sensitive to some phonetic information in non-native phonemes (i.e. SSBE vowels) that is phonologically meaningful in their L1 (i.e. Punjabi-Urdu vowel system).

According to SLM and PAM-L2, for clustered and focalized assimilations a new L2 category can be formed, if L2 contrasts do not overlap with any other L2 phonemes, for example, new categories might be formed for SSBE focalised /ɛ/ and clustered /uː/. However, for SSBE /ɪ/, /ɛ/, and /æ/ new categories will not be formed due to the substantial classification overlap between them.

According to PAM and PAM-L2, non-overlapping and partially overlapping phonemes can be discriminated more accurately than completely overlapping phonemes. For example, focalized assimilation of SSBE /iː/ and /ɛ/ to two separate Urdu vowel categories /iː/ and /eː/ (non-overlapping) suggests that these will be discriminated easily, as compared to when two SSBE vowels /ɪ /and /ɛ/ are assimilated to one Urdu vowel category /eː/ (complete overlap). However, according to PAM-L2, for dispersed assimilations new L2 phonological categories will be formed, because the dispersed assimilation patterns emerged due to the absence of similar phonological categories in the listeners' L1 phonological space. Lastly, the classification overlap scores (as discussed in Section 5.8) suggest that discrimination will be easier for the vowels in a hVba context; however future experiments are required to verify this prediction.

According to PAM, articulatory similarities between L1 and L2 phones play an important role in the perception of L2 phones. Therefore, one possibility for the complete, partial or high versus low overlap scores could be the shared articulatory-phonetic features between the L2 phones (SSBE vowels) and L1 (Urdu vowels). For example, Faris et al (2018) reported that lip rounding and tongue backness played a significant role in the accurate discrimination of the Danish vowel /ɛ/-/o/ contrast by monolingual Australian English speakers. In the present experiment, the degree of similarities and/or differences of articulatory-phonetic features between the L2 phones (SSBE vowels) and L1 (Urdu vowels), do not fully explain the assimilation patterns. For example, as discussed in

Section 5.8, in bVd context, /iː/-/ɪ/ do not show any overlap, and in hVba context /iː/-/ɪ/ show very little overlap. Familiarity (i.e. bVd context) and unfamiliarity (i.e. hVba context) with L2 could be an important factor for these anomalies in the perceptual assimilation patterns.

Quantitative analysis was conducted to explore the differences in goodness ratings (Faris et al. 2018) (see Section 5.4 and 5.5 for further details). Both quantitative analysis and descriptive statistics of ratings given in the bVd context suggested that, for SSBE vowels, listeners used the rating scale to indicate their sensitivity to some cross-language spectral differences. However, in the hVba context the ratings did not suggest sensitivity to cross-language spectral differences, with some exceptions, for example the mean goodness rating for SSBE /ɔː/ was 5.1 (with a higher fit index of 3.3, see Section 5.5 for further details) when it was mapped to Urdu /oː/ and 4.6 when it was mapped to Urdu /uː/. Overall, listeners performed as expected in the bVd context. However, their performance was inconsistent in the hVba context, and the rating task did not provide any additional information. Hence, L2LP predictions based on acoustic similarities between SSBE and Urdu vowel system were not correct. The above results suggest that not only the consonantal context (in line with Levy, 2009) in which the SSBE vowels were produced (e.g. monosyllabic vs. disyllabic, English words vs. nonsense words and bVd vs. hVba), but also the familiarity with the target language (in line with Levy and Strange, 2007) had significant effects on the perceptual assimilation of SSBE vowels by Punjab-Urdu speakers.

In addition, the results from the free classification task show that cross-language perceptual assimilation patterns cannot predict the perceptual similarity between L2 vowels as perceived by L2 learners. For example, the cross-language perceptual assimilation of SSBE central vowels /ʌ/ and /ɜː/, in both bVd and hVba contexts, suggest listeners found these vowels quite similar. However, in the free classification task in hVd context, hierarchical clustering and MDS analysis showed that listeners found these vowels very different, grouping SSBE central open-mid vowel /ʌ/ and SSBE central close-mid vowel /ɜː/ with back and front vowels, respectively.

Similarly, SSBE /ɪ/ and /ɛ/ were assimilated to Urdu vowel /eː/ in both bVd and hVba contexts, which suggest that listeners found these vowels very similar; however, hierarchical clustering and MDS analysis showed that /ɪ/ and /eɪ/ were more similar to each other than /iː/, which was added to this cluster later on; and /ɛ/ was perceived as very different from /ɪ/ and was grouped with /æ/ and /eə/; and /ɛ/ appeared to be more similar to /eə/ than /æ/ or /ɜː/, which were added to the cluster later on. Similarly, in two-dimensional space /ɪ/ was found in the front-high dimension, and /ɛ/ was found in front-low dimension. Thus, we can conclude that cross-language perceptual assimilation patterns cannot predict the perceptual similarity between L2 vowels as perceived by L2 learners.

The results also suggest that cross-language acoustic and phonetic properties of vowels (*spectral vs. temporal parameters,* Strange, et al., 1998; L2LP: Escudero, 2005; Escudero et al., 2014) cannot predict assimilation patterns effectively. For example, given that Urdu distinguishes between short and long vowels, it was predicted that listeners will be able to detect the temporal differences between SSBE tense and lax vowels. However, the results showed that listeners were more sensitive to F1 and F2 than duration, across all three contexts. In addition, the significant differences in temporal as well as spectral properties of SSBE vowels in disyllabic and monosyllabic words suggest that these results cannot be considered predictive of cross-language speech perception in natural speaking conditions.

Another factor for the differences in the assimilation patterns in bVd and hVba and hVd context could be the nature of the tasks (Strange and Shafer, 2008), i.e. cross-language mapping versus free classification. Task 1 involved listening to six repetitions of each SSBE token in a familiar and unfamiliar context and mapping this to the closest Urdu vowel and rating its goodness. Task 2 did not involve the comparison of SSBE vowels to Urdu vowels or goodness ratings. The results suggest that in free classification task, listeners paid more attention to the articulatory features as compared to the cross-language assimilation task. The presence of monophthongs and diphthongs could have also influenced the perception of these SSBE vowels. In addition, the presence of nonce words in bVd and hVd context might have also influenced the perception of those vowels. However, the results showed that in bVd context, the inclusion of a nonce word (e.g. the SSBE vowel /ʊ/ was embedded in a nonce word /bʊd/) did not affect the perception of this vowel as this vowel was mapped to its Urdu counterpart /ʊ/ with a higher percentage. In order to test the effects of these above-mentioned factors comprehensively, future studies can focus on the perception of monophthongs and diphthongs separately and with and without the inclusion of nonce words.

The question of whether PAM can be extended to multilingual learners of English cannot be answered fully. Firstly, discrimination experiments are required to test the predictions that are derived from the cross-language assimilation task in this study. Secondly, a cross-language mapping task to learners' first/second dominant language (i.e. Punjabi in this case) would also help to better understand the perception of SSBE vowels by Punjabi-Urdu speakers. Thirdly, a comparison of the patterns of perceptual assimilation by Punjabi-Urdu speakers with Sindhi-Urdu speakers or Balochi-Urdu speakers could also provide more insights to better understand the applicability of PAM to multilingual learners of English.

As far as SLM is concerned, the effects of age and experience are not applicable (could not be tested) in the present study because all the participants started to learn English at the age of four or six. In addition, they received similar input for learning English (see Section 1.3 and Section 4.1 for further details). Therefore, in order to test the applicability of SLM to multilingual learners of English, two types of groups are needed, i.e. Punjabi-Urdu speakers who started to learn English at the age of 12 or later versus Punjabi-Urdu speakers who started to learn English at the age of 4 or so. Another experiment can test the perceptual assimilation of SSBE vowels by Punjabi-Urdu speakers who received native input versus Punjabi-Urdu speakers who did not receive native input at all (i.e. the participants in the present study). Further follow-up discrimination experiment will be required to test the predictions from the present study. With these additional experiments, we could determine whether SLM and/or PAM could be extended to multilingual learners of English or a new model would be required.

However, based on the above discussed results from the perception experiments, some predictions can be made, especially with regard to PAM and SLM.  The implications of those predictions for Pakistani English are discussed in the next section.

**7.3 Implications for the Teaching of English in Pakistan**

According to PAM and PAM-L2 (Best, 1995; Best and Tyler, 2007), the following predictions can be made about the perceptual difficulties faced by Punjabi-Urdu listeners when learning SSBE. The perceptual assimilation patterns from both contexts suggest that listeners were able to differentiate some of the SSBE vowels successfully; however, they managed to categorise eight out of 19 SSBE vowels in bVd context and only four out of 19 vowels in hVba context more than 50% of the time. The rest of the vowels were mapped to multiple Urdu vowels, and some of those showed substantial classification overlap. According to PAM (Best, 1995) and PAM-L2 (Best and Tyler, 2007), the classification overlap scores as discussed in Section 5.8 suggest that these vowel pairs will be very difficult for Punjabi-Urdu listeners to differentiate (discriminate), and similarly these difficulties will occur in the production of these vowels (SLM, Flege, 1995). These results suggest that, except for SSBE /iː/, /ɛ/, /ɔː/, /aɪ/, /aʊ/ and /ɔɪ/, Punjabi-Urdu listeners will face difficulties in differentiating and producing SSBE vowels accurately.

These predictions and findings are in line with the limited available literature on Pakistani English (Farooq and Mahmood, 2017; Hassan and Imtiaz, 2015; Sheikh, 2012; Bilal et al., 2011a, 2011b and 2011c; Raza, 2008; Mahboob and Ahmar, 2004; Rahman, 1991). According to the literature on Pakistani English (PE), SSBE /æ/ and /eɪ/ are pronounced as [e]; SSBE /ɒ/, /ɔː/ and /əʊ/ are often pronounced as [o] or [a] depending on the social and educational background of the speakers; and SSBE /ɔɪ/ and /əʊ/ are frequently replaced with a short vowel [ʊ] and/or [oː], respectively. In addition, SSBE /ɜː/, /ɒ/ and /ɔː/ are reported to be missing from PE. SSBE /ɔː/ is more often pronounced as [oː].

Based on these findings and predictions, it can be proposed that training materials for English language teachers, and teaching materials for English language learners in Pakistan, can focus on identification patterns and pronunciation practices (listening and speaking activities) for the SSBE vowels that are found to be most confusing for Punjabi-Urdu speakers (Derwing and Munro, 2005). According to SLM (Flege, 1995), exposure to native input would also increase learners' accuracy in discrimination of these vowels and as a result will improve their speaking and listening skills overall. Previous research has shown that English language teachers' lack of knowledge of phonetics can result in poor teaching and learning, and adopting a more student centered-approach (computer assisted language learning) for perceptual training with synthetic speech stimuli can enhance the perception and production of difficult L2 contrasts such as English /iː/-/ɪ/ and /ɛ/-/æ/ (Wang and Munro, 2004).

For pedagogical purposes, English language teachers can be trained to focus not only on the temporal but also on the spectral qualities of the SSBE vowels that are found to be most confusing for Punjabi-Urdu speakers. Studies on the perceptual and auditory training for English vowels (Lengeris, 2018; Lengeris and Hazan, 2010; Lengeris, 2008 (Greek learners of English); Rato, 2014 (Portuguese learners of English); Alshangiti and Evans, 2014 (Arabic learners of English); Iverson et al., 2012 (French learners of English); Iverson and Evans, 2009 (Spanish and German learners of English), (and consonants (Bradlow et al., 1997, Japanese learners of English)) have shown that the perception of L2 vowels significantly improves with effective perceptual and auditory training, which results in better identification, discrimination and production. Thus, English language teachers can be trained to use perceptual and auditory training methods for teaching of English in Pakistan.

### 7.4 Limitations and Future Work

For acoustic and phonetic investigation of Urdu vowels, firstly the data were analysed only for monophthongs and six diphthongs. The data for nasalised vowels were collected; however, this was not analysed. Secondly, the data in the study were collected from highly educated Punjabi-Urdu speakers who speak English fluently. Future research would focus on Punjabi-Urdu speakers with minimum exposure to English. Future research can also focus on other dialects of Urdu (i.e. Urdu spoken in other parts of the country with different L1s, e.g. Saraiki-Urdu speakers, Pashto-Urdu speakers, and Sindhi-Urdu speakers). These cross-dialect comparisons of Urdu would help to define a standard Urdu dialect, if possible. In addition, future studies can use diphthongs in disyllabic or multisyllabic words with stressed and unstressed syllables, to investigate any differences in the quality of these diphthongs. Future perception and syllable identification tests, from native and non-native speakers, will also help to better establish the status of Urdu diphthongs.

For the investigation of the perceptual assimilation of SSBE vowels there were a number of limitations. First of all, the cross-language mapping, goodness rating and auditory free classification experiments were not paired with discrimination and identification experiments; hence the predictions based on PAM and SLM could not be tested. Future studies would focus on discrimination and identification experiments followed by production experiments in order to test predictions based on the findings from the present study. The perceptual assimilation of SSBE vowels by native SSBE speakers would provide a useful basis for comparison with the perceptual assimilation patterns of non-native/second language learners (e.g. Punjabi-Urdu speakers). This would aid understanding of the role of cross-language similar and/or different phonetic and phonological features in the perception and production of SSBE vowels. Future studies should compare natural speech with citation speech to help investigate the predictions proposed by second language perception and learning theories. Although Urdu has short and long vowels, the perceptual assimilation patterns from three contexts for short and long vowels cannot be generalised because the duration was significantly shorter in the hVba context than the bVd and hVd context. Therefore, future studies can focus on materials where nonce words are also monosyllabic, or the target vowels are extracted from connected speech, for both types of context.

Pairing the current English perception experiments with English speech production experiments will define how Punjabi-Urdu speakers process second language sounds. These findings will then be used to validate and expand on current theories of second language perception and acquisition, such as Best's Perceptual Assimilation Model, Flege's Second Language Speech Learning Model. In addition, English speech production experiments will also help to better define the acoustic and phonetic properties of Pakistani English. Such studies will be fundamental for further research on various aspects of Urdu and Pakistani English, and the findings will be helpful for future research on teaching and learning of English in Pakistan.

Future research will also focus on acoustic and phonetic investigation of Urdu consonants. Such research will allow to document, for the first time, the sound patterns of Urdu as spoken in Punjab, Pakistan, and will provide the first thorough treatment in the literature of the Urdu sound system. As well as being the fundamental springboard to spawn further research into Urdu, the findings will have immediate applications in speech recognition and text-to-speech software for Urdu and second language speech perception.

## References

Adank, P., Smits, R. and Van Hout, R., 2004a. A comparison of vowel normalization procedures for language variation research. *The Journal of the Acoustical Society of America*, *116*(5), pp.3099-3107.

Adank, P., Van Hout, R. and Smits, R., 2004b. An acoustic description of the vowels of Northern and Southern Standard Dutch. *The Journal of the Acoustical society of America*, *116*(3), pp.1729-1738.

Adank, P., Van Hout, R. and Velde, H.V.D., 2007. An acoustic description of the vowels of northern and southern standard Dutch II: Regional varieties. *The Journal of the Acoustical Society of America*, *121*(2), pp.1130-1141.

Adler, D., 2005. *Vioplot: Violin plot.* R package version 0.2. [Online]. Available from: http://wsopuppenkiste.wiso.uni-goettingen.de/~dadler [Accessed 10 February 2018].

Aguilar, L., 1999. Hiatus and diphthong: Acoustic cues and speech situation differences. *Speech communication*, *28*(1), pp.57-74.

Ahmed, A. and Nasir, S. (2002). *Phonotactic Distribution of Consonantal Sounds in Urdu. CRULP Annual Student Report, 2002-2003*. Centre for Research in Urdu Language Processing, Pakistan. [Online]. Available from: http://www.crulp.org/Publication/Crulp_report/CR03_10E.pdf [Accessed 10 October 2014]

Ali, A.R. and Hussain, S., 2010. Automatic diacritization for Urdu. In: *Proceedings of the Conference on Language and Technology (CLT10), 22-24 October, 2010.* Islamabad. Pakistan. pp. 105-111.

Alshangiti, W. and Evans, B.G., 2014. Investigating the domain-specificity of phonetic training for second language learning: Comparing the effects of production and perception training on the acquisition of English vowels by Arabic learners of English. In: Fuchs, S., Grice, M., Hermes, A., Lancia, L., Mücke, D. eds. *The 10th International Seminar for Speech Production 2014 Proceedings, 5 - 8 May, 2014.* Cologne, Germany, pp. 9-12.

Antoniou, M., Tyler, M.D. and Best, C.T., 2012. Two ways to listen: Do L2-dominant bilinguals perceive stop voicing according to language mode? *Journal of phonetics*, *40*(4), pp.582-594.

Aoyama, K., Flege, J.E., Guion, S.G., Akahane-Yamada, R. and Yamada, T., 2004. Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese/r/and English/l/and/r. *Journal of Phonetics*, *32*(2), pp.233-250.

Asu, E.L., Lippus, P., Niit, E. and Türk, H., 2012. The acoustic characteristics of monophthongs and diphthongs in the Kihnu variety of Estonian. *Linguistica Uralica*, *48*(3), pp.161-170.

Atagi, E. and Bent, T., 2013. Auditory free classification of nonnative speech. *Journal of phonetics*, *41*(6), pp.509-519.

Atagi, E. and Bent, T., 2016. Auditory free classification of native and nonnative speech by nonnative listeners. *Applied Psycholinguistics*, *37*(2), pp.241-263.

Baker, W. and Trofimovich, P., 2005. Interaction of native-and second-language vowel system (s) in early and late bilinguals. *Language and speech*, *48*(1), pp.1-27.

Baker, W., Trofimovich, P., Mack, M. and Flege, J.E., 2002. The effect of perceived phonetic similarity on non-native sound learning by children and adults. In: Skarabela, B., Fish, S., and Do, H-J, A. eds. *The 26th annual Boston university conference on language development 2001 Proceedings, 2-4 November.* Boston, MA: Cascadilla Press. 26(1), pp. 36-47.

Ball, M.J., 1976. *Towards a Description of the North Welsh Monophthongs*. University of Essex. MA Thesis.

Ball, M.J., 1983. A spectrographic investigation of three Welsh diphthongs. *Journal of the International Phonetic Association*, *13*(2), pp.82-89.

Barr, D.J., Levy, R., Scheepers, C. and Tily, H.J., 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, *68*(3), pp.255-278.

Bates, D., Mächler, M., Bolker, B. and Walker, S., 2015. Fitting Linear Mixed-Effects Models Using *lme4*. *Journal of Statistical Software* [Online], *67*(1), pp.1-48. Available from: doi:10.18637/jss.v067.i01 [Accessed 10 September, 2017].

Baumgardner, R.J., 1995. Pakistani English: Acceptability and the norm. *World Englishes*, [Online], *14*(2), pp.261-271. Available from: https://doi.org/10.1111/j.1467-971X.1995.tb00355.x [ Accessed 22 January 2018].

Best, C.T. and Strange, W., 1992. Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of phonetics*, *20*(3), pp.305-330.

Best, C.T. and Tyler, M.D., 2007. Non-native and second-language speech perception: Commonalities and complementarities. In: Bohn, O-S. and Munro, M.J., eds. *Language experience in second language speech learning: In honour of James Emil Flege*. Amsterdam/Philadelphia: John Benjamins Publishing Company, pp. 13-34.

Best, C.T., 1995. A direct realist view of cross-language speech perception. In: Strange, W. ed. *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research.* Timonium, MD: York Press. pp.171-206.

Best, C.T., Faber, A. and Levitt, A., 1996. Assimilation of non-native vowel contrasts to the American English vowel system. *The Journal of the Acoustical Society of America*, *99*(4), pp.2602-2603.

Best, C.T., McRoberts, G.W. and Sithole, N.M., 1988. Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of experimental psychology: human perception and performance*, *14*(3), p.345.

Bhatti, R., and Mumtaz, B. 2016. Identification of Diphthongs in Urdu and their Acoustic Properties. In: *The Conference on Language and Technology 2016 (CLT 16) Proceedings, 17-18 November, 2016*. Lahore, Pakistan. [Online]. Available from: http://www.cle.org.pk/clt16/) [Accessed 10 October 2017].

Bilal, H.A., Mahmood, M.A. and Saleem, R.M., 2011a. Acoustic analysis of front vowels of Pakistani English. *International Journal of Academic Research*, *3*(6-I), pp.20-27.

Bilal, H.A., Mahmood, M.A. and Saleem, R.M., 2011b. Merger of /iː/ and /ɪ/ in Pakistani English. *International Journal of Linguistics*, *3*(1), pp. E34 (1-12).

Bilal, H.A., Warraich, A.A., Rasheed, T., Fatima, N. and Tiwana, D.S., 2011c. Acoustic Analysis of Central Vowels of PakE. *International Journal of Linguistics*, *3*(1), pp. E45 (1-13).

Boersma, P. and Weenink, D., 2016. Praat. Doing phonetics by computer. [Online]. Available from: http://www.fon.hum.uva.nl/praat/ [Accessed 12 March, 2016].

Bohn, O.S. and Flege, J.E., 1990. Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied psycholinguistics*, *11*(3), pp.303-328.

Bohn, O.S. and Flege, J.E., 1992. The production of new and similar vowels by adult German learners of English. *Studies in Second Language Acquisition*, *14*(2), pp.131-158.

Bohn, O-S. and Munro, M.J., eds., 2007. *Language experience in second language speech learning: In honour of James Emil Flege*. Vol. 17. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Bokhari, R., and Pervez, S., 2003. Syllabification and Re-Syllabification in Urdu. Akhbar-i-Urdu. *CRULP Annual Student Report, 2002-2003*. Centre for Research in Urdu Language Processing, Pakistan [Online]. Available from: http://www.cle.org.pk/Publication/Crulp_report/CR03_11E.pdf [Accessed 10 October 2014].

Bokhari, S., 1985. *Phonology of Urdu language*. Karachi: Royal Book Company.

Bokhari, S., 1991. *Urdu zubaan ka soti nizaam* . Karachi: Royal Book Company.

Bolton, K., 2008. English in Asia, Asian Englishes, and the issue of proficiency. *English Today 94*, *24*(2), pp.3-12.

Boomershine, A., 2013. The perception of English vowels by monolingual, bilingual, and heritage speakers of Spanish and English. In: Howe, C., Blackwell, S.E. and Quesada, M.L. eds. *Selected Proceedings of the 15th Hispanic Linguistics Symposium, 6-9 October, 2011*. University of Georgia. Somerville, MA, USA: Cascadilla Proceedings Project. pp.103-118.

Bosch, L., Costa, A. and Sebastián-Gallés, N., 2000. First and second language vowel perception in early bilinguals. *European Journal of Cognitive Psychology*, *12*(2), pp.189-221.

Bradlow, A., Clopper, C., Smiljanic, R. and Walter, M.A., 2010. A perceptual phonetic similarity space for languages: Evidence from five native language listener groups. *Speech Communication*, *52*(11-12), pp.930-942.

Bradlow, A.R., Pisoni, D.B., Akahane-Yamada, R. and Tohkura, Y.I., 1997. Training Japanese listeners to identify English /r/and /l/. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America* [Online], *101*(4), pp.2299-2310. Available from: doi: 10.1121/1.418276

Broselow, E., Chen, S.I. and Huffman, M., 1997. Syllable weight: convergence of phonology and phonetics. *Phonology*, *14*(1), pp.47-82.

Brown, C.A., 1998. The role of the L1 grammar in the L2 acquisition of segmental structure. *Second Language Research*, *14*(2), pp.136-193.

