Acoustic Analysis of Sindhi Speech - A Pre-curser for an ASR System

by

Ayaz Keerio

Submitted in fulfilment of
the requirements for the
degree of Doctor of
Philosophy

School of Engineering and Design
University of Sussex, UK

November 2010

DECLARATION

I hereby declare that this work has not been submitted in support of an application for another degree or qualification at this or any other university or institution of learning.

Ayaz Keerio
October 2010

ABSTRACT

The functional and formative properties of speech sounds are usually referred to as acoustic-phonetics in linguistics. This research aims to demonstrate acoustic-phonetic features of the elemental sounds of Sindhi, which is a branch of the Indo-European family of languages mainly spoken in the Sindh province of Pakistan and in some parts of India. In addition to the available articulatory-phonetic knowledge; acoustic-phonetic knowledge has been classified for the identification and classification of Sindhi language sounds. Determining the acoustic features of the language sounds helps to bring together the sounds with similar acoustic characteristics under the name of one natural class of meaningful phonemes. The obtained acoustic features and corresponding statistical results for a particular natural class of phonemes provides a clear understanding of the meaningful phonemes of Sindhi and it also helps to eliminate redundant sounds present in the inventory. At present Sindhi includes nine redundant, three interchanging, three substituting, and three confused pairs of consonant sounds. Some of the unique acoustic-phonetic features of Sindhi highlighted in this study are determining the acoustic features of the large number of the contrastive voiced implosives of Sindhi and the acoustic impact of the language flexibility in terms of the insertion and digestion of the short vowels in the utterance. In addition to this the issue of the presence of the affricate class of sounds and the diphthongs in Sindhi is addressed. The compilation of the meaningful language phoneme set by learning their acoustic-phonetic features serves one of the major goals of this study; because twelve such sounds of Sindhi are studied that are not yet part of the language alphabet. The main acoustic features learned for the phonological structures of Sindhi are the fundamental frequency, formants, and the duration — along with the analysis of the obtained acoustic waveforms, the formant tracks and the computer generated spectrograms. The impetus for doing such research comes from the fact that detailed knowledge of the sound characteristics of the language-elements has a broad variety of applications — from developing accurate synthetic speech production systems to modeling robust speaker-independent speech recognizers. The major research achievements and contributions this study provides in the field include the compilation and classification of the elemental sounds of Sindhi. Comprehensive measurement of the acoustic features of the language sounds; suitable to be incorporated into the design of a Sindhi ASR system. Understanding of the dialect specific acoustic variation of the elemental sounds of Sindhi. A speech database comprising the voice samples of the native Sindhi speakers. Identification of the language's redundant, substituting and interchanging pairs of sounds. Identification of the language's sounds that can potentially lead to the segmentation and recognition errors for a Sindhi ASR system design. The research achievements of this study create the fundamental building blocks for future work to design a state-of-the-art prototype, which is: gender and environment independent, continuous and conversational ASR system for Sindhi.

ACKNOWLEDGEMENT

The journey of perusing doctoral studies wouldn't be possible without the help and support of many individuals. First and foremost I would like to thank my supervisors Professor Christopher Chatwin and Dr Ruppert Young for their support and guidance for the completion of this study in the Industrial Informatics Research Group, University of Sussex, UK. Without the expert advice, encouragement, interest and dedicated help of my supervisors this research would not have been completed.

My sincere thanks go to all the academic and admin staff at Sindh University Laar campus for their help and support during the data collection process of this study. I would like to pay my special thanks to Professor Qalandar Shah Lakyari, Hayat Keerio, Abdul Khalique Rajar, Faraz Bhughio, Irfan Rind and Intzar Lashari for their help and support at various stages of the data collection process.

I would like to extend my thanks to my sponsor (University of Sindh, Jamshoro) and the entire academic and admin staff at the Institute of Mathematics and Computer Science for their help and support in various administrative aspects.

I am also thankful to all my friends who had been helping me at various stages of this study specially Achar Ahmedani, Soomar Thebo, Qurban Dahani, Mumtaz Bhanbhro, Hameed Chachar, Iqbal Kunbhar Abdullah Burdi and Suleman Kunbhar.

My sincere gratitude goes to my parents and my brothers Aziz Keerio and Ismail Keerio for their financial support when times were tough.

To my parents

&

My wife Ameerzadi Keerio

TABLE OF CONTENTS

LIST OF SYMBOLS

L_T	Total Loss of signal
μ	Mean
σ	Standard Deviation
ξ_k	Data Points
ξ	Data Set
ξ_{kr}	Nominal Reference
α	Threshold Parameter
γ	Scatter Estimate
ξ_{ko}	An Outlier
V_{man}	Count of Manually Labelled Voiced Samples
V_{meth}	Count of Voiced Samples Obtained as the output of an applied method
C	Consonant Phoneme
c	Speed of Sound
f	Frequency
F_0	Fundamental Frequency
F_1	First Formant
F_2	Second Formant
F_3	Third Formant
F_4	Fourth Formant
kHz	Kilo Hertz
V	Vowel Phoneme
λ	Wavelength

LIST OF ABBREVIATIONS

ASR	Automatic Speech Recognizer
CRULP	Centre for Research in Urdu Language Processing
CVC	Consonant Vowel Consonant
DFT	Discrete Fourier Transform
FFT	Fast Fourier Transform
HCI	Human Computer Interaction
IPA	International Phonetic Alphabet
LPC	Linear Predictive Coding
MAD	Median Absolute Deviation
SR	Silence Region
STD	Standard Deviation
STE	Short Time Energy
UR	Unvoiced Region
VCV	Vowel Consonant Vowel
VOT	Voice Onset Time
VR	Voiced Region
ZCR	Zero Crossing Rate

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Chapter 1

Introduction

1.1 Introduction

Speech processing is a significant research area covering a variety of research topics across various disciplines and applications such as designing sophisticated speech production systems to state of the art speech recognition systems. The research and development progress in this area has an active history of more than sixty years; the accurate and efficient design of a speech recognition system is one of the great research challenges in the field of speech processing (Anusuya, 2009; Furui, 2005). The basic goal, which speech production and recognition systems serve, is to enable efficient Human Computer Interaction (HCI); in addition to this many other social and industrial applications take advantage from such HCI enabled systems including the applications of telephony communication, military, banking, education and health care etc. Speech recognizers based on the utilization of instrumental studies such as acoustic-phonetics, still lack the design of accurate and efficient recognition systems which are proven to be a speaker, gender, age, rate and environment independent recognizers. However significant research and development success has been achieved for the English language. This is the result of sixty years on research and development successes; English at present is the language of global trade, finance and politics etc. (Anusuya, 2009; Hongyan, 2007). The speech processing tools and techniques designed for developed languages such as English, Spanish, Mandarin Chinese etc. are at the stage of dealing with problems at an advanced level such as misrecognition due to the poor quality of input signal through a telephone line, lost information while decoding the coded input speech signals, online interactive recognition effected by small bandwidths or discontinuity problems etc. In this respect languages like Sindhi are far behind. Hence this thesis focuses on research and development for the Sindhi language, particularly in the area of acoustic-phonetics that investigates the physical properties of the sounds and provides the fundamental building blocks for the design of an acoustic-phonetics Automatic Speech Recognizer (ASR) system.

Sindhi is ranked as the world's 50th most spoken language according to various independent languages rankings published in 1996 on the web (http://www.davidpbrown.co.uk/help/top-100-languages-by-population.html, http://www.photius.com/rankings/languages2.html). The population figure taken for this ranking is outdated (19 million speakers according to the 1981 census for the Sindh province of Pakistan), whereas the population figures according to the 1998 census for Sindh province published by the government of Pakistan was 30.4 million speakers (http://www.statpak.gov.pk/). The survival of this widely spoken language is in danger due to various socio-economic and socio-linguistic factors, for further details refer to the work of author (Jennifer, 2005) on this subject.

In its territory Sindhi is competing for survival against economically dominant languages such as Urdu and English, both the languages are used as the second language in schools in Sindh province (Pakistan). Other factors such as: social prestige, poor literacy rate of the speakers, lack of research and development resources, mass migration etc. are hastening the native language speaker's shift to other more prestigious or economically dominant languages and cause a reduction in the number of native speakers for the languages which are comparatively less prestigious (David, 2003; Grimes, 2001). Preserving local language is one way to preserve a community's culture because languages grow up with the community and provide the means to express the cultural facets which require a complex local system of words, sentences, and paragraphs (Grimes, 2001) etc.

A language is to be considered in danger if one or all of the following signs are found: when there are few speakers who speak the language, when there are fewer areas of the language use, when its speakers no longer pass it onto the next generation, negative attitude towards their own language, when the language is structurally complex and difficult to learn and when parents do not impart language to their children as the mother tongue etc. (Jennifer, 2005). When communities come into contact with different cultural backgrounds, they begin to influence each other. The economically dominant communities tend to gain speakers from the poorer communities. In the case of Sindhi, potential threats include the declining number of speakers, loss of prestige, increased urbanization of the population where parents are reluctant to impart Sindhi as the mother tongue to the children, difficult to learn due to the large number of phonetic elements, and poor literacy rate of the language's native speakers (according to the 1998 census published by government of Pakistan the literacy rate in the province Sindh is 45.29%: http://www.sindh.gov.pk/aboutsindh.htm).

In the field of linguistics and phonetics quite few researchers have carried out research for Sindhi; in result of this language lacks literature in all the linguistic and phonetic areas particularly in the area of acoustic-phonetics. The notable work in this respect is the acoustic analysis of the voiced implosives of Sindhi carried out by authors Raza et al. (2004). Note that the voice samples for study were collected from one part of the language's *Utraadi* dialect and hence the acoustic variation of the four other dialects has been overgeneralized. The study carried out by Khawaja et al. (2007) is about the segmentation of Sindhi speech based on tracking the dynamics of formants. The study lacks the information about how the segmentation was achieved for the phoneme classes that share some common acoustic-phonetic characteristics such as the liquids, vowels, diphthongs, and glides etc. Note that the study is based on the voice samples of seven speakers however the relevant information about the speaker's age, sex and dialectical background is not clearly mentioned. Other areas in which Sindhi is mostly studied include the articulatory-phonetics, language's grammer, languages orthography and morphological structures refer to the work of authors (Trumpp, 1872; Grierson, 1919; Advani, 1993; Jatoi, 1996; Jennifer, 2006) in this respect. To the best of our knowledge and literature survey no such study has been undertaken sofar that comprehensively presents the acoustic-phonetic characteristics of the elemental sounds of Sindhi covering the five language dialects spoken in Sindh provice of Pakistan.

This study mainly contributes in the area of acoustic-phonetics for Sindhi by providing the fundamental acoustic-phonetic characteristics of its elemental sounds for the research community worldwide, this fills a huge research and development gap for Sindhi. An online web repository based on the collected voice samples of the native Sindhi speaker's across all five dialects spoken in Sindh province (Pakistan) has been made available by this work. Further contributions that this study makes in the area of acoustic-phonetics are, determining the cross dialect acoustic variation for the vowel phonemes, the classification of the language elemental sounds by determining their acoustic-phonetic properties and by using the available articulatory-phonetics knowledge. The acoustic analysis of such sounds whose phonetic status so far is either ambiguous or confused is one of the main contributions of this study. In chapter IV these sounds are introduced as the redundant, substituting, interchanging and confused pairs of sounds; twelve such sounds are analysed that are not yet part of the language alphabet. In chapter four a brief description and classification of these sounds is given; however the final conclusions about the phonemic status of these sounds will be drawn after the acoustic-phonetic analysis of the consonant sounds of Sindhi discussed in detail in chapter eight.

Above all, this study about determining the acoustic-phonetic characteristics of the elemental sounds of Sindhi fills a research and development gap for the language after the partition of the subcontinent. To the best of our knowledge and survey no such research work has been undertaken that comprehensively studies the elemental sounds of Sindhi acoustically. The impetus for doing such research in this respect comes from the fact that detailed knowledge of the sound characteristics of the language-elements has a broad variety of applications — from developing accurate synthetic speech production systems to modelling robust speaker-independent speech recognition systems (Ansarin, 2004; Lee, 1992).

1.2 Research objectives

Based on the problems discussed above and the description of the work that has been done so far; the areas in which more research work is needed are the main objectives of this research study:

- Determining the acoustic-phonetic characteristics of the language elemental sounds.
- A test bed that summarizes the context and dialect specific acoustic variation of the language vowel phonemes acoustically.
- Building an online web repository for the native speaker voice samples

- Compilation of the meaningful languages phonemes
- To emphasize unique linguistic and phonetic characteristics of Sindhi
- Suggestions and recommendations for the simple and easy to learn sound system of the language

1.3 Thesis outline

The introductory chapter details the existing work done for Sindhi and the areas which require more research. It also explains the motivation for the research, its significance and the rational for this research study to be carried out for the Sindhi language.

Chapter 2

The chapter aims to present an overview of the existing work done for Sindh, and its rich historical and cultural background.

Chapter 3

This chapter provides the details of the field study conducted for the purpose of collecting voice samples from the native Sindhi speakers of five dialects in Sindh province (Pakistan). The five language dialects are discussed in detail by outlining the linguistic diversities and similarities among the dialects. The detailed map of the study is given in which details of the sixteen geographically distinct locations travelled to for data collection are indicated. The core research experiments presented in the following chapters of this thesis utilize the voice samples reported in this chapter.

Chapter 4

This chapter introduces the phonetic elements present in the sound inventory of Sindhi. The inventory of meaningful language phonemes is compiled by finding the minimal pair of words for each phonetic unit of Sindhi. In addition to this nine redundant, three interchanging, three substituting and confused pairs of the consonant sounds are discussed; twelve such sounds are also discussed in this chapter, which are not yet part of the language alphabet.

Chapter 5

Chapter five first discusses signal pre-processing techniques essentially used to remove the unwanted signals from the voice samples such as the ambient noise or the signals that are part of the silence. The techniques used to segment the voiced part of the speech from the unvoiced part based on the utilization of the conventional short-time energy based methods are discussed. In a second step, this chapter mainly discusses how the acoustic-phonetic features are determined from the captured voice samples during the field study?

Chapter 6

This chapter discusses in detail the ten monophthong vowel phonemes of Sindhi. It mainly provides the analysis of the cross dialectical acoustic variation of the vowels. The chapter summarizes the main acoustic parameter values such as the mean vowel duration, the fundamental frequencies, the first four formant frequencies, the acoustic vowel plots, and spectrograms etc. The chapter ends by providing the classification of the ten vowel phonemes by incorporating the acoustic-phonetic knowledge in addition to the articulatory-phonetic knowledge.

Chapter 7

Chapter seven mainly discusses the three types of phonetic units of Sindhi having similar acoustic-phonetic characteristics: the monophthong vowels, diphthongs and the glide consonants. This chapter discuss the presence of diphthongs in Sindhi and presents analysis of the identified diphthong phonemes. A comparative analysis of the diphthongs and glides is given; because the two classes of phonemes are considered, the sounds with gradually changing formant pattern. The acoustic differences between these three types' sounds (vowels, diphthongs and glides) are also presented in this chapter.

Chapter 8

In this chapter acoustic analysis of the six consonant classes (stops, implosives, nasals, fricatives, liquids and flaps) is presented. Acoustic measurements are carried out by putting the consonant phonemes in a Vowel-Cconsonant-Vowel (VCV) frame part of the word utterance. The classification of the consonants is achieved by utilizing the obtained acoustic-phonetic knowledge in addition to the available articulatory-phonetic knowledge of the consonant classes. The classification of the four affricate class of consonants is given special attention in this chapter because these sounds are classified in the past part of the stop class as well as the affricate class. Acoustic analysis of the sounds that are not yet part of the language alphabet is given in this chapter. This chapter also provides the details of the two segments present in the stops (closure and release segments); this information helps to segment and identify these sounds while designing a Sindhi recognizer. Various other issues are addressed in this chapter such as the lower F1 values for nasals, and the formant transitions for the sonorant consonants coming into the vowels and vice versa.

Chapter 9

Chapter nine provides the overall thesis conclusions and the future work directions in addition to a detailed overview of the contributions this work makes to the field.

1.4 Publications contributed

Ayaz Keerio**,** Bhargav Kumar Mitra, Philip Birch, Rupert Young, Chris Chatwin, "On Preprocessing of Speech Signals", In the proceedings of World Academy of Science, Engineering and Technology, volume 47, pp.317-323, November 2008.

Ayaz Keerio, Bhargav Kumar Mitra, Philip Birch, Rupert Young, Chris Chatwin, "On Preprocessing of Speech Signals," International Journal of Signal Processing, vol. 5, no. 3, pp. 216-222, February 2009.

Ayaz Keerio, Muhammad Zeeshan Patoli, Bhargav Kumar Mitra, Chris Chatwin, Rupert Young, Philip Birch "Acoustic Analysis of Diphthongs in Sindhi", Grassroots Biannual Research Journal, Pakistan Studies Centre University of Sindh Jamshoro, Vol. 41(1), June-2010.

Aya**z** Keerio, Chris Chatwin, Rupert Young, "The Phonetic Transcription of Sindhi Language", Poster paper presentation, 2007.

CHAPTER 2

BACKGROUND

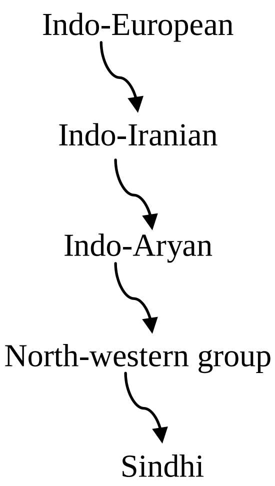
Chapter 2

Background

2.1 Introduction

According to geology and archaeology experts “Sindh, is mainly a sea-born land”, due to natural disasters such as earthquakes and floods etc. this arid-desert land surfaced by pushing the sea level back (Advani, 1993; Pithawalla, 1935). In the modern age this sea born land is known as the ‘Sindh’ province of Pakistan with its rich cultural and historical background. The discovery of Moen-Jo-Daro (Mound of the dead) in 1920s revealed that a highly developed urbanized culture existed in Sindh between 2600 and 1900 B.C known as the Indus valley civilization, which is as old as the ancient Egyptian, Mesopotamian and the Chinese civilizations (Advani, 1993). There were approximately 2000 settlements associated with the Indus valley civilization; Mohenjo-Daro (Mound of the dead) is named the largest one, populated between 35000-40000 people approximately (Corbishley, 1989; William, 2006). Figure 2.1(a) and (b) show the excavated city of Mohenjo-Daro, the city is spread over the area of 1.5 square kilometres (culture and tourism department government of Sindh (Pakistan): http://www.sindh.gov.pk/).

The early history and culture of Sindh is a mystery and almost unknown. The ancient script found in the Mohenjo-Daro site and from the other settlements of the Indus valley civilization is not yet deciphered. The people of the modern age in Sindh province (Pakistan) have somehow preserved and continued their life with a very similar cultural background. This can be observed from the eight artefacts discovered from the excavated Mohenjo-Daro settlement and the life of Sindhi residents in the modern age. Shown in figure 2.2(a) is the picture of a bull, an artefact found from the Mohenjo-Daro site. Artefacts highlight the fact that the bull was an important domestic animal around 1900-2600 B.C. Figure 2.2(b) is a picture of a bull an important domestic animal in the modern age as well. Shown in the figure 2.3(a) is an artefact of a bull cart generally used for loading, figure 2.3(b) is the bull cart being used in modern age in the rural areas of Sindh province (Pakistan). In connection with this the modern Sindhi language evolved over the period of 2500 years. This evolution period counts many foreign invasions, during these rules Sindhi has gathered vocabulary and linguistic features of the languages of its foreign ruler's i.e. the invasion of Achamenians, Persians, Mouryans, Scythians and Greeks, including Alexander the Great 329-324 B.C (Jennifer, 2006). Arab rule began from 711 A.D, during this period Sindhi remained under the heavy influence of Persian until the British rule started in 1843 A.D. During and after the British rule Sindhi was heavily influenced by English; however the impact of English is considered as secondary compared to the Persian, due to its similar script and writing style which is right to left for both the languages (Jennifer, 2006). The peoples of Sindh in communication with foreigners have set strong multilingual foundations for Sindhi; however the family hierarchy for Sindhi, most of the researchers are agree with is, "Sindhi belongs to a North-western sub-group of Indo-Aryan, under the Indo-Iranian branch of Indo-European" (Jatoi, 1996; Jennifer, 2006):



Sindhi is one among the few world languages that uses two scripts such as Hindi and Urdu languages which are considered one language with two scripts (http://www.indianchild.com/hindi.htm). The modern Sindhi uses the *Arabic Nask script* (prominent in the Muslim Sindhi speaker areas) and *Devanagri script* (prominent in Hindu Sindhi speaker areas). However the oral style of speaking among the Hindu and Muslim speakers is similar (Jennifer, 2006; Khawaja, 2004).

There is a huge research and development gap for Sindhi, due to the mass migration after the partition of the sub-continent. The notable pre-partition research work can be witnessed as Sir George A. Grierson’s work titled as linguistic survey of India between 1894-1928, a section in this survey describes the Sindhi language history and geography in detail (Grierson, 1919), Ernest Trumpp’s work Grammar of the Sindhi Language (Trumpp, 1872) and the work of Sir John H. Marshall titled: Mohenjo-Daro and the Indus Civilization, provides a solid linguistic and historical background for Sindhi (Marshall, 1931). It is hard to find such a broad scale comprehensive research study in the field of linguistics and phonetics to-date. In this chapter we tried to explore the historical and cultural background of Sindhi along with discussion of the areas in which it lacks the research work specifically in the area of acoustic-phonetics.

Figure 2.1(a) Mohenjo-Daro (Mound of dead) discovered in 1922.

Figure 2.1(b) Aerial photograph of Mohenjo-Daro (Mound of dead) city spread over 1.5 square kilometre area source: (http://images.google.com/)

(a) Ancient bull picture found in Moen Jo Daro

(b) bull a domestic animal in Sindh province

Figure 2.2 Ancient bull picture and the recent bull picture source: culture and tourism department government of Sindh, Pakistan (www.sindh.gov.pk)

(a) Ancient bull cart model

(b) bull cart in use till date

Figure 2.3(a) a bull cart artefact found in Mohenjo-Daro and (b) the bull is being used till to-date source: (www.images.google.com)

2.2 Standard Sindhi and dialects

There are six dialects of Sindhi out of which five are spoken in Sindh province of Pakistan are: *Middle, Lari, Lasi, Thareli, Utradi* and the sixth dialect is (Kachhi) spoken in Rann of Kutch located in the south western Indian state Gujarat. The *Middle* dialect is geographically important because it is the only dialect that shares the geographical boundary with the rest four dialects of Sindhi spoken in Sindh province of Pakistan. Among the five dialects the standard Sindhi is referred to the language spoken by the speakers of the *Middle* dialect. Officially Sindhi is written and spoken grammatically according to the Middle dialect. The only matter which can be found written and published according to the other dialects is either in poetic form or in other literary forms such as stories etc. This is why the Middle dialect is more influential and dominant than the other dialects. This is the only dialect whose geographical boundaries are extended; the geographical importance of this dialect is that it shares a boundary with all the other dialects of Sindhi. This makes the Middle dialect understandable to the speakers of other dialects and therefore it was given the status of the standard Sindhi (*Middle*) dialect. The geographical boundaries discussed in early published literature for the six dialects of Sindhi refer to the work (Grierson, 1919; Advani, 1993; Jatoi, 1996); at present they show great change in their geography. This change involve varous factors i.e. the socio-economic factors, cross-dialect, cross-language marriages, and mass migration etc. One of the major factors involved is the partition of the sub-continent (Jennifer, 2006), and the division of the country (Pakistan), into four provinces i.e. Sindh, Punjab, Khyber Pakhtunkhwa, and Balochistan. Some parts of the north Sindh province are in comunication with the southern part of Punjab and some western parts of Sindh are in communication with eastern parts of Balochistan. This has a huge impact on the language of peoples communicating with each other. For example the dialect spoken in the Jacobabad, Ghotki, Shikarpur and Larkana districts is slightly different from the *Utradi* dialect; geographically these districts are part of the *Utradi* dialect. The dialect spoken in Sanghar district is associated with the *Middle* dialect; however the dialect of this region is slightly different from the *Middle*. There also exist some community based dialects of Sindhi which are never been studied. For example Sindhi of Jats (name of tribe note here that 'Jat', tribe is different from the 'Jaats', in Punjab), and Sindhi of Dhatis (name of tribe) etc. The study of Sindhi geography is out of the scope of this study; however some findings regarding dialect specific geography is presented in appendix A, which is observed during the data collection process discussed in detail in chapter III.

2.3 Rural and Urban Sindhi

There exists the difference between the rural and urban spoken Sindhi. The rural spoken Sindhi is further divided between the educated and uneducated speakers. The urban Sindhi and the Sindhi of educated people in rural areas can be referred to as the refined modern form of Sindhi, in which the speakers unintentionally add the words of other languages such as Urdu, English and Arabic languages. The uniformity in the language spoken by the literate peoples can be observed across all the dialects; because they speak the taught version of the language which is standard (*Middle* dialect) Sindhi; whereas the Sindhi of uneducated peoples of rural areas speak a non-refined version of language in their natural style.

2.4 Acoustic phonetics

Acoustic phonetics refers to the instrumental study of the speech sounds using the computer generated spectrograms, fundamental frequency, amplitude and durational measurements etc. The continued research and development efforts in the past six decades along with substantial technological advancement in the area of speech production and recognition systems still lack the design of a state of the art, efficient and accurate speech recognizer (Furui, 2005). The level of difficulty to design such systems is very high due to the variable nature of the speech even if the same speaker repeats the utterance result in different speech signals. The major factors that cause variability in speech are the environment change, speaker change and context variation (Anusuya, 2009). Change in the signal caused by any of the above mentioned factors heavily effect the subsequent process of segmentation, and feature extraction etc. There are three basic approaches for the design of a state of the art speech recognizer which are: *Acoustic Phonetic Approach*, *Pattern Recognition Approach*, and *Artificial Intelligence Approach* (Anusuya, 2009; Furui, 2005). The acoustic-phonetic approach based recognizer follow the sequence of steps beginning with spectral measurements known as the feature extraction process followed by segmentation and labelling etc (Anusuya, 2009; Furui, 2005). In other words the basic building blocks involved in the acoustic-phonetic approach based speech recognizer is to segment, locate and label the distinguishable phonetic elements present in the speech signal (Espy-Wilson, 1987; Espy-Wilson, 1994; Hemdal, 1967) and the last step involved in this approach is the modelling and validation of the decoded words from the cluster of labelled phoneme(s). The phonetic elements of a particular language can be classified and characterized by determining their acoustic-phonetic properties referred to as the meaningful phonemes of a language. The sequence of these phonetic units in speech constitutes word utterances of a language. The acoustic characteristics of these phonetic units are highly variable and difficult to determine if produced in combination for a recognizer system. Speech recognizers are therefore classified according to the level of difficulty a system can handle on the basis of the utterances that they can recognize i.e. spontaneous speech recognizer, continuous speech recognizer, connected word recognizer and isolated word recognizer (Anusuya, 2009; Furui, 2005).

A Spontaneous Speech recognizer is thought to be a recognizer with the ability of dealing with the natural sounding speech, not the rehearsed one (Anusuya, 2009). Miller and Weinert (1998) summarized five main properties of spontaneous speech, for details refer to the work of (Miller and Weinert, 1998). Continuous speech recognizer’s operate on naturally spoken language in which word utterances are connected together. Recognizers capable of determining the content of such naturally spoken language are considered one of the difficult systems to be designed (Anusuya, 2009; Furui, 2005). Because such systems need to handle a variety of difficult problems such as determining the word utterance boundaries, dealing with the rate of the speaker's speech, co-articulation the production of one phoneme may affect the articulation of predecessor or successor phoneme(s) (Cynthia, 2009; Levinson, 1990). Isolated word recognizer systems operate on the word utterances clearly marked with a silence or deliberate pause between the utterances of two words by the speaker. These systems are easy to design compared to others because the word utterance segmentation is controlled through the speaker while communicating the speech signals to the recognizer in the start and stop manner (Furui, 2005). Connected word recognizers are quite similar to the isolated word recognizers, but the connected word systems operate on a string of words with a minimized pause between the word utterances (Anusuya, 2009; Dobler, 1992).

Sufficient published literature is available on the subject of acoustic-phonetics for the study of physical properties of sounds in various languages. The vowel phonemes are generally studied more acoustically compared to the consonant phonemes because the former show steady-state pattern in speech (Kent, 2002; Moore, 2003). English language or foreign- accented English is the most studied language acoustically. For example the work of Peterson and Barney's (1952) study of English language vowels, study of Mandarin-accented English vowels (Chen, 2001), Indian-accented English (Maxwell, 2009), Singapore-accented English (Deterding, 2003), Japanese-accented English vowels (Kewley-Porr, 1996) etc. Other languages rigorously studied acoustically are the study of French (Ouni, 2003), Shanghai Chinese (Chen, 2008), Turkish language (Gordon, 2006), Spanish language (Borzone de Manrique, 1979; Borzone de Manrique, 1981; Gurlekian, 1985), Japanese language (Homma, 1973), and the acoustic analysis of modern Hebrew language (Aronson, 1996). Like the English the elemental sounds of the Spanish language have been rigorously studied and compared acoustically to that of English i.e. the comparative study of English and Spanish vowels by (Bradlow, 1995). Other comparative studies include a cross gender and cultural acoustic study of the Japanese and American speaker's carried out in order to explore the pitch differences based on the speaker's familiarity vs. non-familiarity (Yuasa, 2001). The study about the Mandarin Chinese speaking mothers carried out by (Liu, 2001), in which the author tried to investigate the changing acoustic behaviours of speech when a mother speaks to her infant and when she speaks with adults (Liu, 2001). A socio-phonetic study by (Lewis, 2002), in this study author tried to investigate the pitch changes caused due to the speaker variability options such as gender and status of young English female speakers (Lewis, 2002).

The languages use similar script or have a similar writing style as used by Sindhi, and their elemental sounds have been studied acoustically are: the study of modern Persian language vowels (Ansarin, 2004), acoustic analysis of Urdu and Siraiki language at the Centre for Research in Urdu Language Processing (CRULP), study carried out by Amna(2003) and Sarwar(2004). There is no such comprehensive study reported in published literature so far on the subject of acoustic-phonetics of the element sounds of Sindhi or the comparative linguistic and phonetic study that incorporates such socio-economic and socio-phonetic factors. Work reported in the literature to date on the language has been either on articulatory-phonetics or the language's writing system, grammar, dialects and history etc.

Linguistically one of the early published literatures available is Trumpp's (1872) work "the grammar of Sindhi", another study related to the grammar and phonology of Sindhi is Khubchandani's work (Khubchandani, 1961). Sir Grierson's work (Grierson, 1919) linguistic survey of India in which a section describes history and dialectical geography of Sindhi is one of the thorough linguistic studies of Sindhi. Paroo Nihalani's work provides detailed description about the production process of the stop consonants in Sindhi such as the voiced, unvoiced, aspirated and breathy voiced stops of Sindhi (Nihalani, 1974; Nihalani, 1975). Nihalani's study is one of the early studies that talks about the presence of the purely glottalic ingressive implosives in Sindhi, unlike the implosives present in the Hausa (Nihalani, 1986). His research work clearly rejects the idea that there are no true implosives that involve suction of air while producing them, " ... implosives in Sindhi involve ingressive airflow, unlike the implosives in Hausa. The immediate consequence of this fact is that the proposal that there are no true implosives, i.e., those that involve suction, must be rejected", (Nihalani, 1986).

The work of Raza *et. al.*'s (2004) has been on articulatory phonetics of the language's consonantal inventory and the work presents the acoustic analysis of the implosives only; the authors have thoroughly compared their findings with those of the Urdu language. Moreover, the study is based on the voice samples collected from the speakers of only one district (Jacobabad) in Sindh (Pakistan) this region is a small part of the *Utradi* dialect, and hence does not include the diversity of the four distinct dialects of Sindhi in the Sindh province; the only other distinguishable dialect is spoken in the Rann of Kutch—a region in the western state of Gujarat, India (Jennifer, 2006). Jatoi's work, published and written in Sindhi (Jatoi, 1996) explains three major linguistic and phonetic areas of Sindhi: the articulation of the vowel- and consonant-sounds of the language, the language's morphology, and the history of the language. Author (Jennifer, 2006) mainly discussed the language morphology and history in detail, however a section of the article discusses the language's orthography and phonology. Authors Khawaja *et. al.* (2007) in their work proposed a formant based speech segmentation method for Sindhi (Khawaja, 2007). The method is based on tracking the changing dynamics of the resonant frequencies in speech signals. The segmentation is than achieved on the basis of the first three formants. First two formants are adequate to identify the monophthong vowels of Sindhi (the vowel system of Sindhi is discussed in chapter VI in detail); however the segmentation of the consonantal parts in speech on the basis of formants could potentially lead to a false segmentation. In the proposed method it is not clearly mentioned how the voiced consonantal phonemes are differentiated from the vowels, diphthongs and glide phonemes. In case of the formant transitions such as vowel-to-glide and vice versa, vowel-to-stops and vice versa, vowel-to-nasals and vice versa it is not clearly mentioned that how segmentation is achieved. The consonant phoneme /r/ of Sindhi maintain the same formant motion coming into the /r/ sound from the previous sonorant (vowel) sound to the following sonorant sound, in such case the /r/ consonant might be incorrectly segmented as a vowel phoneme. Another factor is the nasal consonants; all the nasals of Sindhi are voiced phonemes with a low first formant (F1) value, such vowels with low F1 values i.e. the high back monophthong vowels may be segmented incorrectly as nasals and vice versa. Chapter VI and VIII in this study discuss in detail such issues related to the vowels, and consonants for the design of a Sindhi speech recognizer.

CHAPTER 3

FIELD STUDY

Chapter 3

Field Study

3.1 Introduction

In this chapter the overall setup of the data collection process is discussed. A comprehensive field study was conducted for the collection of voice samples from native Sindhi speakers living in the Sindh province (Pakistan). There are six regional dialects of Sindhi named as: (i) *Vicholi* ‘Middle’, this dialect is recognized as the standard dialect of Sindhi, (ii) *Thareli,* this dialect is spoken in Thar (a deserted region), (iii) *Lasi,* this dialect is spoken in the Kohistan and Lasbelo regions, (iv) *Lari* this dialect is spoken in lower Sindh (coastal) regions, (v) *Utradi* this dialect is spoken in upper Sindh regions, (vi) *Kachhi,* this dialect is spoken in Rann of Kachh region located in the south western Indian state Gujarat. In this field study, voice samples were collected from the native Sindhi speakers living in five different regions of the Sindh province (Pakistan); the only dialect which is not part of this study is Kachhi spoken in Gujarat, which is a state of India. The core research experiments reported in this work are mainly based on the voice samples collected during the field study. The speech database of this study consists of the voice samples of seventy five male native Sindhi speakers; dividing fifteen male speakers for each dialect spoken in Sindh province of Pakistan. In addition to this the geographical survey of five language dialects was carried out with this study, the details of the survey are given in appendix A.

3.2 Field study

The geographical region selected for voice sample recording is the Sindh province (Pakistan); this is because among the six language dialects five dialects exist in this region (Jennifer, 2006). Seventy five adult male speakers participated in the study, with fifteen male speakers for each dialect region. The selection of the speaker for voice recording was made after an extensive screening procedure. The screening procedure was as follows: (i) participant must be born and brought up in Sindh province (Pakistan), (ii) participant must be living for the last 15 years in that particular region (regional dialect) for which the intended recording is to be made (iii) participant must be Sindhi literate (can read and write Sindhi), (iv) participant must be between 20 to 50 years of age, and (v) participant does not have any speaking or hearing disorders. The selection of each speaker was made after conducting an informal interview with the speaker. The informal interview with speaker served two major goals, first to judge whether the speaker meets all the above mentioned criteria and second was to judge the speaker's pronunciation intelligibility of the target dialect for the recording of the voice sample.

3.3 Recording environment

Two recording environments were observed during the data collection process: (i) open and (ii) closed. Recordings made while sitting with a participant alone in open-air places called "*Jhupri's* or *Otaks*" in the rural areas of the province are treated as open environment recordings, whereas recordings made while sitting with a participant inside a concrete building are treated as closed environment recordings. Speech samples were recorded for three major classes of the language sounds: *vowels*, *consonants* and *diphthongs*. Voice samples for vowels and consonants were recorded in two ways: (i) the isolated utterance of the target sound (either vowel or consonant), and (ii) the word utterances (the target sound embedded in the word medial position). For diphthongs only the word utterance based voice samples were recorded. The word utterances of a target sound help to analyse the impact of leading and following phones in the utterance. In this study words chosen for the voice recording contain CVCv (Consonant Vowel Consonant) phoneme sequence at word medial position, the small v in this respect represent the occurrence of a short vowel at the word final position if the word utterance ends at CVCv phoneme sequence. The selected list of words for each class of sounds was printed on a separate sheet. Each speaker was given a printed sheet for speaking in his natural style of speaking. At least ten instances of each speaker's voice sample were made for the purpose of enabling the best utterance selection for subsequent offline processing. A maximum of four and minimum of three speakers' voice samples were recorded at one recording place of the target dialect during the field study. In this way at least four distant places within one target dialect were visited for recording, except for the *Lasi* and the *Thareli* dialects, for which voice samples were recorded from two distant places. On average the distance between two recording places was thirty kilometres. Figure 3.1 show the detailed map of the places travelled to for the recording sessions during the field study conducted in the Sindh province (Pakistan).

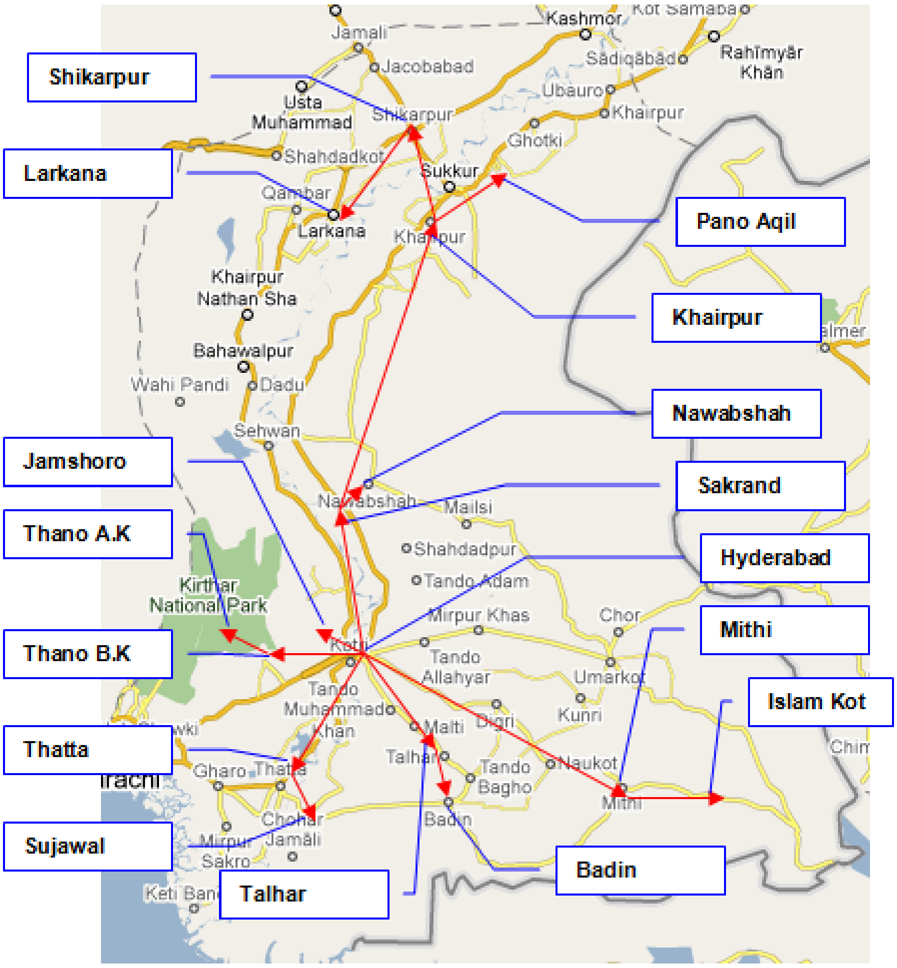


Figure 3.1 Map of the places travelled to for the recording of voice samples in the Sindh province (Pakistan).

3.4 Recording details

A portable solid state digital audio recorder *Marantz PMD660* shown in figure 3.2 was used with an external broadcast quality condenser microphone the *Rode NTG-2* shown in figure 3.3, for the voice samples recording. The speech samples were recorded as single channel (mono) at the sampling rate of 48 kHz; the output of the channel was obtained as 16 bit words. The recorded voice samples were stored as uncompressed

.wav files for subsequent offline processing. On average the mouth to microphone distance was maintained at 4 inches for the voice samples recording.

Figure 3.2 Marantz PMD660, portable digital audio recorder.

Figure 3.3 Rode NTG-2, condenser microphone.

3.5 Details of the distance travelled to various recording places

The first region selected for the voice sample recordings was the standard dialect of Sindhi called '*Middle*'. Hyderabad district of the *Middle* dialect was set as the origin to calculate the approximate road distance travelled to cover all the other recording places for the five language dialects. The approximate road distance to various locations from the origin in kilometres, and the number of speaker voice samples recorded at one recording place are given in Table 3.1

Table 3.1 Approximate road distance travelled to various locations for recording from origin (Hyderabad)

Dialect	Recording point	Distance from Origin - km	Number of speaker's participated for recording
Middle	Hyderabad	--	04
	Jamshoro	18	04
	Sakrand	107	03
	Nawabshah	123	04
Utradi	Khairpur	305	04
	Pano Aqil	357	04
	Shikarpur	376	03
	Larkana	470	04
Lari	Badin	110	04
	Talhar	87	04
	Thatta	102	04
	Sujawal	125	03
Thareli	Mithi	150	07
	Islamkot	240	08
Lasi	Thano Bula Khan	72	08
	Thano Ahmed Khan	85	07

3.6 Voice sample recording for vowels

For the recording of ten monophthong vowels a printed sheet containing the isolated vowel utterances and the word utterances (embedded with a target vowel) was given to each speaker and asked to speak out the isolated vowel utterances first; followed by the word utterances of the CVCv phoneme sequence. The vowel sound surrounded by consonants is the target analysis sound. Table 3.2 below shows the list of isolated vowel utterances and the list of word utterances for the recording of vowel sounds.

Table 3.2 List of the word utterances for vowels (Jatoi, 1996).

S.No	Isolated vowel utterance		Word in Sindhi	Roman Sindhi
1	اِي	iy	سِيرَ	siyra
2	اِ	ih	سِرَ	Sira
3	اي	eh	سيرَ	Sera
4	اَي	ay	سَئيرَ	Sayra
5	اَ	ah	سَرَ	Sara
6	آ	aa	سارَ	Saara
7	اَو	ao	چَوندئو	Chawandao
8	او	o	چَوندو	Chawando
9	اُ	uh	سُرَ	Sura
10	اُو	uo	سُورَ	suora

3.7 Voice sample recording for consonants

Voice samples for consonant sounds were recorded in two ways: (i) the word utterances (a list of fifty words compiled by finding the minimal pair words (discussed in chapter IV) for each consonant sound) and (ii) the utterances of 470 diaphone mono-syllabic words. For the second category of the voice samples for consonant phonmes the length of each word was restricted to two phones (called diaphone-syllabic words) of the structure CV (a consonant followed by a vowel). It was unavoidable to create some meaningless word utterances to meet the two phone word utterance restriction. However some meaningful words in this CV structure were also present i.e the words [ɓə] ٻَ '*two*' , [ɗe] ڏي '*give*', [ʧʱɑ] ڇا '*what*', etc. Our focus here was to reduce the utterance duration (which helps to reduce the overall offline processing cost) and to achieve all possible vowel and consonant combinations for the acoustic analysis of the elemental sounds of Sindhi. Another linguistic property of these diaphone mono-syllabic words is that the composition of Sindhi words can be achived by adding these diaphone syllables. Table 3.3 shows the list of the word utterances for the consonant sounds of Sindhi for recording. Table 3.4 shows the list of diaphone mono-syllabic word utterances of Sindhi.

Table 3.3 List of the word utterances for the consonants

S.No	Word in Sindhi	Roman Sindhi	S.No	Word in Sindhi	Roman Sindhi
1	سَبق	Sabaku	26	غاليچو	galicho
2	رَبِ	raba	27	ڪل	kala
3	سِپَ	Sipa	28	ساره	sarah
4	سڀُ	Sabhu	29	زاري	zari
5	سَتَ	sata	30	ماري	mari
6	سٺُ	sathu	31	پڙُ	paru
7	سٽ	Sata	32	پڙهُ	parhu
8	هَٿ	hatha	33	سَگهَ	gharu
9	فِڪِرُ	fikiru	34	سارو	saro
10	در	dar	35	ماڻي	maney
11	ٻڦ	bafa	36	ماڻهين	manhen
12	آگَ	aaga	37	ڪاش	kashu
13	وَڏو	vado	38	سِڪَ	sika
14	اکر	akhar	39	هرڻ	haran
15	رَڱ	ranga	40	ڪُجهه	Kujhu
16	جَڄ	janga	41	ڪاڻُ	kanu
17	ساڌگي	sadgi	42	ڪَلهه	kalha
18	سڏ	sada	43	سَمَ	sama
19	ڪَڍ	kadha	44	سُمهي	sumhey
20	ٻڍو	bado	45	سنا	sana
21	سَجائي	Sajai	46	سنها	sanha
22	سڃَر	sajar	47	يارُ	yaru
23	سَچُ	Sachu	48	وارُ	waru
24	ڪَڇُ	kachhu	49	سُڌ	sudha
25	خَچر	Khachar	50	ڏَڳو	dhago

Table 3.4 List of diaphone mono-syllabic units of Sindhi

/bə/ بَ	/bɑ/ با	/bɪ/ بِ	/bi/ بِي	/be/ بي	/bʊ/ بُ	/bu/ بوُ	/bɛ/ بَي	/bɔ/ بَو	/bo/ بو
/ɓə/ ٻَ	/ɓɑ/ ٻا	/ɓɪ/ ٻِ	/ɓi/ ٻِي	/ɓe/ ٻي	/ɓʊ/ ٻُ	/ɓu/ ٻوُ	/ɓɛ/ ٻَي	/ɓɔ/ ٻَو	/ɓo/ ٻو
/pə/ پَ	/pɑ/ پا	/pɪ/ پِ	/pi/ پِي	/pe/ پي	/pʊ/ پُ	/pu/ پُو	/pɛ/ پَي	/pɔ/ پَو	/po/ پو
/b^{h}ə/ ڀَ	/b^{h}ɑ/ ڀا	/b^{h}ɪ/ ڀِ	/b^{h}i/ ڀِي	/b^{h}e/ ڀي	/b^{h}ʊ/ ڀُ	/b^{h}u/ ڀُو	/b^{h}ɛ/ ڀَي	/b^{h}ɔ/ ڀَو	/b^{h}o/ ڀو
/tə/ تَ	/tɑ/ تا	/tɪ/ تِ	/ti/ تِي	/te/ تي	/tʊ / تُ	/tu/ تُو	/tɛ/ تَي	/tɔ/ تَو	/to/ تو
/ʈ^{h}ə/ ٺَ	/ʈ^{h}ɑ/ ٺا	/ʈ^{h}ɪ/ ٺِ	/ʈ^{h}i/ ٺِي	/ʈ^{h}e/ ٺي	/ʈ^{h}ʊ/ ٺُ	/ʈ^{h}u/ ٺُو	/ʈ^{h}ɛ/ ٺَي	/ʈ^{h}ɔ/ ٺَو	/ʈ^{h}o/ ٺو
/ʈə/ ٽَ	/ʈɑ/ ٽا	/ʈɪ/ ٽِ	/ʈi/ ٽِي	/ʈe/ ٽي	/ʈʊ / ٽُ	/ʈu/ ٽُو	/ʈɛ/ ٽَي	/ʈɔ/ ٽَو	/ʈo/ ٽو
/t^{h}ə/ ٿَ	/t^{h}ɑ/ ٿا	/t^{h}ɪ/ ٿِ	/t^{h}i/ ٿِي	/t^{h}e/ ٿي	/t^{h}ʊ/ ٿُ	/t^{h} u/ ٿُو	/t^{h}ɛ/ ٿَي	/t^{h}ɔ/ ٿَو	/t^{h}o/ ٿو
/fə/ فَ	/fɑ/ فا	/fɪ/ فِ	/fi/ فِي	/fe/ في	/fʊ/ فُ	/fu/ فُو	/fɛ/ فَي	/fɔ/ فَو	/fo/ فو
/nə/ نَ	/nɑ/ نا	/nɪ/ نِ	/ni/ نِي	/ne/ ني	/nʊ/ نُ	/nu/ نُو	/nɛ/ نَي	/nɔ/ نَو	/no/ نو
/p^{h}ə/ڦَ	/p^{h}ɑ/ ڦا	/p^{h}ɪ/ ڦِ	/p^{h}i/ ڦِي	/p^{h}e/ڦي	/p^{h}ʊ/ ڦُ	/p^{h}u/ ڦُو	/p^{h}ɛ/ڦَي	/p^{h}ɔ/ ڦَو	/p^{h}o/ڦو
/gə/ گَ	/gɑ/ گا	/gɪ/ گِ	/gi/ گِي	/ge/ گي	/gʊ/ گُ	/gu/ گُو	/gɛ/ گَي	/gɔ/ گَو	/go/ گو
/ɠə/ ڳَ	/ɠɑ/ ڳا	/ɠɪ/ ڳِ	/ɠi/ ڳِي	/ɠe/ ڳي	/ɠʊ/ ڳُ	/ɠu/ ڳُو	/ɠɛ/ ڳَي	/ɠɔ/ ڳَو	/ɠo/ ڳو
/k^{h}ə/کَ	/k^{h}ɑ/ کا	/k^{h}ɪ / کِ	/k^{h}i/کِي	/k^{h}e/کي	/k^{h}ʊ/کُ	/k^{h}u/گُو	/k^{h}ɛ/گَي	/k^{h}ɔ/ گو	/k^{h}o/کو
/ŋə/ ڱَ	/ŋɑ/ ڱا	/ŋɪ/ ڱِ	/ŋi/ ڱِي	/ŋe/ ڱي	/ŋʊ/ ڱُ	/ŋu/ ڱُو	/ŋɛ/ ڱَي	/ŋɔ/ ڱَو	/ŋo/ ڱو
/ɲə/ ڃَ	/ɲɑ/ ڃا	/ɲɪ/ ڃِ	/ɲi/ ڃِي	/ɲe/ ڃي	/ɲʊ/ ڃُ	/ɲu/ ڃُو	/ɲɛ/ ڃَي	/ɲɔ/ ڃَو	/ɲo/ ڃو
/də/ دَ	/dɑ/ دا	/dɪ/ دِ	/di/ دِي	/de/ دي	/dʊ / دُ	/du/ دُو	/dɛ/ دَي	/dɔ/ دَو	/do/ دو
/ɗə/ ڏَ	/ɗɑ/ ڏا	/ɗɪ/ ڏِ	/ɗi/ ڏِي	/ɗe/ ڏي	/ɗʊ / ڏُ	/ɗu/ ڏُو	/ɗɛ/ ڏَي	/ɗɔ/ ڏَو	/ɗo/ ڏو
/ɖ^{h}ə/ ڍَ	/ɖ^{h}ɑ/ ڍا	/ɖ^{h}ɪ/ ڍِ	/ɖ^{h}i/ڍِي	/ɖ^{h}e/ڍي	/ɖ^{h}ʊ/ ڍُ	/ɖ^{h}u/ڍُو	/ɖ^{h}ɛ/ڍَي	/ɖ^{h}ɔ/ ڍَو	/ɖ^{h}o/ڍو
/ɖə/ ڊَ	/ɖɑ/ ڊا	/ɖɪ/ ڊِ	/ɖi/ ڊِي	/ɖe/ ڊي	/ɖʊ / ڊُ	/ɖu/ ڊُو	/ɖɛ/ ڊَي	/ɖɔ/ ڊَو	/ɖo/ ڊو
/ʤə/ جَ	/ʤɑ/ جا	/ʤɪ/ جِ	/ʤi/جِي	/ʤe/جي	/ʤʊ/ جُ	/ʤu/جُو	/ʤɛ/جَي	/ʤɔ/جَو	/ʤo/جو
/ʄə/ ڄَ	/ʄɑ/ ڄا	/ʄɪ/ ڄِ	/ʄi/ ڄِي	/ʄe/ ڄي	/ʄʊ/ ڄُ	/ʄu/ ڄُو	/ʄɛ/ ڄَي	/ʄɔ/ ڄَو	/ʄo/ ڄو
/ʧə/ چَ	/ʧɑ/ چا	/ʧɪ/ چِ	/ʧi/چِي	/ʧe/ چي	/ʧʊ/ چُ	/ʧu/ چُو	/ʧɛ/ چَي	/ʧɔ/ چَو	/ʧo/ چو
/ʧ^{h}ə/ ڇَ	/ʧ^{h}ɑ/ ڇا	/ʧ^{h}ɪ/ ڇِ	/ʧ^{h}i/ڇِي	/ʧ^{h}e/ڇي	/ʧ^{h}ʊ/ ڇُ	/ʧ^{h}u/ڇُو	/ʧ^{h}ɛ/ڇَي	/ʧ^{h}ɔ/ ڇَو	/ʧ^{h}o/ و

-Continued on next page -

/xə/ خَ	/xɑ/ خا	/xɪ/ خِ	/xi/ خِي	/xe/ خي	/xʊ/ خُ	/xu/ خُو	/xɛ/ خَي	/xɔ/ خَو	/xo/ خو
/ɣə/ غَ	/ɣɑ/ غا	/ɣɪ/ غِ	/ɣi/ غِي	/ɣe/غي	/ɣʊ/ غُ	/ɣu/ غُو	/ɣɛ/ غَي	/ɣɔ/ غَو	/ɣo/ غو
/rə/ رَ	/rɑ/ را	/rɪ / رِ	/ri/ رِي	/re/ ري	/rʊ/ رُ	/ru/ رُو	/rɛ/ رَي	/rɔ/ رَو	/ro/ رو
/kə/كَ	/kɑ/ كا	/kɪ/ كِ	/ki/كِي	/ke/كي	/kʊ/ كُ	/ku/كُو	/kɛ/كَي	/kɔ/كَو	/ko/كو
/zə/ زَ	/zɑ/ زا	/zɪ/ زِ	/zi/ زِي	/ze/ زي	/zʊ/ زُ	/zu/ زُو	/zɛ/ زَي	/zɔ/ زَو	/zo/ زو
/mə/ مَ	/mɑ/ ما	/mɪ / مِ	/mi/مِي	/me/مي	/mʊ/ مُ	/mu/مُو	/mɛ/ ميَ	/mɔ/ مَو	/mo/مو
/ɽə/ ڙَ	/ɽɑ/ ڙا	/ɽɪ/ ڙِ	/ɽi/ ڙِي	/ɽe/ ڙي	/ɽʊ / ڙُ	/ɽu/ ڙُو	/ɽɛ/ ڙَي	/ɽɔ/ ڙَو	/ɽo/ ڙو
/ɽʰə/ ڙَھ	/ɽʰɑ/ڙھا	/ɽʰɪ/ ڙھِ	/ɽʰi/ڙھِي	/ɽʰe/ڙھي	/ɽʰʊ/ ڙُھ	/ɽʰu/ڙُھو	/ɽʰɛ/ ڙَھ	/ɽʰɔ/ڙَھو	/ɽʰo/ڙھو
/gʰə / گھَ	/gʰɑ/ گھا	/gʰɪ / گھِ	/gʰi / گھِي	/gʰe / گھي	/gʰʊ/ گُھ	/gʰu / گُھو	/gʰɛ / گھَي	/gʰɔ / گھوَ	/gʰo / گھو
/sə/ سَ	/sɑ/ سا	/sɪ/ سِ	/si/ سِي	/se/ سي	/sʊ / سُ	/su/ سُو	/sɛ/ سَي	/sɔ/ سَو	/so/ سو
/ɳə/ ٹَ	/ɳɑ/ ٹا	/ɳɪ / ٹِ	/ɳi/ ٹِي	/ɳe/ ٹي	/ɳʊ / ٹُ	/ɳu/ ٹُو	/ɳɛ/ ٹَي	/ɳɔ/ ٹَو	/ɳo/ ٹو
/ɳʰə / ٹھَ	/ɳʰɑ/ ٹھا	/ɳʰɪ / ٹھِ	/ɳʰi/ ٹھِي	/ɳʰe / ٹھي	/ɳʰʊ / ٹُھ	/ɳʰu/ ٹُھو	/ɳʰɛ/ٹھَي	/ɳʰɔ/ ٹھَو	/ɳʰo/ ٹھو
/ʃə/ شَ	/ʃɑ/ شا	/ʃɪ/ شِ	/ʃi/ شِي	/ʃe/ شي	/ʃʊ / شُ	/ʃu/ شُو	/ʃɛ/ شَي	/ʃɔ/ شَو	/ʃo/ شو
/hə/ هَ	/hɑ/ ها	/hɪ / هِ	/hi/ هِي	/he/ هي	/hʊ/ هُ	/hu/ هُو	/hɛ/ هَي	/hɔ/ هَو	/ho/ هو
/ʤʰə/ جھَ	/ʤʰɑ/ جھا	/ʤʰɪ / جھِ	/ʤʰi / جھِي	/ʤʰe/ جھي	/ʤʰʊ/ جُھ	/ʤʰu/ جُھو	/ʤʰɛ/ جھَي	/ʤʰɔ/ جھَو	/ʤʰo/ جھو
/lə/ لَ	/lɑ/ لا	/lɪ / لِ	/li / لِي	/le/ لي	/lʊ/ لُ	/lu/ لُو	/lɛ/ لَي	/lɔ/ لَو	/lo/ لو
/lʰə/ لھَ	/lʰɑ/ لھا	/lʰɪ / لھِ	/lʰi/ لھِي	/lʰe/لھي	/lʰʊ/ لُھ	/lʰu/لُھو	/lʰɛ/ لھَي	/lʰɔ/لھَو	/lʰo/لھو
/mʰə/ مھَ	/mʰɑ/ مھا	/mʰɪ / مھِ	/mʰi / مھِي	/mʰe/ مھي	/mʰʊ/ مُھ	/mʰu/ مُھو	/mʰɛ/ مھَي	/mʰɔ/ مھَو	/mʰo/ مھو
/nʰə/ نھَ	/nʰɑ/ نھا	/nʰɪ / نھِ	/nʰi / نھِي	/nʰe / نھي	/nʰʊ/ نُھ	/nʰu/ نُھو	/nʰɛ / نھَي	/nʰɔ/ نھَو	/nʰo/ نھو
/jə/ يَ	/jɑ/ يا	/jɪ / يِ	/ji / يِي	/je/ يي	/jʊ/ يُ	/ju/ يُو	/jɛ/ يَي	/jɔ/ يَو	/jo/ يو
/və/ وَ	/vɑ/ وا	/vɪ/ وِ	/vi/ وِي	/ve/ وي	/vʊ/ وُ	/vu/ وُو	/vɛ/ وَي	/vɔ/ وَو	/vo/ وو
/dʰə/ ڏَ	/dʰɑ/ ڏا	/dʰɪ / ڏِ	/dʰi/ڏِي	/dʰe/ڏي	/dʰʊ/ ڏُ	/dʰu/ڏُو	/dʰɛ/ڏَي	/dʰɔ/ ڏَو	/dʰo/ڏو

3.8 Voice sample recording for diphthongs

The presence of the diphthongs in the sound inventory of Sindhi has remained a debateable issue among researchers for five decades. Ernest Trumpp in his book, '*Grammar of the Sindhi Language*', published in 1872, argued that there is no diphthong sound in Sindhi; the two vowels /ɑɪ/, and /ɑʊ/ are loosely pronounced as a-ɪ and a-ʊ (Trumpp, 1872). The author Jatoi in his book (written in Sindhi script) said that there are diphthongs in Sindhi such as diphthong /ɑɪ/ in word l[ɑɪ], /iʊ/ in word p[iʊ] etc. (Jatoi, 1996). The author Bughio in his article discussed the existence of two diphthongs in Sindhi: (i) /ɑɪ/ as non-labialised and (ii) /ɑʊ/ as labialised diphthong (Bhugio, 2006). Due to the issue of diphthong presence controversy it was hard to compile a concrete list of words that carry diphthong sounds of Sindhi. To overcome this issue a list of twenty eight words (say raw words) was compiled; to the best of our knowledge these words may carry a diphthong sound. Table 3.5 shows the list of raw words compiled for the voice samples of the diphthongs.

Table 3.5 List of the raw word utterances for diphthong sounds (Keerio, 2010).

S.No	Sindhi	Roman Sindhi	S.No	Sindhi	Roman Sindhi
1	عينک	aynaka	15	ٻُڌاءِ	budhaey
2	سَئير	sayru	16	خوُشبوُءِ	khushboey
3	عَيب	aybu	17	چَوندئو	chawandao
4	سِيءَ	siya	18	ڀَئو	bhao
5	هِيءَ	hiya	19	هُوءَ	Houa
6	پوءِ	poe	20	ماءُ	mau
7	گوءِ	goe	21	پاءُ	pau
8	هُئي	huiy	22	سِيءُ	siyu
9	پِيتُي	pituiy	23	پِيءُ	piyu
10	سُئِي	suiy	24	جِيءُ	jiyu
11	ٿِئي	thiey	25	ڌِيءَ	dhiya
12	پِئي	piey	26	جُئَر	juar
13	لاءِ	laey	27	ڌُئو	dhuo
14	پاءِ	paey	28	پُئو	puo

3.9 Problems faced by the speakers

Our data collection method was well focused and targeted in the sense that the recording for the voice samples were aimed at three major classes of the language sounds i.e. *vowels*, *consonants* and *diphthongs*. The selection of the word utterances for vowels and consonants were made by finding the minimal pair words for each sound (minimal pair method is discussed in chapter IV). Due to the use of the minimal pair method some words included in the study were odd or strange for the speakers to pronounce. Because such words are very rarely used in general language conversation, for example the words *Manhein* ‘your mother’, and *Chawandao*, ‘you will say’, etc. It was also difficult for speakers to produce isolated utterances of the language sounds (vowels and consonants); especially for the speakers who got their primary education in English medium schools. There was a huge microphone speaking hesistation or shyness, especially in the rural areas of Sindh.

3.10 Conclusion

In this chapter the overall process of data collection during the field study is described. The core research findings and suggestions in subsequent chapters are mainly based on the analysis of the raw data collected during the field study. Voice samples are collected from the five geographically distinct dialects of Sindhi spoken in Sindh province of Pakistan. Seventy five adult male speakers participated in this study with fifteen male speakers for each dialect. Voice samples recordings were made for three major classes of language phonemes (vowels, consonants and diphthongs). At least ten instances of each utterance were made for the purpose of subsequent offline process.

CHAPTER 4

INTRODUCTION TO THE SOUNDS OF SINDHI

Chapter 4

Introduction to the Sounds of Sindhi

4.1 Introduction

In this chapter the meaningful phonemes of Sindhi are identified by finding the minimal pair words for each sound. A minimal pair method is widely used in phonetic studies for the identification of the meaningful phonemes of a particular language. In this study the meaningful phonemes of Sindhi are identified by providing the minimal pair of words for each sound. A pair of words, if differ by one phone at an identical position and the two words also differ in their meaning; the differing phone that cause the meaningful distinction between two words is then identified as a meaningful phoneme of Sindhi. In this way forty eight consonant and ten vowel phonemes of Sindhi are identified.

The nine redundant sounds of Sindhi are also identified in this chapter, this helps to make complex structural compositions of language simple and easy to learn. In addition to the three pairs of interchanging sounds are identified in this chapter. The clear description of these interchanging sounds help in this study to avoid potential segementation and recognition errors, because the speakers unintentionally replace one sound with other; whereas the interchanging sound do not change the contextual meaning of the word in utterance. In Sindhi there are three pairs of substituting consonant sounds, the majority of these substituting sounds are borrowed sounds from other languages such as Arabic and Persian. The words pronounced with substituting sounds are considered phonetically incorrect however the native Sindhi speaker's pronounce words with substituting sounds frequently. The chance of segmentation and recognition error is high when words are pronounced phonetically incorrect. Twelve such sounds are also identified in this chapter that are not yet part of the language alphabet. These sounds are frequently used by the native speakers and need to be identified and type categorized; because the phonemic status of these sounds is unknown. The classification of these sounds is presented in appendix B after the acoustic-phonetic analysis of the consonant sounds in chapter VIII.

4.2 Phonemes of Sindhi

Phonetically the smallest sound unit capable of carrying meaningful distinction between two words is referred to as a meaningful phoneme of the language. For example two sounds /m/, and /s/ are meaningful phonemes of English, because the two sounds carry the meaningful distinction between words '*mat*' and 'sat' (Olive, 1993). In Sindhi there are fifty one sound symbols (the graphic notations) for the representation of the language elemental sounds (Mirza, 2006). A thorough linguistic sciences study is required for Sindhi to propose the concrete set of the language elemental sounds and their corresponding symbolic representation (because it includes redundant, confused, interchanging, and substituting pairs of the consonant sounds); this study mainly focuses on the aspects of acoustic-phonetics (a subfield of phonetics) of the language elemental sounds. A minimal pair method is used to identify the meaningful phonemes of Sindhi and the IPA (International Phonetic Alphabet) notation given in the IPA hand book (IPA, 1999) is used for the symbolic representation of the language phonemes in the subsequent chapters of this thesis. A meaningful phoneme is enclosed with two forward slashes '//' and the transcription in Sindhi is enclosed within the quotation marks "". Table 4.1 below shows the list of consonant sounds of Sindhi by providing the minimal pair of each sound. The conventional criteria used for the selection of a meaningful phoneme is that if two words differ in one phone at an identical position and they also differ in word meaning then the sound that causes the meaningful distinction between two words is a meaningful phoneme of Sindhi.

Table 4.1 Minimal pair words for the consonants of Sindhi (Jatoi, 1996)

S.no	Minimal pair	Transcription (in Sindhi)	Identified phoneme	Sindhi Alphabet
1	bɑrʊ	بارُ	/b/	ب
	ɓɑrʊ	ٻارُ	/ɓ/	ٻ
2	pəlo	پَلو	/p/	پ
	b^{h}əlo	ڀَلو	/b^{h}/	ڀ
3	tɑrʊ	تارُ	/t/	ت
	ʈ^{h}ɑrʊ	ٿارُ	/ʈ^{h}/	ٿ
4	ʈokə	ٽوڪَ	/ʈ/	ٽ
	t^{h}okə	ٺوڪَ	/t^{h}/	ٺ

-Continued on next page -

5	fɪkɪrʊ	فِڪِرُ	/f/	ف
	nɪkɪrʊ	نِڪِرُ	/n/	ن
6	p^{h}ərʊ	ڦرُ	/p^{h}/	ڦ
	gərʊ	گرُ	/g/	گ
7	ɠərʊ	ڳرُ	/ɠ/	ڳ
	k^{h}ərʊ	کرُ	/k^{h}/	ک
8	mʊŋə	مُنگ	/ŋ/	ڱ
	mʊɲə	مُنڃ	/ɲ/	ڃ
9	dərʊ	دَرُ	/d/	د
	ɗərʊ	ڏرُ	/ɗ/	ڏ
10	ɖ^{h}ʊkʊ	ڍڪُ	/ɖ^{h}/	ڍ
	ɖʊkʊ	ڊُڪُ	/ɖ/	ڊ
11	ʤərʊ	جَرُ	/ʤ/	ج
	ʄərʊ	ڄَرُ	/ʄ/	ڄ
12	ʧɪlko	چِلڪو	/ʧ/	چ
	ʧ^{h}ɪlko	ڇِلڪو	/ʧ^{h}/	ڇ
13	xɪlafʊ	خِلاف	/x/	خ
	ɣɪlafʊ	غِلاف	/ɣ/	غ
14	rʊtʊ	رُتُ	/r/	ر
	kʊtʊ	ڪُتُ	/k/	ڪ
15	zari	زاري	/z/	ز
	mari	ماري	/m/	م
16	pəɽʊ	پڙُ	/ɽ/	ڙ
	pəɽ^{h}ʊ	پڙهُ	/ɽ^{h}/	ڙه
17	g^{h}aro	گهارو	/g^{h}/	گه
	saro	سارو	/s/	س
18	maɳe	ماڻي	/ɳ/	ڻ
	maɳ^{h}en	ماڻهين	/ɳ^{h}/	ڻه
19	kaʃʊ	ڪاش	/ʃ/	ش
	kaɳʊ	ڪاڻ	/ɳ/	ڻ
20	harʊ	هارُ	/h/	ه
	ʤ^{h}arʊ	جهارُ	/ʤ^{h}/	جه

-Continued on next page -

21	kələ	ڪَلَ	/l/	ل
	kəlʰə	ڪَلھ	/lʰ/	لھ
22	sʊme	سُمي	/m/	م
	sʊmʰe	سُمهي	/mʰ/	مھ
23	sənɑ	سنا	/n/	ن
	sənʰɑ	سنها	/nʰ/	نھ
24	jɑrʊ	يارُ	/j/	ي
	vɑrʊ	وارُ	/v/	و
25	dʰɑɠo	ڌاڳو	/dʰ/	ڌ
	rɑɠo	راڳو	/r/	ر

4.3 Classification of consonant sounds of Sindhi

Consonants are tend to be easier to describe and classify in groups than vowels by using their articulatory information such as place and manner of the articulation (Kent, 2002; Olive, 1993). Consonants that share some common articulatory characteristics can be grouped together to form a natural class of phonemes, for example in Sindhi the sounds **/p/, پ , /pʰ/, ڦ** , and **/b/,** ب , **/bʰ/,** ڀ can be grouped and classified in one natural class as bilabial (place of articulation) plosive stops (manner of articulation) of Sindhi. The voicing feature of phonology further divides a natural class of sounds into two subclasses as: *voiced* and *unvoiced*. Consonants produced with the vibrating vocal cords are referred to as voiced and the others are referred to as unvoiced sounds of Sindhi. In this chapter the following three main articulatory features are used for the classification of consonant sounds of Sindhi:

- The state of vocal cords: if vibrated the voiced consonant is produced, otherwise the unvoiced consonant is produced
- The place of articulation: which articulators are involved during the production of the consonant sounds and where the constriction in the vocal tract occurs
- The manner of articulation: how the organs are involved in making the consonant sounds i.e. the degree of obstruction complete or partial and the state of the velum (whether the nasal cavity is open or closed)

4.3.1 Stop consonants

Stops are produced by blocking the airstream completely for a moment in the vocal tract followed by the sudden release of the airstream. The complete closure for the air passage through the vocal tract can be achieved by closing the lips and the nasal cavity for the production of stops. One of the unique properties associated with the stops of Sindhi is that if a stop consonant is followed by a vowel sound in an utterance then the stop consonant become's a plosive consonant and is referred to as a plosive stop consonant of Sindhi. The phonemic inventory of Sindhi consists of the largest set of the stop consonants than the other classes of the consonant sounds. Table 4.2 below shows the list of the stop consonants of Sindhi.

4.3.2 Implosive stops

Implosive stops exist in a few of the world language's sound inventory; the notable ones are the Hausa and Swahili spoken in Africa, and the Sindhi spoken in the Indian subcontinent. Among these languages Sindhi posses the large number of contrastive implosive stops. Implosives of Sindhi are purely non-aspirated glottal ingressive sounds produced by lowering the larynx with vibrating vocal cords (Jennifer, 2006; Nihalani, 1986; Raza, 2004). Implosives of Sindhi are produced with the vibrating vocal cords; therefore they are referred to as voiced consonants of Sindhi. To carry out the acoustic analysis of these sounds is one of the greatest motivations of this research work. Table 4.3 below shows the list of implosive stops of Sindhi.

Table 4.2 Stop consonants of Sindhi

Voiced		Unvoiced		Place of articulation	Voiced		Unvoiced		Place of articulation
/b/	ب	/p/	پ	Bilabial	/ʤ/	ج	/ʧ/	چ	Palato-Alveolar
/bʰ/	ڀ	/pʰ/	ڦ	Bilabial	/ʤʰ/	جھ	/ʧʰ/	ڇ	Palato-Alveolar
/d/	د	/t/	ت	Dental	/g/	گ	/k/	ڪ	Velar
/dʰ/	ڌ	/tʰ/	ٿ	Dental	/gʰ/	گھ	/kʰ/	ک	Velar
/ɖ/	ڊ	/ʈ/	ٽ	Retroflex	--	--	--	--	--
/ɖʰ/	ڍ	/ʈʰ/	ٺ	Retroflex	--	--	--	--	--

Table 4.3 Implosive stop consonants of Sindhi

Voiced		Place of articulation	Voiced		Place of articulation
/ɓ/	ٻ	Bilabial	/ ʄ/	ڄ	Palato-Alveolar
/ɗ/	ڏ	Retroflex	/ɠ/	ڳ	Velar

4.3.3 Nasal consonants

Nasals are produced by lowering the velum in the mouth; allowing the air passage through the nasal cavity (Olive, 1993). There are five nasals in Sindhi and they all are voiced (produced with vibrating vocal cords). Table 4.4 below shows the nasal consonants of Sindhi.

Table 4.4 Nasal consonants of Sindhi

Voiced		Place of articulation	Voiced		Place of articulation
/m/	م	Bilabial	/ɳ/	ڻ	Retroflex
/n/	ن	Alveolar	/ɲ/	ڃ	Palatal
/ŋ/	ڱ	Velar	--	--	--

4.3.4 Liquids

Trill and lateral class of consonant sounds are jointly referred to as liquids of Sindhi. Liquids are produced with a partial (much less) constriction of the air in vocal tract, sounds produced with less constriction are frictionless speech sounds (Olive, 1993); thus the liquids of Sindhi are frictionless sounds. The two liquids of Sindhi are categorized as one lateral consonant: **/l/, ل**, and the other is a trill consonant: **/r/, ر** . When the tip of the tongue touches the alveolar ridge in such a way so that the air passes through the sides of the mouth then the lateral consonant **/ l /,ل** of Sindhi is produced. When the tip of the tongue touches the alveolar ridge loosely; so that the air passage causes the vibration at the tip of the tongue, then the trill consonant **/ r /,ر** of Sindhi is produced. One special property associated with liquids of Sindhi is that if they appear at the word beginning or medial position they are characterised as voiced and if they appear at word final positions they are characterized as unvoiced. For example the trill sound **/ r /,ر** at final position in word **'sətər'**, and the lateral sound **/ l /,ل** at final position in word **'ədəl'**, are unvoiced liquid consonants of Sindhi.

4.3.5 Fricatives

Fricatives are produced by bringing articulators close together to make the vocal tract highly constricted. Due to the friction, air passing through the highly constricted area is set into turbulent motion, and produces noisy hissing sounds called fricatives (Olive, 1993). There are nine fricatives in Sindhi out of which five are categorized as voiced and four are unvoiced. The fricative consonant **/h/, ھ** of Sindhi can be either voiced or unvoiced. If it appears at the beginning or in middle position of the utterance it is voiced, and unvoiced if it appear at the final position in the utterance. Table 4.5 shows the list of fricative consonants of Sindhi.

Table 4.5 Fricative consonants of Sindhi

Voiced		**Place of articulation**	**Unvoiced**		**Place of articulation**
/ v /	و	Labio-dental	/f/	ف	Labio-dental
/z/	ز	Alveolar	/s/	س	Alveolar
/j/	ي	Palatal	/ʃ/	ش	Palato-Alveolar
/ɣ/	غ	Velar	/x/	خ	Velar
/h/	ھ	Glottal	/h/	ھ	Glottal

4.3.6 Affricates

Four consonant sounds of Sindhi: (i) / ʧ /, چ (ii) / ʤ /, ج (iii) / $ʧ^h$ /, ڇ (iv) /$ʤ^h$/, جھ are reported as affricate sounds by authors (Jennifer, 2006; Raza, 2004). However the authors (Jatoi, 1996; Ladefoged, 1996; Mirza, 2006) have classified these sounds as stop consonants of Sindhi. There exists some controversy among the researchers regarding the identification and classification of these four sounds of Sindhi. So far the classification of these sounds: (i) / **ʧ** / (ii) / **ʤ** / (iii) / **$ʧ^h$** / (iv) /**$ʤ^h$**/, is deferred until the acoustic analysis of these sounds carried out in chapter VIII.

4.4 Redundant consonant sounds

The following nine redundant alphabet symbols of Sindhi represent nine redundant consonant sounds of the phonemic inventory of Sindhi. The majority of these sounds are the borrowed from Arabic or Persian languages are: (i) /s/, ث (ii) /h/, ح (iii) /z/, ذ, (iv) /s/, ص (v) /z/, ض (vi) /t/, ط (vii) /z/, ظ (viii) /k/, ق, (ix) /?/ ع.

4.5 Confused consonant sounds

There are three pairs of consonant sounds in Sindhi that speakers in general conversation pronounce in such a way so that the listeners are confused in identifying which sound is pronounced. The confused consonant pairs of Sindhi are: /b/, ب sounds like /bʰ/, ڀ of Sindhi and vice versa, the sound /ɖ /, ڊ sounds like /ɖʰ/, ڍ of Sindhi and vice versa, the sound /g/, گ sounds like /gʰ/, گھ of Sindhi and vice versa.

4.6 Interchanging consonants

There are three pairs of consonant sounds in Sindhi that the speakers of different dialects interchange while speaking them. For example the speakers of the Middle dialect pronounce the word fish as: /məʧʰi/,ميڇي with /ʧʰ /, ڇ sound and the speakers of the Utradi dialect change this sound with sound /ʃ/, ش and pronounce the word fish as: /məʃi/, مشي and in similar way the sound / r /, ر is changed with /ɽ/, ڙ and vice versa, and the sound /ʤ/, ج is changed with /z/, ز and vice versa.

4.7 Substituting consonant sounds

There are three substituting pairs of consonants in Sindhi: (i) the sound /pʰ/, ڦ is substituted with /f/, ف and vice versa (ii) the sound /kʰ/, ک is substituted with /x/ خ and vice versa (iii) the sound /ɣ/, غ is substituted with /g/, گ and vice versa.

4.8 Sounds that are not yet part of the language alphabet

There are twelve sounds of Sindhi, which are not yet part of the language alphabet. Two alphabet symbols are used to represent these sounds symbolically. They are: (i) /ɽʰ/, ڙھ (ii) /rʰ/, رھ (iii) /lʰ/, لھ (iv) /nʰ/, نھ (v) /mʰ/, مھ (vi) /ɳʰ /, ڻھ. (vii) /m̶/,

مين (viii) /ʔ/, ء (ix) /ʈr/, ٽر (x) /ɖr/, ڊ ر (xi) /lɑ/, لا (xii) /yɑ/ ي . After the acoustic analysis of the consonant sounds of Sindhi the classification of these sounds will be presented.

4.9 Vowels

Phonetically vowels are classified by tracking the shape changes of the vocal tract during the production of these sounds. The main articulators involved in changing the shape of the vocal tract are: *the tongue, the lips* and *the jaw* (Jones, 1969). These articulators are referred to as moving articulators that move against the fixed articulators in the oral cavity, for the vocal tract shape changes. The lips during the production of vowels are either rounded or remain neutral. The lip protrusion (when lips are pushed forward) during the production of vowels make the vocal tract a bit longer than the normal. The lower jaw can be raised up or lowered down for the closure or opening of the oral cavity. The tongue is the most flexible of the moving articulators. It can be moved forward or backward, and upward or downward during the production of vowels. Vowels are generally produced with open vocal tract and frictionless air flow, the narrowest aperture that the tongue may make without friction is called close and the widest it makes will be called open (Olive, 1993; Raphael, 2006; Stevens, 2000). In this work vowels are identified by finding their minimal pair words. If two words are different in one vowel sound at an identical position and the two words also differ in their meaning then the differing vowel sound is chosen as a meaningful vowel phoneme of Sindhi. In this way ten monophthong vowels of Sindhi are identified. Table 4.6 shows the list of ten vowels of Sindhi.

Table 4.6 Minimal pair words for the vowels of Sindhi (Jatoi, 1996).

S.No	IPA symbol	Sindhi	Example	Sindhi	Translation
1	i	اِي	sirə	سِيرَ	Wave of the sea
	ɪ	اِ	sɪrə	سِرَ	A brick
2	e	اي	serə	سيرَ	Unit of weight (one kilo)
	ɛ	اَي	sɛrə	سَئيرَ	Walk

-Continued on next page -

3	ə	اَ	sərə	سَرَ	Name of plant
	ɑ	آ	sɑrə	سارَ	Remember
4	ɔ	اَو	ʧəvəndɔ	چَوندئو	You will say
	o	او	ʧəvəndo	چَوندو	He will say
5	ʊ	اُ	sʊrə	سُرَ	Tune
	u	اُو	surə	سُورَ	Pain

4.10 Conclusion

This chapter serves three major purposes in this study: (i) the phonemic importance of the language sounds by providing minimal pair words for each sound, (ii) the articulatory description and classification of the consonant and vowel sounds of Sindhi, (iii) the identification of the language sounds that are either redundant, confused, interchanged or they are not yet part of the language alphabet. Until the acoustic analysis of the consonant sounds the phonemic status of the redundant and the sounds that are not yet part of the language alphabet is ambiguous; the acoustic analysis and the classification of these sounds is discussed in chapter VIII, along with the acoustic analysis of the consonant sounds of Sindhi.

CHAPTER 5

PRE-PROCESSING OF SPEECH SIGNALS

Chapter 5

Pre-processing of Speech Signals

5.1 Introduction

This chapter is divided in two parts, first part discusses some statistical outlier-detection based techniques used to remove the unwanted signals from the voice samples such as the ambient noise or the signals that are part of the silence. The process of removing unwated signals from speech is referred to as signal pre-processing. The second part mainly focuses upon how the acoustic-phonetic features are determined for the captured voice samples during the field study? The voice samples collected during the field study for this work are discussed in detail in chapter III.

Pre-processing of speech signals is a crucial step in the development of a robust speech or speaker recognition system. It also helps significantly to reduce the overall complexity of the analysis process and hence improves the overall system efficiency. In this study we present some popular statistical outlier-detection based strategies to segregate the ambient noise signals or the silence/unvoiced parts of the speech signal from the voiced portion of the speech. The proposed methods are based on the utilization of the '3σ edit rule', and the Hampel Identifier which are compared with the conventional techniques: (i) Short-Time Energy (STE) based methods, and (ii) Distribution based methods.

There are numerous speech or speaker recognition systems discussed in the literature i.e. isolated word, connected word, Continuous speech, and spontaneous speech recognizers (Anusuya, 2009; Furui, 2005). Whichever method one is going to implement for the design of a recognizer, determining the acoustic-phonetic features for the speaker speech is considered one of the essential steps involved in the design of a robust speech or speaker recognizer. The main acoustic features of the voice samples concentrated on in this study are the fundamental frequency, the resonant frequencies, and the duration of each phoneme of Sindhi. The mean and the standard deviation of the obtained features for each phoneme are given with the results.

5.2 Pre-processing of speech signals

Pre-processing of speech signals, i.e. segregating the voiced signals from the silence/unvoiced signals is usually advocated as a crucial step in the development of a reliable speech or speaker recognition system. This is because most of the speech or speaker specific attributes are present in the voiced part of the speech signals (Saha, 2005); moreover, extraction of the voiced part of the speech signal by marking and/or removing the silence and unvoiced region leads to a substantial reduction in computational complexity at the later stages (Mitra, 2005; Saha, 2005). Other applications of classifying speech signals into silence/unvoiced region and voiced region, as described in (Saha, 2005), are: Fundamental Frequency Estimation, Formant Extraction or Syllable Marking, Stop Consonant Identification, and End Point Detection for isolated speech signals.

One of the accepted ways of labeling a speech signal is the three state representation: (i) Silence Region (SR) where no speech is produced, (ii) Unvoiced Region (UR), where the resulting waveform is aperiodic or random in nature as the vocal chords do not vibrate, and (iii) Voiced Region (VR) where the resulting waveform is quasi-periodic as the vocal chords are tensed and hence vibrate periodically (Rabiner, 1993; Saha, 2005). It should be made clear that the segmentation of the speech signal into the aforementioned regions is not very rigid; however, it has been noted that small errors in the boundary locations seldom have any significant effect in most of the applications (Rabiner, 1993). Mention should also be made of the fact that as the energy in the unvoiced portion of the speech signal is usually low; this region is usually clubbed together with the silence region and segregated from the voiced part (Saha, 2005).

The short-time energy based methods to segregate the silence/unvoiced region from the voiced portion are in general fast; however, these methods are stymied by the fact that the thresholds needed to implement them are chosen on an *ad hoc* basis. This means that the recognition system has to be retuned every time there is some change in the ambience (Mitra, 2005). It has been noted that when the method is applied with another popular technique to segregate the different portions of the speech signal, namely the zero cross detection rate (ZCR) method, the success achieved is around 65% (Saha, 2005). The second category of methods, i.e. distribution based methods, rely heavily on the distribution of the first few thousand samples of the signal which are assumed to be part of the noise region (Saha, 2005). The methods completely fail if the noise is not well described by the chosen distribution function. The outlier-detection based strategies, on the other hand, can be applied on a constrained database where one class of the data is ideally a small fraction of the other class (Mitra, 2007). Here, also, the '3σ edit rule' depends heavily on the distribution of the entire data; however, as will be discussed below, the strategy based on the Hampel Identifier can be successfully applied to segregate the silence/unvoiced region from the voiced portion. The motivation behind trying the statistical outlier-detection based strategies stems from the fact that if the microphone is kept on for a few seconds before and after the utterance of the word(s), then the samples of the voiced portion can be treated as outliers and the samples of the silence/unvoiced region as inliers.