Bughio, F.A., 2013. *Improving English language teaching in large classes at university level in Pakistan.* University of Sussex. PhD Thesis.

Bundgaard-Nielsen, R.L., Best, C.T. and Tyler, M.D., 2008. The assimilation of L2 Australian English vowels to L1 Japanese vowel categories: Vocabulary size matters. In: *9th Annual Conference of the International Speech Communication Association 2008 proceedings, 22-26 September, 2008.* Brisbane: Australia. p.1177.

Bundgaard-Nielsen, R.L., Best, C.T. and Tyler, M.D., 2011a. Vocabulary size matters: The assimilation of second-language Australian English vowels to first-language Japanese vowel categories. *Applied Psycholinguistics*, *32*(1), pp.51-67.

Bundgaard-Nielsen, R.L., Best, C.T. and Tyler, M.D., 2011b. Vocabulary size is associated with second-language vowel perception performance in adult learners. *Studies in Second Language Acquisition*, *33*(3), pp.433-461.

Cambridge Assessment English. 2016. Cambridge English: Advanced (CAE). [Online]. Available from: https://www.cambridgeenglish.org/exams-and-tests/advanced/ [Accessed 20 May, 2016].

Chitoran, I., 2002. A perception-production study of Romanian diphthongs and glide-vowel sequences. *Journal of the International Phonetic Association*, *32*(2), pp.203-222.

Chládková, K., Escudero, P. and Boersma, P., 2011. Context-specific acoustic differences between Peruvian and Iberian Spanish vowels. *The Journal of the Acoustical Society of America*, *130*(1), pp.416-428.

Clermont, F., 1993. Spectro-temporal description of diphthongs in F1− F2− F3 space. *Speech communication*, *13*(3-4), pp.377-390.

Clopper, C.G., 2008. Auditory free classification: Methods and analysis. *Behaviour Research Methods, 40*(2), pp.575-581.

Clopper, C.G. and Bradlow, A.R., 2008. Perception of dialect variation in noise: Intelligibility and classification. *Language and speech*, *51*(3), pp.175-198.

Clopper, C.G. and Bradlow, A.R., 2009. Free classification of American English dialects by native and non-native listeners. *Journal of phonetics*, *37*(4), pp.436-451.

Clopper, C.G., Mitsch, J.F. and Tamati, T.N., 2017. Effects of phonetic reduction and regional dialect on vowel production. *Journal of Phonetics*, *60*, pp.38-59.

Clopper, C.G. and Pisoni, D.B., 2007. Free classification of regional dialects of American English. *Journal of Phonetics*, *35*(3), pp.421-438.

Clopper, C.G., Pisoni, D.B. and De Jong, K., 2005. Acoustic characteristics of the vowel systems of six regional varieties of American English. *The Journal of the Acoustical Society of America*, *118*(3), pp.1661-1676.

Clopper, C.G., and Turnbull, R., 2018. Exploring variation in phonetic reduction: Linguistic, social, and cognitive factors. In: Cangemi, F., Clayards, M., Niebuhr, O., Schuppler, B., and Zellers, M. eds. *Rethinking reduction: Interdisciplinary perspectives on conditions, mechanisms, and domains for phonetic variation*. Berlin: de Gruyter, pp.25-72.

Colantoni, L., Steele, J., and Escudero, P., 2015. *Second language speech: Theory and Practice*. Cambridge: Cambridge University Press.

Coleman, H., 2010. *Teaching and learning in Pakistan: The role of language in education*. Islamabad: The British Council Pakistan.

Cox, F., 2006. The acoustic characteristics of /hVd/ vowels in the speech of some Australian teenagers. *Australian journal of linguistics*, *26*(2), pp.147-179.

Cunnings, I., 2012. An overview of mixed-effects statistical models for second language researchers. *Second Language Research*, *28*(3), pp.369-382.

DeKeyser, R.M., 2000. The robustness of critical period effects in second language acquisition. *Studies in second language acquisition*, *22*(4), pp.499-533.

Derwing, T.M. and Munro, M.J., 2005. Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly*, *39*(3), pp.379-397.

Deterding, D., 1997. The formants of monophthong vowels in Standard Southern British English pronunciation. *Journal of the International Phonetic Association*, *27*(1-2), pp.47-55.

Deterding, D., 2000. Measurements of the /eɪ/ and /əʊ/ diphthongs of young English speakers in Singapore. In: Brown, A., Deterding, D and Eeling, L. eds. *The English Language in Singapore: Research on Pronunciation*. Singapore: Singapore Association for Applied Linguistics. pp. 93-99.

Deterding, D., 2005. Emergent patterns in the vowels of Singapore English. *English World-Wide*, *26*(2), pp.179-197.

Deterding, D., 2006. The North Wind versus a Wolf: short texts for the description and measurement of English pronunciation. *Journal of the International Phonetic Association*, *36*(2), pp.187-196.

Díaz Granado, M., 2011. *L2 and L3 acquisition of the Portuguese stressed vowel inventory by native speakers of English*. Tucson, AZ: The University of Arizona. PhD Thesis.

Durand, J., 2005. Tense/lax, the vowel system of English and phonological theory. In: Carr, O., Durand, J. and Ewen, C.J. eds. *Headhood, elements, specification and contrastivity: Phonological papers in honour of John Anderson.* Philadelphia, PA: John Benjamins B.V. pp.77-98.

Edwards, H. T. (2003). *Applied Phonetics: The Sounds of American English*. Singular Publishing Group. 3rd edn. Available from: http://books.google.com/books?hl=en&lr=&id=FYMba6II7fgC&oi=fnd&pg=PR9&dq=URDU+CONSONANTAL+AND+VOCALIC+SOUNDS&ots=AscWJ90oEf&sig=azwePdDd6K0E76N3QlMdEQSw7_A [Accessed 10 October 2014] .

Elvin, J. and Escudero, P., 2015. Predicting vowel discrimination accuracy through cross-linguistic acoustic analyses. In: ICPhS. *18th International Congress of Phonetic Sciences, Proceedings, 10-14 August, 2015*. SECC Glasgow: Scotland, UK. [Online]. Available from: https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/proceedings.html [Accessed 22 December, 2017].

Elvin, J., Escudero, P. and Vasiliev, P., 2014. Spanish is better than English for discriminating Portuguese vowels: acoustic similarity versus vowel inventory size. *Frontiers in psychology* [Online], *5*. Article 1188, pp. 1-8. Available from: doi: 10.3389/fpsyg.2014.01188 [ Accessed 28 December 2017].

Elvin, J., Williams, D. and Escudero, P., 2016. Dynamic acoustic properties of monophthongs and diphthongs in Western Sydney Australian English. *The Journal of the acoustical society of America*, *140*(1), pp.576-581.

Escudero, P. and Boersma, P., 2002. The subset problem in L2 perceptual development: Multiple-category assimilation by Dutch learners of Spanish. In: Skarabela, B., Fish, S., and Do, H-J, A. eds. *The 26th annual Boston university conference on language development 2001* Proceedings, *2-4 November.* Boston, MA: Cascadilla Press. 26(1), pp. 208-219.

Escudero, P. and Boersma, P., 2004. Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*, *26*(4), pp.551-585.

Escudero, P. and Chládková, K., 2010. Spanish listeners' perception of American and Southern British English vowels. *The Journal of the Acoustical Society of America*, *128*(5), pp. EL254-EL260.

Escudero, P. and Williams, D., 2011. Perceptual assimilation of Dutch vowels by Peruvian Spanish listeners. *The Journal of the Acoustical Society of America*, *129*(1), pp.EL1-EL7.

Escudero, P., Benders, T. and Lipski, S.C., 2009. Native, non-native and L2 perceptual cue weighting for Dutch vowels: The case of Dutch, German, and Spanish listeners. *Journal of Phonetics*, *37*(4), pp.452-465.

Escudero, P., Broersma, M. and Simon, E., 2013. Learning words in a third language: Effects of vowel inventory and language proficiency. *Language and Cognitive Processes*, *28*(6), pp.746-761.

Escudero, P., Simon, E. and Mitterer, H., 2012. The perception of English front vowels by North Holland and Flemish listeners: Acoustic similarity predicts and explains cross-linguistic and L2 perception. *Journal of Phonetics*, *40*(2), pp.280-288.

Escudero, P., Sisinni, B. and Grimaldi, M., 2014. The effect of vowel inventory and acoustic properties in Salento Italian learners of Southern British English vowels. *The Journal of the Acoustical Society of America*, *135*(3), pp.1577-1584.

Escudero, P.., 2005. *Linguistic perception and second language acquisition: explaining the attainment of optimal phonological categorization*. Utrecht University. The Netherlands: LOT. PhD Thesis.

Evans, B.G. and Alshangiti, W., 2018. The perception and production of British English vowels and consonants by Arabic learners of English. *Journal of Phonetics*, *68*, pp.15-31.

Ewald, O., 2016. *The Estonian Swedish diphthongs /ai au oi ui/: Acoustic characteristics and cross-dialectal variation*. Centre for Language and Literature: Lund University. MA Thesis.

Fabiano-Smith, L. and Goldstein, B.A., 2010. Phonological acquisition in bilingual Spanish–English speaking children. *Journal of Speech, Language, and Hearing Research*. *53*, pp. 160-178.

Faris, M.M., Best, C.T. and Tyler, M.D., 2016. An examination of the different ways that non-native phones may be perceptually assimilated as uncategorized. *The Journal of the Acoustical Society of America*, *139*(1), pp.EL1-EL5.

Farooq, M. and Mahmood, M.A., 2017. Acoustic Behaviour of RP Diphthongs in Pakistani English (PakE). *Balochistan Journal of Linguistics ISSN-No. 2312-5454*. [Online]. *2017(5)*, pp.85-105. Available from: http://www.luawms.edu.pk/wp-content/uploads/2018/04/BJL-2017-Issue-5.pdf#page=85 [Accessed 22 January, 2019].

Farooq, M. and Mumtaz, B., 2016. Urdu Phonological Rules in Connected Speech. In: *The Conference on Language and Technology 2016 (CLT 16) Proceedings, 17-18 November,* *2016.* Lahore, Pakistan. [Online]. Available from: http://www.cle.org.pk/clt16/ [Accessed 10 October 2017].

Farooq, M., 2014. *An Acoustic Phonetic Study of Six Accents of Urdu in Pakistan*. Department of English Language and Literature, School of Social Sciences and Humanities: University of Management and Technology, Johar Town, Lahore, Pakistan. MPhil Thesis.

Fatima, N., Aden, R., 2003. Vowel Structure of Urdu: *CRULP Annual Student Report, 2002-2003*. Centre for Research in Urdu Language Processing, Pakistan. [ [Online]. Available from: http://www.cle.org.pk/Publication/Crulp_report/CR03_04E.pdf [Accessed 10 October, 2014].

Ferozuddin, M., n.d. *Feroz-ul-Lughat.* Karachi, Pakistan: Feroz Sons Limited.

Ferragne, E. and Pellegrino, F., 2010. Formant frequencies of vowels in 13 accents of the British Isles. *Journal of the International Phonetic Association*, *40*(01), pp.1-34.

Flege, J. E., Bohn, O. S., and Jang, S. 1997. Effects of Experience on non-native Speakers' Production and Perception of English Vowels. *Journal of Phonetics*, *25*(4), 437–470.

Flege, J.E. and Hillenbrand, J., 1986. Differential use of temporal cues to the /s/–/z/ contrast by native and non-native speakers of English. *The Journal of the Acoustical Society of America*, *79*(2), pp.508-517.

Flege, J.E. and MacKay, I.R., 2004. Perceiving vowels in a second language. *Studies in second language acquisition*, *26*(1), pp.1-34.

Flege, J.E., 1987. The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of phonetics*, *15*(1), pp.47-65.

Flege, J.E., 1992. Speech learning in a second language. In: Ferguson, C., Menn, L. and Stoel-Gammon, C. eds. *Phonological development: models, research, and implications.* Timonium MD: York Press, pp. 565-604.

Flege, J.E., 1995. Second-language speech learning: Theory, findings, and problems. In: Strange, W. ed. *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research.* Timonium, MD: York Press. pp.233-277.

Flege, J.E., 1999. Age of learning and second language speech. In: Birdsong, D. ed. *Second language acquisition and the critical period hypothesis.* Mahwah, New Jersey, London: Lawrence Erlbaum Associates, Inc, pp. 101-132.

Flege, J.E., 2002. Interactions between the native and second-language phonetic systems. In: Burmeister, P. Piske, T. and Rohde, A. eds. *An integrated view of language development. Papers in honour of Henning Wode*. Trier: Wissenschaftlicher Verlag, pp. 217-244.

Flege, J.E., 2003. Assessing constraints on second-language segmental production and perception. In: Mayer, A and Schiller, N. eds. *Phonetics and phonology in language comprehension and production: Differences and similarities*. Berlin: Mouton, pp.319-355.

Flege, J.E., 2007. *Language experience in second language speech learning.* In: Bohn, O-S. and Munro, M.J., eds., 2007. *Language experience in second language speech learning: In honor of James Emil Flege*. Vol. 17. Amsterdam/Philadelphia: John Benjamins Publishing Company, pp. 13-34.

Flege, J.E., Guion, S., Akahane-Yamada, R. and Downs-Pruitt, J., 1998, June. Categorical discrimination of English and Japanese vowels and consonants by native Japanese and English subjects. In: *The 16th International Congress on Acoustics and the 135th* *meeting of the Acoustical Society of America Proceedings, 1998, 20-26 June 1998.* Washington, USA: Seattle. Vol. 4, pp. 2973-2974.

Flege, J.E., MacKay, I.R. and Meador, D., 1999. Native Italian speakers' perception and production of English vowels. *The Journal of the Acoustical Society of America*, *106*(5), pp.2973-2987.

Flege, J.E., Munro, M.J. and MacKay, I.R., 1995. Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America*, *97*(5), pp.3125-3134.

Flege, J.E., Schirru, C. and MacKay, I.R., 2003. Interaction between the native and second language phonetic subsystems. *Speech communication*, *40*(4), pp.467-491.

Flege, J.E., Takagi, N. and Mann, V., 1996. Lexical familiarity and English-language experience affect Japanese adults' perception of /ɹ/ and /l/. *The Journal of the Acoustical Society of America*, *99*(2), pp.1161-1173.

Flynn, N., 2011. Comparing vowel formant normalisation procedures. *York papers in linguistics*, *2*(11), pp.1-28.

Fourakis, M., Botinis, A. and Katsaiti, M., 1999. Acoustic characteristics of Greek vowels. *Phonetica*, 56(1-2), pp.28–43.

Fowler, C.A., Sramko, V., Ostry, D.J., Rowland, S.A. and Hallé, P., 2008. Cross language phonetic influences on the speech of French–English bilinguals. *Journal of Phonetics*, *36*(4), pp.649-663.

Fox, J. and Weisberg, S., 2011. Multivariate linear models in R. *An R companion to applied regression.* 2nd edn. Los Angeles, Thousand Oaks, CA: Sage.

Fox, R.A. and Jacewicz, E., 2009. Cross-dialectal variation in formant dynamics of American English vowels. *The Journal of the Acoustical Society of America*, *126*(5), pp.2603-2618

Fox, R.A., 1983. Perceptual structure of monophthongs and diphthongs in English. *Language and speech*, *26*(1), pp.21-60.

Fox, R.A., Flege, J.E. and Munro, M.J., 1995. The perception of English and Spanish vowels by native English and Spanish listeners: A multidimensional scaling analysis. *The Journal of the Acoustical Society of America*, *97*(4), pp.2540-2551.

Frieda, E.M., Walley, A.C., Flege, J.E. and Sloane, M.E., 1999. Adults' perception of native and nonnative vowels: Implications for the perceptual magnet effect. *Perception & Psychophysics*, *61*(3), pp.561-577.

Fuchs, R., 2015. You're Not from Around Here, Are You? In: Delais-Roussarie, E., Avanzi, M and Herment, S. eds. *Prosody and Language in Contact: L1 Acquisition, Attrition and Language in Multilingual Situations*. Berlin, Heidelberg: Springer, pp. 123-148.

Gargesh, R., 2004. Indian English: Phonology. In: Schneider, E.W., Burridge, K., Mesthrie, R and Upton, C. eds. *A handbook of varieties of English Varieties of English.* New York: Mouton de Gruyter, pp.992-1002.

Gay, T., 1968. Effect of speaking rate on diphthong formant movements. *The Journal of the Acoustical Society of America*, *44*(6), pp.1570-1573.

Gahl, S. and Baayen, R.H., 2019. Twenty-eight years of vowels: Tracking phonetic variation through young to middle age adulthood. *Journal of Phonetics*, *74*, pp.42-54.

Ghazali, M. A., 2001. *Urdu Syllable Templates. CRULP Annual Student Report, 2001-2002*. Center for Research in Urdu Language Processing. [Online]. Available from:

http://www.cle.org.pk/Publication/Crulp_report/CR02_23E.pdf [Accessed 10 October 2014].

Gilichinskaya, Y.D. and Strange, W., 2010. Perceptual assimilation of American English vowels by inexperienced Russian listeners. *The Journal of the Acoustical Society of America*. [Online] *128*(2), pp.EL80-EL85. Available from: doi: 10.1121/1.3462988 [Accessed 24 May, 2016].

Golestani, N. and Zatorre, R. J., 2009. Individual Differences in the Acquisition of Second Language Phonology. *Brain and Language*, *109*(2), 55–67.

Gottfried, T.L. and Beddor, P.S., 1988. Perception of temporal and spectral information in French vowels. *Language and Speech*, *31*(1), pp.57-75.

Government of the Punjab (GoP). 2015. *Punjab Development Statistics.* Bureau of Statistics. Lahore. [Online]. Available from: http://www.bos.gop.pk/publicationreports [Accessed 29 September, 2017].

Guion, S.G., Flege, J.E., Akahane-Yamada, R. and Pruitt, J.C., 2000. An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *The Journal of the Acoustical Society of America*, *107*(5), pp.2711-2724.

Haddican, B., Foulkes, P., Hughes, V. and Richards, H., 2013. Interaction of social and linguistic constraints on two vowel changes in northern England. *Language Variation and Change*, *25*(3), pp.371-403.

Hallé, P.A., Best, C.T. and Levitt, A., 1999. Phonetic vs. phonological influences on French listeners' perception of American English approximants. *Journal of Phonetics*, *27*(3), pp.281-306.

Hammarberg, B. and Hammarberg, B., 2009. Re-setting the basis of articulation in the acquisition of new languages: A third language case study. In: Hammarberg, B. ed. *Processes in third language acquisition*. Edinburgh: Edinburgh University Press, pp. 74-84.

Hammarberg, B., 2014. Problems in defining the concepts of L1, L2 and L3. In: Otwinowska, A. and De Angelis, G. eds. *Teaching and learning in multilingual contexts: Sociolinguistic and educational perspectives*. Bristle, Buffalo, Toronto: Multilingual Matters, pp. 3-18.

Harrington, J. and Cassidy, S., 1994. Dynamic and target theories of vowel classification: Evidence from monophthongs and diphthongs in Australian English. *Language and Speech*, *37*(4), pp.357-373.

Harrington, J., 2010. *Phonetic analysis of speech corpora*. Wiley-Blackwell. [Online]. Available from: https://www.phonetik.uni-muenchen.de/~jmh/research/pasc010808/pasc.pdf [Accessed 20 December, 2016].

Harrington, J., Cox, F. and Evans, Z., 1997. An acoustic analysis of cultivated, general and broad Australian English speech. *Australian Journal of Linguistics*, *17*, pp.155-84.

Harrington, J., Palethorpe, S. and Watson, C., 2000. Monophthongal vowel changes in Received Pronunciation: an acoustic analysis of the Queen's Christmas broadcasts. *Journal of the International Phonetic Association*, *30*(1-2), pp.63-78.

Harrington, J., Palethorpe, S. and Watson, C., 2000. Monophthongal vowel changes in Received Pronunciation: an acoustic analysis of the Queen's Christmas broadcasts. *Journal of the International Phonetic Association*, *30*(1-2), pp.63-78.

Harris, M.S. and Umeda, N., 1974. Effect of speaking mode on temporal factors in speech: vowel duration. *The Journal of the Acoustical Society of America*, 56(3), pp.1016-1018.

Hassan, S. and Imtiaz, S., 2015. Acoustic Analysis of Back Vowels (/ɒ/ /ɔː/) in Pakistani English. *Global Journal of Human-Social Science Research (G)*. Linguistics and Education. [Online]. Available from: http:// creativecommons.org/licenses/by-nc/3.0/ [Accessed 10 January, 2017].

Hawkins, S. and Midgley, J., 2005. Formant frequencies of RP monophthongs in four age groups of speakers. *Journal of the International Phonetic Association, 35*(02), pp.183–199.

Hernandez, A.E., Dapretto, M., Mazziotta, J. and Bookheimer, S., 2001. Language switching and language representation in Spanish–English bilinguals: An fMRI study. *Neuro Image, 14*(2), pp.510-520.

Hillenbrand, J., Getty, L.A., Clark, M.J. and Wheeler, K., 1995. Acoustic characteristics of American English vowels. *The Journal of the Acoustical society of America*, *97*(5), pp.3099-3111.

Hillenbrand, J.M., 2003. American English: Southern Michigan. *Journal of the International Phonetic Association*, *33*(1), pp.121-126.

Hillenbrand, J.M., 2013. Static and dynamic approaches to vowel perception. In: Morrison, G.S. and Assmann, P.F. eds. *Vowel inherent spectral change*. Berlin Heidelberg: Springer, pp. 9-30.

Hillenbrand, J.M., Clark, M.J. and Nearey, T.M., 2001. Effects of consonant environment on vowel formant patterns. *The Journal of the Acoustical Society of America*, *109*(2), pp.748-763.

House, A.S., 1961. On vowel duration in English. *The Journal of the Acoustical Society of America*, *33*(9), pp.1174-1178.

Hussain, I., 2005. Education, employment and economic development in Pakistan. Inaugural Address at: The conference on education, 15 April, 2005, Woodrow Wilson Centre, Washington D.C. [Online]. pp.1-18. Available from: https://ishrathusain.iba.edu.pk/speeches/humanDevelopment/Edu_Emp_Dev_Apr_15.pdf [Accessed 20 April, 2015].

Hussain, Q., Mahmood, R. and Mahmood, M.A., 2011. Vowel substitution: A comparative study of English loans in Punjabi and Urdu. *International Journal of Linguistics*, *3*(1), pp. E31(1-13).

Hussain, Q., Mahmood, R., and Mahmood, M. A. 2012. Phonological Make-up of English Loanwords. *Journal of Language Teaching and Research*, *3*(5), pp. 838-843.

Hussain, S. 2005. Phonological Processing for Urdu Text to Speech System. In: Yadava, Y, Bhattarai, G, Lohani, RR, Prasain, B and Parajuli, K. eds. *Contemporary issues in Nepalese linguistics*. Linguistics Society of Nepal, Kathmandu, Nepal. Available from: http://www.cle.org.pk/Publication/papers/2005/Phonological_Processing-4-Urdu_Text-to-Speech_System.pdf [Accessed 10 October, 2014].

Hussain, S., 1997. *Phonetic correlates of lexical stress in Urdu*. Communication Sciences and Disorders-Speech and Language Pathology. Evanston: Illinois. PhD Thesis.

Hussain, S., 2004. Letter-to-sound conversion for Urdu text-to-speech system. In: *The workshop on computational approaches to Arabic script-based languages 2004 Proceedings, 28 August, 2004*. Geneva, Switzerland: Association for Computational Linguistics, pp. 74-79.

Ijaz, M. and Hussain, S., 2007., Corpus based Urdu lexicon development. In: *The Conference on Language Technology (CLT07) 2007 Proceedings, 7-11 August, 2007.* University of Peshawar, Pakistan.

Ingram, J.C. and Park, S.G., 1997. Cross-language vowel perception and production by Japanese and Korean learners of English. *Journal of phonetics*, *25*(3), pp.343-370.

Iverson, P. and Evans, B.G., 2007. Learning English vowels with different first-language vowel systems: Perception of formant targets, formant movement, and duration. *The Journal of the Acoustical Society of America*, *122*(5), pp.2842-2854.

Iverson, P. and Evans, B.G., 2009. Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers. *The Journal of the Acoustical Society of America*, *126*(2), pp.866-877.

Iverson, P. and Kuhl, P.K., 1996. Influences of phonetic identification and category goodness on American listeners' perception of /r/and /l/. *The Journal of the Acoustical Society of America*, *99*(2), pp.1130-1140.

Iverson, P., Kuhl, P.K., Akahane-Yamada, R., Diesch, E., Tohkura, Y.I., Kettermann, A. and Siebert, C., 2003. A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87*(1), pp. B47-B57.

Iverson, P., Pinet, M. and Evans, B.G., 2012. Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels. *Applied Psycholinguistics*, *33*(1), pp.145-160.

Jacewicz, E. and Fox, R.A., 2013. Cross-dialectal differences in dynamic formant patterns in American English vowels. In: Morrison, G.S. and Assmann, P.F. eds. *Vowel inherent spectral change*. Berlin Heidelberg: Springer. pp. 177-198.

Jaeger, T.F., 2008. Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of memory and language*, *59*(4), pp.434-446.

Javed, S.T., Hussain, S., Maqbool, A., Asloob, S., Jamil, S. and Moin, H., 2010. Segmentation free Nastalique Urdu OCR. World Academy of Science, Engineering and Technology. *International Journal of Computer and Information Engineering*. *4*(10), pp.1514-1519.

Jia, G., Strange, W., Wu, Y., Collado, J. and Guan, Q., 2006. Perception and production of English vowels by Mandarin speakers: Age-related differences vary with amount of L2 exposure a. *The Journal of the Acoustical Society of America*, *119*(2), pp.1118-1130.

Johnson, K., 2005. Speaker normalization in speech perception. In: Pisoni, D.B. and Remez, R.E. eds. *The handbook of speech perception*. Blackwell Publishing Ltd, pp.363-389.

Johnson, K., Ladefoged, P. and Lindau, M., 1993. Individual differences in vowel production. *The Journal of the Acoustical Society of America*, *94*(2), pp.701-714.

Jones, W.E., 1971. Syllables and word-stress in Hindi. *Journal of the International Phonetic Association*, *1*(2), pp.74-78.

Kachru, B.B. ed., 1992. *The other tongue: English across cultures*. 2nd edn. Urbana and Chicago: University of Illinois Press.

Kachru, B.B., 1998. English as an Asian language. *Links & letters*, (5), pp.89-108.

Kachru, Y. 1987. Hindi-Urdu. In: Comrie, B. ed. *The world's major languages*. 2nd edn. New York: Routledge, pp.53-72.

Kent, R.D., 1993. Vocal tract acoustics. *Journal of Voice, 7*(2), pp.97-117.

Kent, R.D., Read, C. and Kent, R.D., 1992. *The acoustic analysis of speech*. Vol. 58. San Diego: Singular Publishing Group.

Kessler, B. and Treiman, R., 1997. Syllable structure and the distribution of phonemes in English syllables. *Journal of Memory and language*, *37*(3), pp.295-311.

Khan, A.G. and Alward, M.A., 2011. Arabic Loanwords in Urdu: A Linguistic Analysis. *Speech and Context: International Journal of Linguistics, Semiotics and Literary Science*, *1*, pp.31-37.

Khan, H.I., 2012. The evolution of Pakistani English (PakE) as a legitimate variety of English. *International Journal of Applied Linguistics and English Literature*, *1*(5), pp.90-99.

Khurshid, k., Usman, S. A., and Butt, N. J., 2004. Possibility of Existence and Identification of Diphthongs and Triphthongs in the Urdu Language. *CRULP Annual Student Report, 2003-2004.* Center for Research in Urdu Language Processing. [Online] Available from: http://www.cle.org.pk/Publication/Crulp_report/CR04_03E.pdf [Accessed 4 March, 2015].

Khurshid, K., Usman, S.A. and Javaid, N., 2003. Possibility of Existence and Identification of Diphthongs and Triphthongs in Urdu Language. *CRULP Annual Student Report 2003-2004*. Center for Research in Urdu Language Processing. [Online], pp. 16-21. Available from: http://www.cle.org.pk/Publication/Crulp_report/CR04_03E.pdf [Accessed 4 March, 2015].

Kirkpatrick, A., 2011. English as an Asian lingua franca and the multilingual model of ELT. *Language Teaching*, *44*(2), pp.212-224.

Kirtley, M.J., Grama, J., Drager, K. and Simpson, S., 2016. An acoustic analysis of the vowels of Hawai ‘i English. *Journal of the International Phonetic Association*, *46*(1), pp.79-97.

Krueger, C. and Tian, L., 2004. A comparison of the general linear mixed model and repeated measures ANOVA using a dataset with multiple missing data points. *Biological research for nursing*, *6*(2), pp.151-157.

Kruskal, J.B., 1964. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, *29*(1), pp.1-27.

Kuhl, P.K. and Iverson, P., 1995. Linguistic experience and the “perceptual magnet effect”. In: Strange, W. ed. *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research*. Baltimore, MD: York Press, pp.121-154.

Kuhl, P.K., 1991. Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Attention, Perception, & Psychophysics*, *50*(2), pp.93-107.

Kuhl, P.K., 1993. Innate predispositions and the effects of experience in speech perception: The native language magnet theory. In: de Boysson-Bardies, B., de Schonen, S., Jusczyk, P., McNeilage, P. and Morton, J. eds. *Developmental neurocognition: Speech and face processing in the first year of life*. Vol. 69. 1st edn. Dordrecht: Springer. pp. 259-274.

Kumaravadivelu, B., 1994. The Other Tongue: English Across Cultures. *Studies in Second Language Acquisition*, *16*(4), p.485.

Kuznetsova, A., Brockhoff, P.B. and Christensen, R.H.B., 2017. *lmer*Test package: tests in linear mixed effects models. *Journal of Statistical Software*, [Online], *82*(13), pp. 1–26. Available from: doi: 10.18637/jss.v082.i13. [Accessed 20 January, 2018].

Ladefoged, P. and Disner, S.F., 2012. *Vowels and consonants*. 3rd edn. Wiley-Blackwell.

Ladefoged, P. and Johnson, K., 2011. *A course in phonetics*. 6th edn. Wadsworth: Cengage Learning.

Ladefoged, P. and Johnson, K., 2014. *A course in phonetics*. 7th edn. Wadsworth: Cengage Learning.

Ladefoged, P., 1999. American English. In: *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge: Cambridge University Press.

Ladefoged, P., 2003. *Phonetic data analysis: An introduction to fieldwork and instrumental techniques*. Malden, MA: Wiley-Blackwell.

Lado, R., 1957. Linguistics Across Cultures: Applied Linguistics for Language Teachers.

Lengeris, A. and Hazan, V., 2010. The effect of native vowel processing ability and frequency discrimination acuity on the phonetic training of English vowels for native speakers of Greek. *The Journal of the Acoustical Society of America*, *128*(6), pp.3757-3768.

Lengeris, A., 2008. The effectiveness of auditory phonetic training on Greek native speakers' perception and production of Southern British English vowels. In: Botinis, A. ed. *The ISCA Tutorial and Workshop on Experimental Linguistics (ExLing 2008) Proceedings, 25-27 August, 2008*. Athens: Greece, pp. 133-136.

Lengeris, A., 2009. Perceptual assimilation and L2 learning: Evidence from the perception of Southern British English vowels by native speakers of Greek and Japanese. *Phonetica*, *66*(3), pp.169-187.

Lengeris, A., 2018. Computer-based auditory training improves second-language vowel production in spontaneous speech. *The Journal of the Acoustical Society of America*, *144*(3), pp. EL165-EL171.

Lenneberg, E.H., 1967. *Biological Foundations of Language*. New York: Wiley.

Lenth, R.V., 2016. Least-Squares Means: The R Package lsmeans. *Journal of Statistical Software* [Online], *69*(1), pp.1-33. Available from: doi:10.18637/jss.v069.i01 [Accessed 18 August, 2017].

Levy, E.S. and Law, F.F., 2010. Production of French vowels by American-English learners of French: Language experience, consonantal context, and the perception-production relationship. *The Journal of the Acoustical Society of America*, *128*(3), pp.1290-1305.

Levy, E.S. and Strange, W., 2008. Perception of French vowels by American English adults with and without French language experience. *Journal of phonetics*, *36*(1), pp.141-157.

Levy, E.S., 2009. Language experience and consonantal context effects on perceptual assimilation of French vowels by American-English learners of French. *The Journal of the Acoustical Society of America*, *125*(2), pp.1138-1152.

Lindau, M., Norlin, K. and Svantesson, J.O., 1990. Some cross-linguistic differences in diphthongs. *Journal of the International Phonetic Association*, *20*(1), pp.10-14.

Lipińska, D., 2017. The influence of L2 status on L3 pronunciation. *English Insights*, *1*, pp.69-86.

Lively, S.E. and Pisoni, D.B., 1997. On prototypes and phonetic categories: a critical assessment of the perceptual magnet effect in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(6), pp.1665-1679.

Llama, R., Cardoso, W. and Collins, L., 2010. The influence of language distance and language status on the acquisition of L3 phonology. *International Journal of Multilingualism*, *7*(1), pp.39-57.

Lobanov, B.M., 1971. Classification of Russian vowels spoken by different speakers. *The Journal of the Acoustical Society of America*, *49*(2B), pp.606-608.

Lotto, A.J., Kluender, K.R. and Holt, L.L., 1998. Depolarizing the perceptual magnet effect. *The Journal of the Acoustical Society of America*, *103*(6), pp.3648-3655.

Maclagan, M. and Hay, J., 2007. Getting fed up with our feet: Contrast maintenance and the New Zealand English "short" front vowel shift. *Language variation and change, 19*(1), pp.1-25.

Mahboob, A and Ahmar, N.H., 2004. Pakistani English: Phonology. In: Schneider, E.W., Burridge, K., Mesthrie, R and Upton, C. eds. *A handbook of varieties of English Varieties of English.* New York*:* Mouton de Gruyter, pp.1003-1016.

Mahboob, A., 2002. No English, no future! *Language Policy in Pakistan.* In: Obeng, S. G. and Hartford, B. eds. *Political Independence with Linguistic Servitude: The Politics about Languages in the Developing World*. New York: Nova Science Publishers, pp.15-39.

Mahboob, A., 2017. English medium instruction in higher education in Pakistan: Policies, perceptions, problems, and possibilities. In: Walkinsah, I., Fenton-Smith, B. and Humphreys, P. eds. *English as a Medium of Instruction in Higher Education in Asia-Pacific: Issues and Challenges*. London: Springer, pp.71-92.

Mahmood, M.A., 2009. *A corpus-based analysis of Pakistani English*. Department of English Language and Literature. Bahauddin Zakariya University, Multan, Pakistan. PhD Thesis.

Mahmood, R., Hussain, Q., and Mahmood, A., 2011. Phonological Adaptations of English words borrowed into Punjabi. *European Journal of Social Sciences, 22*(2), 234-245.

Man-khongdi, P., Onsuwan, C. and Tantibundhit, C., 2016. Durational and spectral differences in Thai diphthongs and final glides. *The Journal of the Acoustical Society of America*, *140*(4), pp.3109-3109.

Mansoor, S., 2003. Language planning in higher education: Issues of access and equity. *The Lahore Journal of Economics*, 8, pp.17-42.

Mansoor, S., 2004. The status and role of regional languages in higher education in Pakistan. *Journal of Multilingual and Multicultural Development*, *25*(4), pp.333-353.

Marian, V., Blumenfeld, H.K. and Kaushanskaya, M., 2007. The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*. *50*, pp.940-967.

Maxwell, O. and Fletcher, J., 2009. Acoustic and durational properties of Indian English vowels. *World Englishes*, *28*(1), pp.52-69.

Maxwell, O. and Fletcher, J., 2010. The acoustic characteristics of diphthongs in Indian English. *World Englishes*, *29*(1), pp.27-44.

Mayr, R. and Davies, H., 2009. The monophthongs and diphthongs of north-eastern Welsh: An acoustic study. In: *10th Annual Conference of the International Speech Communication Association (INTERSPEECH 2009) Proceedings, 6-10 September, 2009.* Brighton: UK. pp. 2315-2318.

Mayr, R. and Davies, H., 2011. A cross-dialectal acoustic study of the monophthongs and diphthongs of Welsh. *Journal of the International Phonetic Association*, *41*(1), pp.1-25.

Mayr, R., Morris, J., Mennen, I. and Williams, D., 2017. Disentangling the effects of long-term language contact and individual bilingualism: The case of monophthongs in Welsh and English. *International Journal of Bilingualism*, *21*(3), pp.245-267.

McMahon, A., 2002. *An introduction to English phonology*. Edinburgh University Press.

Memon, G.R., 2007. Education in Pakistan: The key issues, problems and the new challenges. *Journal of Management and Social Sciences*, *3*(1), pp.47-55.

Morrison, G.S., 2009. Likelihood-ratio forensic voice comparison using parametric representations of the formant trajectories of diphthongs. *The Journal of the Acoustical Society of America*, *125*(4), pp.2387-2397.

Morrison, G.S., 2013. Theories of vowel inherent spectral change. In: Morrison, G.S. and Assmann, P.F. eds. *Vowel inherent spectral change.* Berlin, Heidelberg: Springer, pp. 31-48.

Muth, C., Bales, K.L., Hinde, K., Maninger, N., Mendoza, S.P. and Ferrer, E., 2016. Alternative models for small samples in psychological research: applying linear mixed effects models and generalized estimating equations to repeated measures data. *Educational and psychological measurement*, *76*(1), pp.64-87.

Mutonya, M., 2008. African Englishes: Acoustic analysis of vowels. *World Englishes*, *27*(3-4), pp.434-449.

Narang, G.C. and Becker, D.A., 1971. Aspiration and nasalization in the generative phonology of Hindi-Urdu. *Language*, pp.646-667.

Nayyar, S., 2002. Syllable stress in Urdu. *CRULP Annual Student Report, 2002-2003*. Centre for Research in Urdu Language Processing, Pakistan. [Online]. Available from: https://pdfs.semanticscholar.org/b533/d167d4c704ea1daee398459586b2adaa1516.pdf [Accessed 28 January, 2016].

Nazar, N. 2002. Syllable Templates in Urdu Language. *CRULP Annual Student Report, 2001-2002*. Center for Research in Urdu Language Processing, pp. 189-194.

NCSS Statistical Software., 2017. *Hierarchical Clustering/Dendrograms*. [Online] Available from: https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Hierarchical_Clustering-Dendrograms.pdf [Accessed 14 March, 2017].

Nishi, K., Strange, W., Akahane-Yamada, R., Kubo, R. and Trent-Brown, S.A., 2008. Acoustic and perceptual similarity of Japanese and American English vowels. *The journal of the Acoustical Society of America*, *124*(1), pp.576-588.

Ogden, R., 2009. *An introduction to English phonetics*. Edinburgh University Press.

Ohala, J.J., 1986. Consumer's guide to evidence in phonology. *Phonology*, *3*, pp.3-26.

Ohala, M., 1983. *Aspects of Hindi phonology*. Vol. 2. Delhi, Varanasi, Patna: Motilal Banarsidass Publisher.

Pakistan Bureau of Statistics (PBS) 2017. *Population Census*. [Online]. Available from: http://www.pbs.gov.pk/content/population-census [Accessed 16 December 2019].

Patkowski, M.S., 1990. Age and accent in a second language: A reply to James Emil Flege. *Applied linguistics*, *11*(1), pp.73-89.

Pätzold, M. and Simpson, A.P., 1997. Acoustic analysis of German vowels in the Kiel Corpus of Read Speech. *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung Universität Kiel*, *32*, pp.215-247.

Peterson, G.E. and Barney, H.L., 1952. Control methods used in a study of the vowels. *The Journal of the acoustical society of America*, *24*(2), pp.175-184.

Phull, D.K. and Kumar, G.B., 2016. Vowel Analysis for Indian English. *Procedia Computer Science*, *93*, pp.533-538.

Pierrehumbert, J.B., 2003. Phonetic diversity, statistical learning, and acquisition of phonology. *Language and speech*, *46*(2-3), pp.115-154.

Piske, T., MacKay, I.R. and Flege, J.E., 2001. Factors affecting degree of foreign accent in an L2: A review. *Journal of phonetics*, *29*(2), pp.191-215.

Polka, L. and Bohn, O.S., 1996. A cross-language comparison of vowel perception in English-learning and German-learning infants. *The Journal of the Acoustical Society of America, 100*(1), pp.577-592.

Politzer-Ahles, S. and Piccinini, P., 2018. On visualizing phonetic data from repeated measures experiments with multiple random effects. *Journal of Phonetics*, *70*, pp.56-69.

PPTAlchemy. 2019. [Online]. Available from: http://www.pptalchemy.co.uk/Powerpoint_drag_and_drop.html [Accessed 18 February 2019].

Rahman, T. 1991. Pakistani English: Some Phonological and Phonetic Features. *World Englishes,* 10(1), pp. 83-95.

Rahman, T. 2015. *Pakistani English: Revised Edition*. Islamabad: National Institute of Pakistan Studies. Quaid-i-Azam University.

Rahman, T., 1996. *Language and Politics in Pakistan*. Karachi, Pakistan: Oxford University Press.

Rahman, T., 2008. Language policy and education in Pakistan. In: Aneel, S.S., Haroon, U.T. and Niazi, I. eds. *Creating Momentum: Today is tomorrow.* Lahore, Pakistan: Sang-e-Meel Publications, pp. 55–84.

Rahman, T., 2010. *Language policy, identity and religion: Aspects of the civilisation of the muslims of Pakistan and North India.* Islamabad, Pakistan: Al-Noor publishers.

Ranjha, M.I., 2012. Urdu Syllable: Templates and Constraints. In: *The Conference on Language and Technology 2012 Proceedings*. Centre for Language Engineering. Al-Khawarizmi Institute of Computer Science. University of Engineering and Technology, Lahore, Pakistan. pp. 29-36.

Rato, A., 2014. Effects of perceptual training on the identification of English vowels by native speakers of European Portuguese. In: *The international symposium on the acquisition of second language speech, 2014 Proceedings*. *Concordia working papers in applied linguistics*. *5*, pp. 529-546.

Rauber, A.S., Escudero, P., Bion, R.A. and Baptista, B.O., 2005. The interrelation between the perception and production of English vowels by native speakers of Brazilian Portuguese. In: *6th Interspeech and 9th European Conference on Speech Communication and Technology (Eurospeech) 2005 Proceedings, 4-8 September, 2005.* Lisboa, Portugal: Curran Associates, Inc, pp. 2913-2916.

Rauf, Z., 1997. Phonology of Delhi Urdu. India: Shirpa Publications.

Raza, A., and Hussain, S., 2010. Automatic diacritization for urdu. In *Proceedings of the Conference on Language and Technology*. [Online]. pp. 105–111. Available from http://www.cle.org.pk/clt10/papers/Automatic%20Diacritization%20for%20Urdu.pdf [Accessed 2 November, 2014].

Raza, A.A., 2009. *Design and Development of an Automatic Speech Recognition System for Urdu*. FAST-National University of Computer and Emerging Sciences, Lahore Pakistan. PhD Thesis.

Raza, W., 2008. Patterns of Pakistani English pronunciation and pedagogic priorities. *Market Forces*, *4*(3), pp. 102-112.

Rehman, I., 2016. *English Language Proficiency Test (ELP-T) adapted from Cambridge Assessment English (CAE, 2016)*. [Google form]. Available from: https://docs.google.com/forms/d/1mra3q37L72kxQYYLAGq8TiFU_lzIZM6DHkZTpTxw834/edit [Accessed 12 May 2016].

Rehman, I., 2016. *Language Experience and Proficiency Questionnaire (LEAP-Q) adapted from Marian, Blumenfeld and Kaushanskaya, 2007*. [Google form]. Available from: https://docs.google.com/forms/d/1VintiTKwLQFrDk_JrMv3reKMPX4qMngujlDlCM3JL4c/edit [Accessed 12 May 2016].

Saleem, A. M., Kabir, H., Riaz, M. K., Rafique, M. M., Khalid, N., and Shahid, S. R. (2002). *Urdu Consonantal and Vocalic Sounds. CRULP Annual Student Report.* Centre for Research in Urdu Language Processing. [Online]. Available from: http://www.cle.org.pk/Publication/Crulp_report/CR02_01E.pdf [Accessed 10 October, 2014].

Sarwar, A., Ahmed, S. and Tarar, A.A., 2003. Diphthongs in Urdu language and analysis of their acoustic properties. *CRULP Annual Student Report 2002-2003,* Center for Research in Urdu Language Processing. [Online]. pp.9-15. Available from: https://pdfs.semanticscholar.org/2d9b/f9d66e3147ad6f13cadd5770f92206a17f17.pdf [Accessed 20 December, 2014].

Schmidt, A.M., 1996. Cross-language identification of consonants. Part 1. Korean perception of English. *The Journal of the Acoustical Society of America*, *99*(5), pp.3201-3211.

Schmidt, R.L., 1999. *Urdu, an Essential Grammar*. London and New York: Routledge.

Schneider, E.W. and Kortmann, B. eds., 2004. *A handbook of varieties of English: a multimedia reference tool*. Vol. 1. Berlin: Walter de Gruyter.

Scovel, T., 1969. Foreign accents, language acquisition, and cerebral dominance. *Language learning*, *19*(3-4), pp.245-253.

Scovel, T., 2000. A critical review of the critical period research. *Annual Review of Applied Linguistics*, *20*, pp.213-223.

Shackle, C. and Snell, R., 1990. *Hindi and Urdu since 1800: a common reader*. London: School of Oriental and African Studies, University of London.

Shahid, S.R., 2002. A Study on Glottal Stops in Urdu. *(CRULP)Annual Student Report 2002-2003*, Center for Research in Urdu Language Processing, [Online]. pp.19-24. Available from: http://www.cle.org.pk/Publication/Crulp_report/CR02_04E.pdf [Accessed 25, November 2014].

Shamim, F., 2008. Trends, issues and challenges in English language education in Pakistan. *Asia Pacific Journal of Education*, *28*(3), pp. 235-249.

Sheikh, Q.A., 2012. An analysis of the vowel sounds of Pakistani English. *Bulletin of Education and Research 34*(1): pp. 1-18.

Sheldon, A. and Strange, W., 1982. The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied psycholinguistics*, *3*(3), pp.243-261.

Simons, G.F., Eberhard, D.M. and Fennig, C.D., 2017. Urdu: A language of Pakistan. In: Simons, G.F., Eberhard, D.M. and Fennig, C.D. eds. *Ethnologue: Languages of the world. SIL International, 20.* [Online]. Available from: https://www.ethnologue.com/language/urd [Accessed 21 January, 2018].