5.2.1 Short-Time Energy (STE) based methods

STE based methods of speech signal segregation utilizes the fact that energy in the voiced region is greater than in the silence/unvoiced region (Atal, 1976; Childers, 1989; Mitra, 2005; Saha, 2005). So, a small window is taken and the energy of the window calculated; if the total energy of the window is more than the chosen threshold, then samples of the window are retrieved, otherwise dropped. Two different experiments were conducted; the first one with a non-overlapping moving window of fixed size, and the other with an overlapping moving window of fixed size. The overlapping moving window technique has been used to prevent valuable information loss (Mitra, 2005). It can be shown that if the all the samples of the captured data are divided into several over-lapping blocks of size T with the distance between the start of two blocks as N, then the total loss (L_T) can be restricted to (Mitra, 2005):

$$0 < L_T < 2(T - T_1) = 2N \tag{5.1}$$

where, in (5.1), T_1 is the number of overlapped samples between two successive positions of the fixed sized window.

Note that in the above expression, the values of N and T should be chosen very carefully in order to maximize the desired system accuracy as well as to minimize the operational counts. In our case, the selection of values has been based on the optimization of the following approximate cost function:

$$f(N,T) = \left(\frac{\lambda - T}{N} + 1\right)T$$

$$\cong \left(\frac{\lambda - T}{N}\right)T \qquad (5.2)$$

where, in (5.2), λ is the total number of samples in the raw data.

It is evident that such a function would always yield the value $f(N,T) = \lambda$ for points with the condition $N = T$; these points are found to be saddle points. However, we have emphasized having overlapping blocks so that the total loss can be minimized as per the system requirement. From a system accuracy point-of-view, the ideal value of N should be 1 and therefore, $N = T$ is not a feasible solution. On the other hand, decreasing the value of N translates into the increasing the number of total unit operations; we have used the following values for our experiments $T = 4410$ (200msec window, sampling rate 22050 samples/second) and $N = 2205$.

Taking an average value of λ, say 308265, the total unit operations for one successful execution of the algorithm were 606335 approximately, and the total loss was restricted to $\left(\frac{2N}{\lambda} \times 100\right) = 1.43\%$ of the entire data.

5.2.2 Distribution based methods

Here, the nature of the background noise has to be assumed, and then the relevant distribution functions are used for the segregation purpose (Rabiner, 2007; Saha, 2005;). If we assume that the ambient noise present in the captured signal is Gaussian in nature, then the uni-dimensional Mahalanobis distance function, which itself is a Linear Pattern

Classifier (Duda, 2000; Sarma, 1978; Saha, 2005) can be used to extract the voiced part from the signal.

In this case, at first a 200msec window is chosen to find out the parameters of the Gaussian distribution; the time duration of the window is chosen considering the fact that the speaker will take more than 200msec to initiate speaking after recording starts (Saha, 2005).

Note that the Gaussian distribution in one dimension is defined as:

$$g(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{(x-\mu)^2}{2\sigma^2}} \qquad (5.3)$$

where, in (5.3), μ is the mean and σ is the standard deviation of the distribution. It can be calculated that the probabilities obey:

$$\Pr[|x-\mu| \le \sigma] \approx 0.68$$

$$\Pr[|x-\mu| \le 2\sigma] \approx 0.95$$

$$\Pr[|x-\mu| \le 3\sigma] \approx 0.997$$

The Mahalanobis distance (r) from x to μ is defined as:

$$r = \frac{|x-\mu|}{\sigma} \qquad (5.4)$$

From the calculations of the probabilities it is evident that there is a probability of 99.7% that the distance, *r*, will be less than 3.

5.2.2.1 Application of the Mahalanobis Distance method

A window of duration 200msec is chosen (4410 samples, sampling rate being 22,050 samples/sec) to calculate the mean and the standard deviation of the distribution. The duration of the window is chosen based on the assumption that the speaker will take more than 200ms of time to initiate the process of speaking; in other words, this means the chosen window contains ambience noise samples. Now the Mahalanobis distance is calculated for each sample after the window. If the distance calculated is more than 3 then the sample is restored as a voice sample, otherwise dropped.

5.2.3 Outlier-detection based strategies

Data points, ξ_k, in a data set, ξ, that do not agree with our expectations based on the bulk of the data are termed as outliers (Hawkins, 1980; Pearson, 2002). The popular automatic outlier-detection approaches depend on two estimates: (1) an estimate of a nominal reference value for the data set, and (2) a scatter estimate of the data. Based on these estimators, outliers can be detected based on the following criterion:

$$\left|\xi_k - \xi_{kr}\right| > \alpha\gamma \quad \Rightarrow \xi_k = \xi_{ko}, \qquad \forall \xi_k \in \xi \tag{5.5}$$

where in (5.5), ξ_{kr} is the nominal reference value of the dataset, α is the threshold parameter, γ the scatter estimate, and ξ_{ko} an outlier.

5.2.3.1 The '3 σ edit rule'

The outlier-detection strategy based on the '3 σ edit rule' considers the mean of the data values of the data set as the nominal reference value and the corresponding standard deviation as an estimate of the scatter:

$$\xi_{kr} = \xi_{mean} = \frac{1}{\mathrm{N}} \sum_{k=1}^{\mathrm{N}} \xi_k \tag{5.6}$$

where in (6.6), N is the total number of observations in the data set.

$$\gamma = \left[\frac{1}{\mathrm{N}-1} \sum_{k=1}^{\mathrm{N}} \left(\xi_k - \xi_{mean}\right)^2 \right]^{1/2} \tag{5.7}$$

Note that if the distribution is assumed to be approximately normal, then the probability of getting a data value greater than three times the standard deviation of the data $(\alpha = 3)$, added to the mean, is around 0.3% (Pearson, 2005). The technique suffers from the fact that both the mean and the standard deviation of the data are very much outlier sensitive (Pearson, 2002). Moreover, it should be kept in mind that the strategy is based on the fact that the underlying distribution is approximately Gaussian (Mitra, 2007).

5.2.3.2 Strategy based on Hampel Identifier

In this case, the outlier resistant median (breakpoint value of 50%) and the median absolute deviation from the median (MAD) scale estimates replace the outlier sensitive mean and standard deviation estimates respectively. The median of a data sequence is obtained as follows (Halin, 1967; Mitra, 2007):

1. The observations are ranked according to their magnitude.
2. If N is odd, the median is taken as the value of the $\left[\frac{(N+1)}{2}\right]^{th}$ ranked observation; otherwise if N is even, the median is taken as the mean of the $\left(\frac{N}{2}\right)^{th}$ and $\left[\left(\frac{N}{2}\right)+1\right]^{th}$ ranked observations.

The MAD scale estimate is defined as:

$$\gamma = MAD_{se} = 1.4826 \times median\{|\xi_k - \xi_{median}|\} \qquad (5.8)$$

where in (5.8), 'the factor 1.4826 was chosen so that the expected value of γ is equal to the standard deviation for normally distributed data' (Pearson, 2002).

The strategy, although quite often very effective in practice (Pearson, 2002), completely fails if more than 50% of the observations are of the same value, then the scale estimate is equal to 0, i.e. every data value greater than the median would then be considered as an outlier. Note, for such cases the boxplot outlier-detection strategy can also be used by replacing the mean with the median and the standard deviation with the interquartile deviation (Pearson, 2005).

5.2.3.3 Application of outlier-detection strategies

The microphone is kept on for a few seconds before and after the utterance of the speech sample by the speaker. This ensures that the silence/unvoiced region forms the major class, and the voiced samples the minor class. The two aforementioned strategies are then applied to demarcate one class from the other, i.e. to segregate the voiced portion from the noise part of the captured raw data.

5.2.4 Results

The different strategies have been tested and compared based on the utterance of (i) a three two-digit combination lock number phrase i.e. 35-72-41 (pronounced as thirty-five/ seventy-two/ forty-one) "پنجتيهہ ، ٻاهتر ، ايڪيتاليهہ", and (ii) a running text read from a paragraph, "*Acoustic signal processing extracts the desired information from speech signals*", "ٻولي جي آوازن جي جاڻ اڪوسٽڪ سگنل پروسيسنگ جي مدد سان حاصل ڪري سگهجي ٿي.". Figure 1 and Figure 2 below show two typical speech waveforms of the two experiments.

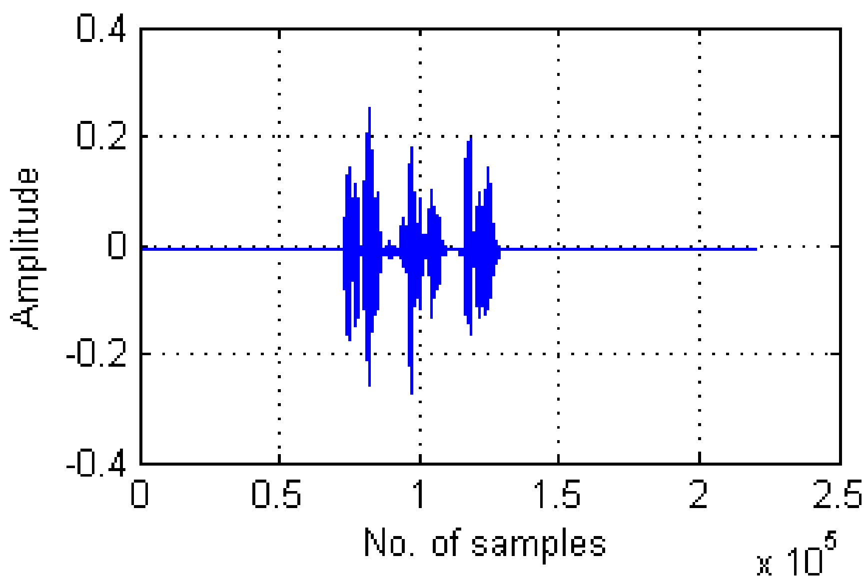


Figure 5.1 Typical speech waveform of the three two digit combinational lock number experiment (35-72-41)

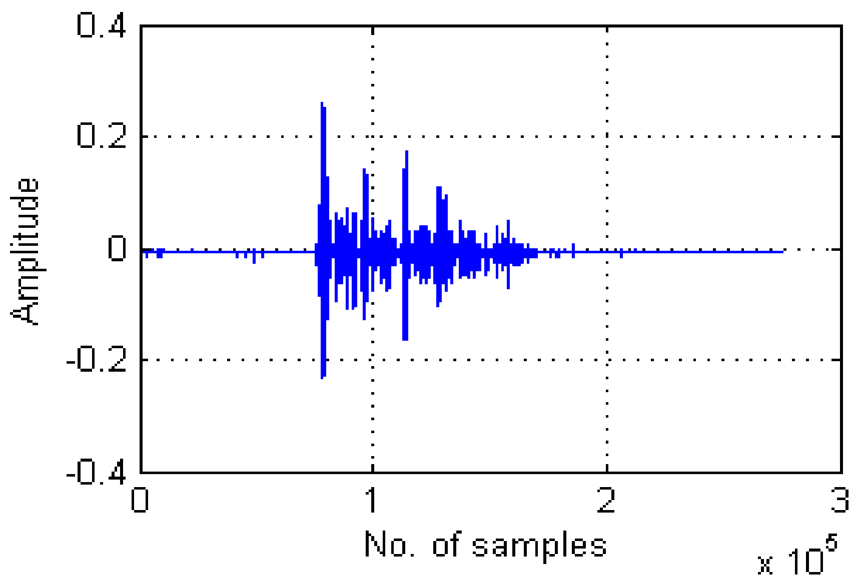


Figure 5.2 Typical speech waveform of the running text experiment: "Acoustic signal processing extracts the desired information from speech signals."

5.2.4.1 Performance criterion

A single 'figure-of-merit', percentage distortion d is defined (Saha, 2005) to analyse the performance of the method:

$$d = \frac{|Vman - Vmeth|}{Vman} \times 100\% \quad (5.9)$$

where, in (5.9) $Vman$ is the count of manually labeled voiced samples, and $Vmeth$ is the count of the voiced samples obtained as the output of an applied method.

Table 5.1 summarize the results by showing the percentage distortion for the voice samples that occurred after applying each strategy for the two experiments using equation (5.9). The STE based strategies i.e. (i) the non-overlapping moving window (of fixed size) and (ii) the overlapping moving window (of fixed size) show encouraging results for both the experiments, however their results for the combination lock number experiment are better. It is obvious from the results obtained that the results of the overlapping moving window (of fixed size) based strategy are better than the non-overlapping moving window (of fixed size) based strategy. Note that the *ad hoc* thresholds required implementing STE based strategies need to be selected carefully. In this work the *ad hoc* thresholds were set manually on a trial and error basis. Figure 5.3

and figure 5.4 below, show the typical results of the non-overlapping moving window (of fixed size) based strategy. Figure 5.5 and figure 5.6 show the typical results of the overlapping moving window (of fixed size) based strategy.

Table 5.1 The summary of average percentage distortion for the obtained voice samples (Keerio, 2009).

Strategy	Combination lock number	Running text
Non-overlapping moving window	4.92%	11.91%
Overlapping moving window	3.66%	8.40%
Mahalanobis Distance	20.91%	9.71%
3σ edit rule	28.17%	30.43%
Hampel Identifier	12.52%	9.68%

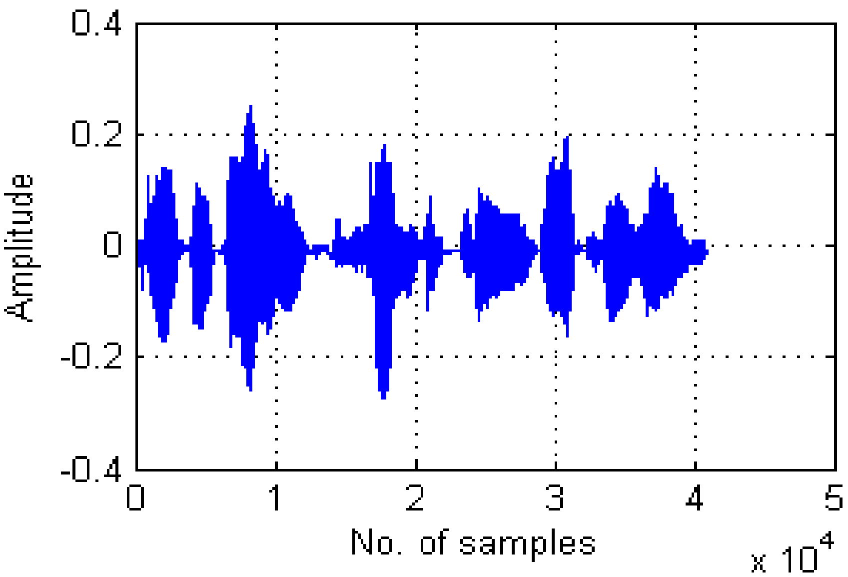


Figure 5.3 Typical output of the non-overlapping moving window (of fixed size) based strategy for the combination lock number experiment.

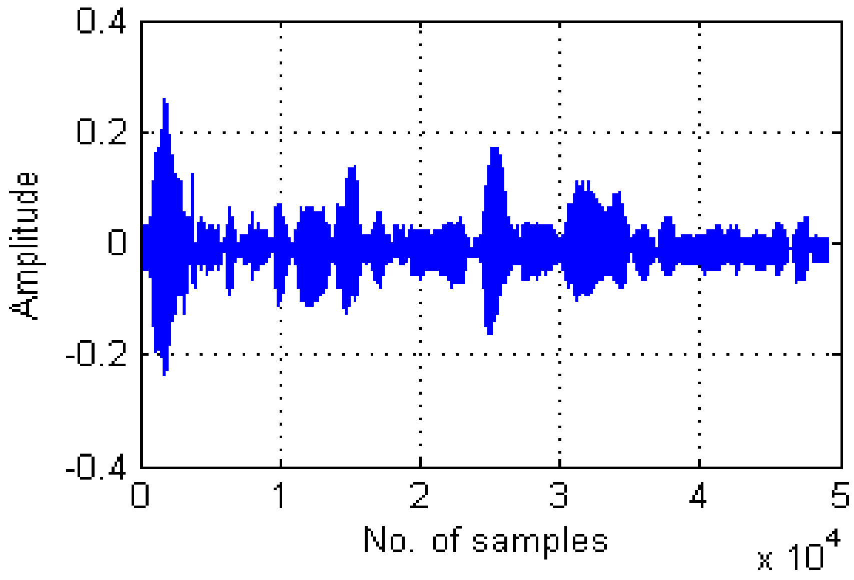


Figure 5.4 Typical output of the non-overlapping moving window (of fixed size) based strategy for the running text experiment.

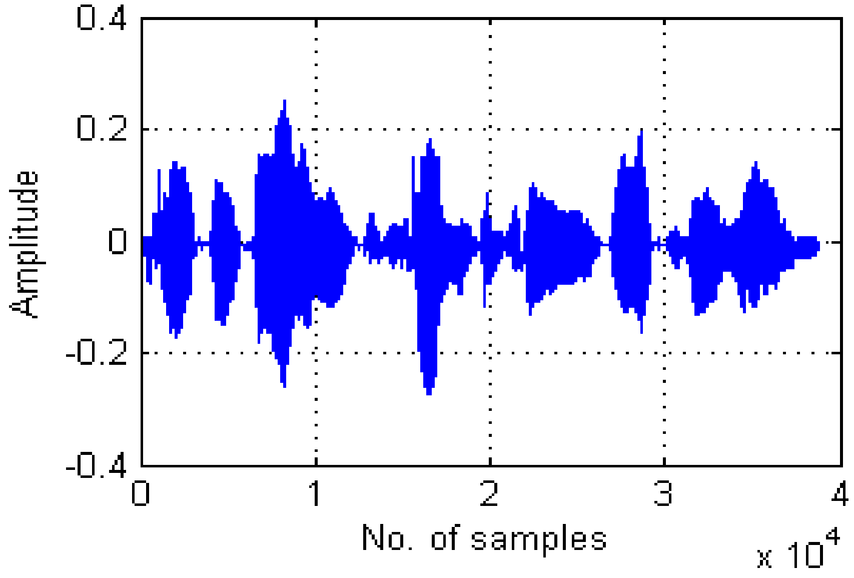


Figure 5.5 Typical output of the overlapping moving window (of fixed size) for the combination lock number experiment.

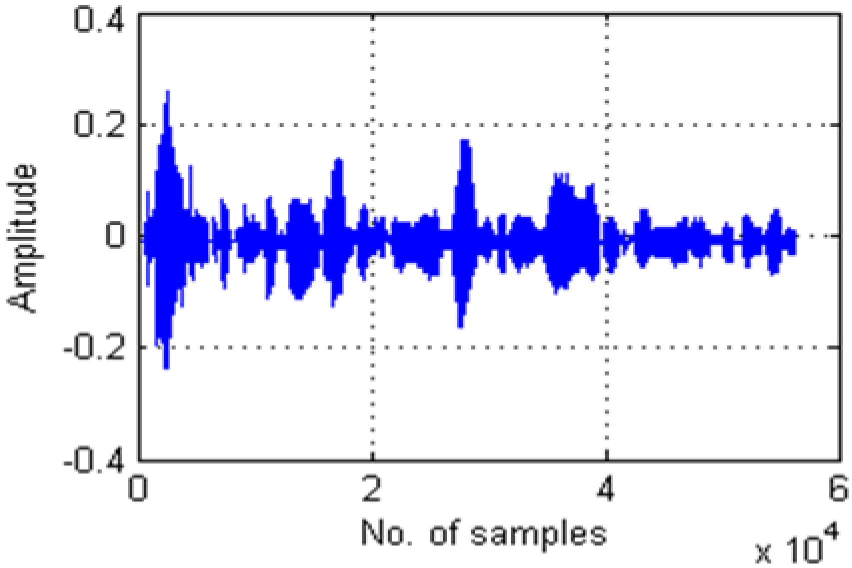


Figure 5.6 Typical output of the overlapping moving window (of fixed size) for the running text experiment.

The distribution based ‘Mahalanobis distance’ method depends on the nature of the noise distribution. If the distribution is not Gaussian in nature it fails completely. For the running text experiment the Mahalanobis distance method showed moderately good results. However it failed completely for some of the samples tested for the combination lock number experiment as shown in figure 5.7, and Table 5.1 (the high percentage distortion 20.91%). Figure 5.8 below shows a typical result of the Mahalanobis Distance based method for the running text experiment.

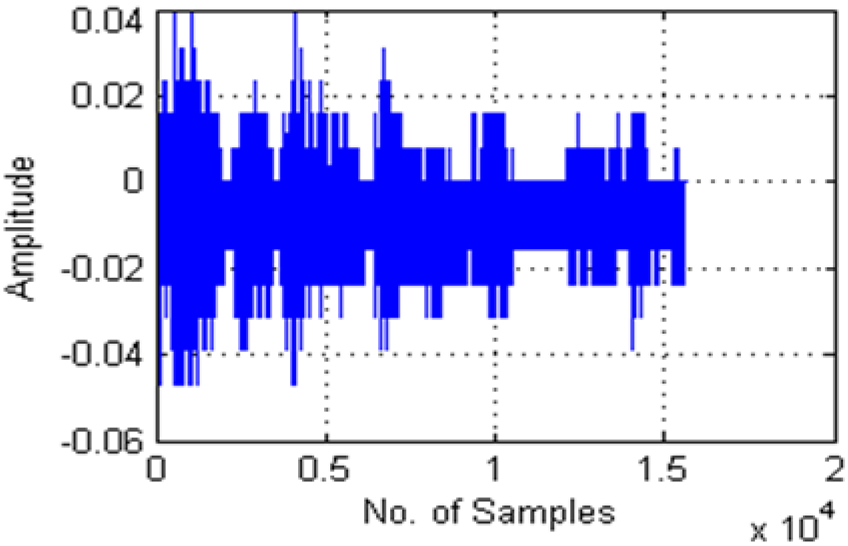


Figure 5.7 Output of the Mahalanobis Distance based method for one of the Combination lock number experiments where it failed completely.

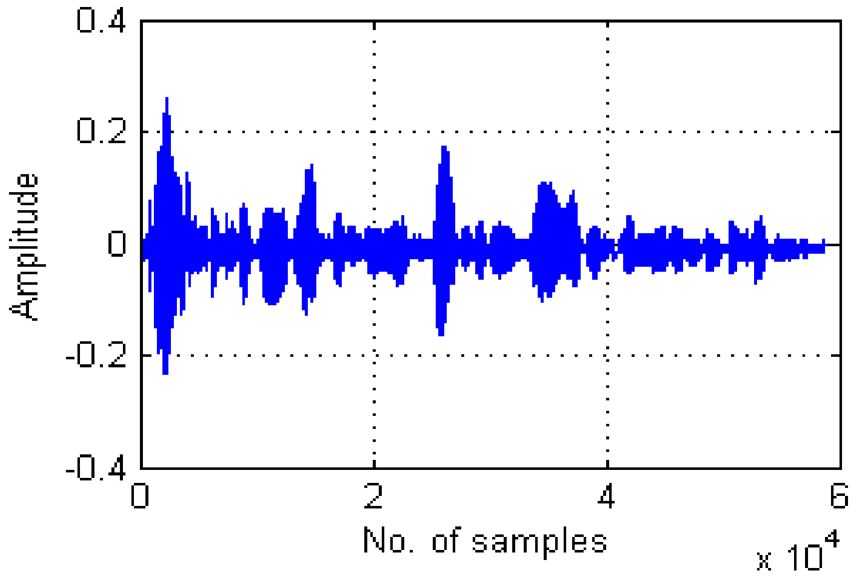


Figure 5.8 Typical output of Mahalanobis Distance based method for the running text experiment.

One of the outlier-detection based strategies the '3 σ edit rule' fails completely for both the experiments showing a very high percentage distortion (as shown in Table 5.1 and Table 5.2). Figure 5.9 and figure 5.10 below show the typical results of the '3 σ edit rule' based method.

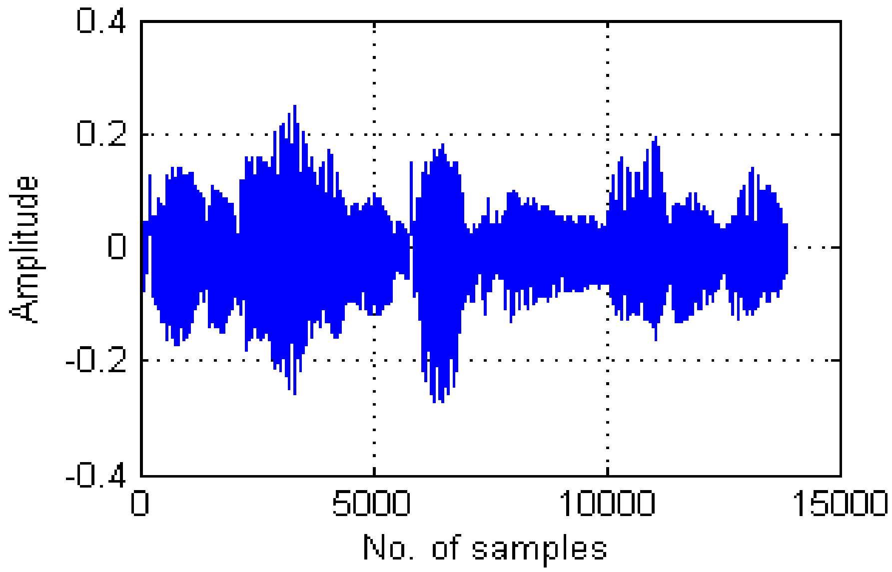


Figure 5.9 Typical output of the '3 σ edit rule' based method for the combination lock number experiment.

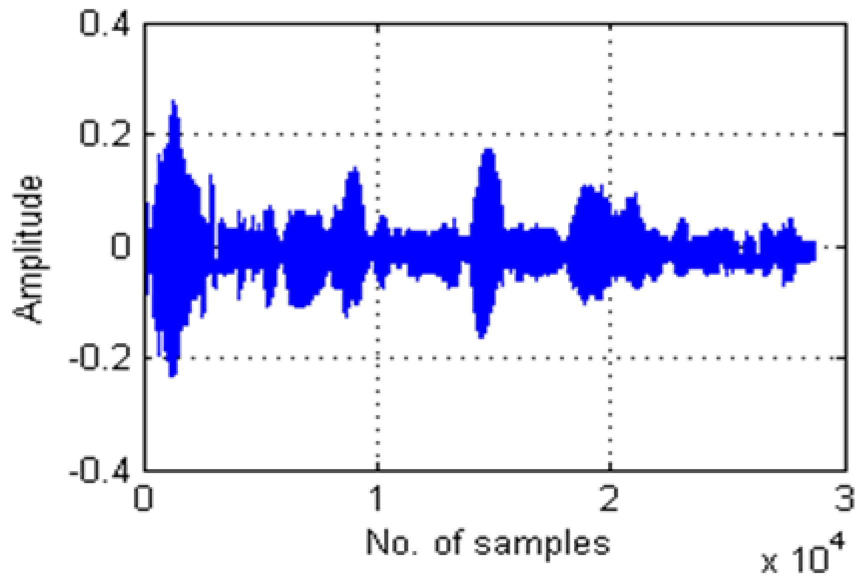


Figure 5.10 Typical output of the '3 σ edit rule' based method for the running text experiment.

However, the outlier-detection based strategy, the Hampel Identifier, show good results for both the experiments. Thus it is observed from the results obtained that the Hampel Identifier based method can be used to segregate the voiced and unvoiced/silence portions of the speech signals. Note that the success rate of this strategy depends on the fact that one class of data should ideally be a small fraction of the other class. Figure 5.11 and Figure 5.12 below show the typical results of the Hampel Identifier based method.

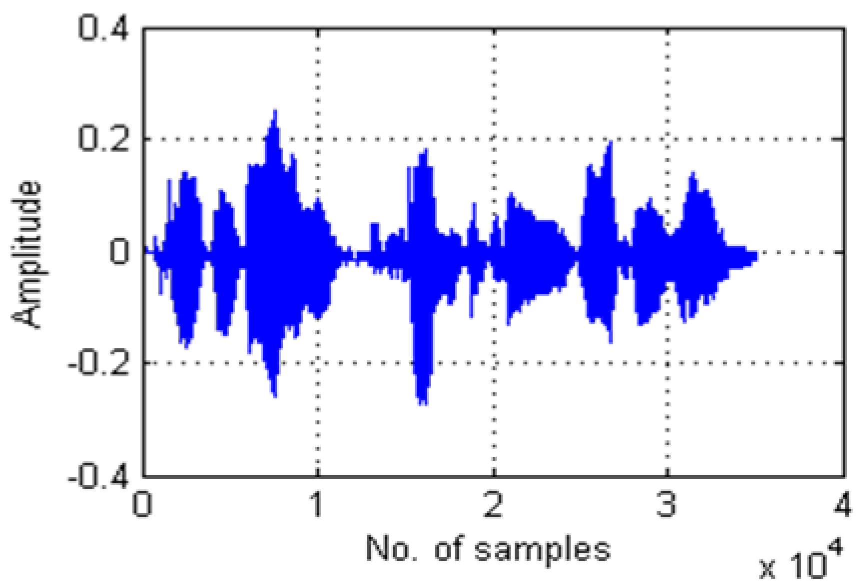


Figure 5.11 Typical output of the Hampel Identifier based method for the combination lock number experiment.

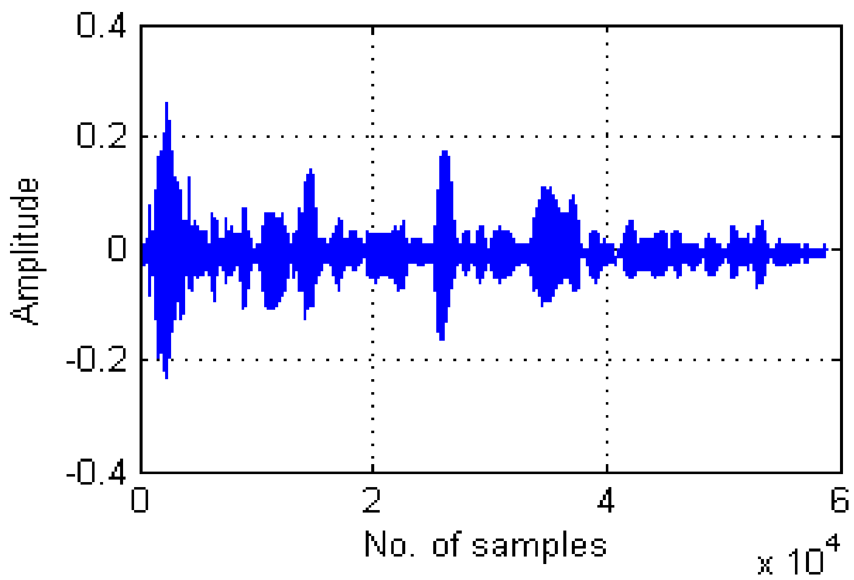


Figure 5.12 Typical output of the Hampel Identifier based method for the running text experiment.

5.3 Acoustic feature extraction and analysis process

An audible speech signal formed by producing a meaningful language phoneme; can be interpreted, identified and distinguished by the native speakers of that language quite comfortably. In order to understand how this perception and recognition process works in humans, it is essential to understand how sounds are produced, perceived, and communicated through the air (Moore, 2008; Patterson, 2007). The articulatory-phonetics, a subfield of phonetics, studies the sound production process; whereas the acoustic-phonetics a subfield of phonetics studies aspects of the physical characteristics of the sounds produced (Stevens, 2000).

Sound travels through the air by causing pressure variations in the air at the speed of 34400 cm/s (1130 feet per second) approximately (Raphael, 2006). Acoustically sounds produced in the same environment may exhibit different properties, such as the sounds can be high or low in pitch, they can be soft or loud or can differ in quality (Madureira, 1994). This section of the chapter discusses acoustic feature extraction techniques used throughout this research to analyze and quantify the acoustic similarities and differences present in the sounds of Sindhi and the factors that cause such acoustic differences.

5.3.1 Periodic vibrations

The duration of one complete cycle is referred to as its period, if the duration of the next and preceding cycle is same the motion of such cycles is called periodic (Olive, 1993; Johnson, 2003). The period of the sound is the time interval required for repetition of the signal. The relationship of frequency and period is an inverse function: the frequency $f = 1/t$, and the time $t= 1/f$, conversely (Olive, 1993). The speech sounds usually having periodic vibrations are called the sonorant (voiced) sounds and the rest are the voiceless (unvoiced) sounds.

5.3.2 Wavelength

The distance a sound travels by completing one cycle is called its wavelength or the wavelength is the distance interval occupied by one complete cycle of vibration (Olive, 1993). The speed of sound is an important variable while measuring the wavelength of the sound wave. The speed of sound in air is 34,400cm approximately, and the wavelength is the speed of sound divided by frequency (Johnson, 2003; Raphael, 2006):

$$\lambda = \frac{c}{f} \qquad (5.10)$$

where, in (5.10) c is the speed of sound f is the frequency and the Greek letter symbol λ is used to denote the wavelength.

The frequency of the sound wave is inversely related to the wavelength of vibration as the frequency of sound increases the wavelength of vibration will decrease (Johnson, 2003; Olive, 1993). Determining the wavelength of the vibration is an important variable in determining the resonant frequencies of the vocal tract; a tube of non-uniform cross sections open at one end (lips) and closed at the other (glottis) (Johnson, 2003; Parsons, 1987). The lowest resonant frequency of a resonator tube has the wavelength, four times the length of the resonator tube. If the length of the vocal tract for an adult male is 17 cm approximately; than the wavelength for its lowest resonance equals to: 4x17= 68 (Raphael, 2006). The resonances of the vocal tract are called formants which are discussed in the next section and usually labelled as the F_1 (first formant) and successively F_2, F_3, F_4 and F_5.

5.3.3 Vocal tract resonances

The resonant frequencies of an acoustic resonator tube (the vocal tract open at one end (lips) and closed at the other (glottis); can be calculated if its length is known. The resonator tube open at one end and closed at the other, resonate at the odd multiples of its lowest resonant frequency i.e the first formant frequency labelled as F_1 (Parsons, 1987). The second resonance frequency (F_2) occurs, three times the resonant frequency of the first and the third (F_3) is five times the first, fourth is 7 times the first and so on (Johnson, 2003). The relationship between the formant number and the odd multiples can be defined as:

$$2n - 1 \qquad (5.11)$$

where, in (5.11) *n* is the number of formant.

The frequency of the sound wave is inversely related to the wavelength of vibration discussed in section 5.2.2; we can derive this relationship from the equation (5.10) as:

$$f = \frac{c}{\lambda} \qquad (5.12)$$

If the length of the acoustic resonator tube (the vocal tract an acoustic resonator tube in an adult male is 17 cm approximately) is known then the wavelength of the lowest resonant frequency (F_1) at which such an acoustic resonator tube vibrates naturally is 4 times the length of the tube (Johnson, 2003). Given below are the equations for obtaining the frequency of first three formants, when the wavelength of the resonator tube is known (Johnson, 2003):

$$F_1 = c / \lambda_1 = c /(4L) = c / 4L \qquad (5.13)$$

$$F_2 = c / \lambda_2 = c /(4 / 3L) = 3c / 4L \qquad (5.14)$$

$$F_3 = c / \lambda_3 = c /(4 / 5L) = 5c / 4L \qquad (5.15)$$

The equations (5.13 - 5.15) are summarized in equation (5.16):

$$Fn = \frac{(2n-1)c}{4L} \qquad (5.16)$$

where, in (5.16) the n is the formant number c is the speed of sound and L is the length of the tube.

5.3.4 Windowing

The speech signals were split into the segments of 20ms (minimum) and 200ms (maximum) for acoustic analysis and the feature extraction process. Because the acoustic characteristics of the speech remain fairly stationary for a very short period of time (between 5ms to 100ms), the speech signals show change in the acoustic characteristics over a long period, when the next sound is being spoken (Flanagan, 1972; Rabiner, 1993; Saha, 2005). Splitting the speech signals over the length of 20ms or more best suits in this study; because the speech sounds of Sindhi analyzed in this study have duration less than 200ms. This segmentation of the speech signals into lengths of 20ms is achieved using the hamming window technique (Jurafsky, 2008). The implemented function hamming(L) returns the Hamming window of length L, where the length L for a 20ms window is obtained as:

$$L = \frac{Sr}{1000} \times 20 \qquad (5.17)$$

where, in (5.17) Sr is the sampling rate of an utterance.

5.3.5 Fourier transform

The obtained speech signals are transformed into the frequency domain at various stages of this study for the purpose of spectral analysis such as computing the power spectrum or the wideband spectrogram figures of the target analysis phonemes of Sindhi etc is discussed in section (5.3.6). The Fourier transform of the speech signal is achieved using the Fast Fourier Transform (FFT) algorithm (Cooley and Tukey, 1965). The DFT of the given signal x is achived by using the equation (5.18):

$$Y(k) = \sum_{n=0}^{N-1} x(n) e^{-2\pi ikn/N} \quad , \quad k = 0, 1, 2, \ldots, N-1 \qquad (5.18)$$

where, in (5.18) the DFT of the signals *x* of length *n* result in another vector of equal length *Y(k)*.

5.3.6 Spectral analysis

Spectral analysis refers to the representation of the time domain signals in the frequency domain, typically by the means of performing a Fourier transform. In the subsequent chapters of this study the spectral information of the corresponding Sindhi phonemes is presented in two ways: the power spectrum and the spectrogram. The short-time power spectrum of the input signal *x* is obtained by performing *n*-point Fast Fourier Transform; if *n* is less than *x,* then *x* is truncated; otherwise if *n* is longer; then *x* is padded with zeros. Spectrograms provide time-varying spectral representation of the speech signals such as the intensity, formant frequencies, formants bandwidth and formants motion etc. The spectrogram of the input signal *x* of length n for all the phonemes of Sindhi is computed by dividing the input signal *x* into windows of specified length *n* and the Fourier transform is applied on each window. In this study *x*, is divided into windows of 128 points as Hamming windowed signals.

5.3.7 Energy

The energy information available for a signal can be efficiently used to segregate the voiced and unvoiced parts of the speech. The segmentation between the vowel and consonant phonemes in this study rely upon the presence and absence of the energy between two types of sounds. The energy of the windowed signals can be obtained as the sum over time of the sample magnitudes defined as:

$$Energy = \sum_{t=t1}^{t2} x^2(t) \tag{5.19}$$

where, in (5.19) the t_1 is the beginning time and t_2 is the ending time of the windowed signals.

5.4 Conclusion

This chapter is devided in two segments the first segment discusses the signal pre-processing methods applied on collected raw voice samples of Sindhi and second segment mainly discuss the process of determining acoustic properties of the language sounds. In first segment three different approaches are presented for the segregation of voiced, unvoiced and silence portions of the speech signals using the statistical properties of the ambient noise. It is observed that the STE based strategies are generally efficient computationally and have shown moderately good results; however these strategies depend heavily on the selection of a suitable *ad hoc* threshold, which is manually set on a trial and error basis in this study. The STE based methods are discussed in section 5.2.1. The distribution based strategies depend on prior knowledge of the noise statistics. The Mahalanobis distance method fails for one of the experiments because the noise distribution was not Gaussian see section 5.2.2. Section 5.2.3 discuss the outlier-detection based strategies, the '3σ edit rule', fails completely for both experiments, while the Hampel Identifier based strategy gives better results and can be used for the segregation of the contaminated ambient noise signals. Thus it can be concluded that the STE based strategies and the Hampel Identifier based strategy give better results for the segregation of voiced, unvoiced and silence portions of the speech signals.

CHAPTER 6

ACOUSTICS OF VOWEL SOUNDS OF SINDHI

Chapter 6

Acoustics of Vowel Sounds of Sindhi

6.1 Introduction

Much research has been done in the past on developing techniques to study the acoustic-phonetic characteristics of the vowels, and in relating those to the other branches of phonetics (like articulatory-phonetics). It has also been noticed that vowel-sounds are studied more in comparison to consonant-sounds as the former show a steady-state acoustic pattern in the speech stream (Kent, 2002; Moore, 2003). However, to the best of our knowledge, no such comprehensive work has ever been undertaken to study the acoustics of the vowel-sounds of Sindhi. Work reported in the literature to date on the language has either been on articulatory-phonetics of the element-sounds, or on the language's writing system, grammar, dialects and history etc. In this respect, note that Raza et. al.'s work reported in their Centre for Research in Urdu Language Processing (CRULP) Annual Report (2003-2004) has been on articulatory-phonetics of the language's consonantal inventory and on acoustic-phonetics of the voiced implosives only; the authors have thoroughly compared their findings with those of Urdu. Moreover, the report is based on samples collected from the native speakers of only one district (Jacobabad) in Sindh, and hence does not include the diversity of the five distinct dialects of the province. Note here, there are five different regional dialects present in the Sindh province in Pakistan; the only other distinguishable dialect is spoken in the Rann of Kutch—a region in the western state of Gujarat, India (Jennifer, 2006). Jatoi's work, written in Sindhi, (Jatoi, 1996) in this field has been on the articulatory phonetics of the vowel- and consonant- sounds of the language, and also on the history of the language. It should also be noted that the vowel plots reported in the CRULP report and in Jatoi's book differ in the vowel positions (and, hence, in vowel qualities) significantly for the three following vowels: (i) /ə/ (ii) /ɑ/ and (iii) /ʊ/. As a part of this research work, a rigorous field study (discussed in detail in chapter III) was undertaken to collect voice samples from the native speakers of Sindh, Pakistan; all the five distinct dialects have been equally represented in the data set so that subtle and major differences in the vowel-sound characteristics can be studied. This chapter presents a thorough acoustic analysis of the vowel sounds of Sindhi; for which voice samples were collected from native male speakers of the five distinct regions of Sindh province of Pakistan.

6.2 Vowel system of Sindhi

Vowels are generally difficult to define in terms of articulatory-phonetics because the main articulatory organs involved in changing the shape of the vocal tract during the production of vowels are the tongue and the lips only (Jones, 1969). The articulator tongue during the formation of vowel sounds, move against the other articulators; for example, the body of the tongue is moved against the hard palate for the production of high front vowels and against the soft palate for the production of high back vowels. The author (Jatoi, 1996) presented three articulatory (tongue, lips and jaw) description of the Sindhi vowels. The articulatory descriptions in this study are based on the Jatoi's work and the definitions available in the literature for English on the subject of articulatory-phonetics. The vowel diagram of the eight primary cardinal vowels described by (Jones, 1969) is used for the reference of vowel space on acoustic vowel plots. In this study the vowel system of Sindhi is described by reffering the eight cardinal vowels and by measuring the acoustic-phonetic properties of Sindhi vowels given in Table 6.1.

Table 6.1 Ten vowel sounds of Sindhi.

IPA symbol	/i/	/ɪ/	/e/	/ɛ/	/ə/	/ɑ/	/ɔ/	/o/	/ʊ/	/u/
Sindhi Symbol	اِي	اِ	اي	اَي	اَ	آ	اَو	او	اُ	اُو

A set of acoustic properties for each vowel of Sindhi measured across all five dialects (15 speaker samples per dialect adding up to 75 speaker voice samples in total). The main acoustic parameters measured are: the mean fundamental frequency (F0), the mean of the first four formants (F1, F2, F3, and F4), and the mean vowel duration. The formant frequencies for the first four formants were obtained using the Burg LPC

(Linear Predictive Coding) algorithm and stored in .txt file format for the subsequent analysis process.

6.3 Vowel duration

Phonologically the vowel system of Sindhi can be divided into two classes according to the vowel length contrast: (i) short (lax) vowels and (ii) long (tense) vowels. The vowel durational differences have been observed for the three long and short vowel pairs of the Sindhi vowel system; these pairs are: (i) /i/ and /ɪ/ (ii) /u/ and /ʊ/ and (iii) /ɑ/ and /ə/. The vowel duration for each recorded vowel utterance in our data set has been measured in two different ways across the five regional dialects. In the first method, the vowel samples have been trimmed by excluding the leading and trailing silence and the contaminated background noise signals from the actual vowel signals, and then the vowel duration calculated. This is achived by utilizing the STE and Hampel Identifier based pre-processing methods as discussed by (Keerio, 2009). In the second method the spectrograms of the vowel utterances are computed and the vowel starting and ending positions are marked manually through the visual inspection of the spectrograms. The two obtained durations for each vowel are compared, and the mean duration is assigned to a target Sindhi vowel.

It is observed that speakers of the same geographical region when produce same vowel; the utterances result in different vowel durations, this is because some speakers speak faster and some speak slowly (Rosner, 1994). In this situation it is barely possible to conclude whether the vowel lies in the class of long or short vowels of the inventory from the computed raw vowel durations. The vowel normalization procedure is adopted to determine whether a vowel is a long or a short vowel as described by (Hongyan, 2007; Lobanov, 1971). In the normalization procedure the z-scores for the entire Sindhi vowel set are computed by subtracting the mean vowel duration from the raw vowel durations. The differences are then divided by the standard deviation of the entire vowel set. The mean of the obtained z-scores for each vowel is then computed; a negative z-score for a vowel categorizes the vowel as a short vowel, and a positive z-score categorizes the vowel as a long vowel. The vowels /ɪ/, /ʊ/, and /ə/ of the Sindhi vowel system produced negative z-scores, and are classified as short vowels; whereas the vowels /i/, /u/, /a/, /æ/, /ɔ/, /o/, and /e/ produced positive z-scores, and are classified as long vowels of the inventory; the mean vowel durations for ten vowels of Sindhi are shown in figure 6.1.

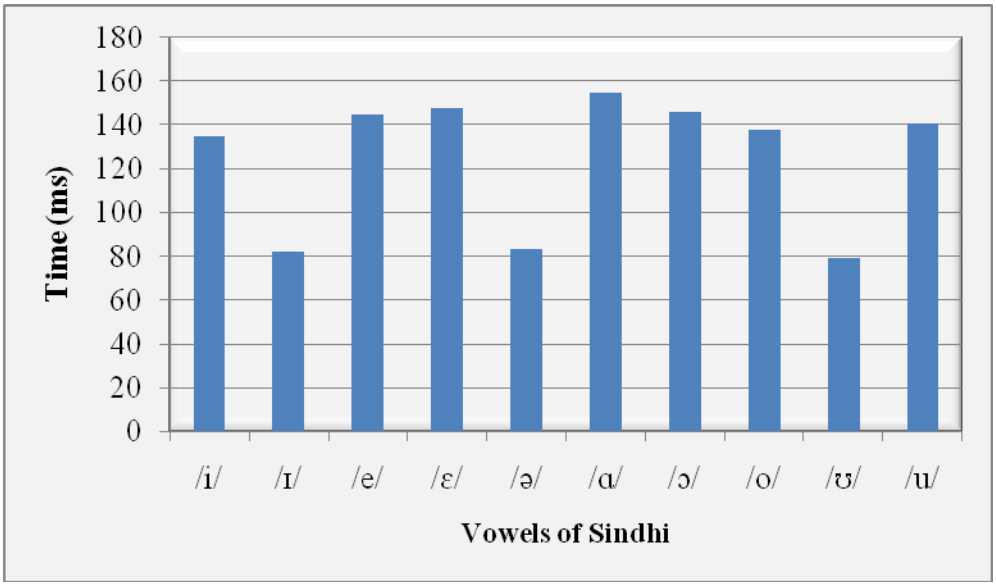


Figure 6.1 The mean vowel duration of the ten vowels of Sindhi.

6.4 Formant plots of Sindhi vowels

Formants provide the fundamental cues to analyze and quantify the perceived vowel quality measures. Formants are usually present in all the voiced phonemes of a language such as vowels, nasals, and approximants etc; however in vowels they are strong and relatively stable over time (Moore, 2003). For acoustic-phonetic analysis the first two formants are considered sufficient; however for speech recognition and synthesis three to five formants are required (Parsons, 1987). In this work, acoustic analysis of the Sindhi vowel system relies on the first two formants. Formants may vary for the same vowel between speakers of different age and sex; they may also vary in more than one instance of the vowel utterance for the same speaker; however that variation may produce a subtle difference in the vowel quality. This variation is due to the fact that the vocal tract during the production of vowels acts as a tube of non-uniform cross sections, approximately 17 cm in length for adult males (Parsons, 1987). The frequencies at which it resonates naturally occur at the odd multiples of 500 Hz approximately (Parsons, 1987; Raphael, 2006). However, since the cross section is non-uniform, the vocal tract resonant frequencies are not equally spaced (Parsons, 1987); therefore it is hardly possible to associate a constant frequency value to these resonances known as formants. However the formants can be defined within the range of frequencies i.e. first formant (F1) ranges between 200-800 Hz, second formant (F2) ranges between 600-2800 Hz and third formant (F3) ranges between 1300-3400 Hz for an adult male speaker (Parsons, 1987). Phoneticians agree that the vowel quality can be quantified acoustically by measuring the lower two resonant frequencies (F1 and F2) of the vocal tract. A variety of published literature is available on the subject of vowel quality and intelligibility measures based on the analysis of the first two formants.

The two basic articulatory characteristics that roughly correspond to the first two formants are: (i) the tongue body displacement in mouth (the height and backness) and (ii) the lip rounding (Ladefoged, 1993; Pfitzinger, 2003). The F1 is closely correlated to the perceived vowel height (Raphael, 2006) in a way that lower the F1 value the higher is the tongue body in the mouth, and the converse is also true. Similarly, the F2 is closely related to the perceived tongue backness and the lip rounding (Raphael, 2006) in a way that the lower the F2 value the more backward is the tongue body in the mouth, and the lips are rounded; conversely the higher the F2 value the more forward is the tongue body in the mouth and the lips remain either neutral or widely spread. There are many languages for which acoustic vowel plots have been drawn based on the first two formants and comparative conclusions are made for their vowel system. To the best of our knowledge no such acoustic phonetic vowel plot based on the first two formants for Sindhi across five language dialects is available in the literature. The purpose of the field study that we have conducted and the acoustic-phonetic analysis carried out on the captured voice samples provides this knowledge and data for the research community. Shown in Table 6.2 below are the mean formant frequencies of the first four formants, the mean fundamental frequency, the mean duration and the standard deviation of ten Sindhi vowels across the five dialects of the Sindh province of Pakistan. The raw formant values for ten monophthong vowels of Sindhi are given in appendix C.

Table 6.2 The mean and STD of the duration, F0, F1, F2, F3 and F4 frequency values in Hz, obtained from the voice samples of 75 adult male speakers of Sindhi.

Dialects		*Middle*		*Utradi*		*Thareli*		*Lari*		*Lasi*	
Vowel	Parameter	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
/i/	F0	160	-	179	-	166	-	166	-	160	-
	F1	256	24	264	28	241	30	282	19	268	22
	F2	2330	48	2341	85	2331	106	2342	96	2409	62
	F3	3033	239	3103	164	3005	271	3110	229	3113	183
	F4	3640	267	3741	73	3645	360	3752	176	3771	115
	Duration	126	14	135	18	145	22	123	17	149	27
/ɪ/	F0	163	-	167	-	132	-	160	-	128	-
	F1	324	17	330	21	325	16	323	15	326	13
	F2	2153	70	2065	81	2124	90	2120	89	2225	113
	F3	2752	250	2600	242	2720	206	2714	149	2754	209
	F4	3669	309	3694	284	3787	331	3765	169	3757	176
	Duration	81	11	89	15	75	23	95	26	71	35
/e/	F0	154	-	144	-	135	-	161	-	160	-
	F1	402	28	371	21	383	21	380	29	366	31
	F2	2105	59	2063	88	2111	83	2142	74	2178	54
	F3	2727	209	2725	186	2739	155	2730	133	2812	218
	F4	3612	269	3704	316	3682	89	3733	86	3583	342
	Duration	135	12	143	14	156	27	142	24	149	19
/ɛ/	F0	130	-	147	-	140	-	155	-	156	-
	F1	472	25	464	37	451	31	492	30	463	21
	F2	1939	75	1835	99	1866	102	1942	105	1948	122
	F3	2551	353	2593	143	2645	174	2619	122	2597	303
	F4	3421	468	3691	211	3641	209	3661	354	3408	486
	Duration	141	17	150	22	146	18	147	27	160	34
/ə/	F0	147	-	145	-	137	-	145	-	157	-
	F1	594	28	587	47	577	37	553	42	567	23
	F2	1490	58	1437	77	1386	65	1419	81	1413	75
	F3	2533	326	2667	230	2700	194	2604	166	2632	197
	F4	3582	400	3735	121	3700	122	3687	190	3591	314
	Duration	83	20	88	29	91	27	76	33	79	37

-Continued on next page -

/ɑ/	F0	132	-	170	-	151	-	140	-	145	-
	F1	755	56	714	50	700	57	714	63	700	62
	F2	1215	55	1147	77	1148	76	1148	73	1181	67
	F3	2533	285	2574	337	2681	105	2566	210	2588	181
	F4	3578	408	3450	422	3758	218	3529	323	3628	133
	Duration	152	08	155	12	148	19	165	24	159	32
/ɔ/	F0	136	-	123	-	152	-	159	-	141	-
	F1	447	28	462	31	449	36	460	34	469	34
	F2	928	45	1016	82	952	61	981	33	1019	44
	F3	2473	190	2505	190	2620	110	2532	111	2566	260
	F4	3505	331	3597	311	3548	228	3679	112	3577	138
	Duration	135	24	144	31	151	21	154	38	148	19
/o/	F0	163	-	151	-	170	-	145	-	148	-
	F1	386	28	387	24	375	20	419	18	377	29
	F2	814	36	785	53	814	73	826	31	821	72
	F3	2481	340	2509	459	2598	182	2597	50	2668	129
	F4	3437	291	3437	393	3566	170	3368	229	3543	277
	Duration	125	13	129	11	137	18	140	26	159	31
/ʊ/	F0	162	-	168	-	157	-	159	-	179	-
	F1	316	13	324	27	328	19	327	22	303	14
	F2	773	57	764	83	870	99	813	70	750	65
	F3	2618	130	2490	128	2470	255	2452	143	2577	159
	F4	3580	131	3504	278	3486	275	3432	262	3647	239
	Duration	72	18	80	21	76	27	89	23	78	29
/u/	F0	165	-	129	-	165	-	162	-	165	-
	F1	279	23	269	25	252	29	286	21	283	19
	F2	679	57	714	81	645	82	731	58	645	83
	F3	2454	245	2391	374	2540	205	2551	147	2160	407
	F4	3518	236	3416	341	3580	331	3485	230	3147	412
	Duration	135	17	142	28	151	21	146	27	135	42

The figures 6.2(a) – 6.2(e) and figure 6.3 below show the F1 x F2 scatter vowel plots for the utterance of the fifteen male speakers per dialect. The ellipses fitted to the data are drawn for each vowel category to cover one vowel data on the vowel plot. It is clear from the vowel plots (figures 6.2(a) - 6.2(e) and figure 6.3) that the vowel positions are relatively stationary across all five dialects of Sindhi. However, due to the different regional dialects a greater variation is observed in the F1 value for the open back (also called low back) vowel /ɑ/ and the mid-open (also called central) vowel /ə/ throughout the five dialects. The high front and high back vowel pairs: /i/-/ɪ/ and /u/-/ʊ/ are observed varying in the F2 frequency value throughout the five dialects. Tight vowel clusters are also evident in figures 6.2(a) – 6.2(e) across all five dialects; moreover the overlapping in the vowel space between the high front vowel pair /i/ and /ɪ/ and the high back vowel pair /u/ and /ʊ/ is obvious for four dialects shown in figures 6.2(a), 6.2(c), 6.2(d), and 6.2(e). The only exception is the *Utradi* dialect which is shown in figure 6.2(b). The neighboring front series vowels /ɪ/ and /e/, back series vowels /ʊ/, /o/ and /ɔ/ overlap for all five dialects this is shown in figures 6.2(a) – 6.2(e). The front vowels /e/ and /ɛ/ only overlap for the *Thareli* dialect, refer to figure 6.2(c). The open back vowel /ɑ/ and mid open vowel /ə/ show clear vowel quality across all five dialects of Sindhi.

The association of the corresponding vowel quality and the formant frequencies is non-linear; so the formant values are converted into the bark units, using the Bark transform formula given in (Hongyan, 2007; Traunmuller, 1990). The distance between two vowel qualities are then computed using the F1 and F2 values (in Bark units). Figures 6.4(a) – 6.4(e) below show the Bark transformed acoustic phonetic vowel plots based on F1 and F2 values for the five dialects of Sindhi. In the figures 6.4(a) – 6.4(e) the F1 values (in Bark) are plotted on the ordinate (vertical axis) and the F2 values (in Bark) are plotted on the abscissa (horizontal axis); the axis values are inverted so that the maximum crosses at the origin. The four corner vowels of the Sindhi vowel system can then be selected with the help of these acoustic phonetic vowel diagrams. By drawing a line between the four corner vowels shown in figures 6.4(a) – 6.4(e); starting from the high front vowel /i/ to /ɛ/, from /ɛ/ to /ɑ/, from /ɑ/ to /u/ and finally from /u/ to back /i/, we get a four sided figure; which covers all the remaining vowels of the Sindhi vowel system. The four sided shape shown in figures 6.4(a) – 6.4(e) has roughly similar shape of the IPA vowel quadrilateral. If we draw a line between two corner vowels i.e. between /ɛ/ and /u/; it will divide the four sided irregular quadrilateral shape in two triangles. Where the first triangle is formed by vowels, starting from /i/ to /ɛ/, from /ɛ/ to /u/ and finally back from /u/ to /i/; the second triangle includes the vowels starting from

/ɛ/ to /ɑ/, from /ɑ/ to /u/ and finally back from /u/ to /ɛ/. The area for the Sindhi vowel system in this way is calculated by calculating the area of the irregular quadrilateral by adding the area of the two triangles. The area of the Sindhi vowel system is shown in Table 6.3 below.

Table 6.3 The area of Sindhi vowel system (in Bark^2 units)

Area of	*Middle*	*Utradi*	*Thareli*	*Lari*	*Lasi*	*Sindhi vowel System*
All ten vowels	18.103	18.2376	17.6939	16.513	17.0388	17.5169
Three short vowels	7.9446	6.3399	7.244	6.3595	7.60149	7.08718

The four corner vowels of Sindhi /i/, /ɛ/, /ɑ/, and /u/ shown in figures 6.4(a) - 6.4(e) can be used as fixed reference points for the description of the other six vowels of the inventory according to the eight cardinal vowels of English described by (Jones, 1969) and the four extreme reference points for vowels discussed in the IPA hand book (IPA, 1999).

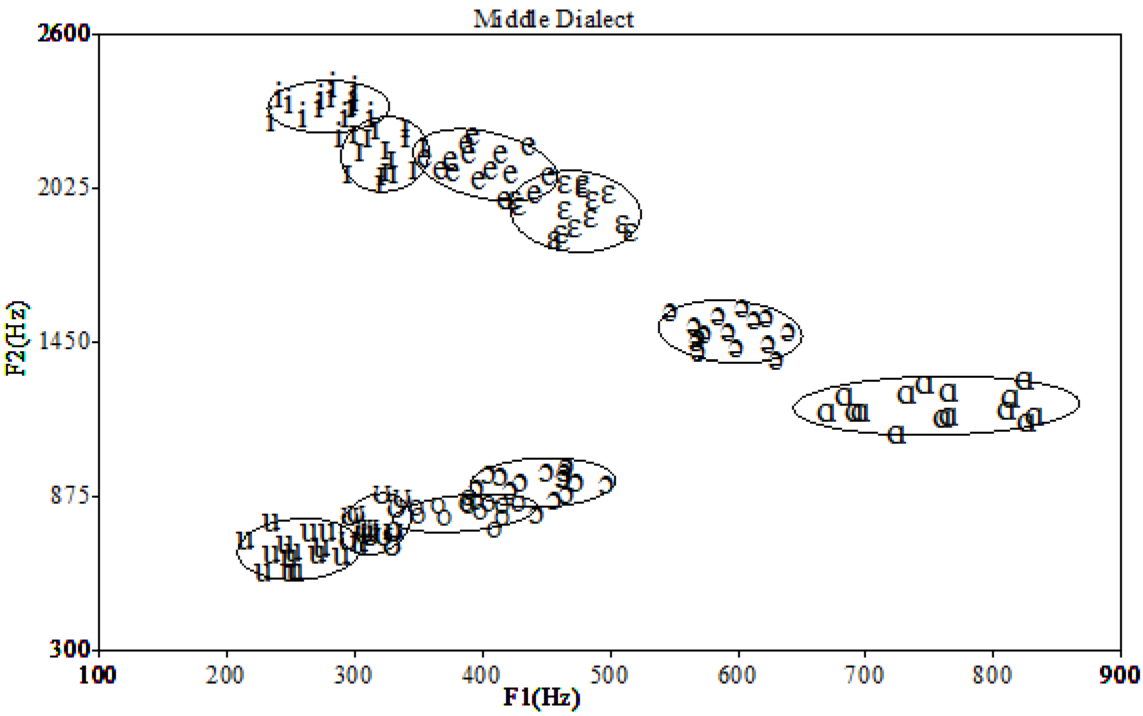


Figure 6.2(a) F1 x F2 scatter vowel plot for fifteen adult male speakers of "Middle" dialect of Sindhi. Ellipses are drawn fitting the vowel space for each vowel.

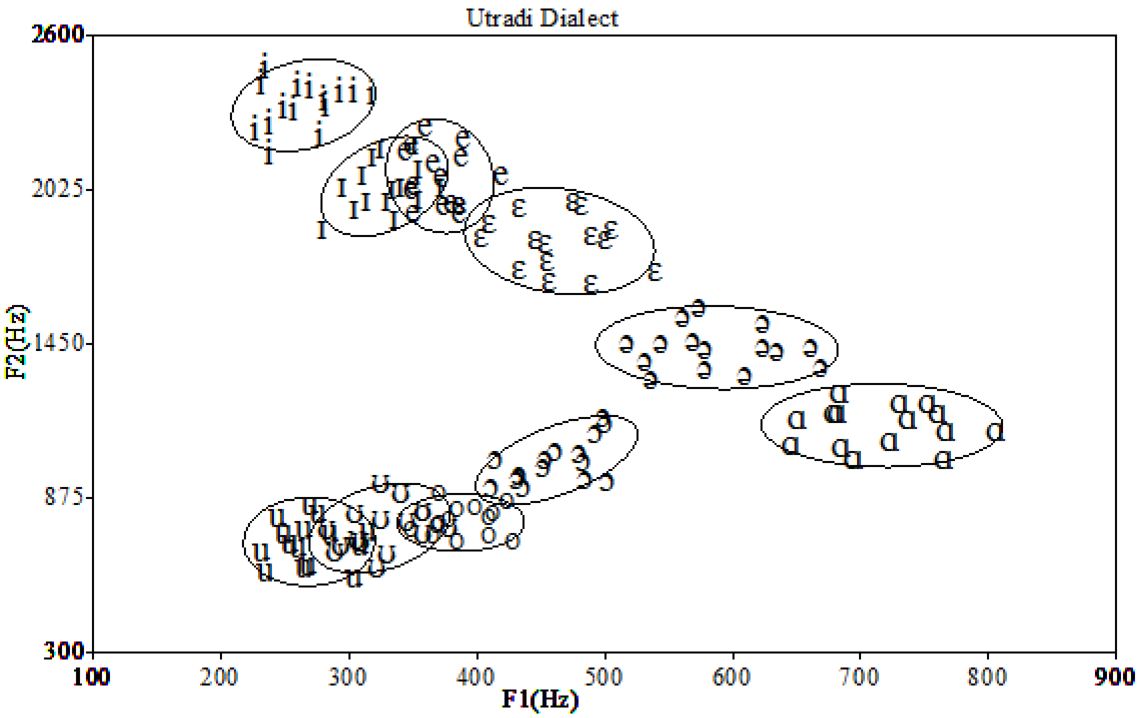


Figure 6.2(b) The F1 x F2 scatter vowel plot for fifteen adult male speakers of "Utradi", dialect of Sindhi. Ellipses are drawn fitting the vowel space for each vowel.

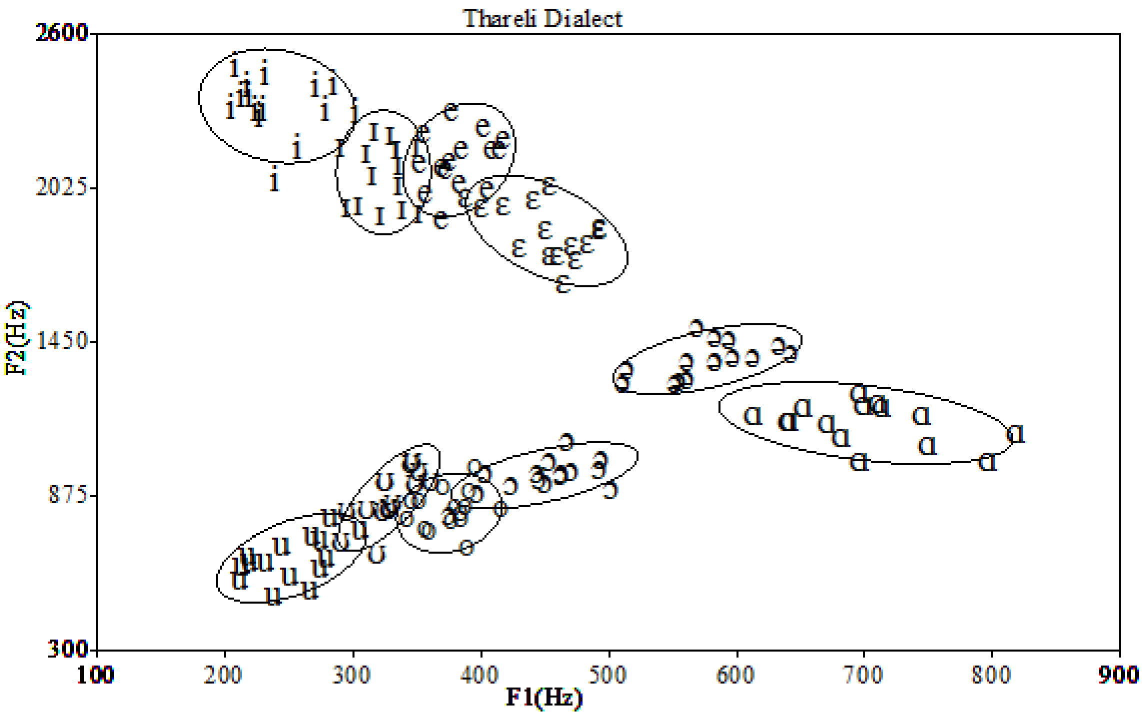


Figure 6.2(c) The F1 x F2 scatter vowel plot for fifteen adult male speakers of "Thareli", dialect of Sindhi. Ellipses are drawn fitting the vowel space for each vowel.

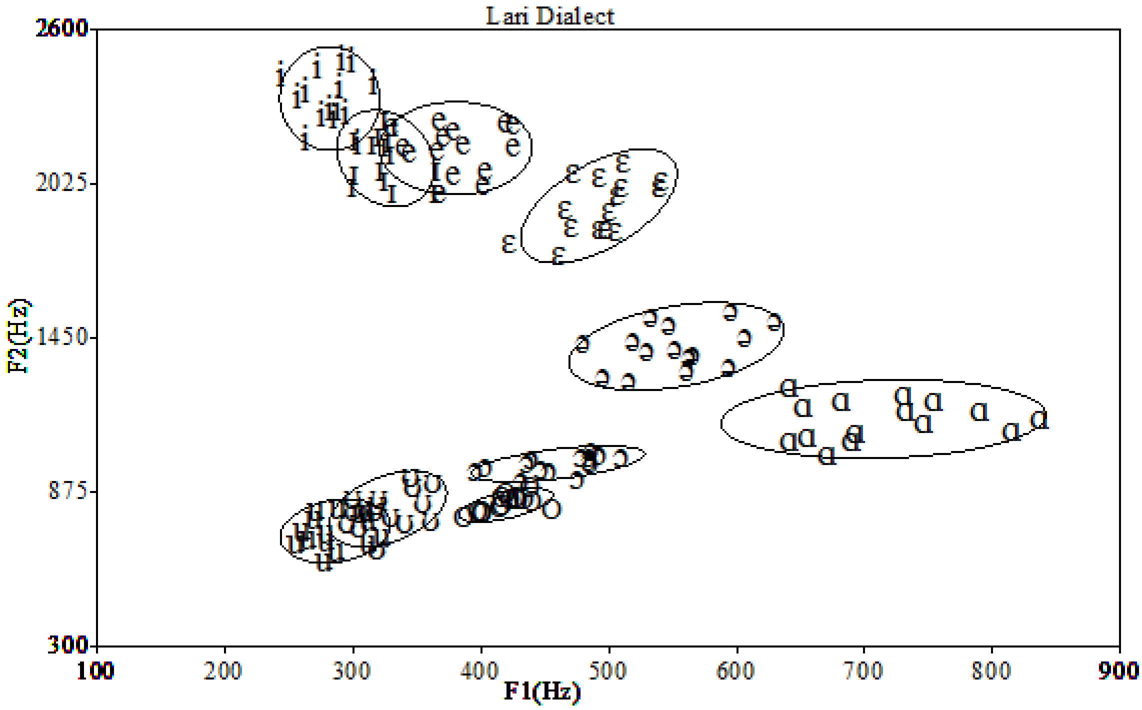


Figure 6.2(d) The F1 x F2 scatter vowel plot for fifteen adult male speakers of "Lari", dialect of Sindhi. Ellipses are drawn fitting the vowel space for each vowel.

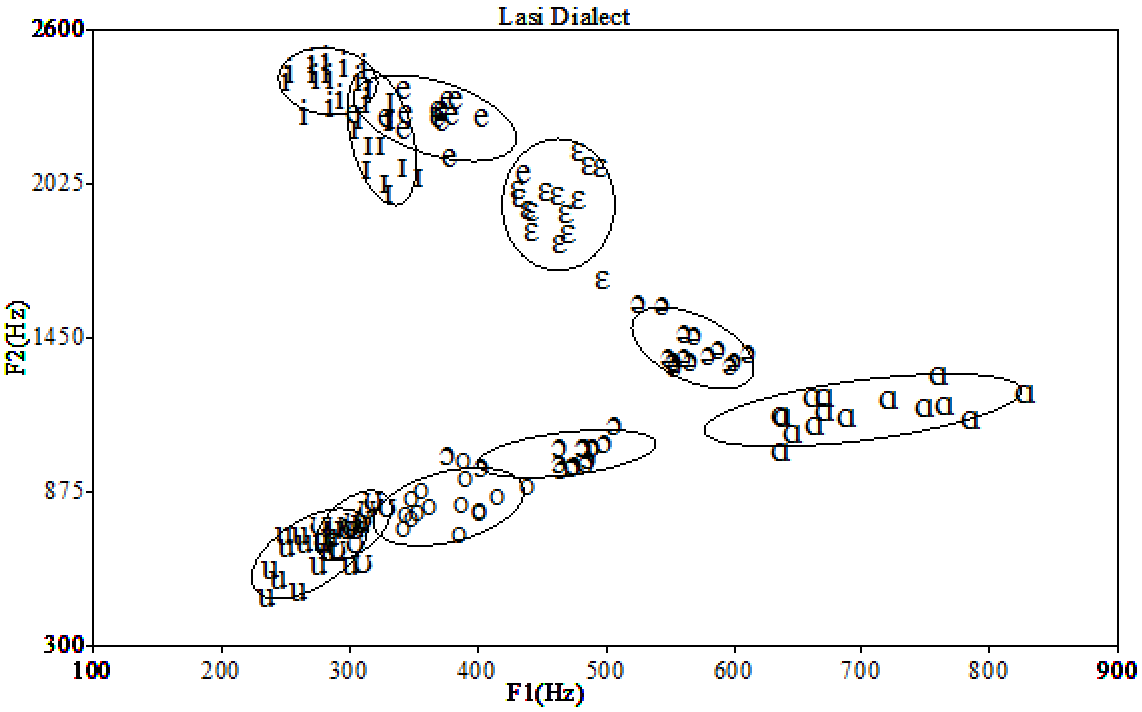


Figure 6.2(e) The F1 x F2 scatter vowel plot for fifteen adult male speakers of "Lasi", dialect of Sindhi. Ellipses are drawn fitting the vowel space for each vowel.

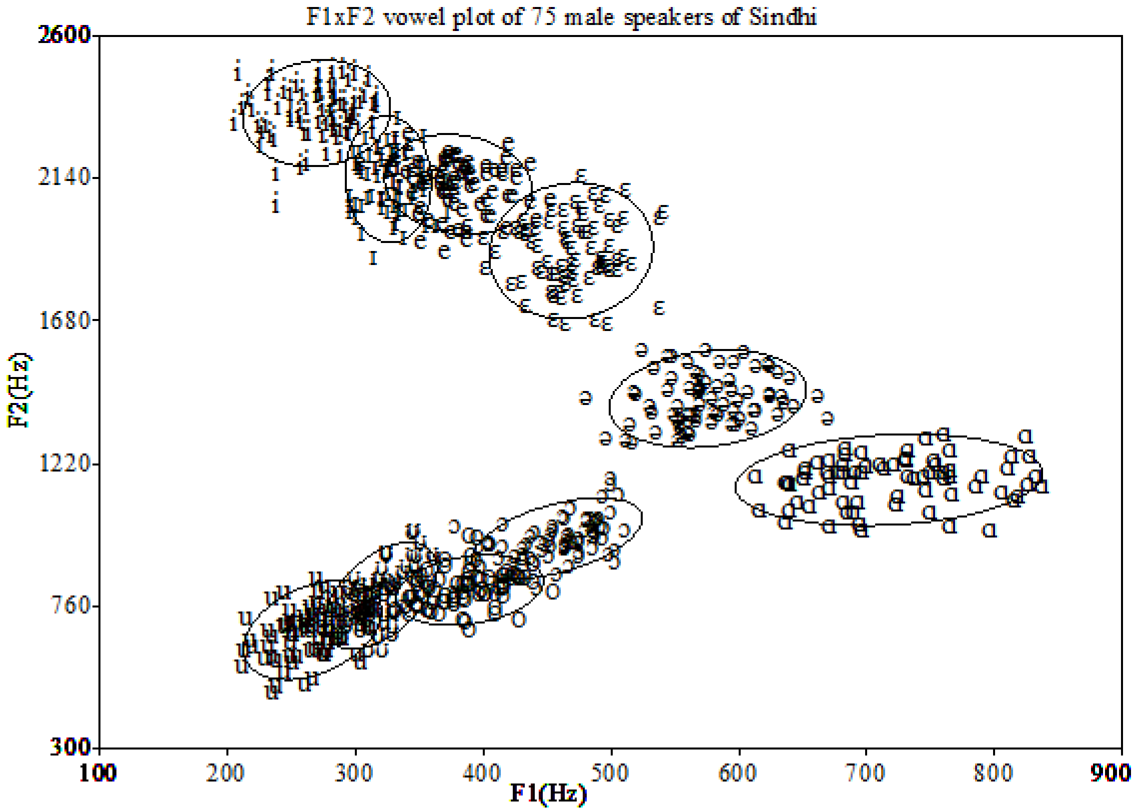


Figure 6.3 The F1 x F2 scatter vowel plot for 75 adult male speakers of Sindhi language. Ellipses are drawn fitting the vowel space for each vowel.

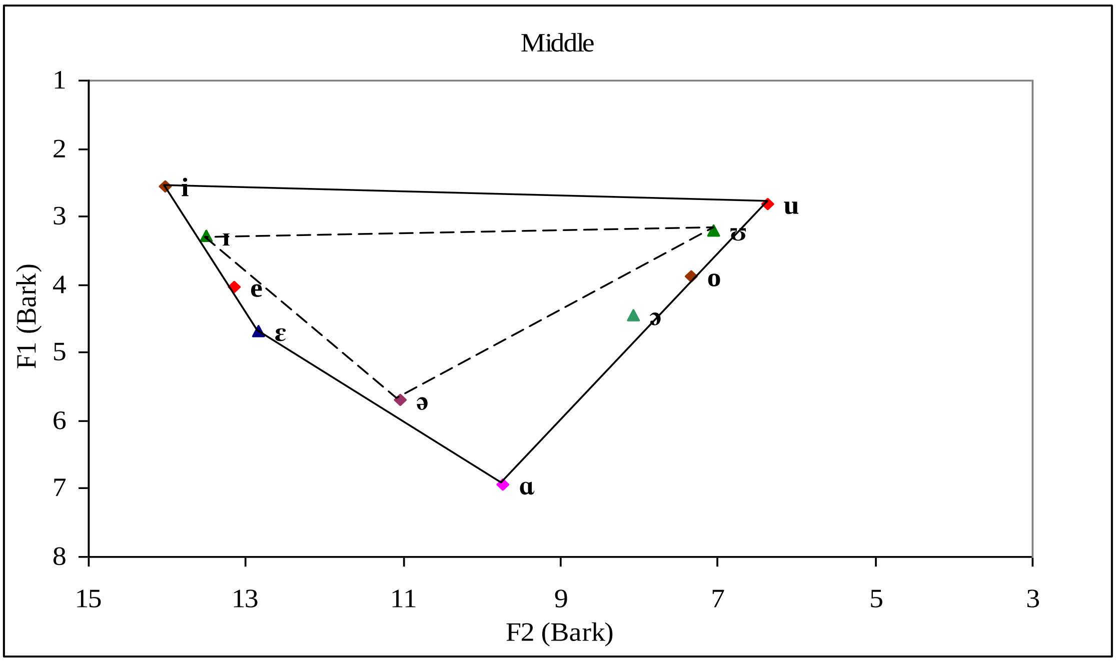


Figure 6.4(a) Acoustic vowel plot of the *Middle* dialect of Sindhi, the axis values are inverted. The solid line connects the four corner vowels and the dashed line connects the three short vowels of the inventory..

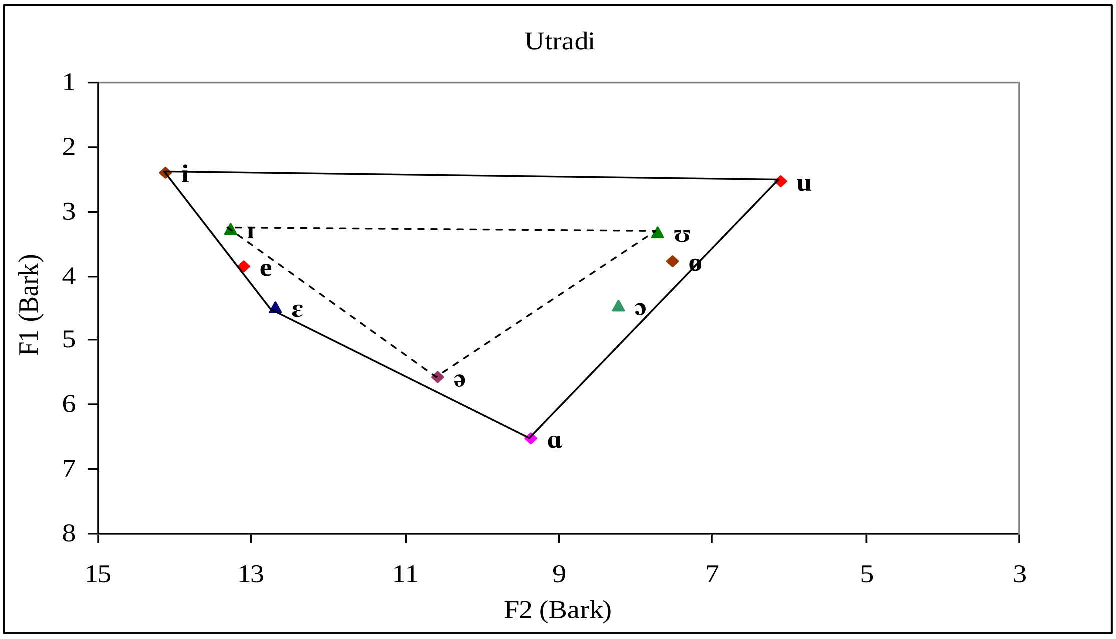


Figure 6.4(b) Acoustic vowel plot of the *Utradi* dialect of Sindhi, the axis values are inverted. The solid line connects the four corner vowels and the dashed line connects the three short vowels of the inventory.

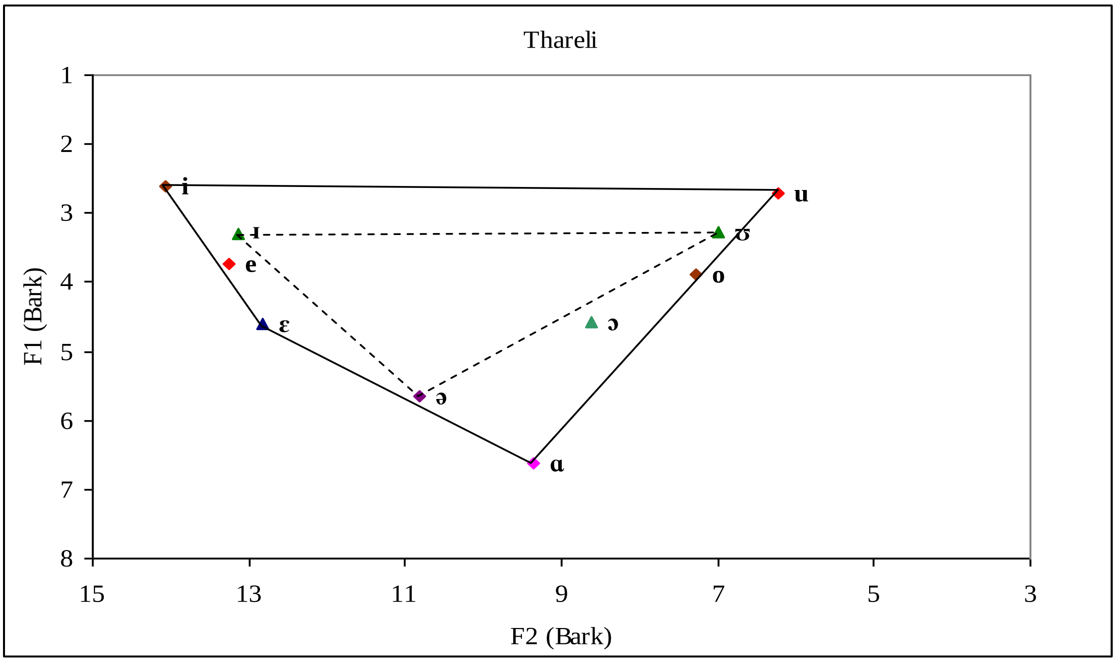


Figure 6.4(c) Acoustic vowel plot of the *Thareli* dialect of Sindhi, the axis values are inverted. The solid line connects the four corner vowels and the dashed line connects the three short vowels of the inventory.

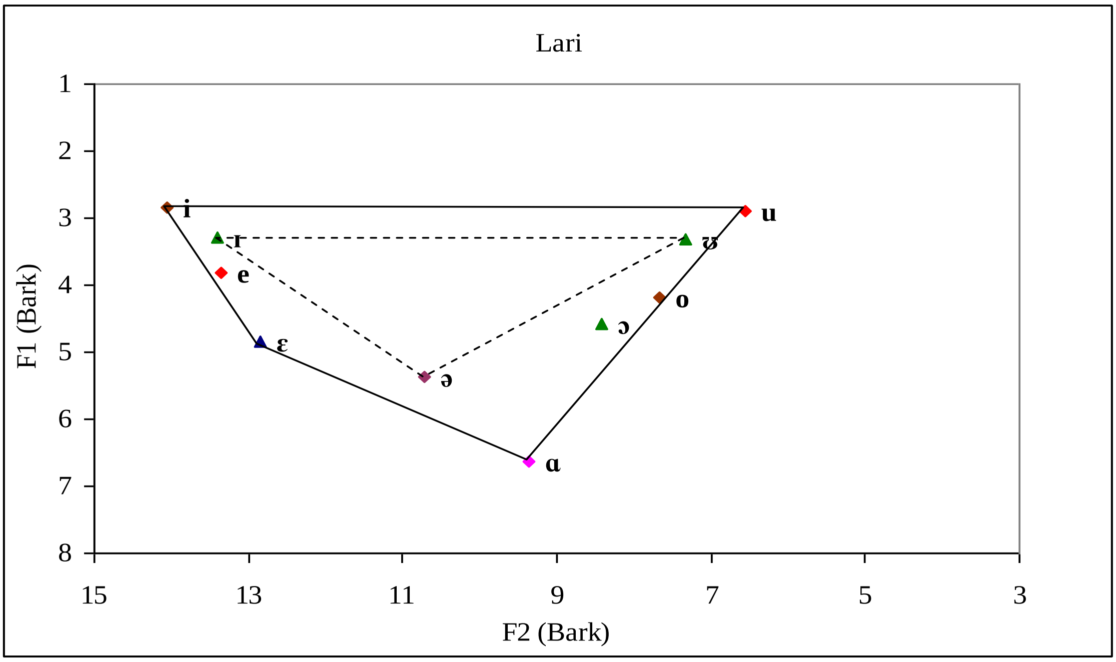


Figure 6.4(d) Acoustic vowel plot of the *Lari* dialect of Sindhi, the axis values are inverted.The solid line connects the four corner vowels and the dashed line connects the three short vowels of the inventory.

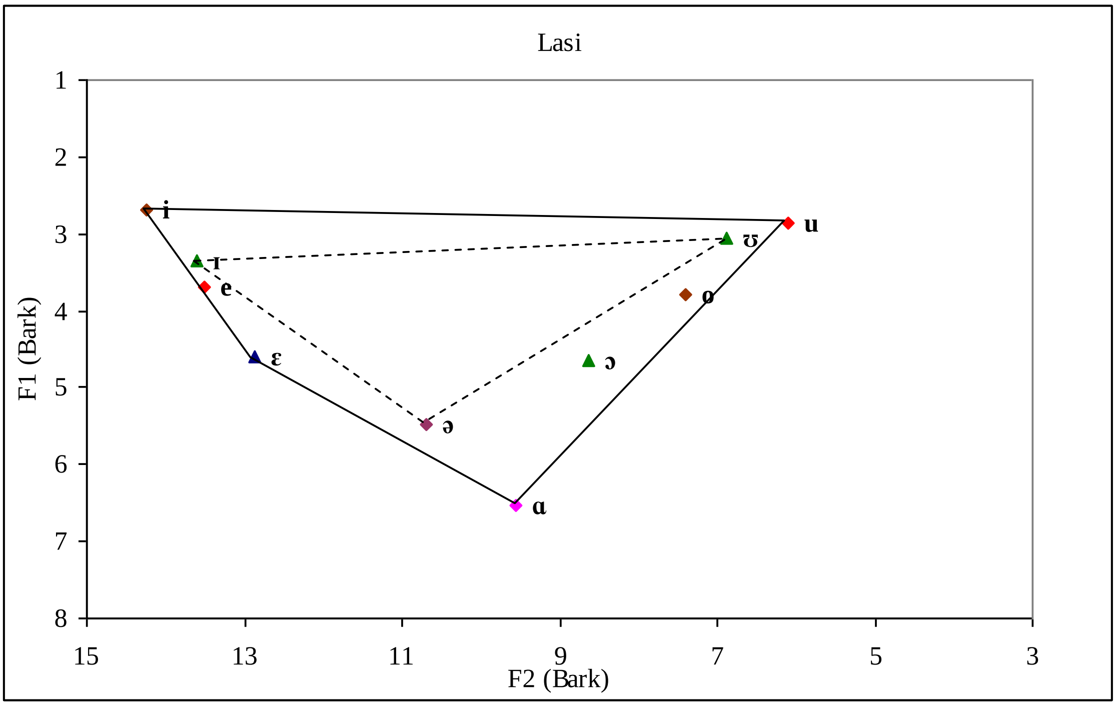


Figure 6.4(e) Acoustic vowel plot of the *Lasi* dialect of Sindhi, the axis values are inverted. The solid line connects the four corner vowels and the dashed line connects the three short vowels of the inventory.