Smith, L.I., 2002. *A tutorial on principal components analysis*. Department of Computer Science, University of Otago. New Zealand. Technical Report OUCS-2002-12.

Sridhar, S.N. 1992. Non-Native English Literatures: Context and Relevance. In: Kachru, B.B. ed. *The other tongue: English across cultures*. Urbana, Chicago, London: University of Illinois Press, pp. 291-306.

Stevens, K.N., Libermann, A.M., Studdert-Kennedy, M. and Öhman, S.E.G., 1969. Cross Language study of vowel perception. *Language and Speech*, *12*(1), pp.1-23.

Strange, W., 1995. Cross-language study of speech perception: A historical review. In: Strange, W. ed. *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research.* Timonium MD: York Press, pp. 3-45.

Strange, W., Akahane-Yamada, R., Fitzgerald, B.H. and Kubo, R., 1996. Perceptual assimilation of American English vowels by Japanese listeners. In: Bunnell, H.T. and Idsardi, W. eds. *4th International Conference on Spoken Language, 1996 (ICSLP 96) Proceedings, 3-6 October, 1996.* Philadelphia, PA, USA: duPont Institute. Vol. 4, pp. 2458-2461.

Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S.A. and Nishi, K., 2001. Effects of consonantal context on perceptual assimilation of American English vowels by Japanese listeners. *The Journal of the Acoustical Society of America*, *109*(4), pp.1691-1704.

Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S.A., Nishi, K. and Jenkins, J.J., 1998. Perceptual assimilation of American English vowels by Japanese listeners. *Journal of Phonetics*, *26*(4), pp.311-344.

Strange, W., Bohn, O.S., Nishi, K. and Trent, S.A., 2005. Contextual variation in the acoustic and perceptual similarity of North German and American English vowels. *The Journal of the Acoustical Society of America*, *118*(3), pp.1751-1762.

Strange, W., Bohn, O.S., Trent, S.A. and Nishi, K., 2004. Acoustic and perceptual similarity of North German and American English vowels. *The Journal of the Acoustical Society of America*, *115*(4), pp.1791-1807.

Strange, W., and Shafer, V. L. (2008). Speech perception in second language learners: The re-education of selective perception. In Hansen Edwards, J. G. and Zampini, M. L. eds.

*Phonology and second language acquisition*. Vol. 36. Philadelphia: John Benjamins Publishing Company, pp. 153–192.

Thomas, E.R. and Kendall, T., 2007. NORM: The vowel normalization and plotting suite. [Online]. Available from: http://lingtools.uoregon.edu/norm/about_normalization1.php [Accessed 10 June, 2015].

Trask, R.L., 2004. *A dictionary of phonetics and phonology*. Routledge.

Tsukada, K., Birdsong, D., Bialystok, E., Mack, M., Sung, H. and Flege, J., 2005. A developmental study of English vowel production and perception by native Korean adults and children. *Journal of Phonetics*, *33*(3), pp.263-290.)

Tyler Kendall and Erik R. Thomas (2014). vowels: Vowel Manipulation, Normalization, and Plotting. R package version 1.2-1. [Online]. Available from: https://CRAN.R-project.org/package=vowels [Accessed 20, June 2016]

Tyler, M.D., Best, C.T., Faber, A. and Levitt, A.G., 2014. Perceptual assimilation and discrimination of non-native vowel contrasts. *Phonetica*, *71*(1), pp.4-21.

Venables, W. N. and Ripley, B. D., 2002. Modern Applied Statistics with S. 4th edn. New York: Springer. ISBN 0-387-95457-0.

Vorperian, H.K. and Kent, R.D., 2007. Vowel acoustic space development in children: A synthesis of acoustic and anatomic data. *Journal of Speech, Language, and Hearing Research*, [Online], *50*(6), pp.1510-1545. Available from: doi:10.1044/1092-4388(2007/104) [Accessed 15 December, 2019].

Wali, A. and Hussain, S., 2007. Context sensitive shape-substitution in nastaliq writing system: Analysis and formulation. In: Sobh, T. ed. *Innovations and Advanced Techniques in Computer and Information Sciences and Engineering*. Dordrecht: Springer, pp. 53-58.

Wali, A., 2001. *The Rules Governing the Writing-Pronunciation Contrast in Urdu: A Phonological study*. *CRULP Annual Student Report*, *2001-2002*. Centre for Research in Urdu Language Processing. [Online]. Available from: from http://crulp.org/Publication/Crulp_report/CR02_29E.pdf [Accessed 10 October, 2014]

Wang, X. and Munro, M.J., 2004. Computer-based training for learning English vowel contrasts. *System*, *32*(4), pp.539-552.

Wang, X., 2006. Mandarin listeners' perception of English vowels: Problems and strategies. *Canadian Acoustics*. *34*(4), pp. 15-26.

Waqar, A. and Waqar, S., 2002. Identification of Diphthongs in Urdu and their Acoustic Properties. *CRULP Annual Student Report, 2002*-2003. Centre for Research in Urdu Language Processing. [Online]. pp.16-26. Available from: http://www.cle.org.pk/Publication/Crulp_report/CR03_04E.pdf [Accessed 4 March, 2015].

Warrich, A.A., Bilal, H.A., Rasheed, T., Fatima, N. and Tiwana, D.S., 2011. Acoustic Analysis of Central Vowels of PakE. *International Journal of Linguistics*, *3*(1), pp.1-13.

Watson, C.I. and Harrington, J., 1999. Acoustic evidence for dynamic formant trajectories in Australian English vowels. *The Journal of the Acoustical Society of America*, *106*(1), pp.458-468.

Wells, J.C., 1982. *Accents of English*. Vol. 1. Cambridge University Press.

West, B.T., Welch, K.B. and Galecki, A.T., 2014. *Linear mixed models: a practical guide using statistical software*. Chapman and Hall/CRC.

Wheaton College Massachusetts., 2016. [Online]. Available from: https://wheatoncollege.edu/wp-content/uploads/2012/08/How-to-Read-a-Dendrogram-Web-Ready.pdf [Accessed 2 Februry 2016].

Wickelmaier, F., 2003. An introduction to MDS. *Sound Quality Research Unit, Aalborg University, Denmark*, *46*(5), pp.1-26.

Williams, D. and Escudero, P., 2014. A cross-dialectal acoustic comparison of vowels in Northern and Southern British English. *The Journal of the acoustical society of America*, *136*(5), pp.2751-2761.

Williams, D.P., 2013. *Cross-language acoustic and perceptual similarity of vowels: The role of listeners' native accents*. Department of Germanic Studies and Department of Human Communication Sciences: University of Sheffield. PhD Thesis.

Winter, B., 2013. Linear models and linear mixed effects models in R with linguistic applications. [Online]. Mercede: University of California. Available from: https://arxiv.org/ftp/arxiv/papers/1308/1308.5499.pdf.

Wormald, J., 2015. Dynamic variation in 'Panjabi-English': Analysis of F1 & F2 trajectories for face /eɪ/ and goat /əʊ/. In: *The Scottish Consortium for ICPhS 2015, ed. The 18th International Congress of Phonetic Sciences (ICPhS) 2005, Proceedings, 10-14 August, 2015*. [Online]. Glasgow: UK: The University of Glasgow. Available from: https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0809.pdf [Accessed 20 April, 2016].

Wrembel, M., 2010. L2-accented speech in L3 production. *International Journal of Multilingualism*, *7*(1), pp.75-90.

Zahid, S., and Hussain, S., 2012. An Acoustic Study of Vowel Nasalization in Punjabi. Centre for Language and Engineering, UET Lahore. In: *The Conference on Language and Technology 2012 Proceedings*. Centre for Language Engineering. Al-Khawarizmi Institute of Computer Science. University of Engineering and Technology, Lahore, Pakistan, pp. 61-66.

Zwicker, E., 1961. Subdivision of the audible frequency range into critical bands (Frequenzgruppen). *The Journal of the Acoustical Society of America*, *33*(2), p.248.

## Appendix 2A

Blumenfeld, & Kaushanskaya (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech Language and Hearing Research, 50* (4), pp. 940-967.

# Language Experience and Proficiency Questionnaire (LEAP-Q)

(adapted by Ishrat Rehman)

| | | | | | |
|---|---|---|---|---|---|
| Last Name | | First Name | | Today's Dat | |
| Age range | **18-24** | **25-34** | **35-44** | **45-54** | **55-64** |
| Name of Institution | | Gender | Female ☐ | Male ☐ | |

**(1)** Please list all the languages you know **in order of dominance**:

| | | | | |
|---|---|---|---|---|
| **1** | **2** | **3** | **4** | **5** |

**(2)** Please list all the languages you know **in order of acquisition** (your native language first):

| | | | | |
|---|---|---|---|---|
| **1** | **2** | **3** | **4** | **5** |

**(3)** Please list what percentage of the time you are *currently* and *on average* exposed to each language.

(*Your percentages should add up to 100%*):

| | | | | | |
|---|---|---|---|---|---|
| List language here: | | | | | |
| **List percentage here:** | | | | | |

**(4)** When choosing to read a text available in all your languages, in what percentage of cases would you choose to read it in each of your languages? Assume that the original was written in another language, which is unknown to you.

*(Your percentages should add up to 100%*):

| | | | | | |
|---|---|---|---|---|---|
| List language here | | | | | |
| **List percentage here:** | | | | | |

**(5)** When choosing a language to speak with a person who is equally fluent in all your languages, what percentage of time would you choose to speak each language? Please report percent of total time.

(*Your percentages should add up to 100%*):

| | | | | | |
|---|---|---|---|---|---|
| List language here | | | | | |
| **List percentage here:** | | | | | |

**(6)** Please name the cultures with which you identify. On a scale from zero to ten, please rate the extent to which you identify with each culture. (Examples of possible cultures include Punjabi, Sindhi, Balochi, Saraiki, Pathn, Muslim, Christian, Hindu, Sikh, etc):

**Culture: ------------------------------------**

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| No Identification | very low | | | | Moderate identification | | | | | Complete identification |

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| No Identification | very low | | | | Moderate identification | | | | | Complete identification |

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| No Identification | very low | | | | Moderate identification | | | | | Complete identification |

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| No Identification | very low | | | | Moderate identification | | | | | Complete identification |

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| No Identification | very low | | | | Moderate identification | | | | | Complete identification |

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| No Identification | very low | | | | Moderate identification | | | | | Complete identification |

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| No Identification | very low | | | | Moderate identification | | | | | Complete identification |

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| No Identification | very low | | | | Moderate identification | | | | | Complete identification |

**(7)** How many years of formal education do you have? _______ ___________________________________

Please check your highest education level:

☐ Less than High School ☐ Some College ☐ Masters

☐ High School ☐ College ☐ Ph.D./M.D./J.D.

☐ Professional Training ☐ Some Graduate School ☐ Other:

**(8)** Travel abroad, if applicable ___ ____________________________________________

If you have ever travelled to another country, please provide name of the country and date of travel and reasons (e.g. holiday, education etc) here.

___________________ ___________________________________________________________________

**(9)** Have you ever had a vision problem☐, hearing impairment☐, language disability☐, or learning disability ☐ ? (Check all applicable). If yes, please explain (including any corrections):

______________________________________ _______________________________________________

**Language: Urdu, Punjabi, Saraiki, Sindhi, Pashto, Balochi, English**

**This is my (native/first second third fourth fifth) language.**

**All questions below refer to your knowledge of Error! Reference source not found..**

**(1)** Age when you…:

| *began acquiring* **Error! Reference source not found.** : | *became fluent* in **Error! Reference source not found.** : | *began reading* in **Error! Reference source not found.**: | *became fluent reading* in **Error! Reference source not found.** : |
|---|---|---|---|
| | | | |

**(2)** Please list the number of years and months you spent in each language environment:

| | Years | Months |
|---|---|---|
| A country where **Error! Reference source not found.** is spoken | | |
| A family where **Error! Reference source not found.** is spoken | | |
| A school and/or working environment where **Error! Reference so found.** is spoken | | |

**(3)** On a scale from zero to ten, please select your *level of* ***proficiency*** in speaking, understanding, and reading **Error! Reference source not found.** from the scroll-down menus:

***Speaking***

0 None, 1 Very low, 2 Low, 3 Fair, 4 Slightly less than adequate, 5 Adequate, 6 Slightly more than adequate, 7 Good, 8 Very good, 9 Excellent, 10 Perfect

***Understanding spoken language***

0 1 2 3 4 5 6 7 8 9 10

None Very low Low Fair Slightly less Adequate Slightly more Good Very good Excellent Perfect

than adequate than adequate

***Reading***

---

0 1 2 3 4 5 6 7 8 9 10

None Very low Low Fair Slightly less Adequate Slightly more Good Very good Excellent Perfect

than adequate than adequate

**(4)** On a scale from zero to ten, please select how much the following factors contributed to you learning **Error! Reference source not found.**:

***Interacting with friends:***

---

0 1 2 3 4 5 6 7 8 9 10

Not a Minimal Moderate Most important

contributor contributor contributor

contributor

***Interacting with family:***

---

0 1 2 3 4 5 6 7 8 9 10

Not a Minimal Moderate Most important

contributor contributor contributor

contributor

***Reading:***

---

0 1 2 3 4 5 6 7 8 9 10

Not a Minimal Moderate Most important

contributor contributor contributor

contributor

***Language tapes/self-instruction:***

---

0 1 2 3 4 5 6 7 8 9 10

Not a Minimal Moderate Most important

contributor contributor contributor

contributor

***TV:***

---

0 1 2 3 4 5 6 7 8 9 10

Not a Minimal Moderate Most important

contributor contributor contributor

contributor

***Listening to the radio:***

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Not a important contributor | Minimal contributor | | | | Moderate contributor | | | | | Most contributor |

**(5)** Please encircle to what extent you are currently exposed to **Error! Reference source not found.** in the following contexts:

***Interacting with friends:***

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Never | Almost never | | | | Half of the time | | | | | Always |

***Interacting with family:***

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Never | Almost never | | | | Half of the time | | | | | Always |

***Watching TV:***

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Never | Almost never | | | | Half of the time | | | | | Always |

***Listening to radio/music:***

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Never | Almost never | | | | Half of the time | | | | | Always |

***Reading:***

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Never | Almost never | | | | Half of the time | | | | | Always |

***Language lab/self-instruction:***

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Never | Almost never | | | | Half of the time | | | | | Always |

**(6)** In your perception, how much of a foreign accent do you have in this language ( **Error! Reference source not found.** )?

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| None | Almost none | very light | Light | Some | Moderate | | Considerable | Heavy | very heavy | Extremely Pervasive |

**(7)** Please rate how frequently others identify you as a non-native speaker based on your <u>accent</u> in this language **(**Error! Reference source not found.:

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Never | Almost never | | | | Half of the time | | | | | Always |

## Social Background:

**Q1.** What town you were raised in?

---

**Q.2.** On a scale from zero to ten where would you place this town/city?

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Rural/Village | | | | Quite far from the nearest city | | | | Quite close to the nearest city | | Urban/City |

**Q.3 (a)** Please put a large **"X"** on the rung where you think you stand at this time in your life, relative to other people in your country.

**Q.3 (b)** Please put a large **"X"** on the rung where you think you stand at this time in your life, relative to other people in your community/village/town/city.

## Appendix 2B

### Table 2B (i) Oral monophthongs in standard carrier phrases and full sentences

| Phonetic transcription | *Gloss* | Urdu sentence<br>Standard carrier phrases (CP)<br>Full sentence (FS) |
|---|---|---|
| /mɛ̃ ɪsɛ bi:t̪ ek bɑr kəhõ gɪ/<br>/kəi gʰent̪e bi:t̪ gəe/ | *"I will say ----- once"*<br>*"many hours have passed"* | میں اسے بِیت ایک بار کہوں گی<br>کئی گھنٹے بِیت گئے |
| /mɛ̃ ɪsɛ bɪd̪ ek bɑr kəhõ gɪ/<br>/je t̪ɔ bɪd̪ he/ | *"this is strange"* | میں اسے بِد ایک بار کہوں گی<br>یہ تو بِد ہے |
| /mɛ̃ ɪsɛ bɪk ek bɑr kəhõ gɪ/<br>/jɛ zəmin nə̃hĩ bɪk səkt̪ɪ/ | *"this land cannot be sold"* | میں اسے بک ایک بار کہوں گی<br>یہ زمین نہیں بک سکتی |
| /mɛ̃ ɪsɛ peʈ ek bɑr kəhõ gɪ/<br>/mɛre peʈ mẽ dərd hɛ/ | *"I have stomach-ache"* | میں اسے پیٹ ایک بار کہوں گی<br>میرے پیٹ میں درد ہے |
| /mɛ̃ ɪsɛ bɛ:d̪ ek bɑr kəhõ gɪ/<br>/jɛ bɛ:d̪ kɑ d̪rʌxt̪ hɛ/ | *"this is a willow tree"* | میں اسے بید ایک بار کہوں گی<br>یہ بید کا درخت ہے |
| /mɛ̃ ɪsɛ bəd̪ or bʌd̪ ek bɑr kəhõ gɪ/<br>/je t̪ɔ bʌd̪ kɑrɪ hɛ/ | *"this is bad"* | میں اسے بد ایک بار کہوں گی<br>یہ تو بد کاری ہے |
| /mɛ̃ ɪsɛ bæ:t̪ ek bɑr kəhõ gɪ/<br>/hər ʃəxs ne ɑp ki bæ:t̪ kɪ/ | *"everyone followed him"* | میں اسے بیعت ایک بار کہوں گی<br>ہر شخص نے آپ کی بیعت کی |
| /mɛ̃ ɪsɛ bɑ:d̪ ek bɑr kəhõ gɪ/<br>/ɑp ke bɑ:d̪ ʊs kɑ nəmbər he/ | *"his turn is after yours"* | میں اسے بعد ایک بار کہوں گی<br>آپ کے بعد اس کا نمبر ہے |
| /mɛ̃ ɪsɛ pɔ:d̪ ek bɑr kəhõ gɪ/<br>/jɛ ʊn kə pɔ:d̪ hɛ/ | *"he is their descendant"* | میں اسے پود ایک بار کہوں گی<br>یہ ان کا پود ہے |
| /mɛ̃ ɪsɛ pu:t̪ ek bɑr kəhõ gɪ/<br>/jɛ mɛrə pu:t̪ hɛ/ | *"he is my son"* | میں اسے پوت ایک بار کہوں گی<br>یہ میرا پوت ہے |
| /mɛ̃ ɪsɛ bʊd̪ʰ ek bɑr kəhõ gɪ/<br>/ɑp kɔ bʊd̪ʰ kɔ dʒɑnɑ tʃahɪje/ | *"you should go on Wednesday"* | میں اسے بدھ ایک بار کہوں گی<br>آپ کو بدھ کو جانا چاہیے |
| /mɛ̃ ɪsɛ bu:d̪ʰ ek bɑr kəhõ gɪ/<br>/vɔ badɪ bu:d̪ʰ vɑlɑ hɛ/ | *"he is a wise man"* | میں اسے بودھ ایک بار کہوں گی<br>وہ بڑی بودھ والا ہے |
| /mɛ̃ ɪsɛ ku:d̪ ek bɑr kəhõ gɪ/<br>/vɔ divɑr sɛ ku:d̪ geə/ | *"he jumped over the wall"* | میں اسے کود ایک بار کہوں گی<br>وہ دیوار سے کود گیا |
| /mɛ̃ ɪsɛ su:t̪ ek bɑr kəhõ gɪ/<br>/jɛ tʃɑrpɑɪ su:t̪ ki hɛ/ | *"this cot is made with yarn"* | میں اسے سوت ایک بار کہوں گی<br>یہ چارپائی سوت کی ہے |

**Table 2B (ii) Nasal vowels in standard carrier phrases and sentences**

| Phonetic transcription | Gloss | Urdu sentence<br>Standard carrier phrases (CP)<br>Full sentence (FS) |
|---|---|---|
| /mɛ̃  ɪsɛ bĩt̪ ɛk bɑr kəhõ gɪ/ | | میں اسے بنت ایک بار کہوں گی |
| /je merɪ bĩt̪ hɛ/ | *“she is my daughter”* | یہ میری بنت ہے |
| /mɛ̃  ɪsɛ pẽt̪ʰ ɛk bɑr kəhõ gɪ/ | | میں اسے پنتھ ایک بار کہوں گی |
| /ek pẽt̪ʰ d̪ɔ kɑdʒ/ | *“kill two birds with one stone”* | ایک پنتھ دو کاج |
| /mɛ̃  ɪsɛ bɔ̃d̪ ɛk bɑr kəhõ gɪ/ | | میں اسے بند ایک بار کہوں گی |
| /d̪ərɪɑ ke kɪnɑre bɔ̃d̪ bɑ̃d̪ dɪə geɑ/ | *“a wall was built on the river bank”* | دریا کے کنارے بند باندھ دیا گیا |
| /mɛ̃  ɪsɛ pɔ̃k ɛk bɑr kəhõ gɪ/ | | میں اسے پنکھ ایک بار کہوں گی |
| | *“this is a feather”* | یہ ایک پنکھ ہے |
| /mɛ̃  ɪsɛ sɔ̃d̪ ɛk bɑr kəhõ gɪ/ | | میں اسے سند ایک بار کہوں گی |
| | *“this is my certificate”* | یہ میری سند ہے |
| /mɛ̃  ɪsɛ bæ̃ːd ɛk bɑr kəhõ gɪ/ | | میں اسے بینڈ ایک بار کہوں گی |
| /jɛ ɪskə bæ̃ːd hɛ/ | *“this is his group”* | یہ اس کا بینڈ ہے |
| /mɛ̃  ɪsɛ bɑ̃ːd̪ ɛk bɑr kəhõ gɪ/ | | میں اسے باندھ ایک بار کہوں گی |
| /d̪ərɪɑ ke kɪnɑre bɔ̃d̪ bɑ̃ːd̪ dɪə geɑ/ | *“a wall was built on the river bank”* | دریا کے کنارے بند باندھ دیا گیا |

## Appendix 2C

## Pronunciation mistakes (errors) in production recordings

FS1_R3
LS18_ bu:nd sounds like nasal
FS1_R5
LS12_ bu:nd sounds like a nasal
FS2_R1
LS24_bi:t is pronounced as bæt
FS2_R2
LS5 and NS5_pɔ:d is pronounced as pu:d
LS25_bi:t is pronounced as bæt
*Note*: a little bit noise in the background, rest was all good, no mistakes except for the ones mentioned above.
FS3_R1
LS3 and NS3_pu:t is pronounced as pɔ:t
LS15_ su:t is pronounced as sɔ:t
LS17 and NS17_ bɪd is pronounced as b^d
LS21 and NS21_ bu:dh is pronounced as bɔ:dh
FS3_R2
LS5 and NS5_ pɔ:d is pronounced as pu:d
FS3_R3
LS22 and NS22_ pɔ:d is pronounced as pu:d
LS28_ geə does not follow /k/ rather follows /ɛ/
FS3_R4
LS18 and NS18_ pɔ:d is pronounced as pu:d
FS3_R5
LS24_ pu:t is pronounced as pɔ:t
FS4_R1
LS3 and NS3_ pu:t is pronounced as pɔ:t
LS7_ geə is pronounced as kɪə
LS17 and NS17_bɪd is pronounced as b^d
LS25_ bɪk is pronounced as bʊk
FS4_R2
LS25_bi:t is pronounced as bæt
LS19 and NS19_ bɪd is pronounced as b^d
FS4_R3
LS10_ pu:t is pronounced as pɔ:t
LS30 and NS30_ bɪd is pronounced as bɪd
FS4_R4
FS4_R5
LS21 and NS21_ bɪd is pronounced as b^d
FS5_R1
LS17 and NS17 bɪd is pronounced as b^d
FS5_R2
LS5 andNS5_pɔ:d is pronounced as pu:d
LS19 and NS19_ bɪd is pronounced as b^d
LS25_bi:t is pronounced as bæt

FS5_R3
LS30 and NS30_ bɪd is pronounced as b^d
FS5_R4
LS18 and NS19_ pɔ:d is pronounced as pu:d
FS5_R5
LS26 and NS26_ pɔ:d is pronounced as pu:d
FS6_R1
LS3 and NS3_ pu:t is pronounced as pɔ:t
LS12_ bu:nd is pronounced as bɔ:nd
LS15 and NS15_ su:t is pronounced as sɔ:t
LS24_ bi:t IS pronounced as bæt
FS6_R2
LS2 and NS2_ su:t is pronounced as sɔ:t
LS10_ bɪk is pronounced as bʊk
LS19 and NS19_bɪd id pronounced as b^d
LS29 and NS29_ pu:t is pronounced as pɔ:t
FS6_R3
LS4_ bi:t is pronounced as bæt
LS10 and NS10_ pu:t is pronounced as pɔ:t
LS17_ bɪk is pronounced as bʊk
LS20 and NS20_ su:t is pronounced as sɔ:t
FS6_R4
LS10 and NS10_ su:t is pronounced as sɔ:t
FS6_R5
LS24 and NS24_ pu:t is pronounced as pɔ:t
LS27 and NS27_ su:t is pronounced as sɔ:t

FS7_R1
LS3_ pu:t is pronounced as pɔ:t

LS15 and NS15_ su:t is pronounced as sɔ:t
LS24_bi:t is pronounced as bæt
LS25_bɪk is pronounced as bʊk

FS7_R2
LS2 and NS2_ su:t is pronounced as sɔ:t
LS3_ ku:d is pronounced as kɔ:d
LS10_ bɪk is pronounced as bʊk
LS19 and NS19_bɪd is pronounced as b^d
LS25_ bi:t is pronounced as bæt

FS7_R3
LS4_ bi:t is pronounced as bæt
LS17_ bɪk is pronounced as bʊk
LS20_ su:t is pronounced as sɔ:t

FS7_R4
LS10_ su:t is pronounced as sɔ:t
LS15_ bi:t is pronounced as bæt
LS20_pɛt is pronounced as pi:t

FS7_R5
LS24 and NS24_ pu:t is pronounced as pɔ:t
LS26_pɔ:d is pronounced as bu:dh
LS27_ su:t is pronounced as sɔ:t
LS30_ bi:t is pronounced as bæt

FS8_bu:nd sounds like a nasal vowel.

FS9 R1_ Incomplete and most sounds are incorrect

FS9_R2_ Bu:nd sounds nasal here too.
Mispronounced bu:dh (As bʊd) and pu:t (ɑs pɔ:t) and bɪd (as b^d)

FS9_R3_ bu:nd sounded like nasal in natural sentence and like non nasal in CP
Mispronounced pu:t (as pɔ:t)
Pronounced bu:dh as bu:nd and bʊd
LS28_ geə followed ejk instead of kahun…

FS9_R4
Pɪə followed ejk instead of kahun…
LS8_ geə followed ejk instead of kahun…
LS11 and 12_pu:t is pronounced as pɔ:t
NS23_ bu:nd sounds like nasal vowel

FS9_R5
LS30_ pronounced as bæt instead of bi:t
LS12 and NS12_ bu:nd, sounds like a nasal vowel

FS10_ R1
LS3 is missing- and NS3 was pronounced as pɔ:t instead of pu:t
Baby noise in the background
LS12 and NS12 sounds like a nasal bu:nd
LS17 and NS17 bɪd was pronounced as b^d
LS21 and NS21_ bu:dh heavy noise in the background
NS22_ heavy noise in the background
LS24_ Bi:t was pronounced as bæt, heavy noise in the background of NS24

FS10_R2
LS19 and NS19_ bɪd was pronounced as b^d
NS20 sounds like nasal bɑ:ndh· however there is a lot of baby noise in the background
LS29 and NS29_ pu:t was pronounced as pɔ:t

FS10_R3
LS10 and NS10 pu:t is pronounced as pɔ:t

FS10_R4

LS5 and NS5_ heavy noise in the background
NS8_ noise in the background
LS11 and NS11_ pu:t is pronounced as pɔ:t
NS13_ heavy noise in the background

FS10_R5
LS24 and NS24 pu:t is pronounced as pɔ:t

FS11_R1
NS20_ pənkh sounds like nasal
LS30 and NS30_ pɔ:d is pronounced as pu:d (only mistake in whole recording)

FS11_R2
LS5 and NS5_ pɔ:d is pronounced as pu:d

FS11_R3
LS22 and NS22_ p:d is pronounced as pu:d
LS27 and NS27_bæd is pronounced as bɛd

FS11_R4
LS18 and NS18_ pɔ:d is pronounced as pu:d
LS25 and NS25_bæd is pronounced as bɛd

FS11_R5
LS20 and NS20_bæd is pronounced as bɛd
LS26 and NS26_ p:d is pronounced as pu:d

MS1_R1
LS2 and NS2_bæt is pronounced as bəɪt

MS1_R2
LS18 and NS18_ bæt is pronounced as bəɪt

MS1_R3

LS29 and NS29_ bæt is pronounced as bəɪt
MS1_R4
LS7 and NS7_ bæt is pronounced as bəɪt
MS1_R5
LS28 and NS28_ bæt is pronounced as bəɪt

MS2_R1
LS15 and NS15_ su:t is pronounced as sɔ:t
LS16_p^nth is pronounced as pɛnth
Note: bɪk and pɔ:d are missing (there are 28 sentences in this set of recording)

MS2_R2
LS4_ p^nth is pronounced as pɛnth
LS19_ geə does not follow consonant /k/ rather it follows vowel /ɛ/
LS23_bi:t is pronounced as bæt
MS2_R3
LS4_ bi:t is pronounced as bæt
LS18_ p^nth is pronounced as pɛnth
LS25_bæd is pronounced as bi:d
LS26_ geə does not follow consonant /k/ rather it follows vowel /ɛ/

MS2_R4
LS7_ geə does not follow consonant /k/ rather it follows vowel /ɛ/
MS2_R5
LS2_ geə does not follow consonant /k/ rather it follows vowel /ɛ/
LS19 and NS19_bæd is pronounced as bɛd

MS3_R1
LS3 and NS3_pu:t is pronounced as pɔ:t
LS6_pɛt is pronounced as pi:t
LS8_NS8_bæd is pronounced as bɛd
LS15 and NS15_su:t is pronounced as sɔ:t
LS16 and NS16_ p^nth is pronounced as pɪnth
LS24_bi:t is pronounced as bɛd
Note: bɪk is missing in this recording (there are 29 SENTENCES)

MS3_R2
LS2 and NS2_ su:t is pronounced as sɔ:t
LS4 and NS4_ p^nth is pronounced as pɪnth
LS18 and NS18_bɪd is pronounced as b^d
LS24_ bi:t is pronounced as bæt
LS28 and NS28_ pu:t is pronounced as pɔ:t
Note: bɪk is missing in this recording (there are 29 SENTENCES)

MS3_R3
LS4_ bi:t is pronounced as bæt
LS10_ pu:t is pronounced as pɔ:t
LS18 and NS18_ p^nth is pronounced as pɪnth
LS21 and NS21_ pɔ:d is pronounced as pu:d
LS26 and NS26_bæd is pronounced as bɛd
LS27_ geə does not follow consonant /k/ rather it follows vowel /ɛ/
LS29 and NS29_ bɪd is pronounced as b^d
Note: bɪk is missing in this recording (there are 29 SENTENCES)

MS3_R4
LS1 and NS1_ p^nth is pronounced as pɪnth
LS7_ geə does not follow consonant /k/ rather it follows vowel /ɛ/
LS14_ bi:t is pronounced as bæt
LS26 and NS26_ bɪd is pronounced as b^d
Note: bɪk is missing in this recording (there are 29 SENTENCES)

MS3_R5
LS2_ geə does not follow consonant /k/ rather it follows vowel /ɛ/
LS15_bɪnt is pronounced as bənt
LS23 and NS23_ pu:t is pronounced as pɔ:t
Note: bɪk is missing in this recording (there are 29 SENTENCES)

MS4_R1
LS25_ bi:t is pronounced as bæt
LS26_bɪk is pronounced as bək
MS4_R2
LS19 and NS19_bɪd is pronounced as b^d
LS25_ bi:t is pronounced as bæt
MS4_R3
LS4_ bi:t is pronounced as bæt
MS4_R4
LS24_ b^d is pronounced as bɪd
MS4_R5
LS30_ bi:t is pronounced as bæt

MS5_R1
LS2 and NS2_bæt is pronounced as bæit
LS3_ pu:t is pronounced as pɔ:t
LS7_ noise in the background (at the onset)
LS9_bɔɪ is pronouncrd as bʊe
LS21_ /b/ in bu:dh is aspirated
LS23_pɑʊ is pronounced as pɑʊn
LS25_bɪk is pronounced as bʊk
MS5_R2

LS4_p^nth is pronounced as pɛnth
LS11_ bɔɪ is pronounced as bʊe
LS15_b^d is pronounced as bɪd
LS18 and NS18_ bæt is pronounced as bəɪt
LS24_ bæd is pronounced as bəɪd
Note: fan noise in the background throughout this recording
MS5_R3
LS4_ bi:t is pronounced as bæt
LS6_ bɔɪ is pronounced as bʊe
LS28_ geə does not follow consonant /k/ rather it follows vowel /ɛ/ and noise in the background (mobile ringtone)
MS5_R4
LS7 and NS7_ bæt is pronounced as bəɪt
LS11 and NS11_pu:t is pronounced as pɔ:t
LS15 and NS15_ bi:t is pronounced as bæt
LS26_ bɔɪ is pronounced as bʊe
LS28 and NS28_bu:dh is pronounced with aspirated /b/
MS5_R5
LS5_ bɔɪ is pronounced as bʊe
LS16_ bɪnt is pronounced as bənt
LS18_ b^d is pronounced as bɪd
LS24 and NS23_pu:t is pronounced as pɔ:t
Note: this fifth recording was done extremely fast.

MS6_R1
LS17_ bɪd is pronounced as b^d
LS21 and NS21_ bu:dh is pronounced as bʊd
LS24_ bi:t is pronounced as bæt
LS30_pɔ:d is pronounced as pu:d
MS6_R2
LS5 and NS5_ pɔ:d is pronounced as pu:d
LS21_geə does not follow consonant /k/ rather it follows vowel /ɛ/
LS25_ bi:t is pronounced as bæt
MS6_R3
LS22_ pɔ:d is pronounced as pu:d
LS28_ geə does not follow consonant /k/ rather it follows vowel /ɛ/
MS6_R4
LS8_ geə does not follow consonant /k/ rather it follows vowel /ɛ/
MS6_R5
LS2_ geə does not follow consonant /k/ rather it follows vowel /ɛ/
LS18_b^d is pronounced as bɪd

MS7_R1
LS17 and NS17_bɪd is pronounced as b^d
LS21 _ bu:dh is pronounced in two syllables as bɔ:dh and bɔ:dha
LS22_ku:d is pronounced as kɔ:d
MS7_R2
LS19 and NS19_ bɪd is pronounced as b^d
LS21 AND NS21 noise in the background
MS7_R3
NS24_ heavy noise in the background
MS7_R4
NS30_ heavy noise in the background
MS7_R5
NS2_ not clear as very heavy noise in the background
LS26 and NS26_ pɔ:d is pronounced as pu:d

MS8_R1
LS4_ bint sounds like nasal
LS7_ pronounced geə as keə
LS9_ bɔɪ is pronounced as bʊɛ
LS10_ b^nd sounds like a nasal but not in NS10
LS12 and NS12_ sounds like nasal bu:nd
LS19 sounds like nasal bænd
LS30 and NS30_ pɔ:d is pronounced as pu:d
MS8_R2
LS16_ b^d is pronounced as bɪd
LS21_ geə is pronounced as keə
MS8_R3
LS4_ bi:t is pronounced as bæt
LS17_ bɪk is pronounced as bək
NS22_ pɔ:d is pronounced as pu:d
MS8_R4
LS2_Sɑ:nd sounds like nasal
LS5_ bɪk is pronounced as bək
LS18 and LS19_ pɔ:d is pronounced as pu:d
LS23_ bu:nd sounds like nasal
MS8_R5
LS13_ bɪk is pronounced as bək
LS26 and NS26_ pɔ:d is pronounced as pu:d

MS9_R1
LS16_ p^nth is pronounced as p^n^th
LS17 and NS17_ bɪd is pronounced as b^d
LS24_ bi:t is pronounced as bæt
LS25_ bɪk is pronounced as bək
MS9_R2
LS6 and NS6_ bu:nd sounds like nasal

MS9_R3

NS1_ pənkh sounds nasal
LS3_ b^d is pronounced as bɪd
LS4_ bi:t is pronounced as bæt
LS8_b^nd sounds nasal
MS9_R4
All nasal vowels sound quite nasal in this recording--- No mispronunciation

MS9_R5

LS9 and LS9-_ sa:nd sounds like nasal
LS18_ b^d is pronounced as bɪd
NS12_ bu:nd sounds nasal

MS10_ R1

NS16_p^nth sounds like nasal
NS26_Sɑːnd sounds like nasal

MS10_R2

Clicking noise in the background throughout the recording, however there are no mispronunciations.
NS4_ p^nth sounds like nasal
LS6 and NS6_ bu:nd sounds like nasal
LS8_ sɑːnd sounds like nasal
LS28 and NS28_ bænd sounds like nasal

MS10_R3

LS10_ pu:t is pronounced as pɔːt
LS9 and NS9_sɑːnd sounds like nasal
NS15_ bænd sounds like nasal
NS19_p^nth sounds like nasal

MS10_R4

LS14 and NS14_bænd sounds like nasal
LS23_ bu:nd sounds like nasal

MS10_R5 (this fifth recording is done in extreme rush) - speaker spoke very fast))

LS2_geə follow a vowel sound in CP instead of a consonant /k
NS5_ bɔɪ is not fully pronounced.
NS9_ sɑːnd sounds like nasal

MS11_R1

LS3 and NS3_puːt is pronounced as pɔːt

LS25_ bɪk is pronounced as bək

LS12 nas NS12_ bu:nd sounds like nasal

MS11_R2

NS6_ bu:nd sounds like nasal
LS15_b^d is pronounced as bɪd
LS23_ bu:dh sounds like bo:dh

MS11_R3

LS4_ bi:t is pronounced as bæt
LS10 and NS10_ pu:t is pronounced as pɔːt
LS22 and NS22_pɔːd is pronounced as puːd

MS11_R4

LS15_ bi:t is pronounced as bæt
LS24_b^d is pronounced as bɪd
LS23 and NS23_ bu:nd sounds like nasal

MS11_R5

LS18_b^d is pronounced as bɪd
LS26 and NS26_pɔːd is pronounced as puːd
LS12 and NS12_bu:nd sounds like nasal

MS12_R1

LS3 and NS3_puːt is pronounced as pɔːt
LS16 and NS16_p^nth is pronounced as pɛnthd
LS17 and NS17_bɪd is pronounced as b^d
LS1_sənəd is pronounced as sɪn

## Appendix 2D

**Table 2D (i): Total number of tokens per vowel by individual male (M) and female (F) speakers in CP and FS**

| Vowel | iː | ɪ | eː | ɛ | æ | ɑː | ɔː | o | ʊ | uː | ə | ʌ | ɑɪ | ɑʊ | ɔɪ | ɪə | eə | ʊə |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Tokens | 182 | 366 | 218 | 213 | 239 | 217 | 167 | 84 | 222 | 777 | 419 | 247 | 220 | 218 | 212 | 220 | 217 | 220 |
| F | 93 | 183 | 109 | 106 | 128 | 108 | 85 | 61 | 113 | 364 | 211 | 134 | 110 | 109 | 108 | 110 | 109 | 110 |
| M | 89 | 183 | 109 | 107 | 111 | 109 | 82 | 23 | 109 | 413 | 208 | 113 | 110 | 109 | 104 | 110 | 108 | 110 |
| CP | 80 | 180 | 108 | 103 | 136 | 110 | 83 | 50 | 109 | 379 | 211 | 121 | 111 | 108 | 103 | 109 | 107 | 110 |
| FS | 102 | 186 | 110 | 110 | 103 | 107 | 84 | 34 | 113 | 398 | 208 | 126 | 109 | 110 | 109 | 111 | 110 | 110 |
| FS1 | 10 | 20 | 10 | 11 | 9 | 10 | 10 | NA | 10 | 40 | 20 | 10 | 10 | 10 | 10 | 10 | 10 | 10 |
| FS2 | 8 | 20 | 10 | 8 | 14 | 10 | 8 | NA | 10 | 40 | 20 | 10 | 10 | 10 | 10 | 10 | 10 | 10 |
| FS3 | 9 | 18 | 10 | 10 | 11 | 10 | 4 | 4 | 10 | 34 | 21 | 12 | 10 | 10 | 10 | 10 | 10 | 10 |
| FS4 | 8 | 10 | 10 | 10 | 12 | 10 | 10 | 3 | 10 | 35 | 20 | 14 | 10 | 10 | 10 | 10 | 9 | 10 |
| FS5 | 9 | 14 | 10 | 10 | 11 | 10 | 4 | NA | 10 | 39 | 20 | 17 | 10 | 10 | 10 | 10 | 10 | 10 |
| FS6 | 8 | 16 | 10 | 10 | 12 | 10 | 10 | 18 | 10 | 22 | 26 | 12 | 10 | 10 | 10 | 10 | 10 | 10 |
| FS7 | 5 | 15 | 9 | 10 | 15 | 10 | 9 | 10 | 10 | 29 | 12 | 12 | 10 | 10 | 10 | 10 | 10 | 10 |

| | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| FS8 | 10 | 18 | 10 | 10 | 10 | 10 | 10 | 9 | 10 | 30 | 20 | 10 | 10 | 10 | 10 | 10 | 10 | 10 |
| FS9 | 7 | 16 | 10 | 10 | 13 | 8 | 10 | 8 | 13 | 26 | 18 | 13 | 10 | 9 | 8 | 10 | 10 | 10 |
| FS10 | 9 | 16 | 10 | 10 | 11 | 10 | 10 | 9 | 10 | 29 | 24 | 14 | 10 | 10 | 10 | 10 | 10 | 10 |
| FS11 | 10 | 20 | 10 | 7 | 10 | 10 | NA | NA | 10 | 40 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 |
| MS1 | 10 | 20 | 10 | 10 | NA | 10 | 10 | NA | 10 | 40 | 20 | 10 | 10 | 10 | 10 | 10 | 10 | 10 |
| MS2 | 8 | 10 | 10 | 8 | 12 | 10 | NA | 2 | 10 | 38 | 20 | 10 | 10 | 10 | 10 | 10 | 10 | 10 |
| MS3 | 4 | 4 | 9 | 8 | 14 | 10 | 8 | 11 | 10 | 29 | 15 | 15 | 10 | 10 | 10 | 10 | 10 | 10 |
| MS4 | 6 | 18 | 10 | 10 | 14 | 10 | 10 | NA | 10 | 40 | 20 | 10 | 10 | 10 | 10 | 10 | 10 | 10 |
| MS5 | 8 | 19 | 10 | 9 | 6 | 9 | 10 | 5 | 10 | 34 | 18 | 8 | 10 | 9 | 5 | 10 | 10 | 10 |
| MS6 | 8 | 20 | 10 | 11 | 11 | 10 | 6 | 1 | 10 | 37 | 20 | 10 | 10 | 10 | 10 | 10 | 10 | 10 |
| MS7 | 10 | 16 | 10 | 10 | 10 | 10 | 8 | NA | 10 | 38 | 15 | 14 | 10 | 10 | 10 | 10 | 10 | 10 |
| MS8 | 9 | 18 | 10 | 10 | 11 | 10 | 3 | NA | 10 | 40 | 20 | 9 | 10 | 10 | 9 | 10 | 8 | 10 |
| MS9 | 8 | 19 | 10 | 10 | 12 | 10 | 10 | NA | 10 | 40 | 20 | 10 | 10 | 10 | 10 | 10 | 10 | 10 |
| MS10 | 10 | 20 | 10 | 10 | 10 | 10 | 10 | 1 | 10 | 39 | 20 | 10 | 10 | 10 | 10 | 10 | 10 | 10 |
| MS11 | 8 | 19 | 10 | 11 | 11 | 10 | 7 | 3 | 9 | 38 | 20 | 7 | 10 | 10 | 10 | 10 | 10 | 10 |