In order to determine the relative vowel quality change in vowel space a simple formant ranking method is used, which provides a simple and direct comparison in terms of individual formant frequency values (Martland, 1996). Figure 6.5, illustrates the vocal tract shape for eight English vowels; this is used for the articulatory reference of the tongue position in the vocal tract during the production of vowel sounds. The ten Sindhi vowels are separated as front and back series vowels by computing the mid point value for F2 (in Bark); because F2 is related to the perceived vowel quality as front or back vowels. The vertical line shown in figure 6.6 is drawn for the mid point value of F2 that separates the two vowel groups (front, back) of the Sindhi vowel system.

The two vowel groups are then ranked separately, first by F1 values (in Bark) and second by F2 values (in Bark) refer Tables 6.4, 6.5 and 6.6. In Table 6.4, the five front series vowels are ranked in an ascending order by F1 values and in Table 6.5 the five back series vowels are ranked in an ascending order by F1 values. In Table 6.6, the ten vowels are ranked by the F2 values in descending order for all five dialects of Sindhi. The vowel at index one in Table 6.4 and 6.5, is an extremely high vowel among both front and back series vowels and the vowel at index one in Table 6.6, is an extremely forward vowel, and the vowel at index ten in Table 6.6, is an extremely backward vowel. This association is made according to the fact that the first formant is mainly responsive to the perceived vowel height, and the second formant is mainly responsive to the perceived vowel backness (Ladefoged, 1993; Martland, 1996; Pfitzinger, 2003; Raphael, 2006); hence the relationship can be established between the ranked vowel positions in Tables 6.4, 6.5 and 6.6 and the relative vowel quality showed in the acoustic vowel plot figures 6.4(a) - 6.4(e).

Analysis of the ranked tables for F1 and F2 values enable us to classify the Sindhi vowel system comfortably. The vowel /i/ retains number one position in Tables 6.4 and 6.6, which means that the tongue body is raised to its maximum and displaced far forward in the oral cavity for this vowel throughout the five dialects of Sindhi articulatory tongue position for this vowel is shown in Figure 6.5. The vowel /ɪ/ is at index position two in Table 6.4 and at index position two in Table 6.6 except for the *Lari* dialect, which means that the tongue body is raised up in the direction of the hard palate and displaced to the forward position the articulatory tongue position for this vowel is shown in Figure 6.5. The vowel /e/ is at index position three in Tables 6.4 and 6.6 except for the *Lari* dialect. The speakers of *Lari* dialect displace the tongue body further forward for this vowel compared to the vowel /ɪ/, refer Table 6.6. The vowel /ɛ/ is at index position four in Tables 6.4 and 6.6; the tongue body is raised up from the normal position to the middle position for this Sindhi vowel. This is shown in acoustic vowel plots of figures 6.4(a) - 6.4(e), the vowel lies on the mid horizontal line computed for the vowel height over F1 values (in Bark) see figure 6.6. The vowel /ə/ is at index position five in Tables 6.4 and 6.6, which means the vowel, is a central vowel among the ten vowels of Sindhi. The body of the tongue is raised from the normal position to the middle position in terms of both height and backness for this vowel. The vowel /ɑ/ is at index position five in Table 6.5, which means the height of the tongue body is lowest for this vowel among the ten vowels and it is at index position six in Table 6.6, which means the vowel is a middle vowel see figure 6.6. The tongue articulatory position for this vowel is shown in figure 6.5. The vowel /ɔ/ is at index position four in Table 6.5 and at index position seven in Table 6.6. The tongue body is displaced backward and upward in the direction of the soft palate, for this vowel shown in figures 6.5 and 6.6. The vowel /o/ is at index position three in Table 6.5 and at index position eight in Table 6.6, except for the Thareli dialect. The tongue body is displaced backward and upward in the direction of the soft palate, for this vowel see figures 6.5 and 6.6. The speakers of the Thareli dialect displace the tongue body further backward for the vowel /o/ compared to the vowel /ʊ/ of Sindhi this is shown in Table 6.6. The vowel /ʊ/ is at index position two in Table 6.5 and at index position nine in Table 6.6 except for the Thareli dialect. The body of the tongue for this vowel is displaced backward and upward in the direction of the soft palate. The tongue articulatory position for this vowel is shown in figure 6.5. The vowel /u/ is at index position one in Table 6.5 and index position ten in Table 6.6. The tongue body is displaced far backward and upward for this vowel among the ten vowels; the tongue articulatory position for this vowel is shown in figure 6.5.

Table 6.4 Five front series vowels of Sindhi ranked in an ascending order for the F1 value (in Bark).

Index	Middle	Utradi	Thareli	Lari	Lasi
1	i	i	i	i	i
2	ɪ	ɪ	ɪ	ɪ	ɪ
3	e	e	e	e	e
4	ɛ	ɛ	ɛ	ɛ	ɛ
5	ə	ə	ə	ə	ə

Table 6.5 Five back series vowels of Sindhi ranked in an ascending order for the F1 value (in Bark).

Index	Lari	Lasi	Middle	Thareli	Utradi
1	u	u	u	u	u
2	ʊ	ʊ	ʊ	ʊ	ʊ
3	o	o	o	o	o
4	ɔ	ɔ	ɔ	ɔ	ɔ
5	ɑ	ɑ	ɑ	ɑ	ɑ

Table 6.6 Ten vowels of Sindhi ranked in descending order for the F2 value (in Bark).

Index	Middle	Utradi	Thareli	Lari	Lasi
1	i	i	i	i	i
2	ɪ	ɪ	ɪ	e	ɪ
3	e	e	e	ɪ	e
4	ɛ	ɛ	ɛ	ɛ	ɛ
5	ə	ə	ə	ə	ə
6	ɑ	ɑ	ɑ	ɑ	ɑ
7	ɔ	ɔ	ɔ	ɔ	ɔ
8	o	o	ʊ	o	o
9	ʊ	ʊ	o	ʊ	ʊ
10	u	u	u	u	u

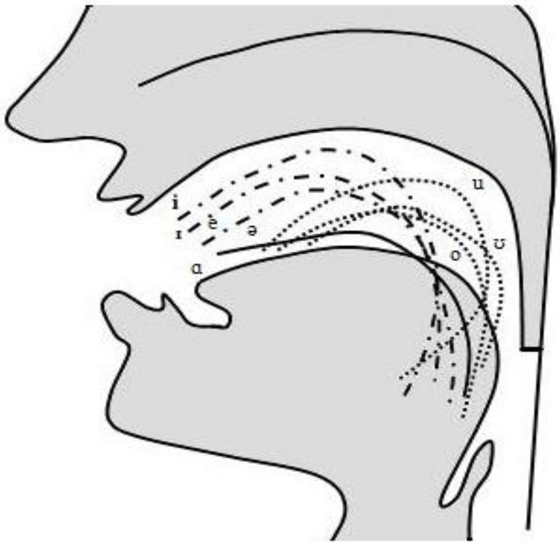


Figure 6.5 The vocal tract shape for the vowels /i/ in word 'heed', /ɪ/ in 'hid', /e/ in 'pet', /ə/ in 'away', /ɑ/ in 'father', /o/ in 'four', /ʊ/, in 'put', /u/ in 'food'.

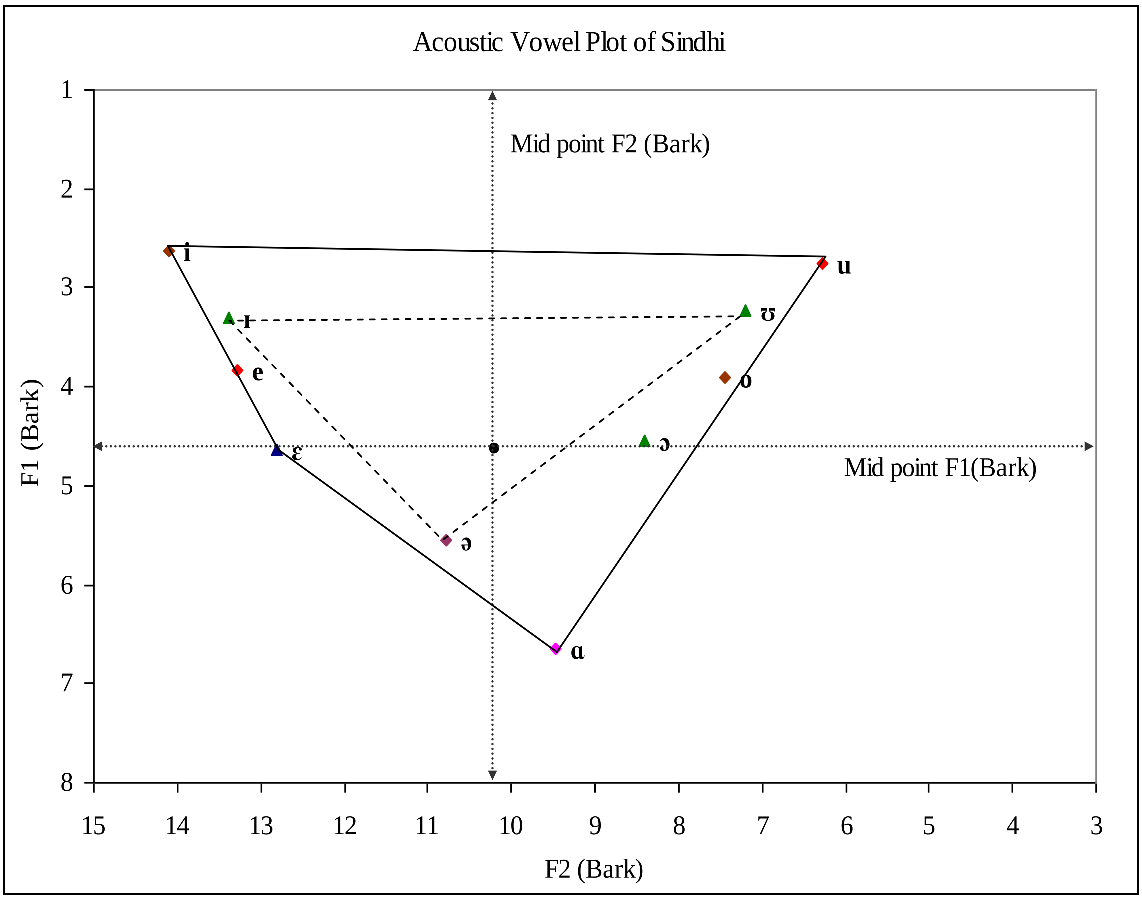


Figure 6.6 Acoustic vowel plot of the Sindhi vowel system, the axis values are inverted. The solid line connects the four corner vowels and the dashed line connects the three short vowels of the inventory.

6.5 Classification of Sindhi vowels

Vowels traditionally in articulatory-phonetics are identified referring to the displacement of the tongue body and rounding of the lips. This articulatory description is relatively unspecific in answering the questions like how exactly the tongue body is displaced forward, backward and the degree of height and lips rounding? The above discussion in this chapter answers such questions by incorporating the acoustic features of the language vowels in addition to the available articulatory knowledge. In this section we will show how acoustic-phonetic knowledge of the vowels in Sindhi enables us to provide accurate description and classification of Sindhi vowels.

The vowel quality and intelligibility depends on the shape of the vocal tract open at one end and closed at the other end for the passage of air during the production of vowel sounds (Jones, 1969; Raphael, 2006; Rosner, 1994). The size and shape of the vocal tract may vary for the speakers of different age and sex; however the shape of the vocal tract for the same speaker varies for the production of different vowels**.** The main articulators involved in changing the shape of the air passage in the oral cavity are the tongue and the lips (Jones, 1969). Linguistically, in this study, Sindhi vowels are classified referring to the acoustic phonetic vowel diagrams shown in Figures 6.4(a) – 6.4(e) for five Sindhi dialects and Figure 6.6 for the whole Sindh region. Articulatory the movement of the tongue body upward in the direction of the hard palate is relative to the high front or close front vowels; whereas the movement of the tongue body in the direction of the soft palate is relative to the high back or close back vowels. The tongue body movement downward is relative to the open vowels of Sindhi. The lips can be in one of the three states: rounded, spread or neutral and the mouth is either widely opened or narrowed downed during the production of vowels in Sindhi. The following are the three articulatory description of the ten vowels of Sindhi, referring to the obtained acoustic-phonetic features of Sindhi vowel system:

i. **/i/, اِي :** this vowel has the lowest F1 value and the highest F2 value among the entire vowel system which makes this vowel a high front vowel. Articulatory: the body of the tongue is displaced far forward and upward in the direction of the hard palate shown in figure 6.5. The lips are widely spread, and the mouth is narrowed down for the passage of air. The vowel /i/ has shown the maximum separation between F1 and F2 among the ten vowels of Sindhi. The vowel has stable F1 and F2 frequencies, with most of the energy lying in the F1 band rather than the second, third and fourth formants as shown in the spectrogram figure 6.7. Since the tongue body is displaced forward and upward in the direction of the hard palate, this vowel can be classified as *high-front* or *close-front* long vowel of Sindhi.

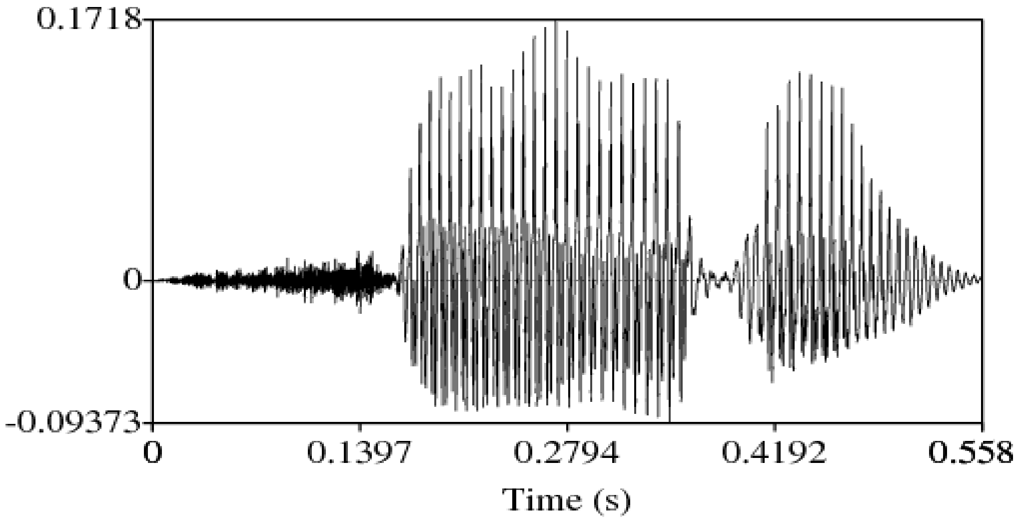


Figure 6.7(a) A typical waveform of the word utterance /sirə/ spoken by speaker of middle dialect, containing the front-close vowel /i/ of Sindhi.

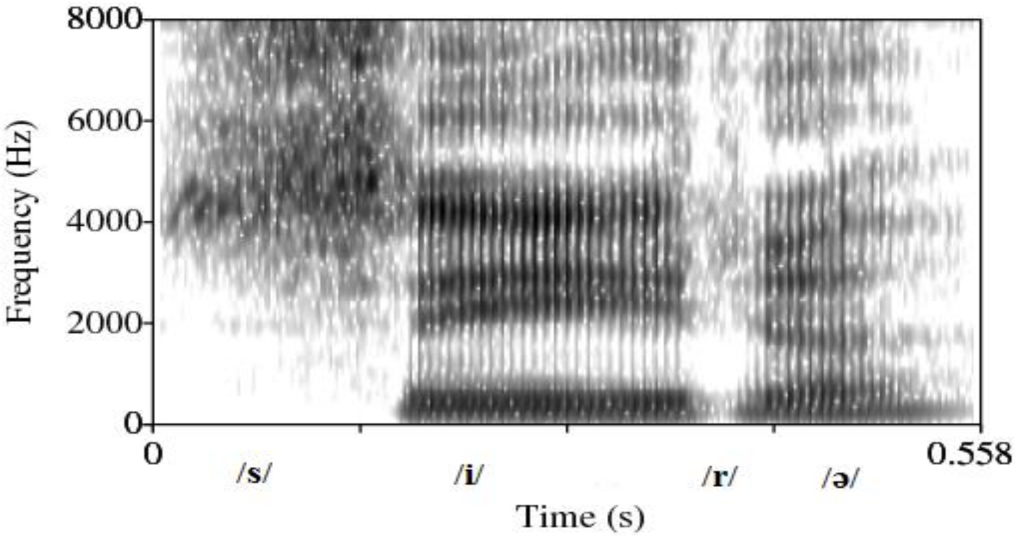


Figure 6.7(b) A typical spectrogram of the word utterance /sirə/ spoken by speaker of middle dialect, containing the front-close vowel /i/ of Sindhi.

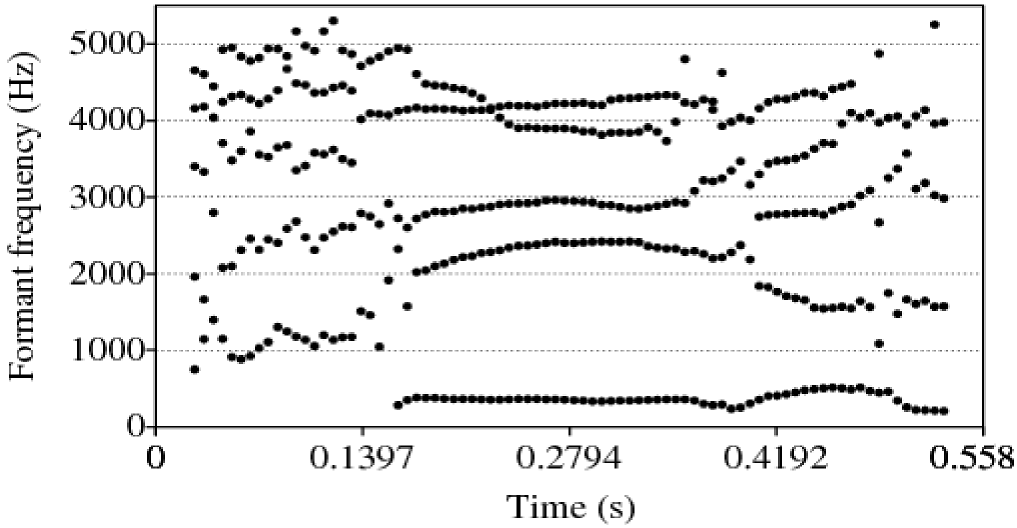


Figure 6.7(c) Formant tracks of the word utterance /sirə/, containing the front-close vowel /i/ of Sindhi.

ii. **/ɪ/, ِ :** this vowel has the 2nd lowest F1 value and the 2nd highest F2 value among the close front vowels. In articulatory terms, the body of the tongue is displaced upward and forward in the direction of the hard palate. The lips are spread, and the mouth is narrowed down for the passage of air during the production of this vowel. The vowel has a stable F1 with most of the energy, and has weak and unstable F2, F3 and F4 as shown in the spectrogram figure 6.8. The displacement of the tongue body is shown in figure 6.5 for this vowel of Sindhi. Since the tongue body is displaced forward and upward in the direction of hard palate and the vowel has a negative z-scored mean duration (discussed in section 6.3 above), this vowel can be classified as *high-front* or *close-front* short vowel of Sindhi.

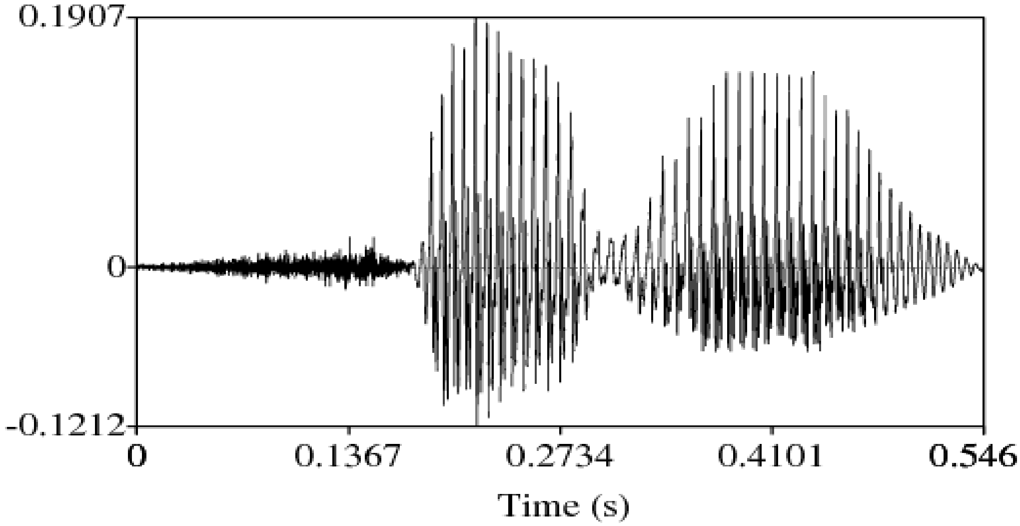


Figure 6.8(a) A typical waveform of the word utterance /sɪrə/ spoken by speaker of middle dialect, containing the front-close vowel /ɪ/ of Sindhi.

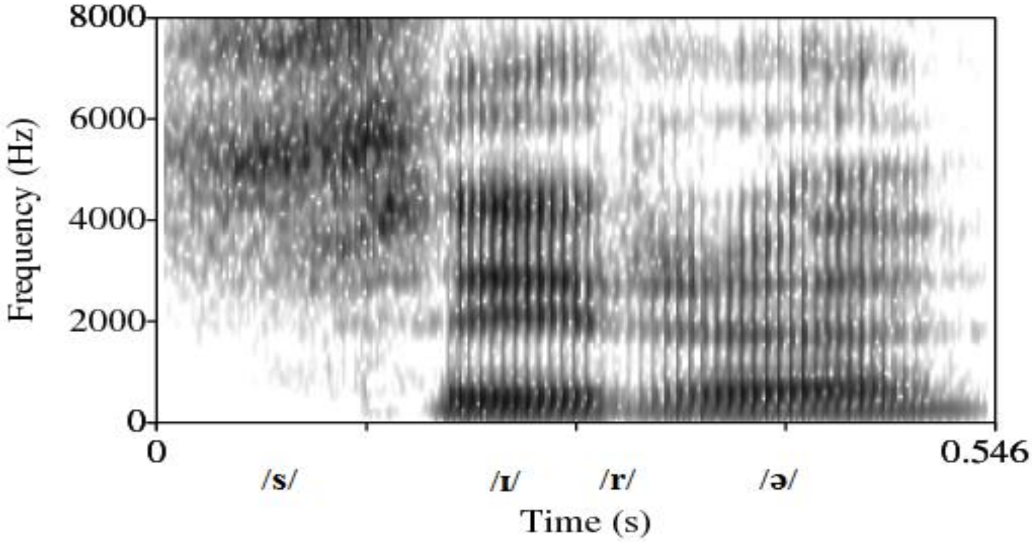


Figure 6.8(b) A typical spectrogram of the word utterance /sɪrə/ spoken by speaker of middle dialect, containing the front-close vowel /ɪ/ of Sindhi.

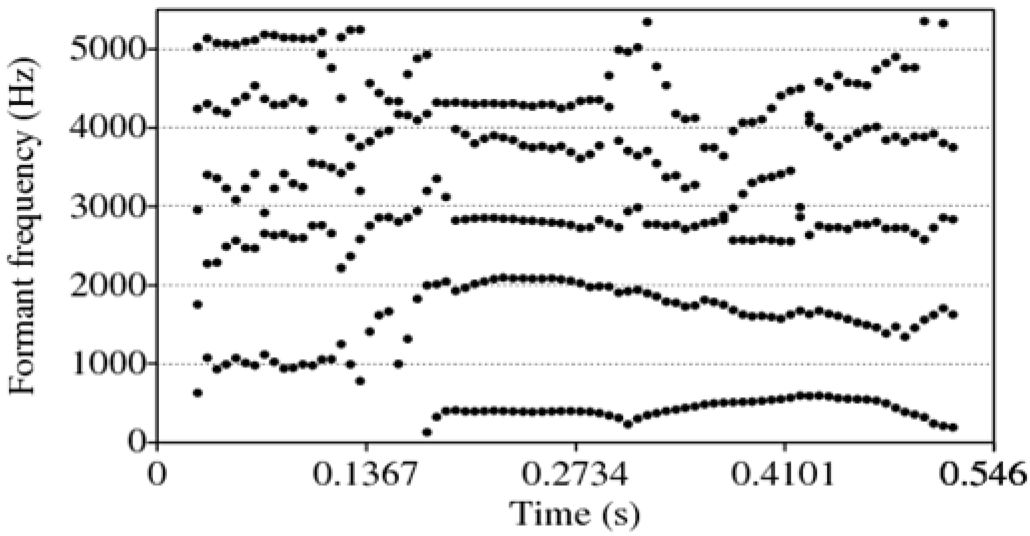


Figure 6.8(c) Formant tracks of the word utterance /sɪrə/ containing the front-close vowel /ɪ/ of Sindhi.

iii. **/e/, اي :** this vowel has a high mean F2 value and a middle range F1 value which is more close to the high front vowels. In articulatory terms, the body of the tongue is displaced forward and upward in the direction of the hard palate for this vowel of Sindhi. The lips remain neutral, and the mouth remains half closed for this vowel. The vowel has stable F1 and F2, with most of the energy present in F1 band shown in the spectrogram figure 6.9. The tongue body displacement for this vowel is shown in figure 6.5. Since the tongue body is displaced forward and upward in the direction of the hard palate this vowel can be classified as *mid-close front* long vowel of Sindhi.

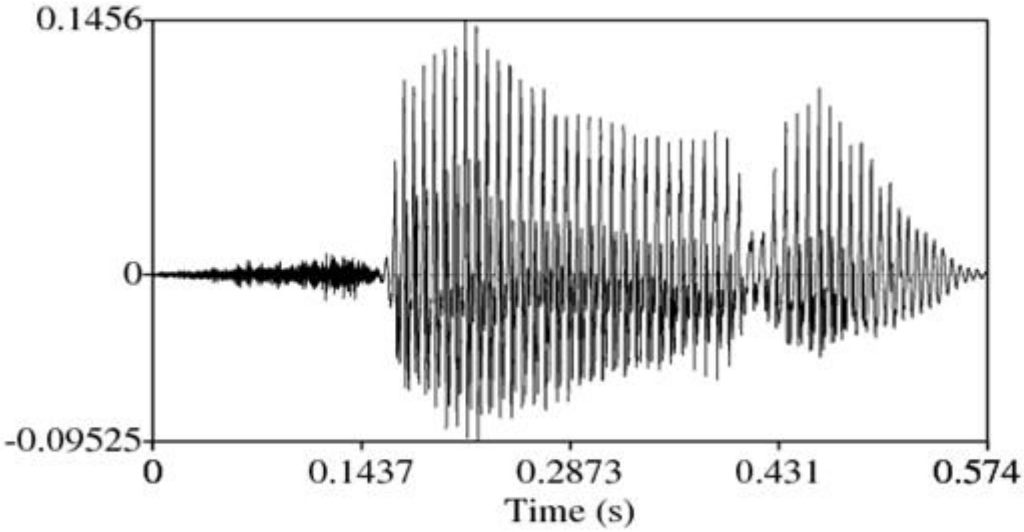


Figure 6.9(a) A typical waveform of the word utterance /serə/ spoken by speaker of middle dialect, containing the mid-close front vowel /e/ of Sindhi.

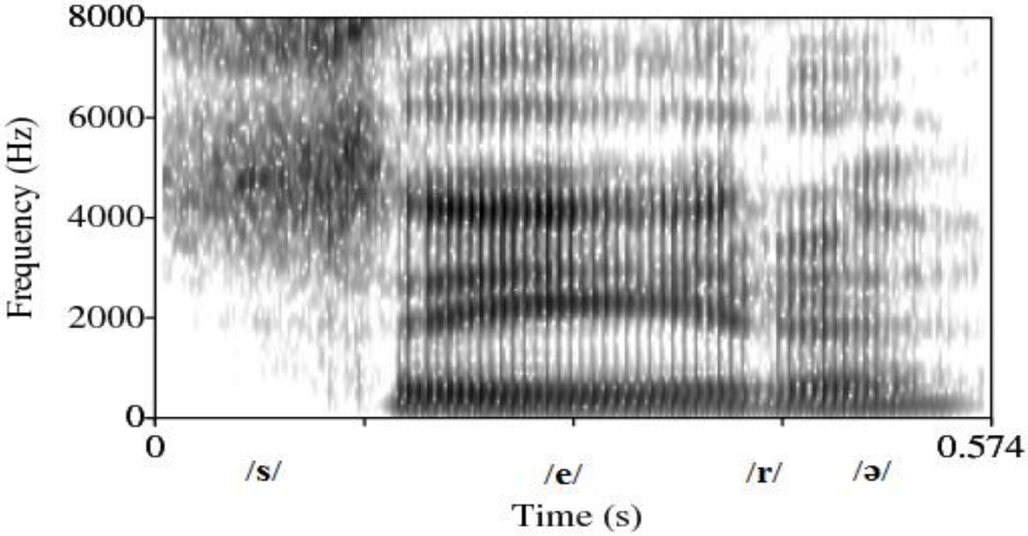


Figure 6.9(b) A typical spectrogram of the word utterance /serə/ spoken by speaker of middle dialect, containing the mid-close front vowel /e/ of Sindhi.

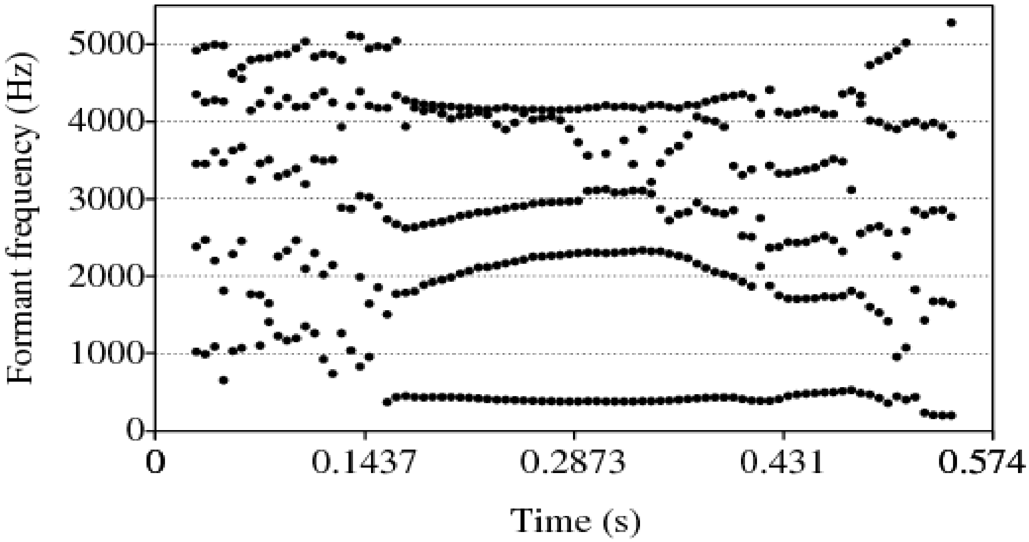


Figure 6.9(c) Formant tracks of the word utterance /serə/ containing the mid-close front vowel /e/ of Sindhi.

iv. **/ɛ/, أي :** this vowel has a high F2 frequency value and the middle range F1 value which is neither near to the close front vowels nor near to the open front vowels. In articulatory terms, the body of the tongue is displaced forward and upward in the direction of the hard palate from neutral position to the mid high position in mouth. The lips remain neutral, and the mouth remains half opened. The vowel has shown stable F1 and F2 values with most of energy present in the F1 band this is shown in the spectrogram figure 6.10. The vowel lies on the mid horizontal line drawn for the F1 (in Bark) shown in figure 6.6. Since the tongue body is displaced forward and upward in the direction of the hard palate up to the middle position in mouth this vowel can be classified as a *mid-open front* long vowel of Sindhi.

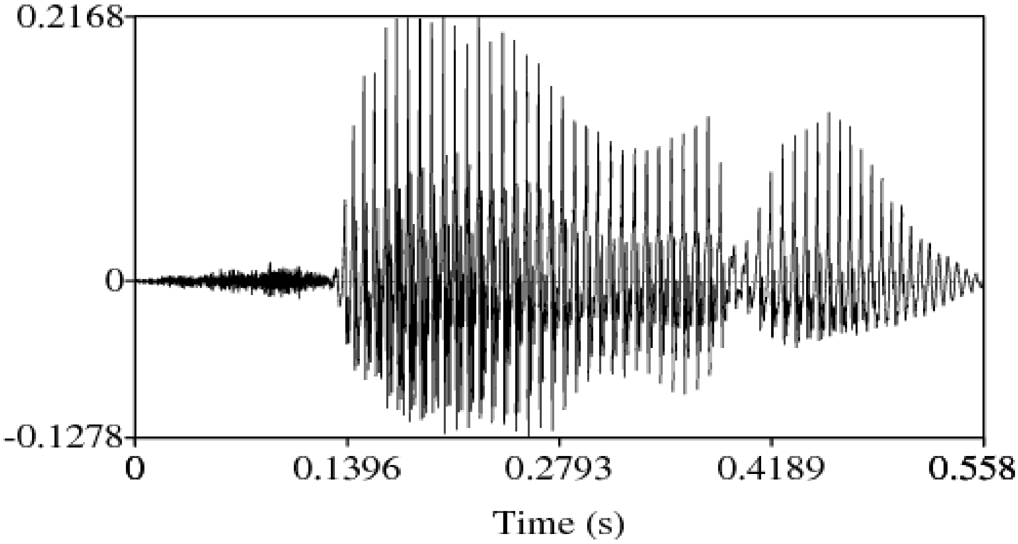


Figure 6.10(a) A typical waveform of the word utterance /sɛrə/ spoken by speaker of middle dialect, containing the mid-open front vowel /ɛ/ of Sindhi.

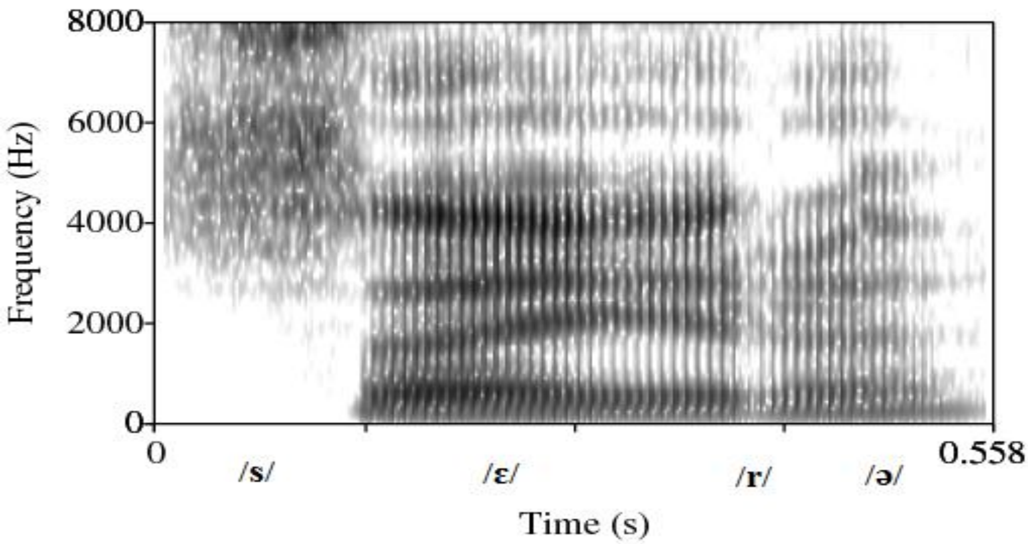


Figure 6.10(b) A typical spectrogram of the word utterance /sɛrə/ spoken by speaker of middle dialect, containing the mid-open front vowel /ɛ/ of Sindhi.

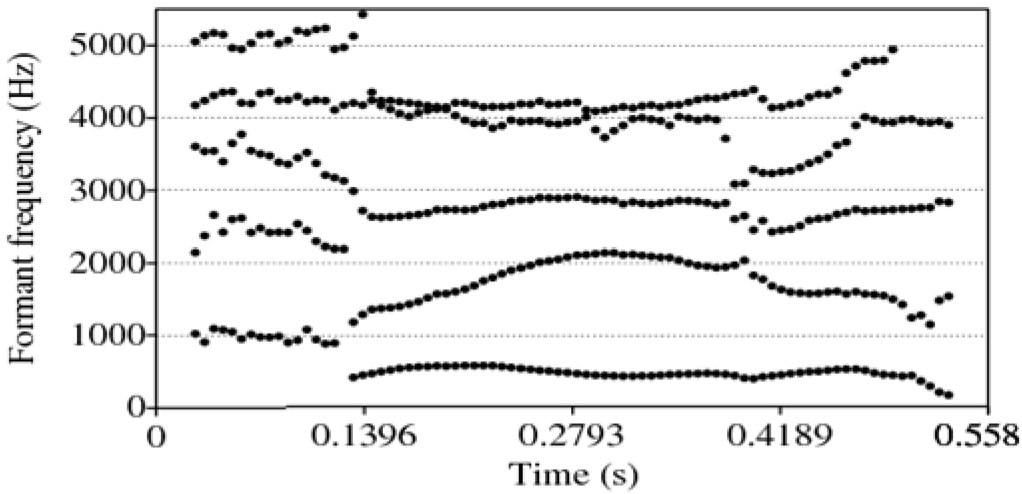


Figure 6.10(c) Formant tracks of the word utterance /sɛrə/ containing the mid-open front vowel /ɛ/ of Sindhi.

v. /ə/, اَ **:** this vowel has a high F1 frequency value and a middle range F2 value which makes this vowel a central vowel. In articulatory terms, the body of the tongue is raised up in the direction of hard palate from the neutral position to the middle position in the mouth. The lips remain neutral, and the mouth remains opened. The vowel has stable F1 and F2 with most of the energy in F1. This vowel has shown equal separation distance between F1, F2 and F3 shown in the spectrogram figure 6.11. The vowel lies near to the vertical mid line drawn for F2 values shown in figure 6.6. Since the tongue body remains in the middle for this vowel and has negative z-scored mean duration (discussed in section 6.3 above), this vowel can be classified as an *open-central* short vowel of Sindhi.

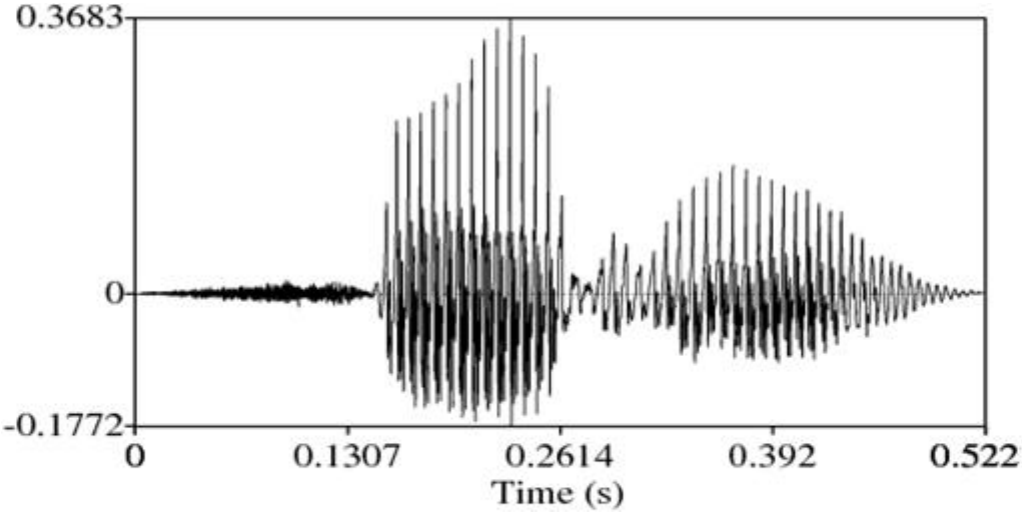


Figure 6.11(a) A typical waveform of the word utterance /sərə/ spoken by speaker of middle dialect, containing the open-central vowel /ə/ of Sindhi.

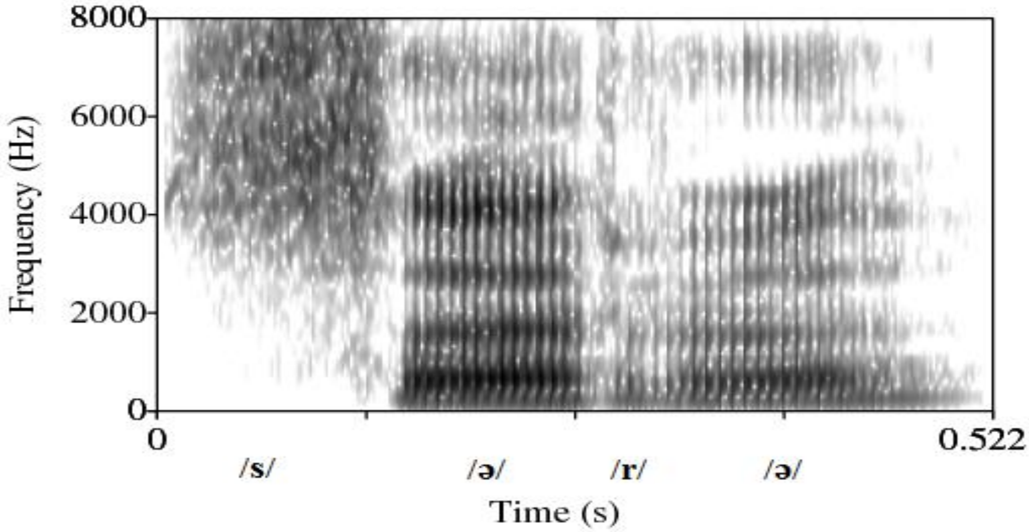


Figure 6.11(b) A typical spectrogram of the word utterance /sərə/ spoken by speaker of middle dialect, containing the open-central vowel /ə/ of Sindhi.

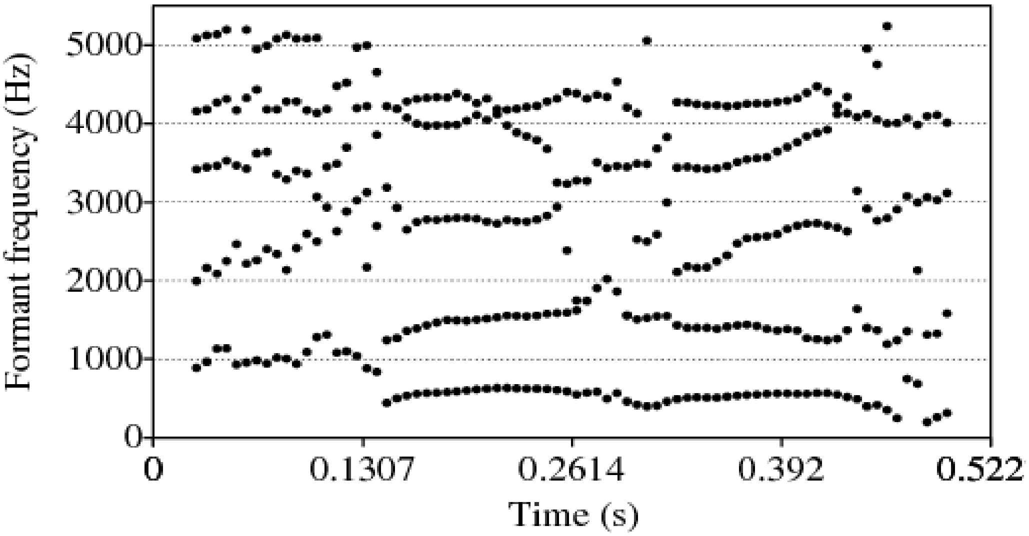


Figure 6.11(c) Formant tracks of the word utterance /sərə/ containing the open-central vowel /ə/ of Sindhi.

vi. **/ɑ/, آ :** this vowel has the highest F1 frequency value among the ten vowels of Sindhi and a mid range F2 value near to the back vowels. In articulatory terms the tongue body is displaced backward and downwards for this vowel shown in figure 6.5. The lips remain neutral and the mouth remains widely opened during the production of this vowel. The vowel has stable F1 and F2 values with most of the energy present in F1 band shown in the spectrogram figure 6.12. Since the tongue body remains downward and backward this vowel can be classified as an *open-back* long vowel of Sindhi.

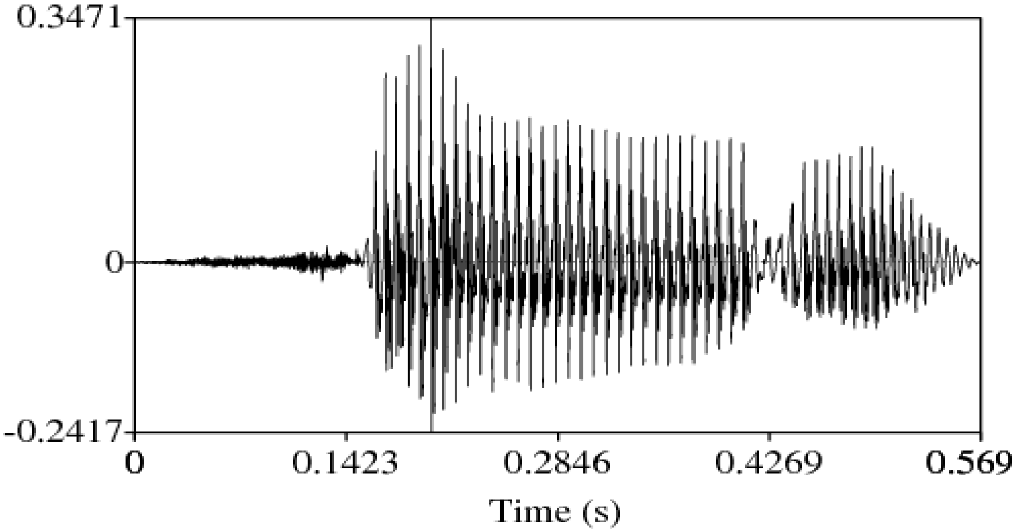


Figure 6.12(a) A typical waveform of the word utterance /sɑrə/ spoken by speaker of middle dialect, containing the open-back vowel /ɑ/ of Sindhi.

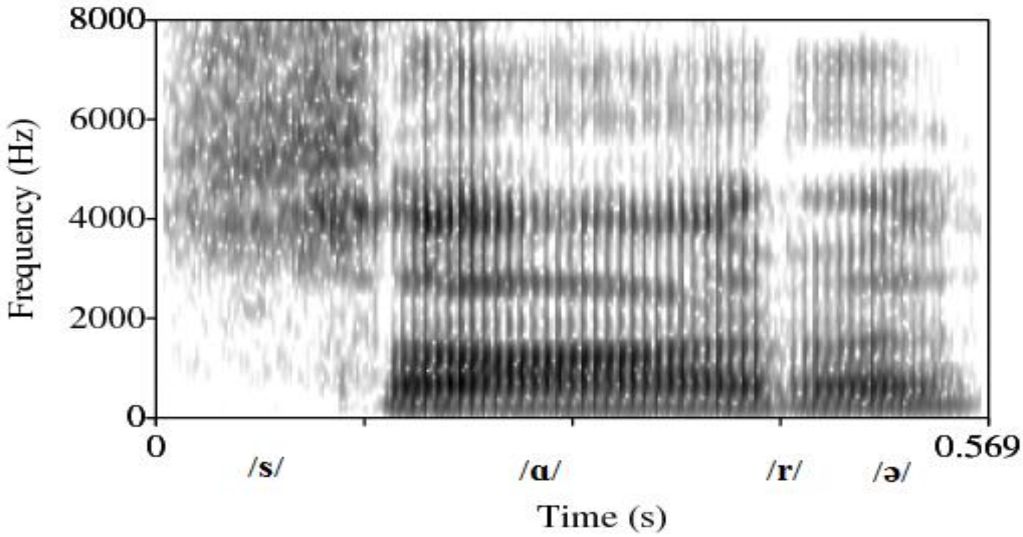


Figure 6.12(a) A typical spectrogram of the word utterance /sɑrə/ spoken by speaker of middle dialect, containing the open-back vowel /ɑ/ of Sindhi.

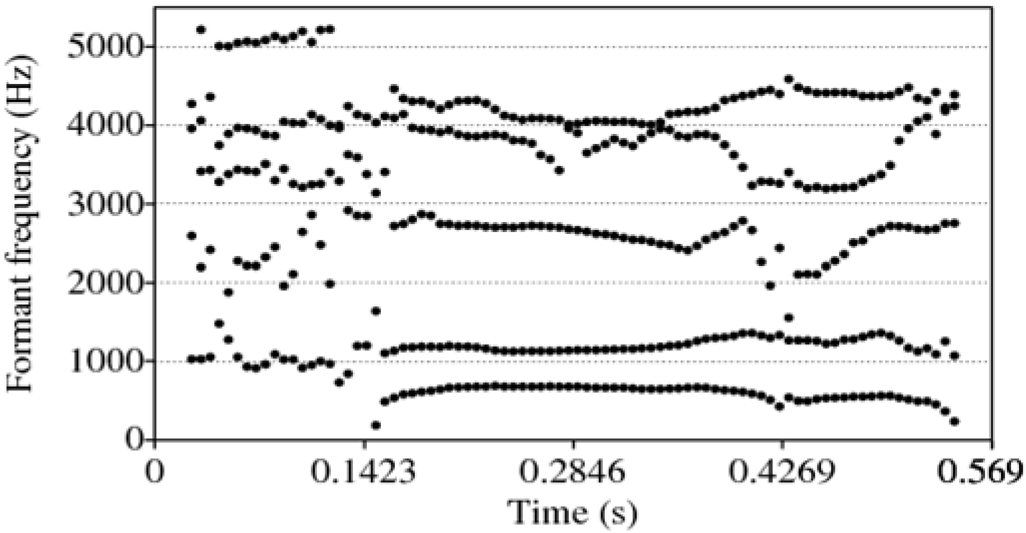


Figure 6.12(c) Formant tracks of the word utterance /sɑrə/ containing the open-back vowel /ɑ/ of Sindhi.

vii. **/ɔ/, اَو :** this vowel has a middle range F1 value and a small F2 value which makes this vowel a back vowel. In articulatory terms, the body of the tongue is displaced backward and upward in the direction of the soft palate. The lips are rounded and the mouth remains half opened for this vowel. The vowel has stable F1 and F2 with most of the energy present in the F1 band this is shown in the spectrogram figure 6.13. The vowel lies very close to the mid horizontal line drawn for the F1 (in Bark) as shown in figure 6.6. Since the tongue body is displaced backward and upward in the direction of the soft palate this vowel can be classified as a *mid-open back* long vowel of Sindhi.

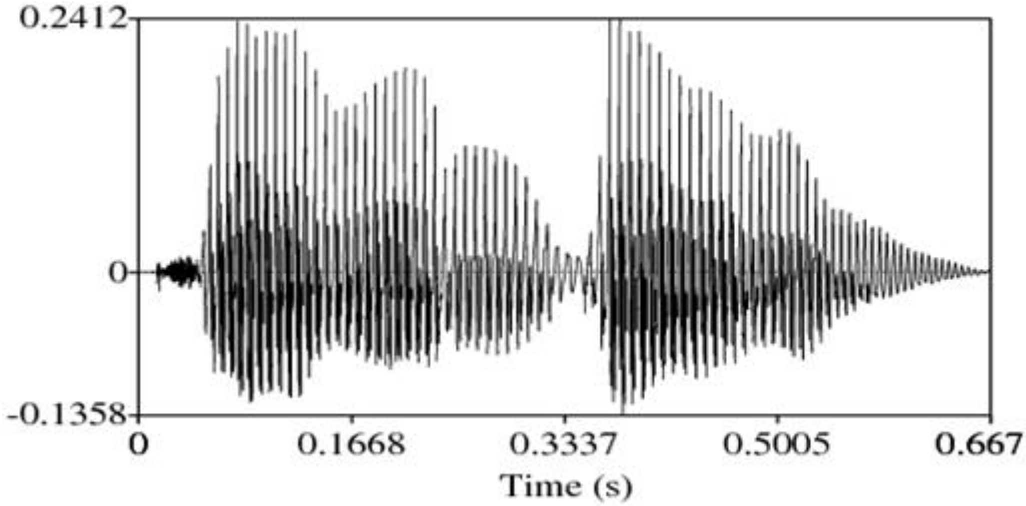


Figure 6.13(a) A typical waveform of the word utterance /ʧəvədɔ/ spoken by speaker of middle dialect, containing the mid-open back vowel /ɔ/ of Sindhi.

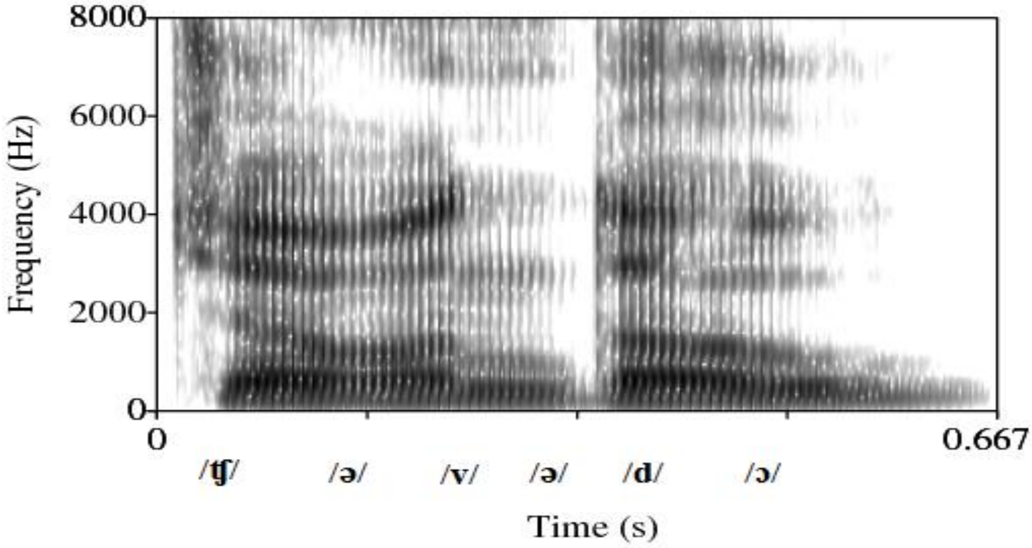


Figure 6.13(b) A typical spectrogram of the word utterance /ʧəvədɔ/ spoken by speaker of middle dialect, containing the mid-open back vowel /ɔ/ of Sindhi.

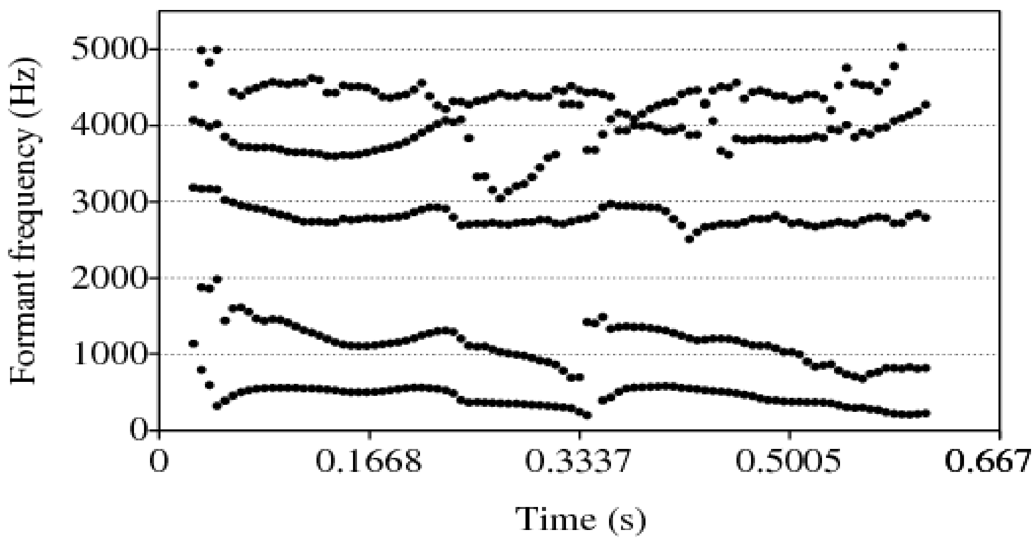


Figure 6.13(c) Formant tracks of the word utterance /ʧəvədɔ/ containing the mid-open back vowel /ɔ/ of Sindhi.

viii. **/o/, او :** this vowel has a middle range F1 value and a small F2 value which make this vowel a back series vowel. The tongue body is displaced upward and backward in the direction of the soft palate. The lips are rounded and the mouth remains half closed during the production of this vowel. The vowel has stable F1 and F2 values with most of the energy in F1 as shown in the spectrogram of figure 6.14; the tongue height for this vowel is shown in figure 6.5. The vowel lies above the mid horizontal line drawn for the F1 (in Bark) shown in figure 6.6. Since the body of the tongue is displaced upward and backward in the direction of the soft palate this vowel can be classified as a *mid-close back* long vowel of Sindhi.

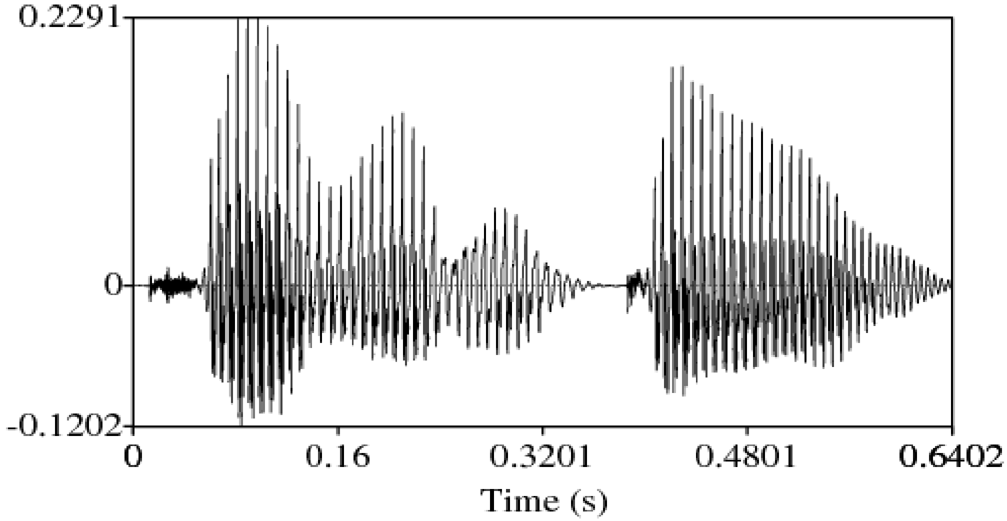


Figure 6.14(a) A typical waveform of the word utterance /ʧəvədo/ spoken by speaker of middle dialect, containing the mid-close back vowel /o/ of Sindhi.

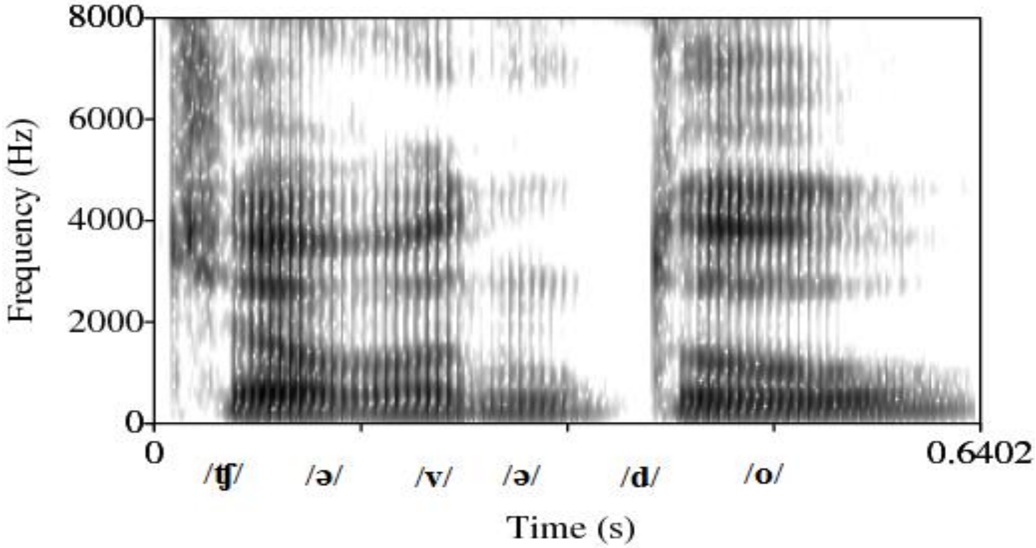


Figure 6.14(b) A typical waveform of the word utterance /ʧəvədo/ spoken by speaker of middle dialect, containing the mid-close back vowel /o/ of Sindhi.

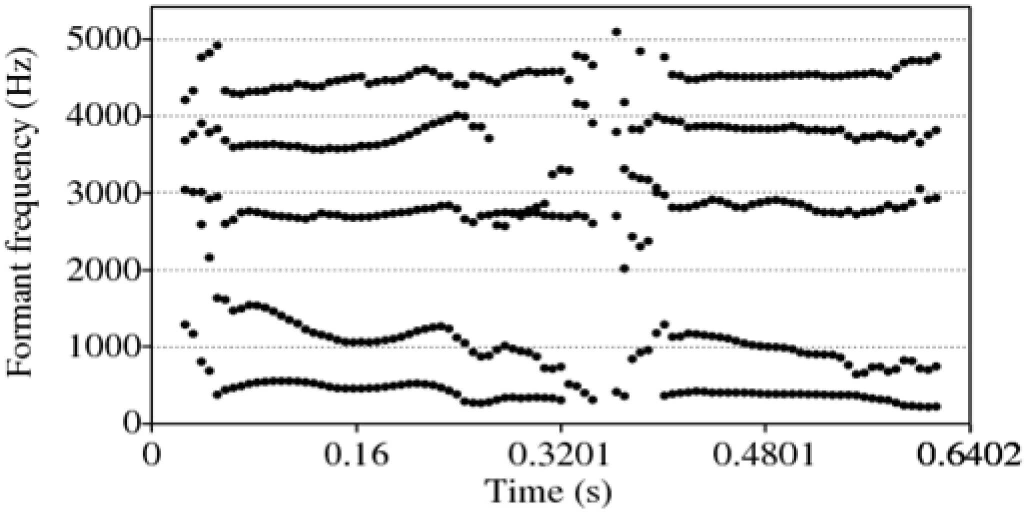


Figure 6.14(c) Formant tracks of the word utterance /ʧəvədo/ containing the mid-close back vowel /o/ of Sindhi.

ix. **/ʊ/,** أُ : this vowel has small F1 and F2 values, which make this vowel a high back series vowel. Articulatory: the tongue body is displaced upward and backward in the direction of the soft palate. The lips are rounded and the mouth is narrowed down during the production of this vowel. Only F1 is stable over time; most of the energy lies in the first formant with little in the third and fourth formants as shown in the spectrogram in figure 6.15, the tongue height for this vowel is shown in figure 6.5. Since the body of the tongue is displaced upward and backward in the direction of soft palate and the vowel has negative z-scored mean duration, this vowel can be classified as a *high-back* or *close-back* short vowel of Sindhi.

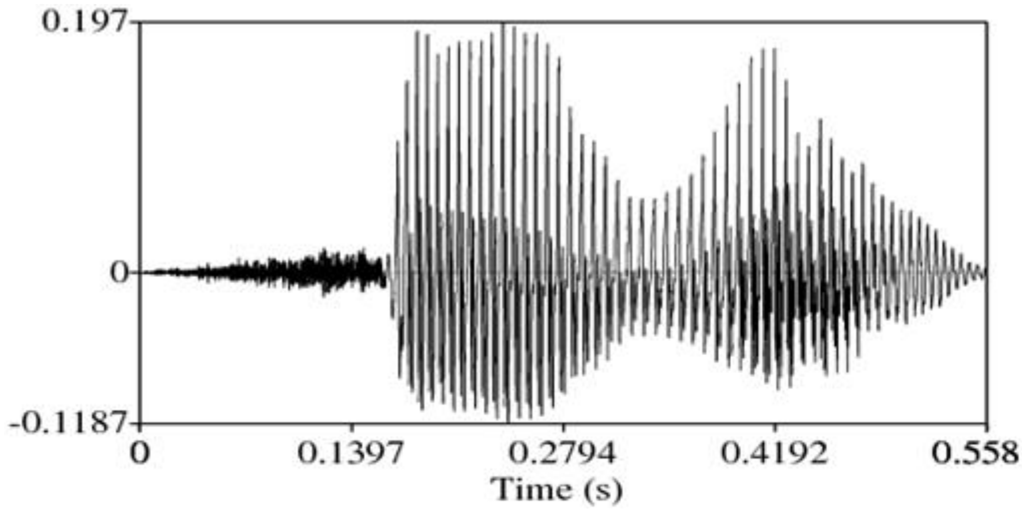


Figure 6.15(a) A typical waveform of the word utterance /sʊrə/ spoken by speaker of middle dialect, containing the close-back short vowel /ʊ/ of Sindhi.

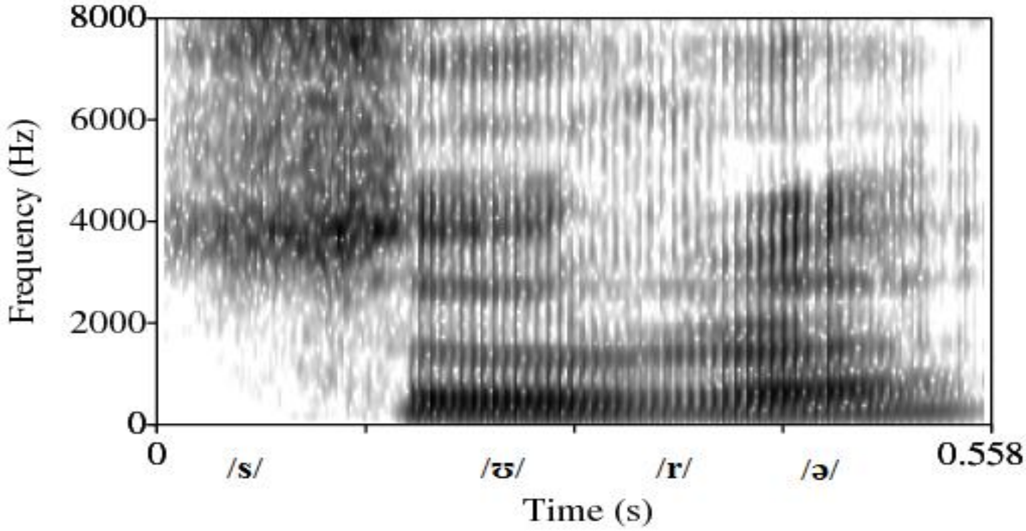


Figure 6.15(b) A typical spectrogram of the word utterance /sʊrə/ spoken by speaker of middle dialect, containing the close-back short vowel /ʊ/ of Sindhi.

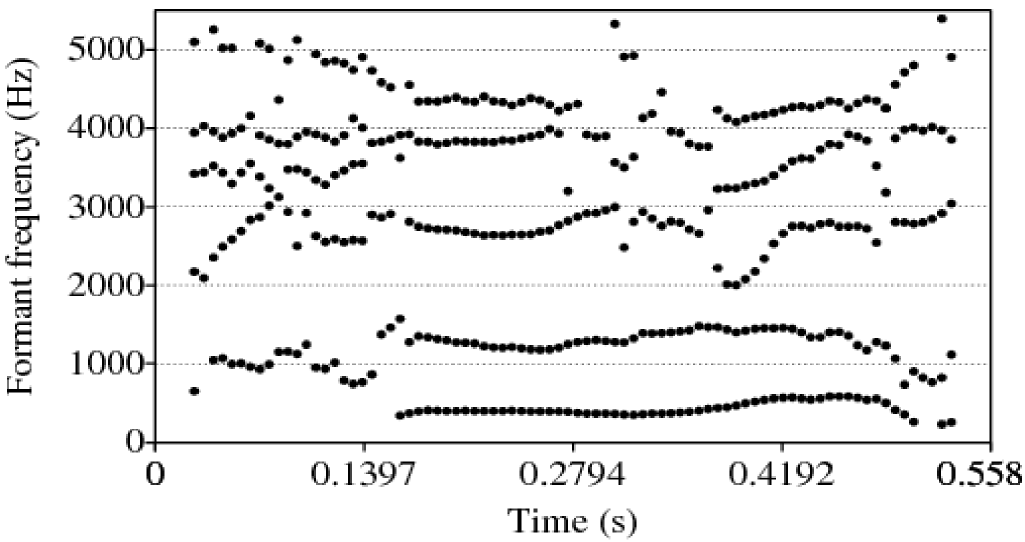


Figure 6.15(c) Formant tracks of the word utterance /sʊrə/ containing the close-back short vowel /ʊ/ of Sindhi.

x. **/u/, اُو :** this vowel has the lowest F1 and F2 values among the back series vowels, which make this vowel the high back or close back vowel of Sindhi. Articulatory: the tongue body is displaced far backward and upward in the direction of the soft palate. The lips are rounded and the mouth simultaneously narrowed downed during the production of this vowel. The vowel has only F1 value stable over time, most of the energy lies in the first formant with very little energy in the third and fourth formants as shown in the spectrogram in figure 6.16, the tongue height is shown in figure 6.5. Since the body of the tongue is displaced backward and upward in the direction of the soft palate this vowel can be classified as a *high-back* or *close-back* long vowel of Sindhi.

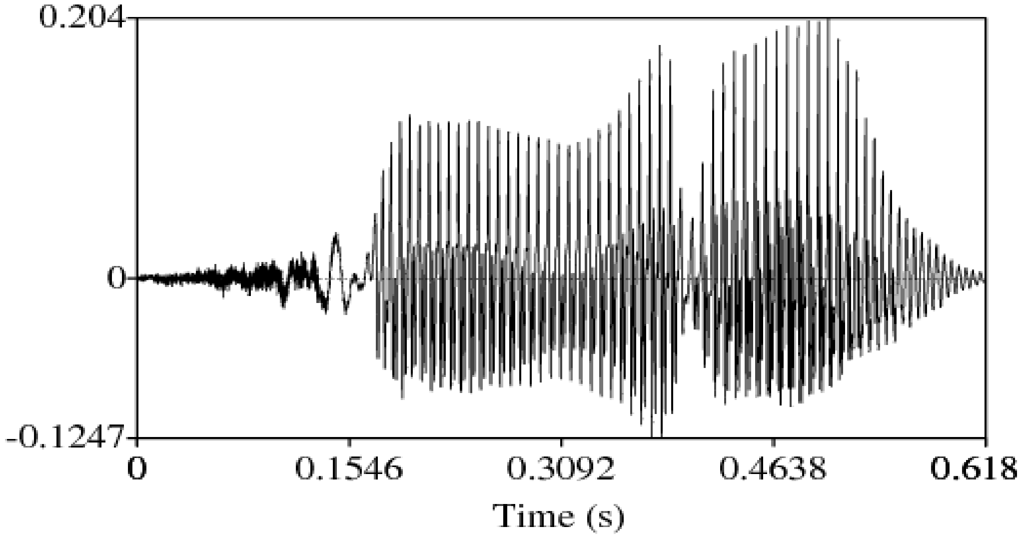


Figure 6.16(a) A typical waveform of the word utterance /surə/ spoken by speaker of middle dialect, containing the close-back vowel /u/ of Sindhi.

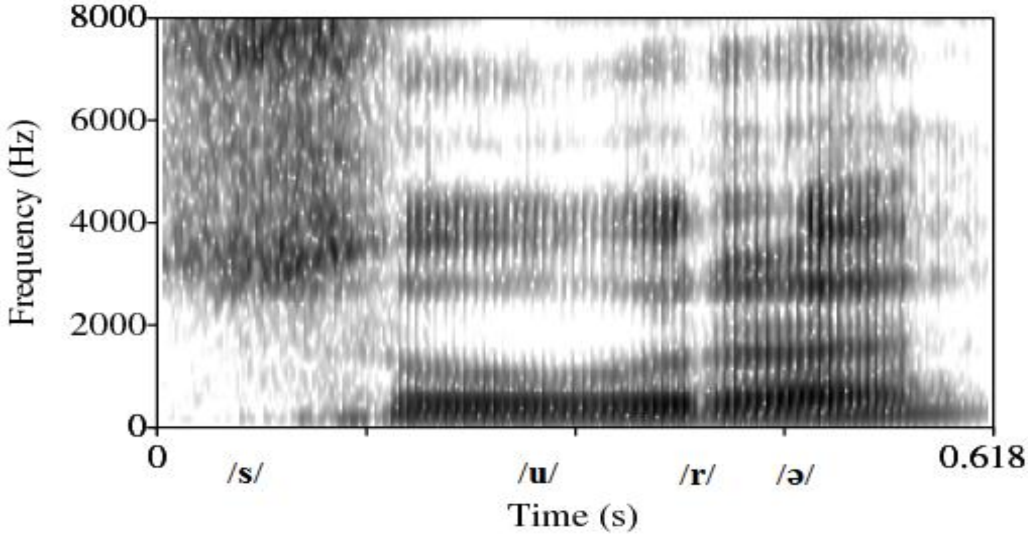


Figure 6.16(b) A typical spectrogram of the word utterance /surə/ spoken by speaker of middle dialect, containing the close-back vowel /u/ of Sindhi.

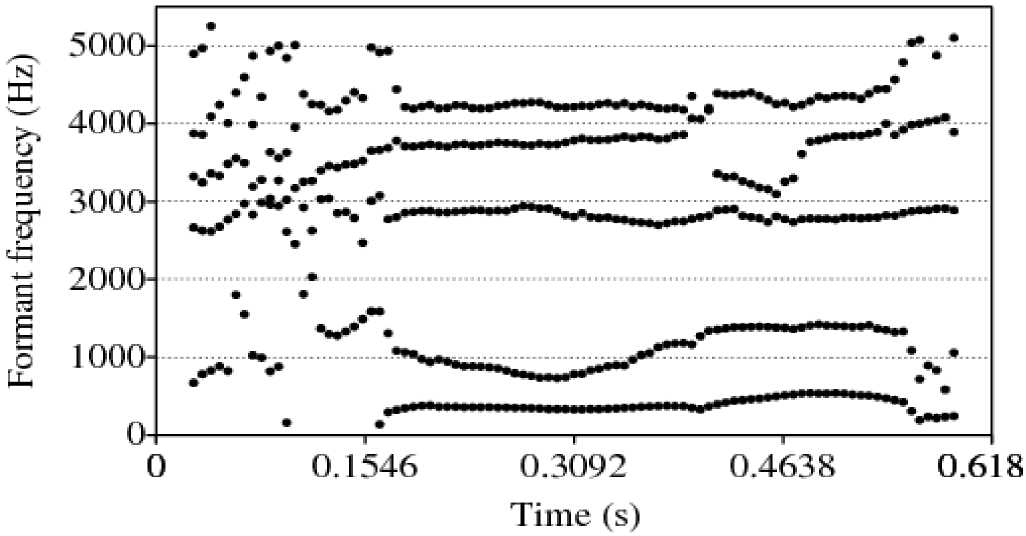


Figure 6.16(c) Formant tracks of the word utterance /surə/ spoken by speaker of middle dialect, containing the close-back vowel /u/ of Sindhi.

6.6 Conclusion

In this chapter a thorough study has been conducted to analyze the ten monophthong vowels of Sindhi. The reported acoustic parameters of the sounds include the fundamental frequency, the average durations, and the first four formant values. Based on the average durations, the vowels present in the language inventory have been categorized either long (tense) and short (lax) vowels. A cross dialect acoustic variation for the five language dialects is presented by plotting F1 x F2 acoustic vowel plots shown in section 6.4. The first two formants provide relative information about the perceived vowel quality and intelligibility. Clear vowel qualities in this respect are uniformly observed for the open back and open central vowels; whereas the highest degree of overlap is observed for close front and close back lax and tense vowel pairs. The vowel /ɛ/ has shown the distinct vowel quality for *Lari* and *Lasi* dialect.

The four extreme reference (cornering) vowels are identified by analysing the F1 x F2 acoustic vowel plots are: /i/, /ɛ/, /ɑ/ and /u/. It is observed that the vowel space area for the three short vowels is uniformly smaller compared to the vowel space area measured for the four extreme reference (cornering) vowels. It is also observed that the *Utradi* and *Middle* dialect has the largest vowel space area and the *Lari* dialect has the smallest vowel space area. It is therefore concluded that the speakers of the *Utradi* and *Middle* dialect with largest vowel space area produce more intelligible speech than the *Lari* dialect. Using the first two formant frequency values, acoustic vowel plots and knowledge about the relationships between acoustic phonetics and articulatory phonetics, a three articulatory classification of the language vowel system is presented in section 6.5. From the vowel plots presented in the chapter, an idea about the vocal exercise that needs to be performed to produce the ten monophthong vowel-sounds of Sindhi and the intelligibility of each of the vowel-sounds can be formed.

CHAPTER 7

ACOUSTIC ANALYSIS OF DIPHTHONGS AND GLIDES OF SINDHI

Chapter 7

Acoustic Analysis of Diphthongs and Glides of Sindhi

7.1 Introduction

In this chapter we will start by analysing the motion of the vocal tract and the associated formant structure as one vowel or vowel like sound changes to another without hiatus. Three such types of sounds (the vowels, diphthongs and glides) are present in Sindhi for which the formant structure is stable and predictable. This chapter mainly discusses and presents the acoustic-phonetic analysis of the diphthongs and glides of Sindhi. However understanding the corresponding acoustic similarities and differences between three types of sounds (the vowels, diphthongs and glides) of Sindhi is the subject of discussion in this chapter.

The difference between oral vowels and glides is quite straight forward; the glides do not maintain a steady-state; whereas the vowels maintain a steady-state (Kent, 2002; Raphael, 2006). Unlike the oral vowels the difference between diphthongs and glides is observed by analysing the transition segment of the two sound types. The transition segment of diphthongs is slower and more gradual compared to the vowel-to-glide transition (Olive, 1993). Because the formation of the vocal tract during the production of glides moves in and out very quickly; therefore the glides are considered as sharp transitory vowels (Raphael, 2006).

One of the early study talks about the diphthongs of Sindhi is Ernest Trumpp's book, 'Grammar of the Sindhi Language', in which author has mentioned that there are no diphthongs in Sindhi; the vowels [ɑɪ], and [ɑʊ] are loosely pronounced as ɑ-ɪ and ɑ-ʊ (Trumpp, 1872). The presence of diphthongs in Sindhi became contentious in the literature; when author Jatoi (Jatoi, 1996) in his book (written in Sindhi script) reported that diphthongs are present in Sindhi. For example the diphthong [ɑɪ], in word 'l[ɑɪ]', and the diphthong [iʊ], in word 'p[iʊ]' etc. The author Bughio in (Bhugio, 2006) discussed the presence of two diphthongs in the phonemic inventory of Sindhi: (i) [ɑɪ] a non-labialised descending diphthong with a glide from [ɑ] to [ɪ] and (ii) [ɑʊ] a labialised descending diphthong with a glide from [ɑ] to [ʊ]. The presence of two diphthongs discussed in (Bhugio, 2006) is also discussed in (Jennifer, 2006); the author has associated the diphthong realization in Sindhi with the speaker's religious affiliation such as Muslim Sindhi speakers and Hindu Sindhi speakers. The Muslim Sindhi speakers produce the short diphthongs [ɑɪ], and [ɑʊ], whereas the lax vowels [ɛ], and [ɔ] are typical for Hindu Sindhi speakers. Two phonetic behaviours are associated with the glides /w/ and /y/ of Sindhi: (i) at the word initial position or at the word final or medial position, if followed by a consonant, glides function as independent phonemes (ii) the glides if preceded or followed by a vowel, at the word final or medial position and the two sounds are produced together form a diphthong phoneme of Sindhi (http://tdil.mit.gov.in/sindhidesignguideoct02.pdf; site visited on 10th October 2009). The two viewpoints can be concluded from the literature regarding the presence of diphthongs in Sindhi. First when a short vowel at the word medial or final position (orthographically represented with the symbol 'Hamzo'), if immediately followed or preceded by a vowel, could potentially form a diphthong phoneme of Sindhi. Second, if a glide is preceded or followed by a vowel at the word medial or final position it can potentially form a diphthong phoneme of Sindhi. In order to carryout acoustic-phonetic analysis of Sindhi diphthongs according to the above mentioned two perspectives, the selection of the words (for voice sample recording) were made in a way so that these words contain the syllable(s) having a short vowel with 'hamzo', either preceded or followed by a vowel or the syllables containing the glide consonant followed or preceded by a vowel. A syllable in Sindhi must have a vowel that serves as a syllable nucleus; the consonants in Sindhi cannot form a syllable. The two adjacent vowels in the syllable of Sindhi are paired as long and short vowels. The syllables of the following structure can be constructed in Sindhi (Jatoi, 1996):

- V: only long vowel individually can construct a syllable i.e. /ɑ/, آ "come"
- Cv: consonant followed by a short vowel i.e. /nə/, نَہ "no"
- CV: consonant followed by a long vowel i.e. /mɑn/, مان "I"
- CVv: consonant followed by two successive vowels; in this syllable the final vowel is a short vowel i.e. /huə/, هوء "she"
- CVC: a vowel surrounded by consonants i.e. /sɑn/, سان "with"
- CCV: two consecutive consonants followed by a vowel i.e. /kyɑ/, ڪيا "did"

- CCVC: two consecutive consonants followed by a vowel and a consonant i.e. /qyɑs/, کياس "forgive"
- CVCC: syllable ends with two consecutive consonants i.e. /hərj/, هرج "problem"

Note: symbol C = consonant, V= long vowel, and v = short vowel.

There is a need to carry out more research work on the subject of diphthongs in Sindhi; even though it is to be assumed that diphthongs do exist in Sindhi; their identification is yet to be registered and type categorized. In the subsequent sections of this chapter the acoustic-phonetic analysis of the diphthongs of Sindhi along with the comparative analysis of the glides is presented. The results in the subsequent sections for the diphthong and glide phonemes of Sindhi are mainly based on the speech samples taken from the word utterances containing the syllable(s) with 'hamzo', either preceded or followed by a vowel or the syllables containing the glide consonant followed or preceded by a vowel.

7.2 Glides

Glides are vowel-like sounds; but unlike the vowels they do not form the prominent part of the speech signals in a syllable which is why they cannot form syllables individually in Sindhi as vowels do. However this consonant class of sounds posses a few vowel-like acoustic characteristics such as the periodic output waveform and well defined formant structure; due to the well defined formant structure glides are usually referred to as the semi-vowels (Ioana, 2002; Martínez Celdrán, 2004; Padgett, 2008). Sindhi includes two glide consonants in the phonemic inventory articulated at two places of articulation: the bilabial /w/و and the palato-alveolar /y/ ي. Glides are the sounds considered with a gradually changing formant pattern in the same way the pattern changes for the diphthongs (Ladefoged, 1993). This is why the acoustic analysis of the glide consonants is carried out in this chapter together with the analysis of the diphthongs. As discussed above the glides do not maintain a steady-state, this characteristic differentiates them from the monophthong vowels; however it is difficult to differentiate them from the diphthongs as both sounds are referred to as the sounds gradually change the formant pattern. Glides are sharp transitory semi-vowels; therefore the transition segment duration for these sounds is shorter compared to the transition segment for diphthongs; because the vocal tract configurations during the production of glides move in and out quickly (Aguilar, 1999; Olive, 1993). A glide in a syllable if it precedes or follows the vowel it can potentially result in a diphthong phoneme of Sindhi; therefore the voice sample recording for the analysis of glides follows the VCV phoneme sequence. In the VCV phoneme sequence the consonant surrounded by vowels is the target analysis glide consonant in the word utterance. This VCV phoneme sequence helps to analyse the formant transition coming into the consonant sound (the glide consonant) from the preceding sonorant sound (vowel) and the transitions going away from the glide consonant to the following sonorant sound (vowel). Figure 7.1 below shows the waveform of the word utterance /həwɑ/ هَوا containing the bilabial glide /w/ of Sindhi. The waveform segment for the glide /w/ shows the periodic signal activity and the presence of the energy in the lower frequency region is shown in the spectrogram of figure 7.2; this indicates that the bilabial glide /w/ is a voiced phoneme of Sindhi. The first formant motion due to the low F1 is downward (coming into the glide) and upward (going away from it); whereas the F2 transitions remain unchanged at the vowel junctures for this sound. The glide /w/ showed sharp transitory formant transitions; therefore the two vocalic elements for the segment of the glide consonant cannot be witnessed in the word utterance /həwɑ/ هَوا see figure 7.2 and figure 7.3. The results show that the glide /w/ of Sindhi, if followed or preceded by a vowel, forms an independent phoneme of Sindhi.

Figure 7.4 below shows the waveform of the word utterance /vəyo/ ويو containing the palato-alveolar glide /y/ of Sindhi. The waveform segment for the glide /y/ shows the periodic signal activity and the presence of the energy in the lower frequency region, see the spectrogram in figure 7.5; this indicates that the palato-alveolar glide /y/ is a voiced phoneme of Sindhi. Due to the high F2 the second formant motion is upward (coming into the glide) and downward (going away from it); whereas the F1 motion is downward and upward. The glide /y/ of Sindhi shows sharp transitory formant transitions, this is shown in figure 7.6 - the formant tracks. The results show that the glide /y/ of Sindhi, if followed or preceded by a vowel forms an independent phoneme of Sindhi. The acoustic differences between diphthongs and glides are discussed in section 7.2 along with the acoustic analysis of diphthongs.

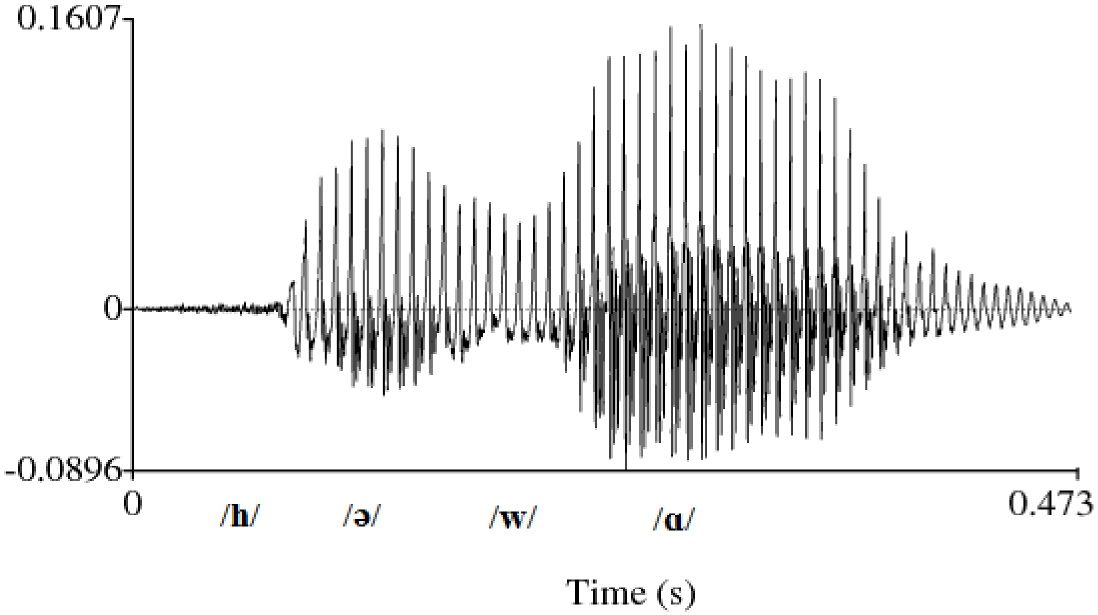


Figure 7.1 The waveform of the word utterance /həwɑ/ هَوا containing the bilabial glide /w/ of Sindhi.

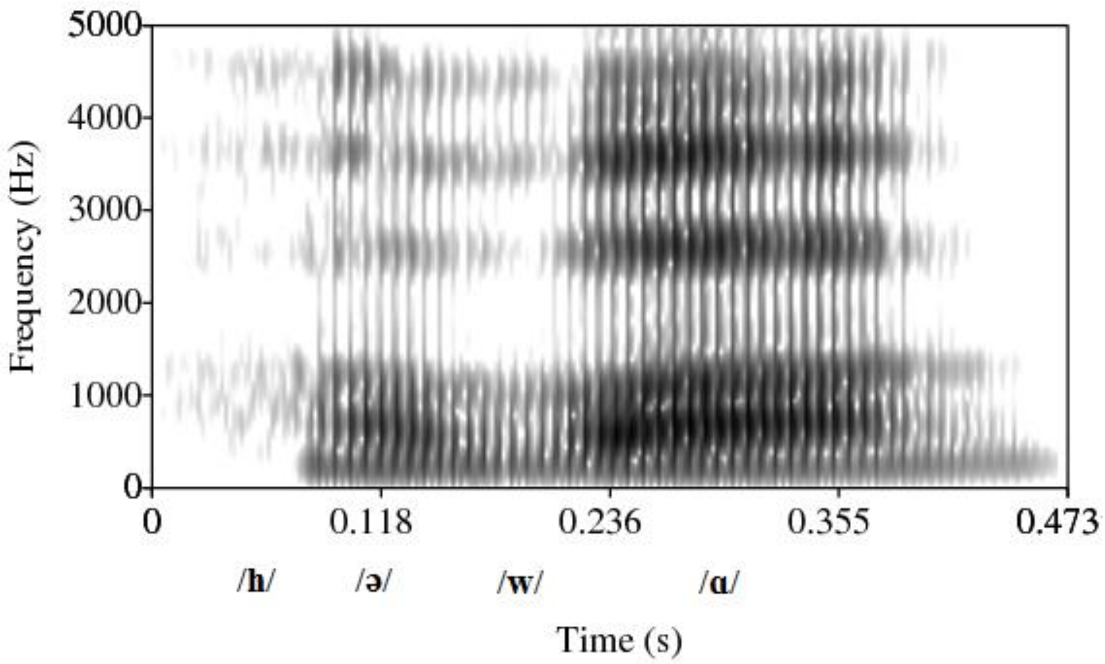


Figure 7.2 The spectrogram of the word utterance /həwɑ/ هَوا containing the bilabial glide /w/ of Sindhi.

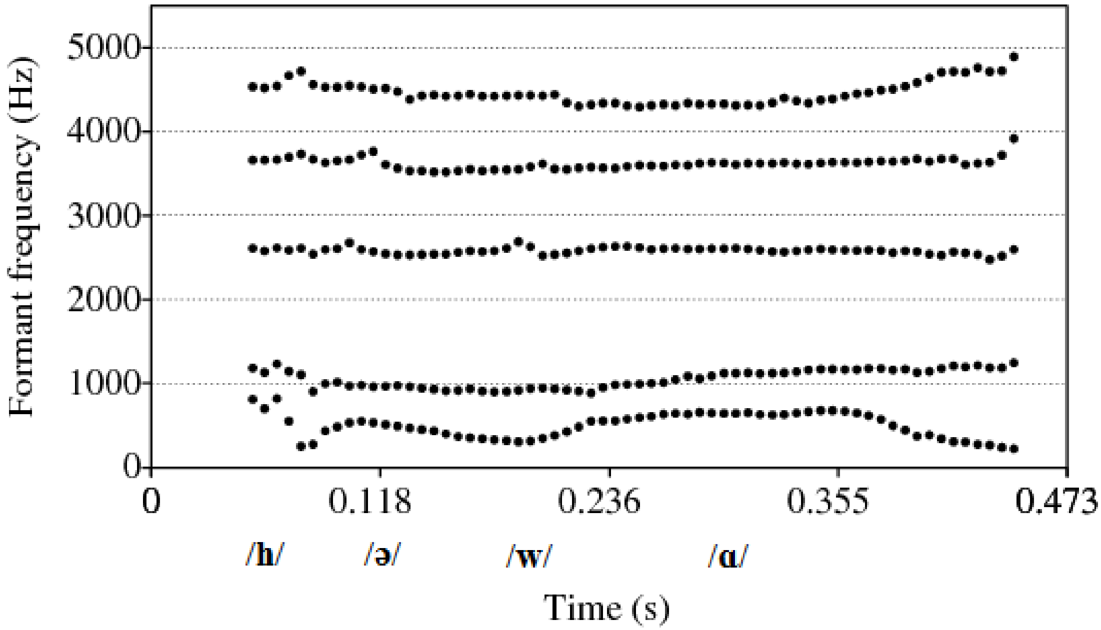


Figure 7.3 The formant transitions coming into the bilabial glide /w/ from the preceding vowel sound /ə/ and going away from it to the following vowel sound /ɑ/ of Sindhi.

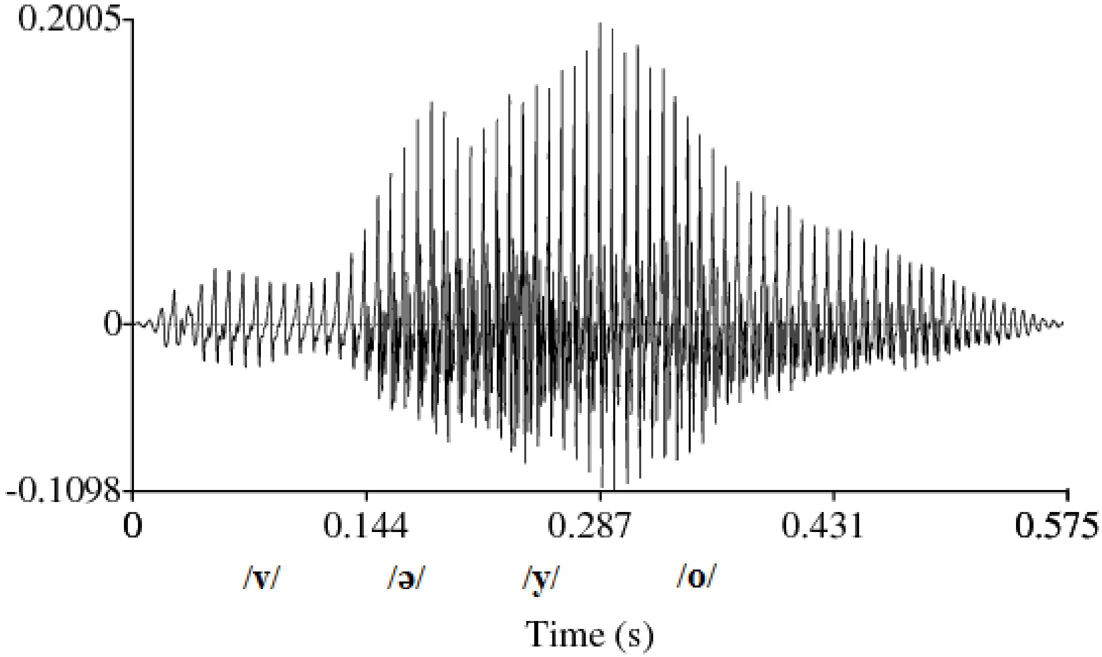


Figure 7.4 The waveform of the word utterance /vəyo/ويو , containing the palato-alveolar glide /y/ of Sindhi.

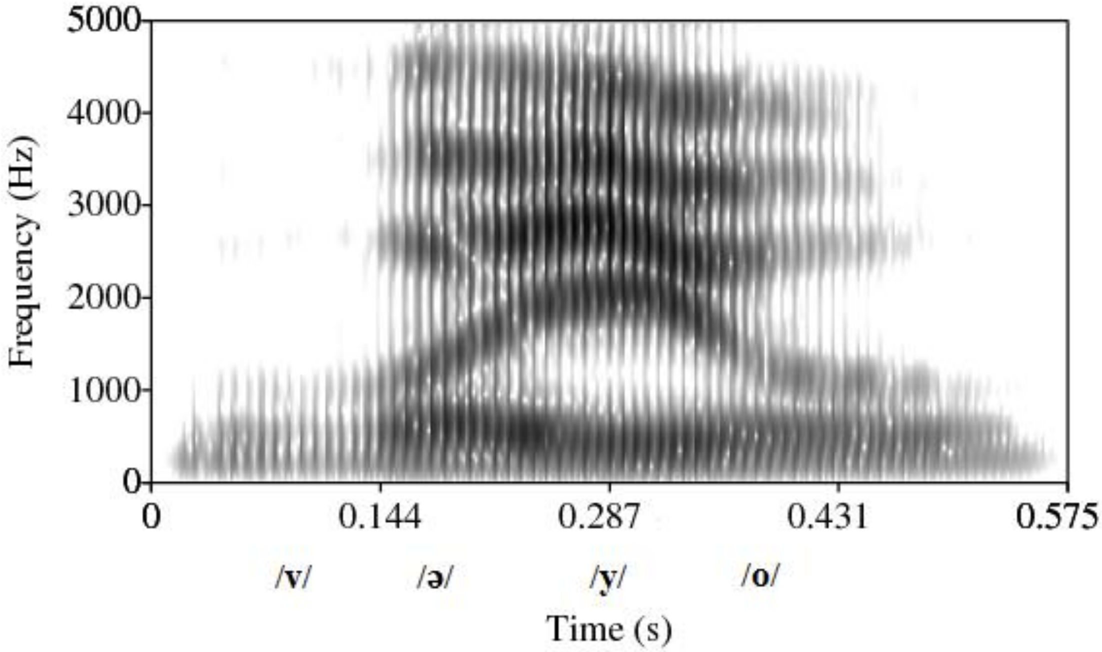


Figure 7.5 The spectrogram of the word utterance /vəyo/ويو , containing the palato-alveolar glide /y/ of Sindhi.

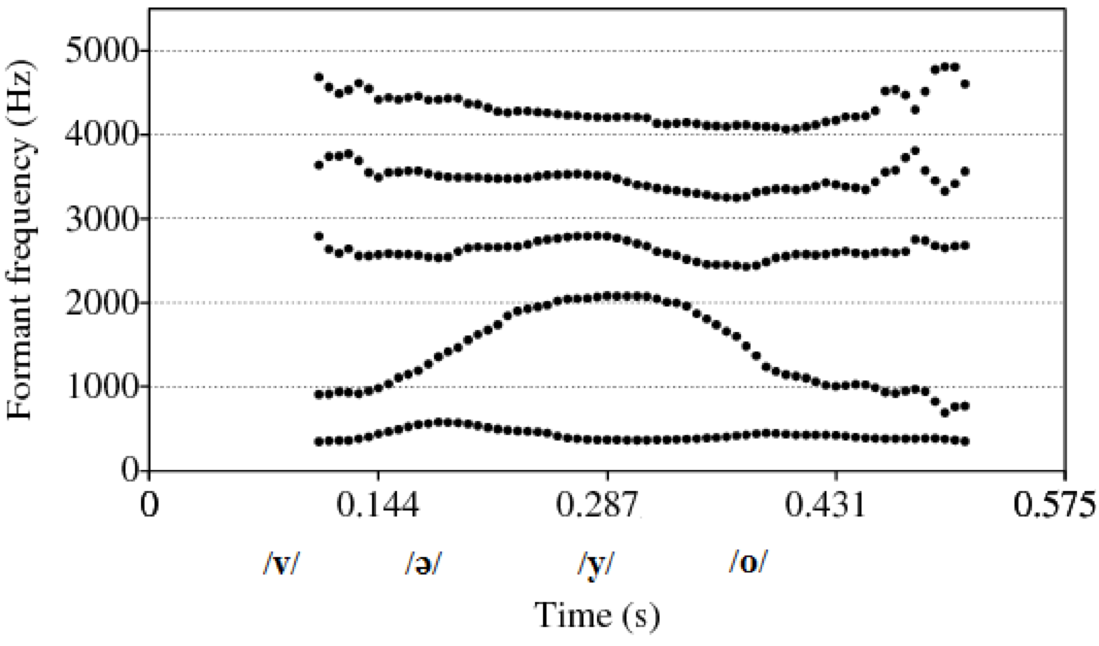


Figure 7.6 The formant transitions coming into the palato-alveolar glide /y/ from the preceding vowel /ə/ and going away to the following vowel /o/ of Sindhi.

7.3 Diphthongs

The formation of two consecutive vowel sounds in the oral cavity for one syllable results in a glide from the start-vowel to the end-vowel, usually referred to as a diphthong. Diphthongs are also defined as independent vowel-glides; which means that the articulators set in the position of one vowel for articulation, move in the direction of another vowel for articulation (Jones, 1969). Diphthongs are well correlated with the moving pattern of the formant frequency plots, starting at the on-glide section of the plot and ending at the off-glide section. For example, this feature can be observed in the English diphthong [aɪ] which involves plot changes from the on-glide [a] segment to the off-glide [ɪ] segment (Kent, 2002). Diphthongs are also considered as vowel sounds which inherit the qualities of two vowels. This can be observed in the diphthong sound [eɪ], of English, ending with the articulators set for the vowel [ɪ] in the oral cavity that entails tongue movement forward and upward from the position set while pronouncing [e] (Raphael, 2006). The same can be observed for [aɪ], and [ɔɪ], etc. diphthongs of English (Raphael, 2006). On the other hand, the diphthongs [aʊ] and [oʊ] of English ending with the articulators set for the vowel [ʊ] in the oral cavity entail tongue movement backward and upward simultaneously with lip protrusion (Raphael, 2006). Another class of sounds possess similar acoustic characteristics like vowels are the glide consonants; with a few similar acoustic characteristics i.e. the periodic output waveforms and well defined formant structure, usually these sounds are referred to as the semi-vowel sounds (Ioana, 2002; Padgett, 2008). The glides having well defined and predictable formant structure could potentially lead to confusion for the identification of the three classes of sounds: the glides, diphthongs and the monophthong vowels (Olive, 1993). The difference between glides and vowels is easy to define; because the later class of sounds maintain steady-state. However the difference between glides and diphthongs is difficult to determine; because the two sounds do not maintain steady-state, rather they are the sounds with a gradually changing formant pattern. We will rely upon the analysis of the transition segment of the glides and diphthongs for their identification.

7.3.1 Voice sample recording for diphthongs

The voice sample recording for the analysis of the diphthongs of Sindhi was made in a different way to the speech samples recorded for the monophthong vowels and the consonant sounds. Due to the issue of diphthong presence controversy (discussed in the introduction of the chapter) it was hard to compile a concrete list of words that carry diphthongs of Sindhi. Therefore the list of twenty eight such words, with the help of Sindhi dictionary, was compiled to the best of our perception that these words may contain a diphthong sound. The equal number of other non-diphthong words were added in the list of twenty eight selected words and randomized. Speakers were given the randomized list of the isolated words and each word embedded in a carrier sentence for the recording of the voice samples. Speakers were asked to read out the two lists for the voice sample recording. Ten instances of each speaker's voice samples were recorded and stored in .wav file format for subsequent offline processing. Only one speaker voice samples were recorded at one time to avoid any pronunciation imitation. Tools used for recording include Marantz (PMD660) a solid state digital audio recorder see figure 3.2, used with an external broadcast quality lightweight condenser microphone the Rode NTG-2 see figure 3.3. The Speech samples were recorded as a single channel with a sampling rate of 48 kHz, and stored in .wav file format, 16 bits per sample.

The voice samples for the glide consonants were recorded in a similar way; the voice samples recorded for the other consonant sounds during the field study discussed in detail in chapter III.

7.3.2 Acoustic analysis of diphthongs

Two experiments were set for the acoustic analysis of the obtained speech signals from the raw word list of twenty eight words that may carry language diphthong sounds: (i) analysis based on the speech samples of the isolated word utterances and (ii) the analysis based on the speech samples recorded as a carrier sentence containing the target diphthong word from the raw word list. The reason to carry out analysis in two ways is that there are temporal differences, when speakers speak isolated words compared to the words embedded in a carrier sentence. A diphthong is the combination of two vocalic elements with three significant analysis points: (i) on-glide (the duration of the first vocalic element) (ii) transition (the duration of the shift between first and second vocalic elements), and (iii) off-glide (the duration of the second vocalic element) (Gay, 1968). In order to quantify the acoustic characteristics of diphthongs and glides in Sindhi; the main acoustic parameters measured are the first two formants (F1, F2), and the duration of three analysis points (on-glide, off-glide and the transition segment). The speech samples obtained in two ways are gathered and the signals of the diphthong and glide segments were extracted from the word utterances and analyzed. The segment boundaries for the diphthong and glide phonemes in the word utterances were marked manually by inspecting the spectrograms of the word utterances. Spectrograms of the corresponding twenty eight word utterances, shown in Table 7.1, were computed for the analysis of the diphthongs of Sindhi.

Note that the diphthongs listed in this study may not be the only diphthongs present in the language inventory; because Sindhi is a flexible language in terms of altering the word pronunciation with a vowel insertion or digestion in between two consonants i.e. the word *hikɽo* "one" and the word *j^hupɽi* "shack" can be alternatively pronounced as *hikəɽo* and *j^hupɪɽi* by inserting the short vowels /ə/, and /ɪ/ respectively (Jennifer, 2006). The other factor is the rate of the speaker speech varies from speaker to speaker; some speak slowly, while others may speak the same utterance quickly, the slow utterance of the glides may result in a diphthong sound (Raphael, 2006). The first two formants provide fundamental cues for the perception of the monophthong vowels; however the second formant is considered one of the most fundamental acoustic cues for the perception of diphthongs (Aiza, 2004; Borzone de Manrique, 1979; Jha, 1985). The acoustic analysis and the conclusions on the basis of the obtained results in the subsequent sections are drawn keeping in view the following characteristics of diphthongs in Sindhi:

- the diphthongs should have two vocalic targets (Ladefoged, 2001)
- the type of the diphthong falling or rising
- the transition between two vocalic elements is not a hiatus
- the transition between two vocalic elements is a diphthong transition or it is a vowel-to-glide transition.

Table 7.1 List of the words potentially contain the diphthong phoneme of Sindhi (Keerio, 2010).

IPA Symbol	S.no	Phonetic transcription	Sindhi
əy	1	əynəkə	عينڪ
	2	səyrʊ	سَئير
	3	əybʊ	عَيبُ
ɪə	4	siə	سِيءَ
oɪ	5	hiə	هِيءَ
	6	poɪ	پوءِ
	7	goɪ	گوءِ
ʊi	8	hʊi	هُئي
	9	pitʊi	پِيٺُي
ɪe	10	t^{h}ɪe,	ٿِئي
	11	pɪe	پِئي
ɑɪ	12	lɑɪ,	لاءِ
	13	pɑɪ,	پاءِ
	14	ɓʊd^{h}ɑɪ	ٻُڌاءِ
uɪ	15	khushbuɪ	خوُشبوُءِ
əw	16	ʧəvəndəw	چَوندئو
	17	b^{h}əw	ڀَئو
uə	18	huə	هُوءَ
ɑʊ	19	mɑʊ	ماءُ
	20	pɑʊ	پاءُ
iʊ	21	siʊ	سِيءُ
	22	piʊ	پِيءُ
	23	ʤiʊ	جِيءُ
ɪə	24	d^{h}ɪə	ڌِيءَ
ʊə	25	sʊər	سُئر
	26	ʤʊər	جُئر
ʊo	27	pʊo	پُئو
	28	d^{h}ʊo	ڌُئو

To accomplish the above mentioned characteristics of diphthongs a set of acoustic properties for each diphthong were measured based on the speech signals of the twenty eight raw word utterances. The formant frequencies for the first four formants (F1, F2, F3, and F4) were obtained using the Burg LPC (linear predictive coding) algorithm. The presence of the two vocalic elements in diphthongs and glides are determined by computing the z-scores for the first two formants according to the method described by authors (Hongyan, 2007; Lobanov, 1971). The z-scores for first two formants of each diphthong and two glides were computed by subtracting the mean of the first formant from the raw F1 values and the mean of the second formant from the raw F2 values. The differences were then divided by the standard deviation of the raw formants values. The reason for computing the z-scores of the first two formants is to determine whether there are two vocalic elements corresponding to the diphthong phoneme or the speech signals belong to glide or other non-diphthong phoneme. In this study the obtained negative (-ve) z-scores belong to one vocalic element and the positive (+ve) z-scores belong to the other vocalic element. If all the obtained z-scores are either +ve or -ve or they are very irregular, (say +ve, –ve and again +ve, -ve and vice versa), then the phoneme is considered a non-diphthong phoneme of Sindhi. Figures 7.7 and 7.8 below show the computed z-scores for two glides of Sindhi and figures 7.9 to figure 7.15 show the z-scores for the diphthongs of Sindhi.

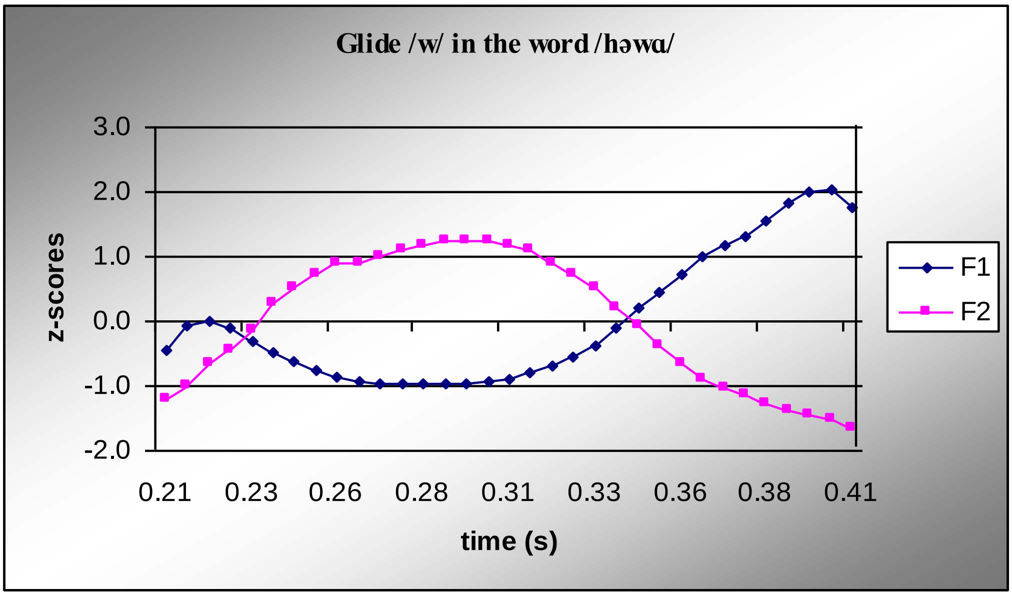


Figure 7.7 Typical first two formants z-scores of the glide /w/ present in the word /həwɑ/ هوا of Sindhi. Note that the z-scores for both F1 and F2 are irregular.

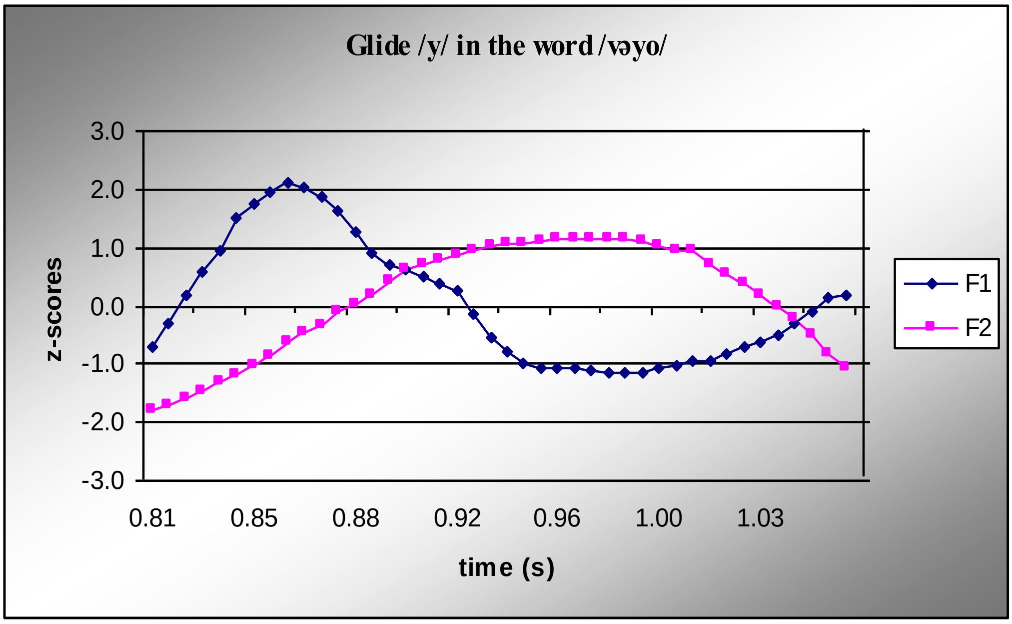


Figure 7.8 Typical first two formants z-scores of the glide /y/ present in the word /vəyo/ويو of Sindhi. Note that the z-scores for both F1 and F2 are irregular.

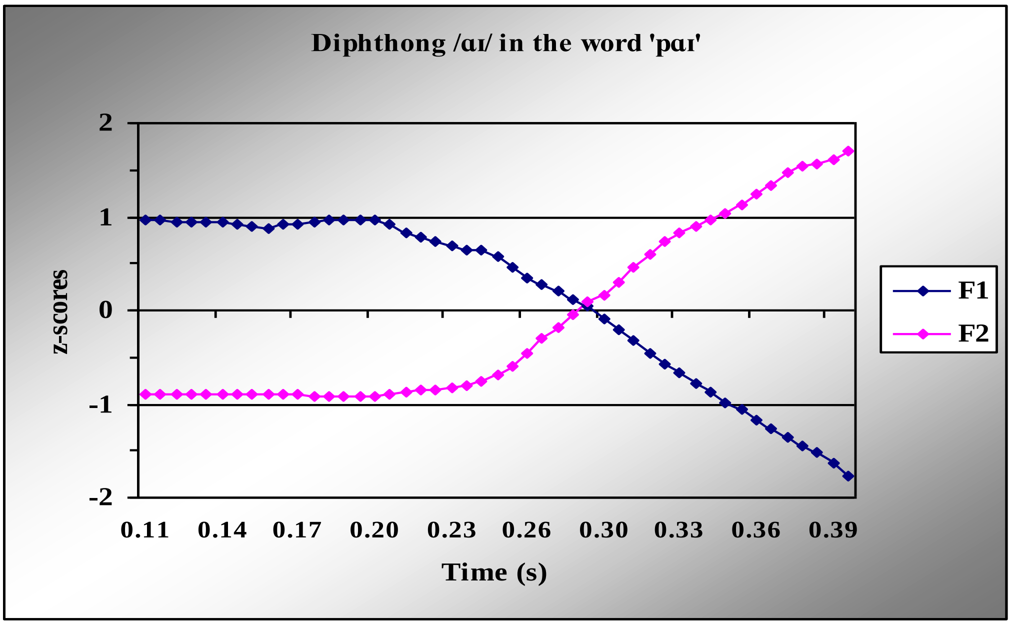


Figure 7.9 Typical first two formants z-scores of the diphthong [ɑɪ] present in the word utterance ‘p[ɑɪ]’, of Sindhi.Note that F1 starts with +ve z-scores and moves gradually to the –ve z-scores and in similar way the F2 z-scores move gradually from –ve to +ve z-scores.

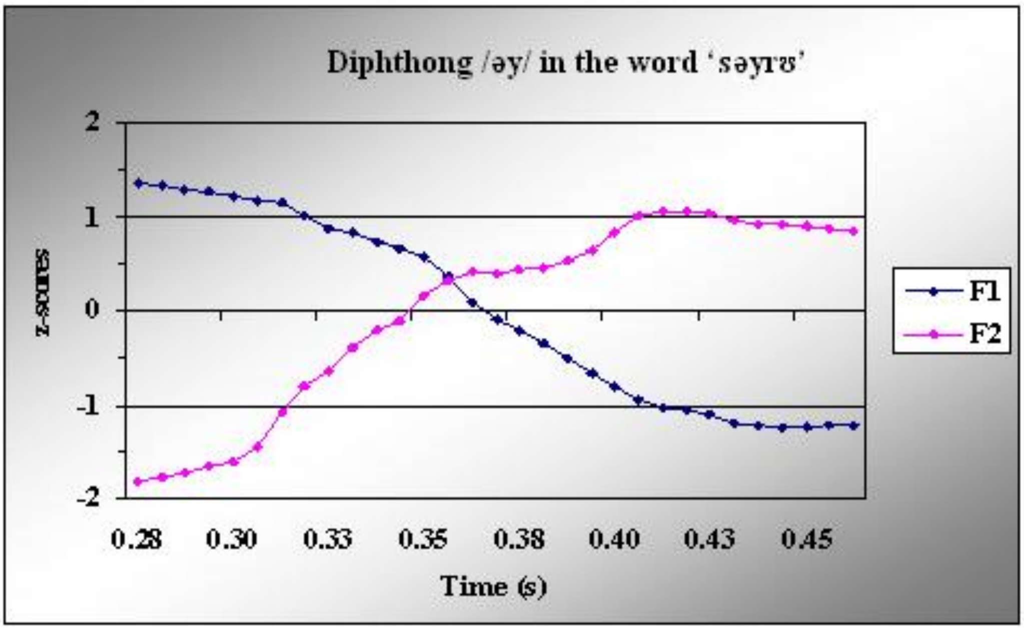


Figure 7.10 Typical first two formants z-scores of the diphthong [əy] present in the word utterance 's[əy]rʊ', of Sindhi. Note that F1 starts with +ve z-scores and moves gradually to the –ve z-scores and in similar way the F2 z-scores move gradually from –ve to +ve z-scores.

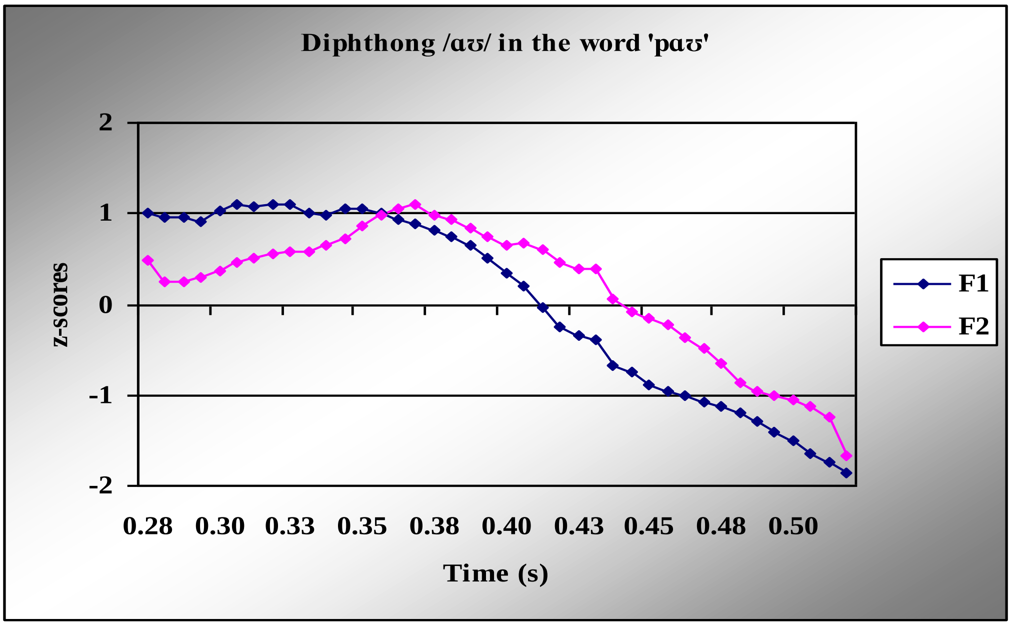


Figure 7.11 Typical first two formants z-scores of the diphthong [ɑʊ] present in the word utterance 'p[ɑʊ]', of Sindhi. Note that both F1 and F2 start with +ve z-scores and move gradually to the –ve z-scores.

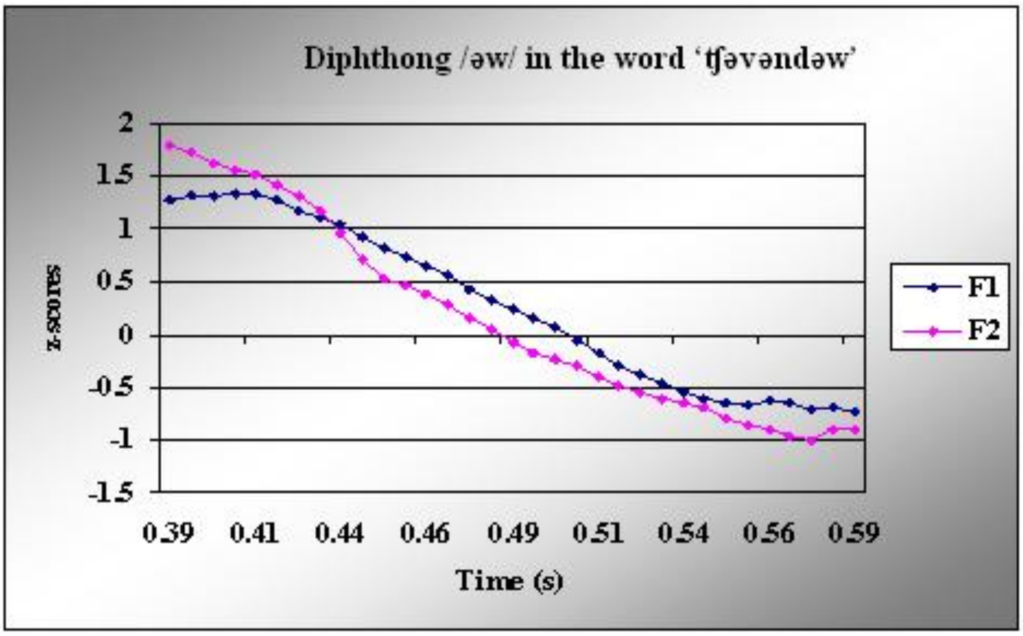


Figure 7.12 Typical first two formants z-scores of the diphthong [əw] present in the word utterance 'ʧəvənd[əw]', of Sindhi. Note that both F1 and F2 start with +ve z-scores and move to the –ve z-scores.

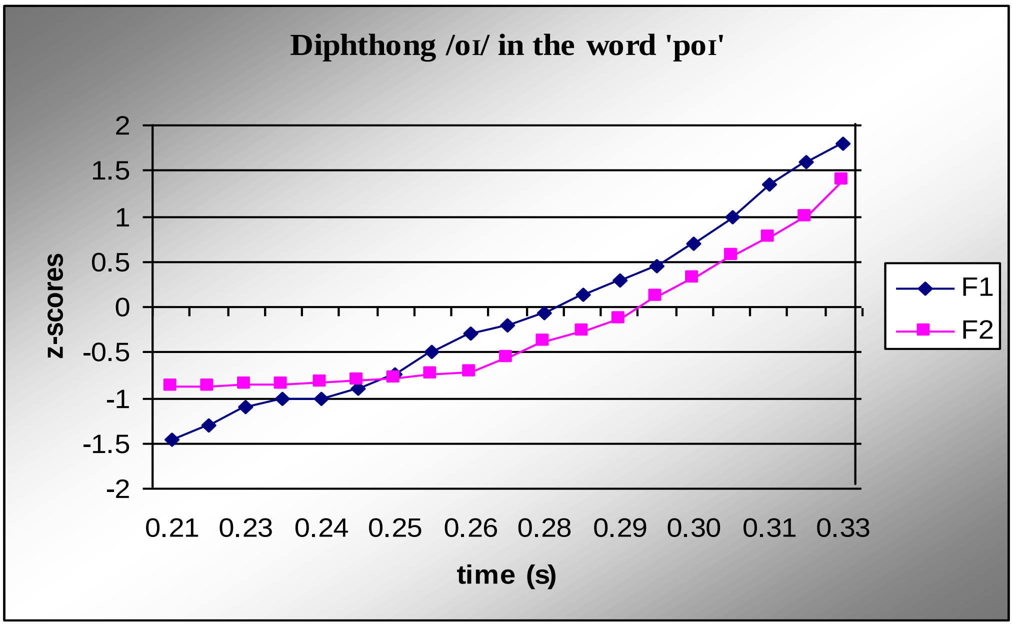


Figure 7.13 Typical first two formants z-scores of the diphthong [oɪ] present in the word utterance 'p[oɪ]', of Sindhi. Note that both F1 and F2 start with -ve z-scores and gradually move to the +ve z-scores.

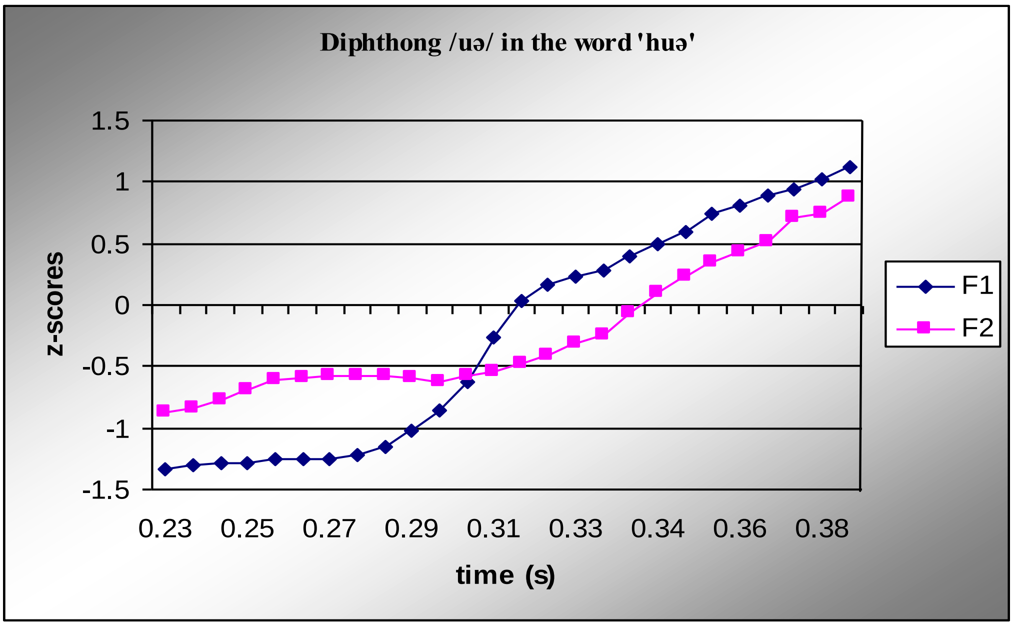


Figure 7.14 Typical first two formants z-scores of the diphthong [uə] present in the word utterance 'h[uə]', of Sindhi. Note that both F1 and F2 start with -ve z-scores and gradually move to the +ve z-scores.

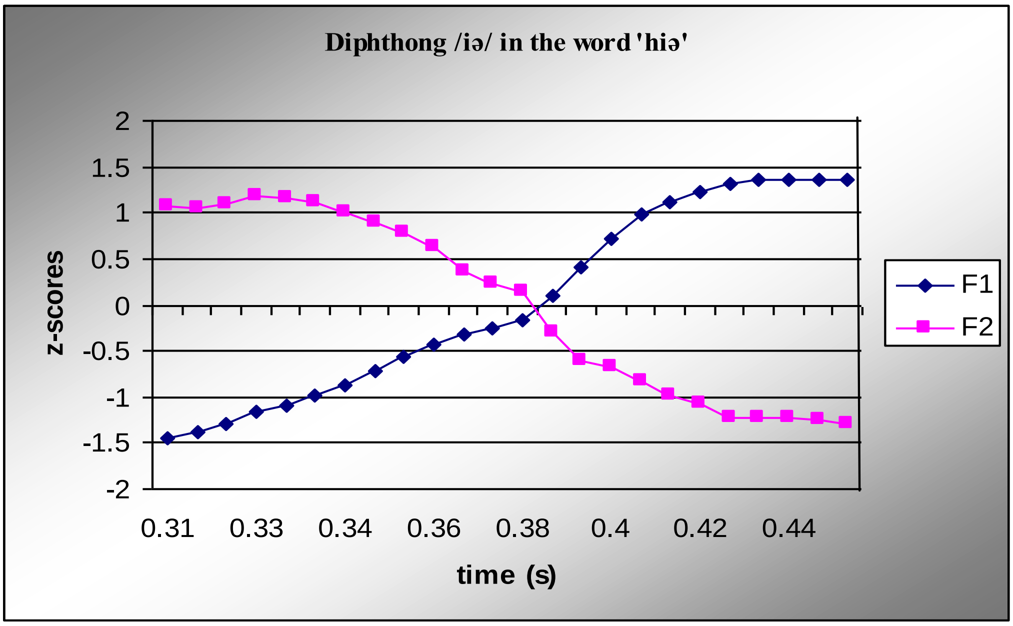


Figure 7.15 Typical first two formants z-scores of the diphthong [iə] present in the word utterance 'h[iə]', of Sindhi. Note that the F1 starts with –ve z-scores and move gradually to the +ve z-scores and F2 start with +ve z-scores and gradually move to the -ve z-scores.

Figures 7.9 and 7.10 show that the diphthongs with the transition from open back vowel (first vocalic element) to the close front vowel (second vocalic element) yield +ve F1 z-scores for the first vocalic element and -ve z-scores for the second vocalic element and the corresponding F2 z-scores for the first vocalic element are -ve and +ve, respectively for the second vocalic element. Figures 7.11 and 7.12 show that the diphthongs with the transition from open back vowel (first vocalic element) to the close back vowel (second vocalic element) yield +ve both F1 and F2 z-scores for the first vocalic element and –ve both F1 and F2 z-scores for the second vocalic element. Figures 7.13 and 7.14 show that the diphthongs with the transition from close back vowel (first vocalic element) to the close front vowel (second vocalic element) yield -ve both F1 and F2 z-scores, for the first vocalic element and +ve both F1 and F2 z-scores for the second vocalic element. Figure 7.15 shows that the diphthongs with the transition from close front vowel (first vocalic element) to the open mid vowel (second vocalic element) yield -ve F1, z-scores for the first vocalic element and +ve z-scores for the second vocalic element and the F2 z-scores are +ve for first vocalic element and -ve for the second vocalic element.

A significant difference among the z-scores of the diphthongs in Sindhi is observed; however the sign change from +ve to –ve and vice versa for the first two formants among two vocalic elements in diphthongs is of more importance. If there is a sign transition between the (F1, F2) z-scores of the two vocalic elements (first and second) the presence of a diphthong phoneme in the utterance is guaranteed. Whereas the irregular sign transition of the z-scores among the two vocalic elements confirm the presence of a non-diphthong phoneme in the utterance as shown in figures 7.7 and 7.8 for two glide consonants of Sindhi. In other words more than one F1 and F2 z-score crossovers prove that the phoneme is not a diphthong of Sindhi. Table 7.2 below summarizes the results obtained for the twenty eight words analysed for the diphthongs of Sindhi.

Table 7.2 List of the diphthong words after performing acoustic analysis (Keerio, 2010).

IPA Symbol	S.no	Phonetic transcription	Sindhi	Diphthong
əy	1	əynəkə	عينڪ	Yes
	2	səyrʊ	سَئير	Yes
	3	əybʊ	عَيبُ	Yes
ɪə	4	siə	سِيءَ	Yes
	5	hiə	هِيءَ	Yes
oɪ	6	poɪ	پوءِ	Yes
	7	goɪ	گوءِ	Yes
ʊi	8	hʊi	هُئي	No
	9	pitʊi	پِيٽُي	No
ɪe	10	t^hɪe,	ٿِئي	No
	11	pɪe	پِئي	No
ɑɪ	12	lɑɪ,	لاءِ	Yes
	13	pɑɪ,	پاءِ	Yes
	14	ɓʊd^hɑɪ	ٻُڌاءِ	Yes
uɪ	15	khushbuɪ	خوُشبوُءِ	No
əw	16	ʧəvəndəw	چَوندئو	Yes
	17	b^həw	ڀَئو	Yes
uə	18	huə	هُوءَ	Yes
ɑʊ	19	mɑʊ	ماءُ	Yes
	20	pɑʊ	پاءُ	Yes
iʊ	21	siʊ	سِيءُ	Yes
	22	piʊ	پِيءُ	Yes
	23	ʤiʊ	جِيءُ	Yes
ɪə	24	d^hɪə	ڌِيءَ	No
	25	sʊər	سُئر	No

-Continued on next page -

ʊə	26	ʤʊər	جُئر	No
ʊo	27	pʊo	پُئو	No
	28	d^hʊo	ڏُئو	No

Diphthongs are considered as moving voicing elements from the beginning vowel to the target vowel (Gay, 1968; Raphael, 2006). The acoustic F1xF2 vowel plot of figure 7.16 shows how formants gradually move for the two vocalic elements of diphthongs and in which direction. The solid and dashed line arrows in figure 7.16 below indicate the direction of the formants movement from the first vocalic element (the onset) to the second vocalic element (the offset vowel). The solid line arrows are drawn for the direction of all falling diphthongs from the first vocalic element (the onset) to the second vocalic element (the offset vowel) and the dashed line arrows are drawn for the direction of all rising diphthongs from the first vocalic element (the onset) to the second vocalic element (the offset vowel) of Sindhi diphthongs. The count of diphthongs in Sindhi, shown in figure 7.16 below, may vary; because of the fact that Sindhi is a flexible language and the rate of the speaker's speech for example the slow utterance of the glides and the dialect specific insertion and digestion of the short vowels could potentially lead to the increase or decrease in the count of the language diphthongs. The formant movement of the diphthongs /ɑɪ/, /əy/, and /oɪ/ of Sindhi typically starts with the open vowel and end with the close-front vowel this is shown in figures 7.16 and 7.17. The formant movement of the diphthongs [ɑʊ] and [əw] of Sindhi typically starts with the open vowel and end with the close-back vowel this is shown in figures 7.16 and 7.18. The formant movement of the diphthongs [iʊ] and [iə] of Sindhi typically starts with the close-front and end with the open-back vowel this is shown in figures 7.16 and 7.19. The formant movement shown in the spectrogram figures 7.17-7.19 for the diphthongs of Sindhi show the gradually changing formant pattern without hiatus or the formant discontinuities for the two vocalic elements.

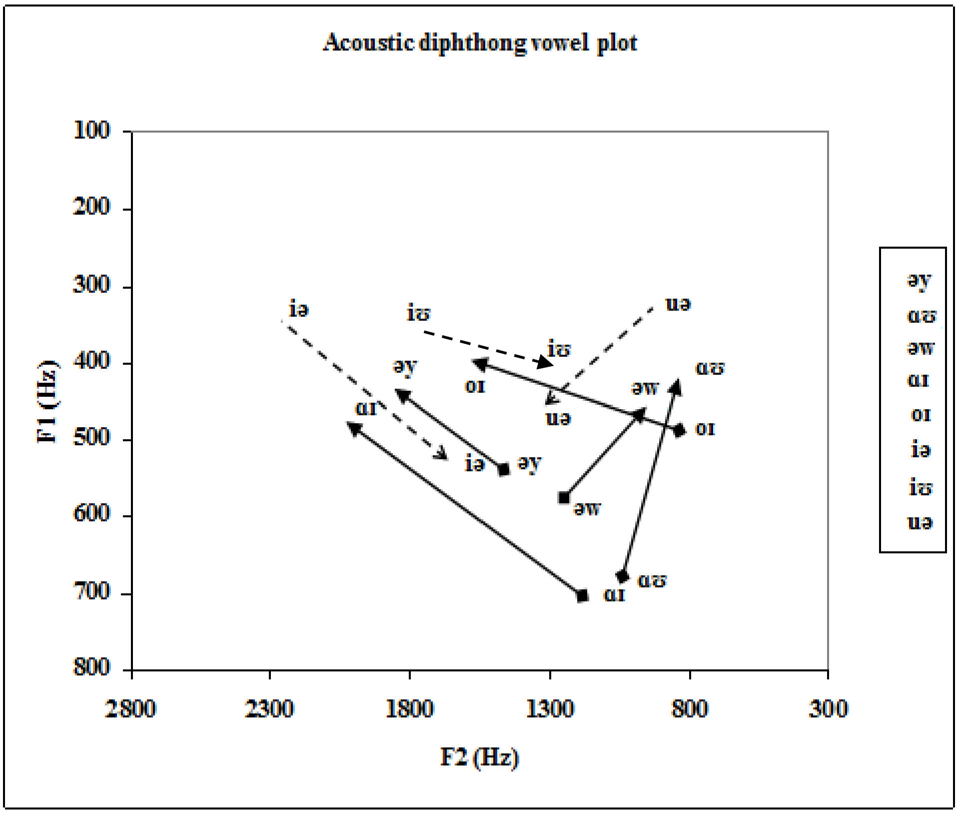


Figure 7.16 Acoustic F_1 x F_2 diphthong vowel plot of Sindhi.

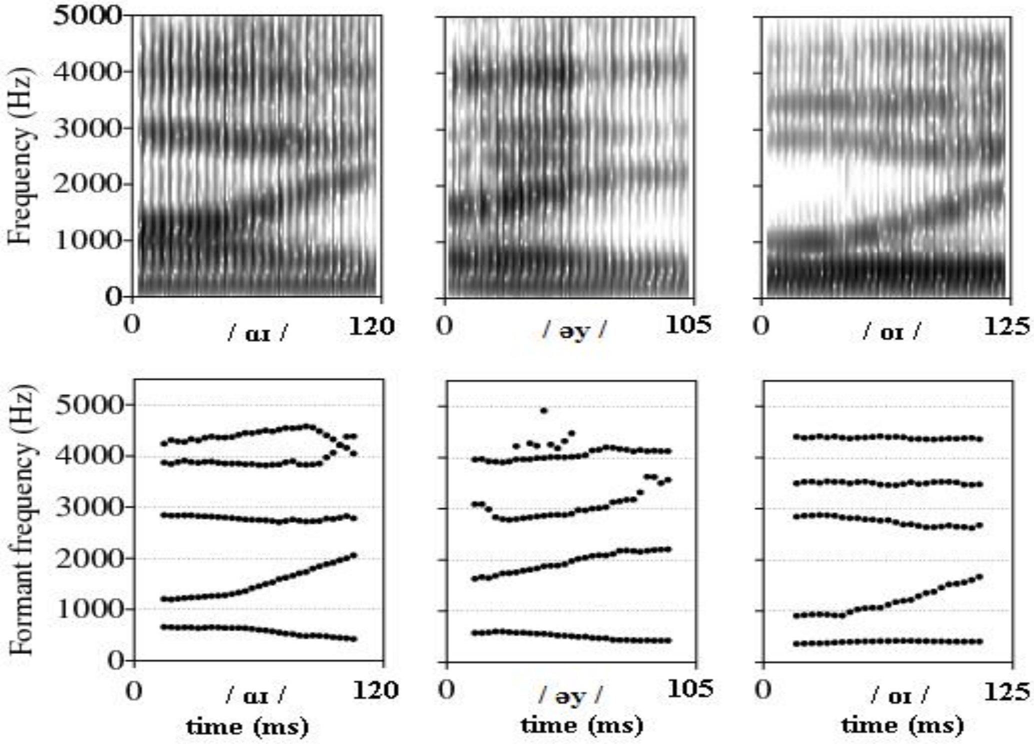


Figure 7.17 Spectrogram of three diphthongs /aɪ/, /əy/, and /oɪ/ of Sindhi along with the formant tracks.

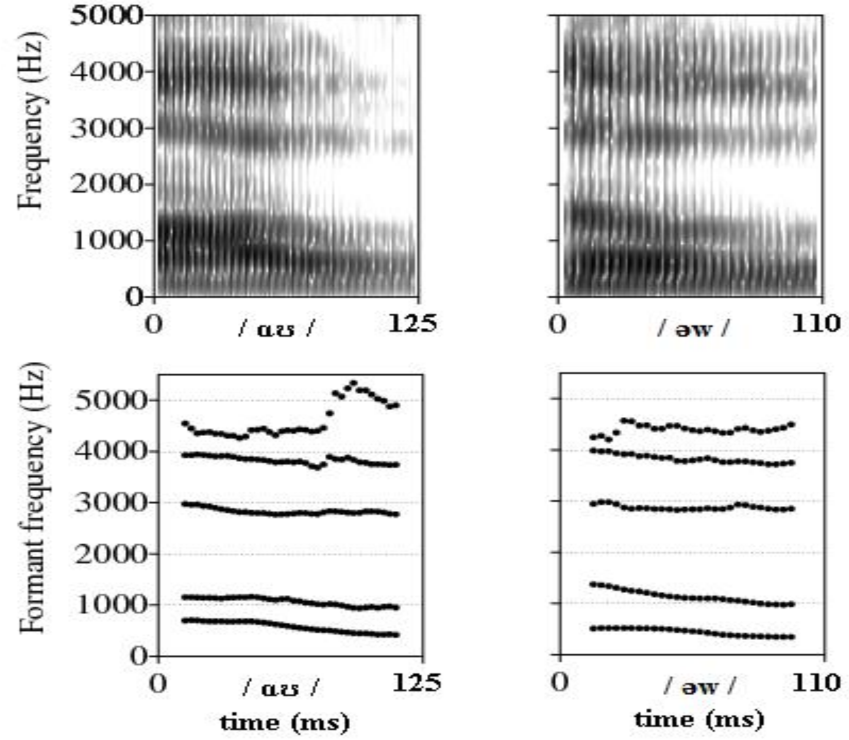


Figure 7.18 Spectrogram of two diphthongs /ɑʊ/, and /əw/ of Sindhi along with the formant tracks

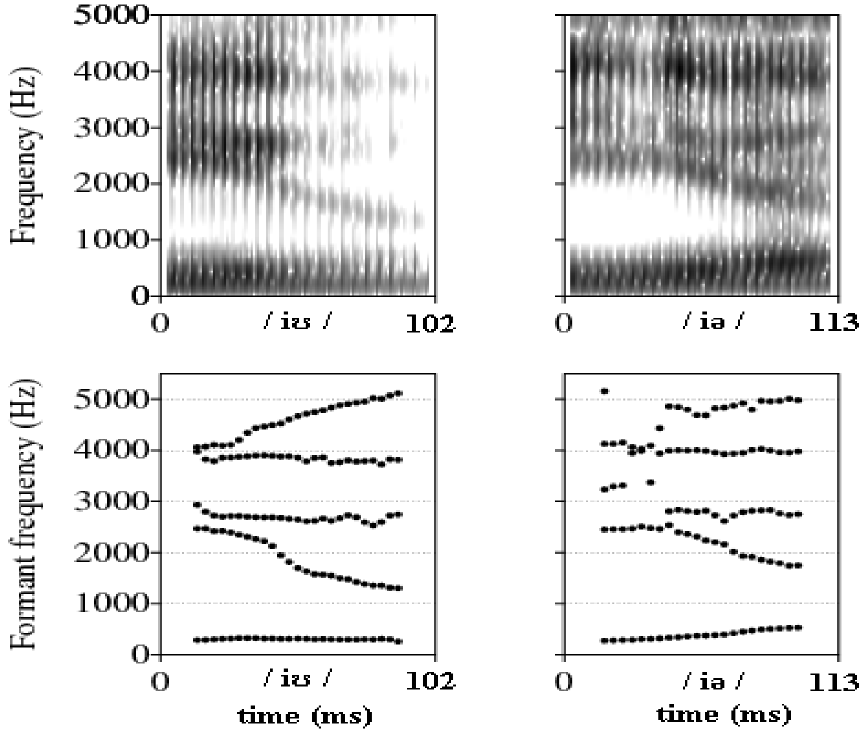


Figure 7.19 Spectrogram of two diphthongs /iʊ/ and /iə/ of Sindhi along with the formant tracks

7.4 Conclusion

Two types of Sindhi sounds (diphthongs and glides) are analysed and compared acoustically in this chapter. The experiments based on the captured voice samples were conducted to analyse the phonetic realization of diphthong phonemes in Sindhi. Diphthongs and glides are considered sounds with moving formant pattern, therefore the formant transitions for these sounds were analysed at three critical analysis points (onglide, offglide and transition). The two vocalic elements present in diphthong phonemes are determined by analysing the sign transition in computed z-scores for the first two formants. It is observed from the computed z-scores for first two formants in section 7.3.2 that the glide consonants have more than one crossovers and hence donot form a diphthong phoneme in Sindhi, regardless of its position in utterance; whereas the computed z-scores for diphthong phonemes show one sign transition (crossover) for the computed z-scores of first two formants. The diphthongs with the transition from open back (first vocalic element) to the close front vowel (second vocalic element) have positive F1 z-scores for the first vocalic element, and negative for second vocalic element; whereas the F2 z-scores for the first vocalic element are negative and positive for the second vocalic element. The diphthongs with the transition from open back (first vocalic element) to the close back vowels (second vocalic element) have positive both F1 and F2 z-scores for the first vocalic element and negative both F1 and F2 z-scores for the second vocalic element. The diphthongs with the transition from close front (first vocalic element) to the open back vowel (second vocalic element) have negative F1, z-scores for the first vocalic element and positive F1, z-scores for the second vocalic element and the F2 z-scores are positive for first vocalic element and negative F2 z-scores for the second vocalic element.

It is observed that the presence of diphthongs in Sindhi mainly depends on the pronunciation and the rate of speaker’s speech. In addition to this the language’s flexibility of inserting and digesting a short vowel in the utterances could potentially increase or decrease the count of diphthongs in Sindhi. The obtained acoustic analyses results in this study suggest that when a short vowel immediately follows a long vowel (Vv) in a syllable at the word final position, the combined production of the two sounds result in a diphthong phoneme of Sindhi. The analysis results also show that a short and a long vowel (vV) phoneme sequence in a syllable or in a word utterance do not form a diphthong sound of Sindhi. The diphthongs [əy] and [əw] are special in a sense that how speakers are pronouncing them; for example the word ‘s[əy]rə’, is quite frequently pronounced as ‘s[e]rə’, and the word ‘s[əw]k^ho’, is pronounced as ‘sok^ho’. Phonetically the latter pronunciation of the two words is wrong and hence both the words in the latter pronunciation do not form a diphthong phoneme of Sindhi.

CHAPTER 8

ACOUSTIC ANALYSIS OF CONSONANT SOUNDS OF SINDHI

Chapter 8

Acoustic Analysis of Consonant Sounds of Sindhi

8.1 Introduction

In this chapter the acoustic-phonetic characteristics of the consonant phonemes of Sindhi are presented. In order to measure the acoustic features of the consonants each consonantal phoneme was framed in a VCV sequence of phonemes in the word utterances (the word utterances were selected according to the minimal pair method discussed in detail in chapter IV). The acoustic analysis for the consonantal phonmes is carried out in two ways: first the consonantal phonemes embedded in the word utterance and pronounced as isolated word utterances; and the second is the words embedded in the meaningful language sentences read by speakers as continuous speech samples. The two way analysis of these sounds enables us to highlight the acoustic variation of these sounds when produced by changing the sounds or the sequence of sounds in the utterances referred to as the contextual variety of these sounds in Sindhi. The two way analysis of these sounds helps to determine the changing acoustic behavior of the consonants when vocal tract configurations are set to produce only the target analysis phoneme part of the isolated word utterance and when the vocal tract configurations are set to produce the next utterance once the production of the target phoneme is finished part of the carrier sentence read by speakers. When the vocal tract configurations are set to produce only the target analysis phoneme; there is enough time for the moving articulators to produce the utterane and come back to the normal position. Whereas in continuous speech the articulators when they produce one utterance can not move quickly to come back to the normal position and then move for the production of the next utterance in sequence. In this situation the moving articulators move from the position of the previous utterance to the production of next utterance in sequence. Therefore the most significant acoustic difference between the two types of speech signals is the duration; however the other differences are also observed and are discussed in the subsequent sections of this chapter. The mean acoustic parameter values and standard deviation are given in a separate table for each class of the consonantal phonemes present in Sindhi.

The history of phonetic research regarding the classification of the consonantal phonemes of Sindhi shows that these sounds have been classified so far on the basis of the available articulatory-phonetics knowledge which refers to the work of authors (Jatoi, 1996; Jennifer, 2006; Raza, 2004). To the best of our knowledge and literature survey there is no evidence that the acoustic-phonetics based knowledge of these sounds has been incorporated for the identification and classification of these sounds phonetically. The importance of the acoustic-phonetic features has been widely accepted in determining the place of articulation and the voicing feature of the consonant sounds in phonetics (Ali, 2001; Zue, 1979). Therefore the importance of the acoustic-phonetic based knowledge for the identification and classification of the consonantal phonemes of Sindhi cannot be avoided especially when determining the place of articulation and the voicing feature of these sounds. This work incorporates the obtained acoustic-phonetic features of Sindhi consonants in addition to the available articulatory-phonetics knowledge for their identification and classification. In addition to this the acoustic-phonetic characteristics obtained for these four consonants (/ʤ/ ج, /$ʤ^h$/ جھ , /ʧ/ چ, /$ʧ^h$/ ڇ) of Sindhi will enable us to associate them with a proper consonant class; because the classification of these four consonants have been controversially reported in the literature for example the authors Raza et al.(2004) and Jennifer(2006) have associated them with the members of the affricate class of consonant sounds; whereas the authors (Jatoi, 1996; Ladefoged, 1996; Mirza, 2006) associated them as the members of the stop class of the consonants in Sindhi. The only notable work available in the published literature on the subject of the acoustic-phonetic analysis of the implosive consonants of Sindhi is by Raza et al.(2004). The study carried out by Raza et al.(2004) is mainly focused on the acoustic analysis of the implosive consonants of Sindhi; however the speech samples for the analysis were only collected from the district (Jacobabad) of the *Utraddi* dialect. The speech samples for the analysis did not include the other four dialects (*Middle, Lasi, Thareli, and Lari*) of Sindhi mainly spoken in the Sindh province of Pakistan. Therefore the acoustic analysis for the implosive consonants of Sindhi presented in Raza et al.(2004) are not yet sufficient to be incorporated into the design of the today's state-of-the-art Sindhi ASR system.

The lack of such a study that comprehensively presents the acoustic-phonetic analyses of the language consonant sounds provides the motivation to conduct this study. In chapter III the data collection process for the acoustic analyses of the consonant sounds of Sindhi is discussed in detail, it covers all five dialects of Sindhi spoken in the Sindh province of Pakistan. In chapter IV the classification of the consonant sounds is discussed on the basis of the three articulatory qualities of these sounds i.e the voicing feature, the place of articulation and the manner of the articulation. The consonant sounds that share some common articulatory characteristics are also grouped in one natural class i.e. the stops, nasals, fricatives etc in chapter IV. In this chapter the acoustic analysis for each class of the consonant sounds is given individually keeping in view the fact that the consonant sounds that are the members of one natural class share some common articulatory as well as the acoustic-phonetic characteristics (Olive, 1993; Kent, 2002; Stevens, 2000). The grouping of the consonant sounds that share some common characteristics formed seven major consonant classes of Sindhi named as: stops (plosives and implosives), nasals, fricatives, affricates, flaps, glides and liquids. The main acoustic parameters measured for each consonant class are: the fundamental frequency, formant frequencies, duration, and the energy etc. The two glide consonants of Sindhi are discussed in chapter VII with the discussion of the diphthongs; because of the fact that the acoustic characteristics of these sounds correlate well with the diphthong phonemes. This grouping of the consonant sounds leaves twenty consonant sounds member of the stop class as the largest consonant class in Sindhi, four are the members of implosive class, five are the members of nasal class, seven are the members of fricative class, one consonant is the member of each flap, lateral and trill class and two consonants are the members of glide class discussed in chapter VII. The acoustic analyses presented in the subsequent sections of this chapter are based on the speech samples collected during the field study discussed in chapter III.

8.2 Stop consonants of Sindhi

In this section we will investigate the acoustic-phonetic characteristics of the consonant sounds members of the stop class of Sindhi. The analysis is mainly focused upon determining the three main acoustic characteristics of the stops as: (i) whether the stop consonant is voiced/ unvoiced, (ii) whether the stop consonant is aspirated or non-aspirated and (iii) whether the stop consonant is plosive stop or implosive stop or it is the member of the affricate class of Sindhi. The stops in Sindhi are produced at five places of the articulation, they are: (i) bilabial stops produced with both the lips include the phones: /b/ ب, /b^h/ ڀ, /p/ پ, and /p^h/ ڦ of Sindhi (ii) dental stops produced with the tip of the tongue touching behind the upper teeth include the phones: /d/ د, /d^h/ ڌ, /t/ ت, and /t^h/ ٿ of Sindhi (iii) retroflex stops produced with the tip of the tongue curled up in the direction of the hard palate but do not touch it include the phones: /ɖ/ ڊ, /ɖ^h/ ڍ, /ʈ/ ٽ, and /ʈ^h/ ٺ of Sindhi (iv) palato-alveolar stops are produced by raising the front of the tongue behind the alveolar ridge in the direction of the hard palate include the phones: /ʤ/ ج, /ʤ^h/ جھ, /ʧ/ چ, and /ʧ^h/ ڇ of Sindhi (v) velar stops produced by raising the back of the tongue in the direction of the soft palate include the phones: /g/ گ, /g^h/ گھ, /k/ ڪ, and /k^h/ ک of Sindhi. There are three critical analysis points for all the stops in Sindhi: (i) the closure segment where the articulators completely block the air passage through the vocal tract (ii) the release segment where the burst of the air immediately released after the closure segment with an expulsion of breath and sound (iii) the formant transitions coming into the stop closure from the preceding sound and the transitions going away from it to the immediately following sound. Acoustically the role of the formants becomes important if the formant transitions are sharp and stable particularly in the closure segment of the voiced stops; otherwise they are no more helpful for the classification and identification of the stop consonants (Ali, 2001; Stevens, 1974). The role of the rise in the amplitude of the modulation signal at the beginning of the release segment of the stops can play an important role for the identification of the unvoiced stops particularly (Ohde, 1983). The use of the burst amplitude for the identification of the voiced stops can be helpful if the speech signals of the closure segment are weaker than the signals of the release segment.

In the subsequent sections of this chapter the acoustic analysis results and discussions are presented keeping in view mainly the role of the formants, the burst amplitude, the presence or absence of the energy and the duration of the two segments (closure and release) for the classification and identification of the stop consonants of Sindhi.

8.2.1 Voiced bilabial stops

Bilabial stops of Sindhi are articulated with both the lips; tightly closed during the closure segment and relaxed during the release segment of the stop. When lips are tightly closed together the complete closure of the vocal tract occurs and the air pressure builds up behind the lips in the oral cavity. The sudden release of the air pressure (with an expulsion of breath) when two lips are relaxed, four bilabial stops of Sindhi produced in this way are two non-aspirated plosives /b/ ب , and /p/ پ and two aspirated plosives /b^h/ ڀ and /p^h/ ڦ. Figure 8.1(a) shows the typical waveform of the bilabial stop /b/ of Sindhi extracted manually from the word utterance /səbəkʊ/ سَبق and figure 8.1(b) shows the typical waveform of the bilabial stop /b^h/ of Sindhi extracted manually from the word utterance /səb^{h}ʊ/ سَڀُ. The waveforms shown for the closure and release segments of the stops in this chapter are extracted manually from the word utterances embedded in a carrier sentence by inspecting the waveforms and spectrograms of the utterances visually. The manual extraction of these segments is preferred over the automatic segmentation for the sake of achieving a highly accurate segmentation mark for the closure and release segments.

The phoneme sequence for the word utterances was organized in such a way so that the target stop phoneme follows the VCV phoneme sequence. This VCV phoneme sequence for the analysis of the target stop phoneme is adopted for three main reasons: (i) to avoid creating any consonant clustered speech signals (ii) the stops when surrounded by vowels (as VCV phoneme sequence) in an utterance and articulated; there happens a complete vocal tract closure (Stevens, 1993; Stevens, 1998) and (iii) the analysis of the formant transitions coming into the closure segment from the preceding sonorant sound (the vowel in this case) and the formant transitions going away from the stop consonant to the sonorant sound (the following vowel in this case). The waveforms shown for the stop consonants in the subsequent sections of this chapter are manually extracted from the word utterance to show only the speech signals corresponding to the VCV phoneme sequence. The two critical segments (the closure and release) of each stop consonant are marked with arrows; the speech signals between the points ***a*** and ***b*** mark the closure segment and the signals between the points ***b*** and ***c*** mark the release segment of the stop consonant. The significant difference between the closure and the release segments for all the stop consonants of Sindhi observed uniformly is the duration; that is the duration of the closure segment is much longer compared to the duration of the release segment. This difference for the two bilabial stops /**b**/ and /$\mathbf{b^h}$/ of Sindhi is shown in figures 8.1(a) and (b). The slow varying periodic signal activity in the closure segment of the two stops /**b**/ and /$\mathbf{b^h}$/ indicate that these are voiced phonemes of Sindhi. The duration of the closure segment (signals between points ***a*** and ***b***) is much longer compared to the release segment (signals between points ***b*** and ***c***) of the two sounds /**b**/ and /$\mathbf{b^h}$/ is shown in the extracted short duration waveforms of the figure 8.2. By comparing the two stops /$\mathbf{b^h}$/ and /**b**/ it is observed that the stop /$\mathbf{b^h}$/ have shown longer duration for both segments (the closure and the release) than the stop /**b**/. The release segment of the stop **/b/ ب**, is smaller and is barely identifiable as shown in the waveform of figure 8.1(a) the signals between the points ***b*** and ***c***.

The fact that the bilabial stops are relatively weak compared to the alveolar and velar stops in English (Ali, 2001; Edwards, 1981; Fant, 1973; Hunt, 1989); is true for the bilabial stops of Sindhi as well. The bilabials of Sindhi are relatively weak compared to the dental, retroflex and velar stops of Sindhi. The presence of the weak energy in the lower frequency regions below 1500 Hz in the wideband spectrogram of the figure 8.3 along with the short length vertical periodic striations in the closure segment of the two stops /**b**/ and /$\mathbf{b^h}$/ of Sindhi indicate that the two sounds are voiced. The stable formant transitions are only observed for the first formant with most of the energy present in it; whereas the F2 is stable only at the start of the closure segment and becomes unstable as it moves towards the end of closure see the spectrogram of the figure 8.3 along with the formant tracks. The F1 is very low for the two stops; therefore the F1 motion is downward (the formant transitions coming into the stop closure from the preceding vowel) and upward (the formant transitions going away from it to the following vowel). The F2 transitions for the closure segment of the two stops are ambiguous and unstable; therefore the F2 motion can not be defined for these stops as shown in figure 8.3.

Stop consonants when followed by a sonorant sound (vowel) then the duration of the release segment, the signals between the point's ***b*** and ***c*** are referred to as the voice-onset-time (VOT) (Ali 2001, Olive 1993). The term VOT hereafter will be interchangeably used for the interval of the release segment throughout this study; because the target analysis stop phoneme in this study is framed in a VCV phoneme sequence. In English VOT for the bilabial stops is relatively shorter than the velar and alveolar stops (Ali, 2001; Edwards, 1981). However the VOT for the bilabial stop **/b^h/** of Sindhi is observed longer compared to the VOT of other stops; whereas the VOT of the bilabial stop /**b**/ is smaller than the VOT of all the other stops of Sindhi. The larger VOT for the bilabial **/b^h/** of Sindhi is due to the fact that the bilabial **/b^h/** is aspirated stop consonant and speakers usually prolong the aspirated sounds.

The stops of Sindhi show a great deal of acoustic variation among the speakers of the different dialects as well as the speakers of the same dialect when the sequence of sounds or sounds changed in the utterance. This variation in phonetics is considered as a contextual variety of the stop sounds (Ali, 2001; Olive, 1993). The acoustic features variation due to the contextual variety of these sounds, at least ten instances of each utterance was captured from the speakers for the purpose of the subsequent offline analysis process. The mean acoustic parameter values and the standard deviation for each stop consonant of Sindhi are given in the Table 8.1.

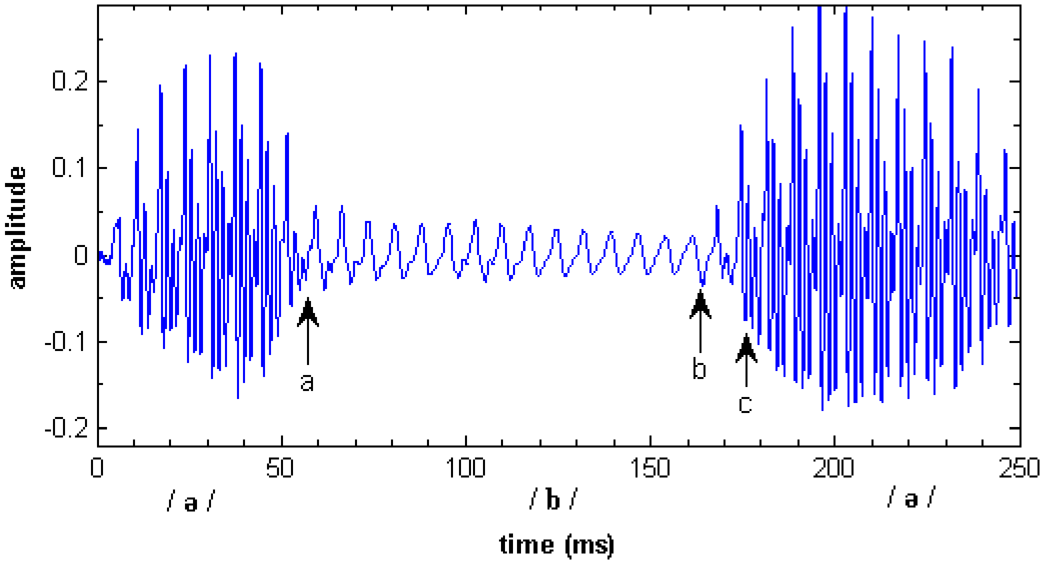


Figure 8.1(a) Waveform of the bilabial voiced stop /b/ of Sindhi, taken from the word utterance /səbəkʊ/ سَبق.

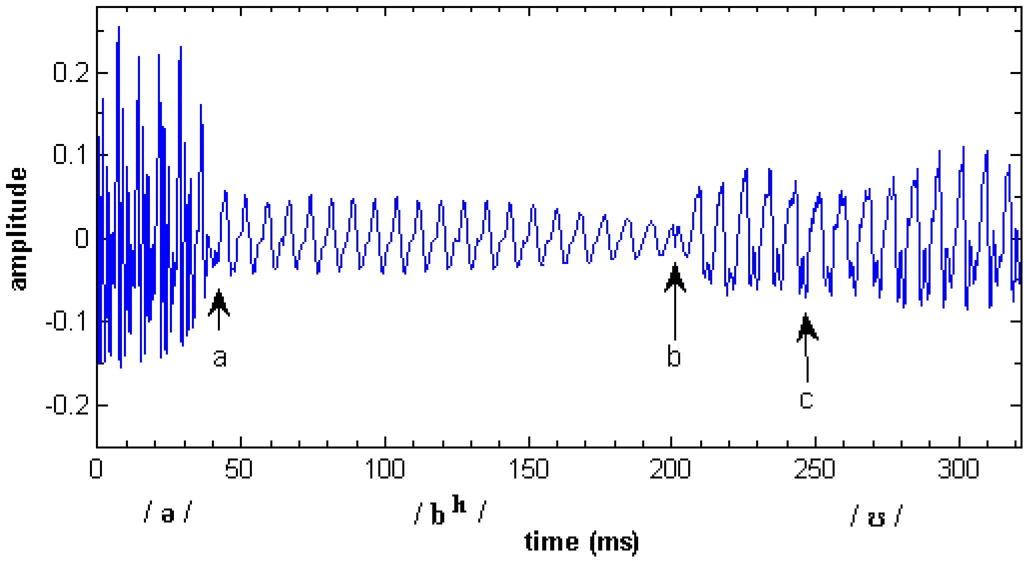


Figure 8.1(b) Waveform of the aspirated bilabial voiced stop /b^h^/ of Sindhi, taken from the word utterance /səbʰʊ/ سڀُ.

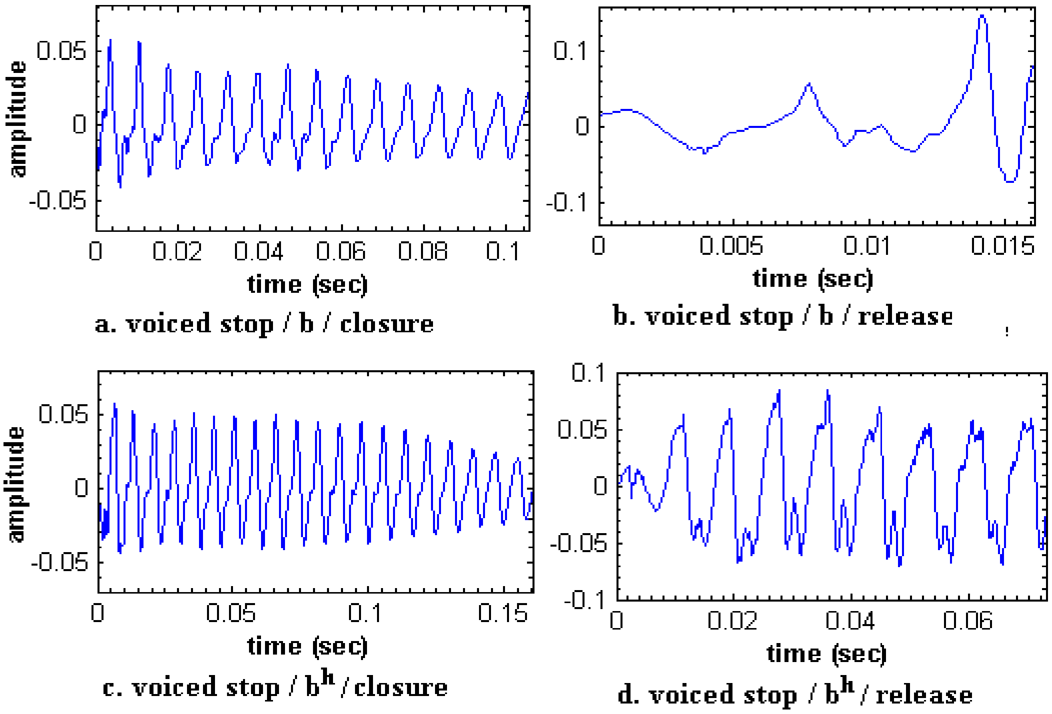


Figure 8.2 Extracted waveforms of the closure and release segments of the two voiced bilabial stops /b/ and /b^h/ of Sindhi.

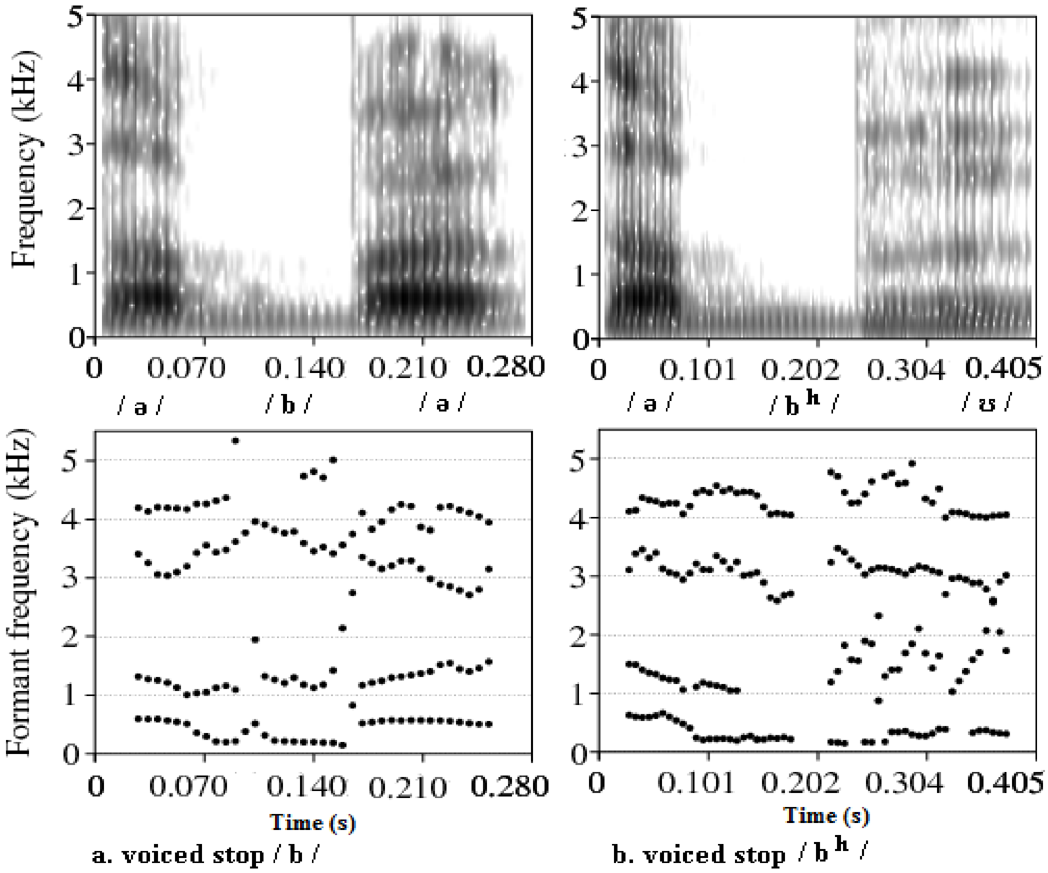


Figure 8.3 Spectrogram along with the formant tracks for the two voiced stops /b/ and /b^h/ of Sindhi.

8.2.2 Unvoiced bilabial stops

The typical waveform of the figure 8.4(a) shows the speech signals of the bilabial stop /p/ پ, of Sindhi extracted manually from the word utterance /sɪpə/ سِپ , and the figure 8.4(b) shows the speech signals of the bilabial stop **/p^h/** ڦ, of Sindhi extracted manually from the word utterance /ɓɑp^{h}ə/ ٻاڦ. By analyzing the waveforms of the figures 8.1 and 8.4 the most significant difference between the two waveforms is the presence of a periodic signal in the closure segment of the figure 8.1 and the absence of the periodic signal from the closure segment of the figure 8.4 (the signals between the points ***a*** and ***b*** shown in the waveforms); this is due to the voicing difference between the stops; the stops /**b^h**/, and /**b**/ are voiced and the stops **/p/,** and /**p^h**/ are unvoiced in Sindhi. This difference can be precisely seen in the extracted short duration waveforms of the closure and release segments of the two stops shown in figures 8.2 and 8.5. The other differences include the absence of the energy across the wideband spectrogram for the closure segments of the two stops **/p/** and /**p^h**/ compared to the closure segments of the stops /**b**/ and /**b^h**/, where energy is present in the lower frequency regions. The formant transitions for the stops **/p/** and /**p^h**/ are ambiguous and unstable; hence cannot be used for the identification of the two sounds. The presence of the weak energy in higher frequency regions (non-periodic vertical striations) for the stop /**p^h**/ shows the presence of the aspirated speech signals; therefore the stop /**p^h**/ is an unvoiced aspirated stop phoneme of Sindhi see the spectrogram of figure 8.6(b).