# Appendix 2E

## Praat Script for Urdu monophthongs and diphthongs

```
# TextGrid tier with intervals containing whole vowels.
vowel_tier = 1
# TextGrid tier with intervals containing initial+transition+final intervals of diphthongs.
diphthong_tier = 3
# Fraction along vowel duration to start taking samples (inclusive).
sample_start = 0.2
# Fraction along vowel duration to stop taking samples (inclusive).
sample_end = 0.8
# Number of sample points to get formant values along vowel interval.
sample_count = 7
# Formant resolution
time_step = 0.01
maximum_number_of_formants = 5
window_length = 0.025
preemphasis_from = 30

# Male
root$ = "/media/August 2016-Praat data/Male"
output$ = "/media/August 2016-Praat data/results/results_m.txt"
maximum_formant = 5000
@readVowels

# Female
root$ = "/media/August 2016-Praat data/Female"
output$ = "/media/August 2016-Praat data/results/results_f.txt"
maximum_formant = 5500
@readVowels


procedure readVowels
	printline -------------------------------------------------------------------------
	Create Strings as directory list... folders 'root$'
	folder_count = Get number of strings

	fullpaths_count = 0
	for folder_num from 1 to folder_count
		select Strings folders
		folder$ = Get string... folder_num

		Create Strings as file list... files 'root$'/'folder$'/*.wav
		file_count = Get number of strings
		for file_num from 1 to file_count
			file_name$ = Get string... file_num
			fullpath$ = "'root$'/'folder$'/'file_name$'"
			fullpaths_count = fullpaths_count + 1
			fullpaths$[fullpaths_count] = fullpath$
			file_names$[fullpaths_count] = file_name$
			printline 'fullpath$'
		endfor

		select Strings files
		Remove
	endfor

	select Strings folders
	Remove
```

```
	resultline$ = "Subject	Gender	Repetition	Vowel	Duration"
	for sample_num from 1 to sample_count
		resultline$ = resultline$ + "	F1_'sample_num'	F2_'sample_num'
	F3_'sample_num'"
	endfor
	resultline$ = resultline$ + "	diph_trans_start_time	diph_trans_end_time
	diph_trans_start_F1	diph_trans_start_F2	diph_trans_start_F3
	diph_trans_end_F1	diph_trans_end_F2	diph_trans_end_F3"
	resultline$ = resultline$ + newline$
	fileappend "'output$'" 'resultline$'

	for fullpath_num from 1 to fullpaths_count
		file_name$ = file_names$[fullpath_num]
		fullpath$ = fullpaths$[fullpath_num]
		textgrid$ = left$(fullpath$, length(fullpath$)-4) + ".TextGrid"

		subject$ = left$(file_name$, length(file_name$)-7)
		repetition$ = mid$(file_name$, length(file_name$)-4, 1)
		gender$ = left$(file_name$, 1)

		printline Processing 'subject$' (gender 'gender$') repetition 'repetition$':
		printline --'fullpath$'
		printline --'textgrid$'

		if fileReadable (textgrid$)

			Read from file... 'fullpath$'
			soundname$ = selected$ ("Sound", 1)
			To Formant (burg)... time_step maximum_number_of_formants
maximum_formant window_length preemphasis_from

			Read from file... 'textgrid$'
			interval_count = Get number of intervals... vowel_tier
			for interval_num from 1 to interval_count
				vowel$ = Get label of interval... vowel_tier interval_num
				vowel$ = replace_regex$ (vowel$, "\n", "", 0)
				vowel$ = replace_regex$ (vowel$, "\r", "", 0)
				# Check if interval label is non-empty (which we take to mean it
contains a vowel).
				if vowel$ <> ""
					# Get the interval's start and end time (in milliseconds):
					start = Get starting point... vowel_tier interval_num
					end = Get end point... vowel_tier interval_num

					# Duration (in seconds) of whole vowel interval.
					duration = (end - start)
					duration_ms = duration*1000

					# Add first few columns to the row that will be written to the
output text file.
					resultline$ = "'subject$'	'gender$'	'repetition$'
	'vowel$'	'duration_ms'"

					# Duration (in seconds) from first sample to last sample.
					sampling_duration = duration * (sample_end - sample_start)
					# Distance (in seconds) between each sample.
					sampling_delta = sampling_duration / (sample_count - 1)
					# Point in time to start sampling.
					sampling_start = start + duration * sample_start

					# Add remaining columns for samples along the vowel.
```

```
for sample_num from 0 to (sample_count-1)
	sample_time = sampling_start + sampling_delta *
sample_num

	# Get the formant values at the interval
	select Formant 'soundname$'
	f1 = Get value at time... 1 sample_time Hertz
Linear
	f2 = Get value at time... 2 sample_time Hertz
Linear
	f3 = Get value at time... 3 sample_time Hertz
Linear
	resultline$ = resultline$ + "	'f1'	'f2'
	'f3'"
endfor

select TextGrid 'soundname$'

# Get interval in diphthong tier corresponding to start time of
whole vowel.
interval_before_diph_transition = Get interval at time...
diphthong_tier start
# End time of above interval is start time of diphthong
transition interval (if applicable).
diph_transition_start = Get end point... diphthong_tier
interval_before_diph_transition

if diph_transition_start > end
	# If transition start is after vowel end, then we have
jumped to a different vowel
	# (i.e. there was no diphthong transition for this
vowel), so just put some zeros.
	resultline$ = resultline$ + "	0	0
	0	0	0	0	0	0"
else
	# The diphthong transition interval is the next
interval along.
	diph_transition_interval =
interval_before_diph_transition + 1

	# Assert that the interval contains some text.
	diph$ = Get label of interval... diphthong_tier
diph_transition_interval
	diph$ = replace_regex$ (diph$, "\n", "", 0)
	diph$ = replace_regex$ (diph$, "\r", "", 0)

	if diph$ = ""
		printline ERROR - SHOULD BE
DIPHTHONG IN 'file_name$' @ 'vowel$'

	else
		# Start of transition interval.
		diph_transition_start = Get starting point...
diphthong_tier diph_transition_interval

		# End of transition interval.
		diph_transition_end = Get end point...
diphthong_tier diph_transition_interval

		# Add transition times to row.
		diph_transition_start_ms =
(diph_transition_start - start) * 1000
		diph_transition_end_ms =
(diph_transition_end - start) * 1000
```

```
resultline$ = resultline$ + "
'diph_transition_start_ms'  'diph_transition_end_ms'"

# Midpoint of diphthong's first vowel.
diph_begin_mid = start +
(diph_transition_start - start)/2
# Midpoint of diphthong's second vowel.
diph_end_mid = end - (end -
diph_transition_end)/2

select Formant 'soundname$'

# Get formants for diphthong's first vowel.
f1 = Get value at time... 1 diph_begin_mid
Hertz Linear
f2 = Get value at time... 2 diph_begin_mid
Hertz Linear
f3 = Get value at time... 3 diph_begin_mid
Hertz Linear
# Add formants for diphthong's first vowel
to row.
resultline$ = resultline$ + "           'f1'
'f2'       'f3'"

# Get formants for diphthong's second
vowel.
f1 = Get value at time... 1 diph_end_mid
Hertz Linear
f2 = Get value at time... 2 diph_end_mid
Hertz Linear
f3 = Get value at time... 3 diph_end_mid
Hertz Linear
# Add formants for diphthong's second
vowel to row.
resultline$ = resultline$ + "           'f1'
'f2'       'f3'"
endif
endif

# Save result to text file:
resultline$ = resultline$ + newline$
fileappend "'output$'" 'resultline$'
select TextGrid 'soundname$'
endif
endfor
else
printline --'textgrid$' not found
endif
endfor
endproc
```

**Appendix 2F**

**Table 2F (i): Mean formant frequencies with standard deviation for the first two formants of 12 monophthongs at the seven equidistant points in time i.e. 20% 30% … 80% in Hz.**

| | Vowel | 20% | SD | 30% | SD | 40% | SD | 50% | SD | 60% | SD | 70% | SD | 80% | SD |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | iː | 329 | 37 | 329 | 38 | 328 | 37 | 329 | 39 | 332 | 42 | 333 | 42 | 337 | 43 |
| | ɪ | 390 | 37 | 394 | 37 | 397 | 37 | 398 | 37 | 396 | 38 | 392 | 39 | 384 | 44 |
| | eː | 411 | 35 | 411 | 37 | 412 | 39 | 413 | 41 | 415 | 41 | 417 | 41 | 415 | 41 |
| | ɛ | 555 | 70 | 586 | 79 | 600 | 84 | 608 | 88 | 611 | 92 | 601 | 95 | 578 | 90 |
| | æ | 561 | 67 | 593 | 74 | 610 | 84 | 621 | 92 | 620 | 104 | 608 | 109 | 571 | 99 |
| F1 | ɑː | 652 | 96 | 678 | 110 | 694 | 112 | 702 | 110 | 709 | 111 | 703 | 109 | 663 | 99 |
| | ɔː | 486 | 66 | 489 | 70 | 492 | 75 | 496 | 80 | 500 | 78 | 505 | 78 | 503 | 77 |
| | o | 491 | 84 | 496 | 92 | 499 | 97 | 501 | 98 | 502 | 101 | 510 | 101 | 512 | 98 |
| | ʊ | 398 | 36 | 400 | 38 | 403 | 40 | 407 | 40 | 407 | 41 | 405 | 41 | 400 | 45 |
| | uː | 377 | 56 | 374 | 64 | 373 | 48 | 371 | 47 | 371 | 45 | 372 | 44 | 372 | 45 |
| | ə | 562 | 84 | 577 | 88 | 581 | 89 | 578 | 92 | 569 | 93 | 551 | 92 | 519 | 90 |
| | ʌ | 569 | 84 | 589 | 91 | 600 | 95 | 603 | 96 | 596 | 95 | 578 | 92 | 543 | 87 |
| | iː | 2372 | 312 | 2395 | 316 | 2415 | 344 | 2439 | 326 | 2450 | 341 | 2451 | 333 | 2420 | 325 |
| | ɪ | 2061 | 276 | 2104 | 288 | 2137 | 293 | 2147 | 315 | 2167 | 300 | 2169 | 309 | 2155 | 320 |
| | eː | 2151 | 307 | 2198 | 309 | 2222 | 314 | 2199 | 329 | 2199 | 315 | 2190 | 322 | 2171 | 308 |
| | ɛ | 1889 | 269 | 1908 | 247 | 1899 | 246 | 1888 | 246 | 1888 | 230 | 1882 | 224 | 1868 | 217 |
| | æ | 1856 | 288 | 1876 | 260 | 1871 | 256 | 1872 | 260 | 1879 | 239 | 1874 | 250 | 1862 | 247 |
| F2 | ɑː | 1210 | 162 | 1216 | 156 | 1214 | 158 | 1218 | 162 | 1254 | 170 | 1322 | 174 | 1419 | 180 |
| | ɔː | 861 | 124 | 855 | 128 | 862 | 138 | 893 | 185 | 953 | 222 | 1017 | 169 | 1145 | 187 |
| | o | 949 | 259 | 928 | 234 | 933 | 239 | 941 | 229 | 950 | 174 | 1025 | 184 | 1135 | 183 |
| | ʊ | 974 | 158 | 993 | 165 | 1033 | 168 | 1090 | 163 | 1151 | 164 | 1209 | 181 | 1268 | 200 |
| | uː | 938 | 279 | 924 | 283 | 918 | 289 | 918 | 274 | 956 | 253 | 1017 | 256 | 1099 | 230 |
| | ə | 1636 | 236 | 1639 | 240 | 1639 | 245 | 1633 | 253 | 1643 | 243 | 1633 | 245 | 1609 | 234 |
| | ʌ | 1310 | 171 | 1356 | 178 | 1402 | 181 | 1453 | 178 | 1507 | 175 | 1542 | 175 | 1570 | 177 |

**Table 2F (ii): Mean formant frequencies with standard deviation for the first two formants of 12 monophthongs at the seven equidistant points in time i.e. 20% 30% … 80% in Lobanov normalised**

| | Vowel | 20% | SD | 30% | SD | 40% | SD | 50% | SD | 60% | SD | 70% | SD | 80% | SD |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | iː | -1.35 | 0.30 | -1.36 | 0.30 | -1.36 | 0.29 | -1.35 | 0.30 | -1.33 | 0.32 | -1.32 | 0.33 | -1.29 | 0.35 |
| | ɪ | -0.83 | 0.30 | -0.80 | 0.30 | -0.78 | 0.30 | -0.77 | 0.30 | -0.78 | 0.30 | -0.82 | 0.31 | -0.88 | 0.34 |
| | e | -0.66 | 0.23 | -0.65 | 0.25 | -0.65 | 0.27 | -0.63 | 0.29 | -0.62 | 0.28 | -0.61 | 0.29 | -0.63 | 0.29 |
| | ɛ | 0.54 | 0.47 | 0.78 | 0.48 | 0.90 | 0.50 | 0.97 | 0.53 | 0.98 | 0.56 | 0.90 | 0.60 | 0.71 | 0.60 |
| | æ | 0.56 | 0.44 | 0.81 | 0.42 | 0.94 | 0.46 | 1.01 | 0.47 | 0.99 | 0.54 | 0.89 | 0.60 | 0.59 | 0.56 |
| F1 | ɑː | 1.31 | 0.50 | 1.52 | 0.54 | 1.64 | 0.53 | 1.71 | 0.51 | 1.77 | 0.50 | 1.72 | 0.49 | 1.40 | 0.48 |
| | ɔː | -0.02 | 0.48 | 0.01 | 0.52 | 0.03 | 0.55 | 0.06 | 0.58 | 0.09 | 0.57 | 0.13 | 0.54 | 0.10 | 0.51 |
| | o | -0.18 | 0.46 | -0.15 | 0.54 | -0.13 | 0.55 | -0.12 | 0.56 | -0.11 | 0.58 | -0.05 | 0.60 | -0.03 | 0.57 |
| | ʊ | -0.76 | 0.26 | -0.74 | 0.26 | -0.72 | 0.26 | -0.69 | 0.25 | -0.69 | 0.25 | -0.71 | 0.25 | -0.76 | 0.27 |
| | uː | -0.95 | 0.44 | -0.97 | 0.51 | -0.99 | 0.35 | -1.00 | 0.33 | -1.00 | 0.31 | -0.99 | 0.29 | -0.99 | 0.30 |
| | ə | 0.58 | 0.54 | 0.70 | 0.55 | 0.74 | 0.56 | 0.71 | 0.59 | 0.63 | 0.60 | 0.48 | 0.61 | 0.22 | 0.62 |
| | ʌ | 0.62 | 0.49 | 0.77 | 0.53 | 0.86 | 0.56 | 0.89 | 0.55 | 0.83 | 0.54 | 0.68 | 0.52 | 0.39 | 0.47 |
| | | | | | | | | | | | | | | | |
| | iː | 1.64 | 0.40 | 1.69 | 0.47 | 1.74 | 0.53 | 1.79 | 0.49 | 1.80 | 0.51 | 1.80 | 0.47 | 1.74 | 0.43 |
| | ɪ | 1.02 | 0.31 | 1.10 | 0.33 | 1.17 | 0.34 | 1.19 | 0.39 | 1.23 | 0.34 | 1.23 | 0.36 | 1.20 | 0.40 |
| | e | 1.20 | 0.33 | 1.30 | 0.34 | 1.34 | 0.35 | 1.30 | 0.40 | 1.30 | 0.38 | 1.28 | 0.39 | 1.25 | 0.37 |
| | ɛ | 0.69 | 0.30 | 0.73 | 0.24 | 0.72 | 0.27 | 0.71 | 0.33 | 0.71 | 0.31 | 0.70 | 0.30 | 0.67 | 0.30 |
| | æ | 0.63 | 0.31 | 0.67 | 0.25 | 0.67 | 0.32 | 0.68 | 0.35 | 0.69 | 0.31 | 0.69 | 0.36 | 0.67 | 0.40 |
| F2 | ɑː | -0.61 | 0.20 | -0.60 | 0.20 | -0.60 | 0.22 | -0.59 | 0.24 | -0.52 | 0.24 | -0.39 | 0.23 | -0.20 | 0.22 |
| | ɔː | -1.27 | 0.18 | -1.29 | 0.18 | -1.27 | 0.19 | -1.21 | 0.29 | -1.10 | 0.38 | -0.98 | 0.22 | -0.74 | 0.24 |
| | o | -1.17 | 0.44 | -1.20 | 0.39 | -1.19 | 0.41 | -1.18 | 0.39 | -1.17 | 0.25 | -1.03 | 0.27 | -0.83 | 0.26 |
| | ʊ | -1.06 | 0.21 | -1.03 | 0.21 | -0.95 | 0.21 | -0.84 | 0.20 | -0.72 | 0.20 | -0.61 | 0.24 | -0.50 | 0.28 |
| | uː | -1.12 | 0.50 | -1.15 | 0.51 | -1.16 | 0.51 | -1.16 | 0.47 | -1.09 | 0.44 | -0.96 | 0.44 | -0.81 | 0.36 |
| | ə | 0.20 | 0.27 | 0.20 | 0.29 | 0.21 | 0.31 | 0.20 | 0.33 | 0.22 | 0.33 | 0.20 | 0.35 | 0.15 | 0.29 |
| | ʌ | -0.43 | 0.21 | -0.34 | 0.22 | -0.25 | 0.22 | -0.16 | 0.21 | -0.05 | 0.18 | 0.02 | 0.17 | 0.07 | 0.17 |

**Appendix 2G**

**Table 2G (i): Mean Frequencies of F1, F2 and F3 in hertz and standard deviation of 12 oral monophthongs at the vowel midpoint for female and male speakers in Carrier Phrases (CP)**

| | CP- Female speakers | | | | | | | | CP – Male speakers | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Vowel | F1 | SD | F2 | SD | F3 | SD | Duration | SD | F1 | SD | F2 | SD | F3 | SD | Duration | SD |
| iː | 347 | 49 | 2661 | 361 | 3347 | 279 | 166 | 26 | 313 | 39 | 2315 | 156 | 2935 | 354 | 161 | 36 |
| ɪ | 421 | 52 | 2323 | 356 | 2918 | 252 | 92 | 21 | 386 | 22 | 1984 | 190 | 2568 | 204 | 84 | 11 |
| eː | 420 | 45 | 2466 | 290 | 2923 | 229 | 179 | 29 | 399 | 30 | 2050 | 199 | 2548 | 303 | 164 | 40 |
| ɛ | 667 | 82 | 2046 | 242 | 2803 | 265 | 218 | 38 | 568 | 43 | 1758 | 138 | 2550 | 251 | 201 | 42 |
| æ | 682 | 90 | 2025 | 239 | 2796 | 300 | 213 | 38 | 569 | 42 | 1740 | 171 | 2538 | 277 | 191 | 27 |
| ɑː | 785 | 100 | 1270 | 100 | 2792 | 402 | 224 | 36 | 649 | 66 | 1088 | 139 | 2618 | 247 | 200 | 40 |
| ɔː | 501 | 86 | 941 | 145 | 2972 | 296 | 200 | 37 | 471 | 68 | 812 | 110 | 2671 | 200 | 185 | 37 |
| o | 501 | 100 | 931 | 162 | 2961 | 248 | 192 | 40 | 461 | 64 | 814 | 148 | 2682 | 154 | 168 | 29 |
| ʊ | 429 | 46 | 1169 | 154 | 2939 | 202 | 100 | 20 | 402 | 31 | 980 | 105 | 2626 | 135 | 88 | 14 |
| uː | 403 | 52 | 965 | 294 | 2918 | 219 | 164 | 38 | 355 | 32 | 775 | 133 | 2579 | 180 | 149 | 33 |
| ə | 624 | 89 | 1780 | 207 | 2877 | 439 | 80 | 21 | 557 | 56 | 1473 | 177 | 2608 | 314 | 66 | 13 |
| ʌ | 695 | 72 | 1556 | 126 | 2828 | 313 | 122 | 18 | 588 | 45 | 1312 | 119 | 2583 | 148 | 105 | 14 |

**Table 2G (ii) Mean Frequencies of F1, F2 and F3 in hertz and standard deviation of 12 oral monophthongs at the vowel midpoint for female and male speakers in Full sentences (FS)**

| | FS - Female speakers | | | | | | | | FS - Male speakers | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Vowel | F1 | SD | F2 | SD | F3 | SD | Duration | SD | F1 | SD | F2 | SD | F3 | SD | Duration | SD |
| iː | 341 | 26 | 2552 | 369 | 3182 | 318 | 129 | 30 | 314 | 29 | 2238 | 133 | 2761 | 276 | 119 | 19 |
| ɪ | 399 | 31 | 2333 | 252 | 2857 | 197 | 75 | 24 | 386 | 28 | 1947 | 210 | 2525 | 190 | 68 | 12 |
| eː | 418 | 51 | 2351 | 268 | 2804 | 210 | 129 | 24 | 416 | 31 | 1932 | 226 | 2345 | 233 | 116 | 23 |
| ɛ | 637 | 110 | 2016 | 253 | 2796 | 270 | 161 | 26 | 564 | 54 | 1735 | 147 | 2513 | 217 | 149 | 25 |
| æ | 662 | 92 | 2005 | 214 | 2803 | 243 | 167 | 18 | 550 | 39 | 1658 | 190 | 2474 | 223 | 153 | 20 |
| ɑː | 752 | 118 | 1365 | 99 | 2809 | 372 | 172 | 39 | 622 | 46 | 1152 | 146 | 2600 | 210 | 164 | 47 |
| ɔː | 520 | 91 | 950 | 146 | 2946 | 328 | 168 | 26 | 489 | 66 | 864 | 268 | 2653 | 152 | 157 | 25 |
| o | 518 | 115 | 976 | 181 | 2879 | 236 | 151 | 21 | 519 | 70 | 1084 | 488 | 2610 | 326 | 120 | 32 |
| ʊ | 410 | 39 | 1196 | 135 | 2896 | 182 | 76 | 21 | 384 | 29 | 1006 | 134 | 2613 | 148 | 70 | 15 |
| uː | 374 | 43 | 1055 | 283 | 2860 | 233 | 128 | 28 | 356 | 46 | 895 | 282 | 2593 | 219 | 121 | 21 |
| ə | 600 | 111 | 1797 | 222 | 3020 | 293 | 81 | 22 | 529 | 70 | 1478 | 187 | 2606 | 270 | 65 | 14 |
| ʌ | 603 | 94 | 1590 | 100 | 2920 | 255 | 77 | 24 | 512 | 48 | 1312 | 141 | 2637 | 193 | 68 | 21 |

## Appendix 3A

**Table 3A (i): Mean formant frequencies with standard deviation for the first three formants of six diphthongs at the seven equidistant points in time i.e. 20%, 30% …, 80% in Hertz pooled over speakers and contexts**

| | | 20% | SD | 30% | SD | 40% | SD | 50% | SD | 60% | SD | 70% | SD | 80% | SD |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | ɑe | 704 | 122 | 711 | 129 | 673 | 115 | 612 | 100 | 553 | 83 | 511 | 76 | 476 | 69 |
| | ɑʊ | 646 | 97 | 640 | 95 | 616 | 89 | 581 | 74 | 543 | 66 | 507 | 61 | 475 | 59 |
| | eɐ | 457 | 90 | 488 | 83 | 530 | 80 | 583 | 104 | 624 | 113 | 644 | 120 | 650 | 136 |
| F1 | ɪɐ | 341 | 43 | 353 | 42 | 394 | 49 | 464 | 67 | 556 | 102 | 615 | 113 | 628 | 124 |
| | oe | 428 | 46 | 440 | 52 | 448 | 54 | 449 | 53 | 441 | 50 | 435 | 46 | 428 | 50 |
| | ʊɑ | 415 | 46 | 454 | 64 | 518 | 81 | 577 | 99 | 617 | 104 | 637 | 107 | 637 | 111 |
| | | | | | | | | | | | | | | | |
| | ɑe | 1330 | 182 | 1447 | 184 | 1594 | 201 | 1770 | 231 | 1910 | 264 | 2021 | 291 | 2079 | 342 |
| | ɑʊ | 1137 | 140 | 1106 | 133 | 1065 | 124 | 1022 | 133 | 973 | 117 | 930 | 133 | 913 | 163 |
| | eɐ | 2117 | 280 | 2018 | 274 | 1947 | 221 | 1798 | 221 | 1687 | 189 | 1597 | 174 | 1559 | 171 |
| F2 | ɪɐ | 2409 | 347 | 2414 | 311 | 2320 | 286 | 2075 | 266 | 1801 | 207 | 1620 | 173 | 1510 | 161 |
| | oe | 955 | 172 | 1032 | 182 | 1225 | 238 | 1500 | 295 | 1772 | 326 | 1942 | 335 | 2046 | 321 |
| | ʊɑ | 1048 | 191 | 981 | 155 | 1015 | 137 | 1090 | 144 | 1145 | 140 | 1190 | 154 | 1236 | 162 |
| | | | | | | | | | | | | | | | |
| | ɑe | 2642 | 315 | 2610 | 328 | 2597 | 323 | 2635 | 330 | 2654 | 285 | 2697 | 306 | 2700 | 309 |
| | ɑʊ | 2699 | 303 | 2718 | 293 | 2744 | 293 | 2774 | 300 | 2786 | 303 | 2789 | 298 | 2782 | 294 |
| F3 | eɐ | 2745 | 302 | 2717 | 284 | 2692 | 291 | 2674 | 299 | 2667 | 303 | 2648 | 326 | 2681 | 323 |
| | ɪɐ | 3012 | 408 | 2995 | 396 | 2843 | 341 | 2720 | 307 | 2696 | 323 | 2649 | 335 | 2643 | 351 |
| | oe | 2723 | 232 | 2690 | 235 | 2630 | 241 | 2576 | 235 | 2570 | 259 | 2610 | 279 | 2644 | 289 |
| | ʊɑ | 2720 | 241 | 2763 | 244 | 2780 | 276 | 2759 | 296 | 2738 | 295 | 2745 | 293 | 2740 | 287 |