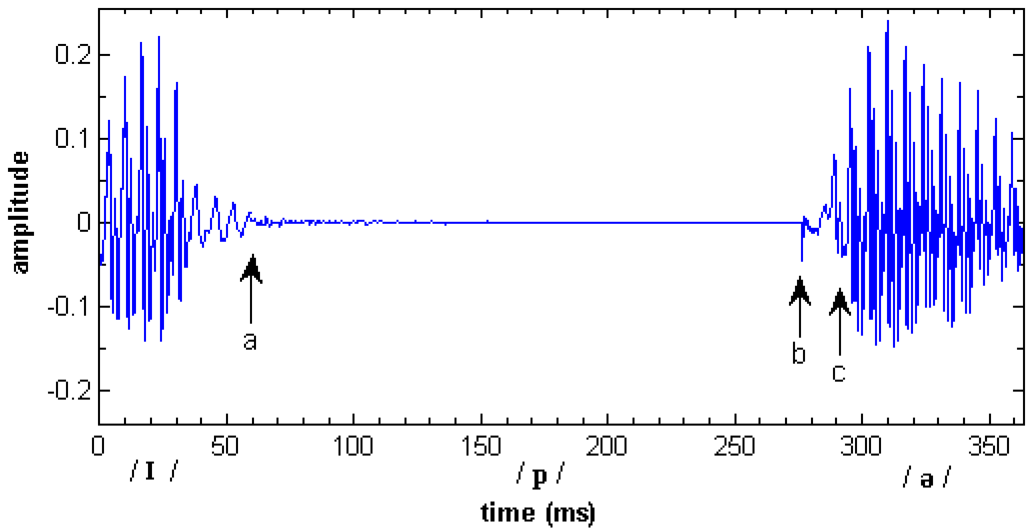


Figure 8.4(a) Wavform of the bilabial unvoiced stop /p/ of Sindhi, taken from the word utterance /sɪpə/ سِپ.

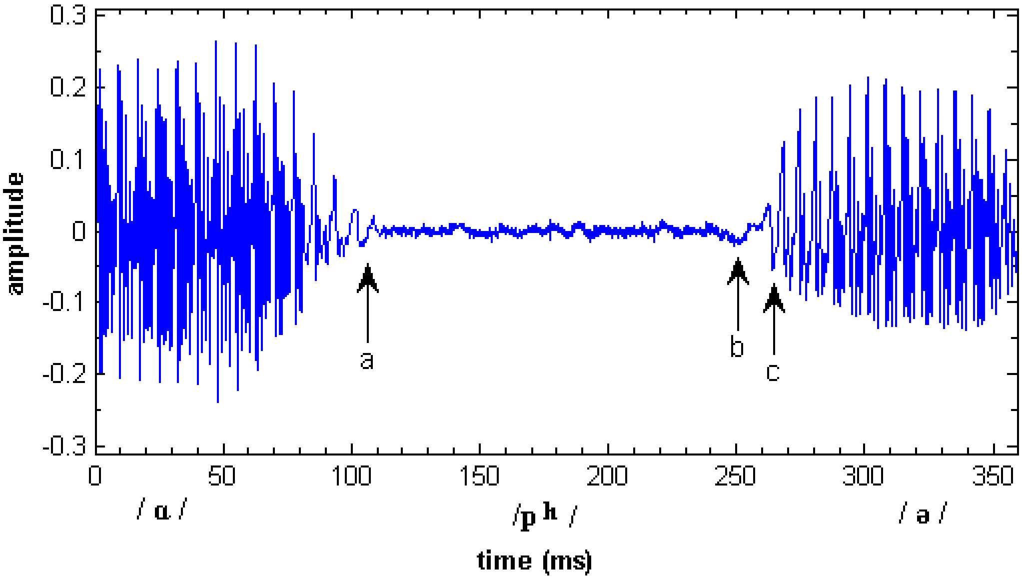


Figure 8.4(b) Waveform of the bilabial unvoiced stop /p^h/ of Sindhi, taken from the word utterance /ɓɑp^hə/ ٻڦ.

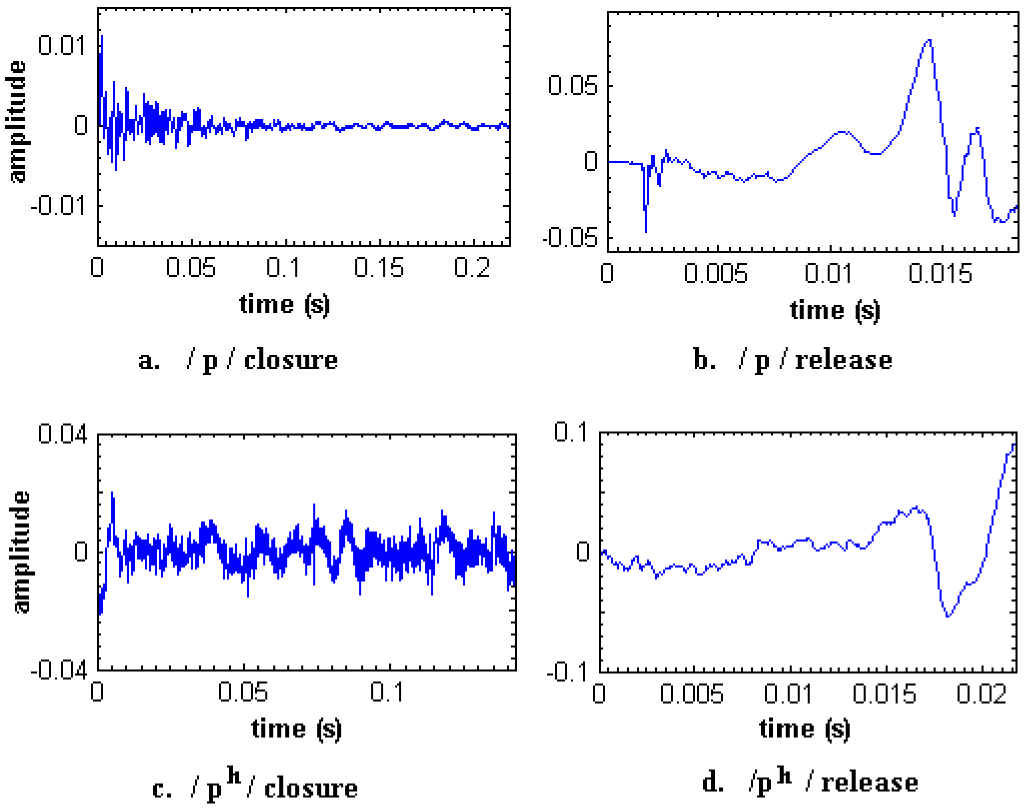


Figure 8.5 Extracted waveforms of the closure and release segments of the two unvoiced bilabial stops /p/ and /p^h/ of Sindhi.

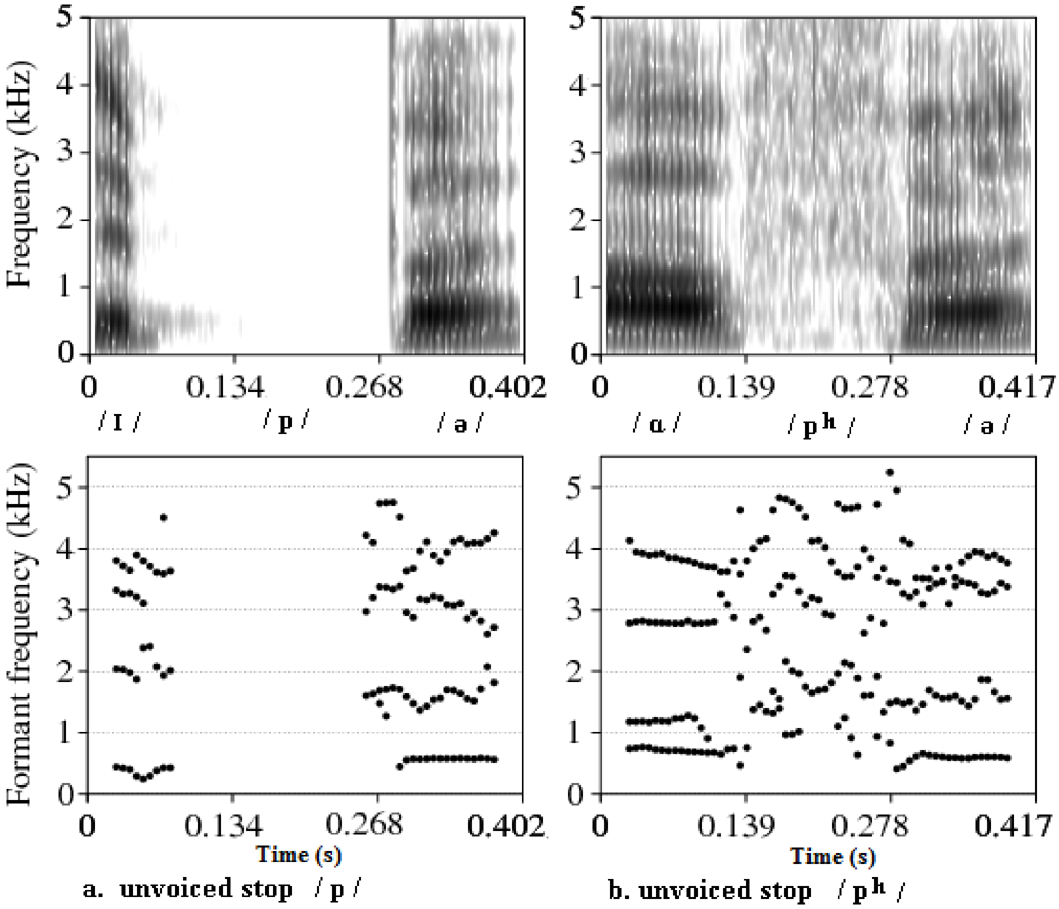


Figure 8.6 Spectrogram along with the formant tracks for the two unvoiced stops /p/ and /p^h/ of Sindhi.

8.2.3 Voiced dental Stops

The dental stop consonants are articulated with the tip of the tongue by placing it behind the upper or lower teeth (Ladefoged, 1993). In Sindhi dental stops are articulated by placing the tip of the tongue behind the upper teeth in such a way so that the mid of the tongue body touches the lower teeth. The articulator tongue when placed in such a way between upper and lower teeth it blocks the air flow in the vocal tract this builds up air pressure behind this closure point in the mouth when dental stops are produced (Ali, 2001; Stevens, 1993). Figures 8.7(a) and (b) below show the waveform of the two dental stops /**d**/ د, and /**d**^{h}/ ڌ, of Sindhi. The dental stop /**d**^{h}/ has a longer closure and release segments than the dental stop /**d**/, see figures 8.7 and 8.8. This is due to the fact that the dental /**d**^{h}/ is an aspirated plosive stop and the dental /**d**/ is a non-aspirated plosive stop of Sindhi. The common acoustic feature between two sounds is the presence of the periodic signals in the closure segment. This indicates that the two stops are voiced phonemes of Sindhi. Due to the slow varying periodic signal activity the energy is present in the lower frequency regions; however a very little energy is present in the higher frequency regions for the stop /**d**^{h}/ which is due to the presence of aspirated signals in this sound as shown in the spectrogram of the figures 8.9(a) and (b). The dental stops /**d**/, and /**d**^{h}/, have shown stable both F1 and F2 transitions, however the F1 for two sounds is very low. Due to the low F1 the first formant motion is downward (coming into the closure) and upward (going away from stop) for both the dental stops. The mean acoustic parameter values and the standard deviation are given in the Table 8.1 for these stops of Sindhi.

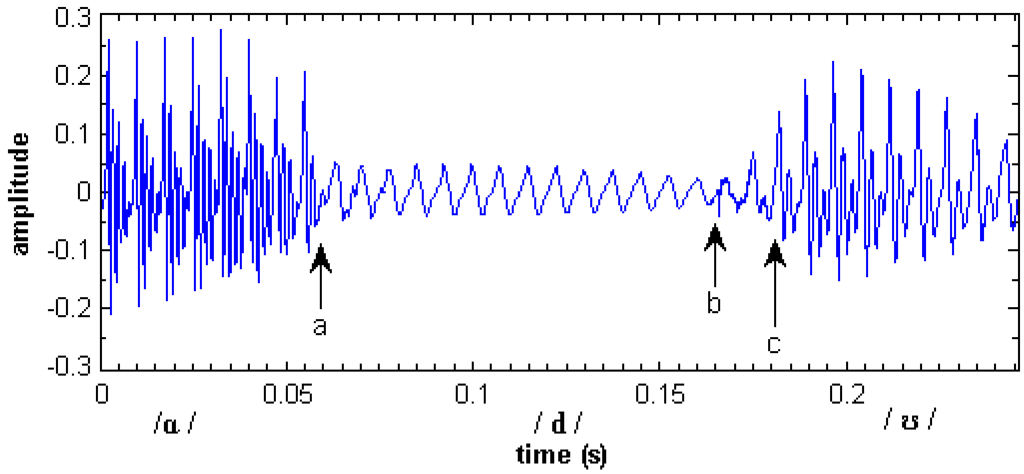


Figure 8.7(a) Waveform of the dental voiced stop /d/ of Sindhi, taken from the word utterance /sɑdʊgi/ ساڌگي.

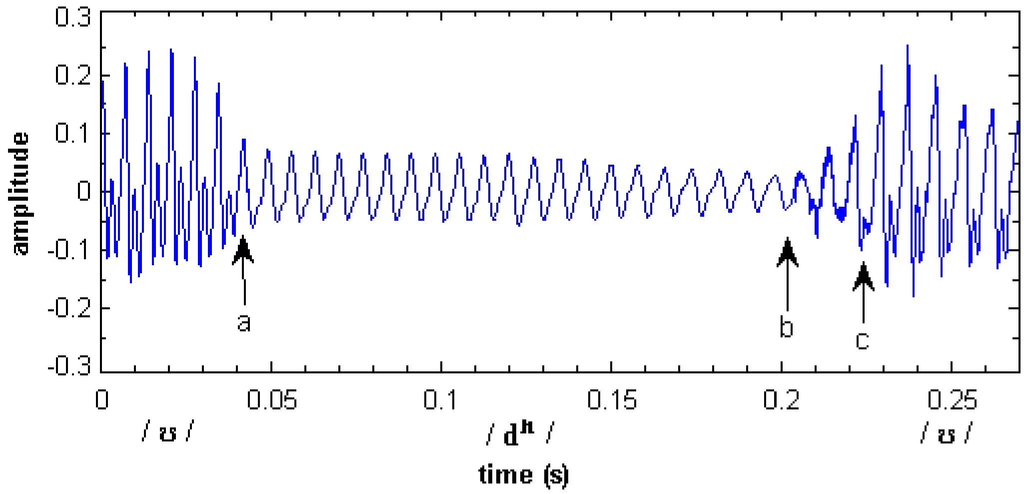


Figure 8.7(b) Waveform of the dental voiced stop /d^h/ of Sindhi, taken from the word utterance /sʊdhʊ/ سڌُ.

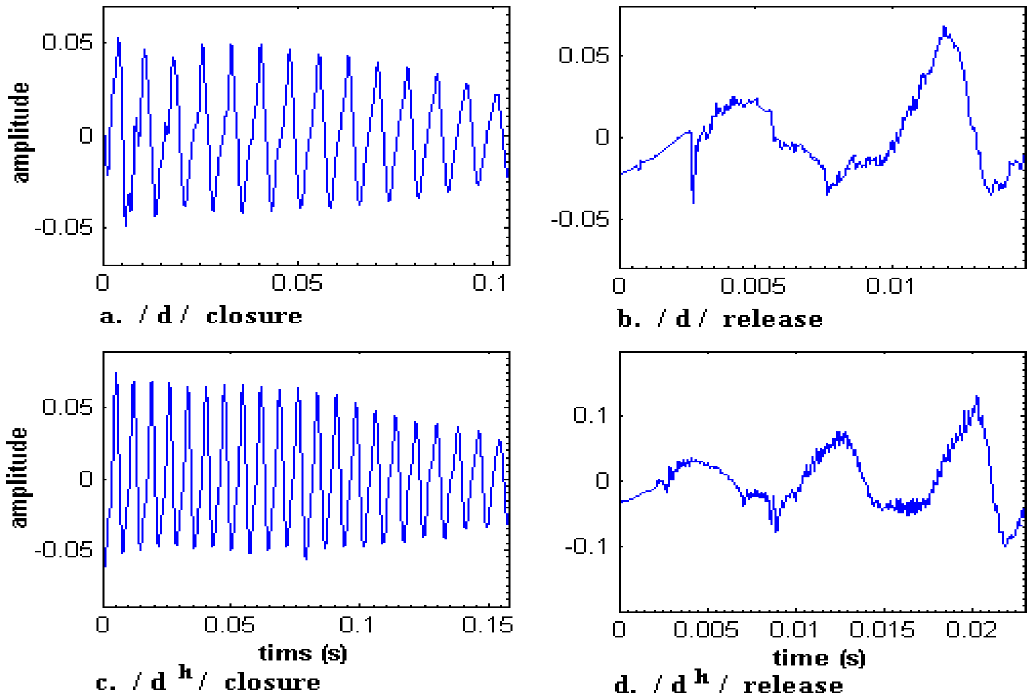


Figure 8.8 Extracted waveforms of the Closure and release segments of the two dental voiced stops /d/ and /dʰ/of Sindhi.

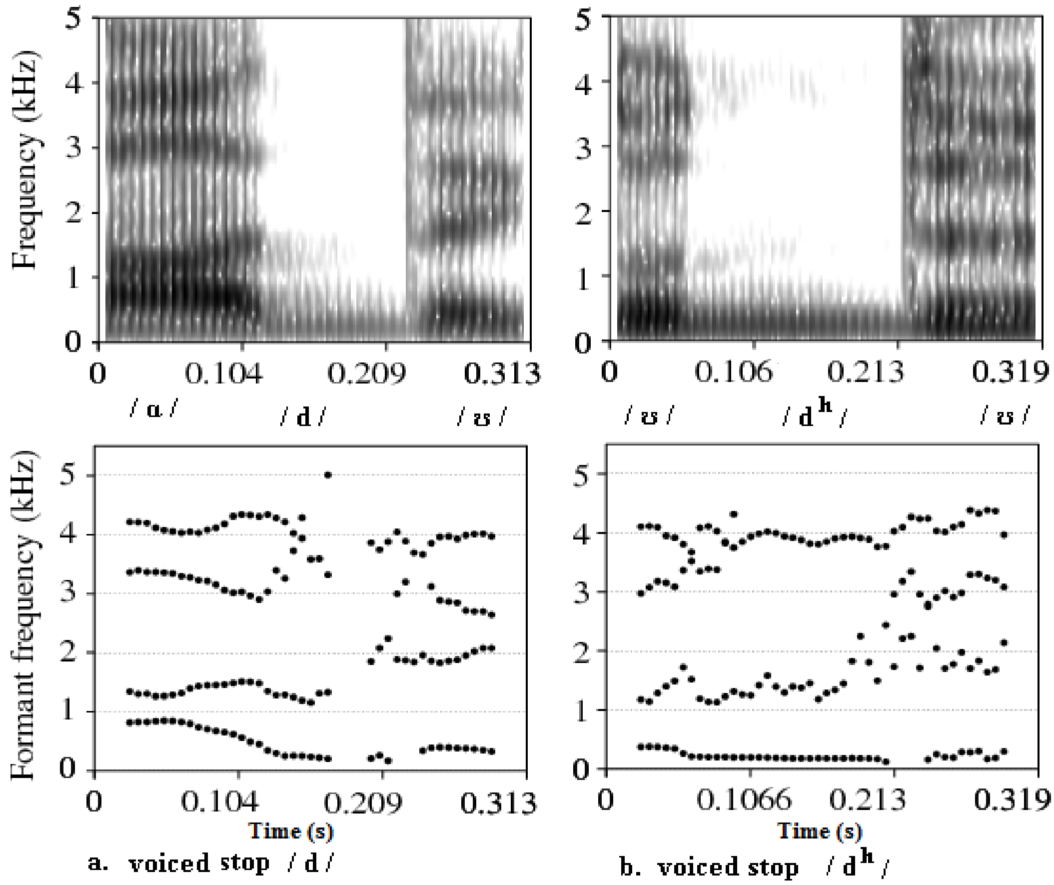


Figure 8.9 Spectrogram along with the formant tracks for the two voiced stops /d/ and /dʰ/ of Sindhi.

8.2.4 Unvoiced dental stops

Figures 8.10(a) and (b) below show the waveforms of the two dental stops **/t/** and **/t^h**/ of Sindhi. There is no periodic signal activity in the closure segments of the two sounds; this indicates that the two sounds produced are unvoiced phonemes of Sindhi see figures 8.10 and 8.11. The closure segment of the two unvoiced dental stops is longer than the closure segment of the voiced dentals and from both the voiced and unvoiced bilabial stops of Sindhi. There is no presence of energy either in lower or higher frequency regions for the closure segment across the wide band spectrogram of the two sounds; this is the common acoustic feature for all the unvoiced stops of Sindhi. The two unvoiced stops **/t/** and **/t^h**/ have shown ambiguous and unstable formant transitions and hence cannot be used for the identification of these sounds see figure 8.12(a) and (b). The mean acoustic parameter values and the standard deviation are given in Table 8.1 for these two stops of Sindhi.

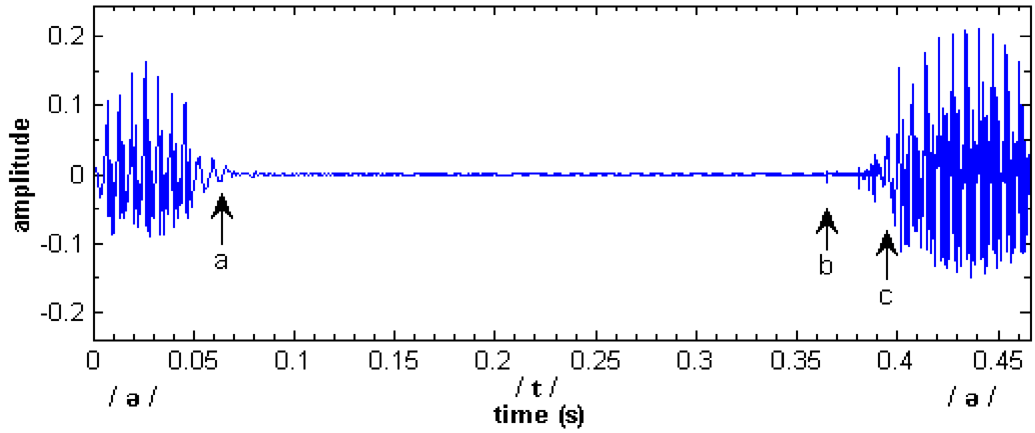


Figure 8.10(a) Waveform of the unvoiced dental stop /t/ of Sindhi, taken from the word utterance /sətə/ سَتَ.

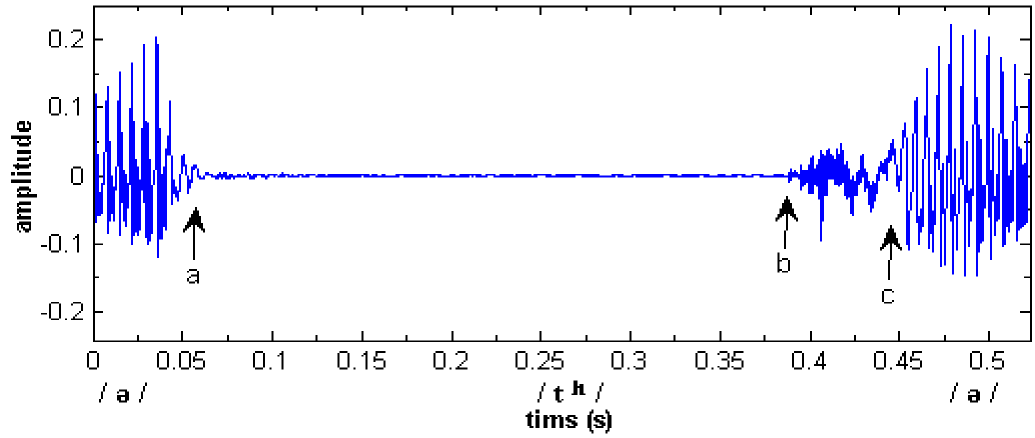


Figure 8.10(b) Waveform of the unvoiced dental stop /t^h/ of Sindhi, taken from the word utterance /hət^{h}ə/ هَٿ.

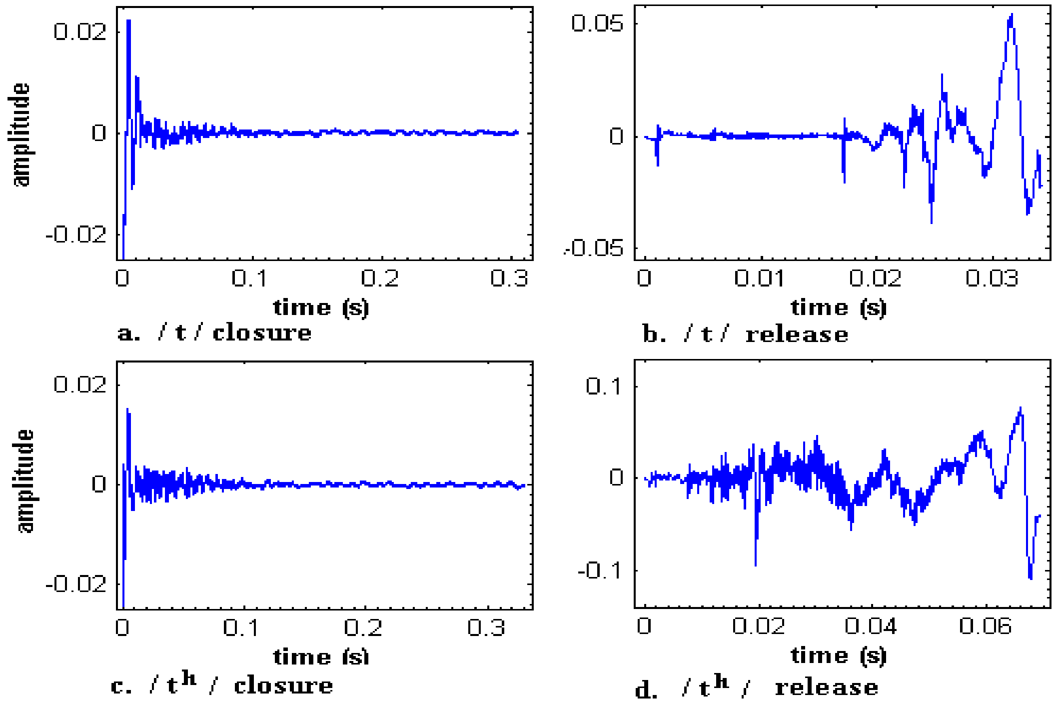


Figure 8.11 Extracted waveforms of the two unvoiced dental stops /t/ and /tʰ/ of Sindhi.

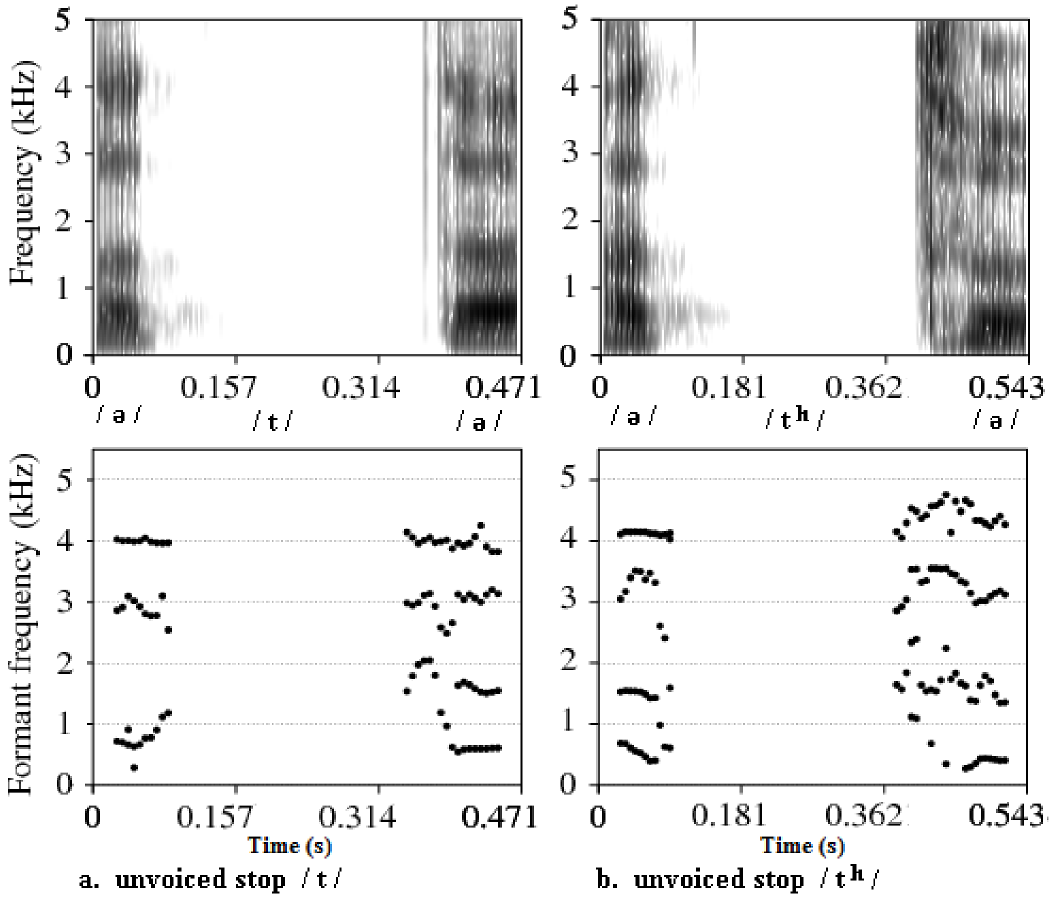


Figure 8.12 Spectrogram along with the formant tracks of the two unvoiced stops /t/ and /tʰ/ of Sindhi.

8.2.5 Voiced retroflex stops

These sounds are produced with the tip of the tongue curled up in the direction of the hard palate and builds up the air pressure behind this closure point in the oral cavity. Figures 8.13(a) and (b) show the waveform of the two retroflex stops /ɖ/ and /ɖʰ/ of Sindhi. The periodic signal activity and the presence of the energy in the lower frequency range indicate that the two sounds are voiced phonemes of Sindhi. The spectrogram for the stop /ɖʰ/ shows little energy present in the higher frequency regions this is because of the presence of the aspirated signals in the release segment of the stop /ɖʰ/ shown in the spectrogram of the figure 8.15. The F1 transitions are stable for both the stops /ɖ/ and /ɖʰ/; whereas the F2 transitions are ambiguous and unstable. The two stops have low F1; therefore the F1 motion for the two stops is downward (coming into the closure segment) and upward (going away from the stop). The mean acoustic parameter values and the standard deviation for these two voiced retroflex stops are given in Table 8.1.

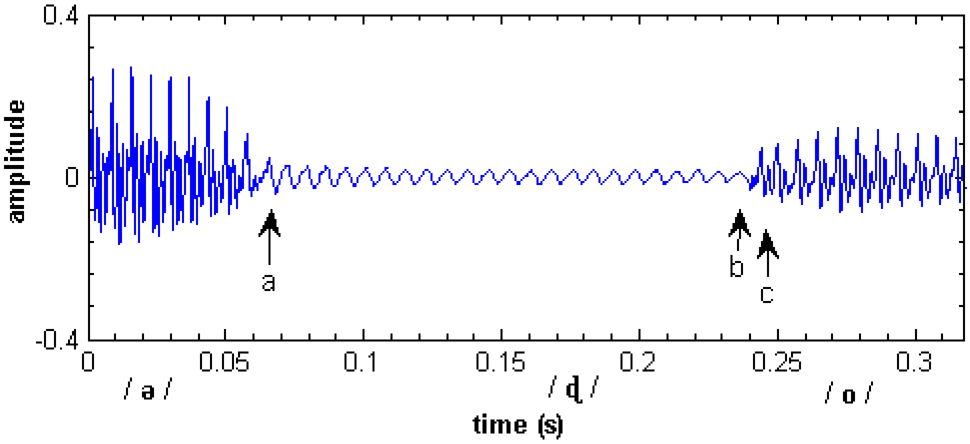


Figure 8.13(a) Waveform of the voiced stop /ɖ/ of Sindhi, taken from the word utterance /bəɖo/ ٻڏو .

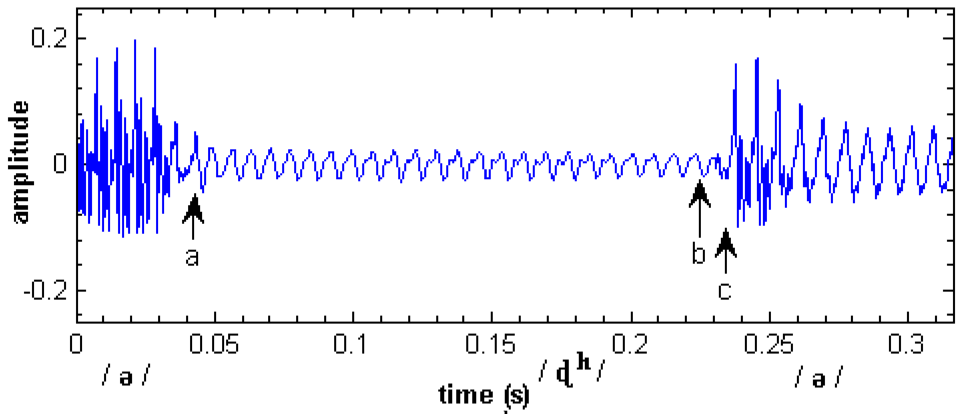


Figure 8.13(b) Waveform of the voiced stop /ɖʰ/ of Sindhi, taken from the word utterance /kəɖʰə/ ڪڍ .

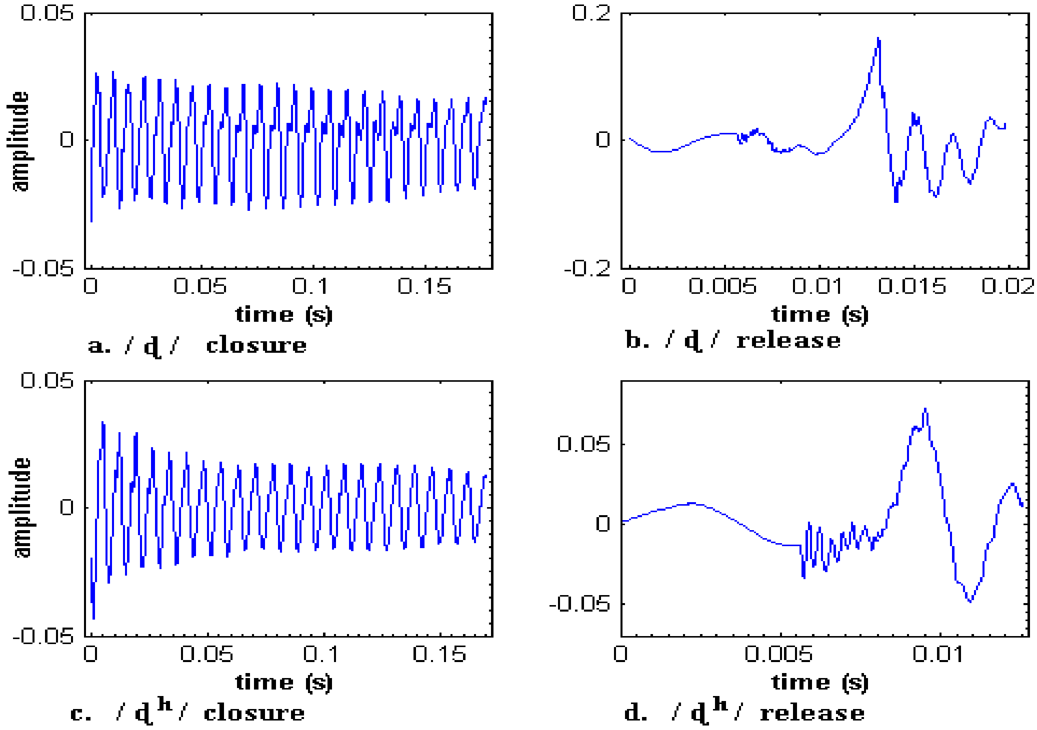


Figure 8.14 Closure and release segments of the voiced retroflex stops /ɖ/, and /ɖʰ/ of Sindhi extracted manually from the word utterance.

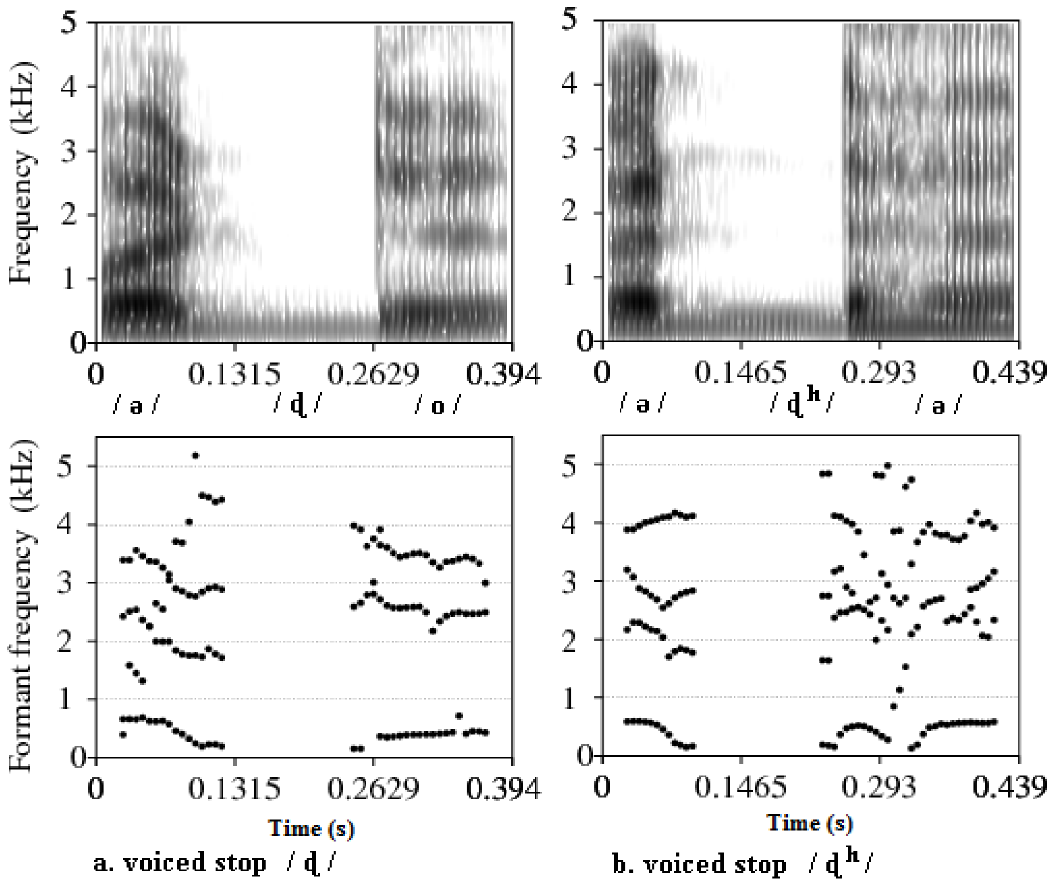


Figure 8.15 Spectrogram along with the formant tracks for the two voiced stops /ɖ/, and /ɖʰ/ of Sindhi.

8.2.6 Unvoiced retroflex stops

Figures 8.16(a) and (b) show the waveforms of the two retroflex stops /ʈ/ and /$ʈ^h$/ of Sindhi. There is no periodic signal activity during the closure segment of the two stops and also the energy is absent from the lower frequency regions shown in the wideband spectrogram of the figure 8.18. This indicates that the two stops /ʈ/ and /$ʈ^h$/ are unvoiced retroflex phonemes of Sindhi. A transient burst of noisy signals in the release segment of the stop /$ʈ^h$/ is the sign of the aspirated signals in the speech sound shown in the waveform of figure 8.17(d) and the spectrogram of figure 8.18(b). The formant transitions for the two stops are ambiguous and unstable for the closure segments; hence cannot be used for the identification of these stops. The mean acoustic parameter values and the standard deviation for the retroflex stops are given in Table 8.1.

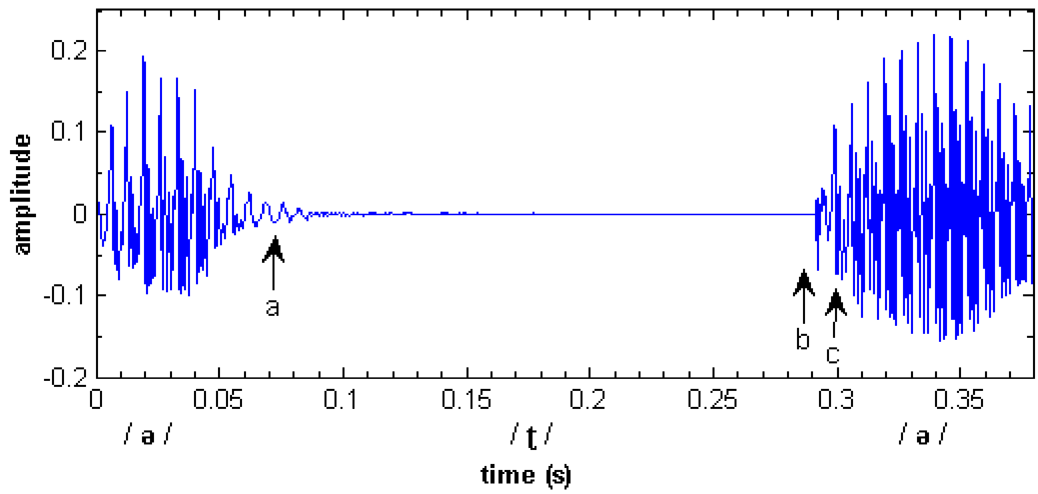


Figure 8.16(a) Waveform of the retroflex unvoiced stop /ʈ/ of Sindhi, taken from the word utterance /səʈə/ سَٽَ .

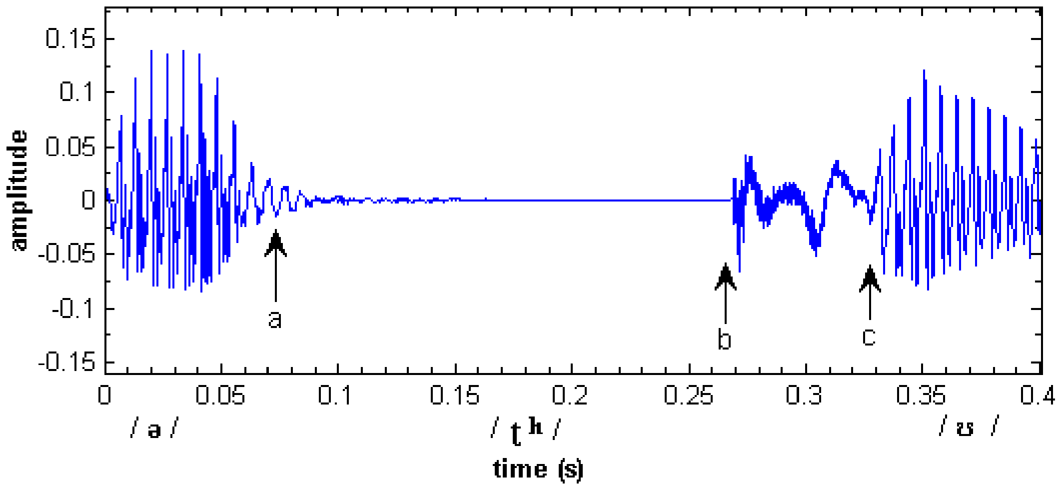


Figure 8.16(b) Waveform of the retroflex unvoiced stop /$ʈ^h$/ of Sindhi, taken from the word utterance /$səʈ^hʊ$/ سَٿُ .

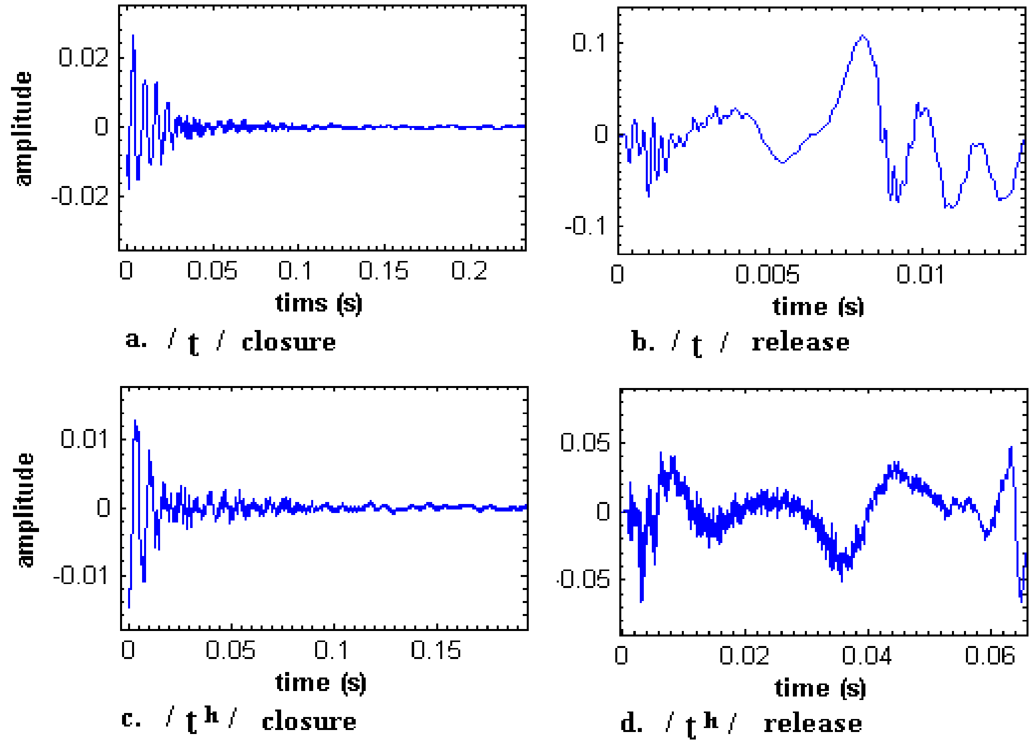


Figure 8.17 Extracted waveforms of the closure and release segments of the two unvoiced retroflex stops /ʈ/ and /ʈ^h/ of Sindhi.

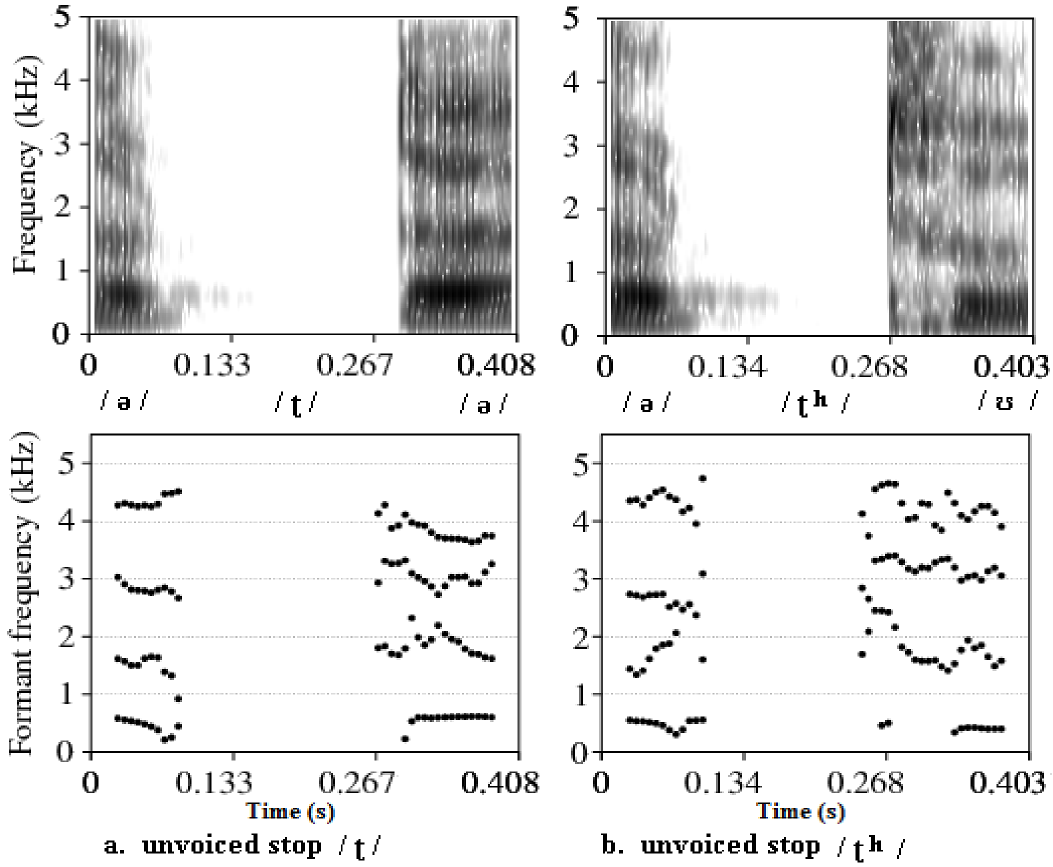


Figure 8.18 Spectrogram along with the formant tracks for the two unvoiced stops /ʈ/ and /ʈ^h/ of Sindhi.

8.2.7 Voiced palato-alveolar stops

Palato-alveolar stops are produced by raising the tongue body in the direction of the hard palate in a way so that it touches the back of the alveolar ridge and blocks the air passage through the oral cavity. This builds up air pressure behind this closure point in the vocal tract and when the air pressure is suddenly released it produces the plosive stops /dʒ/ ج and /ʤ^{h}/ جھ , of Sindhi. Figures 8.19(a) and (b) below show the waveforms of these two palato-alveolar stops of Sindhi. The slow varying periodic signal activity shown in the waveform and the presence of the energy in the lower frequency regions for the closure segment shown in the wideband spectrogram of the figure 8.21 indicate that these are voiced phonemes of Sindhi. The analysis results show that the closure and release segments of the stop /ʤ^{h}/ are longer in duration along with an audible burst of noise signals in the two segments of this sound compared to the stop /ʤ/ shown in the waveform figure 8.20(d) and the spectrogram of the figure 8.21(b). The presence of the noise signals in the release segment of this sound potentially leads to a confusion that such signals are normally present in the kind of hushing or hissing sounds such as fricatives. However this burst of noise signals is transient; whereas the duration of hushing or hissing sound in fricatives is longer than the duration of the noise signals present in the release segment of these two palato-alveolar stops of Sindhi. It is not clear yet that these two phonemes are members of the stop class or they should be associated to the affricate class of consonants. So far we classify these sounds as aspirated voiced stops until the acoustic analysis of the fricatives of Sindhi; because the affricate consonants are the sounds that posses the characteristics of both the stops and fricatives. The role of formants for the identification of the two palato-alveolar stops /ʤ^{h}/, and /ʤ/ of Sindhi is significant, because the two stops have shown sharp and stable F1 transitions shown in the spectrogram of the figure 8.21. The two stops have very low F1 so that the F1 motion can be defined as downward (coming into closure) and upward (going away from stop) for these two voiced stops of Sindhi. The mean acoustic parameter values and the standard deviation are given in Table 8.1 for these two palato-alveolar stop consonants of Sindhi.

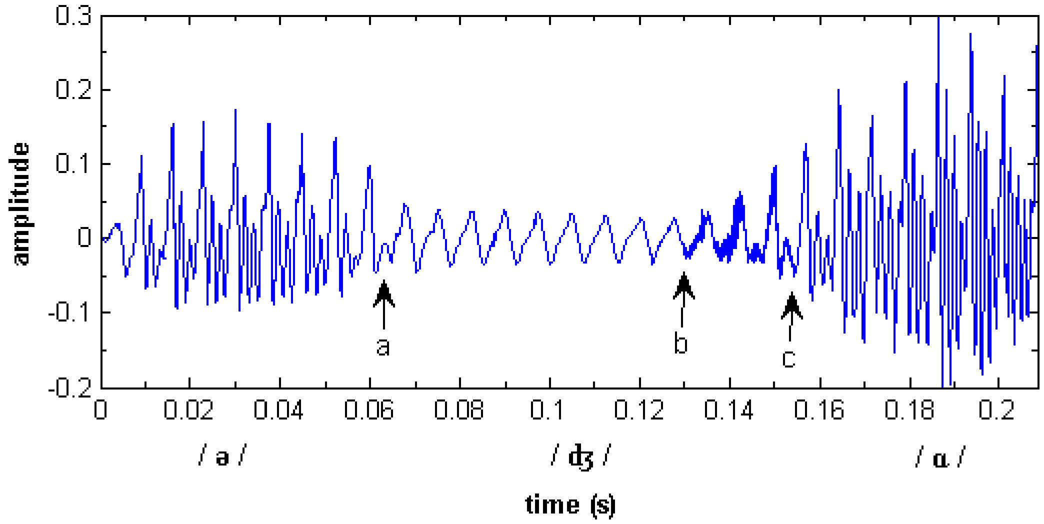


Figure 8.19(a) Waveform of the voiced palato-alveolar stop /dʒ/ ج, of Sindhi, taken from the word utterance /sədʒɑi/ سَجائي .

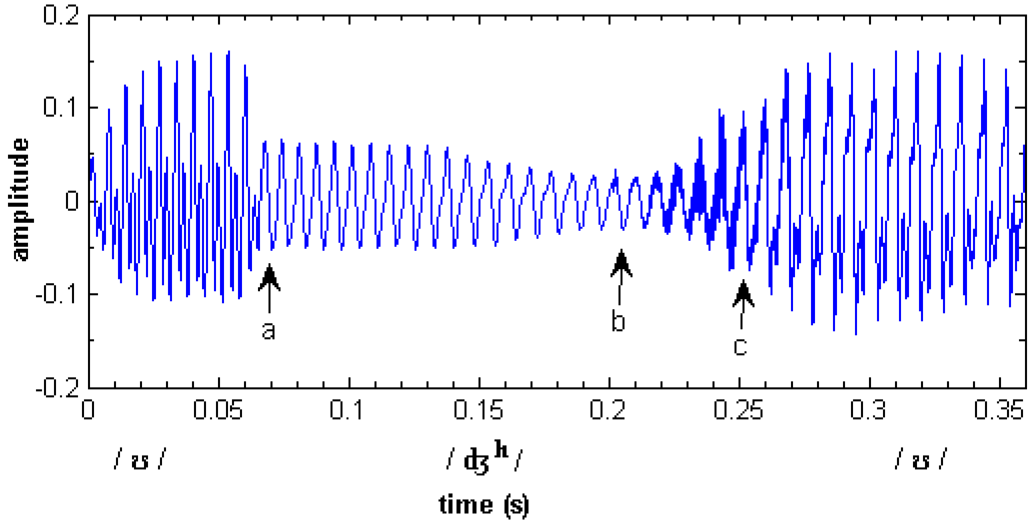


Figure 8.19(b) Waveform of the aspirated voiced palato-alveolar stop /dʒʰ/ جھ, of Sindhi, taken from the word utterance /kʊdʒʰʊ/ ڪُجھ .

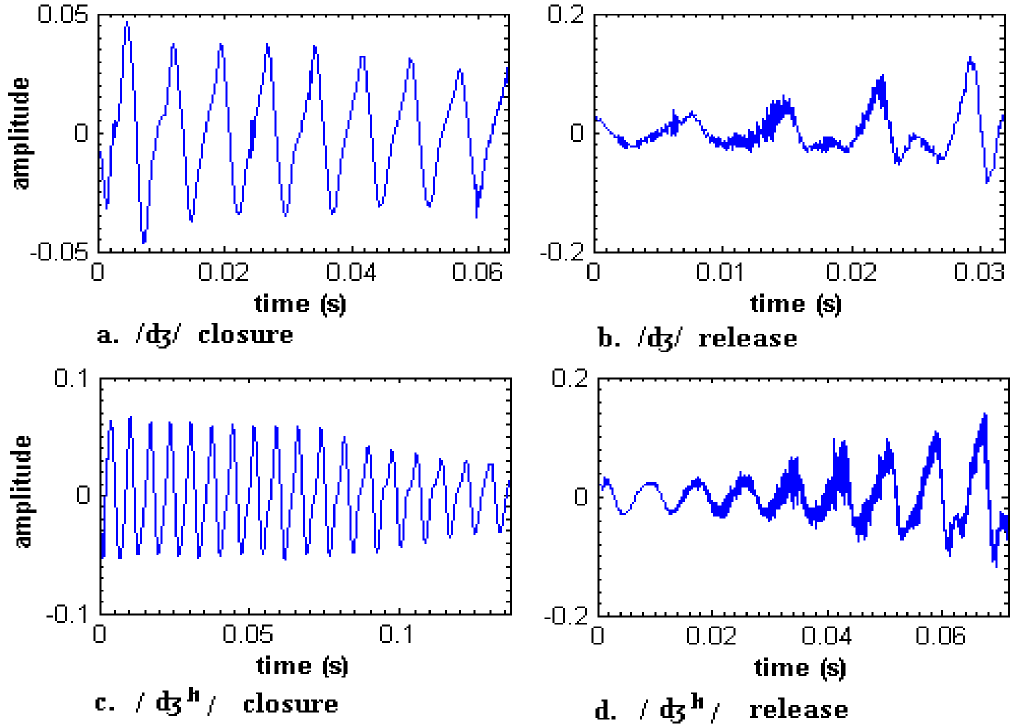


Figure 8.20 Extracted waveforms of the closure and release segments of the two voiced palato-alveolar stops /dʒ/ and /dʒ^h/ of Sindhi.

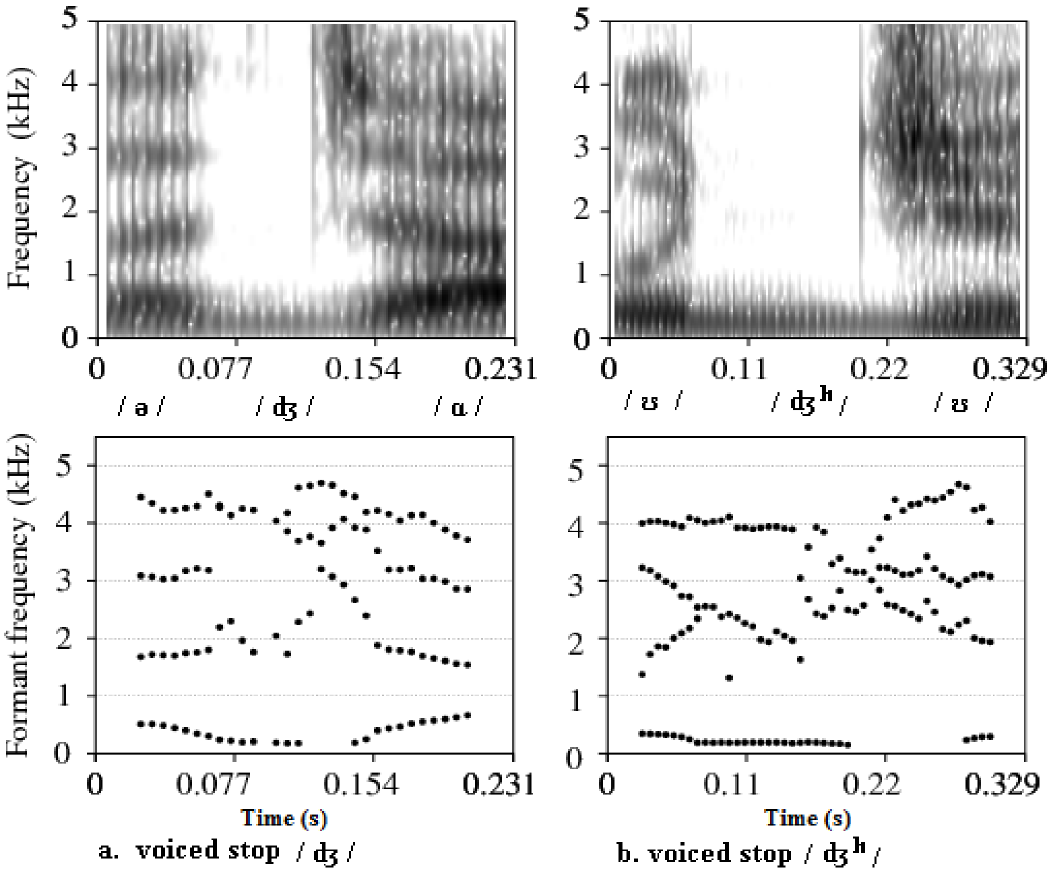


Figure 8.21 Spectrogram along with the formant tracks for the two voiced palato-alveolar stops /dʒ/ and /dʒ^h/ of Sindhi.

8.2.8 Unvoiced palato-alveolar stops

Figures 8.22 (a) and (b) show the waveforms of the two palato-alveolar stops /ʧ/ ڄ, and /ʧʰ/ ڇ, of Sindhi. There is no periodic signal activity in the closure segments of the two stops and the absence of the energy in the lower frequency regions confirm that the two sounds are unvoiced phonemes of Sindhi as shown in the short duration waveforms of the figure 8.23 and the spectrogram of the figure 8.24. The release segment of the two stops /ʧ/ and /ʧʰ/ starts with an audible burst of noise signals; this indicates that the two sounds are either the aspirated sounds or the fricative sounds. The burst of noise signals for the two sounds is relatively strong and longer in duration compared to the aspirated signals present in the other aspirated stops of Sindhi. The release segment of the two stops start with the mark of the dark vertical striation followed by the burst of the energy present in the high frequency region this characteristic of these two unvoiced stops indicate that the two sounds are either affricates (sounds that start as stop and end as fricative) or the aspirated plosive stops. However we classify these sounds so far as aspirated unvoiced stops until the acoustic analysis of the fricative sounds. The formant transitions for the two stops /ʧ/ and /ʧʰ/ are ambiguous and unstable and hence can not be used to identify these sounds see the spectrogram of the figure 8.24 along with the formant tracks.

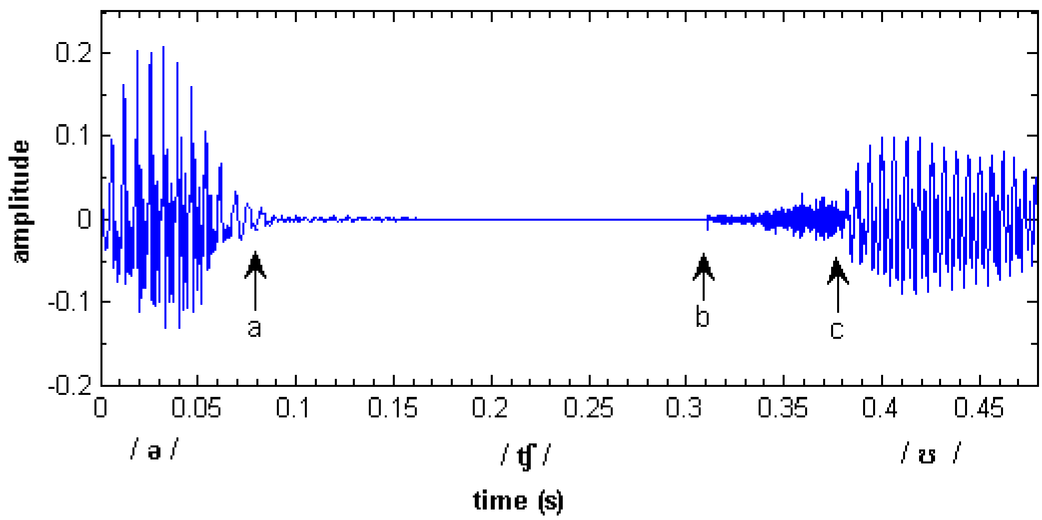


Figure 8.22(a) Waveform of the palato-alveolar unvoiced stop /ʧ/ of Sindhi, taken from the word utterance /səʧʊ/ سَچُ .

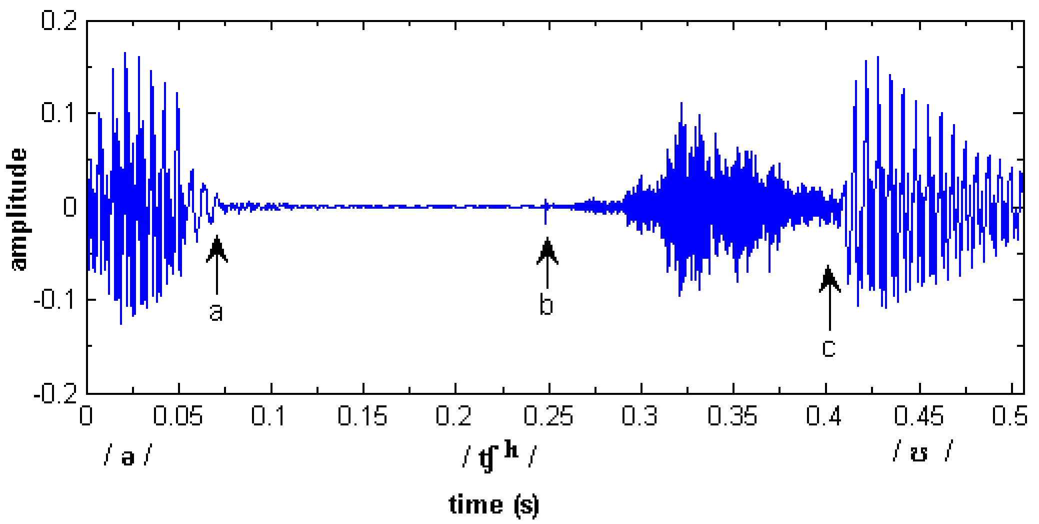


Figure 8.22(b) Waveform of the palato-alveolar unvoiced stop /ʧ^{h}/ of Sindhi, taken from the word utterance /kəʧ^{h}ʊ/ ڪڇُ .

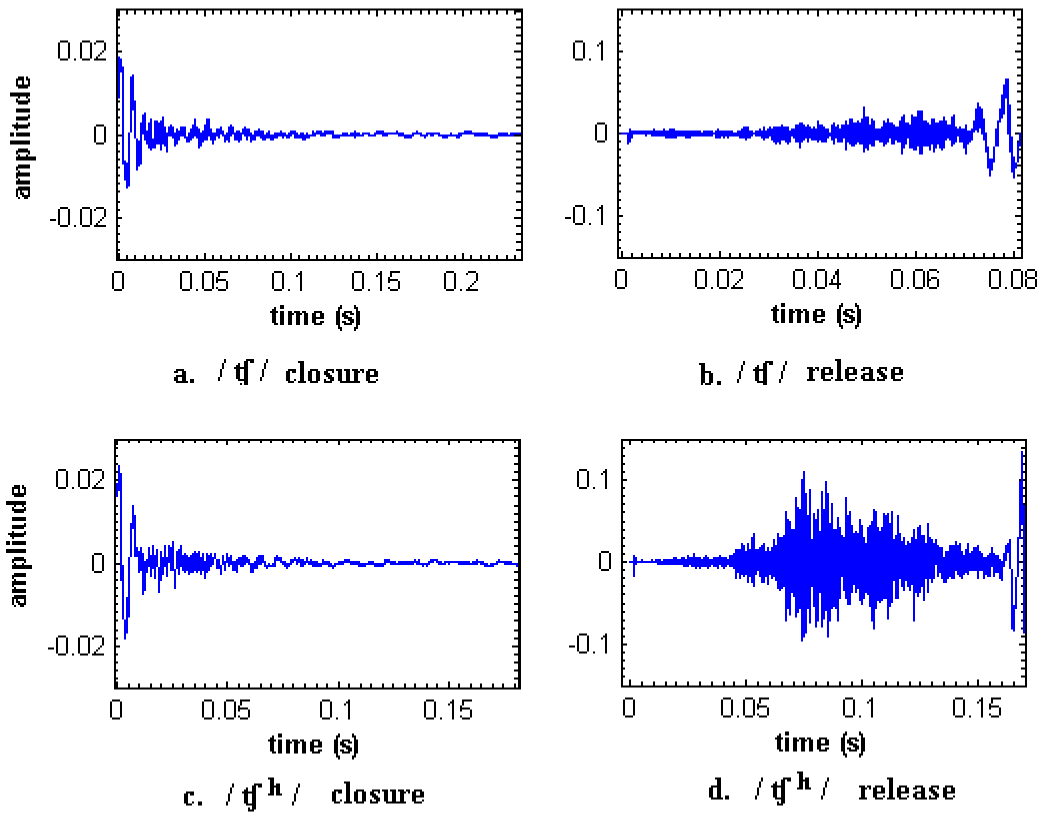


Figure 8.23 Extracted waveforms of the closure and release segments of the two unvoiced palato-alveolar stops /ʧ/ and /ʧ^{h}/ of Sindhi.

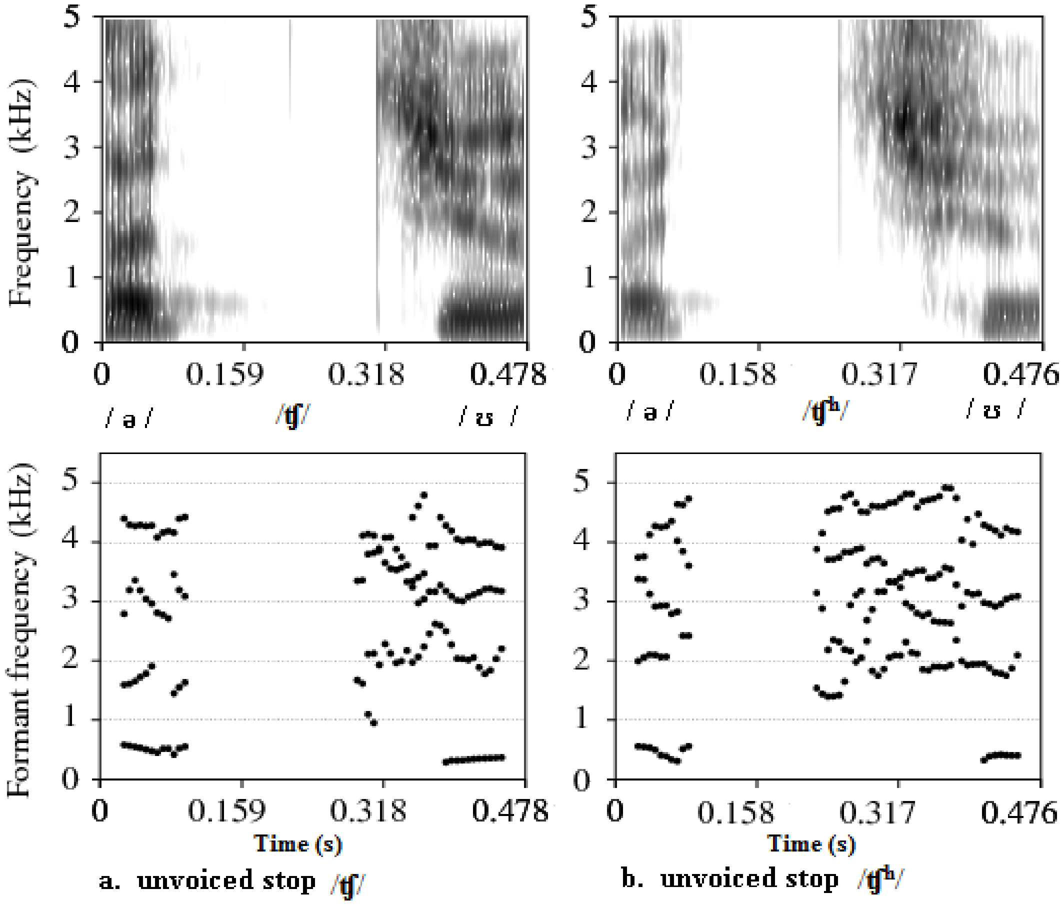


Figure 8.24 Spectrogram along with the formant tracks for the two unvoiced palato-alveolar stops /ʧ/ and /ʧ^h/ of Sindhi.

8.2.9 Voiced velar stops

The velar stops of Sindhi are produced by raising the back of the tongue in the direction of the soft palate. Figures 8.25(a) and (b) show the waveforms of the two velar stops /g/ and /g^h/ of Sindhi. The presence of the slow varying periodic signal activity and the energy in the lower frequency range indicate that the two stops are voiced phonemes of Sindhi. This is shown in the waveform figure 8.26 and the spectrogram figure 8.27. The formant transitions for both F1 and F2 are stable for the velar stop /g/, whereas the formant transitions for the velar stop /g^h/ are only stable for F1 shown in figure 8.27, the formant tracks. The two sounds have shown low F1; therefore the formant motion can be defined as downward (coming into the closure) and upward (going away from it). The presence of the aspirated signals in the release segment of the velar stop /g^h/ confirm this sound as an aspirated voiced velar stop of Sindhi see figures 8.26(d) and spectrogram of the figure 8.27(b).

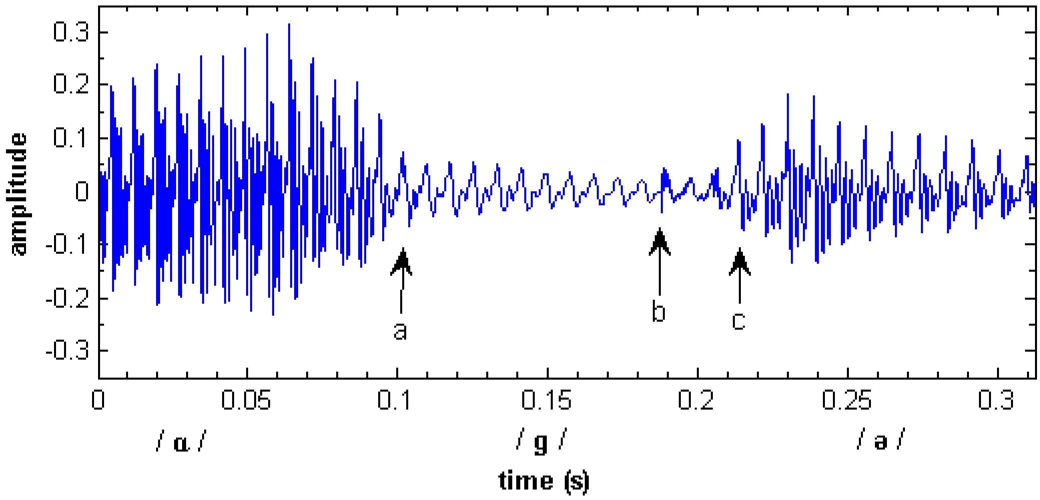


Figure 8.25(a) Waveform of the voiced velar stop /g/ of Sindhi, taken from the word utterance /ɑgə/ آگَ .

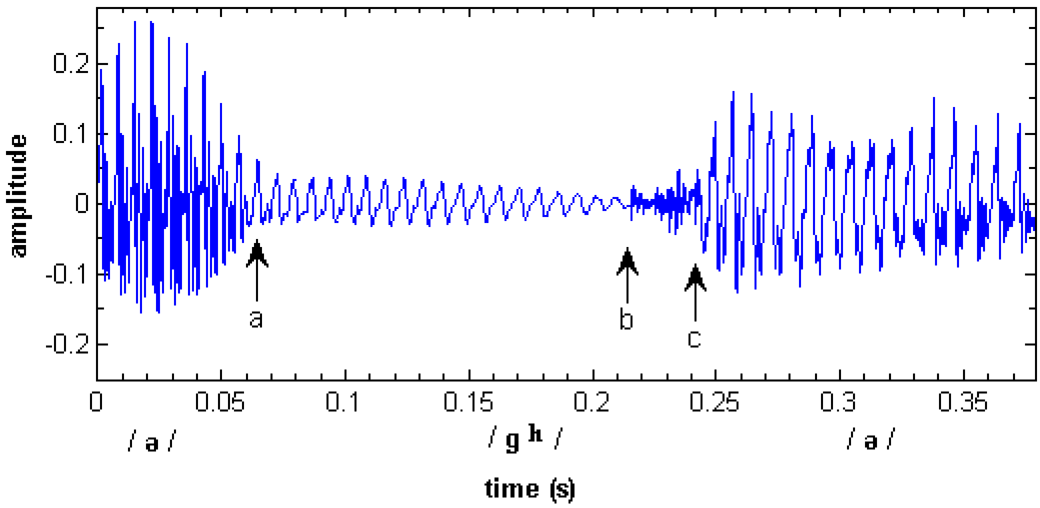


Figure 8.25(b) Waveform of the voiced velar stop /g^h/ of Sindhi, taken from the word utterance /səg^hə/ سَگھَ .

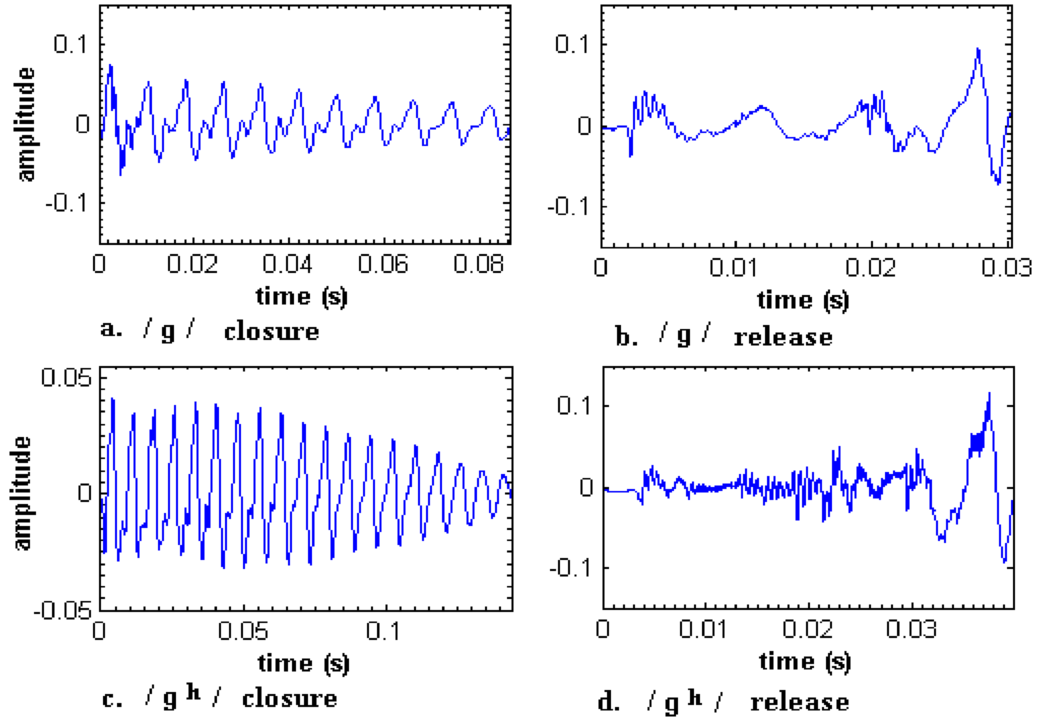


Figure 8.26 Extracted waveforms of the closure and release segments of the two voiced velar stops /g/ and /g^h/ of Sindhi.

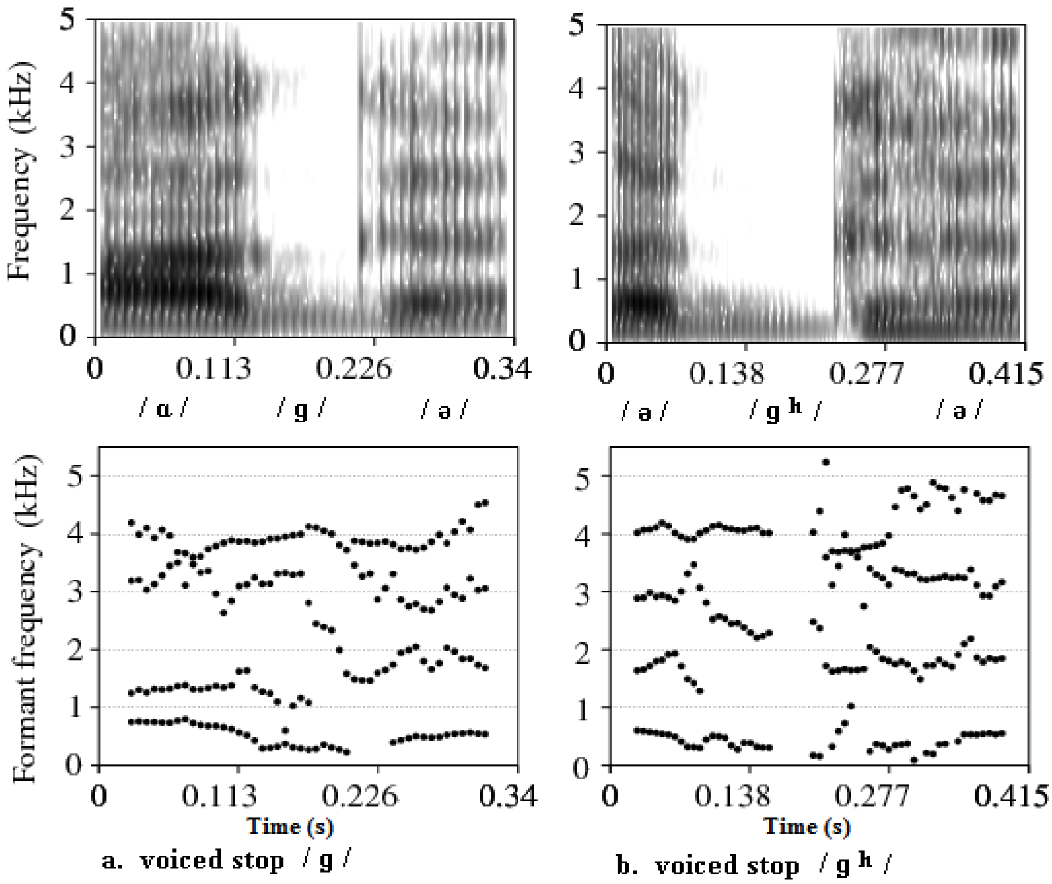


Figure 8.27 Spectrogram along with the formant tracks of the two voiced velar stops /g/ and /g^h/ of Sindhi.

8.2.10 Unvoiced velar stops

Figures 8.28 (a) and (b) below show the waveforms of the two velar stops **/k/** ڪ, and **/kʰ/** ک of Sindhi. There is no periodic signal activity in the closure segment of the two stops and the energy is absent from the lower frequency regions; this indicates that the two stops are unvoiced phonemes of Sindhi. The release segment of the two stops shows the dark vertical line across the wideband spectrogram (a separation mark between the two segments closure and release). The release segment of the stop **/kʰ/** starts with an audible burst of noise signals for a very small duration this is the sign of aspirated speech signals in this sound; therefore the stop **/kʰ/** is an aspirated unvoiced velar stop of Sindhi, see figure 8.29. The formant transitions for the two stops are ambiguous and are not stable for both the stops **/k/**, and **/kʰ/** see the spectrogram of the figure 8.30 along with the formant tracks.

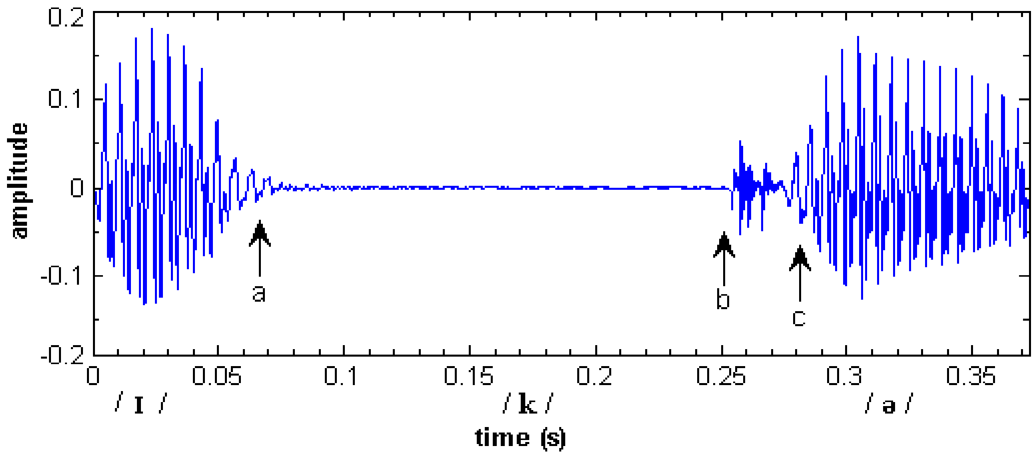


Figure 8.28(a) Waveform of the unvoiced velar stop /k/ of Sindhi, taken from the word utterance /sɪkə/ سِڪَ .

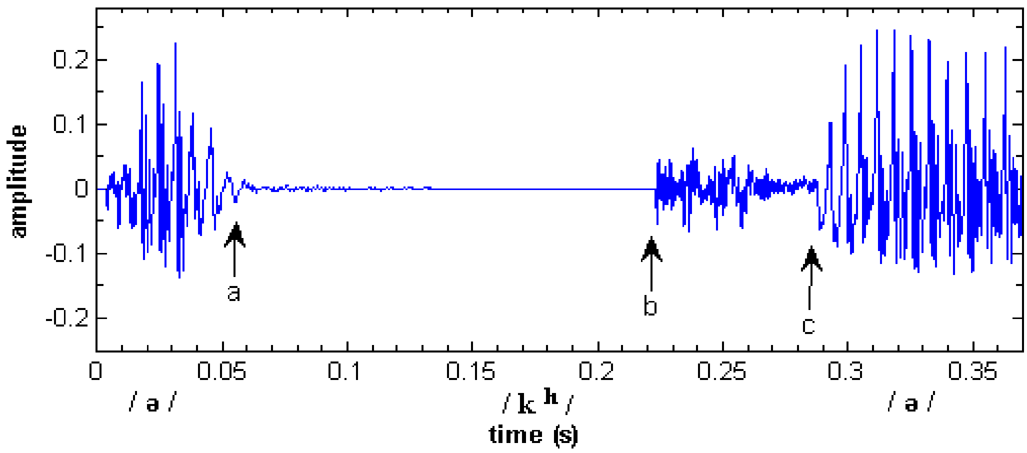


Figure 8.28(b) Waveform of the unvoiced velar stop /kʰ/ of Sindhi, taken from the word utterance /əkʰər/ اکَرُ .

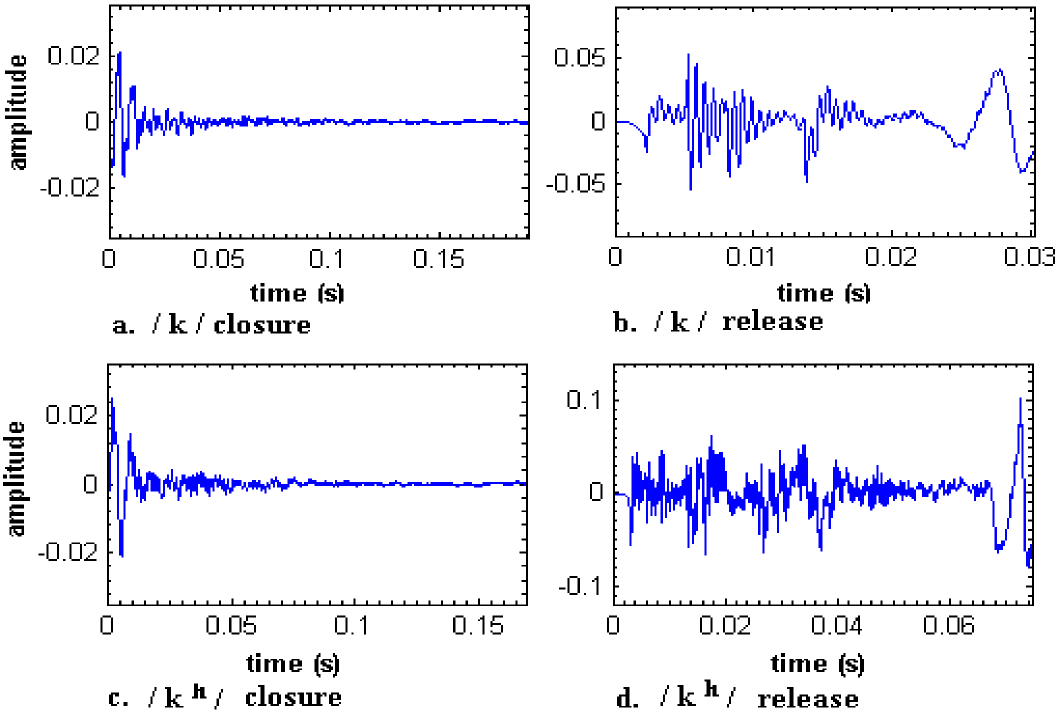


Figure 8.29 Extracted waveforms of the closure and release segments of the two unvoiced velar stops /k/ and /kʰ/ of Sindhi.

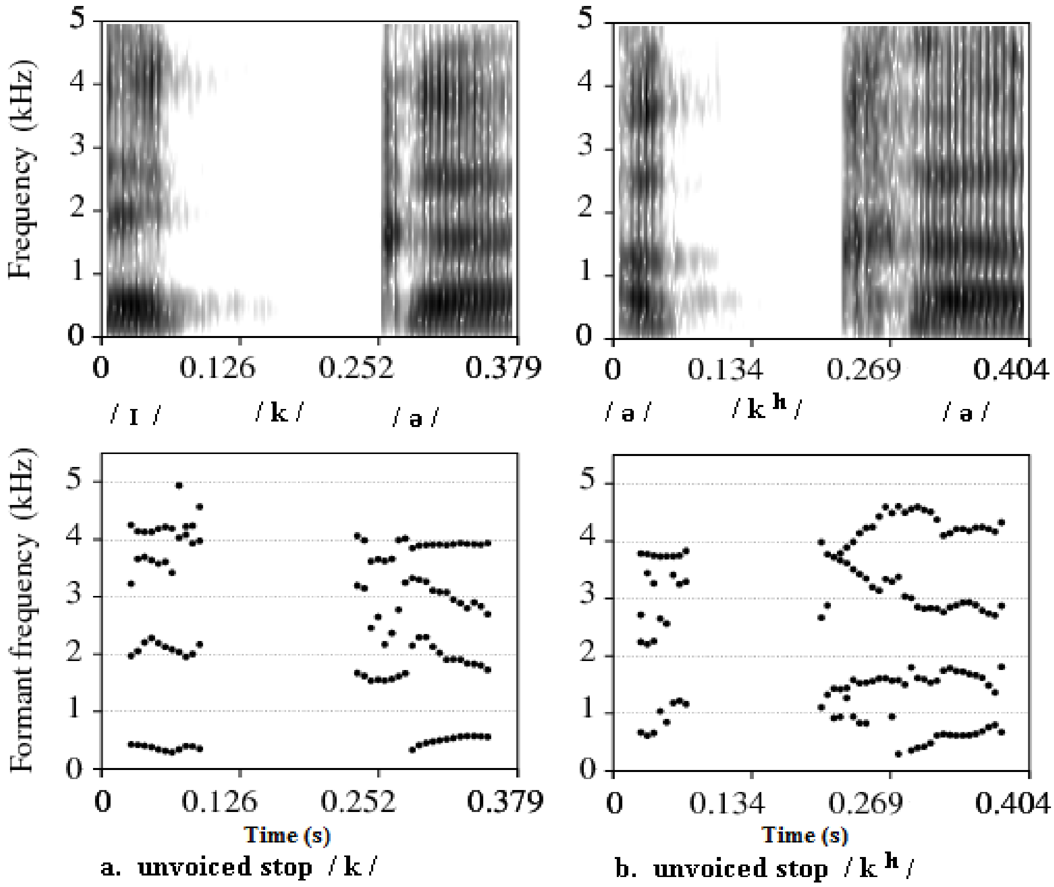


Figure 8.30 Spectrogram along with the formant tracks of the two unvoiced velar stops /k/ and /kʰ/ of Sindhi.

Table 8.1 Mean and standard deviation of the acoustic parameters: F0, first four formants, duration of the stop consonants of Sindhi.

Stop	Segment	parameter	Duration	F0	F1	F2	F3	F4	Intensity (dB)
/b/ب	Closure	Mean	0.127	138	246	1073	2876	3891	60
		STD	0.0179	3	76	131	367	176	3
	Release	Mean	0.0145	147	-	-	-	-	66
		STD	0.0015	1.4	-	-	-	-	4
/p/ پ	Closure	Mean	0.2323	-	-	-	-	-	28
		STD	0.0112	-	-	-	-	-	7
	Release	Mean	0.018	8183	-	-	-	-	63
		STD	0.002	154	-	-	-	-	5
/b^h/ڀ	Closure	Mean	0.151	127	208	1072	2949	4071	61
		STD	0.0166	5	33	304	183	138	2
	Release	Mean	0.048	122	-	-	-	-	65
		STD	0.0112	3	-	-	-	-	2
/p^h/ڦ	Closure	Mean	0.1298	-	-	-	-	-	47
		STD	0.0108	-	-	-	-	-	1
	Release	Mean	0.0223	13460	-	-	-	-	58
		STD	0.0049	211	-	-	-	-	7
/d/ د	Closure	Mean	0.095	133	300	1446	3073	4156	61
		STD	0.011	1.7	108	153	106	98	1
	Release	Mean	0.019	142	-	-	-	-	64
		STD	0.01	1	-	-	-	-	2
/t/ ت	Closure	Mean	0.2833	-	-	-	-	-	25
		STD	0.014	-	-	-	-	-	5
	Release	Mean	0.0257	9788					52
		STD	0.008	14					10
	Closure	Mean	0.298	-	-	-	-	-	25
		STD	0.013	-	-	-	-	-	6

-Continued on next page -

/tʰ/ ت	Release	Mean	0.0587	4317	-	-	-	-	60
		STD	0.007	40	-	-	-	-	4
/ɖ/ ڈ	Closure	Mean	0.154	140	307	1704	2611	3099	56
		STD	0.010	5	74	23	208	437	1
	Release	Mean	0.0147	-	-	-	-	-	64
		STD	0.0035	-	-	-	-	-	2
/ʈ/ ٹ	Closure	Mean	0.185	-	-	-	-	-	53
		STD	0.018385	-	-	-	-	-	1
	Release	Mean	0.0145	9051	-	-	-	-	57
		STD	0.0035	38	-	-	-	-	1.6
/ɖʰ/ ڈھ	Closure	Mean	0.166	135	241	1769	2693	4113	52
		STD	0.0155	2	41	25	16	6	0.7
	Release	Mean	0.018	-	-	-	-	-	63
		STD	0.0057	-	-	-	-	-	1.4
/ʈʰ/ ٹھ	Closure	Mean	0.16	-	-	-	-	-	57
		STD	0.021	-	-	-	-	-	2
	Release	Mean	0.0705	3396	-	-	-	-	61
		STD	0.0106	21	-	-	-	-	1
/ʤ/ج	Closure	Mean	0.0745	132	219	1653	2991	4232	61
		STD	0.0101	3	45	257	226	77	2
	Release	Mean	0.0306	-	-	-	-	-	63
		STD	0.0045	-	-	-	-	-	3
/ʧ/ چ	Closure	Mean	0.2033	-	-	-	-	-	28
		STD	0.0205	-	-	-	-	-	6
	Release	Mean	0.0656	3832	-	-	-	-	47
		STD	0.01137	209	-	-	-	-	5

-Continued on next page -

/ʤ^h/ جھ	Closure	Mean	0.1335	133	223	1862	2908	4057	64
		STD	0.0078	10	30	136	152	179	2
	Release	Mean	0.0585	-	-	-	-	-	62
		STD	0.0049	-	-	-	-	-	4
/ʧ^h/ڇ	Closure	Mean	0.1603	-	-	-	-	-	28
		STD	0.0145	-	-	-	-	-	5
	Release	Mean	0.107	3537	-	-	-	-	54
		STD	0.009	290	-	-	-	-	9
/g/ گ	Closure	Mean	0.0823	125	265	1072	2565	3934	61
		STD	0.0051	1.5	34	167	120	76	1
	Release	Mean	0.022	-	-	-	-	-	54
		STD	0.003	-	-	-	-	-	2
/k/ڪ	Closure	Mean	0.174	-	-	-	-	-	28
		STD	0.0101	-	-	-	-	-	4
	Release	Mean	0.082	6601	-	-	-	-	58
		STD	0.12	108	-	-	-	-	2
/g^h/گھ	Closure	Mean	0.1323	130	304	1533	2765	4044	59
		STD	0.0191	3	113	749	500	367	3
	Release	Mean	0.0263	-	-	-	-	-	56
		STD	0.0047	-	-	-	-	-	5
/k^h/ ک	Closure	Mean	0.1585	-	-	-	-	-	31
		STD	0.0148	-	-	-	-	-	6
	Release	Mean	0.063	1531	-	-	-	-	58
		STD	0.0056	36	-	-	-	-	4

8.3 Implosives of Sindhi

The implosives are the ingressive sounds in the sense that they take air into the mouth during the production in vocal tract. These sounds are usually difficult to produce compared to the ejective sounds for which air flows out of the mouth when produced in the vocal tract. Implosives are difficult to produce this is why they are present only in ten percent of the world's languages (Ladefoged, 1996). Among the ten percent of the world's languages the majority of the languages include voiced implosives; however a few of the world's languages are reported having both voiced as well as voiceless implosives such a language is Owerri Igbo spoken in Nigeria and includes both voiced and unvoiced implosives (Ladefoged, 1996). The other languages which include the voiceless implosives are the Seereer-Siin language branch of the Niger-Congo family spoken in Senegal (Laughlin Mc, 2005), the Lendu language spoken in Democratic Republic of Congo and Uganda (Demolin, 1995), and the Owerri Igbo language spoken in Nigeria (Ladefoged, 1996) etc. Sindhi language primarily spoken in India and Pakistan exhibits a large number of the contrastive implosive consonants among these languages at four places of the articulation they are: the bilabial /ɓ/ ٻ, the retroflex /ᶑ/ ڏ, the Palato-Alveolar /ʄ/ ڄ and the velar /ɠ/ ڳ. The large number of the contrastive voiced implosives at four places of articulation is one of the unique linguistic-phonetic properties of Sindhi language. The fact that the implosives are difficult sounds to be produced than the ejectives, in the production of implosives the vocal folds are lowered down so that they suck air in and phonetically they are referred to as the glottal suction stops (Ladefoged, 1996). The authors (Nihalani, 1986; Raza, 2004) have reported that the implosives of Sindhi are purely non-aspirated glottal ingressive sounds produced by lowering the larynx with vibrating vocal cords.

In this section the acoustic-phonetic analysis of the implosive consonants of Sindhi is presented. The analysis results and discussions are mainly based on the voice samples collected during the field study for the implosive consonants discussed in detail in chapter III. The voice samples were recorded as a carrier sentence read by the native Sindhi speakers. The words that contain the target analysis implosive sound follow the VCV phoneme sequence at word medial position.

8.3.1 Bilabial implosive /ɓ/ ٻ

Figure 8.31 shows the typical waveform and spectrogram of the bilabial voiced implosive /ɓ/ ٻ, present in the word utterance /rəɓə/ رَٻ, of Sindhi. The slow varying periodic signal activity in the waveforms shown in figures 8.31 and 8.32(a) with the presence of the energy in the lower frequency range (the energy below 1 kHz) indicate that a voiced sound is produced. The energy present in the lower frequency range is weak for this bilabial implosive /ɓ/ ٻ of Sindhi compared to the other sonorant consonantal phonemes of Sindhi. The waveform of the bilabial /ɓ/ shows that the amplitude of the modulation signal keeps on increasing during the closure segment of this sound. This characteristic of the implosive stop /ɓ/ is opposite to its contrastive bilabial plosive stop /b/ of Sindhi. For which the amplitude of the modulation signals keep on decreasing as it moves to the end of the closure segment. This difference is shown in the manually extracted short waveforms of the word utterances shown in figures 8.32(a) is the waveform of the bilabial implosive /ɓ/ (b) is the waveform of the bilabial plosive stop /b/ and (c) is the waveform of the aspirated bilabial plosive stop /b^h/ of Sindhi. Note that the amplitude of the modulation signals gradually increase for the implosive /ɓ/; whereas it showed gradual decrease for the contrastive plosive stops /b/ and /b^h/ of Sindhi. Acoustically the gradual increase in the amplitude of the modulation signal can play an important role for the identification of the implosives of Sindhi in addition to the other acoustic parameters given in Table 8.2 for the implosive consonants of Sindhi.

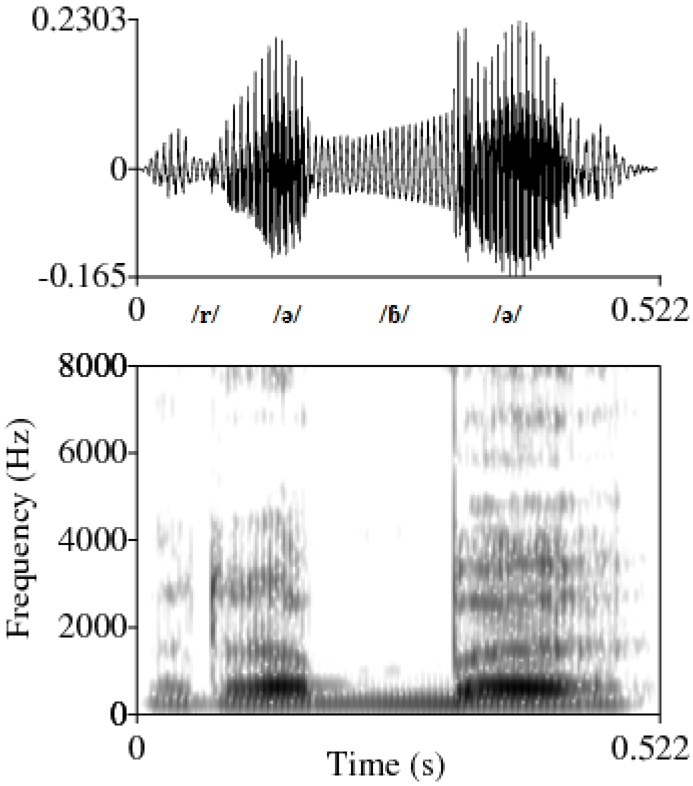


Figure 8.31 Waveform and the spectrogram of the word utterance /rəɓə/ رَٻ, containing the bilabial implosive /ɓ/ of Sindhi.

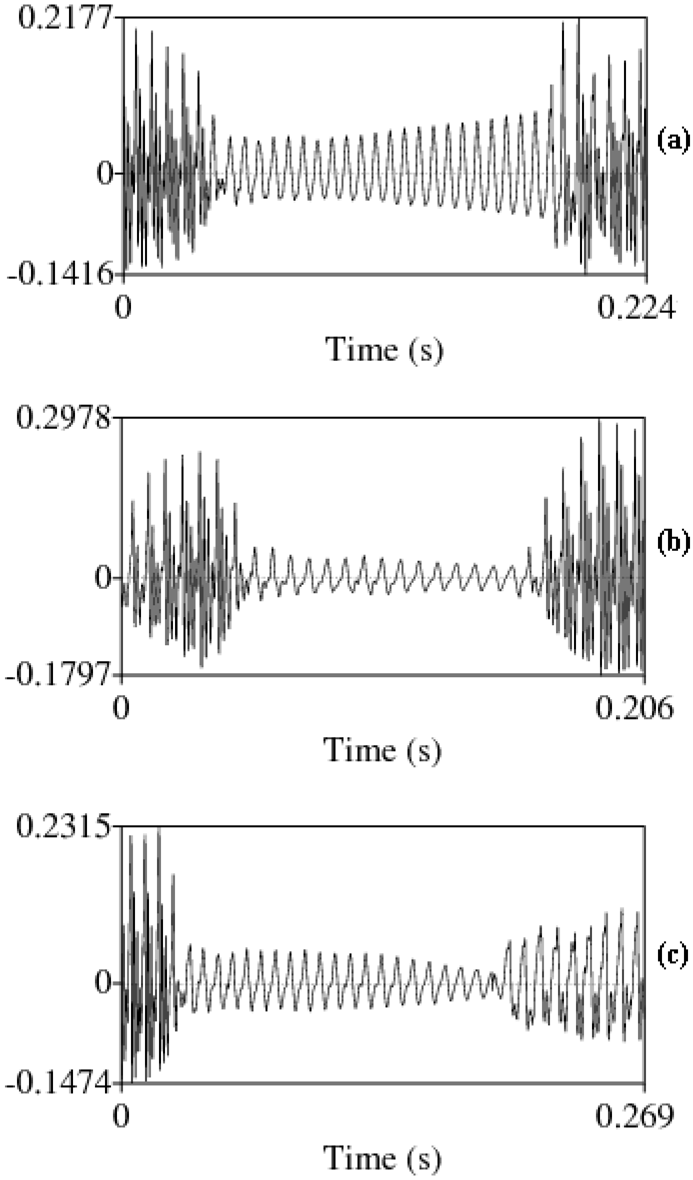


Figure 8.32 (a) the short waveform of the bilabial implosive /ɓ/ of Sindhi shows the increase in the amplitude (b) the short waveform of the bilabial plosive stop /b/ shows the decrease in the amplitude (c) the short waveform of the aspirated bilabial plosive stop /b^{h}/ also shows the decrease in the amplitude.

8.3.2 Retroflex Implosive /ᶑ/ ڏ

Figure 8.33 shows the typical waveform and the spectrogram of the retroflex voiced implosive /**ᶑ**/ ڏ, present in the word utterance /**səᶑə**/ سَڏ, of Sindhi. The slow varying periodic signal activity shown in the figures 8.33 and 8.34(a), and the presence of the energy in the lower frequency regions (the energy below 1 kHz) confirms that the retroflex /**ɗ**/ is a voiced phoneme of Sindhi. The sharp dark vertical striation immediately after the closure segment across the wideband spectrogram of the figure 8.33 indicates the release of the closure segment for retroflex /**ɗ**/. However the discontinuation in the vertical striation is seen for this sound above the 4 kHz of the frequency. The amplitude of the modulation signals keeps on increasing in the closure segment of this sound. Whereas the relative amplitude of the modulation signals for the contrastive retroflex plosive stops keeps on decreasing with respect to time as the speech signals move towards the end of the closure segment this is shown in the waveforms of the figures 8.34(a) is the waveform of the retroflex implosive /**ɗ**/ (b) is the waveform of the retroflex plosive stop /**ɖ**/ and (c) is the waveform of the aspirated retroflex plosive stop /**ɖʰ**/, of Sindhi at word medial position.

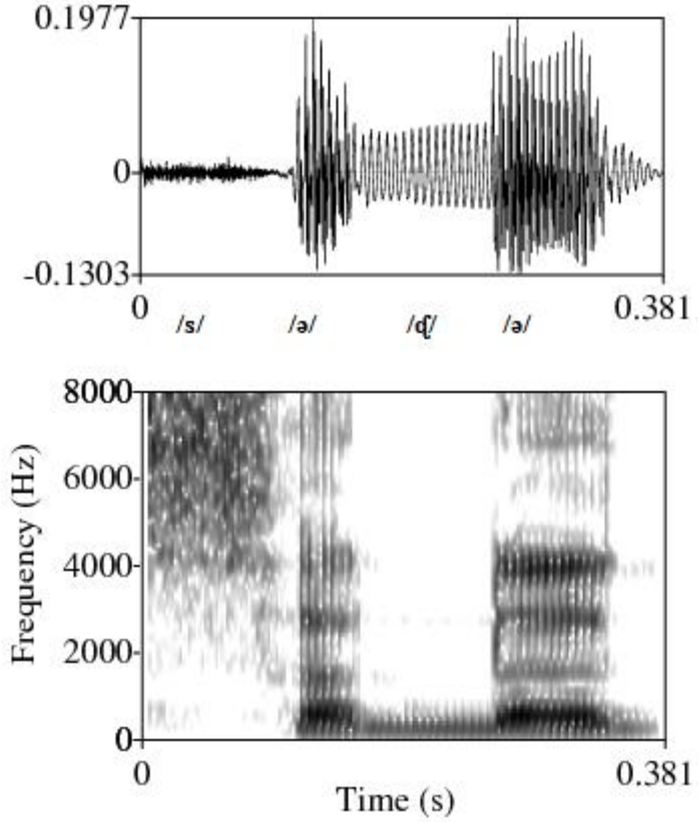


Figure 8.33 Waveform and the spectrogram of the word utterance /səɗə/ سڏ, containing the retroflex implosive /ɗ/ of Sindhi.

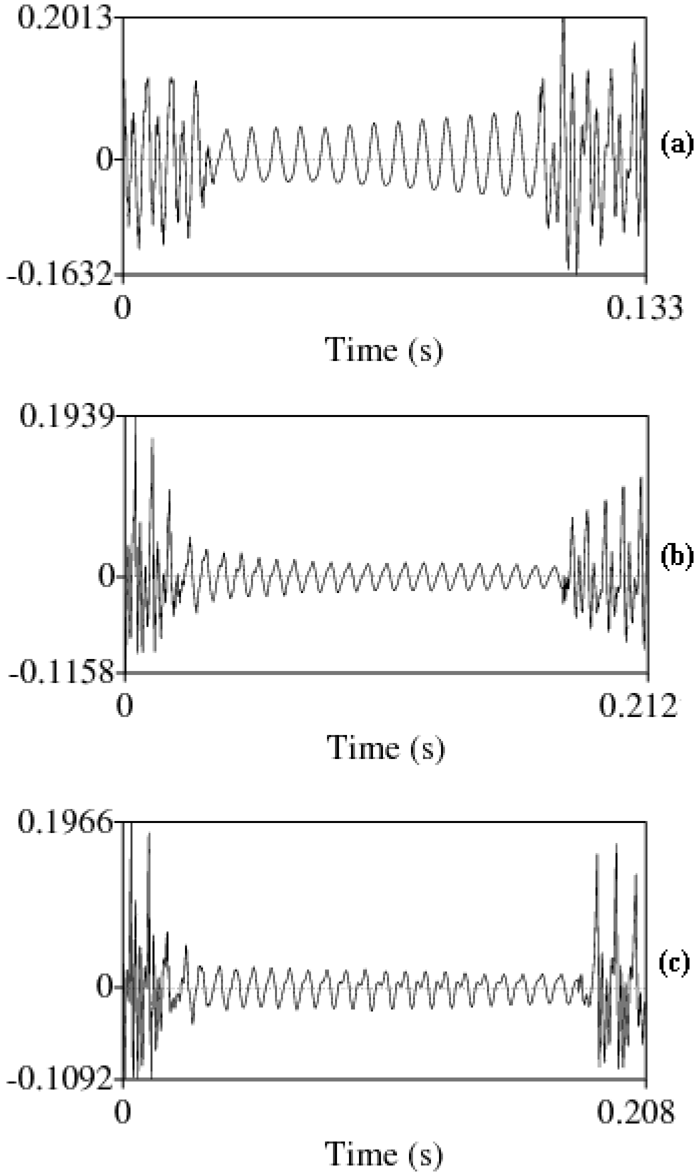


Figure 8.34 (a) the short waveform of the retroflex implosive /ɗ/ of Sindhi shows the increase in the amplitude (b) the short waveform of the retroflex plosive stop /ɖ/, shows the decrease in the amplitude (c) the short waveform of the aspirated retroflex plosive stop /ɖʰ/, shows the decrease in the amplitude.

8.3.3 Palato-alveolar implosive /ʄ/ ڄ

Figure 8.35 shows the typical waveform and the spectrogram of the palato-alveolar implosive /ʄ/ ڄ, present in the word utterance /səʄər/ سَڄَر, of Sindhi. The slow varying periodic signal activity shown in the waveform of figure 8.36(a) and the presence of the energy in the lower frequency regions (the energy below 1 kHz) confirm that the palato-alveolar implosive /ʄ/ is a voiced phoneme of Sindhi. The sharp dark vertical striation at the end of the closure segment across the wideband spectrogram of figure 8.35 indicates the release of the closure segment. The discontinuation in the dark vertical striation across the wideband spectrogram of this sound is not seen as the discontinuation observed for the retroflex implosive /ᶑ/ of Sindhi. One of the major differences observed for the palato-alveolar implosive /ʄ/ compared to the other three implosives of Sindhi is the amplitude of the modulation signal does not increase gradually for this sound as it keeps on increasing for other implosives of Sindhi this is shown in figures 8.35 and 8.36(a). On the other hand the decrease in the amplitude of the modulation signals for the contrastive palato-alveolar plosive stops is observed uniformly, this is shown in the waveforms of figure 8.36 (a) is the waveform of the palato-alveolar implosive /ʄ/, (b) is the waveform of the palato-alveolar plosive stop /ʤ/ and (c) is the waveform of the aspirated palato-alveolar plosive stop /$ʤ^h$/ of Sindhi at the word medial position.

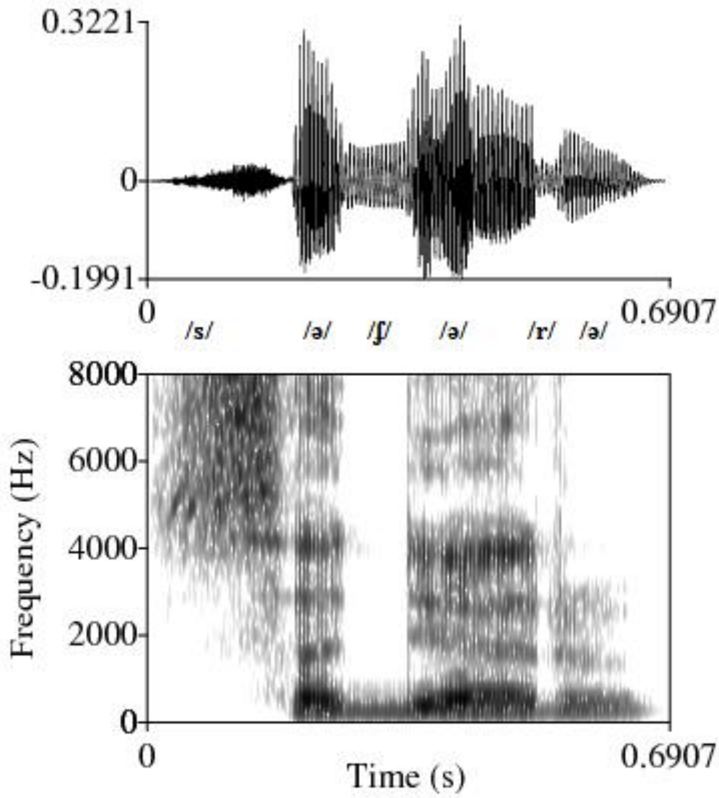


Figure 8.35 Waveform and the spectrogram of the word utterance /səʄər/ سڄر, containing the palato-alveolar implosive /ʄ/ of Sindhi.

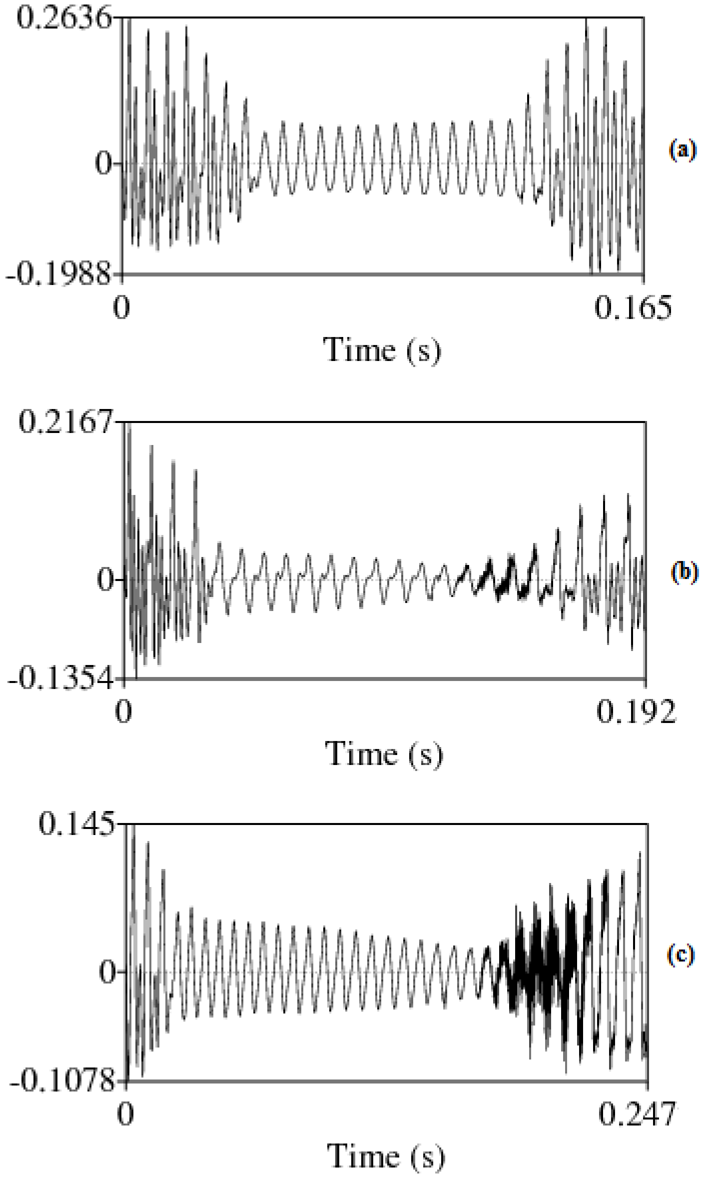


Figure 8.36 (a) the short waveform of the palato-alveolar implosive /ʄ/ of Sindhi do not show the increase in the amplitude (b) the short waveform of the palato-alveolar plosive stop /ʤ/, shows decrease in the amplitude (c) the short waveform of the aspirated palato-alveolar plosive stop /ʤ^h/, shows the decrease in the amplitude.

8.3.4 Velar implosive /ɠ/ ڳ

Figure 8.37 shows the typical waveform and the spectrogram of the velar implosive /ɠ/ ڳ, present in the word utterance /ɖ^{h}əɠo/ ڍڳو, of Sindhi. The slow varying periodic signal activity in the waveform of the signal shown in figures 8.37 and 8.38(a) and the presence of the weak energy in the lower frequency regions (the energy below 1 kHz) confirms that the velar implosive /ɠ/ is a voiced phoneme of Sindhi. The sharp dark vertical striation immediately after the closure segment across the wideband spectrogram of the figure 8.37 marks the release of the closure segment for this implosive sound. The discontinuation in the dark vertical striation is seen for the velar implosive /ɠ/ above the 4 kHz of the frequency. The amplitude of the modulation signal keeps on increasing during the closure segment of the implosive /ɠ/ and the relative amplitude of the modulation signals for the contrastive velar plosive stops keep on decreasing with respect to time as these sounds move towards the end of the closure segment this is shown in the waveforms of figure 8.38 (a) is the increasing amplitude waveform of the velar implosive /ɠ/ , (b) is the decreasing amplitude waveform of the velar plosive stop /g/ and (c) is the decreasing amplitude waveform of the aspirated velar plosive stop /g^{h}/ of Sindhi at word medial position.

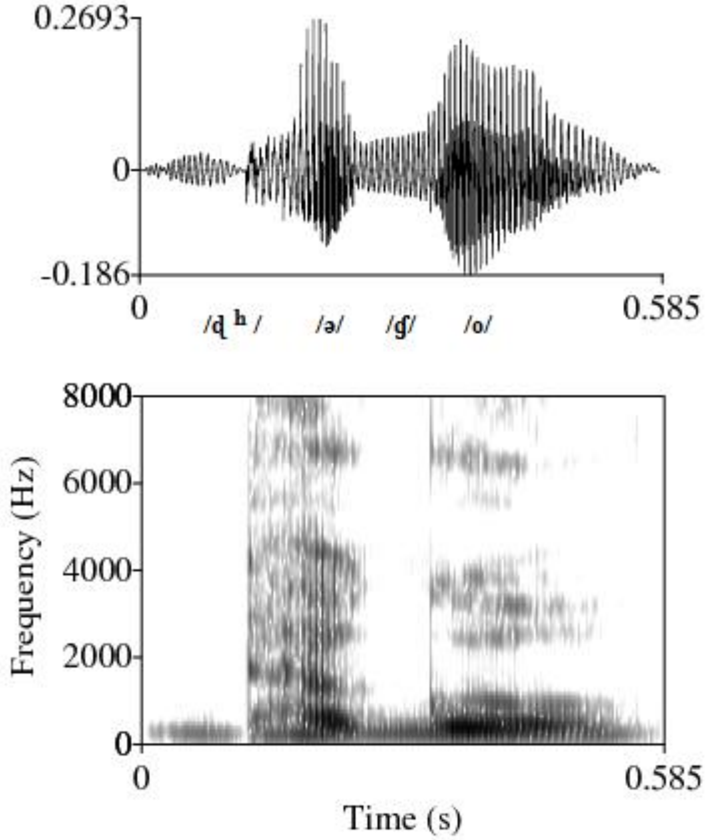


Figure 8.37 Waveform and the spectrogram of the word utterance /ɖ^{h}əɠo/ ڍَڳو, containing the velar implosive /ɠ/ of Sindhi.

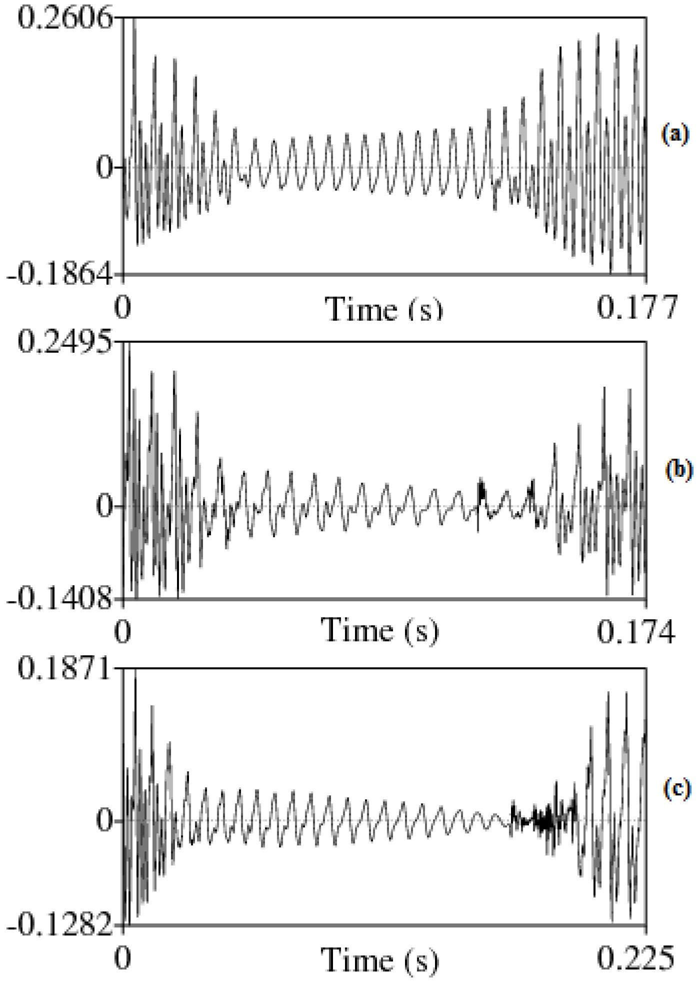


Figure 8.38 (a) the short waveform of the velar implosive /ɠ/, show's the gradually increasing amplitude (b) the short waveform of the velar plosive stop /g/, show's the decreasing amplitude and (c) the short waveform of the aspirated velar plosive stop $/g^h/$, show's the decreasing amplitude.

Table 8.2 Mean and standard deviation of the acoustic paramteres: F0, first four formants and duration of the implosive stops of Sindhi.

Implosive consonant	Parameter	Duration	F0	F1	F2	F3	F4	Intensity (dB)
/ɓ/ ٻ	Mean	0.097	156	189	944	2826	4055	66
	STD	0.005	3	4	167	207	157	--
/ɗ/ ڏ	Mean	0.094	167	213	1282	2752	4139	67
	STD	0.004	2	5	125	64	91	--
/ʄ/ ڄ	Mean	0.074	164	205	1346	2901	4106	70
	STD	0.0075	4	6	412	530	333	--
/ɠ/ ڳ	Mean	0.082	164	233	1245	2988	3951	68
	STD	0.0102	7	15	186	650	198	--

8.4 Nasal consonants of Sindhi

Nasals are produced by lowering the velum that allows air to escape through the nasal cavity. The nasals like the stop consonants are produced with the complete closure of the oral cavity; however the two sounds are different in the sense that the air passes through the nasal cavity during the production of nasals, whereas during the production of stops air passage is completely blocked and is suddenly released after the closure; because the velic port also remain closed during the production of stops (Olive, 1993). The air pressure that builds up during the production of stop consonants do not occur during the production of nasal sounds in the vocal tract; because the air passes through the opened velic port and hence resonate in the nasal cavity to produce '*murmur',* kind of nasal sounds. Sindhi includes nasals at five places of articulation, these are: the bilabial **/m/** م , the alveolar **/n/** ن , the retroflex /ɳ/ ڻ , the palatal /ɲ/ ڃ , and the velar **/ŋ/** ڱ. The acoustic waveforms of the five nasal consonants of Sindhi are shown in figures 8.39 for the bilabial /**m**/, 8.42 for the alveolar **/n/**, 8.45 for the retroflex /ɳ/, 8.48 for the palatal **/ɲ/,** and 8.51 for the velar **/ŋ/**. The waveforms of the nasal sounds have a slow varying periodic signal activity with the presence of energy in the lower frequency regions this confirms that all five nasals of Sindhi are voiced phonemes; like the three nasals of English they are usually referred to as the sonorant (voiced) phonemes (Johnson, 2003; Ladefoged, 2005).

In the production of the nasal sounds the formation of the resonator tube (vocal tract) through the nasal cavity is longer compared to the tube formed for the oral sounds. The vocal tract tube is approximately 17cm long in adult males and the length of the nasal cavity (from uvula to nostrils) is approximately 13cm long (Parsons, 1987). The vocal tract length from glottis to velum is approximately 8.5cm; branching off the nasal cavity with the main resonator tube provides a resonator tube with a length of 21.5cm=8.5cm+13cm (Parsons, 1987). Therefore the sounds produced with the nasal cavity absorb more acoustic energy compared to the non-nasals; because of the fact that the longer resonator tube absorbs more acoustic energy. This phenomenon in acoustic-phonetics is referred to as the decrease in the amplitude of the modulation signal or the damped waveforms along with the lower resonant frequencies (Johnson, 2003;

Kurowski, 1987). Increased damping due to the larger resonator tube also results in a larger formants bandwidth (Johnson, 2003). The nasal sounds of Sindhi have shown strongly damped waveforms and hence a longer formants bandwidth; this is shown in the spectrogram and the spectrum figures of the five nasals of Sindhi. The spectrogram figures help to analyze how frequency changes in time for nasal sounds and the spectrum figures show how amplitude of the harmonics (the resonances) change in frequency for nasals of Sindhi. A high energy level is observed for F1 compared to the higher formants uniformly in all five nasals of Sindhi. The energy for the higher formants is weak but still some resonance structure is observable in the spectrogram figures. The resonance structure for nasals of Sindhi is stable at least for the first three formants; this acoustic characteristic is only seen in the glide and the nasal consonants of Sindhi. The weak energy for the higher formants (the damped resonances) is due to the anti-resonances caused by the side branches i.e. the nasal sinuses and the oral cavity (Johnson, 2003; Kurowski, 1987). In the subsequent sections of this chapter we will present the acoustic analysis of the nasal phonemes of Sindhi keeping in view the acoustic factors involved in the production of the nasal sounds in detail. The acoustic analysis is based on the voice samples recorded for the nasal sounds during the field study discussed in detail in chapter III. The selected word utterances for recording contain the VCV phoneme sequence at the word medial position. This VCV phoneme sequence in utterance helps to analyze the motion of the formants coming into the nasal sounds and going away from them.

8.4.1 Bilabial nasal /m/ م

Figure 8.39 below shows the typical waveform of the bilabial nasal /**m**/ م, of Sindhi present in the word utterance /**səmə**/ سَمَ. The waveform shows the presence of slow varying periodic signals this indicates that the nasal /**m**/ of Sindhi is a voiced phoneme. Articulatory nasal /**m**/ م is different from the rest of the nasals in such a way that the whole oral cavity becomes the side branch; because the oral cavity occlusion happens at the lips for this nasal. This also introduces the longest anti-resonance side branch in the main resonator tube for the production of the bilabial nasals (Johnson, 2003) such as the bilabial nasal /**m**/ م of Sindhi. The decrease in the energy above

700Hz and again above 2500Hz is typical for bilabial nasal /**m**/ م of Sindhi shown in the spectrogram of figure 8.41. This decrease in the energy is caused by the anti-resonances of the side branches referred to as energy damping. The decrease in the energy above 700 Hz for nasal /**m**/ indicate that the anti-resonant side branches have some how cancelled out the F2 transitions shown in figure 8.40 in which the F2 can be seen strongly damped. The damping in energy increases the spectral bandwidth and decreases the spectral peak (Olive, 1993; Johnson, 2003). This decrease in the harmonic peak and increase in the F2 bandwidth is shown for bilabial nasal /**m**/ in the spectrum illustrated in figure 8.40. The formants motion coming into the nasal sound and going away from it can play an important role for the identification of the nasal sounds (Olive, 1993). The F1 and F2 motion for the bilabial nasal /**m**/ can be defined as the downward (the motion coming into the nasal /**m**/) and the upward (the motion going away from it); because of the low F1 and F2 values for this nasal sound shown in figure 8.41 the formant tracks. The mean acoustic parameter values and standard deviation is given in Table 8.3 for this nasal sound.

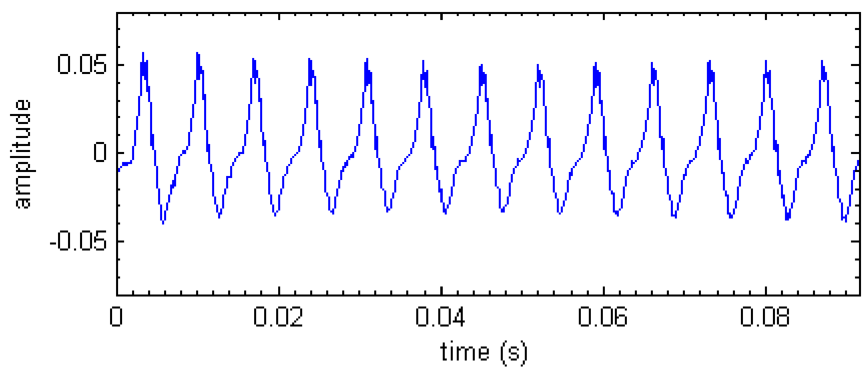


Figure 8.39 Waveform of the bilabial nasal /m/ م of Sindhi, taken from the word utterance /səmə/ سَمَ .

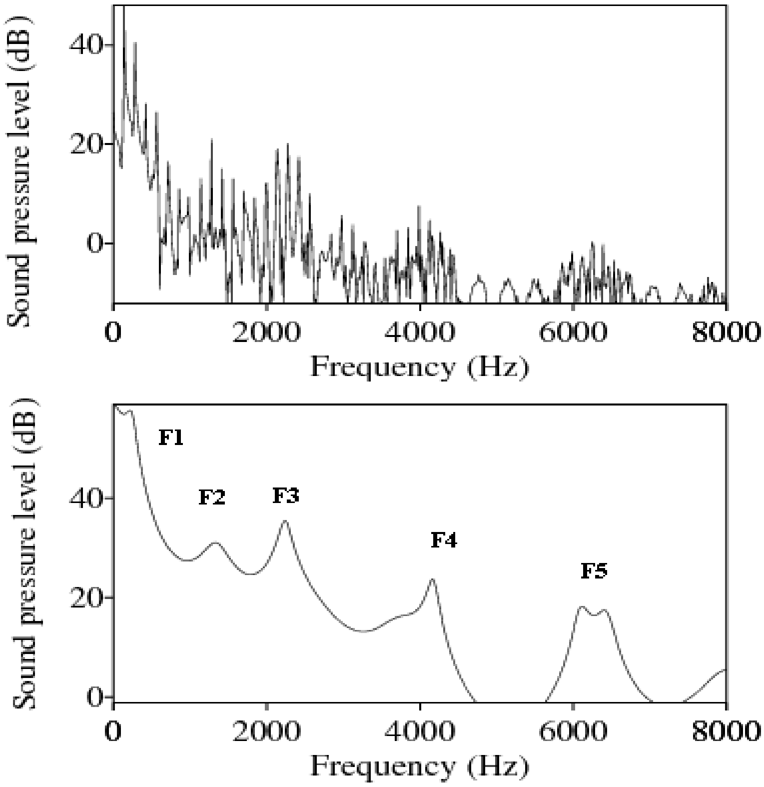


Figure 8.40 Spectrum of the nasal consonant /m/ م , of Sindhi, along with the spectrum slice showing the harmonic peaks.

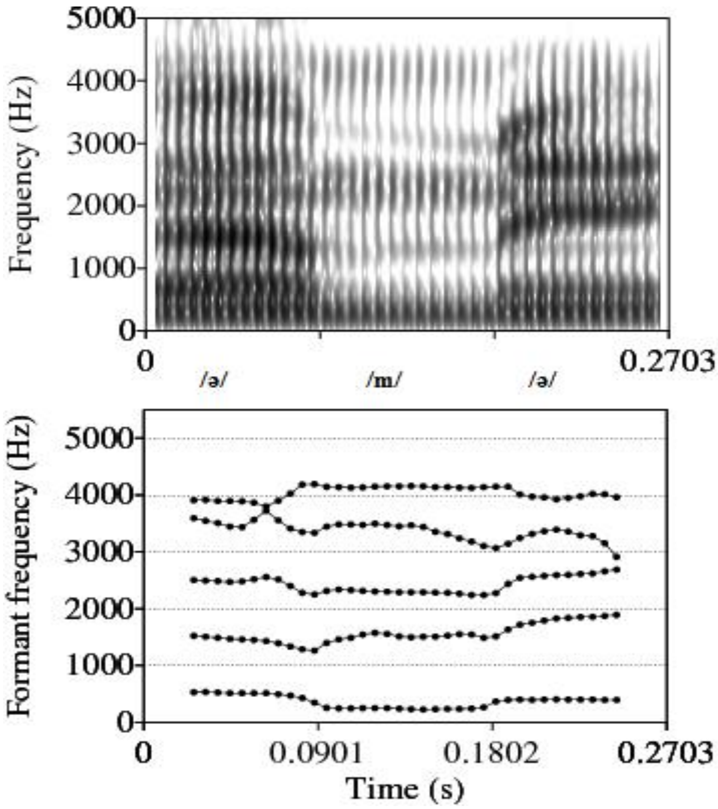


Figure 8.41 Spectrogram of the bilabial nasal /m/ م , of Sindhi taken from the word /səmə/ سَمَ, along with the formant tracks.

8.4.2 Alveolar nasal /n/ ن

Figure 8.42 below shows the waveform of the alveolar nasal /**n**/ ن, of Sindhi extracted manually from the word utterance /**sənɑ**/ سَنا. The presence of slow varying periodic signals and the energy in the lower frequency regions shown in the spectrogram in figure 8.44 confirm this alveolar nasal /**n**/ a voiced phoneme of Sindhi. The decrease in the energy for this sound is observed above 800Hz and again above 1600Hz for this sound. The anti-resonances caused by side branches have cancelled out the resonances of the third formant for this sound. The F3 comes closer to the F4 and is strongly damped compared to the F3 of the bilabial nasal /**m**/ for which F2 is strongly damped see spectrogram figures 8.41 and 8.44 for the comparison of these nasals of Sidnhi. Due to the low F1 and F2 the formants motion coming into the alveolar nasal /**n**/ is downward (coming into the nasal /**n**/ from the preceding sound) and upward (going away to the immediately following sound) for this nasal see figure 8.44 the formant tracks for alveolar nasal /**n**/ of Sindhi. The mean acoustic parameter values and standard deviation is given in Table 8.3 for this nasal sound.

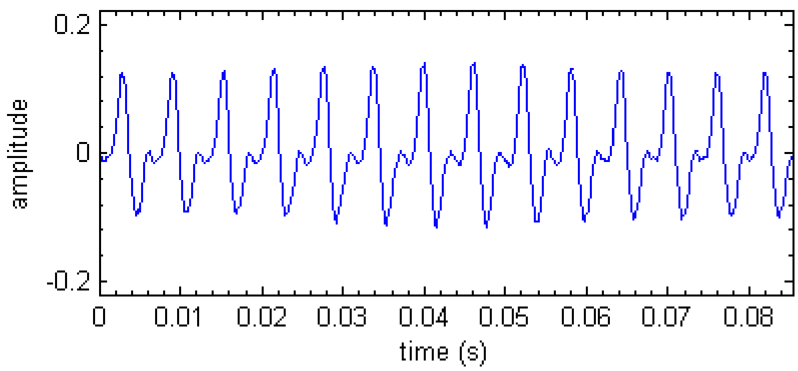


Figure 8.42 Waveform of the alveolar nasal /n/ ن of Sindhi, taken from the word utterance /sənɑ/ سَنا .

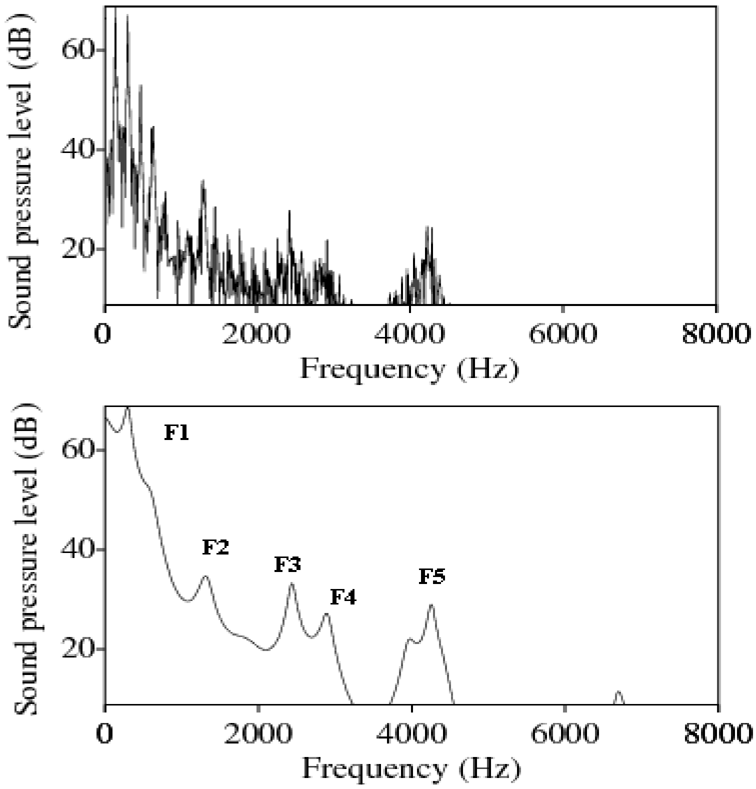


Figure 8.43 Spectrum of the alveolar nasal /n/ ن of Sindhi, along with the spectrum slice showing the harmonic peaks.

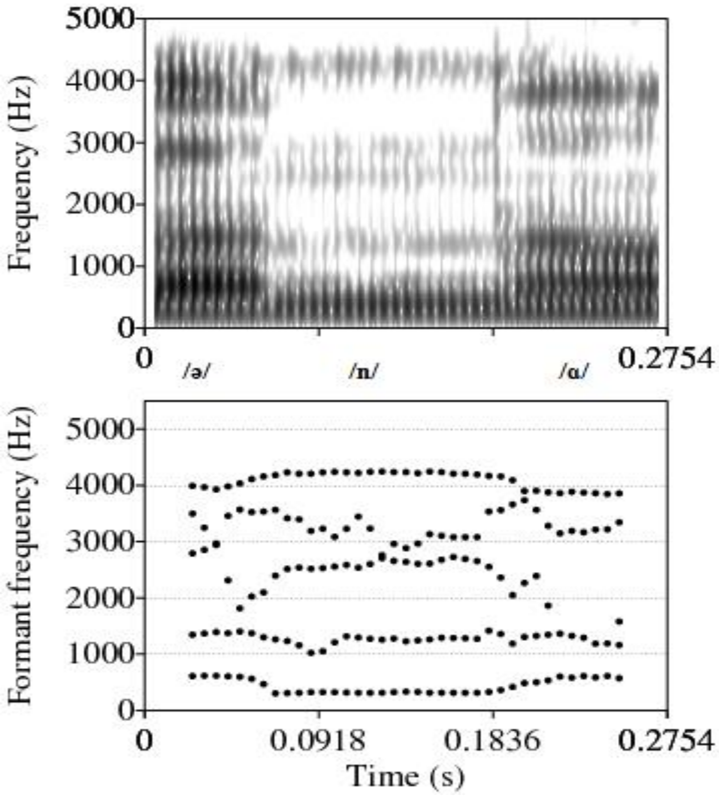


Figure 8.44 Spectrogram of the alveolar nasal /n/ ن , of Sindhi taken from the word /sənɑ/ سَنا, along with the formant tracks.

8.4.3 Palatal nasal /ɲ/ ڃ

Figure 8.45 below shows the waveform of the palatal nasal /ɲ/ ڃ, of Sindhi extracted manually from the word utterance /ʃəɲə/ ڇَڃَ. The presence of slow varying periodic signals and the energy in the lower frequency regions confirm that this is a voiced phoneme of Sindhi. The F2 is strongly damped and has been cancelled out by the anti-resonances caused by side branches. The F2 for this nasal comes closer to the F3 as shown in the spectrum figure 8.46 and in the spectrogram figure 8.47 of the palatal nasal /ɲ/ of Sindhi. The nasal /ɲ/ has a low and stable F1 transitions; therefore the F1 motion for this nasal is downward (transitions coming into the nasal /ɲ/) and upward (going away from it), whereas the F2 is not stable for this nasal and have shown discontinuities at the vowel junctures; therefore it is difficult to define the F2 motion for this nasal sound of Sindhi. The mean acoustic parameter values and standard deviation is given in Table 8.3 for this nasal sound.

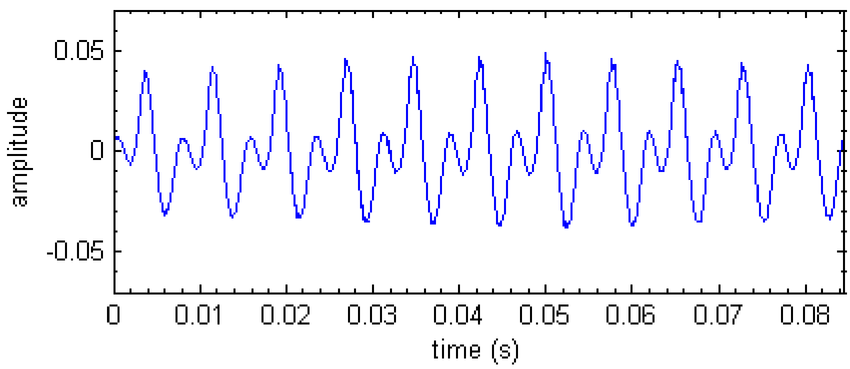


Figure 8.45 Waveform of the palatal nasal /ɲ/ ڃ of Sindhi, taken from the word utterance /ʃəɲə/ ڇَڃَ .

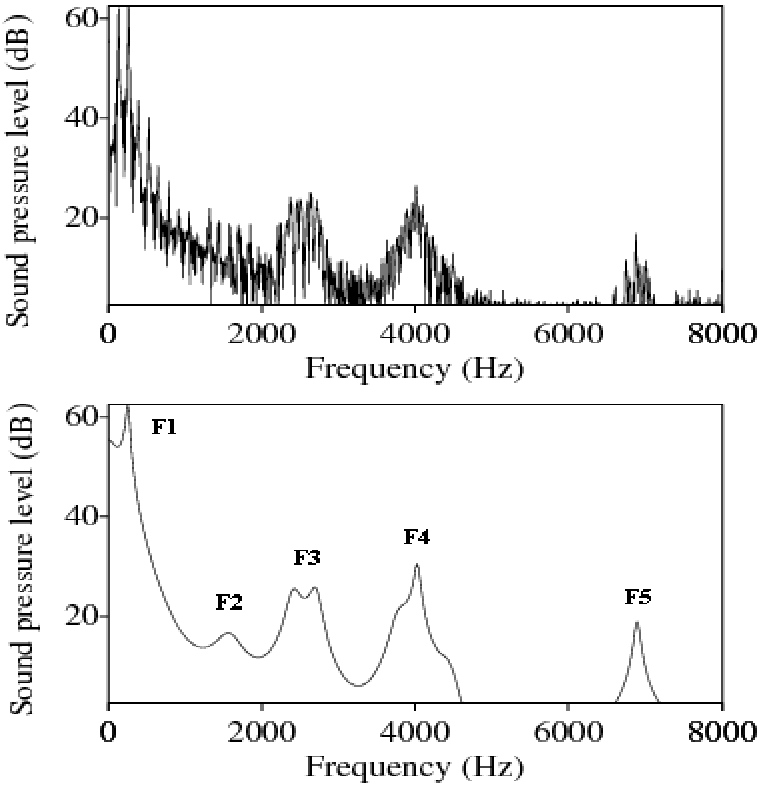


Figure 8.46 Spectrum of the palatal nasal /ɲ/ ڃ, of Sindhi along with the spectrum slice showing the harmonic peaks.

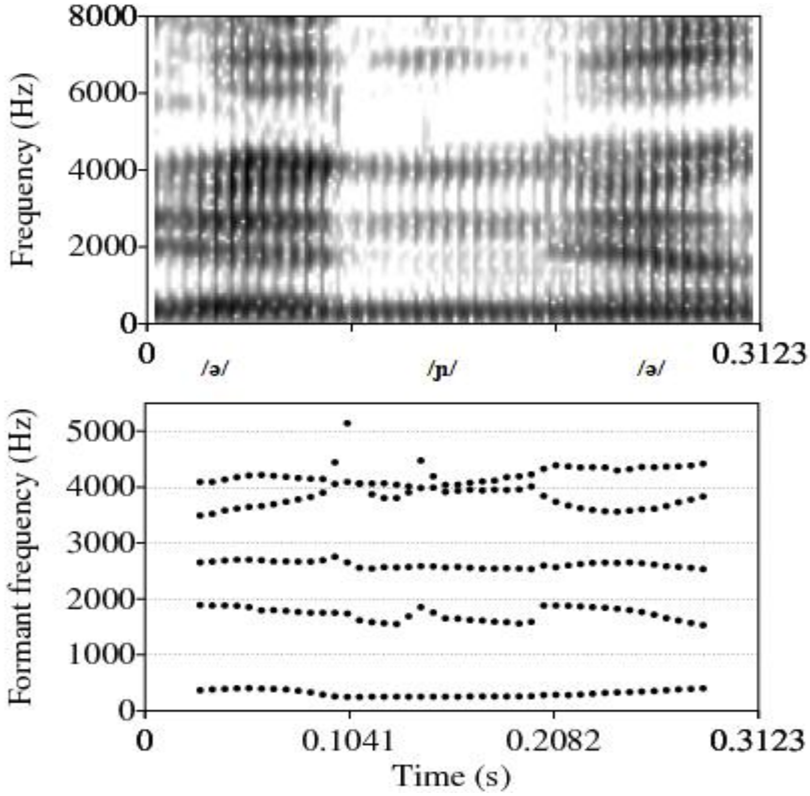


Figure 8.47 Spectrogram of the palatal nasal /ɲ/ ڃ of Sindhi, taken from the word /ɟəɲə/ ڄڃ, along with the formant tracks.

8.4.4 Retroflex nasal /ɳ/ ڻ

Figure 8.48 below shows the waveform of the retroflex nasal /ɳ/ ڻ, of Sindhi extracted manually from the word utterance /kaɳʊ/ ڪاڻُ. Among the five nasals of Sindhi only the retroflex nasal /ɳ/ has shown significant decrease in the amplitude of the wave signals at the middle of the utterance for a very short duration compared to the other nasals of Sindhi, see the waveform signals under the horizontal line (bar) in figure 8.48. The retroflex nasal /ɳ/ is also different in the formants motion compared to the other nasals of Sindhi. The motion for F_1, F_3 and F_4 is downward (while coming into the nasal /ɳ/) and upwards (going away from it); whereas the F_2 motion for nasal /ɳ/ is upward (coming into) and downward (going away from it) see the spectrogram in figure 8.50. Despite the additional decrease in the amplitude of the modulation signals of nasal /ɳ/, the waveform signals still show the slow varying periodic signal activity and the presence of energy in the lower frequency regions, this indicates that the nasal /ɳ/ of Sindhi is a voiced phoneme. Another distinguishing acoustic characteristic observed for this retroflex nasal /ɳ/ is the downward motion of the third formant having the frequency value below 2 kHz. The F_3 for this nasal is strongly damped and comes closer to F_2 see figure 8.50. The mean acoustic parameter values and standard deviation is given in Table 8.3 for this nasal sound of Sindhi.

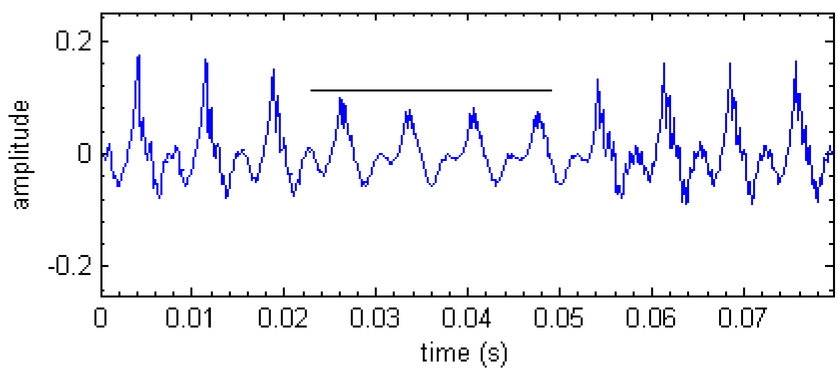


Figure 8.48 Waveform of the retroflex nasal /ɳ/ ڻ of Sindhi, taken from the word utterance /kaɳʊ/ ڪاڻُ .

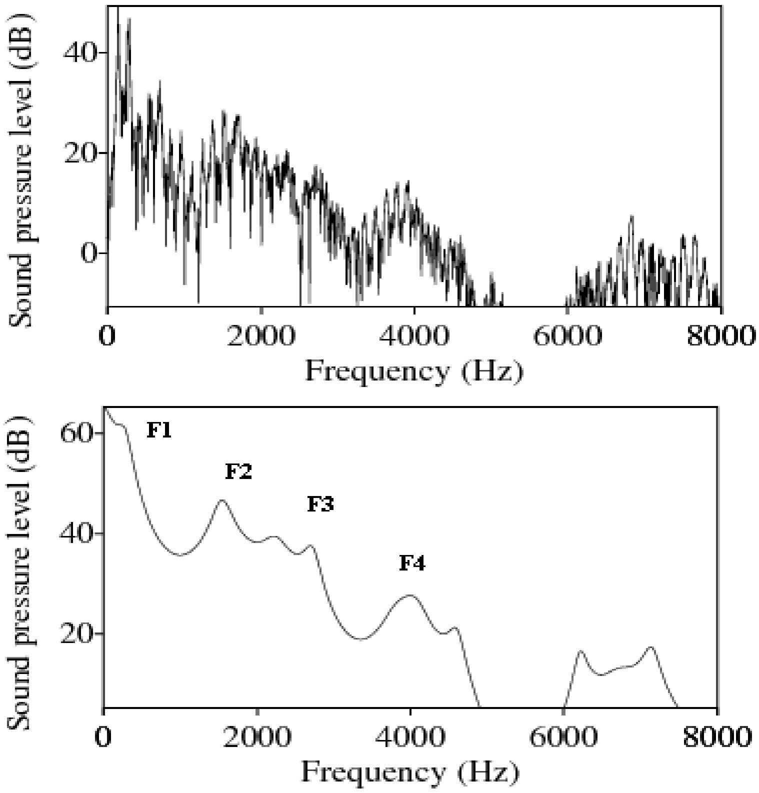


Figure 8.49 Spectrum of the retroflex nasal /ɳ/ ڻ of Sindhi along with the spectrum slice showing the harmonic peaks.

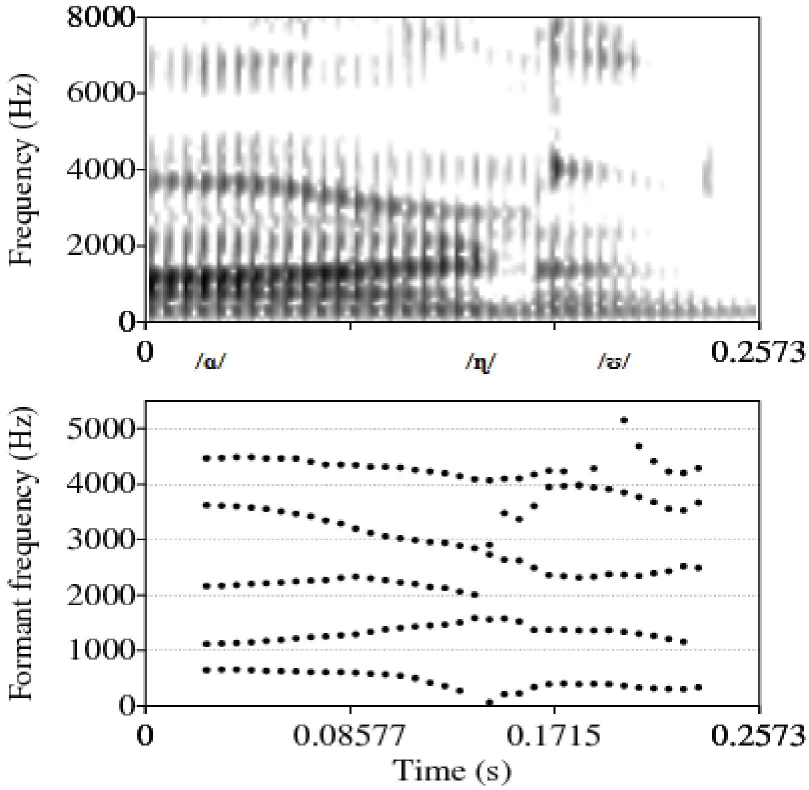


Figure 8.50 Spectrogram of the retroflex nasal /ɳ/ ڻ of Sindhi, taken from the word utterance /kɑɳʊ/ ڪاڻُ, along with the formant tracks.

8.4.5 Velar nasal /ŋ/ ڱ

Figure 8.51 below shows the waveform of the velar nasal **/ŋ/** ڱ, of Sindhi extracted manually from the word utterance **/rəŋə/** رَڱ. The presence of slow varying periodic signals and the energy in the lower frequency regions confirm this velar nasal /**ŋ**/ is a voiced phoneme of Sindhi. The F2 for this nasal sound is strongly damped and comes closer to the F3, see figure 8.52. This is the only nasal sound of Sindhi for which the second formant resonances are observed above 2 kHz. This sound has a low F_1; therefore the F_1 motion is downward (coming into the nasal sound /**ŋ**/) and upward (going away form it); while a very high F_2 for this nasal sound results in the upward motion (coming into the nasal) and downward (going away from it) for this nasal sound see the spectrogram of the figure 8.53. The mean acoustic parameter values and standard deviation is given in Table 8.3 for this nasal sound of Sindhi.

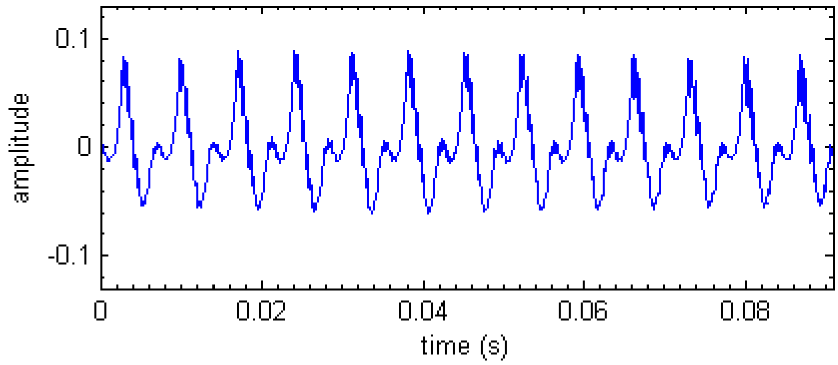


Figure 8.51 Waveform of the nasal consonant /ŋ/ ڱ of Sindhi, taken from the word utterance /rəŋə/ رَڱ .

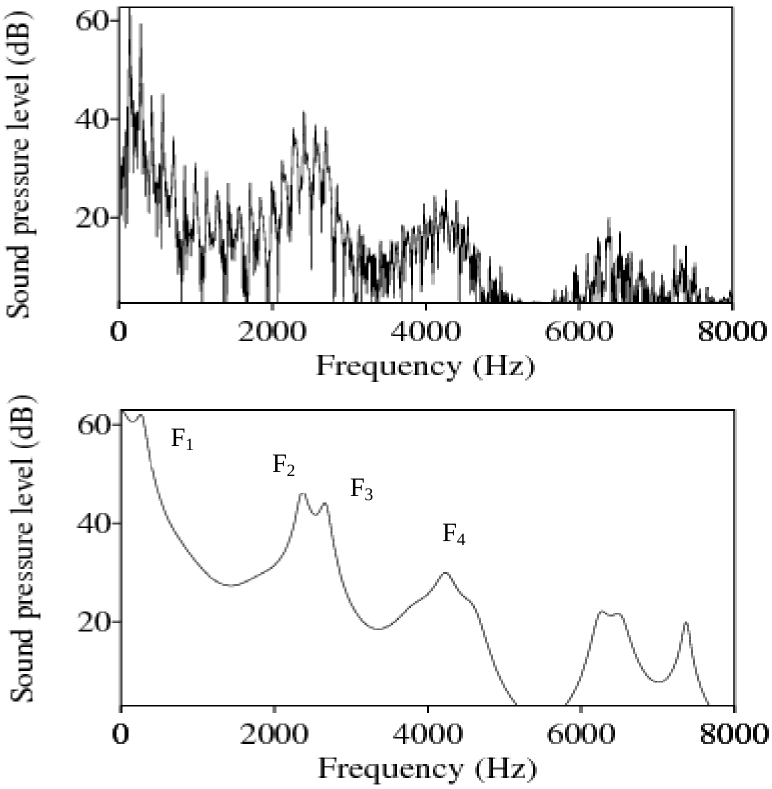


Figure 8.52 Spectrum of the nasal consonant /ŋ/ ڱ of Sindhi along with the spectrum slice showing the harmonic peaks.

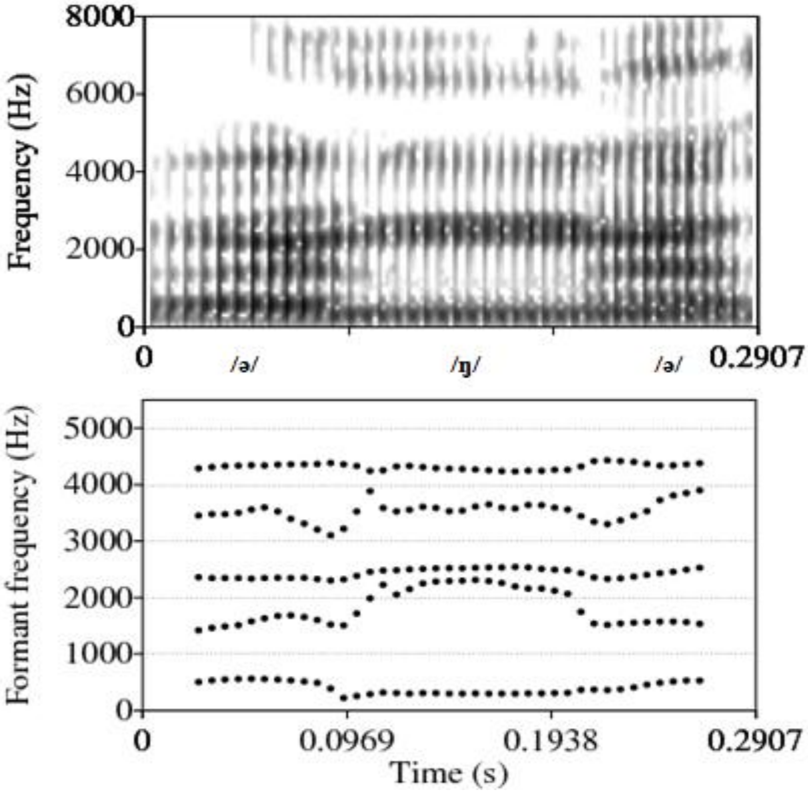


Figure 8.53 Spectrogram of the nasal consonant /ŋ/ ڱ of Sindhi, taken from the word utterance /rəŋə/ رَڱ , along with the formant tracks.

Table 8.3 Mean and standard deviation of the acoustic parameters: F0, first four formants and duration of the nasal consonants of Sindhi.

Nasal consonant	Parameter	Duration	F_0	F_1	F_2	F_3	F	Intensity (dB)
/m/ م	Mean	0.0973	143	238	1550	2291	3429	62
	STD	0.0156	1	18	51	35	164	0.07
/n/ ن	Mean	0.103	161	320	1233	2608	3140	70
	STD	0.0192	2	6	84	70	191	0.136
/ɳ/ ڻ	Mean	0.0642	136	291	1660	2497	3690	60
	STD	0.0065	0.018	55	48	39	119	12
/ɲ/ ڃ	Mean	0.0936	128	256	1676	2561	3985	61
	STD	0.0069	2	1	97	14	41	0.412
/ŋ/ ڱ	Mean	0.096	140	306	2209	2518	3590	64
	STD	0.011798	3	5	89	19	40	0.18

8.5 Fricative consonants of Sindhi

Fricatives are the sounds produced with highly constricted vocal tract; when air escapes through the constricted area in the vocal tract then such hissing or hushing fricative sounds are produced (Roach, 2009). During the production of fricative sounds articulators are brought close together to achieve high enough vocal tract constriction; however the complete vocal tract closure does not happen as it occurs in the production of stop consonants. This difference makes the fricatives the continuant consonants of Sindhi; which means that the speakers can continue making them without interruption until enough air can be ejected from the lungs, whereas the stop consonants cannot be continued in this way and they are referred to as non-continuant consonants of Sindhi. Sindhi includes eight fricative consonants at five places of articulation which are: the bilabial: /f/ ف and /v/ و, the alveolar: /s/ س and /z/ ز, the palatal: /ʃ/ ش , the velar: /x/ خ and /ɣ/ غ, and the glottal /h/ ھ. In this section the acoustic-phonetic analysis of the fricatives of Sindhi is presented in addition to the available articulatory-phonetics knowledge for their identification and classification. Acoustically the prediction of the voicing feature for fricatives of Sindhi is relatively difficult compared to the other consonants particularly the voiced fricatives of Sindhi. In the production of voiced fricatives the air passage is random in a sense that the air passage increase and decrease with vibrating vocal folds when it flows through the highly constricted vocal tract during the production of the voiced fricatives (Maniwa, 2009; Olive, 1993). This behaviour of increasing and decreasing air passage for the voiced fricatives results in quasi-periodic waveforms for these sounds. Whereas in the production of the unvoiced fricatives air passes uniformly through the highly constricted vocal tract; this results in a non-periodic speech waveform for these sounds (Olive, 1993; Tabain, 2001). Therefore the prediction of the voicing feature for the fricatives of Sindhi in this study depends upon the analysis of the acoustic waveforms and the spectral analysis. Three figures are shown for each fricative sound of Sindhi; the waveform figure in which the occurrence of the periodic signals in the speech waveform are analyzed to determine the voicing feature, the power spectrum figure in which the peak at the fundamental frequency and the local variations in the harmonics are analyzed, and the spectrogram figure in which the energy distribution and the formant transitions are analyzed. The acoustic analysis is mainly based on the speech samples collected during the field study discussed in chapter III. The speech samples for the fricative sounds were captured in a CVC phoneme sequence at the word initial position. Finally in Table 8.4 the mean acoustic parameter values and the standard deviation for the fricative class of sounds of Sindhi are given.

8.5.1 Bilabial voiced fricative /v/ و

Figure 8.54 below shows the acoustic waveform of the fricative /**v**/ و, of Sindhi extracted manually from the word utterance **/vəɖo/** وَڏو. The waveform starts with the weak signals along with friction noise signals and gradually turns into the slow varying periodic signals with significant increase in the amplitude of the modulation signals. This gradual increase in the amplitude of signals along with the periodicity and the presence of the energy in the lower frequency regions confirm this that the bilabial fricative /**v**/ is a voiced phoneme of Sindhi, see the waveform figure 8.54 and the spectrogram in figure 8.56. The voiced fricatives show large random variations in the amplitude of the spectra (the harmonics) and a peak at the fundamental frequency

(Maniwa, 2009; Olive, 1993). The Spectrum, figure 8.55, for the fricative /**v**/ و, of Sindhi shows the peak at the fundamental frequency (F0, 147 Hz) and this peak diminish roughly after six harmonics below 1 kHz. The formant transitions for this sound are only clear and stable for the first formant, the transitions for the second, third and fourth formant are ambiguous and unstable; hence can not be used for the identification of this sound the formant tracks are shown in figure 8.56 for the bilabial fricative /v/ of Sindhi.

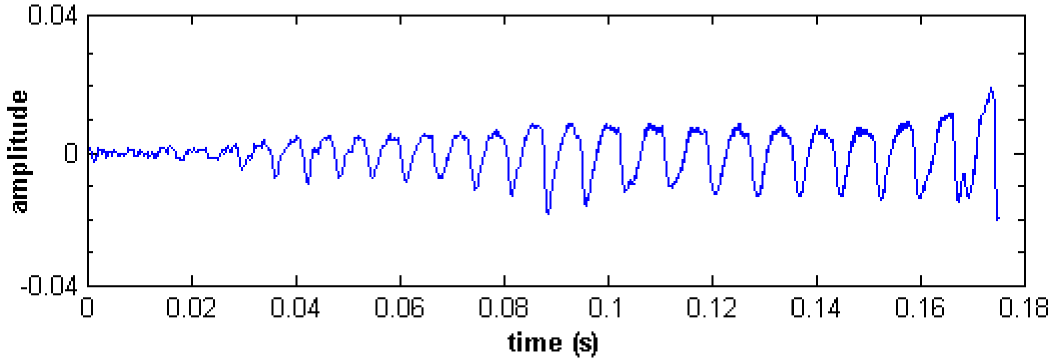


Figure 8.54 Waveform of the fricative consonant /v/ of Sindhi, taken from the word utterance /vəɖo/ وڏو.

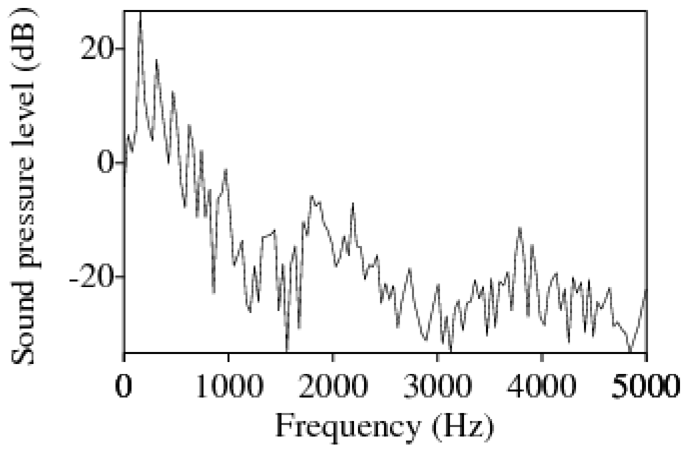


Figure 8.55 Spectrum of the fricative consonant /v/ of Sindhi shows peak at the fundamental frequency and large local variations in the amplitude.

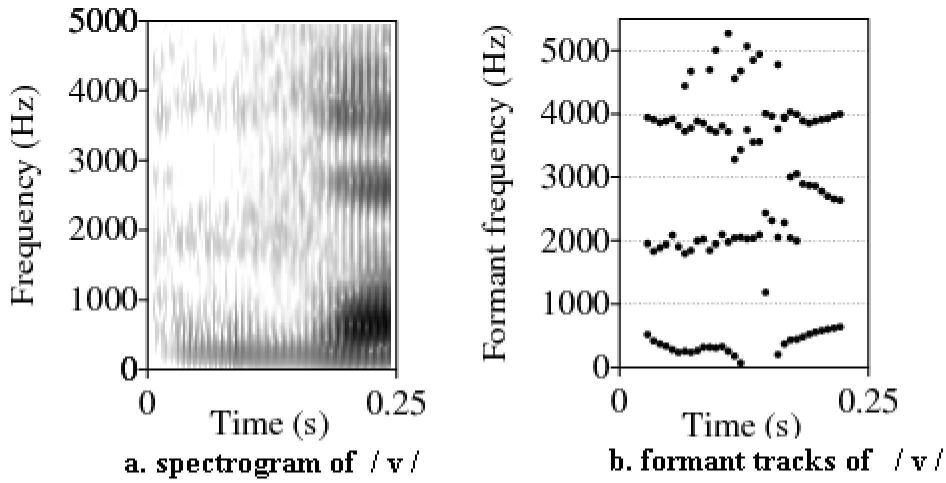


Figure 8.56 (a) Spectrogram of the fricative consonant /v/ followed by a vowel sound (b) the formant tracks.

8.5.2 Bilabial unvoiced fricative /f/ ف

Figure 8.57 below shows the waveform of the fricative sound /f/ ف, of Sindhi taken from the word utterance /fıkır/ فِڪر. The waveform shows the strong friction noise signals, whereas the presence of the periodic signals can not be witnessed in the sound waveform. The absence of the periodic signals and the energy below 1 kHz confirm this fricative sound an unvoiced phoneme of Sindhi. The presence of the energy in the higher frequency range across the wideband spectrogram is a common acoustic feature for the voiceless fricative sounds of Sindhi. There is no harmonic peak below 1 kHz for this sound in the spectrum of figure 8.58 as it was observed for the bilabial fricative /**v**/ of Sindhi. The formant transitions are ambiguous and they are not stable for this sound and hence can not be used for the identification of this sound.

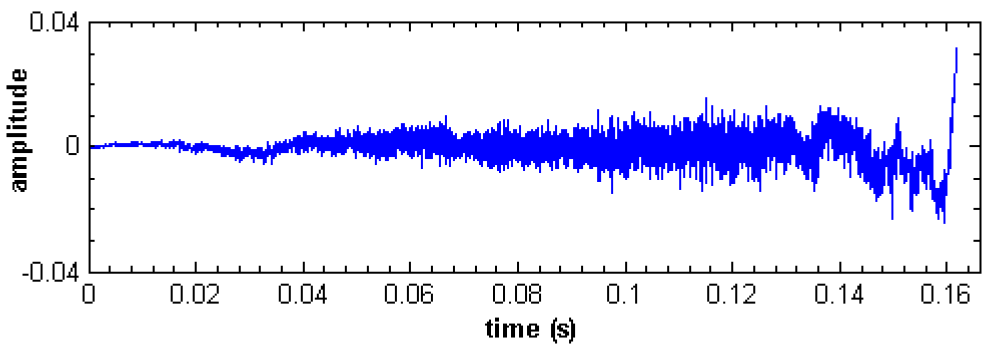


Figure 8.57 Waveform of the fricative consonant /f/ of Sindhi, taken from the word utterance /fıkır/ فِڪر.

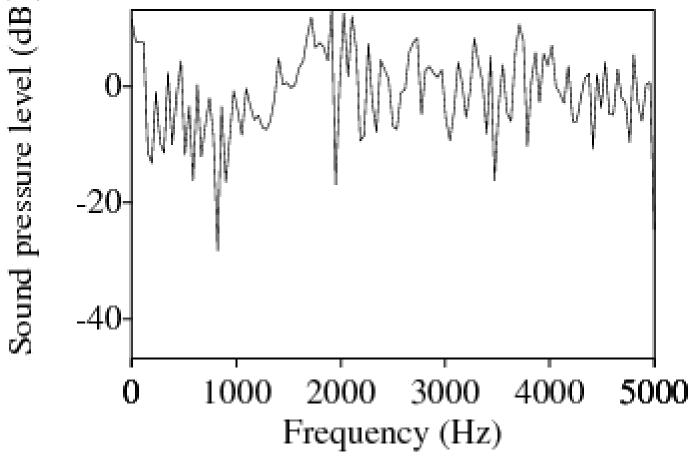


Figure 8.58 Spectrum of the fricative /f / of Sindhi that do not show peak at the fundamental frequency.