**Table 3A (ii): Mean formant frequencies with standard deviation for the first two formants of six diphthongs at the seven equidistant points in time i.e. 20%, 30%, … ,80% in Lobanov normalised pooled over speakers and contexts**

| | Vowel | 20% | SD | 30% | SD | 40% | SD | 50% | SD | 60% | SD | 70% | SD | 80% | SD |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | ɑe | 1.71 | 0.58 | 1.76 | 0.64 | 1.47 | 0.59 | 0.99 | 0.58 | 0.51 | 0.55 | 0.17 | 0.52 | -0.11 | 0.50 |
| | ɑʊ | 1.26 | 0.46 | 1.22 | 0.48 | 1.03 | 0.47 | 0.76 | 0.45 | 0.45 | 0.44 | 0.16 | 0.43 | -0.10 | 0.42 |
| | eɐ | -0.30 | 0.68 | -0.04 | 0.62 | 0.31 | 0.56 | 0.74 | 0.73 | 1.05 | 0.70 | 1.20 | 0.70 | 1.24 | 0.81 |
| F1 | ɪɐ | -1.28 | 0.34 | -1.16 | 0.31 | -0.82 | 0.33 | -0.24 | 0.44 | 0.50 | 0.57 | 0.97 | 0.56 | 1.06 | 0.62 |
| | oe | -0.52 | 0.33 | -0.42 | 0.35 | -0.35 | 0.35 | -0.35 | 0.33 | -0.41 | 0.32 | -0.46 | 0.27 | -0.52 | 0.31 |
| | ʊɑ | -0.62 | 0.34 | -0.29 | 0.45 | 0.23 | 0.53 | 0.70 | 0.57 | 1.02 | 0.55 | 1.18 | 0.55 | 1.17 | 0.57 |
| | ɑe | -0.38 | 0.26 | -0.15 | 0.26 | 0.13 | 0.29 | 0.47 | 0.34 | 0.75 | 0.35 | 0.96 | 0.37 | 1.06 | 0.47 |
| | ɑʊ | -0.75 | 0.21 | -0.81 | 0.21 | -0.89 | 0.20 | -0.97 | 0.24 | -1.06 | 0.21 | -1.14 | 0.25 | -1.18 | 0.29 |
| F2 | eɐ | 1.15 | 0.41 | 0.96 | 0.41 | 0.82 | 0.32 | 0.54 | 0.39 | 0.33 | 0.38 | 0.16 | 0.36 | 0.09 | 0.36 |
| | ɪɐ | 1.72 | 0.51 | 1.73 | 0.40 | 1.54 | 0.35 | 1.08 | 0.41 | 0.56 | 0.37 | 0.20 | 0.30 | -0.02 | 0.28 |
| | oe | -1.10 | 0.30 | -0.95 | 0.29 | -0.58 | 0.37 | -0.05 | 0.47 | 0.47 | 0.48 | 0.80 | 0.47 | 1.00 | 0.41 |
| | ʊɑ | -0.92 | 0.28 | -1.05 | 0.24 | -0.98 | 0.22 | -0.84 | 0.23 | -0.74 | 0.22 | -0.65 | 0.24 | -0.56 | 0.25 |

# Appendix 5A

## Table 5A (i): Responses to LEAP-Q (via Google forms)

## LEAP-Q - About personal and linguistic background

| 3. What is your gender? | 4. What is your age group? | 1. How many languages do you speak? | 2. What is your first language? | 3. What is your second language? | 4. Which language do you speak at home? | 5. Which language do you speak with friends at | 6. Please list all the languages you know in order of dominance. |
|---|---|---|---|---|---|---|---|
| Female | 18 - 24 | 4 | Urdu | Punjabi | Urdu | English | urdu,pushto,english and pujabi |
| Female | 18 - 24 | 3 | Punjabi | Urdu | Punjabi | Urdu | PUNJABI URDU ENGLISH |
| Male | 18 - 24 | 4 | Sariki | Sariki | Sariki | Urdu | SARIKI, PUNJABI, URDU, ENGLISH, MEWATI,RANGRI |
| Male | 18 - 24 | 3 | Punjabi | English | Punjabi | Urdu | Punjabi, Urdu, English |
| Male | 18 - 24 | 4 | Mewati | Urdu | Mewati | Urdu | Mewati, Punjabi, Urdu and English. |
| Female | 18 - 24 | 3 | Urdu | English | Urdu | English | Urdu , English , Punjabi |
| Female | 18 - 24 | 3 | Urdu | Urdu | Urdu | Urdu | Urdu English punjabi |
| Female | 18 - 24 | 3 | Punjabi | English | Urdu | Urdu | Urdu Punjabi English |
| Female | 18 - 24 | 3 | Urdu | English | Urdu | Urdu | Urdu, English, Punjabi |
| Female | 18 - 24 | 3 | Urdu | English | Urdu | English | Urdu , English , Punjabi |
| Female | 18 - 24 | 3 | Urdu | Punjabi | Urdu | Urdu | urdu punjabi English |
| Female | 18 - 24 | 3 | Urdu | English | Urdu | Urdu | English, Urdu, Punjabi |
| Female | 18 - 24 | 2 | Urdu | English | Urdu | Urdu | English, Urdu , Punjabi |
| Female | 18 - 24 | 4 | Punjabi | Urdu | Punjabi | English | Urdu,english,punjabi, and siraiki |
| Female | 18 - 24 | 4 | Punjabi | Urdu | Punjabi | English | Urdu, english,punjabi,and siraiki |
| Female | 18 - 24 | 2 | Urdu | English | Urdu | Urdu | Urdu, English |
| Female | 18 - 24 | 3 | Urdu | English | Urdu | Urdu | Urdu, English and Punjabi |
| Female | 18 - 24 | 3 | Punjabi | Urdu | Punjabi | Urdu | Urdu, English,punjabi |
| Female | 18 - 24 | 3 | Urdu | English | Urdu | Urdu | English, Urdu, Punjabi,Fay languagae |
| Female | 18 - 24 | 3 | Hindko | Urdu | Urdu | Urdu | English, Urdu, Hindko |

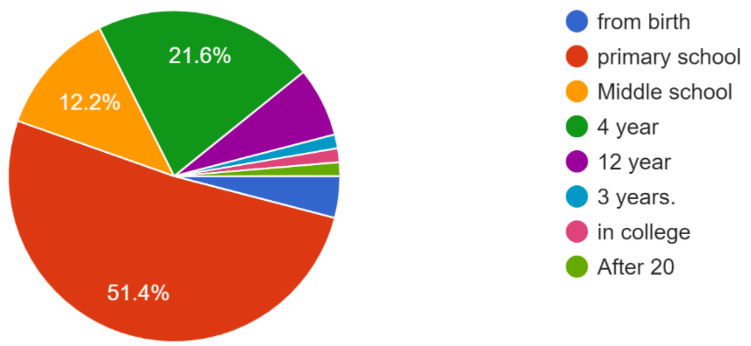
| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Female | 18 - 24 | 3 | Urdu | Punjabi | Urdu | English | URDU PUNJABI ENGLIG |
| Female | 18 - 24 | 3 | Urdu | Punjabi | Urdu | English | ENGLISH, URDU,PUNJABI |
| Female | 18 - 24 | 3 | Urdu | Punjabi | Punjabi | Urdu | Urdu Punjabi English |
| Female | 18 - 24 | 3 | Punjabi | Urdu | Punjabi | Urdu | Urdu, english, punjabi |
| Female | 18 - 24 | 3 | Urdu | English | Urdu | English | english,urdu,punjabi |
| Female | 18 - 24 | 3 | Urdu | Punjabi | Punjabi | English | Punjabi, English, Urdu |
| Female | 18 - 24 | 4 | Urdu | English | Urdu | Urdu | English. Punjabi. Urdu |
| Female | 18 - 24 | 3 | Urdu | English | Urdu | Urdu | Urdu, English , Punjabi |
| Female | 18 - 24 | 2 | Urdu | English | Urdu | Urdu | URDU , ENGLISH |
| Female | 18 - 24 | 4 | Punjabi | Urdu | Urdu | Urdu | Urdu, English, Punjabi, Pushto |
| Female | 18 - 24 | 3 | Urdu | English | Urdu | Urdu | urdu,english,punjabi |
| Female | 18 - 24 | 3 | Urdu | Punjabi | Punjabi | Urdu | english, urdu, punjabi |
| Female | 18 - 24 | 2 | Urdu | English | Urdu | Urdu | URDU, ENGLISH, PUNJABI, ARABIC |
| Female | 18 - 24 | 3 | Punjabi | Urdu | Punjabi | Urdu | Punjabi Urdu English |
| Female | 18 - 24 | 3 | Sariki | Urdu | Urdu | Urdu | English, Urdu,Saraiki or punjabi |
| Female | 18 - 24 | 3 | Urdu | English | Urdu | Urdu | Urdu English Punjabi |
| Female | 18 - 24 | 3 | Urdu | Punjabi | Urdu | Punjabi | Urdu, Punjabi, English |
| Female | 18 - 24 | 3 | Urdu | Punjabi | Urdu | Urdu | Urdu English Punjabi |
| Female | 18 - 24 | 3 | Urdu | Punjabi | Urdu | Urdu | Urdu, English & Punjabi |
| Female | 18 - 24 | 3 | Urdu | English | Punjabi | Urdu | Urdu english punjabi |
| Female | 18 - 24 | 3 | Urdu | Punjabi | Urdu | Urdu | Urdu Punjabi English |
| Female | 18 - 24 | 3 | Urdu | Punjabi | Urdu | Urdu | Urdu Punjabi English |
| Male | 18 - 24 | 3 | Punjabi | Urdu | Punjabi | Punjabi | Punjabi , Urdu , English |
| Male | 18 - 24 | 3 | Punjabi | Urdu | Punjabi | Urdu | English .urdu and pujabi |
| Female | 18 - 24 | 3 | Punjabi | Urdu | Urdu | Urdu | Urdu, English and Punjabi |
| Female | 18 - 24 | 4 | Punjabi | Urdu | Punjabi | Urdu | Punjabi,  urdu, english, arabic |
| Male | 18 - 24 | 3 | Punjabi | Urdu | Punjabi | Urdu | English, Urdu, Punjabi. |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Male | 18 - 24 | 2 | Urdu | English | Urdu | Urdu | Urdu and english |
| Female | 18 - 24 | 3 | Punjabi | Urdu | Punjabi | Urdu | Punjabi, Urdu, English |
| Female | 18 - 24 | 3 | Urdu | English | Urdu | Urdu | Urdu, English, Punjabi |
| Female | 18 - 24 | 3 | Urdu | English | Urdu | English | urdu , english, punjabi |
| Female | 18 - 24 | 3 | Punjabi | Urdu | Punjabi | Urdu | Punjabi Urdu English |
| Male | 18 - 24 | 5 | mewati | Urdu | mewati and punjabi | Urdu | mewati(a language derived from rahistani language), urdu, Punjabi English, raghistani |
| Male | 18 - 24 | 3 | Urdu | English | Sariki | Urdu | Urdu , English , punjabi and sarakii |
| Male | 18 - 24 | 3 | Urdu | English | Punjabi | Urdu | Urdu, English, Punjabi |
| Male | 18 - 24 | 3 | Urdu | Punjabi | Urdu | Urdu | Urdu, Punjabi, English |
| Male | 18 - 24 | 3 | Punjabi | Punjabi | Punjabi | Urdu | punjabi |
| Male | 18 - 24 | 3 | Punjabi | Urdu | Punjabi | English | Punjabi, Urdu, English |
| Female | 18 - 24 | 2 | Urdu | English | Urdu | English | english...urdu |
| Male | 18 - 24 | 3 | Punjabi | Urdu | Punjabi | Urdu | 1 punjabi<br>2 urdu<br>3 english |
| Male | 18 - 24 | 3 | Punjabi | Urdu | Punjabi | Urdu | PUNJABI,URDU,ENGLISH |
| Female | 18 - 24 | 3 | Punjabi | English | Punjabi | Urdu | Punjabi urdu english |
| Female | 18 - 24 | 2 | Urdu | English | Urdu | English | english ..urdu |
| Female | 18 - 24 | 3 | Punjabi | Urdu | Punjabi | Urdu | Punjabi Urdu English |
| Female | 18 - 24 | 3 | Punjabi | Urdu | Punjabi | Urdu | Punjabi Urdu English |
| Female | 18 - 24 | 4 | Punjabi | Urdu | Punjabi | Urdu | Punjabi, Urdu, Arabic, English, Persian |
| Female | 18 - 24 | 3 | Urdu | Punjabi | Urdu | Urdu | Urdu Punjabi English |
| Male | 18 - 24 | 3 | Punjabi | Urdu | Punjabi | English | Punjabi urdu english |
| Male | 18 - 24 | 3 | Punjabi | English | Punjabi | Punjabi | Punjabi, Urdu,Enlish |
| Male | 18 - 24 | 3 | Urdu | Punjabi | Urdu | Urdu | Urdu Punjabi English |
| Male | 18 - 24 | 3 | Urdu | English | Urdu | Urdu | English |
| Male | 18 - 24 | 3 | Punjabi | Urdu | urdu and punjabi | urdu and english | Urdu,Punjabi and English |
| Male | 18 - 24 | 3 | Punjabi | Urdu | Urdu | Urdu | punjabi, urdu, english |

**LEAP-Q- About Language Acquisition**

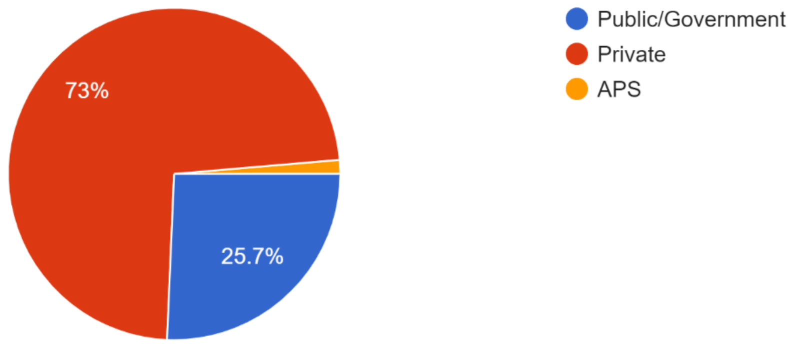
1. What was your age when you began to learn English?

74 responses



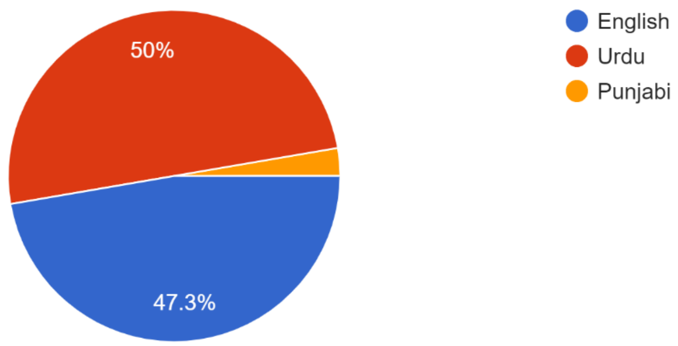

2. What type of primary school did you attend?

74 responses



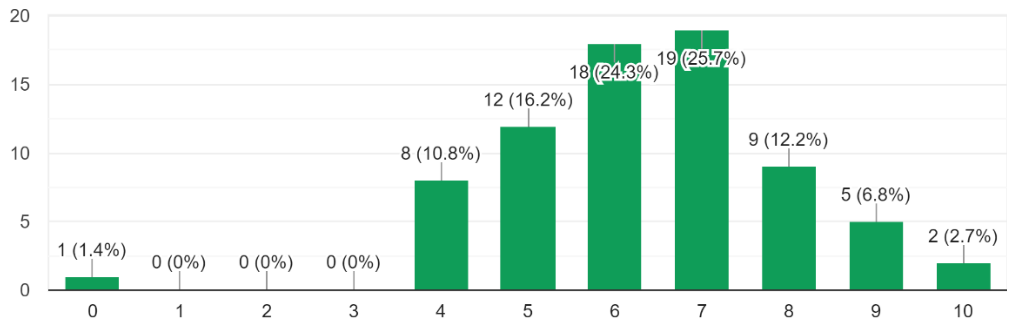

3. What was the medium of instruction in your primary school?

74 responses



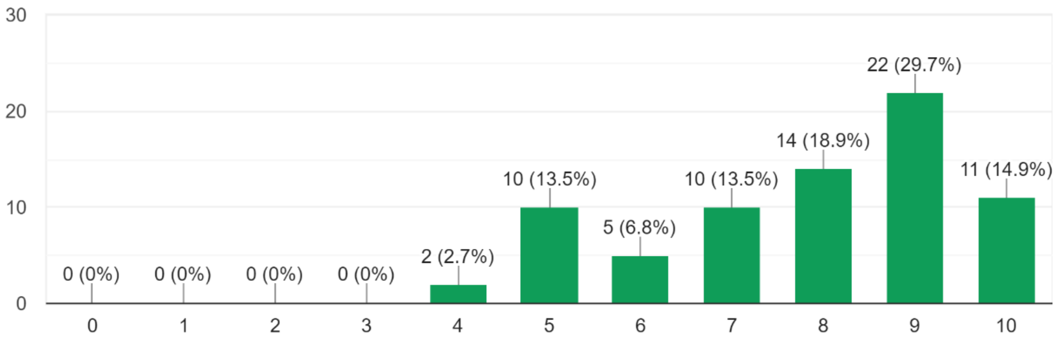
4. On a scale from zero to ten, please select your level of proficiency in speaking English.

74 responses



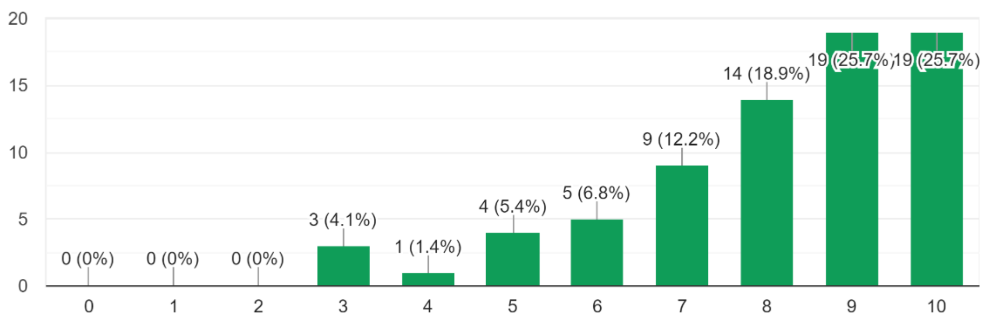

5. On a scale from zero to ten, please select your level of proficiency in understanding spoken English.

74 responses



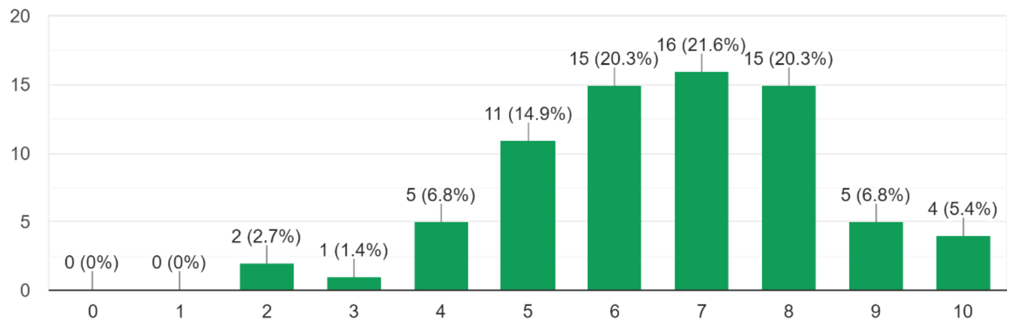

6. On a scale from zero to ten, please select your level of proficiency in reading English.

74 responses



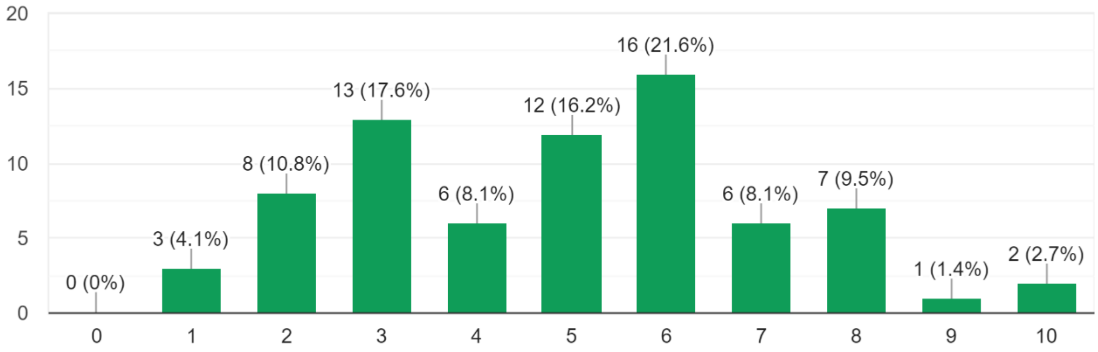
7. On a scale from zero to ten, please choose to what extent you are currently exposed to English when interacting with friends:

74 responses



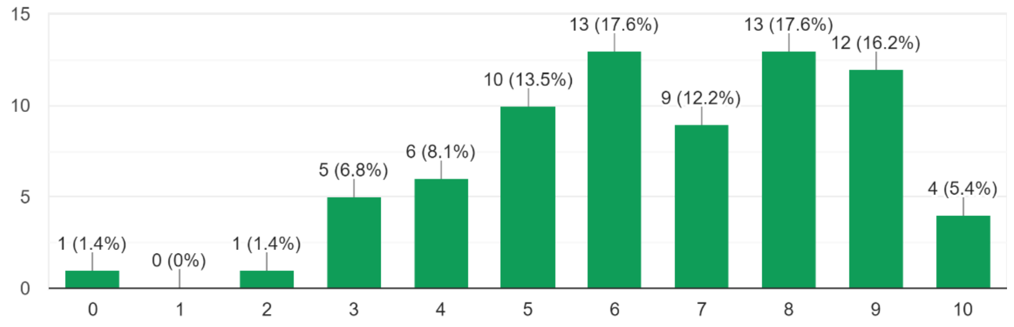

8. On a scale from zero to ten, please choose to what extent you are currently exposed to English when interacting with family:

74 responses



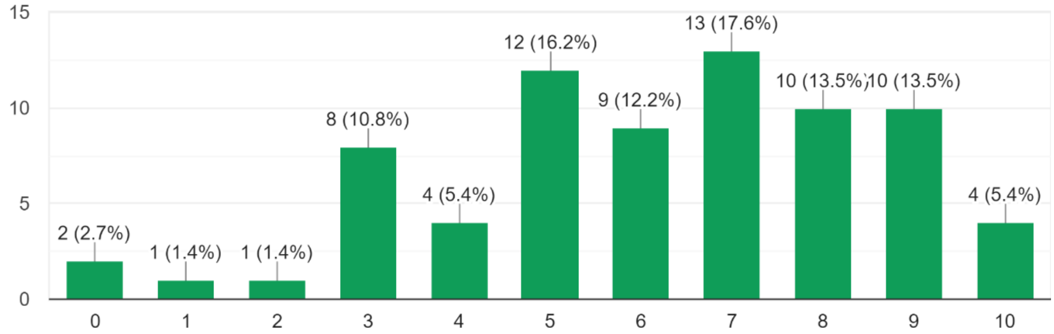

9. On a scale from zero to ten, please choose to what extent you are currently exposed to English when watching TV:

74 responses



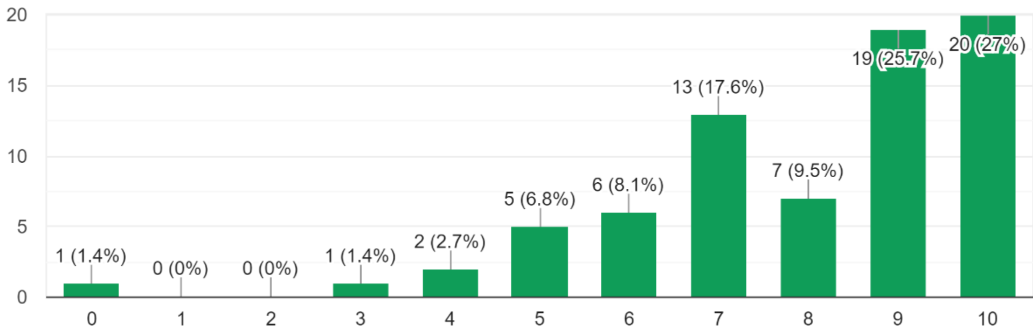
10. On a scale from zero to ten, please choose to what extent you are currently exposed to English when listening to radio/music:

74 responses



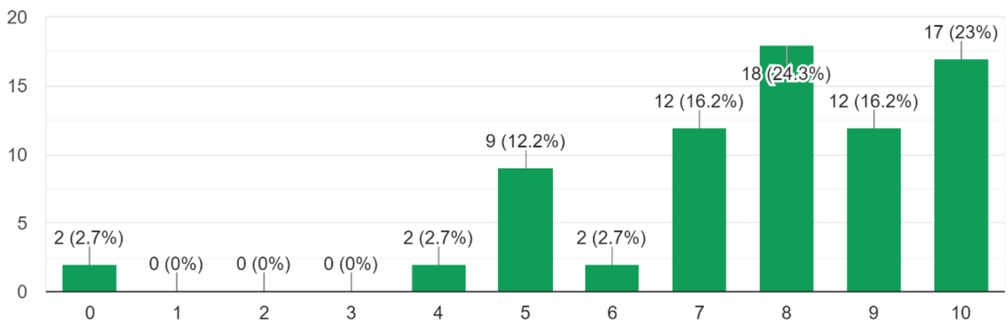

11. On a scale from zero to ten, please choose to what extent you are currently exposed to English when reading:

74 responses



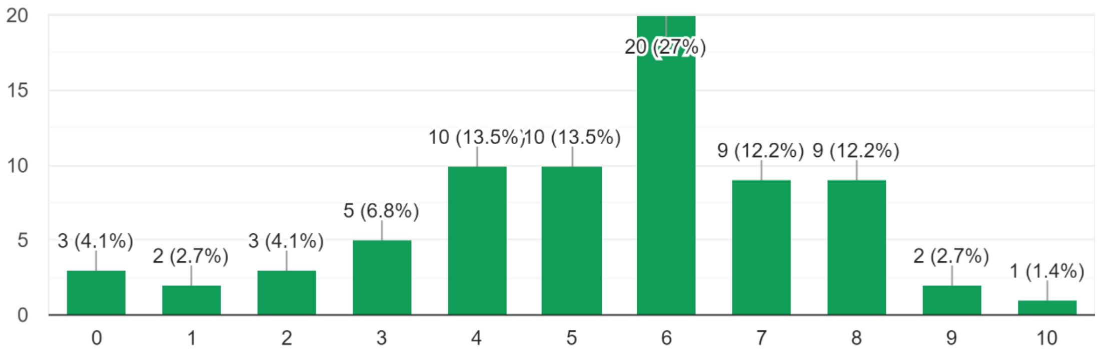

12. On a scale from zero to ten, please choose to what extent you are currently exposed to English for a language course.

74 responses



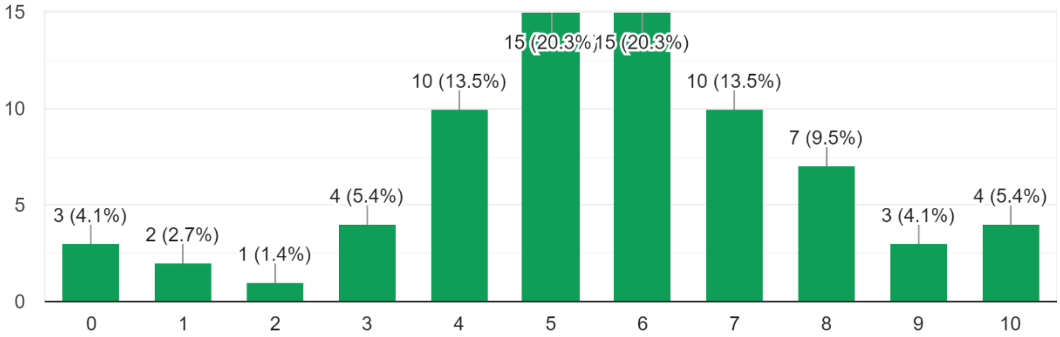
13. On a scale from zero to ten, please choose in your perception, how much of a foreign accent do you have in the English language?

74 responses




14. On a scale from zero to ten, please choose how frequently others identify you as a non-native speaker based on your accent in English.

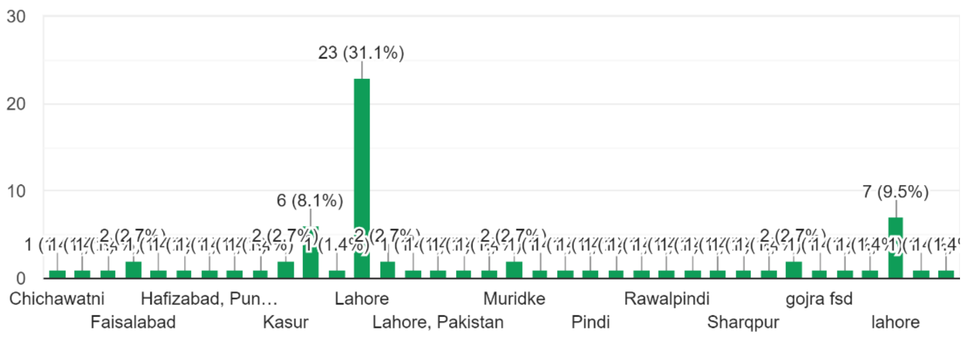
74 responses




**LEAP-Q - About Social Background**

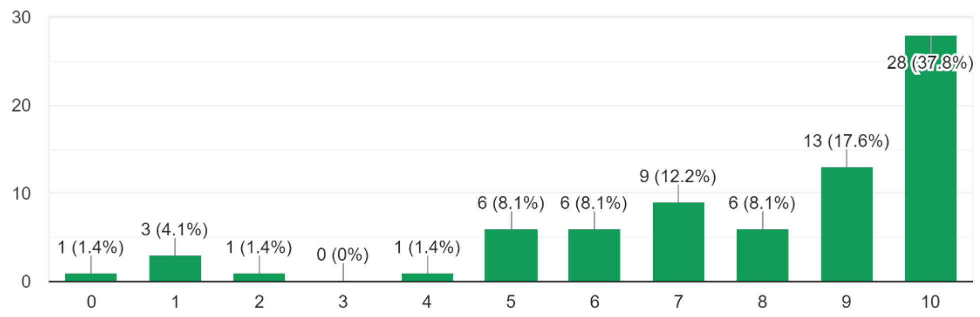
1. What town/city were you raised in?

74 responses



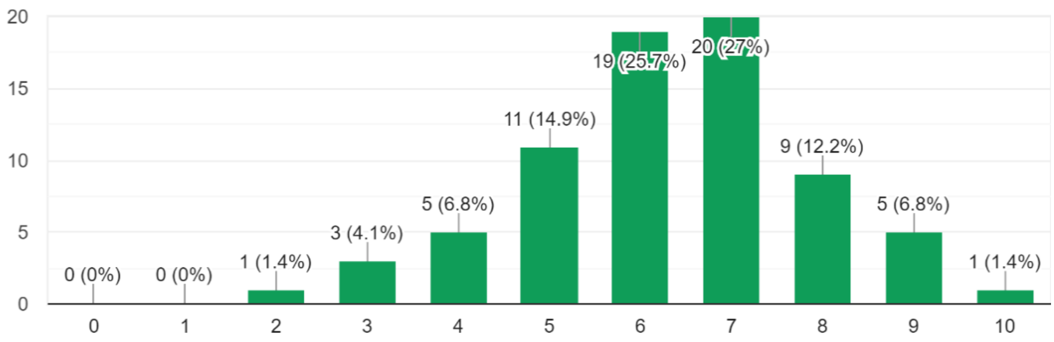
2. On a scale from zero to ten where would you place this town/city?

74 responses



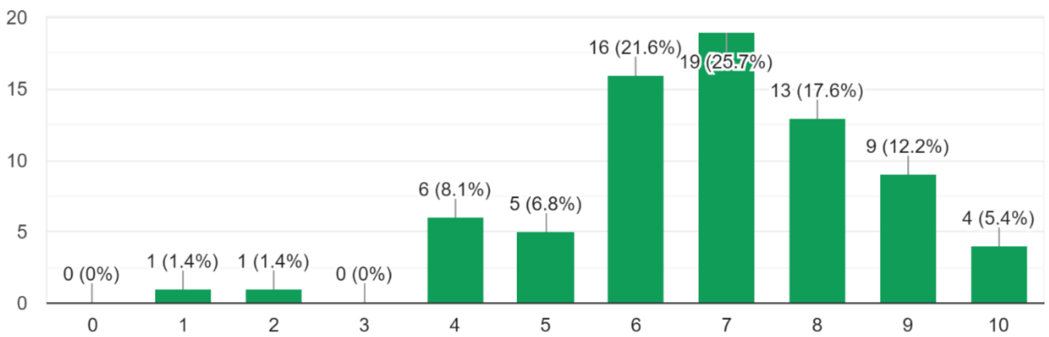

3. On a scale from zero to ten where do you think you stand at this time in your life, relative to other people in your country?

74 responses




4. On a scale from zero to ten where do you think you stand at this time in your life, relative to other people in your village/town/city?

74 responses



## Appendix 5B

## Pratt Script for SSBE stimuli

```
# TextGrid tier with intervals containing whole vowels.
vowel_tier = 1
# Fraction along vowel duration to start taking samples (inclusive).
sample_start = 0.2
# Fraction along vowel duration to stop taking samples (inclusive).
sample_end = 0.8
# Number of sample points to get formant values along vowel interval.
sample_count = 7
# Formant resolution
time_step = 0.01
maximum_number_of_formants = 5
window_length = 0.025
preemphasis_from = 30

# Male
root$ = "/Working/Documents/R/raw/Praat/RP/M"
output$ = "/Working/Documents/R/raw/Praat/RP/results_m.txt"
maximum_formant = 5000
@readVowels

# Female
root$ = "/Working/Documents/R/raw/Praat/RP/F"
output$ = "/Working/Documents/R/raw/Praat/RP/results_f.txt"
maximum_formant = 5500
@readVowels


procedure readVowels
	printline ------------------------------------------------------------------------
	Create Strings as directory list... folders 'root$'
	folder_count = Get number of strings

	fullpaths_count = 0
	for folder_num from 1 to folder_count
		select Strings folders
		folder$ = Get string... folder_num

		Create Strings as file list... files 'root$'/'folder$'/*.wav
		file_count = Get number of strings
		for file_num from 1 to file_count
			file_name$ = Get string... file_num
			fullpath$ = "'root$'/'folder$'/'file_name$'"
			fullpaths_count = fullpaths_count + 1
			fullpaths$[fullpaths_count] = fullpath$
			file_names$[fullpaths_count] = file_name$
			printline 'fullpath$'
		endfor

		select Strings files
		Remove
	endfor

	select Strings folders
	Remove


	resultline$ = "Subject	Gender	Repetition	Vowel	Duration"
	for sample_num from 1 to sample_count
		resultline$ = resultline$ + "	F1_'sample_num'	F2_'sample_num'	F3_'sample_num'"
	endfor
	resultline$ = resultline$ + newline$
	fileappend "'output$'" 'resultline$'

	for fullpath_num from 1 to fullpaths_count
		file_name$ = file_names$[fullpath_num]
		fullpath$ = fullpaths$[fullpath_num]
		textgrid$ = left$(fullpath$, length(fullpath$)-4) + ".TextGrid"

		subject$ = left$(file_name$, length(file_name$)-7)
		repetition$ = mid$(file_name$, length(file_name$)-4, 1)
		gender$ = left$(file_name$, 1)
```

```
			printline Processing 'subject$' (gender 'gender$') repetition 'repetition$':
			printline --'fullpath$'
			printline --'textgrid$'

			if fileReadable (textgrid$)

				Read from file... 'fullpath$'
				soundname$ = selected$ ("Sound", 1)
				To Formant (burg)... time_step maximum_number_of_formants maximum_formant
window_length preemphasis_from

				Read from file... 'textgrid$'
				interval_count = Get number of intervals... vowel_tier
				for interval_num from 1 to interval_count
					vowel$ = Get label of interval... vowel_tier interval_num
					vowel$ = replace_regex$ (vowel$, "\n", "", 0)
					vowel$ = replace_regex$ (vowel$, "\r", "", 0)
					# Check if interval label is non-empty (which we take to mean it contains a vowel).
					if vowel$ <> ""
						# Get the interval's start and end time (in milliseconds):
						start = Get starting point... vowel_tier interval_num
						end = Get end point... vowel_tier interval_num

						# Duration (in seconds) of whole vowel interval.
						duration = (end - start)
						duration_ms = duration*1000

						# Add first few columns to the row that will be written to the output text
file.
						resultline$ = "'subject$'	'gender$'	'repetition$'	'vowel$'
	'duration_ms'"

						# Duration (in seconds) from first sample to last sample.
						sampling_duration = duration * (sample_end - sample_start)
						# Distance (in seconds) between each sample.
						sampling_delta = sampling_duration / (sample_count - 1)
						# Point in time to start sampling.
						sampling_start = start + duration * sample_start

						# Add remaining columns for samples along the vowel.
						for sample_num from 0 to (sample_count-1)
							sample_time = sampling_start + sampling_delta * sample_num

							# Get the formant values at the interval
							select Formant 'soundname$'
							f1 = Get value at time... 1 sample_time Hertz Linear
							f2 = Get value at time... 2 sample_time Hertz Linear
							f3 = Get value at time... 3 sample_time Hertz Linear
							resultline$ = resultline$ + "	'f1'	'f2'	'f3'"
						endfor

						# Save result to text file:
						resultline$ = resultline$ + newline$
						fileappend "'output$'" 'resultline$'
						select TextGrid 'soundname$'
					endif
				endfor
			else
				printline --'textgrid$' not found
			endif
		endfor
endproc
```

## Appendix 5C

**Table 5C (i): Mean frequencies of F1, F2 and F3 in Hertz and mean duration in milliseconds for the SSBE vowels in the bVd context**

| SSBE Vowel | F1(Hz) | F2(Hz) | F3(Hz) | Duration(ms) |
|---|---|---|---|---|
| iː | 290 | 2540 | 3160 | 306 |
| ɪ | 438 | 2173 | 2698 | 188 |
| ɛ | 615 | 1916 | 2563 | 213 |
| æ | 869 | 1587 | 2672 | 283 |
| ɑː | 681 | 1118 | 2728 | 323 |
| ɒ | 581 | 918 | 2785 | 182 |
| ɔː | 465 | 719 | 2797 | 328 |
| ʊ | 447 | 1162 | 2521 | 165 |
| uː | 318 | 1699 | 2300 | 336 |
| ʌ | 684 | 1267 | 2667 | 186 |
| ɜː | 610 | 1522 | 2557 | 315 |

**Table 5C (ii): Mean frequencies of F1, F2 and F3 in Hertz and mean duration in milliseconds for the SSBE vowels in the hVba context**

| SSBE Vowel | F1(Hz) | F2(Hz) | F3(Hz) | Duration(ms) |
|---|---|---|---|---|
| iː | 325 | 2559 | 3254 | 170 |
| ɪ | 437 | 2135 | 2719 | 107 |
| ɛ | 628 | 1808 | 2514 | 116 |
| æ | 881 | 1519 | 2766 | 128 |
| ɑː | 660 | 1151 | 2747 | 227 |
| ɒ | 576 | 947 | 2752 | 113 |
| ɔː | 434 | 771 | 2776 | 208 |
| ʊ | 449 | 1271 | 2389 | 107 |
| uː | 352 | 1773 | 2302 | 186 |
| ʌ | 665 | 1283 | 2578 | 108 |
| ɜː | 601 | 1466 | 2587 | 213 |

## Appendix 6A

**Table 6A (i): Mean frequencies of F1, F2 and F3 in Hertz and mean duration in milliseconds for the SSBE vowels in the hVd context**

| SSBE Vowel | F1(Hz) | F2(Hz) | F3(Hz) | Duration(ms) |
|---|---|---|---|---|
| iː | 294 | 2486 | 3080 | 310 |
| ɪ | 436 | 2215 | 2646 | 183 |
| ɛ | 608 | 2012 | 2697 | 172 |
| æ | 870 | 1628 | 2692 | 208 |
| ɑː | 675 | 1113 | 2588 | 298 |
| ɒ | 593 | 960 | 2801 | 184 |
| ɔː | 468 | 728 | 2815 | 285 |
| ʊ | 471 | 1405 | 2458 | 167 |
| uː | 306 | 1841 | 2372 | 311 |
| ʌ | 694 | 1290 | 2747 | 170 |
| ɜː | 609 | 1539 | 2551 | 300 |

## Appendix 6B

**Table 6B (i): Matrix representation of clustered pairs for one participant.**

| | FA | FB | FC | FD | FE | FF | FG | FH | FI | FJ | FK | FL | FM | FN | FO | FP | FQ | FR | FS | MA | MB | MC | MD | ME | MF | MG | MH | MI | MJ | MK | ML | MM | MN | MO | MP | MQ | MR | MS |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| FA | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 |
| FB | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 |
| FC | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 |
| FD | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 |
| FE | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| FF | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 |
| FG | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| FH | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 |
| FI | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 |
| FJ | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 |
| FK | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 |
| FL | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| FM | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 |
| FN | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| FO | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 |
| FP | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 |
| FQ | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| FR | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| FS | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| MA | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| MB | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| MC | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| MD | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 |
| ME | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| MF | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| MG | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| MH | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 |
| MI | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 |
| MJ | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 |
| MK | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 |
| ML | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 |
| MM | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 |
| MN | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 |
| MO | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 |
| MP | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 |
| MQ | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 |
| MR | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 |
| MS | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 |

## Appendix 6C

**Table 6C (i): Proximity matrix representation: the sum of all the individual participants' matrices**

| | iː | ɪ | ɛ | æ | ɑː | ɔː | ɒ | ʊ | uː | ɜː | ʌ | eɪ | aɪ | aʊ | əʊ | ɔɪ | ɪə | eə | ʊə |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| iː | 40 | 228 | 261 | 263 | 267 | 264 | 269 | 265 | 247 | 268 | 267 | 238 | 264 | 267 | 262 | 265 | 246 | 263 | 257 |
| ɪ | 228 | 40 | 230 | 236 | 265 | 266 | 267 | 263 | 261 | 252 | 261 | 196 | 259 | 269 | 259 | 269 | 195 | 236 | 248 |
| ɛ | 261 | 230 | 52 | 172 | 265 | 266 | 266 | 265 | 263 | 222 | 255 | 230 | 263 | 268 | 265 | 265 | 246 | 145 | 257 |
| æ | 263 | 236 | 172 | 56 | 254 | 262 | 253 | 268 | 265 | 208 | 252 | 221 | 262 | 258 | 253 | 266 | 235 | 146 | 256 |
| ɑː | 267 | 265 | 265 | 254 | 54 | 206 | 199 | 254 | 258 | 239 | 211 | 263 | 256 | 241 | 253 | 259 | 266 | 264 | 268 |
| ɔː | 264 | 266 | 266 | 262 | 206 | 50 | 150 | 242 | 251 | 252 | 236 | 271 | 265 | 234 | 232 | 240 | 266 | 269 | 261 |
| ɒ | 269 | 267 | 266 | 253 | 199 | 150 | 60 | 242 | 255 | 247 | 228 | 268 | 264 | 232 | 241 | 245 | 263 | 262 | 261 |
| ʊ | 265 | 263 | 265 | 268 | 254 | 242 | 242 | 60 | 204 | 245 | 231 | 264 | 261 | 252 | 220 | 260 | 263 | 265 | 253 |
| uː | 247 | 261 | 263 | 265 | 258 | 251 | 255 | 204 | 66 | 257 | 256 | 264 | 258 | 254 | 220 | 256 | 264 | 269 | 232 |
| ɜː | 268 | 252 | 222 | 208 | 239 | 252 | 247 | 245 | 257 | 72 | 226 | 250 | 257 | 256 | 223 | 257 | 240 | 218 | 248 |
| ʌ | 267 | 261 | 255 | 252 | 211 | 236 | 228 | 231 | 256 | 226 | 56 | 267 | 255 | 255 | 250 | 261 | 271 | 263 | 263 |
| eɪ | 238 | 196 | 230 | 221 | 263 | 271 | 268 | 264 | 264 | 250 | 267 | 46 | 264 | 265 | 255 | 264 | 195 | 214 | 248 |
| aɪ | 264 | 259 | 263 | 262 | 256 | 265 | 264 | 261 | 258 | 257 | 255 | 264 | 40 | 248 | 241 | 212 | 264 | 262 | 261 |
| aʊ | 267 | 269 | 268 | 258 | 241 | 234 | 232 | 252 | 254 | 256 | 255 | 265 | 248 | 42 | 245 | 230 | 267 | 268 | 251 |
| əʊ | 262 | 259 | 265 | 253 | 253 | 232 | 241 | 220 | 220 | 223 | 250 | 255 | 241 | 245 | 98 | 239 | 250 | 262 | 251 |
| ɔɪ | 265 | 269 | 265 | 266 | 259 | 240 | 245 | 260 | 256 | 257 | 261 | 264 | 212 | 230 | 239 | 44 | 267 | 268 | 254 |
| ɪə | 246 | 195 | 246 | 235 | 266 | 266 | 263 | 263 | 264 | 240 | 271 | 195 | 264 | 267 | 250 | 267 | 46 | 226 | 171 |
| eə | 263 | 236 | 145 | 146 | 264 | 269 | 262 | 265 | 269 | 218 | 263 | 214 | 262 | 268 | 262 | 268 | 226 | 58 | 252 |
| ʊə | 257 | 248 | 257 | 256 | 268 | 261 | 261 | 253 | 232 | 248 | 263 | 248 | 261 | 251 | 251 | 254 | 171 | 252 | 88 |

**Appendix 6D**

**Table 6D (i): Coordinates for each SSBE diphthong in three-dimensional space**

| Diphthong | x-axis | y-axis | z-axis |
|---|---|---|---|
| ɪə | -112.92 | 15.67 | -17.39 |
| ʊə | -76.35 | 99.45 | -28.84 |
| eɪ | -134.34 | -42.43 | 11.79 |
| ɔɪ | 139.99 | -7.62 | -11.50 |
| əʊ | 68.12 | 34.67 | 6.89 |
| eə | -100.66 | -99.14 | 23.28 |
| aʊ | 113.75 | 33.29 | 90.25 |
| aɪ | 102.41 | 33.87 | -74.46 |