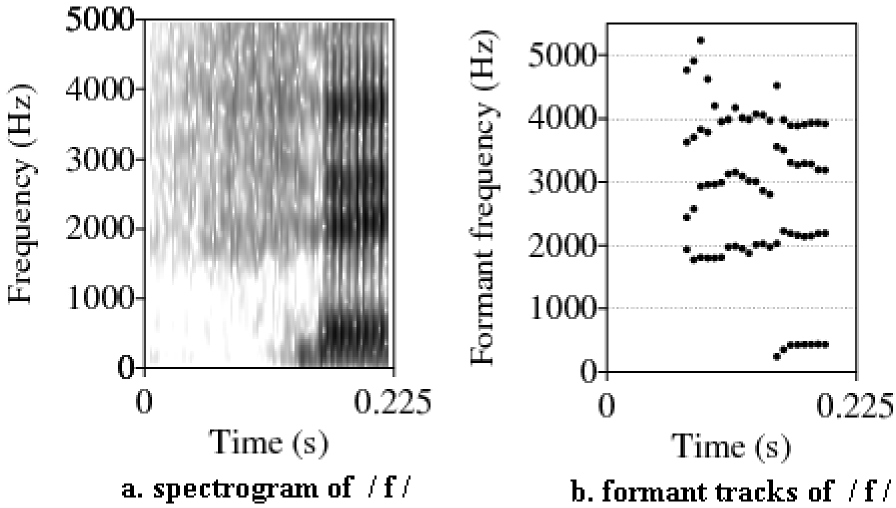


Figure 8.59 (a) Spectrogram of the fricative /f/ of Sindhi followed by a vowel (b) formant tracks.

8.5.3 Alveolar voiced fricative /z/ ز

Figure 8.60 below shows the waveform of the fricative /z/ ز, of Sindhi extracted manually from the word utterance /zɑri/ زاري. The waveform shows the weak periodic signal activity at the start of the waveform and gradually turn into the strong friction noise signals from mid till the end of the waveform; however the periodicity in the waveform continues till the end (the signals specifically starting from the time 0.08 sec till the end). The occurrence of periodic signals and the presence of the energy in the lower frequency range is the indication of the voiced sound; hence the fricative /**z**/ of Sindhi is a voiced phoneme. The peak at the fundamental frequency diminish roughly after six harmonics below 1 kHz shown in the spectrum of this sound, see figure 8.61. The peak at fundamental frequency and roughly six harmonics are common features observed in all the voiced fricatives of Sindhi. The formant transitions are only clear and stable for the first formant; the transitions for second, third and fourth formants are ambiguous and hence can not be used for the identification of this fricative sound of Sindhi see figure 8.62 (b) for the formant tracks.

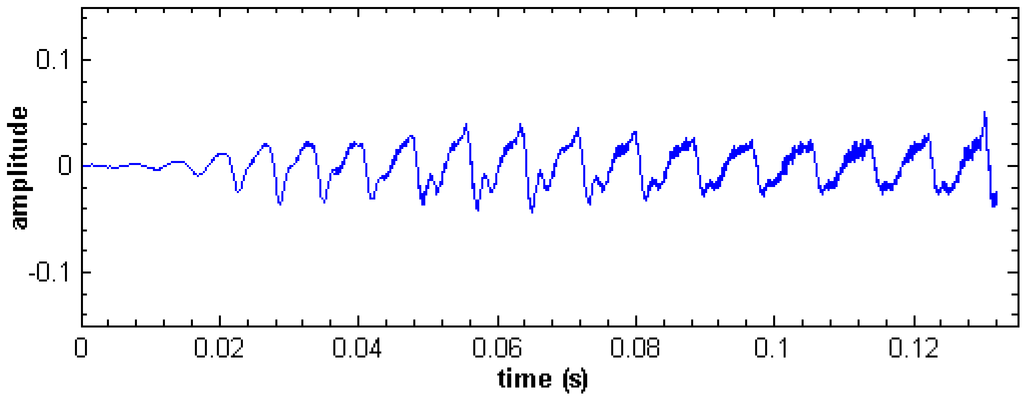


Figure 8.60 Waveform of the fricative consonant /z/ of Sindhi, taken from the word utterance /zɑri/ زاري.

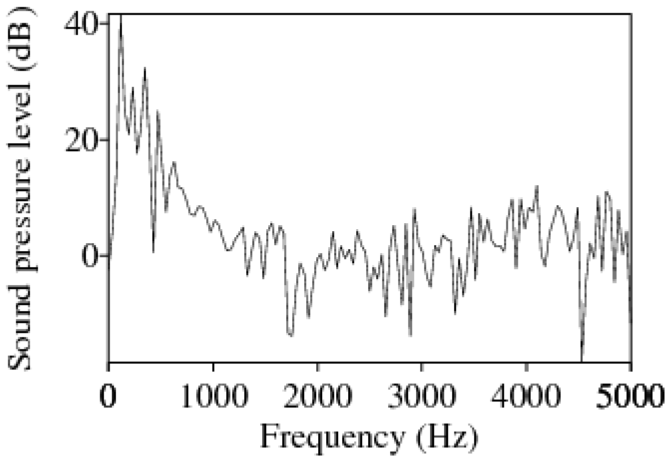


Figure 8.61 Spectrum of the fricative consonant /z/ of Sindhi shows peak at the fundamental frequency and roughly six harmonics below 1 kHz.

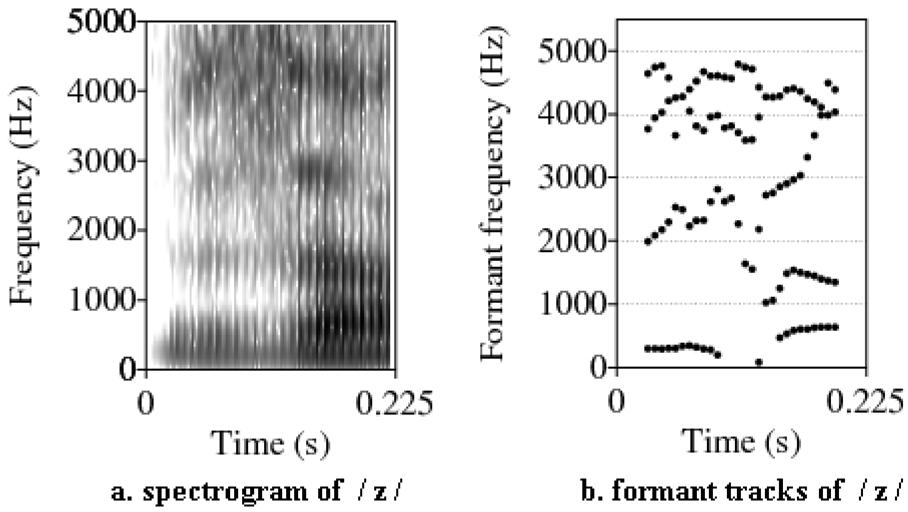


Figure 8.62 (a) Spectrogram of the fricative consonant /z/ followed by a vowel sound (b) the formant tracks.

8.5.4 Alveolar unvoiced fricative /s/ س

Figure 8.63 below shows the waveform of the fricative sound /**s**/ س of Sindhi extracted manually from the word utterance /sɑro/ سارو. The strong friction noise signals are observed at the mid of the waveform compared to the start and end of the speech wave. No periodic signal activity traced in the waveform and the absence of the energy in the lower frequency regions (energy is absent below the 1500 Hz for this sound) is the indication of the unvoiced sound; hence the fricative /**s**/ of Sindhi is an unvoiced phoneme. There is no peak at the fundamental frequency and the harmonics below 1 kHz shown in the spectrum of figure 8.64; this is a common acoustic property for the unvoiced fricatives of Sindhi. The high concentration of the energy above 4 kHz is the additional acoustic feature observed for the alveolar fricative /**s**/ of Sindhi shown in figures 8.64 and 8.65. The acoustic features for the fricatives of Sindhi may vary from speaker to speaker; however this high concentration of the energy above 4 kHz is uniformly observed among the speakers of the five dialects spoken in the Sindh province of Pakistan. The formant transitions are not clear and stable for this fricative and hence can not be used for the identification of this fricative sound of Sindhi see figure 8.65(b) for the formant tracks.

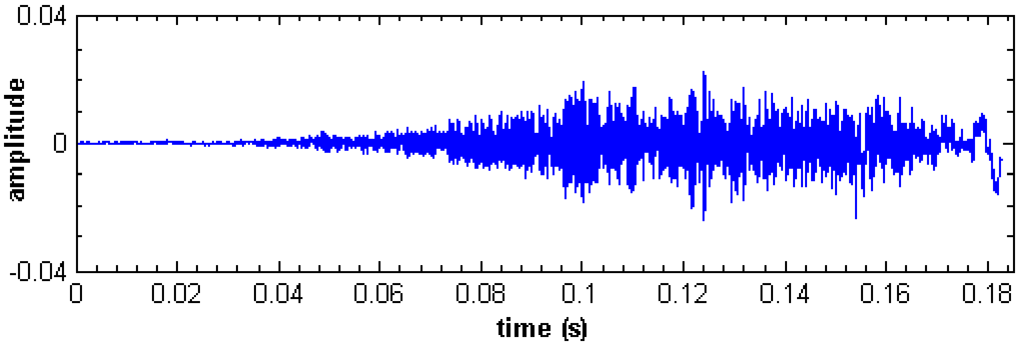


Figure 8.63 Waveform of the fricative consonant /s/ of Sindhi taken from the word utterance /sɑro/ سارو.

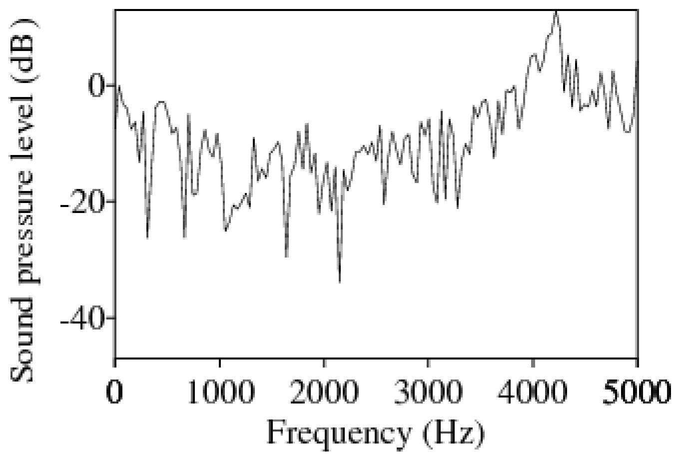


Figure 8.64 Spectrum of the fricative consonant /s/ of Sindhi shows the high concentration of energy above 4 kHz.

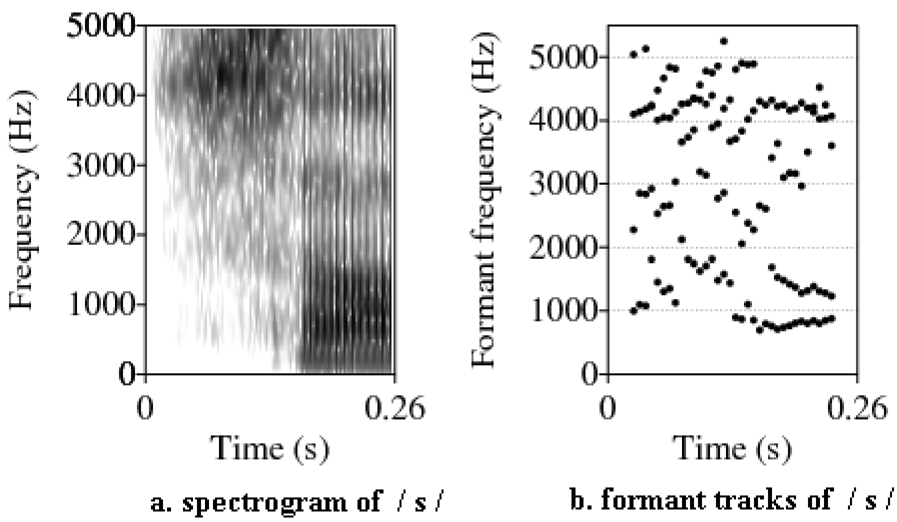


Figure 8.65 (a) Spectrogram of fricative consonant /s/ of Sindhi followed by a vowel (b) the formant tracks.

8.5.5 Palatal unvoiced fricative /ʃ/ ش

Figure 8.66 below shows the waveform of the palatal fricative /ʃ/ ش of Sindhi extracted manually from the word utterance /kɑʃʊ/ ڪاشُ. The strong friction noise signals in the waveform are only observed for this fricative of Sindhi from the start till the end of waveform. There is no periodic signal activity in the waveform of the figure 8.66 and the absence of the energy from the lower frequency range shown in the spectrogram of figure 8.68 is the indication of the unvoiced sound; hence the fricative /ʃ/ of Sindhi is an unvoiced phoneme. One of the acoustic differences between the unvoiced alveolar and unvoiced palatal fricatives of Sindhi is the high concentration of energy. The high concentration of energy for fricative /ʃ/ is observed above 1500 Hz and the high concentration of the energy for fricative /s/ is observed above 4 kHz see figures 8.64 and 8.65 for the alveolar fricative /s/ and figures 8.67 and 8.68 for the palatal fricative /ʃ/ of Sindhi. This difference between alveolar /s/ and palatal /ʃ/ is due to the level of constriction during the production of the two sounds; the vocal tract constriction for palatal fricative /ʃ/ is high compared to the alveolar /s/ of Sindhi. There is no peak at the fundamental frequency and the harmonics below the 1 kHz in the spectrum of the figure 8.67 for the palatal /ʃ/ of Sindhi. The formant transitions are ambiguous and unstable for the palatal /ʃ/ of Sindhi, hence can not be used for the identification of this sound see figure 8.68(b) for the formant tracks.

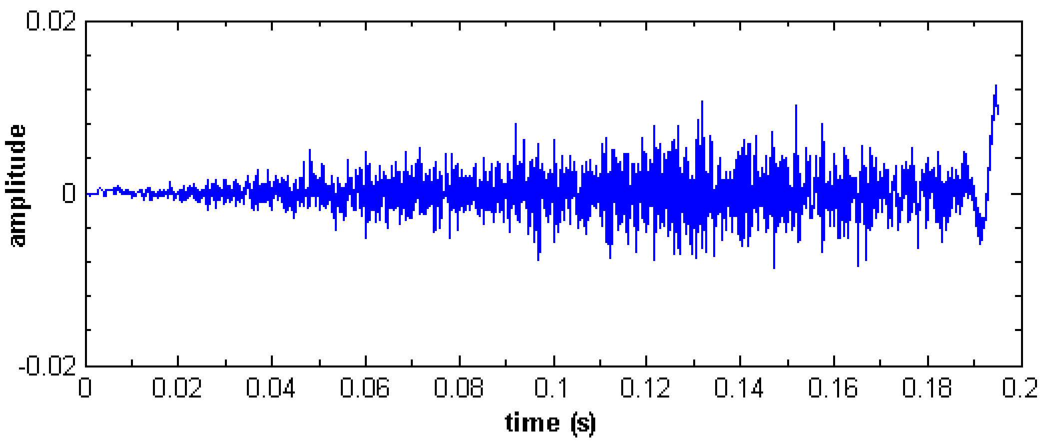


Figure 8.66 Waveform of the fricative consonant /ʃ/ of Sindhi, taken from the word utterance /kɑʃʊ/ ڪاشُ.

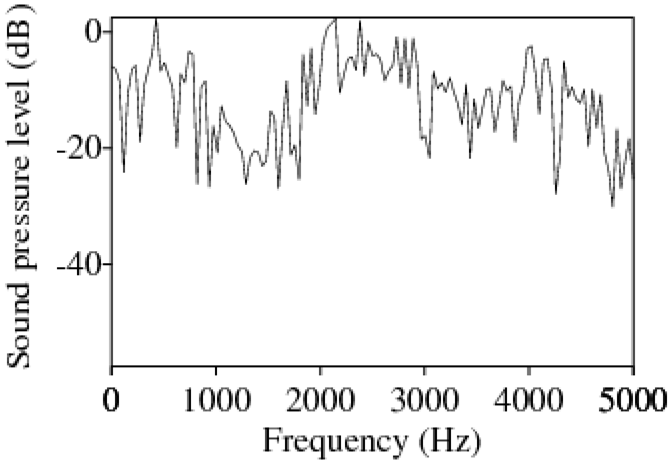
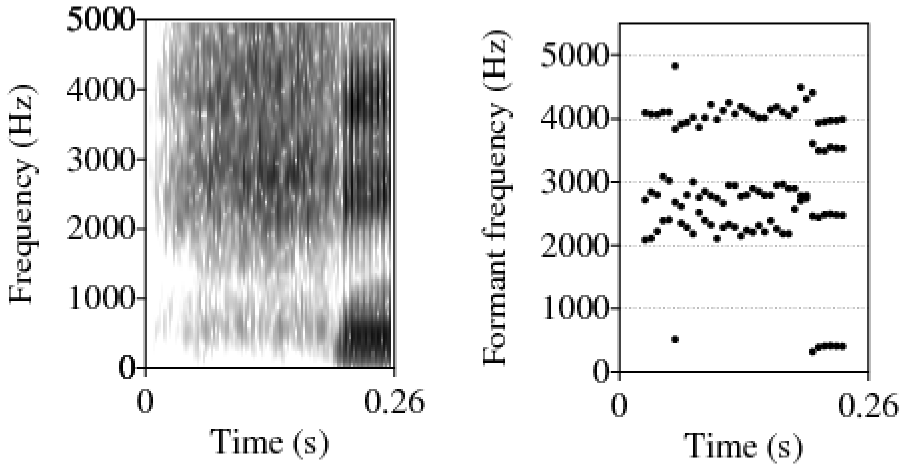


Figure 8.67 Spectrum of the fricative consonant /ʃ/ of Sindhi do not show peak at the fundamental frequency.



a. spectrogram of / ʃ / **b. formant tracks of / ʃ /**

Figure 8.68 (a) Spectrogram of the fricative consonant /ʃ/ of Sindhi followed by a vowel (b) the formant tracks.

8.5.6 Velar voiced fricative /ɣ/ غ

Figure 8.69 below shows the waveform of the velar fricative /ɣ/ غ, of Sindhi extracted manually from the word utterance /ɣaliʧo/ غاليچو. The waveform shows the friction noise signals from start till end. The prediction of the voicing feature for this fricative sound can not be judged from the waveform of this sound. However the spectrum and spectrogram of this sound shown in figures 8.70 and 8.71 are analyzed to determine the voicing feature of this sound. The spectrum figure shows the peak at the fundamental frequency and this peak diminish roughly after six harmonics and the presence of the energy in the lower frequency range shown in the wideband spectrogram enable us to safely conclude that this is a voiced phoneme of Sindhi. The fact that this is a borrowed Arabic phoneme; most of the Sindhi speakers pronounce this fricative wrong phonetically as the velar stop /g/ rather than the velar /ɣ/. The additional characteristic of the velar /ɣ/ of Sindhi observed are the two additional peaks in the spectrum (after the peak at F0) of this sound, first at around 1500 Hz and second at around 4 kHz see figure 8.70. The formant transitions of this sound are relatively unstable and ambiguous compared to the other voiced fricatives; the bilabial /v/ and alveolar /z/ of Sindhi.

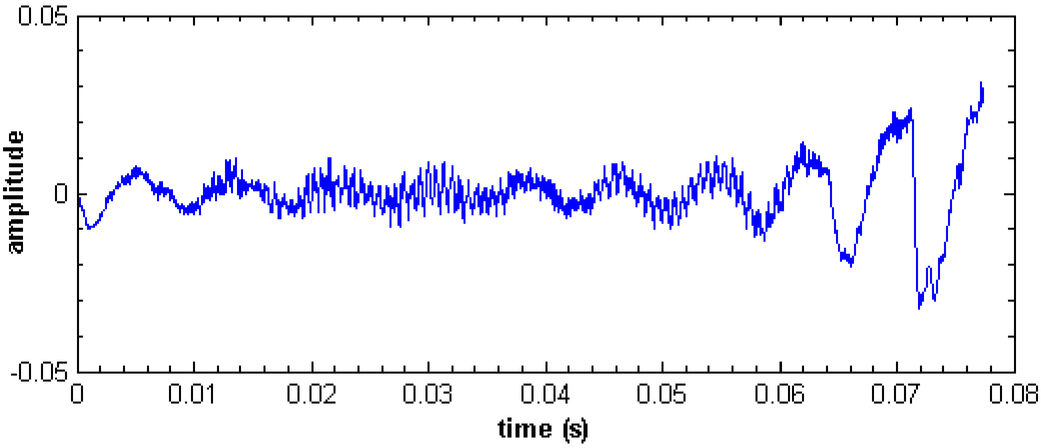


Figure 8.69 Waveform of the fricative consonant /ɣ/ غ of Sindhi, taken from the word utterance /ɣalitʃo/ غاليچو.

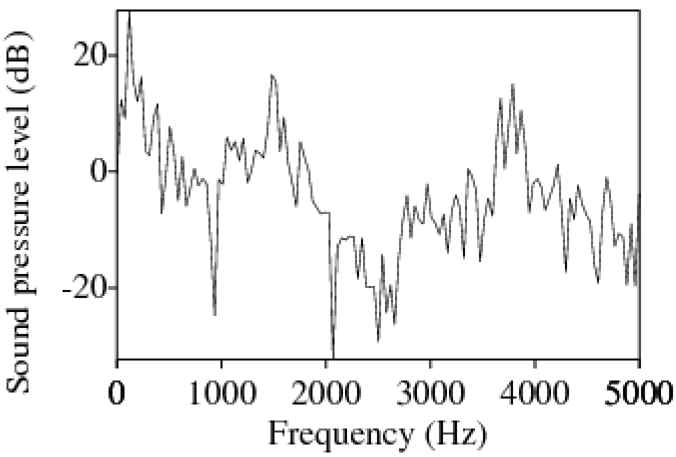


Figure 8.70 Spectrum of the fricative consonant /ɣ/ غ of Sindhi shows peak at the fundamental frequency and roughly six harmonics below 1 kHz.

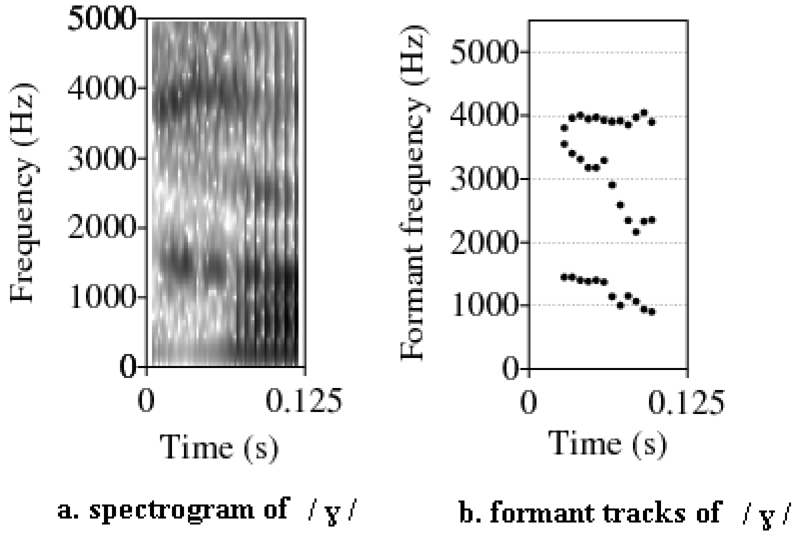


Figure 8.71 (a) Spectrogram of the fricative consonant /ɣ/ غ of Sindhi followed by a vowel (b) the formant tracks.

8.5.7 Velar unvoiced fricative /x/ خ

Figure 8.72 below shows the waveform of the fricative sound /**x**/ خ of Sindhi extracted manually from the word utterance /xəʧər/ خَچر. The waveform starts with the strong friction noise signals without periodic signal activity. The absence of the energy in the lower frequency regions and no peak at the fundamental frequency along with the harmonics below 1 kHz for this sound confirm this sound as an unvoiced phoneme of Sindhi; see spectrum and spectrogram figures 8.73 and 8.74. This unvoiced velar fricative shows a harmonic peak at around 1500 Hz, this is the additional acoustic characteristic observed for both the voiced velar /ɣ/ and unvoiced velar /x/ fricatives of Sindhi. The formant transitions for this sound are ambiguous and unstable and hence can not be used for the identification of this sound see figure 8.74(b) for the formant tracks.

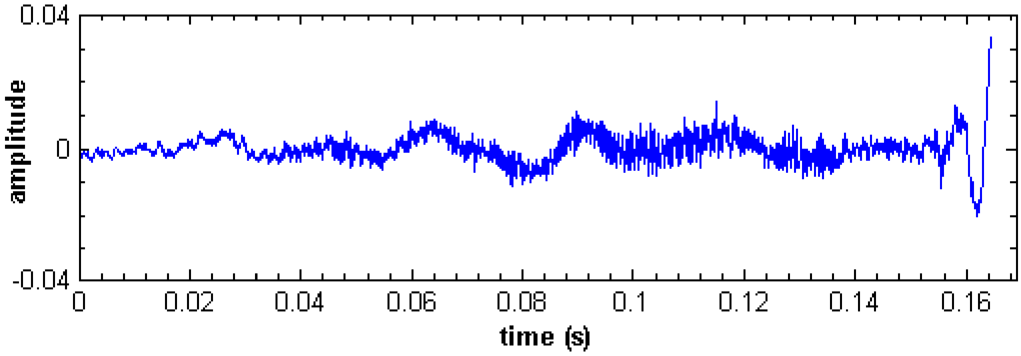


Figure 8.72 Waveform of the fricative consonant /x/ خ of Sindhi, taken from the word utterance /xətʃər/ خچر .

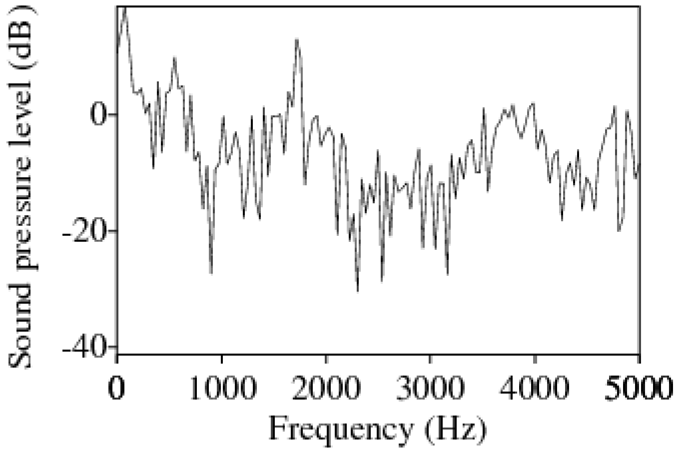


Figure 8.73 Spectrum of the fricative consonant /x/ خ of Sindhi shows peak above 1.5 kHz.

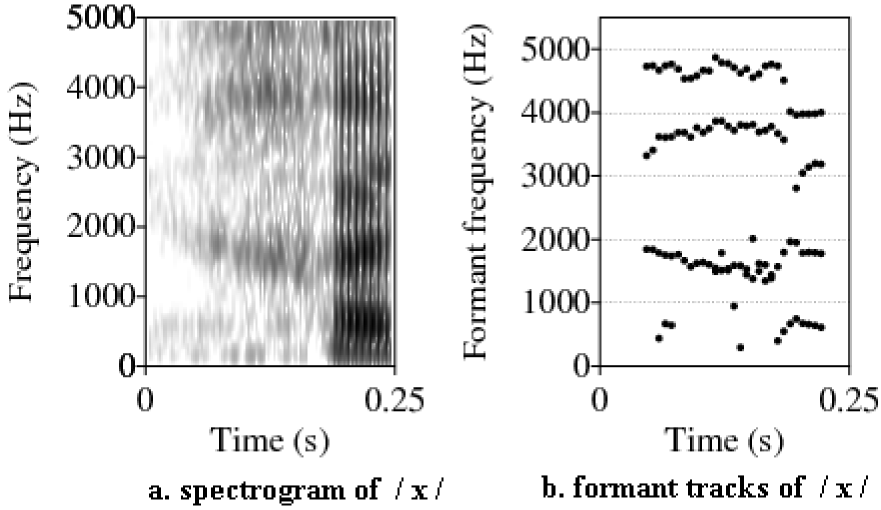


Figure 8.74 (a) Spectrogram of the fricative consonant /x/ خ of Sindhi followed by a vowel (b) the formant tracks.

8.5.8 Glottal fricative /h/ ھ

Figure 8.75 below shows the waveform of the glottal fricative /h/ ھ of Sindhi extracted manually from the word utterance /hərəɳ/ هرڻ. There is no periodic signal activity in the waveform of figure 8.75 and no peak at the fundamental frequency or the harmonic peaks below 1 kHz see the spectrum of figure 8.76, are the indications that the fricative /h/ is an unvoiced phoneme of Sindhi. All of these acoustic features support the conclusion that this is an unvoiced phoneme of Sindhi, but this conclusion leads to controversy regarding the voicing feature of this fricative sound discussed in the published literature. The glottal fricative /h/ of Sindhi when produced at word initiation or medial positions is referred to as a *voiced* phoneme and *unvoiced* phoneme when it appears at word final positions (Jatoi, 1996). The recordings for this fricative sound were made by changing the position and the sequence of phonemes as word initial, medial and word final position. The periodic signal activity and the presence of energy in the lower frequency range were not observed when ten instances of this sound were analyzed by changing the sequence and position of this sound within the word utterances. The peak at the fundamental frequency in the spectrum and roughly six harmonics below 1 kHz observed for all the voiced fricatives /v/, /z/, and /ɣ/ of Sindhi are not observed for any instance of this glottal fricative /h/ of Sindhi. The formant transitions are ambiguous and unstable for this glottal fricative /h/ and can not be used for the identification of this sound see figure 8.77(b) the formant tracks. Therefore in this study we refer to this glottal fricative /h/ of Sindhi as an unvoiced phoneme regardless of its appearance (position) in the word utterance.

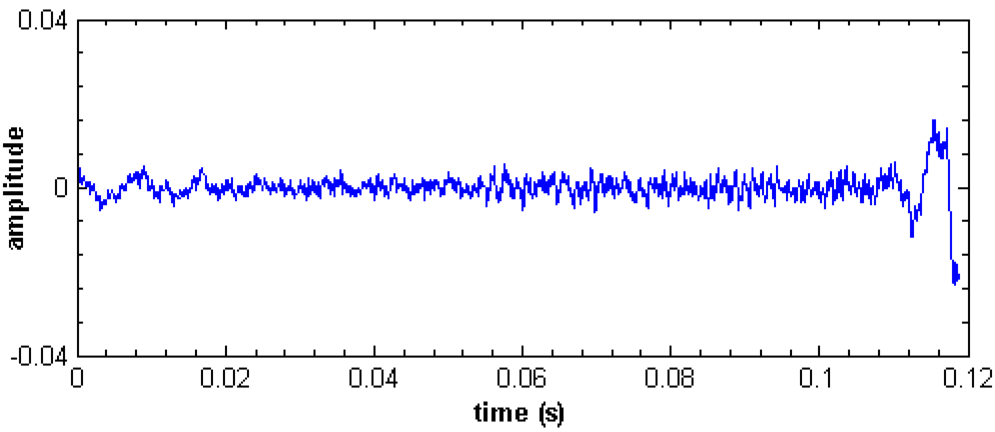


Figure 8.75 Waveform of the fricative consonant /h/ ھ of Sindhi, taken from the word utterance /hərəɳ/ هرڻ.

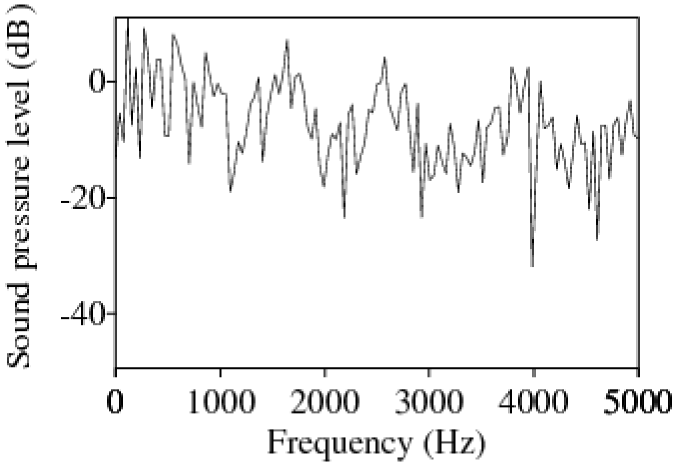


Figure 8.76 Spectrogram of the fricative consonant /h/ ھ of Sindhi do not show peak at the fundamental frequency.

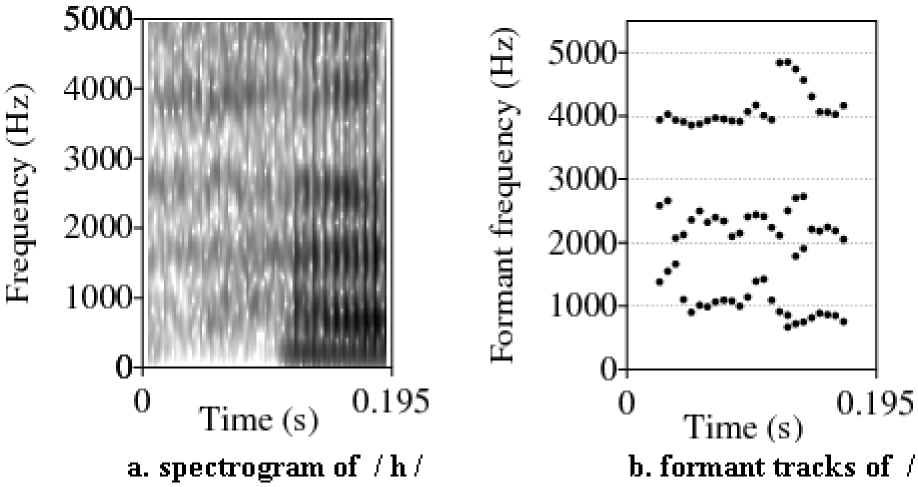


Figure 8.77 (a) Spectrogram of the fricative consonant /h/ ھ of Sindhi followed by a vowel (b) the formant tracks.

Table 8.4 Mean and standard deviation of the acoustic parameters: F0, first four formants and duration of the fricative consonants of Sindhi.

Fricative	parameter	Duration	F0	F1	F2	F3	F4	Intensity (dB)
/v/ و	Mean	0.187	147	310	1634	2708	3956	41
	STD	0.026211	13	94	355	651	358	7
/f/ ف	Mean	0.1705	5958	-	-	-	-	39
	STD	0.014978	97	-	-	-	-	2
/z/ ز	Mean	0.132667	143	245	1709	3344	4270	55
	STD	0.007095	14	42	8	297	137	1

-Continued on next page -

/s/ س	Mean	0.15625	4264	-	-	-	-	38
	STD	0.008539	188	-	-	-	-	5
/ɣ/ غ	Mean	0.123333	128	1104	2405	3559	4349	49
	STD	0.020207	5	440	844	548	494	5
/x/ خ	Mean	0.178333	1818	-	-	-	-	46
	STD	0.003512	18	-	-	-	-	4
/ʃ/ ش	Mean	0.22375	3440	-	-	-	-	33
	STD	0.02227	53	-	-	-	-	4
/ h / ه	Mean	0.102667	-	-	-	-	-	38
	STD	0.008622	-	-	-	-	-	2

8.6 Liquid consonants of Sindhi

The lateral class of sound /**l**/ ل, and a trill class of sound /**r**/ ر, in Sindhi are jointly referred to as the liquid consonants of Sindhi. These sounds are produced with a partial constriction in the vocal tract; therefore they are phonetically referred to as frictionless speech sounds (Olive, 1993). The two liquids of Sindhi are articulated at one place of articulation referred to as the alveolars. When the tip of the tongue touches the alveolar ridge in such a way (centred) so that the air passes through the two sides in mouth in this way the lateral sound /**l**/ of Sindhi is produced. When the tip of the tongue loosely touches the alveolar ridge, allowing the air stream to flow over the middle of the tongue rather than the sides; the air passage in this way causes vibration at the tip of the tongue and produces the trill sound /**r**/ of Sindhi. Two voicing characteristics are associated with liquids of Sindhi, these are: (i) if they appear at word medial or initial position, they are considered voiced phonemes and (ii) unvoiced if they appear at word final position. In this section we will present acoustic-phonetic analysis of the liquids of Sindhi. The analysis results are mainly based on the speech samples collected during the field study discussed in chapter III.

8.6.1 Alveolar trill /r/ ر

Figure 8.78 below shows four waveforms of the trill sound /r/ ر, of Sindhi extracted manually from the isolated word utterances and the words embedded in a carrier sentence read by native Sindhi speakers. The four waveforms are shown for trill /r/ because it varies acoustically when produced within isolated word utterances and when produced as a phrase medially. This sound also showed varying acoustic features when followed by a short or long vowel in the utterance. These differences are highlighted in the four waveforms of this sound: figure 8.78(a) is the waveform extracted from the word utterance /dərə/ در embedded in a carrier sentence read by speakers.

The waveform shows the slow varying periodic signals for a very short duration with relatively small amplitude of the modulation signal for the segment of the trill sound /r/ in the word utterance. This difference of the sharp drop in the amplitude of the signals and the presence of the weak energy for a very short duration can be seen in the waveform and spectrogram of figure 8.79 for the word utterance /dərə/ در. The speech signals of /r/ sound do not repeat when the vocal tract configurations are to be set for the production of the next sound in a sentence; this is shown in the waveform of the figure 8.78(a) and 8.79. Figure 8.78(b) shows another variation of this sound when produced within the isolated utterance of the word /sərə/ سَر of Sindhi. When the vocal tract configurations are not required to be set for the production of the next sound and hence the vocal tract configurations are only set and end with the trill sound /r/; the repeated occurrence of this sound signals are observed in the speech waveforms this is shown in the figures 8.78(b) and 8.80. The mean duration for the trill sound /r/ in the utterances where speech signals for this sound do not repeat is 0.0256 seconds, whereas the mean duration in the utterance where /r/ sound signals are repeated is 0.018sec for the first occurrence and 0.015sec for the second occurrence of the /r/ sound signals.

The second occurrence is uniformly observed smaller in duration compared to the duration of the first occurrence of signals for this sound. The waveform shown in figure 8.78(c) which shows another variation of this sound, the three consecutive occurrences of the /r/ sound signals when produced within the word utterance /dərʊ/ دُر in isolation.

The three repetitions of this sound are very rarely observed; only two out of ten speakers have produced such /r/ sound where three occurrences are observed this is shown in the waveforms of figures 8.78(c) and 8.81 for the word utterance /dərʊ/ دَرُ. However the two occurrences of the /r/ sound signals were observed among the utterances of seven speakers out of ten.

The analysis results show that when the trill sound /r/ is followed by a short vowel and the word utterance ends with this combination (the /r/ sound followed by a short vowel) then the repetitive occurrence of the /r/ sound signals can be seen in the speech wave. This is due to the fact that the tip of the tongue continues vibrating until the articulators are relaxed to the normal position; because the articulators are not required to move for the production of the next sound. Figure 8.78(d) shows the waveform of the /r/ sound followed by a long vowel in the word utterance /sɑro/ سارو of Sindhi. The trill sound /r/ if followed by a long vowel then the repetitive occurrence of the /r/ sound signals do not happen in either case; whether this sound is produced within the isolated word utterance or it is produced within a carrier sentence this is shown in the waveforms of figures 8.78(d) and 8.82.

Another difference observed for the trill sound /r/ compared to the other consonants of Sindhi is that it does not exhibit formant discontinuities at the vowel junctures (the formant transitions coming into the trill sound /r/ from the preceding sound and going away from it to the following sound); rather the transitions are more continuous at the vowel junctures for trill sound compared to the highly constricted consonant phonemes such as fricatives and stops (Olive, 1993; Zhou, 2008).

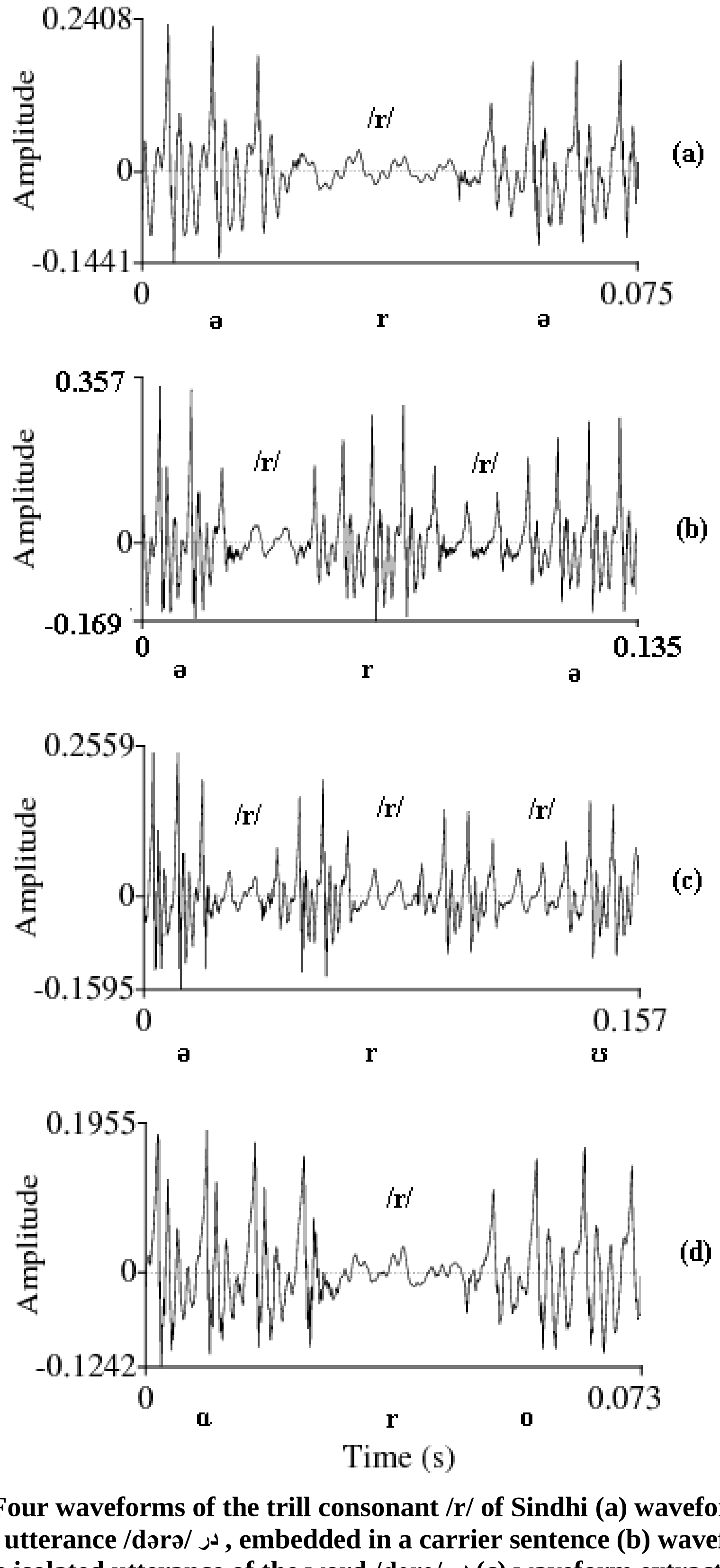


Figure 8.78 Four waveforms of the trill consonant /r/ of Sindhi (a) waveform extracted from the word utterance /dərə/ در , embedded in a carrier sentence (b) waveform extracted from the the isolated utterance of the word /dərə/ در (c) waveform extracted from the isolated utterance of the word /dərʊ/ دَرُ (d) waveform extracted from the word utterance /sɑro/ سارو , embedded in a carrier sentence.

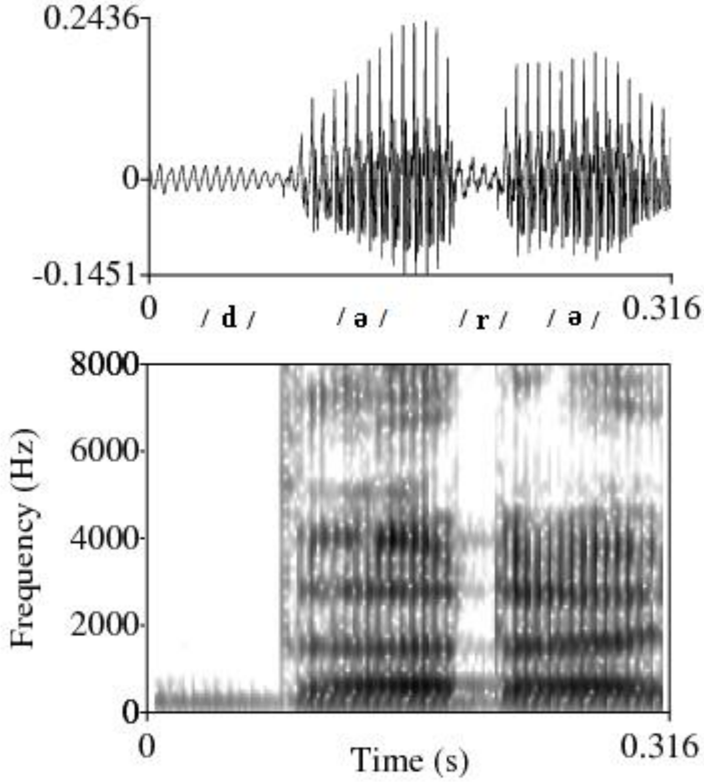


Figure 8.79 Waveform and the spectrogram of the word utterance /dərə/ در , embedded in a carrier sentence.

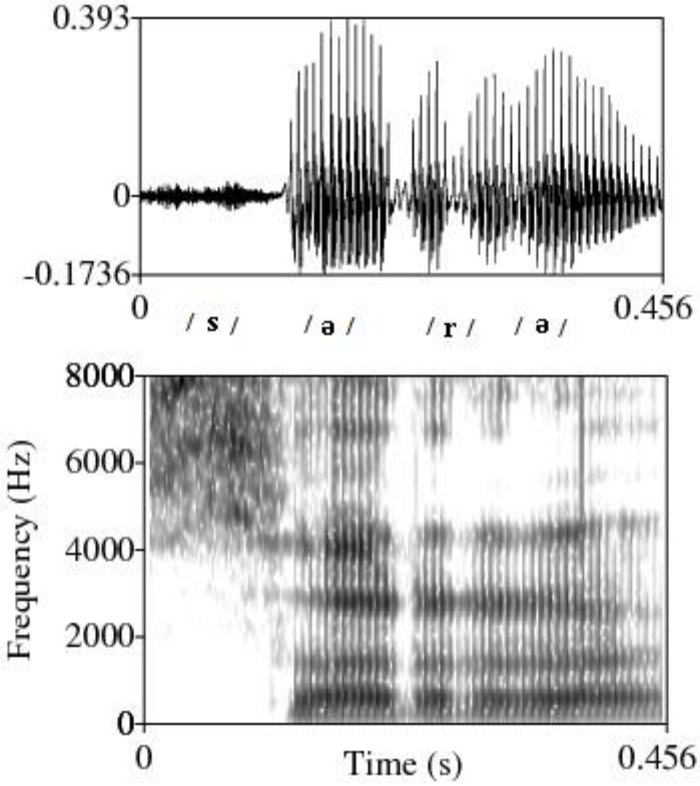


Figure 8.80 Waveform and the spectrogram of the word utterane /sərə/ سَرَ, of Sindhi pronounced in isolation.

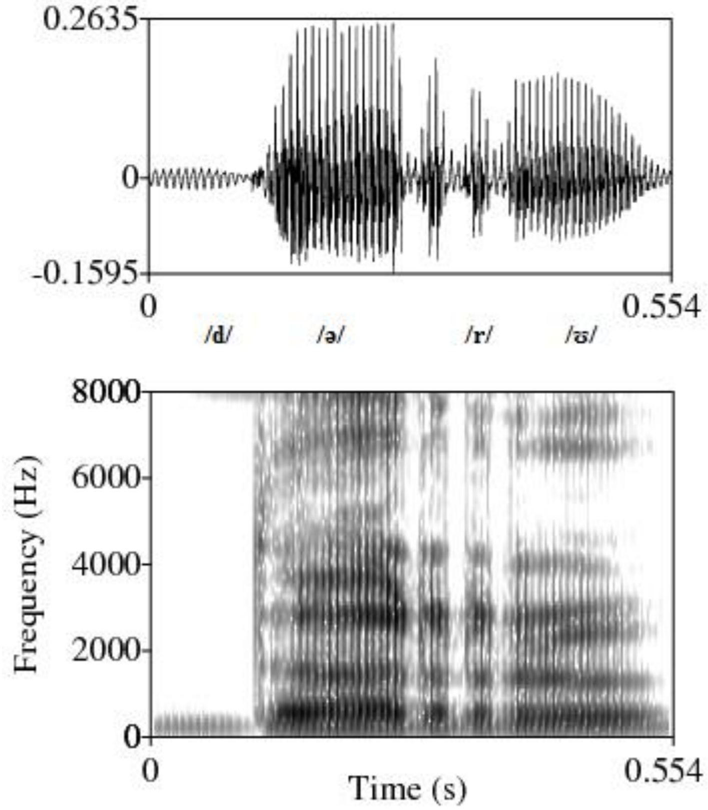


Figure 8.81 Waveform and the spectrogram of the word /dərʊ/ دَرُ of Sindhi pronounced in isolation.

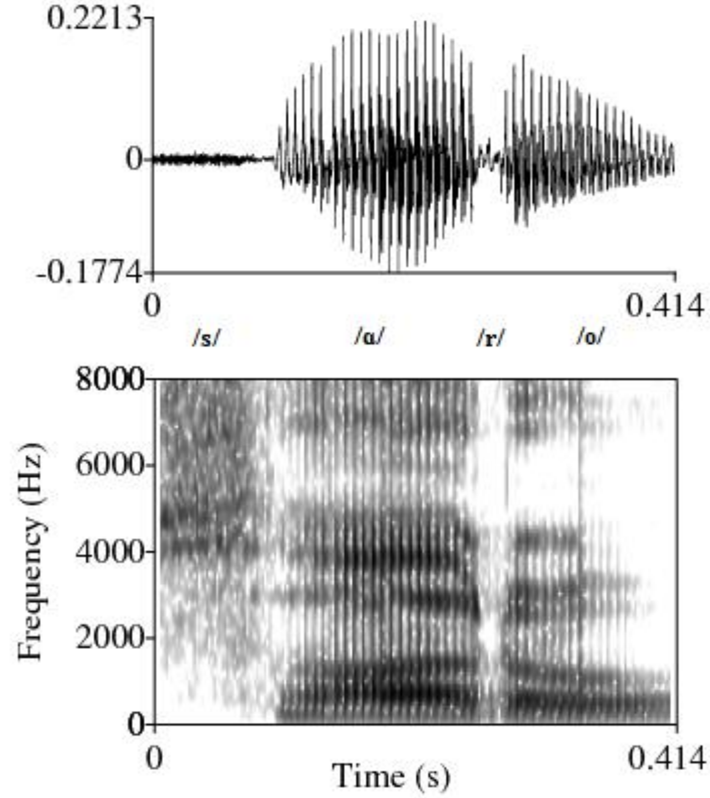


Figure 8.82 The waveform and the spectrogram of the word /sɑro/ سارو of Sindhi embedded in a carrier sentence.

8.6.2 Alveolar lateral /l/ ل

Figure 8.83 below shows the waveform and spectrogram of the lateral phoneme /l/ of Sindhi present in the word utterance /kələ/ ڪَلَ. The lateral sound /l/ shows a vowel like waveform and the presence of energy across wideband spectrogram including the energy in the lower frequency range usually seen for vowels. The slow varying periodic signal activity at the word medial position in the utterance of word /kələ/ ڪَلَ, confirm this sound a voiced phoneme of Sindhi. The liquid /l/ has a formant configuration with a low F1; therefore the F1 motion for this sound is downward (the transitions coming into the lateral /l/ from the preceding sound) and the upward motion (for the transitions going away from it to the following sound). The lateral consonant /l/ has shown stable both F1 and F2 values and can be used for the identification of this sound. One special acoustic property observed for the liquids of Sindhi is their occurrence at word final position make them voiceless sounds as the trill sound /r/ ر , at the final position in the word /sətər/ سَتر, and the lateral sound /l/ ل , at the final position in the word /ədəl/ عَدل. The mean acoustic parameter values for the two liquids of Sindhi are given in Table 8.5.

Table 8.5 Mean and standard deviation of the acoustic parameters: F0, first four formants and duration of the Liquid consonants of Sindhi.

Liquid Consonants	Parameter	Duration	F0	F1	F2	F3	F4	Intensity (dB)
/l/ ل	Mean	0.091	130	350	1550	2952	3785	63
	STD	0.055	0.72	5	33	52	90	--
/r/ ر	Mean	0.0256	149	495	1584	2841	3744	64
	STD	0.046	7	21	138	61	206	--

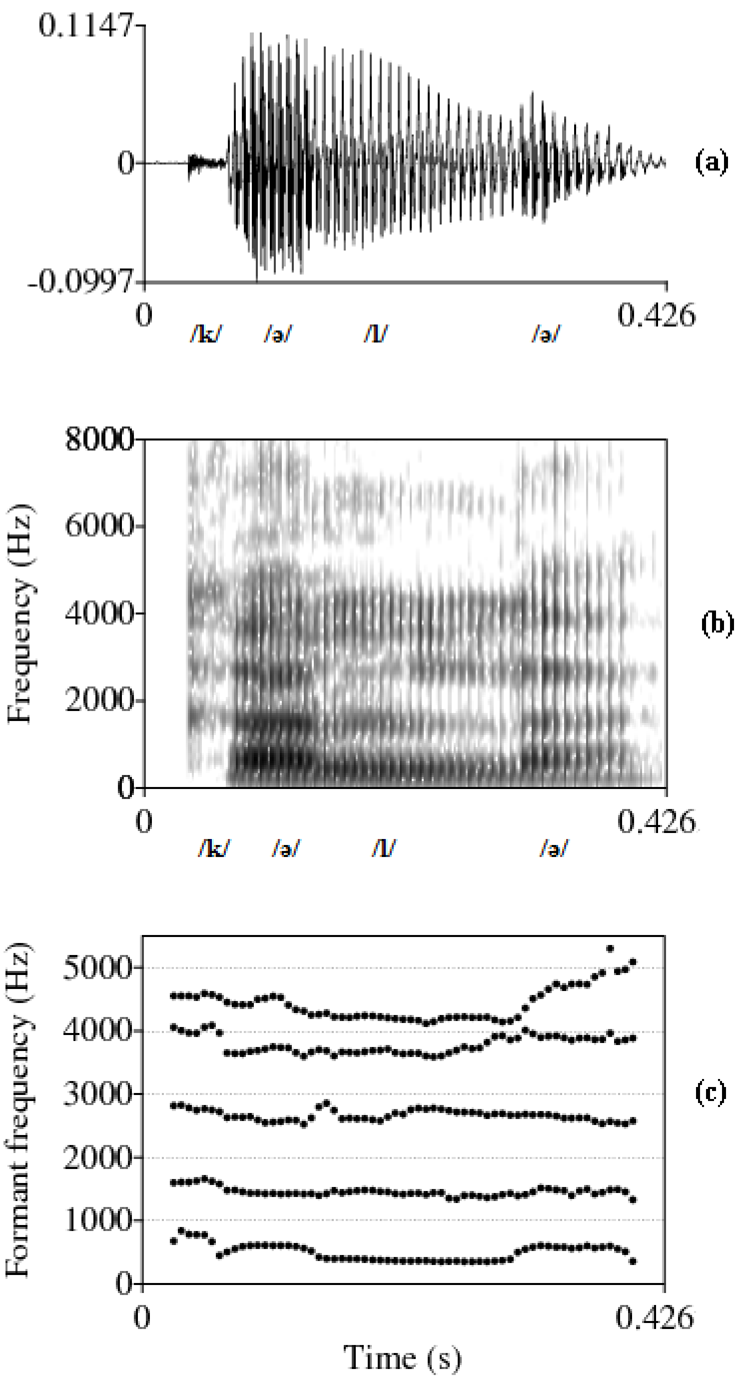


Figure 8.83 (a) Waveform of the word utterance /kələ/ کَلَ, containing the lateral phoneme /l/ of Sindhi, (b) the spectrogram and (c) the formant tracks.

8.7 Retroflex flap consonant /ɽ/ ڙ

Flaps are produced with the front of the tongue curled up in the direction of the hard palate. The flap /ɽ/ ڙ, of Sindhi is articulated with one moving articulator; the tongue that moves against the hard palate. Figure 8.84(a) shows the waveform of the retroflex flap /ɽ/ ڙ of Sindhi present in the word utterance /pəɽʊ/ پڙُ at the word medial position. The slow varying periodic signal activity and the presence of the energy in the lower frequency regions shown in the spectrogram, figure 8.84(b), indicate that this is a voiced phoneme of Sindhi. The F1 for retroflex flap /ɽ/ is relatively stable and have strong energy compared to the higher formants. Due to the low F1 the formant motion for this sound can be defined as the downward (coming into the retroflex flap /ɽ/) and upward (going away from it). The F3 showed discontinuities having the downward motion with the frequency value below 2 kHz. The Table 8.6 below shows the mean acoustic parameter values and standard deviation for the retroflex flap /ɽ/ ڙ of Sindhi.

Table 8.6 Mean and standard deviation of the acoustic parameters: F0, first four formants and duration of the flap /ɽ/ ڙ of Sindhi.

Retroflex consonant	Parameter	Duration	F0	F1	F2	F3	F4	Intensity (dB)
/ɽ/ ڙ	Mean	0.038	160	291	1650	1978	3577	64
	STD	0.5	7	20	170	115	101	--

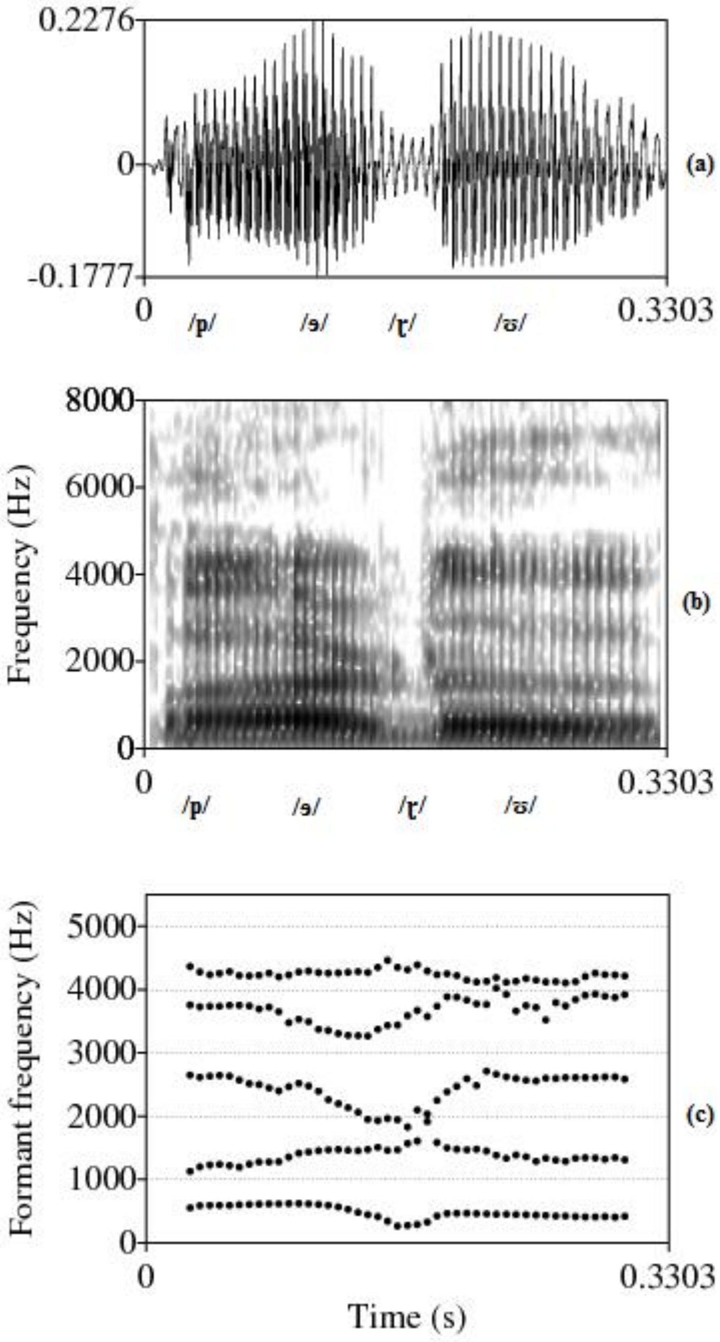


Figure 8.84 (a) Waveform (b) the spectrogram (c) the formant tracks, of the word utterance /pɘɽʊ/ پڙُ containing the retroflex consonant /ɽ/ of Sindhi.

8.8 Sounds that are not yet part of the language alphabet

Figure 8.85(a) and (b) show the acoustic waveform and the spectrogram of the word utterance /sʊm^{h}e/ containing the bilabial nasal /m^h/ of Sindhi. The damped periodic waveform signals between the points ***a*** and ***b*** in figure 8.85(a) indicate that this is a voiced phoneme of Sindhi. Another significant acoustic characteristic shown in the waveform is the presence of the aspirated signals showed between point's ***b*** and ***c*** in figure 8.85(a). The formants motion for the sound /m^h/ is difficult to determine because of the aspirated signals the formants are ambiguous and unclear; however due to the low F1 (common in all the nasal sounds of Sindhi) the formant motion is downward (coming into the sound) and upward (going away from it) this is showed in spectrogram in figure 8.85(b). The formant discontinuities are observed for the segment of the aspirated signals (signals between point's ***b*** and ***c***) shown in figure 8.85(b), the formant tracks. The sound /m^h/ is not yet present in the language alphabet; however the phonemic importance of this sound has been proven by finding the minimal pair of words. This sound is frequently spoken by the speakers of the five dialects of Sindhi; therefore in this study this sound is considered as a meanigfull phoneme and classified as the aspirated voiced bilabial nasal phoneme of Sindhi. The classification of the consonant sounds of Sindhi is given in appendix B.

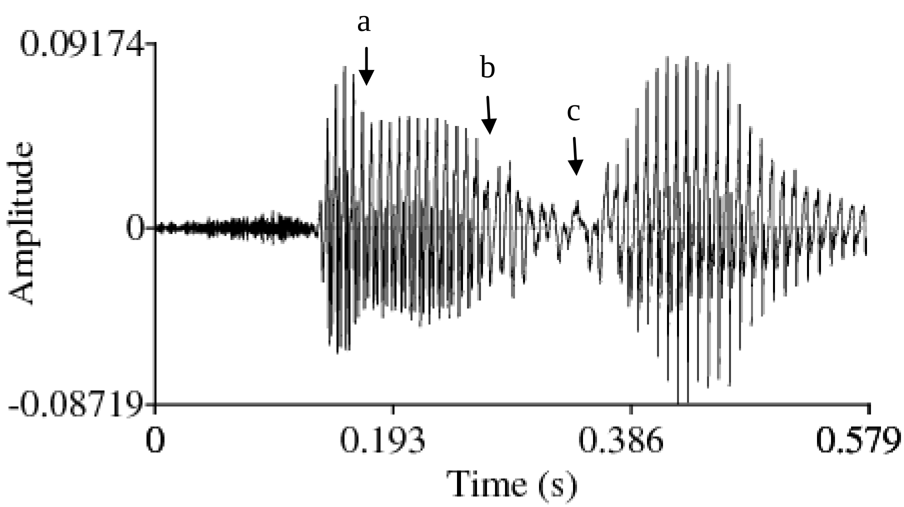


Figure 8.85(a) Acoustic waveform of the aspirated bilabial nasal /m^h/ of Sindhi, taken from the word utterance /sʊm^{h}e/.

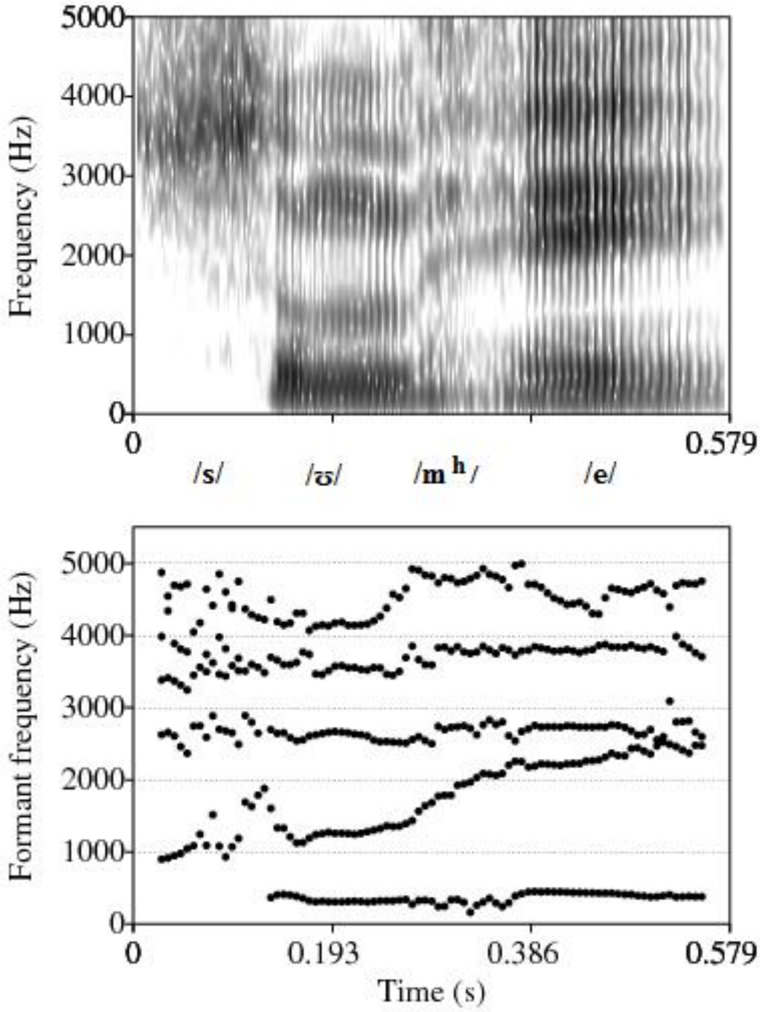


Figure 8.85(b) Spectrogram of the word utterance /sʊm^{h}e/ containing the aspirated bilabial aspirated nasal /m^{h}/ of Sindhi.

Figure 8.86(a) and (b) show the acoustic waveform and the spectrogram of the word utterance /sən^{h}ɑ/ containing the alveolar nasal /n^{h}/ of Sindhi. The damped periodic waveform signals between the points ***a*** and ***b*** in the figure 8.86(a) indicate that this is a voiced phoneme of Sindhi. The signals between point's ***b*** and ***c*** indicate the presence of aspirated signals shown in the waveform. The formants motion for the sound /n^{h}/ is difficult to determine because of the aspirated signals at the end of the sound except for the first formant. Due to the low F1 the formant motion is downward (coming into the sound) and upward (going away from it) this is shown in spectrogram in figure 8.86(b). The formant discontinuities are observed for the segment of aspirated signals (signals between point's ***b*** and ***c***) shown in figure 8.86(a) and (b), the formant tracks. This sound is also not yet part of the language alphabet; however the phonemic importance of this sound has been proven by finding the minimal pair of words. This sound is frequently spoken by the speakers of the five dialects of Sindhi; therefore in this study this sound is considered as a meanigfull phoneme and is classified as the aspirated voiced alveolar nasal phoneme of Sindhi.

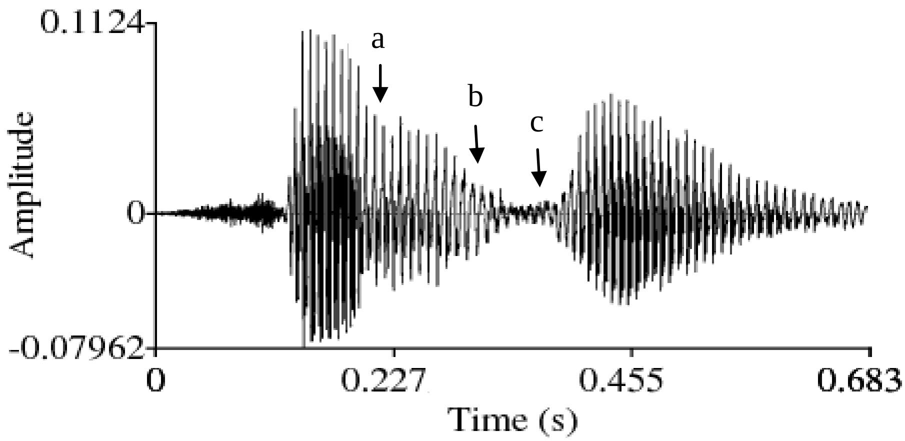


Figure 8.86(a) acoustic waveform of the aspirated alveolar nasal /n^h/ of Sindhi, taken from the word utterance /sən^hɑ/.

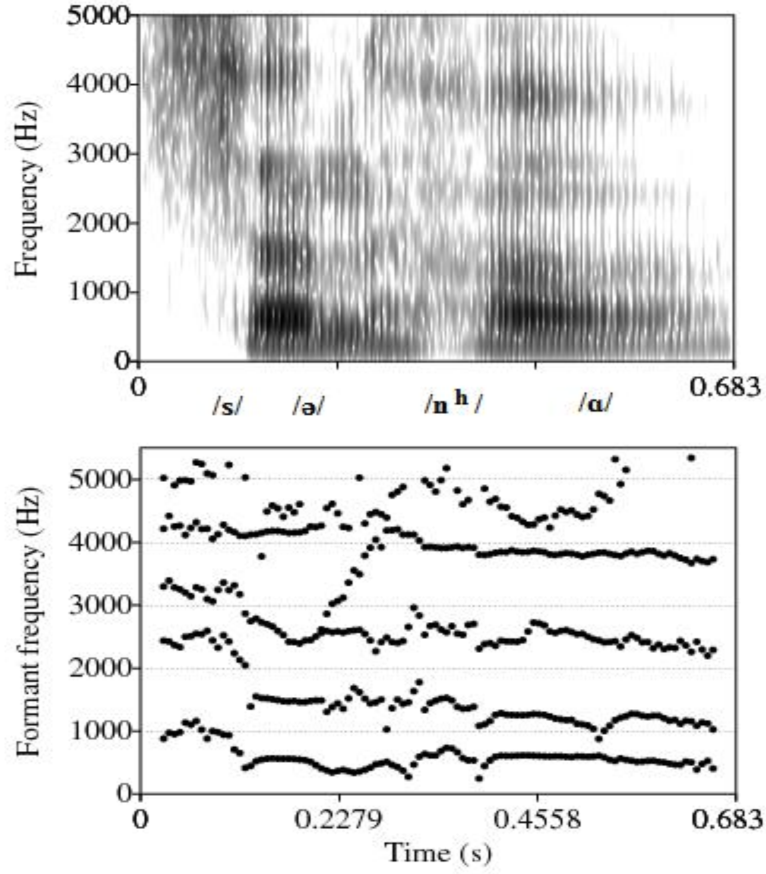


Figure 8.86(b) spectrogram of the word utterance /sən^hɑ/ containing the aspirated alveolar nasal /n^h/ of Sindhi.

Figure 8.87(a) and (b) show the acoustic waveform and the spectrogram of the word utterance **/mɑɳ^{h}e/** containing the retroflex nasal /**ɳ^h**/ of Sindhi. The acoustic waveform shown in figure 8.87(a) have shown the periodic signal activity and the aspirated signals between the points ***a*** and ***b***. It is difficult to indicate the separation mark between the aspirated signals and the signals part of the nasal sound. However the spectrogram figure shown in figure 8.87(b) is analzsed to conclude the classification of this sound, the classification of this sound is given in appendix B. This sound has shown formant transitions for a very short duration and the downward motion for the first and third formants. The third formant transitions are observed below 2 kHz, which is observed for the retroflex nasal /ɳ/ as well discussed in section 8.4.4. The periodic signal activity and the stable F1 value indicate that this is a voiced phoneme of Sindhi. The presence of aspirated signals leads to classify this sound as aspirated voiced phoneme of Sindhi. This sound is also not yet part of the language alphabet; however the phonemic importance of this sound has been proven by finding the minimal pair of words and this sound is frequently spoken by the speakers of the five dialects of Sindhi. Therefore in this study this sound is considered as a meanigfull phoneme and is classified as the aspirated voiced retroflex nasal consonant of Sindhi.

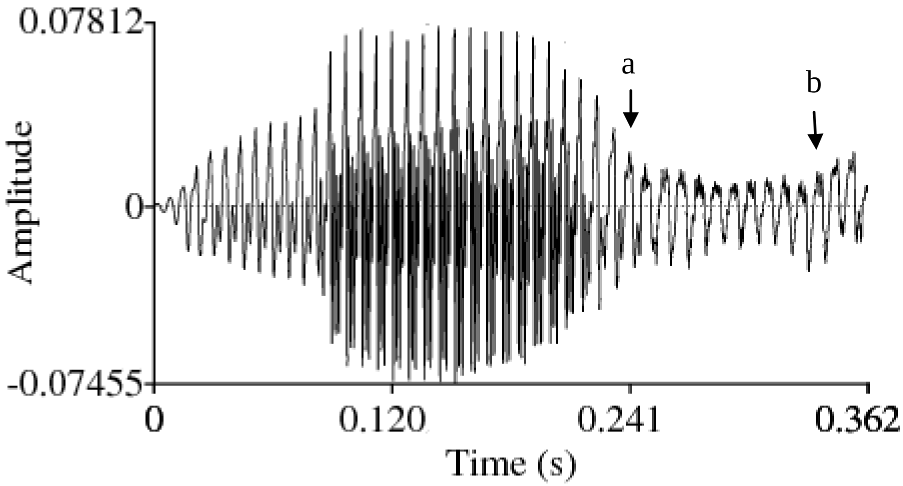


Figure 8.87(a) acoustic waveform of the retroflex nasal /ɳ^h/ of Sindhi, taken from the word utterance /mɑɳ^{h}e/.

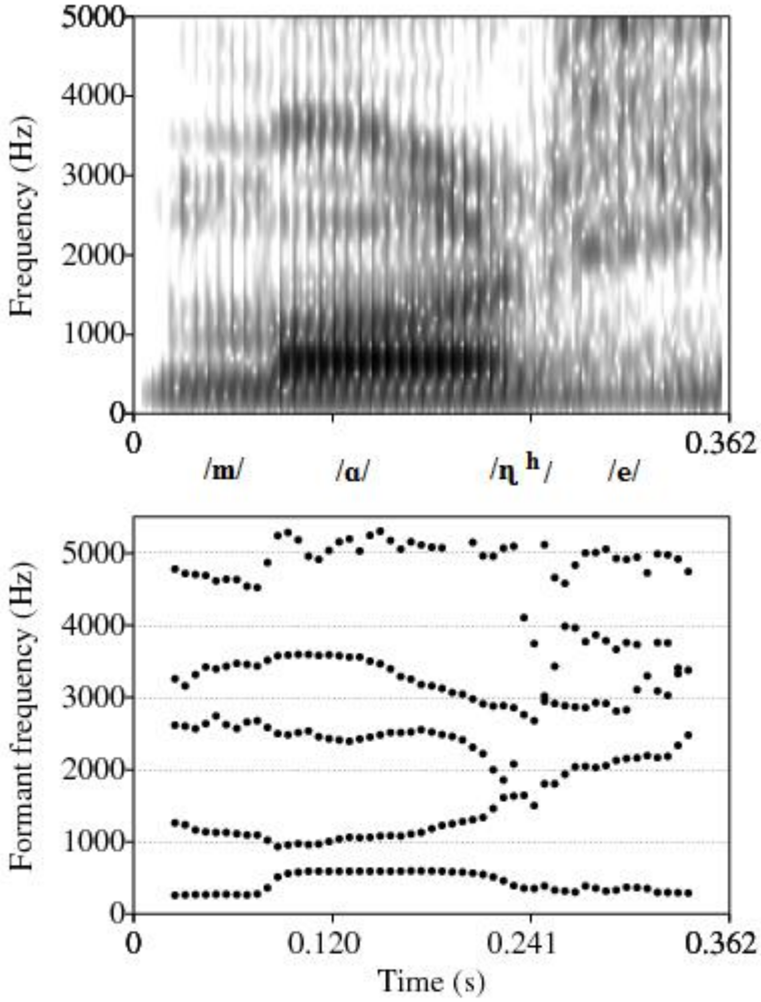


Figure 8.87(b) spectrogram of the word utterance /mɑɳ^{h}e/ containing the aspirated retroflex nasal /ɳ^{h}/ of Sindhi.

Figure 8.88(a) and (b) show the acoustic waveform and the spectrogram of the word utterance **/sɑr^{h}ə/** containing the alveolar trill /r^{h}/ of Sindhi. The short duration periodic signals activity is observed in the waveform shown between the points ***a*** and ***b*** in the figure 8.88(a). The occurrence of weak signals for a very short duration and the continuity in the formant motion (that is the formant motion from the previous sonorant (vowel) sound does not change) indicate that this is a trill class of phoneme. The periodic signal activity and the presence of energy in the lower frequency range indicate that this is a voiced phoneme of Sindhi. The occurrence of the aspirated signals shown between the point’s ***b*** and ***c*** indicate that this is also an aspirated voiced phoneme of Sindhi. This sound is also not yet part of the language alphabet; however the phonemic importance of this sound has been proven by finding the minimal pair of words and this sound is frequently spoken by the speakers of the five dialects of Sindhi. Therefore in this study this sound is considered as a meanigfull phoneme and is classified as the aspirated voiced alveolar trill consonant of Sindhi.

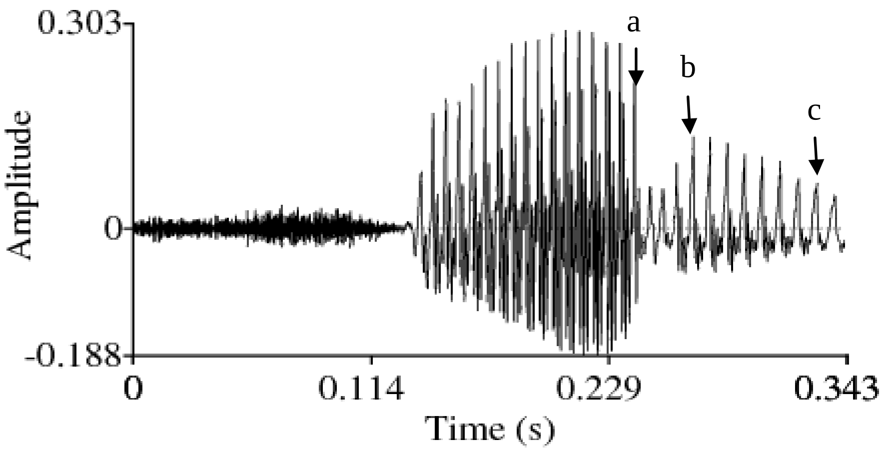


Figure 8.88(a) acoustic waveform of the alveolar tril /r^h/ of Sindhi, taken from the word utterance /sɑr^{h}ə/.

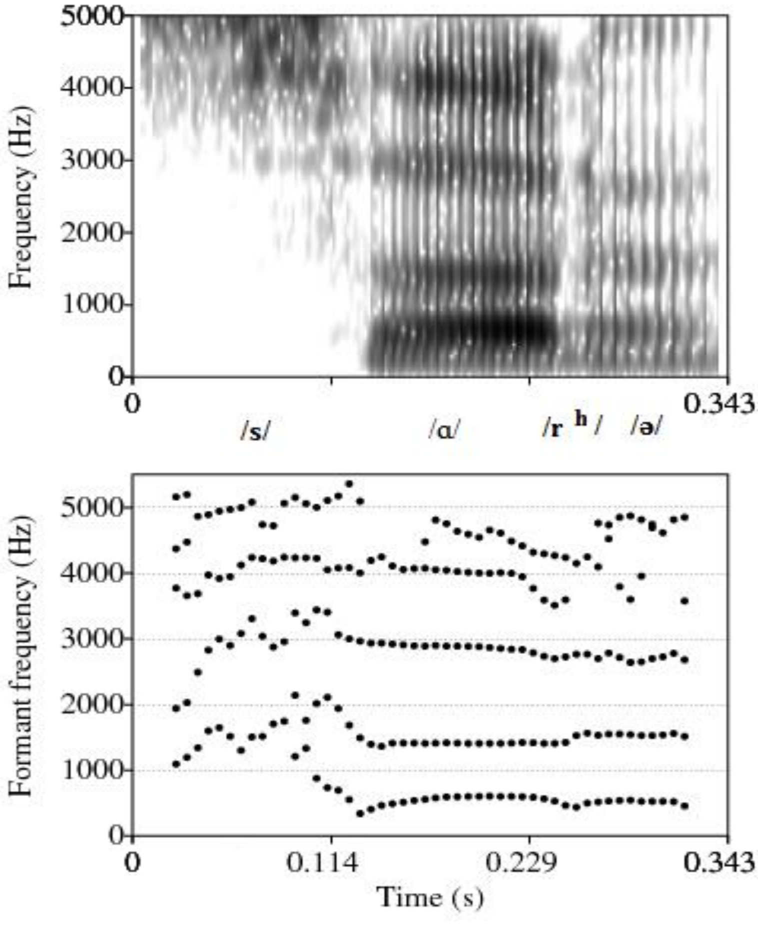


Figure 8.88(b) spectrogram of the word utterance /sɑrhə/ containing the aspirated alveolar trill /r^h/ of Sindhi.

Figure 8.89(a) and (b) show the acoustic waveform and the spectrogram of the word utterance /kəl^h/ containing the alveolar lateral /l^h/ of Sindhi. The periodic signal activity shown between the point's ***a*** and ***b*** in figure 8.89(a) indicate that this is a voiced phoneme of Sindhi. The signals between point's ***b*** and ***c*** indicate the presence of aspirated signals in the waveform. The sound has shown stable formant transitions, which are common for the lateral class of consonants; however the formant discontinuities are observed for the segment of the aspirated signals shown in the spectrogram in figure 8.89(b). Due to the low F1 the formant motion is downward (coming into the sound). This sound is also not yet part of the language alphabet; however the phonemic importance of this sound has been proven by finding the minimal pair of words and this sound is frequently spoken by the speakers of the five dialects of Sindhi. Therefore in this study this sound is considered as a meanigfull phoneme and is classified as the aspirated voiced alveolar lateral consonant of Sindhi.

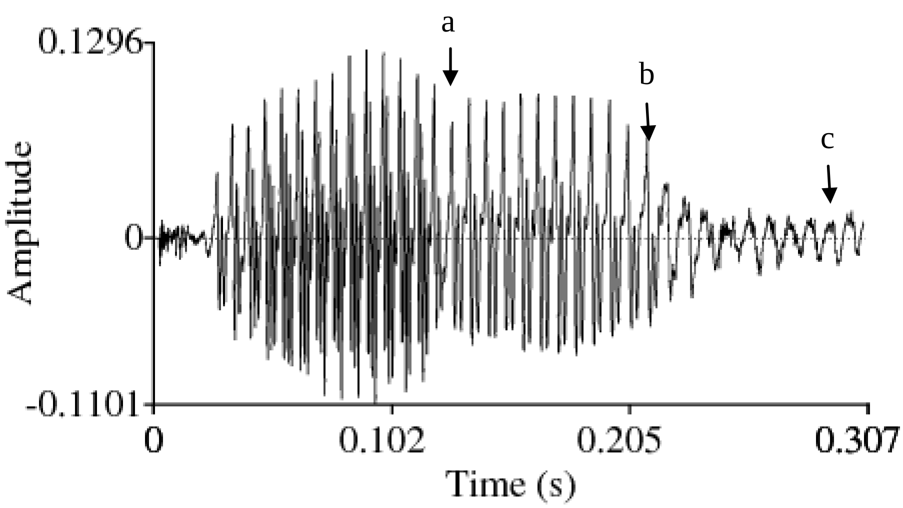


Figure 8.89(a) acoustic waveform of the alveolar lateral /l^h/ of Sindhi, taken from the word utterance /kəl^h/.

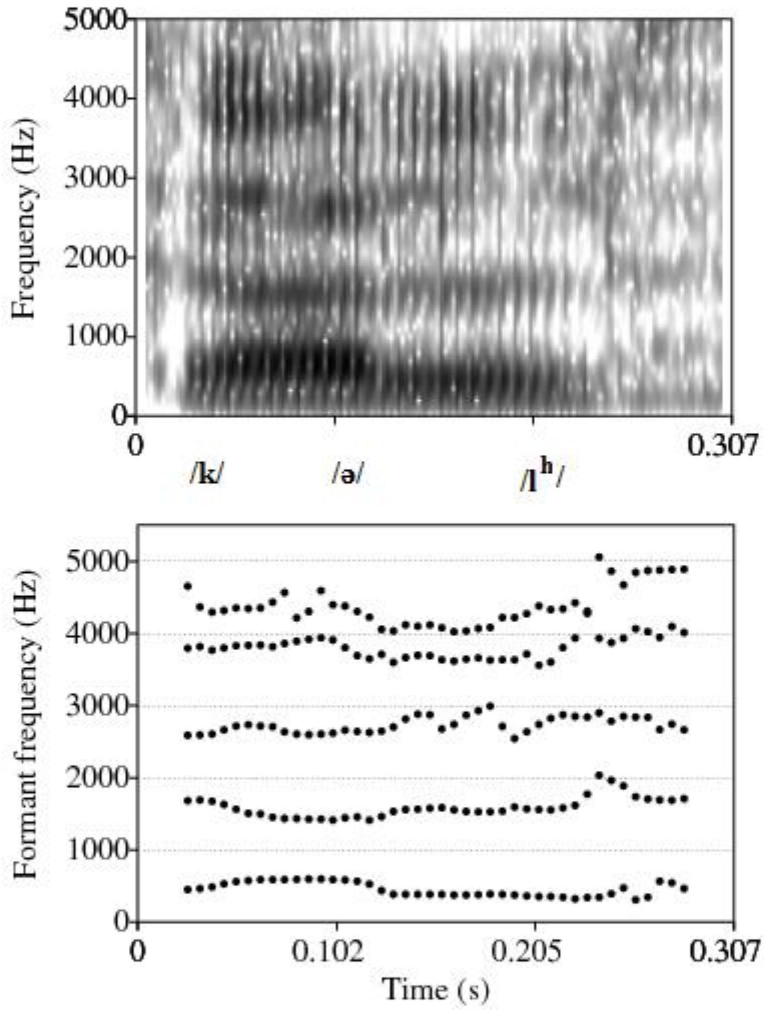


Figure 8.89(b) spectrogram of the word utterance /kəl^h/ containing the aspirated alveolar lateral /l^h/ of Sindhi.

Figure 8.90(a) and (b) show the acoustic waveform and the spectrogram of the word utterance **/pəɽ^h/** containing the retroflex flap /**ɽ^h**/ of Sindhi. The short duration periodic signals activity between the point's ***a*** and ***b*** in the figure 8.90(a) indicate that this is a voiced phoneme of Sindhi. The occurrence of weak signals for a very short duration and the discontinuity in the formant motion are common acoustic characteristics for the flap consonants of Sindhi. The signals between point's ***b*** and ***c*** indicate the presence of aspirated signals in the waveform; therefore the sound /**ɽ^h**/ is an aspirated sound. Due to the discontinuities in the formant transitions it is difficult to define the higher formants motion for this sound; however due to the low F1 the F1 motion is downward coming in to the sound see figure 8.90(b), the formant tracks. This sound is also not yet part of the language alphabet; however the phonemic importance of this sound has been proven by finding the minimal pair of words and the sound is frequently spoken by the speakers of the five dialects of Sindhi. Therefore in this study this sound is considered as a meanigfull phoneme and is classified as the aspirated voiced retroflex flap consonant of Sindhi.

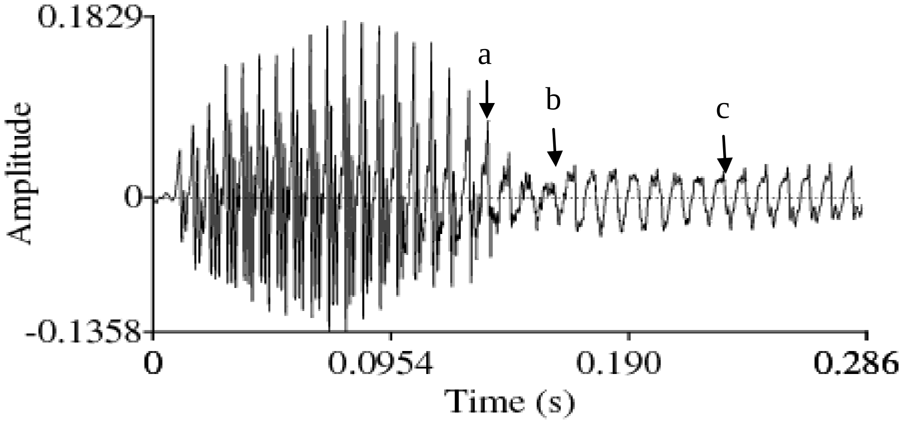


Figure 8.90(a) acoustic waveform of the aspirated retroflex flap /ɽʰ/ of Sindhi, taken from the word utterance /pəɽʰ/.

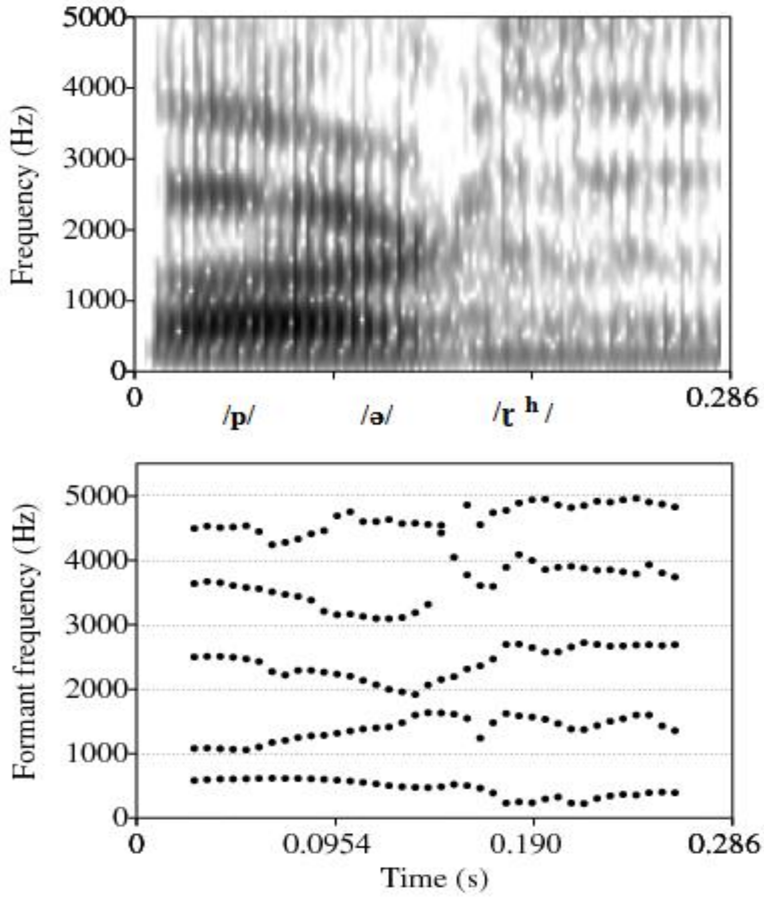


Figure 8.90(b) Spectrogram of the word utterance /pəɽʰ/ containing the aspirated retroflex flap /ɽʰ/ of Sindhi.

8.9 Conclusion

A thorough acoustic-phonetic analysis of the consonantal phonemes of Sindhi is presented in this chapter. It was observed that the acoustic properties of the closure and release segment provide fundamental acoustic cues for the identification and segmentation of the stop consonants of Sindhi discussed in section 8.2. The four stop consonants (/ʧ/ چ, /ʤ/ ج, /ʧʰ/ ڇ, /ʤʰ/ جھ) of Sindhi in this respect showed the acoustic properties of that affricate class of sounds; therefore given in appendix B these sounds are classified as members of affricate class see section 8.2.7 and 8.2.8. The nasals of Sindhi have shown the damped periodic waveforms and almost equally spaced resonant frequencies. Nasal consonants of Sindhi are discussed in section 8.4 in detail. The only exception is retroflex nasal /ɳ/ of Sindhi for which the third formant frequency is observed below 2 kHz.

The separate acoustic cues can be used for the voiced and unvoiced fricative consonants of Sindhi; because the high concentration of energy (in the higher frequency range) and high fundamental frequency are common acoustic cues for unvoiced fricatives (section 8.5), whereas the spectral peak at fundamental frequency and roughly this peak diminshes after six harmonics are common acoustic cues for all the voiced fricatives of Sindhi. Liquid class of consonants is a bit difficult to identify and segment; because the lateral /l/ of Sindhi showed vowel like acoustic characteristics such as periodic waveform and stable formants, whereas the trill /r/ showed formant continuities if surrounded by the sononrant sounds (vowels). This sound also showed repeated occurrence of speech signals especially when word utterance end with one of the short vowels this discussed in detail in section 8.6.1.

The sounds that are not yet part of the language alphabet have shown acoustic properties as an aspirated version of the following consonant phonemes: /ɽ/, /r/, /l/, /n/, /m/, and /ɳ/; therefore these sounds: /ɽʰ/, /rʰ/, /lʰ/, /nʰ/, /mʰ/, and /ɳʰ/ are classified as aspirated consonants of Sindhi (refer section 8.8); the classification of these sounds is given in appendix B. Sound symbols /lɑ/, لا and /yɑ/ ی represent two sounds of Sindhi as: /l/ ل, and /ɑ/ آ ; therefore the symbols /lɑ/ لا and /yɑ/ ی donot represent an independent phoneme of Sindhi discussed in section 8.8. The sounds /ʈr/, ٽر and /ɖr/, ڊ ر are the variant pronunciations of the sounds /ʈ/ ٽ, and /ɖ/ ڊ, mostly uttered by the speakers of the middle and Utradi dialects. The minimal pair of words for the two sounds can not be found in Sindhi; therefore the two sounds do not represent meaningful phonemes in Sindhi.

CHAPTER 9

CONCLUSION

Chapter 9

9.1 Conclusion

This research has presented a comprehensive set of acoustic properties of the elemental sounds, required for the design of an acoustic-phonetics ASR system for Sindhi. The research begins by finding the areas in which the language needs more research work to be carried out. The knowledge gained by critically examining the existing work helped to devise a step by step framework for the completion of this study. The major contributions of the research have been reported in the individual chapters of the thesis. This section provides a summary of the findings and suggests a future programme is given in next section (9.2).

For the purpose of acoustic analysis voice samples of the native Sindhi speakers were gathered from the five geographically distinct dialects of the language spoken in Sindh province (Pakistan). In addition to the dialectical variation other factors considered in the study include the speaker's age, education, environment, speaking and hearing disorders, and the duration of a speakers stay in a particular dialectical region. An independent recording session was conducted to capture the voice samples for each class of phonemes. This helped to drive the conclusions based on the research findings in groups or clusters of sounds; because the sounds of one class share some common acoustic-phonetic characteristics such as vowels having a clear and stable formant pattern, two vocalic elements for diphthongs, and damped waveforms for nasals etc.

Three different approaches are presented for the segregation of the voiced and unvoiced/silence portions of the speech using the acoustic-phonetic characteristics of the individual phonetic units of Sindhi. It is observed from the results that the STE based strategies have shown moderately good results and can be used for the purpose of segregation of the voiced and unvoiced/silence portions of the speech signals and are generally efficient computationally; however these strategies depend heavily on the selection of a suitable *ad hoc* threshold, which is manually set on a trial and error basis.

In this study separate thresholds were used for three types of the speech signals (i) speech signal part of sonorant phonemes i.e. the vowels and the voiced consonants, (ii) speech signal part of the unvoiced consonantal phonemes and (iii) the threshold for the silence or ambient noise signals. It should be noted here that there exists a high sensitivity between the signals of the unvoiced phonemes and background silence; if thresholds for these two types of signals are not selected carefully this may result in the false segmentation of the two types of signals. In order to avoid such false selection or rejection of the speech signals STE based methods are implemented and tested in two ways first is the Moving Window based technique, and second with an Overlapped Moving Window based technique. The Overlapped Moving Window based technique is preferred over the non-overlapped Moving Window based technique; because in this technique selection or rejection of speech signals depends on two consecutive windows for the selected duration. The statistical outlier-detection based strategies, ie: the '3σ edit rule', and the Hampel Identifier, are also used to segregate the unwanted contaminated noise signals from the actual speech signals i.e. the clipping or unexpected ambient noise signals. The Hampel Identifier based strategy gives better results and can be used for the segregation of the contaminated noise signals from the actual speech signals. The trade-off associated with the implementation of this technique in this study is that the method requires the microphone to be kept on for a few seconds before and after the utterance as part of the silence signals, this incurs an additional pre-processing overhead for the system. Hence, the STE overlapped moving window and the outlier-detection method the Hampel Identifier are used in an optimal combination to achieve reliable and accurate signal processing.

The other parameters used for the segmentation and identification of each phonetic class are summarized below. For the vowel phonemes the parameters measured are the fundamental frequency, duration, and first four formants. Based on the measured average vowel durations, vowels of Sindhi have been categorized as seven long (tense) and three short (lax) vowels. The obtained values of the first two formants helped to present two dimensional articulatory classification of the vowels as: high vs low and front vs back, where the first formant is closely co-related with the vowel height and the second formant is closely co-related with its backness. Acoustic properties in addition to the articulatory knowledge of the vowels provide a clear picture and understanding about the production process of these sounds. The acoustic vowel plots presented provide an idea about the vocal exercise that needs to be performed to produce ten monophthong vowels of Sindhi and to determine vowel intelligibility that is how clearly each vowel is articulated?

For the identification and segmentation of the unvoiced stop class the acoustic properties of their release segment can be used and are given in Table 8.1; whereas tracking the dynamics of the formant transitions (formant motion coming into the stop from the previous sound and going away from it) can be helpful for the segmentation and identification of the voiced stops of Sindhi. The relative increase in the amplitude of the modulation signal is one of the important acoustic cues for the identification and segmentation of implosive stops. For the nasals low F1 value (low F1 values are determined for the five nasals of Sindhi), the damped periodic waveforms, tracking the dynamics of the first and third formant transitions (the second formant transitions for some nasals are ambiguous and cannot be used), and presence of the weak energy in the higher frequency range can be used. One exception for the nasal class of sounds is the third formant transition for the retroflex nasal /ɳ/ ڻ, of Sindhi, which is lower than 2 kHz; whereas the third formant transitions for all other nasals are above 2 kHz. The high concentration of energy and the absence of the energy in the lower frequency range are the main acoustic cues for the identification and segmentation of the unvoiced fricative class of phonemes and the peak at the fundamental frequency, first formant and the presence of energy in lower frequency range can be used for the identification and segmentation of the voiced fricative class of phonemes in Sindhi. The liquid class phonemes in Sindhi are considered more difficult to be identified and segmented, when surrounded by the sonorant phonemes (vowels) such as VCV sequence of phonemes; because the lateral liquid /l/ of Sindhi have shown slow varying periodic acoustic waveform and stable formants like vowels. The distance between the first four formants can be used for the identification and segmentation of this class of phoneme; however the chance of false segmentation for this class of phoneme is high, when immediately followed or preceded by a vowel phoneme. Liquid /r/ of Sindhi; on the other hand does not exhibit formant discontinuities when surrounded by vowel phonemes (VCV phoneme sequence) in utterance. However the significant decrease in the amplitude of the modulation signal and repeated occurrence of the /r/ phoneme signals, for a very short period can be used for the identification and segmentation of this class of phoneme. For the flap /ɽ/ of Sindhi presence of weak energy, shorter in duration and low F1 values are the parameters used for the identification and segmentation in Sindhi.

The presence of diphthongs in Sindhi is a debateable issue in the reported published literature; however the core research findings regarding this phonetic class of Sindhi presented in this study are summarized below. The phonetic realization of a diphthongal phoneme in Sindhi heavily depends upon the pronunciation and the rate of the speaker’s speech. The language’s flexibility of insertion and digestion of the short vowel phonemes while pronouncing the word utterance(s) make it difficult to identify the occurrence of a diphthong in utterance. Not only this, but the dialect specific pronunciation variation potentially leads to the increase or decrease in the count of diphthong phonemes in Sindhi. For example considering the dialect specific acoustic variation; the speakers of *Utradi* and *Middle* dialects pronounce the word ‘l[ɑɪ]’; with the diphthong phoneme /ɑɪ/ at the word final position; whereas the speakers of *Lari* dialect pronounce it alternatively as ‘l[əe]’, without the diphthong phoneme /ɑɪ/ at the word final position. The acoustic analysis results in this regard suggest that when a short vowel immediately follows a long vowel (Vv) in a syllable at the word final position, the two sounds produced together result in a diphthong phoneme of Sindhi such as in words ’p[iʊ]’, ‘h[iə]’, ‘p[ɑɪ]’ etc. contain diphthong phonemes at the word final position. The opposite occurrence of the short and long vowel phonemes as (vV) in a syllable at the word final position if pronounced together do not form a diphthong phoneme of Sindhi such as words ‘p[əi]’, and ‘k^{h}ɑd^{h}[ʊi]’ etc. In this case the articulators set for the ending vowel /i/ do not entail tongue movement from the positions set for the first vowel /ə/ and /ʊ/ in the above mentioned two words respectively. The acoustic analysis results in this respect show that the anticipated diphthongs /əi/ in word ‘p[əi]’, and /ʊi/ in word ‘k^{h}ɑd^{h}[ʊi]’, present a longer steady-state formant pattern for the beginning vocalic elements (/ə/ and /ʊ/ respectively) without the off-glide transition before the ending vocalic element /i/ in this case;

Therefore the two vowels produced in this respect (vV) do not form a diphthong phoneme of Sindhi.

The two diphthongs /əy/ and /əw/ are special and analyzed separately; because of their variant pronunciation in utterance. The phonetic importance of the two diphthongs heavily depend upon how speakers pronounce them in word utterance; for example the word 's[əy]rə', in Sindhi is quite frequently pronounced alternatively as 's[e]rə', and the word 's[əw]k^{h}o', alternatively pronounced as 'sok^{h}o'. Phonetically the latter pronunciation of the two words is incorrect and hence both the words in the latter pronunciation do not form a diphthong phoneme of Sindhi.

Analysis of the phonemes /ʧ/ چ, /ʤ/ ج, /ʧ^h/ ڇ, /ʤ^h/ جھ, show that these sounds correlate well with the affricate class of sounds acoustically; whereas in the studies based on articulatory-phonetics these sounds are classified as members of the stop class. The presence of the frication noise signals in the release segment and the closure interval before the release segment provide significant evidence to classify these sounds as members of the affricate class of sounds see figures 8.19, 8.20 and 8.21 for voiced phonemes and figures 8.22, 8.23 and 8.24 for unvoiced phonemes. Phonetically affricates are defined as the sounds with the characteristics of both stop and fricative classes. In the case of the four above mentioned sounds the closure interval relates these sounds with the stop class and the release segment relates them to the fricative class of sounds; and hence the sounds are classified as members of the affricate class of Sindhi.

When analysed acoustically the word utterances for the following sounds of Sindhi pronounced by native Sindhi speakers, it is observed that the sounds (i) /s/, ث, and /s/, ص produced similar acoustic properties to those obtained for the phoneme /s/, س of Sindhi and in a similar way the sounds (ii) /z/, ذ , /z/, ض and /z/, ظ have produced similar acoustic properties to those obtained for the phoneme /z/, ز , the sound (iii) /t/, ط has similar acoustic properties to those obtained for the phoneme /t/, ت , the sound (iv) /k/, ق , produced similar acoustic properties to those obtained for the phoneme /k/, ڪ , and the sound (v) /h/, ح, produced similar acoustic properties to those obtained for the phoneme /h/, ه and the sound /?/ ع produced similar acoustic properties to those obtained for glottal /?/ ء of Sindhi. Therefore the sounds /s/, ث, ص; /z/, ذ , ض, ظ; /t/, ط ; /k/, ق; /h/, ح and /?/ ع of Sindhi are considered redundant sounds.

The sounds used interchangeably are due to dialect and language flexibility; however it should be noted that the false identification and segmentation cannot be avoided if proper attention is not given while segmenting these interchanging phonemes. These interchanging pairs of sounds are: (i) the sound /ʧ^h/, ڇ is interchanged with /ʃ/, ش and vice versa, (ii) the sound /r/, ر is interchanged with /ɽ/, ڙ and vice versa, and (iii) the sound /ʤ/, ج is interchanged with /z/, ز and vice versa.

The sounds used alternatively as substituting sounds are due to the phonetically incorrect pronunciation of these sounds; pairs of such substituting sounds are: (i) / p^h /, ڦ and / f /, ف (ii) / k^h /, ک and /x/ خ (iii) / ɣ /, غ and / g /, گ. There is a need to design an audio visual guide for the correct pronunciation of these sounds phonetically; this task is left as part of the future enhancement of the study.

The sounds that are not yet part of the language alphabet are: (i) /ɽ^h/, ڙھ (ii) /r^h/, رھ (iii) /l^h/, لھ (iv) /n^h/, نھ (v) /m^h/, مھ and (vi) /ɳ^h/, ڻھ; the phonemic importance of these sounds is proven by finding the minimal pairs of words for each sound; acoustically these sounds are analysed and classified as: the sound /ɽ^h/, ڙھ is an aspirated version of the phoneme /ɽ/, sound /r^h/, رھ, is an aspirated version of phoneme /r/, sound /l^h/, لھ is an aspirated version of phoneme /l/, sound /n^h/, نھ is an aspirated version of phoneme /n/, sound /m^h/, مھ is an aspirated version of phoneme /m/, and sound /ɳ^h/, ڻھ is an aspirated version of /ɳ/; therefore in this study these sounds are classified as aspirated sounds of Sindhi the classification of the consonants of Sindhi is given in appendix B. The sounds /lɑ/, لا and /yɑ/ ی represent two sound symbols of Sindhi they are: sounds /l/ ل, and /ɑ/ آ, for the combined symbol /lɑ/, لا and sounds /y/ ي, /ɑ/ آ, for the combined symbol /yɑ/ ی;

therefore the two symbols (/lɑ/ لا and /yɑ/ ى) do not represent independent phoneme of Sindhi. The sounds /ʈr/, ٽر and /ɖr/, ڊر are the variant forms of the sounds /ʈ/ ٽ, and /ɖ/ ڊ, mostly uttered by the speakers of the middle and Utradi dialects of Sindhi. The minimal pairs of words for these two sounds can not be found in Sindhi; therefore the two sounds can not be considered as meaningful phonemes of Sindhi.

In summary the major research achievements and contributions this study provides in the field are:

- The classification of the elemental sounds of Sindhi by incorporating their acoustic-phonetic knowledge in addition to the available articulatory-phonetic knowledge
- Comprehensive measurement of the acoustic features of the elemental sounds of Sindhi suitable to be incorporated into the design of a Sindhi ASR system.
- A cross dialect understanding of the Sindhi vowel system including the vowels intelligibility, quality and the vowel length contrasts.
- Understanding of the context dependant acoustic variation for the consonant classes of Sindhi.
- A speech database comprising the voice samples of the native Sindhi speakers, captured in two ways, for all the phonetic categories of Sindhi: (i) voice samples part of the isolated phoneme utterances and (ii) voice samples part of the phonemes embedded in the word utterances.
- Identification of the language redundant sounds acoustically and the sounds that are not yet part of the language alphabet. This helps to make complex language structural compositions simple and easy to learn.
- A clear description and understanding of the language sounds that can potentially lead to the false segmentation and recognition of Sindhi speech.
- A hybrid prototype for the segmentation of Sindhi speech capable of handling the segmentation and recognition issues for the three major classes of input signals for a Sindhi ASR system.

The research achievements listed above creates the fundamental building blocks for future work to design a state-of-the-art prototype, which is: gender and environment independent, continuous and conversational ASR system for Sindhi speech.

.

9.2 Future work

There are several areas of research in the field of speech processing for which research work in this thesis can be extended. The first and most direct contribution of this work will be designing a state of the art acoustic-phonetic model for designing a Sindhi ASR system and a synthetic Sindhi speech production system. The sound characteristics are highly variable acoustically; common factors that cause this variation are the speaker's age and sex or the environmental and contextual changes of the speech as discussed in the introductory chapters of the thesis. The tools and techniques implemented with this study meet the levels of the cluster and gender specific segmentation and identification of Sindhi speech. The design of optimized algorithms capable of handling the advanced level problems involved in the design and implementation process of continuous, connected and conversational speech recognizers etc. are the main areas that require future research. The recognition systems capable of handling misrecognition problems due to the poor quality of input signals, small bandwidth communication channels, and communication failures due to the link discontinuities etc. are common factors for system failure and misrecognition problems yet to be addressed in the future. In addition to these common misrecognition problems; there are some specific misrecognition problems associated with Sindhi speech such as the lack of a compiled database of such sounds, words or phrases pronounced alternatively, list of the words in which a short vowel either inserted or digested, the sounds having cross cluster similar acoustic properties (such as liquids and vowels), and the sounds that are produced phonetically incorrect need to be identified and type categorized; even though the majority of these sounds are borrowed sounds from Arabic, Persian and Urdu languages etc. Acoustic analysis in the areas of context-driven and infant directed speech will set new future research dimensions for Sindhi in connection to the work carried out in this study. The study about the analysis of the co-articulated sounds (in which the signals of one phoneme overlap the signals of neighbouring phonemes) is an important future work extension for this study; because the immediate impetus of such a study (analysis of co-articulated sounds) helps to prevent recognition systems from the potential false segmentation and misrecognition problem. The co-articulated sounds in an utterance make it difficult to segment and locate the accurate boundaries of the individual phonemes.

Sindhi enabled HCI systems required for the telephony, banking and healthcare industries are also left for part of the future enhancements to this work. The acquired acoustic-phonetic features in this study provide sufficient acoustic-phonetic knowledge of the sound characteristics to be incorporated in the design of such HCI enabled systems; however this requires minor changes in the design interface modules as per the target system requirements.

The neighbouring areas of speech recognition would benefit from this study by making a few changes required by the target system; such as designing synthetic speech production systems, the study of Sindhi accented English speech and the areas of clinical speech studies i.e. the speakers having speaking or hearing disorders. The required interface changes for designing these systems could be moderate but the overall basic prototype of the system design can be extended from the work carried out in this study.

This study is relevant to several speech processing areas as discussed above; however the study has direct implications for the acoustic-phonetic analysis of the female and infant-directed Sindhi speech. The tools and techniques used with this study can be directly applied for the acquisition of the acoustic-phonetic sound characteristics from the captured voice samples of infant and female Sindhi speakers. The acoustic-phonetics study of the female and infants directed Sindhi speech will enable future enhancement of this work in designing a gender independent Sindhi recognition system. Whereas the acoustic analysis of the elemental sounds of Sindhi presented in this study provides the elementary ground work for the design of a gender specific speech recognition system for Sindhi. However, state of the art algorithms need to be designed to meet the cross cluster and cross gender level segmentation and recognition for Sindhi speech.

A generic acoustic-phonetic model is suggested in this respect for the design of a Sindhi ASR system. The design of such a generic ASR system for Sindhi can be achieved by using the utterance distribution based method such as a diaphone or mono-syllabic acoustic-phonetic model discussed in chapter III, part of the future work implementations of the current study. With these future work implementations its is expected to overcome the shortcomings observed in this study such as the problems involved in the design of continuous and conversational recognizers and the problems associated with the process of automatic labelling and segmentation of the phonetic units along with the solution of several misrecognition problems related to the Sindhi speech recognition.

References

Advani, B. M. (1993). *Sindhi Boli ji Tarikh* (4th ed.). Jamshoro: Sindhi Adabi Board.

Aguilar, L. (1999). Hiatus and diphthong: Acoustic cues and speech situation differences. *Speech Communication , Vol. 28* (1), pp. 57-74.

Aiza, S., Sana, A., & Aymen, A. (2004). *Diphthongs in Urdu Language and Analysis of their Acoustic Properties.* Center for Research in Urdu Language Processing (CRULP), pp. 9-15.

Ali, A., Van der Spiegel, J., & Mueller, P. (2001). Acoustic phonetic features for the automatic classification of stop consonants. *IEEE. Transaction on. Speech and audio processing , Vol. 9*, pp. 833-841.

Amna, L. (2003). *Phonemic Inventory of Siraiki Language and Acoustic Analysis of Voiced Implosives.* Center for Research in Urdu Language Processing (CRULP), pp. 89-97.

Ansarin, A. A. (2004). An acoustic analysis of modern Persian vowels. *In The 9th SPECOM-2004 Conference* (pp. 315-318). Saint-Petersburg, Russia: Speech and Computer.

Anusuya, M. A. (2009). Speech Recognition by Machine: A Review. *International Journal of Computer Science and Information Security (IJCSIS) , Vol. 6* (3), pp. 181-205.

Aronson, L. R. (1996). An acoustic analysis of modern Hebrew vowels and voiced consonants. *Journal of Phonetics , Vol. 24* (2), pp. 283-293.

Atal, B., & Rabiner, L. (1976). A pattern recognition approach to voiced-unvoiced-silence classification with applications to speech recognition. *Acoustics, Speech and Signal Processing, IEEE Transactions on , Vol. 24* (3), pp. 201-212.

Bhugio, M. Q. (2006). The Diachronic Sociolinguistic Situation in Sindh. *Web Journal on Cultural Patrimony , Vol. 1*, pp. 89-106

Borzone de Manrique, A.M (1979). Acoustic analysis of the Spanish diphthongs. *Phonetica*, *Vol. 36*, pp. 194–206.

Borzone de Manrique, A.M & Massone, M. (1981). Acoustic analysis and perception of Spanish fricative consonants. *Journal of Acoustic Society of America, Vol. 69*, pp. 1145-1153.

Bradlow, A. (1995). A comparative acoustic study of English and Spanish vowels. *Journal of the Acoustical Society of America , Vol. 97* (3), pp. 1916-1924.

Chen, Y., Robb, M., Gilbert, H., & Lerman, J. (2001). Vowel production by Mandarin speakers of English. *Clinical Linguistics & Phonetics , Vol. 15*, pp. 427–440.

Chen, Y. (2008). The acoustic realization of vowels of Shanghai Chinese. *Journal of Phonetics , Vol. 36* (4), pp. 629-648.

Childers, D. G., Hahn, M., & Larar, J. N. (1989). Silent and voiced/unvoiced/mixed excitation (four-way) classification of speech. *IEEE Transactions on, Acoustics, Speech and Signal Processing , Vol. 37* (11), pp. 1771-1774.

Cooley, J. W. and J. W. Tukey. (1965). An Algorithm for the Machine Computation of the Complex Fourier Series. *Mathematics of Computation, Vol. 19,* pp. 297-301.

Corbishley, M. (1989). *In Search of the Past: Secret cities.* London: Evans Bros.

Cynthia, M. C., & Darnieder, L. M. (2009). Perceptual learning of co-articulation in speech. *Journal of Memory and Language , Vol. 61* (3), pp. 412-422

David, K. M. (2003). Reasons for Language Shift in Peninsular Malaysia. *In Investing in Innovation: Humanities Social and Science. Vol (7),* pp. 111-114. Ramlah: Universiti Putra Malaysia Press, Serdang, Selangor, Malaysia.

Demolin, D. (1995). The phonetics and phonology of glottalized consonants in Lendu. Phonology and Phonetic Evidence. *Papers in Laboratory Phonology IV, ed. by Bruce Connell and Amalia Arvanti*, pp. 368-385. Cambridge: Cambridge University Press

Deterding, D. (2003). An instrumental study of the monophthong vowels of Singapore English. *English World Wide* , *Vol. 24* (1), pp. 1-16.

Dobler, S., Meyer, P., & Ruehl, H. (1992). A robust connected-words recognizer. *International Conference on, Acoustics, Speech, and Signal Processing. Vol. 1*, pp. 245-248. ICASSP-92, IEEE, 1992.

Duda, R. O., Hart, P. E., & G., S. D. (2000). *Pattern Classification* (2 ed.). John Wiley and Sons Inc.

Edwards, T. J. (1981). Multiple features analysis of intervocalic english plosives. *Journal of the Acoustical Society of America , Vol. 69* (2), pp. 535–547.

Espy-Wilson, C. (1987). *An acoustic phonetic approach to speech recognition: Application to the semivowels.* Massachusetts Institute of Technology: Ph.D thesis.

Espy-Wilson, C. (1994). A feature-based semivowel recognition system. *Jouranl of Acoustic Society of America , Vol. 96,* pp. 65–72.

Fant, G. (1973). *Speech Sounds and Features.* Cambridge, Mass: MIT Press.

Flanagan, J. L. (1972). *Speech Analysis, Synthesis, and Perception* (2 ed.). Springer-Verlag, New York.

Furui, S. (2005). 50 years of Progress in speech and Speaker Recognition Research. *ECTI Transactions on Computer and Information Technology , Vol. 1* (2), pp. 64-74.

Gay, T. (1968). Effects of speaking rate on diphthong formant movements. *Journal of the Acoustics Society of America ,Vol. 44*, pp. 1570-1573.

Gordon, M., & Ayla, A. (2006). Phonetic structures of Turkish Kabardian. *Journal of the International Phonetic Association , Vol. 36* (2), pp. 159-186.

Grierson, G. A. (1919). Indo-Aryan Family: Northwestern Group, Specimens of Sindhi and Lahnda. In *Linguistic Survey of India* (Vols. Vol-VII). Calcutta: Superintendent Government Printing.

Grimes, B. F. (2001). Global Language Viability. *In Endangered Languages of the Pacific Rim, Lectures on Endangered Languages: 2, from Kyoto Conference 2000: Osahito Miyaoka* (pp. (pp. 45–61)). Kyoto: ELPR publisher, Osaka Gakuin University.

Gurlekian, J., Guirao, M., & Franco, H. (1985). Acoustic characteristics and perception of Spanish stop consonants. *J. Acoust. Soc. Japan , Vol. 65* (36), pp. 271–278.

Halin, G., & Shafiro, S. S. (1967). *Statistical Models in Engineering.* USA: Wiley.

Hawkins, D. M. (1980). *Identification of Outliers.* London, Angleterre : Chapman and Hall.

Hemdal, J. F., & Hughes, G. W. (1967). A feature based computer recognition program for the modeling of vowel perception. *In: Models for the Perception of Speech and Visual Form, W. Wathen-Dunn.* Ed. Cambridge: M.I.T. Press.

Homma, Y. (1973). An Acoustic Study of Japanese Vowels. *Study of Sounds 16* (pp. 347-368). Tokyo: Phonetic Society of Japan.

Hongyan, W. (2007). *Mutual intelligibility of Chinese, Dutch and American speakers of English.* Ph.D dissertation, Graduate School of Linguistics: Netherlands.

Hunt, M. (1989). A Comparison of Several Acoustic Representations for Speech Recognition with Degraded and Undegraded Speech. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, (pp. 262-265).

Ioana, C. (2002). A perception-production study of Romanian diphthongs and glide-vowel sequences. *Journal of the International Phonetic Association , Vol. 32*, pp. 203-222.

IPA. (1999). *Handbook of the International Phonetic Association: a guide to the use of the International Phonetic Alphabet.* Cambridge: Cambridge University Press.

Jatoi, A. N. (1996). *Sindhi Boli jo Lisani Jaezo* (3 ed.). Jamshoro: Sindh Text Book Board.

Jennifer, S. C. (2005). Language, Identity and the Survival of Culture. *Conference on Language, Terrorism and Ideology: The Sindhi Struggle in Pakistan.* Washington, DC. October 2005.

Jennifer, S. C. (2006). The Sindhi language. In *In K. Brown (ed.) Encyclopedia of Language and. Linguistics* (2 ed., Vol. 11, pp. 384-386). Oxford: Elsevier.

Jha, S. K. (1985). Acoustic analysis of the Maithili diphthongs. *Journal of Phonetics, Vol. 13*, pp. 107-115.

Johnson, K. (2003). *Acoustic and auditory phonetics* (2 ed.). Blackwell Publishing.

Jones, D. (1969). *An outline of English phonetics* (9 ed.). England: W. Heffer & Sons Ltd., Cambridge.

Jurafsky, D., & Martin, J. (2008). *Speech and Language Processing.* Prentice Hall.

Keerio, A., Mitra, B., Birch, P., Young, R., & Chatwin, C. (2009). On preprocessing of speech signals. *International Journal of Signal Processing , Vol. 5* (3), pp. 216-222.

Keerio, A., Patoli, M. Z., Mitra, B. K., Chatwin, C., Young, R., & Birch, P. (2010). Acoustic Analysis of Diphthongs in Sindhi. *GRASSROOTS , 41* (1), pp. 1-18.

Kent, R. D., & Charles, R. (2002). *The acoustic analysis of speech* (2 ed.). Singular Publishing Group.

Kewley-Porr, D., Akahane-Yamada, R., & Aikaea, K. (1996). Intelligibility and acoustic correlates of Japanese accented English vowels. *Fourth International Conference on Spoken Language: ICSLP-96, Vol. 1*, pp. 450-453.

Khawaja, M. A., & Najmi, G. H. (2004). Acoustic Analysis of Phonetics of Arabic Script Sindhi Language to evaluate Vowel-Consonant Segmentation. *Journal of Independent Studies and Research (JISR) , Vol. 2* (2), pp. 15-26.

Khawaja, M., & Najmi, N. (2007). Segmentation of Sindhi Speech using Formants. *IEEE International Conference on Signal Processing and Communications, . ICSPC* , (pp. 796-799).

Khubchandani, L. (1961). *The phonology and morphophonemics of Sindhi.* University of Pennsylvania: M.A. thesis.

Kurowski, K., & Blumstein, S. (1987). Acoustic properties for place of articulation in nasal consonants. *Journal of the Acoustical Society of America, Vol. 81*, pp. 1917–1927.

Ladefoged, P. (1993). *A course in phonetics* (3 ed.). Harcourt College Publishers, New York.

Ladefoged, P., & Maddieson, I. (1996). *The Sounds of the World's Languages.* Oxford: Blackwell.

Ladefoged, P. (2001). *Vowels and consonants: an introduction to the sounds of languages.* Wiley-Blackwell publishers.

Ladefoged, P. (2005). *A Course in Phonetics* (5 ed.). Thomson/Wadsworth Publishers.

Laughlin Mc, F. (2005). Voiceless implosives in Seereer-Siin. *Journal of the International Phonetic Association, Vol. 35*, pp. 201-214.

Lee, C. H., Giachin, E., Rabiner, L. R., Pieraccini, R., & Rosenberg, A. E. (1992). Improved acoustic modelling for large vocabulary speech recognition. *Comput. Speech Lang. , Vol. 6,* pp. 103–127.

Levinson, S. E., Ljolje, A., & Miller, L. G. (1990). Continuous speech recognition from a phonetic transcription. *In the Proceedings of ICASSP.*

Lewis, J. (2002). *Social influences on female speakers' pitch.* Ph.D dissertation, University of California, Berkeley.

Liu, H.-M. (2001). *The acoustic-phonetic characteristics of infant-directed speech in Mandarin Chinese and their relation to infant speech perception in the first year of life.* Unpublished doctoral dissertation, University of Washington.

Lobanov, B. M. (1971). Classification of Russian vowels spoken by different speakers. *Journal of the Acoustical Society of America , Vol. 49* (2B), pp. 606-608.

Madureira, S. (1994). Pitch Patterns in Brazilian Portuguese: an Acoustic-Phonetic Analysis. *In Proceddings of the Fifth Australian International Conference on Speech Science and Technology,* (pp. 156-159).

Maniwa, K., Jongman, A., & Wade, T. (2009). Acoustic characteristics of clearly spoken English fricatives. *Journal of Acoustical Society of America , 125* (6), pp. 3962–3973.

Marshall, J. H. (1931). *Mohenjo-daro and the Indus Civilization* (Vol. Vol. I).

Martínez Celdrán, E. (2004). Problems in the classification of approximants. *Journal of the International Phonetic Association, Vol. 34*, pp. 201-210.

Martland, P., Whiteside, S., Beet, S., & Baghai-Ravary, L. (1996). Analysis of Ten Vowel Sounds Across Gender and Regional Cultural Accent. *In The Fourth International Conference on Spoken Language, ICSLP 96, Vol. 4*, pp. 2231–2234.

Maxwell, O., & Janet, F. (2009). Acoustic and durational properties of indian english vowels. *World Englishes , Vol. 28* (1), pp. 52-69.

Miller, J., & Weinert, R. (1998). *Spontaneous spoken language: syntax and discourse.* Oxford: Clarendon Press.

Mirza, Q. B. (2006). *Sindhi Viya Karan* (3 ed.). Sindhi Adabi board, Jamshoro.

Mitra, A., Chatterjee, B., & Mitra, B. K. (2005). Identification of Primitive Speech Signals using TMS320C54X DSP Processor. *In Proceedings of the Eleventh National Conference on Communications (NCC), IIT- Kharagpur, India,* (pp. 286-290).

Mitra, A., Mitra, B. K., & Chatterjee, B. (2005). Recognition of Isolated Speech Signals using Simplified Statistical Parameters. *Enformatika Transactions, Vol. 8,* pp. 151-154.

Mitra, B. K., Young, R., & Chatwin, C. (2007). On shadow elimination after moving region segmentation based on different threshold selection strategies. *Optics and Lasers in Engineering , Vol. 45* (11), pp. 1088-1093.

Moore, B. C. (2003). *An introduction to the psychology of hearing* (5 ed.). Emerald Group Publishing.

Moore, B. C., Lorraine, K. T., & Marslen-Wilson, W. (2008). Introduction. The perception of speech: from sound to meaning. *Phil. Trans. R. Soc. Biological Sciences ,Vol. 363*, pp. 917-921.

Nihalani, P. (1974). Lingual articulation of stops in Sindhi. *Phonetica , Vol. 30*, pp. 197-212.

Nihalani, P. (1975). Air Flow Rate in the Production of Stops in Sindhi. *Phonetica , Vol. 31*, pp. 198-205.

Nihalani, P. (1986). Phonetic implementation of implosives. *Language and Speech, Vol. 29*, pp. 253-262.

Ohde, R. N., & Stevens, K. N. (1983). Effect of burst amplitude on the perception of stop consonant place of articulation. *Journal of the Acoustical Society of America, Vol. 74* (3), pp. 706–714.

Olive, J. P., Alice, G., & John. C. (1993). *Acoustics of American English Speech: a dynamic approach.* New York: Springer-Verlag.

Ouni, S., & Yves, L. (2003). A study of the French vowels through the main constriction of the vocal tract using an acoustic-to-articulatory inversion method. *In the 15th International Congress of Phonetic Sciences*, (pp. 2901–2904). Barcelona, Spain.

Padgett, J. (2008). Glides, vowels, and features. *Lingua , Vol. 118* (12), pp. 1937-1955.

Parsons, T. W. (1987). *Voice and Speech Processing.* McGraw-Hill.

Patterson, R., Smith, D., Dinther, R. V., & Walters, T. (2007). Size Information in the Production and Perception of Communication Sounds. *In: Auditory Perception of Sound Sources.* Eds: W.A. Yost, R.R. Fay, &A. N. Popper, Springer, New York.

Pearson, R. K. (2002). Outliers in process modeling and identification. *IEEE Transactions on, Control Systems Technology , Vol. 10* (1), pp. 55-63.

Pearson, R. K. (2005). Mining imperfect data dealing with contamination and incomplete records. *Philadelphia: SIAM.*

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America , Vol. 24*, pp. 175-184.

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

Pithawalla, M. B. (1935). *Historical geography of Sindh.* Institute of Sindhology, University of Sindh Jamshoro.

Rabiner, L. R., & Juang, B. H. (1993). *Fundamentals of speech recognition.* Prentice Hall Signal Processing Series.

Rabiner, L. R., & Schafer, R. W. (2007). *Introduction to Digital Speech Processing* (Vol. 1). Foundations and Trends in Signal Processing.

Raphael, L. J., Gloria, J. B., & Katherine, S. H. (2006). *Speech science primer: physiology, acoustics, and perception of speech* (5 ed.). Lippincott Williams & Wilkins.

Raza, S., Agha, F. Z., & Usman, R. (2004). *Phonemic inventory of Sindhi and acoustic analysis of voiced implosives.* Center for Research in Urdu language Processing (CRULP).

Roach, P. (2009). *English phonetics and phonology* (4 ed.). Cambridge University Press.

Rosner, B. S., & Pickering, J. (1994). *Vowel perception and production.* Oxford: Oxford University Press.

Saha, G., Chakroborty, S., & Senapati, S. (2005). A new Silence Removal and End Point Detection Algorithm for Speech and Speaker Recognition Applications. *In Proc. of Eleventh National Conference on Communications (NCC), IIT-Kharagpur, India,* (pp. 291-295).

Sarma, V., & Venugopal, D. (1978). Studies on pattern recognition approach to voiced-unvoiced-silence classification. *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '78, Vol. 3,* pp. 1-4.

Sarwar, A., Sana, A., & Aymen, A. T. (2004). *Diphthongs in Urdu Language and Analysis of their Acoustic Properties.* Center for Research in Urdu Language Processing (CRULP).

Stevens, K. N., & Klatt, D. H. (1974). Role of formant transitions in the voiced–voiceless distinction for stops. *Journal of the Acoustical Society of America , Vol. 55*, pp. 653–659.

Stevens, K. N. (1993). Models for the production and acoustics of stop consonants. *Speech Communication , 13*, pp. 367–375.

Stevens, K. N. (1998). *Acoustic Phonetics.* Cambridge, Mass: MIT Press.

Stevens, K. N. (2000). *Acoustic Phonetics.* MIT Press; New edition edition.

Tabain, M. (2001). Variability in fricative production and spectra: Implications for the Hyper- and Hypo- and Quantal theories of speech production. *Lang. Speech, Vol. 44*, pp. 58–93.

Traunmuller, H. (1990). Analytical expressions for the tonotopic sensory scale. *Journal of the Acoustical Society of America , 88* (1), pp. 97–100.

Trumpp, E. (1872). *Grammar of the Sindhi Language.* London: Trubner and Co. and Leipzig: FA Brockhaus, reprinted in 1970, Biblio Verlag, Osnabruck.

William, D. J., & Jackson, S. J. (2006). *The essential world history* (3 ed.). Belmont, CA : Thomson Wadsworth.

Yuasa, I. (2001). *Politeness, emotion & gender: A sociophonetic study of voice pitch modulation.* Berkeley: University of California, Doctoral dissertation.

Zhou, X., Espy-Wilson, C. Y., Boyce, S., Tiede, M., Holland, C., & Choe, A. (2008). A magnetic resonance imaging-based articulatory and acoustic study of “retroflex” and “bunched” American English /r/. *Journal of the Acoustical Society of America, 123* (6), pp. 4466–4481.

Zue, V. W. (1979). *Acoustic characteristics of stop consonants: A controlled study.* D.Sc. thesis, Mass. Inst. Technol., Cambridge.

Appendices

Appendix A:

Dialectical geography of Sindhi

Studies about the dialectical geography suggest five dialects of Sindhi are spoken in Sindhi province (Pakistan) and one dialect is spoken in Gujarat, a state of India. During the data collection process of this study we came across with some interesting facts regarding the inter-geographical boundaries of Sindhi. District Sanghar in the studies is described as part of the Middle dialect; however the dialect spoken in this region is either an independent new dialect or it is a sub-dialect of the *Middle* (dialect with in dialect). The location of this region is highlighted with the circle on the map of Sindhi Province, shown in figure 1a.

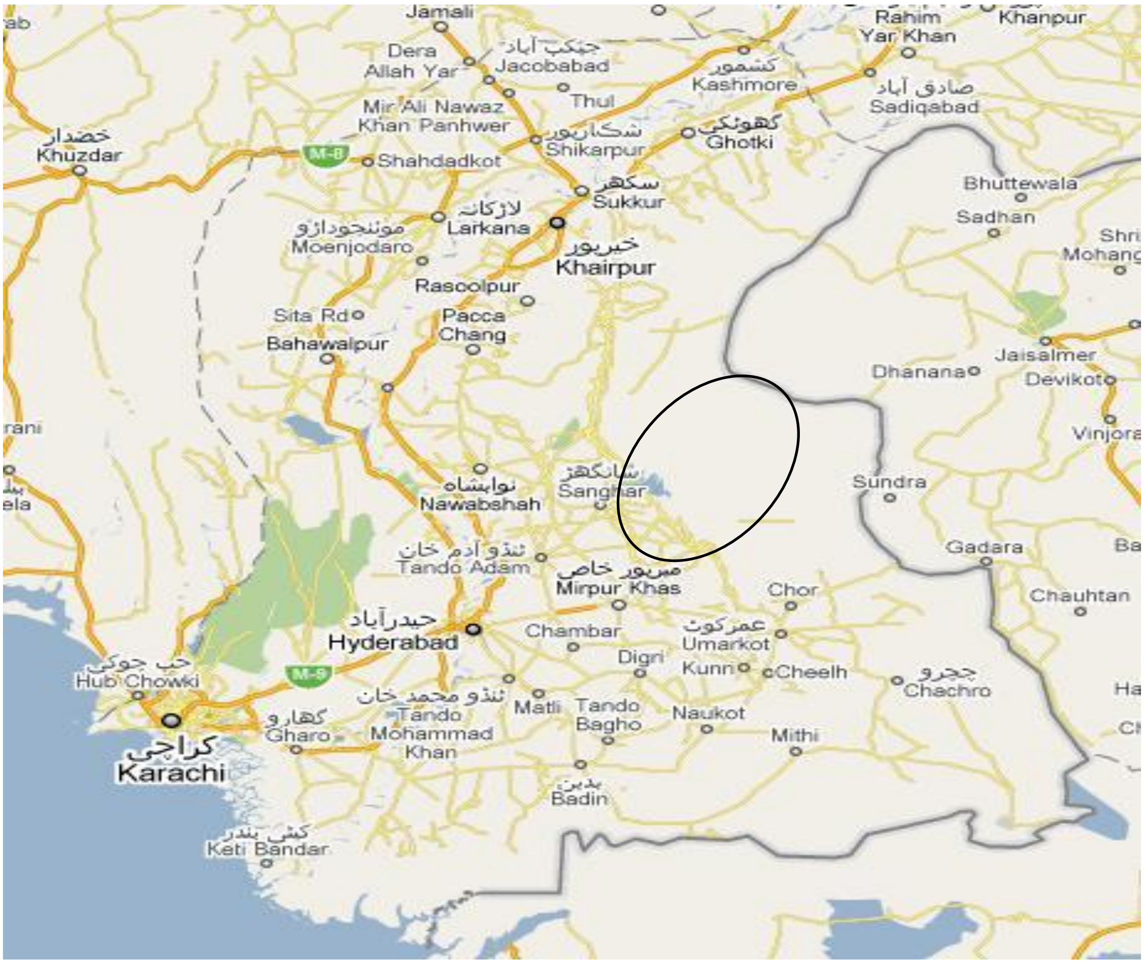


Figure 1a. District Sanghar is part of the *Middle* dialect.

District Umerkot is geographically part of Thareli dialect; however its geography is divided between two dialects. Its western boundaries are in communication with the districts Mirpurkhas and Hyderad and therefore the language of speakers is under the influence of the Middle dialect and the eastern boundaries are in communication with the district Tharparkar and are under the influence of Thareli dialect. Shown in figure 1b, the dotted circle shows the eastern part and the solid circle shows the western part.

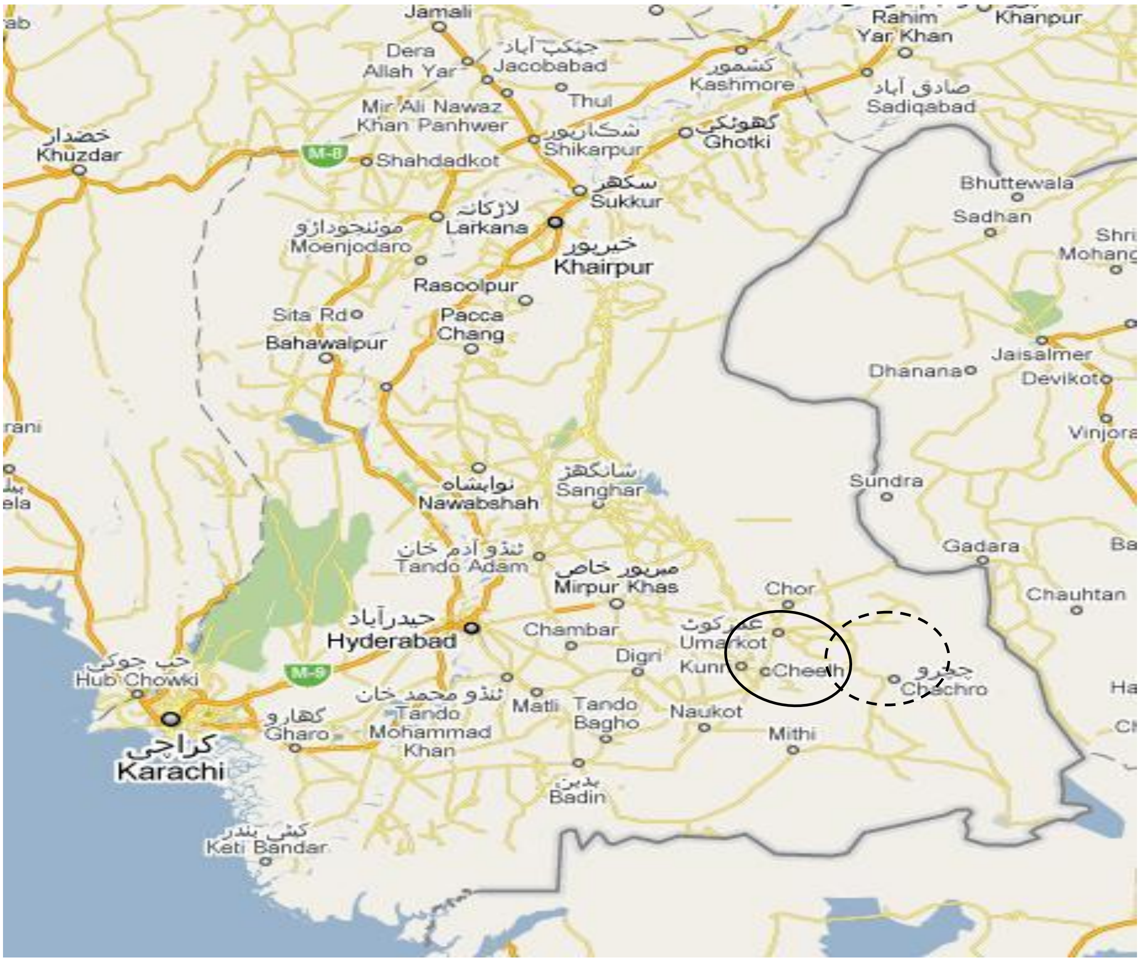


Figure 1b. District Umerkot is part of the *Thareli* dialect.

The district Tando Muhammad Khan is part of the Middle dialect like district Umerkot. The southern part of the district is in communication with *Lari* dialect and is under influence of *Lari* dialect; whereas the northern part is in communication with *Middle* dialect and is more influenced by it. Shown in figure 1c, the dotted circle shows the southern part and the solid line circle shows the northern part of the district.

Figure 1c. District Tando Muhammad Khan is part of the *Middle* dialect.

District Larkana and Shikarpur districts are part of the *Utradi* dialect at present. The peoples of two districts speak a slightly different dialect to the *Utradi* dialect. Figure 1d shows the geographical region (Shikarpur and Larkana) for the *Utradi* dialect. A geographically based survey is required to revise the newly set dialectical boundaries due to various factors. The need to study and recognise such tribe based dialects spoken in Sindh province (Pakistan) i.e. the Jats (a tribe), the Barochky (which is well known as Siraki but the Barochky spoken in this region is entirely different than the one spoken in Punjab province of Pakistan, Dhatki spoken in some parts of Thar desert etc.

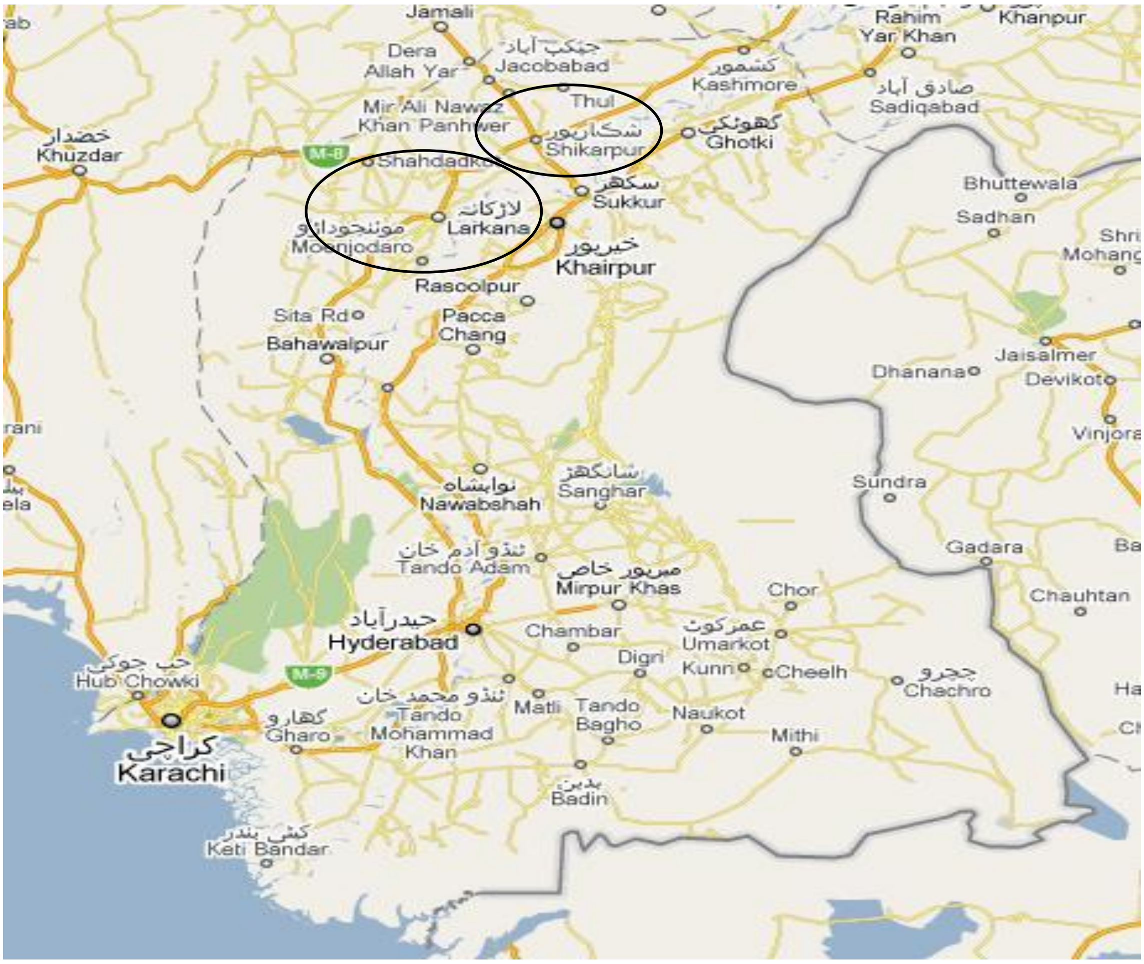


Figure 1d. District Shikarpur and Larkana of the *Utradi* dialect.

Appendix B:

Classification of the consonantal phonemes of Sindhi

	Bilabial		Dental		Alveolar		Retro-flex		Palato-Alveolar		Palatal		Velar		Glottal
	U	V	U	V	U	V	U	V	U	V	U	V	U	V	
Stops	پ p	ب b	ت t	د d			ٽ ʈ	ڊ ɖ					ڪ k	گ g	ء ʔ
A	ڦ p^h	ڀ b^h	ٿ t^h	ڌ d^h			ٺ $ʈ^h$	ڍ $ɖ^h$					ک k^h	گھ g^h	
I	- -	ٻ ɓ					- -	ڏ ɗ	- -	ڄ ʄ			- -	ڳ ɠ	
Nasals	- -	م m			- -	ن n	- -	ڻ ɳ			- -	ڃ ɲ	- -	ڱ ŋ	
A	- -	مھ m^h			- -	نھ n^h	- -	ڻھ $ɳ^h$							
Fricatives	ف f	و v			س s	ز z			ش ʃ	- -			خ x	غ ɣ	ه h
Affricates									چ ʧ	ج ʤ					
A									ڇ $ʧ^h$	جھ $ʤ^h$	`				
Trills					- -	ر r									
A					- -	رھ r^h									
Laterals					- -	ل l									

A			- لھ - l^h					
Flaps				- ڑ - ɽ				
A				- ڑ - $ɽ^h$				
Glides	- و - v					- ي - y		

Appendix C:

Raw formant values of the ten monophthong vowels of Sindhi across five dialects of Sindhi.

Vowel	Dialect	No. of Speakers	F1 (Hz)	F2 (Hz)	F3 (Hz)	F4 (Hz)
/i/	Middle	1	235	2391	3167	3910
		2	249	2368	3072	3731
		3	216	2327	3122	3747
		4	280	2280	3158	3730
		5	246	2339	3059	3874
		6	304	2295	2725	3010
		7	250	2328	3195	3859
		8	252	2346	2943	3662
		9	235	2405	3307	3712
		10	265	2295	2937	3650
		11	255	2289	2940	3718
		12	230	2364	3278	3559
		13	274	2351	3158	3542
		14	272	2216	2355	3059
		15	291	2361	3084	3837
	Utradi	1	260	2413	3117	3655
		2	250	2331	3005	3611
		3	278	2221	2905	3807
		4	281	2361	3155	3762
		5	305	2390	2948	3761
		6	233	2416	3414	3844
		7	227	2250	2993	3675
		8	258	2325	3478	3842
		9	236	2479	3030	3860
		10	283	2337	3168	3751
		11	239	2155	2981	3688
		12	293	2386	3210	3686
		13	270	2406	3120	3732
		14	318	2378	3044	3743
		15	238	2267	2991	3705

Vowel	Dialect	No. of Speakers	F1 (Hz)	F2 (Hz)	F3 (Hz)	F4 (Hz)
/i/	Thareli	1	218	2402	2932	3734
		2	271	2406	3227	3764
		3	225	2311	3200	4152
		4	258	2175	2428	3228
		5	228	2292	3017	3867
		6	206	2322	3176	4204
		7	213	2369	3122	4214
		8	240	2055	2388	3200
		9	303	2304	2867	3192
		10	220	2367	3248	3215
		11	231	2315	3147	3408
		12	279	2318	3195	3713
		13	285	2410	3153	3725
		14	209	2477	3101	3585
		15	233	2450	2879	3483
	Lari	1	263	2369	3156	3766
		2	257	2345	3193	3854
		3	276	2278	3112	3212
		4	317	2398	2977	3842
		5	295	2286	2725	3745
		6	286	2270	2867	3645
		7	289	2298	2826	3853
		8	283	2302	2882	3763
		9	245	2423	3184	3876
		10	290	2385	3101	3955
		11	273	2452	3447	3795
		12	300	2481	3387	3808
		13	291	2484	3518	3882
		14	305	2173	3098	3647
		15	263	2191	3188	3641
	Lasi	1	235	2441	3157	3612
		2	250	2444	3036	3796
		3	251	2418	3071	3808
		4	276	2440	3064	3781
		5	284	2408	3191	3853
		6	279	2285	3104	4043
		7	285	2462	2917	3920
		8	289	2458	3725	3860
		9	309	2316	3029	3742
		10	245	2413	3131	3670
		11	260	2343	3156	3750
		12	239	2327	3017	3725
		13	262	2418	3012	3704
		14	265	2483	3085	3613
		15	302	2480	3001	3690

Vowel	Dialect	No. of Speakers	F1 (Hz)	F2 (Hz)	F3 (Hz)	F4 (Hz)
/ɪ/	Middle	1	320	2045	2668	3739
		2	346	2098	2444	3185
		3	354	2183	3029	3735
		4	341	2255	2977	3894
		5	295	2085	2221	3053
		6	325	2090	2805	3896
		7	299	2229	2870	3711
		8	317	2248	2290	3029
		9	305	2162	2974	3743
		10	311	2215	2881	3893
		11	340	2220	2928	3759
		12	325	2171	2907	3857
		13	331	2084	2703	3792
		14	321	2079	2727	3927
		15	330	2132	2869	3833
	Utradi	1	313	2083	2543	3827
		2	318	2151	2690	3483
		3	351	2198	2779	3593
		4	340	2041	2605	3787
		5	310	2082	2826	3576
		6	372	2038	2537	4057
		7	295	2036	2514	3975
		8	330	1983	2543	3827
		9	354	2109	2800	3732
		10	335	2118	2814	3714
		11	334	2029	1845	2827
		12	305	2056	2543	3827
		13	355	1990	2605	3787
		14	325	2182	2826	3576
		15	315	1883	2543	3827
	Thareli	1	338	1949	2533	3806
		2	320	2074	2535	3742
		3	302	2176	2879	3537
		4	335	2040	2554	3939
		5	321	2223	2629	3712
		6	330	2221	2788	3774
		7	351	2130	2506	3831
		8	334	2175	2892	4427
		9	305	1963	2477	3845
		10	334	2175	2892	4427
		11	335	2113	2715	3028
		12	310	2155	2830	3649
		13	295	2056	2512	3710
		14	350	2177	2881	3653
		15	317	2243	3184	3726

Vowel	Dialect	No. of Speakers	F1 (Hz)	F2 (Hz)	F3 (Hz)	F4 (Hz)
/ɪ/	Lari	1	320	2204	2720	3391
		2	321	2079	2727	3927
		3	330	2132	2869	3833
		4	327	2184	2673	3802
		5	363	1990	2547	3530
		6	316	2173	2858	3893
		7	324	2165	2690	3854
		8	335	2075	2668	3739
		9	300	2018	2739	3994
		10	301	2061	2602	3714
		11	325	2264	2831	3862
		12	333	2237	2986	3642
		13	325	2036	2514	3975
		14	331	1994	2440	3620
		15	301	2189	2848	3710
	Lasi	1	332	2272	2679	3525
		2	315	2380	2963	3824
		3	331	2262	2712	3894
		4	315	2380	2963	3824
		5	308	2271	3093	3632
		6	328	2235	2758	3505
		7	333	2337	2986	3642
		8	331	1994	2440	3620
		9	332	2272	2679	3525
		10	316	2173	2858	3893
		11	344	2080	2528	3963
		12	325	2171	2907	3857
		13	330	2055	2512	3710
		14	354	2277	2433	4085
		15	304	2229	2799	3863
/e/	Middle	1	415	2155	2601	3623
		2	389	2156	2799	3720
		3	418	1984	2778	3623
		4	393	2123	2806	3799
		5	398	2061	2517	3654
		6	422	2085	2755	3658
		7	388	2193	2725	3202
		8	442	2009	2151	2810
		9	436	2188	2809	3686
		10	452	2073	2911	3792
		11	376	2133	2809	3677
		12	407	2097	2564	3622
		13	377	2089	2806	3925
		14	368	2098	3079	3676
		15	356	2144	2795	3722

Vowel	Dialect	No. of Speakers	F1 (Hz)	F2 (Hz)	F3 (Hz)	F4 (Hz)
/e/	Utradi	1	389	2117	2943	3916
		2	387	1972	2634	3792
		3	388	2150	2741	3764
		4	349	2027	2516	3990
		5	360	2164	2901	3458
		6	351	1938	2577	3772
		7	374	1973	2653	2804
		8	366	2126	2936	3566
		9	349	2195	2925	3583
		10	386	1942	2592	3882
		11	420	2085	2926	3546
		12	345	2172	2964	3523
		13	373	2071	2581	4001
		14	350	2047	2554	3898
		15	380	1979	2444	4077
	Thareli	1	418	2208	2865	3689
		2	375	2129	2705	3598
		3	357	2011	2530	3757
		4	405	2024	2499	3697
		5	415	2169	2838	3742
		6	369	2107	2502	3715
		7	377	2215	2921	3864
		8	352	2123	2788	3682
		9	373	2101	2815	3564
		10	383	2042	2743	3480
		11	369	1907	2502	3715
		12	409	2170	2768	3729
		13	356	2134	2885	3644
		14	402	2155	2912	3705
		15	385	2175	2815	3657
	Lari	1	378	2062	2622	3617
		2	371	2203	2782	3680
		3	402	2020	2447	3862
		4	403	2072	2506	3859
		5	331	2205	2676	3767
		6	426	2147	2813	3773
		7	368	2165	2948	3822
		8	425	2168	2832	3778
		9	419	2254	2843	3710
		10	368	1994	2638	3656
		11	386	2173	2773	3750
		12	379	2221	2840	3822
		13	340	2165	2766	3627
		14	366	2147	2735	3629
		15	345	2142	2738	3652

Vowel	Dialect	No. of Speakers	F1 (Hz)	F2 (Hz)	F3 (Hz)	F4 (Hz)
/e/	Lasi	1	369	2180	2923	3590
		2	344	2094	2716	3623
		3	304	2190	2831	3601
		4	403	2185	2933	3819
		5	379	2128	2242	2690
		6	373	2225	2884	3581
		7	372	2215	2983	3853
		8	384	2163	3002	3922
		9	369	2209	2938	3644
		10	381	2190	2898	3595
		11	342	2285	2504	2921
		12	436	2069	2564	3612
		13	328	2175	2965	3572
		14	342	2230	2804	3861
		15	377	2144	3000	3871
/ɛ/	Middle	1	465	1950	2648	3729
		2	425	1985	2753	3605
		3	472	1879	1829	2730
		4	457	1835	2106	2586
		5	485	1918	2406	3492
		6	478	2048	2738	3635
		7	516	1867	1975	2736
		8	462	1862	3009	3732
		9	463	1825	2396	2734
		10	510	1895	2646	3677
		11	486	1985	2941	3819
		12	464	2042	2773	3795
		13	429	1967	2800	3822
		14	499	2006	2823	3810
		15	479	2027	2427	3423
	Utradi	1	410	1905	2703	3681
		2	454	1833	2609	3875
		3	456	1684	2467	4024
		4	475	1986	2749	3592
		5	482	1973	2737	3672
		6	505	1879	2703	3851
		7	489	1683	2390	3806
		8	446	1841	2230	3386
		9	455	1765	2631	3668
		10	403	1854	2571	3797
		11	433	1734	2491	3939
		12	539	1726	2617	3426
		13	500	1844	2669	3342
		14	434	1964	2681	3843
		15	490	1856	2654	3465

Vowel	Dialect	No. of Speakers	F1 (Hz)	F2 (Hz)	F3 (Hz)	F4 (Hz)
/ɛ/	Thareli	1	454	1775	2505	3712
		2	493	1874	2614	3766
		3	441	1985	2734	3673
		4	450	1870	2732	3758
		5	401	1953	2810	3323
		6	388	1990	2848	3456
		7	460	1778	2581	3879
		8	454	2039	2701	3638
		9	465	1674	2716	3708
		10	483	1824	2434	3692
		11	430	1809	2483	3675
		12	474	1765	2288	3083
		13	417	1967	2921	3764
		14	471	1815	2521	3714
		15	493	1885	2787	3724
	Lari	1	539	2015	2554	2735
		2	505	1849	2306	3006
		3	541	2035	2689	3612
		4	462	1759	2456	3859
		5	466	1935	2583	3865
		6	500	1925	2659	3642
		7	493	1861	2626	3619
		8	497	1861	2608	3632
		9	422	1805	2659	3754
		10	472	2065	2611	3810
		11	511	2108	2776	3971
		12	470	1863	2550	3640
		13	492	2054	2740	3750
		14	510	2012	2752	3957
		15	506	1987	2717	4070
	Lasi	1	469	1921	2278	2648
		2	439	1935	2846	3681
		3	463	1984	2839	3075
		4	470	1841	2507	3963
		5	465	1804	2515	2831
		6	478	2153	2794	3843
		7	443	1927	2750	2790
		8	432	2008	2894	3752
		9	443	1854	1879	2867
		10	479	1968	2157	2867
		11	453	1999	2851	3729
		12	434	1979	2880	3736
		13	497	1671	2419	3760
		14	496	2087	2669	3813
		15	486	2095	2691	3771

Vowel	Dialect	No. of Speakers	F1 (Hz)	F2 (Hz)	F3 (Hz)	F4 (Hz)
/ə/	Middle	1	586	1552	2243	2758
		2	593	1494	2402	3937
		3	623	1544	2553	3382
		4	614	1535	2412	3376
		5	570	1417	2427	3683
		6	574	1483	2613	3613
		7	640	1496	2687	3916
		8	569	1467	2619	3754
		9	568	1453	2986	3819
		10	600	1430	2786	3695
		11	630	1385	1568	2609
		12	548	1567	2526	3794
		13	604	1580	2722	3828
		14	625	1445	2656	3800
		15	566	1514	2809	3772
	Utradi	1	635	1424	2846	3643
		2	669	1369	2779	3561
		3	662	1442	2655	3554
		4	569	1465	2913	3723
		5	536	1325	2464	3825
		6	561	1549	2681	3903
		7	532	1390	2384	3892
		8	518	1456	2368	3846
		9	579	1433	2674	3649
		10	545	1458	2391	3746
		11	575	1591	2717	3789
		12	624	1534	3016	3581
		13	625	1440	3017	3867
		14	610	1334	2357	3792
		15	579	1358	2751	3662
	Thareli	1	595	1461	2595	3802
		2	512	1299	2460	3562
		3	515	1350	2354	3541
		4	553	1296	2422	3649
		5	583	1471	2753	3834
		6	569	1511	2790	3733
		7	613	1397	3021	3844
		8	643	1410	2781	3508
		9	584	1383	2941	3827
		10	597	1394	2866	3811
		11	562	1382	2509	3576
		12	633	1442	2778	3636
		13	562	1317	2773	3616
		14	556	1308	2806	3724
		15	583	1380	2664	3846

Vowel	Dialect	No. of Speakers	F1 (Hz)	F2 (Hz)	F3 (Hz)	F4 (Hz)
/ə/	Lari	1	534	1533	2517	3690
		2	607	1456	2632	3882
		3	547	1498	2454	3277
		4	561	1326	2445	3918
		5	595	1347	2488	3921
		6	567	1385	2679	3649
		7	564	1382	2950	3826
		8	596	1552	2647	3765
		9	516	1294	2451	3644
		10	480	1435	2687	3455
		11	553	1410	2400	3496
		12	520	1443	2864	3720
		13	530	1405	2655	3900
		14	496	1306	2444	3635
		15	630	1518	2761	3531
	Lasi	1	553	1363	2378	3357
		2	561	1383	2293	3079
		3	569	1460	2729	3795
		4	612	1393	2606	3782
		5	581	1390	2702	3749
		6	554	1345	2659	3738
		7	525	1585	2874	3879
		8	545	1572	2840	3993
		9	588	1412	2728	3730
		10	598	1351	2838	3547
		11	601	1367	2807	3452
		12	554	1367	2325	3316
		13	566	1365	2370	2886
		14	549	1383	2701	3714
		15	562	1468	2630	3861
/ɑ/	Middle	1	760	1175	2788	4008
		2	732	1265	2399	3237
		3	683	1251	2392	3256
		4	745	1301	2534	3412
		5	764	1267	2570	3472
		6	813	1247	2661	3881
		7	809	1205	2725	3984
		8	765	1181	2522	3692
		9	724	1114	2257	2869
		10	669	1196	2750	3626
		11	695	1196	2710	3626
		12	824	1310	2721	4066
		13	825	1157	2720	3889
		14	831	1180	2595	3900
		15	691	1194	1664	2752

Vowel	Dialect	No. of Speakers	F1 (Hz)	F2 (Hz)	F3 (Hz)	F4 (Hz)
/ɑ/	Utradi	1	759	1194	2324	2805
		2	752	1231	2704	2961
		3	805	1130	2731	3764
		4	683	1269	2774	3737
		5	650	1175	2562	3553
		6	615	1075	2639	3723
		7	730	1235	2733	3727
		8	679	1197	2717	3962
		9	736	1172	2775	2978
		10	765	1022	1453	2745
		11	694	1024	2645	3652
		12	681	1196	2795	4017
		13	722	1092	2498	3493
		14	766	1125	2763	3080
		15	685	1068	2499	3567
	Thareli	1	795	1010	2855	3710
		2	818	1114	2828	3354
		3	695	1007	2844	3528
		4	695	1258	2686	3686
		5	682	1097	2661	3799
		6	639	1158	2667	3949
		7	671	1146	2676	3826
		8	710	1223	2716	4127
		9	699	1214	2590	4040
		10	713	1208	2691	4100
		11	749	1064	2564	3574
		12	744	1180	2533	3615
		13	612	1182	2758	3680
		14	640	1155	2612	3712
		15	652	1211	2541	3681
	Lari	1	753	1219	2470	3661
		2	730	1238	2449	3614
		3	640	1071	2308	3739
		4	789	1182	2214	3723
		5	655	1082	2326	3726
		6	652	1197	2822	3403
		7	815	1105	2555	2955
		8	837	1152	2627	3250
		9	692	1095	2479	3271
		10	671	1019	2553	3557
		11	746	1135	2964	4034
		12	689	1064	2554	2858
		13	640	1267	2880	3799
		14	681	1218	2639	3648
		15	732	1180	2661	3710

Vowel	Dialect	No. of Speakers	F1 (Hz)	F2 (Hz)	F3 (Hz)	F4 (Hz)
/ɑ/	Lasi	1	765	1203	2454	3706
		2	827	1248	2724	3687
		3	748	1193	2386	3668
		4	670	1181	2632	3654
		5	720	1223	2561	3642
		6	636	1030	2634	3555
		7	785	1152	2586	3643
		8	646	1100	2941	3502
		9	662	1129	2858	3444
		10	760	1315	2791	3937
		11	637	1159	2386	3665
		12	636	1163	2384	3622
		13	687	1160	2349	3355
		14	661	1229	2608	3733
		15	670	1235	2537	3615
/ɔ/	Middle	1	465	974	2405	3667
		2	395	910	2716	3588
		3	422	900	2706	3473
		4	405	959	2527	3184
		5	466	991	2483	3692
		6	464	947	2515	3632
		7	430	933	2180	3246
		8	415	957	2400	3696
		9	442	817	2150	3250
		10	497	928	2558	2706
		11	457	868	2678	3792
		12	466	890	2172	3150
		13	474	934	2610	3820
		14	463	950	2618	3893
		15	450	973	2384	3800
	Utradi	1	452	994	2269	3746
		2	485	950	2800	3540
		3	437	918	2687	3139
		4	502	943	2078	2722
		5	415	1025	2445	3700
		6	411	916	2508	3422
		7	499	1178	2709	3771
		8	500	1154	2718	3716
		9	480	1048	2455	3688
		10	454	1023	2446	3757
		11	493	1120	2675	3904
		12	433	960	2443	3954
		13	483	1017	2535	3553
		14	461	1052	2419	3668
		15	431	946	2397	3685

Vowel	Dialect	No. of Speakers	F1 (Hz)	F2 (Hz)	F3 (Hz)	F4 (Hz)
/ɔ/	Thareli	1	468	1079	2666	3875
		2	470	968	2545	3519
		3	450	925	2596	3950
		4	463	962	2442	3554
		5	444	965	2759	2921
		6	454	1007	2551	3540
		7	502	902	2550	3692
		8	375	808	2743	3575
		9	404	964	2767	3505
		10	463	955	2789	3515
		11	493	981	2699	3667
		12	495	1012	2604	3428
		13	398	888	2465	3435
		14	424	920	2560	3538
		15	446	944	2573	3512
	Lari	1	404	970	2383	3641
		2	440	998	2429	3775
		3	485	990	2461	3781
		4	448	959	2441	3701
		5	453	956	2609	3474
		6	492	1018	2412	3507
		7	487	985	2495	3835
		8	485	1008	2620	3763
		9	396	956	2633	3722
		10	431	916	2666	3744
		11	479	1006	2526	3574
		12	475	925	2575	3664
		13	436	986	2755	3640
		14	487	1032	2407	3812
		15	510	1010	2579	3555
	Lasi	1	474	969	2389	3715
		2	481	1042	2348	3607
		3	377	1019	2675	3611
		4	506	1125	2670	3774
		5	499	1063	2904	3580
		6	486	1003	2770	3814
		7	475	980	2176	3280
		8	484	1045	2378	3404
		9	466	966	2678	3538
		10	489	1041	2908	3606
		11	465	986	2955	3575
		12	404	970	2383	3641
		13	485	989	2597	3414
		14	464	1044	2531	3536
		15	490	1043	2130	3568

Vowel	Dialect	No. of Speakers	F1 (Hz)	F2 (Hz)	F3 (Hz)	F4 (Hz)
/o/	Middle	1	388	841	2791	3579
		2	389	829	2444	3595
		3	350	796	2759	3462
		4	332	743	2688	3602
		5	396	845	1960	3214
		6	405	844	2158	3197
		7	348	822	2792	3815
		8	370	785	2517	3661
		9	365	836	2766	3652
		10	389	861	2724	3675
		11	410	746	1794	2617
		12	415	789	1978	3339
		13	415	835	2576	3322
		14	429	842	2590	3317
		15	398	808	2688	3517
	Utradi	1	398	836	2293	3658
		2	371	883	2670	3678
		3	358	811	2287	3755
		4	413	813	2545	3609
		5	365	729	1020	2672
		6	369	765	2953	3282
		7	424	858	2673	3409
		8	349	755	2288	3785
		9	429	707	2839	3748
		10	379	789	3093	3715
		11	410	732	2790	3678
		12	384	822	3142	3824
		13	410	796	2821	2903
		14	370	770	2577	2965
		15	385	710	1646	2881
	Thareli	1	384	779	2819	3834
		2	345	845	2443	3533
		3	352	851	2501	3696
		4	390	887	2406	3617
		5	380	823	2675	3758
		6	371	901	2788	3774
		7	395	975	2442	3623
		8	342	783	2499	3428
		9	357	744	2869	3642
		10	359	735	2729	3360
		11	377	774	2844	3438
		12	385	803	2365	3582
		13	387	824	2344	3613
		14	416	816	2629	3325
		15	389	674	2620	3269

Vowel	Dialect	No. of Speakers	F1 (Hz)	F2 (Hz)	F3 (Hz)	F4 (Hz)
/o/	Lari	1	386	776	2599	3307
		2	433	844	2570	3355
		3	416	816	2629	3325
		4	401	791	2592	3360
		5	401	794	2523	3157
		6	416	851	2587	3205
		7	397	796	2625	3228
		8	419	862	2607	3310
		9	427	852	2509	3712
		10	415	812	2546	3067
		11	439	887	2647	3338
		12	429	842	2590	3317
		13	440	835	2576	3322
		14	422	842	2669	4001
		15	455	802	2688	3517
	Lasi	1	344	780	2900	3589
		2	389	981	2795	3727
		3	401	791	2592	3360
		4	386	703	2821	4340
		5	401	794	2523	3157
		6	416	851	2587	3205
		7	390	919	2756	3612
		8	362	809	2681	3635
		9	348	840	2714	3480
		10	353	787	2842	3507
		11	439	887	2647	3338
		12	349	763	2571	3627
		13	356	870	2532	3673
		14	388	816	2510	3516
		15	343	726	2550	3389
/ʊ/	Middle	1	331	745	2565	3685
		2	313	763	2664	3647
		3	329	695	2569	3715
		4	308	741	2670	3365
		5	297	806	2397	3426
		6	332	839	2754	3840
		7	312	748	2747	3456
		8	323	735	2726	3569
		9	313	734	2767	3528
		10	307	761	2710	3523
		11	330	746	2706	3557
		12	338	873	2623	3609
		13	295	715	2348	3427
		14	301	814	2513	3632
		15	322	887	2521	3730

Vowel	Dialect	No. of Speakers	F1 (Hz)	F2 (Hz)	F3 (Hz)	F4 (Hz)
/ʊ/	Utradi	1	283	761	2367	3216
		2	357	750	2654	3214
		3	378	770	2400	3677
		4	345	794	2339	3783
		5	309	719	2327	2980
		6	321	623	2721	3660
		7	340	893	2356	3579
		8	325	930	2486	3683
		9	305	823	2466	3711
		10	329	669	2510	3548
		11	294	699	2554	3657
		12	288	679	2606	3637
		13	324	801	2620	3647
		14	358	826	2367	3657
		15	306	723	2585	2915
	Thareli	1	295	820	2387	3686
		2	290	710	2556	3796
		3	325	931	2362	3342
		4	325	934	2821	3747
		5	348	923	2621	3211
		6	347	1008	2532	3331
		7	351	967	2489	3692
		8	345	1000	2530	3742
		9	360	932	2557	3843
		10	331	847	2361	3360
		11	318	665	2430	3172
		12	323	818	2685	3333
		13	328	835	1709	2916
		14	328	829	2716	3441
		15	311	832	2299	3688
	Lari	1	317	807	2442	3562
		2	340	762	2485	3531
		3	360	772	2279	3167
		4	347	893	2172	3093
		5	305	744	2272	3737
		6	363	914	2636	3865
		7	300	803	2479	3700
		8	319	831	2374	2978
		9	345	930	2530	3742
		10	355	839	2621	3328
		11	318	665	2430	3172
		12	295	761	2576	3397
		13	320	850	2449	3358
		14	329	783	2369	3336
		15	299	853	2672	3516

Vowel	Dialect	No. of Speakers	F1 (Hz)	F2 (Hz)	F3 (Hz)	F4 (Hz)
/ʊ/	Lasi	1	305	685	2822	4252
		2	289	756	2519	3674
		3	310	616	2208	3350
		4	290	665	2414	3543
		5	318	844	2691	3694
		6	305	736	2663	3647
		7	330	820	2660	3594
		8	277	757	2762	3386
		9	298	740	2529	3751
		10	314	810	2549	3811
		11	279	698	2599	3255
		12	303	765	2476	3698
		13	310	777	2537	3884
		14	318	843	2773	3656
		15	302	747	2463	3524
/u/	Middle	1	301	666	2379	3849
		2	241	600	2786	3045
		3	273	720	2258	3849
		4	235	749	2563	3652
		5	299	705	2719	3645
		6	314	701	2363	3437
		7	298	646	2447	3317
		8	250	669	2576	3276
		9	284	786	2335	3549
		10	294	748	2734	3639
		11	261	599	2399	3624
		12	282	596	1847	3257
		13	301	690	2720	3740
		14	289	661	2456	3601
		15	275	659	2234	3294
	Utradi	1	316	753	2375	3060
		2	305	583	2356	3287
		3	236	601	757	2716
		4	270	852	2510	3745
		5	232	681	2626	3701
		6	265	621	2866	3768
		7	309	700	2655	3625
		8	276	823	2370	3724
		9	260	698	2648	3673
		10	250	750	2335	3699
		11	245	808	2535	3569
		12	268	626	2424	3032
		13	265	762	2537	3487
		14	286	750	2420	3141
		15	255	708	2457	3027

Vowel	Dialect	No. of Speakers	F1 (Hz)	F2 (Hz)	F3 (Hz)	F4 (Hz)
/u/	Thareli	1	245	697	2550	3418
		2	282	795	2478	3814
		3	267	530	2715	3937
		4	220	635	2313	3400
		5	306	753	2419	3857
		6	269	734	2314	3824
		7	275	717	2280	3735
		8	214	621	2453	3148
		9	233	635	2960	3114
		10	212	568	2442	3075
		11	219	654	2422	3686
		12	279	647	2714	3112
		13	238	505	2625	3905
		14	275	609	2525	3771
		15	251	578	2896	3911
	Lari	1	279	696	2477	3514
		2	271	805	2516	3527
		3	310	799	2562	3113
		4	260	728	2575	3153
		5	313	769	2633	3388
		6	306	798	2539	3754
		7	270	777	2535	3693
		8	280	725	2527	3644
		9	256	684	2176	2995
		10	323	710	2877	3732
		11	288	810	2511	3531
		12	265	705	2715	3543
		13	312	687	2633	3480
		14	287	655	2496	3672
		15	278	622	2498	3546
	Lasi	1	255	486	1684	2534
		2	282	718	2060	3218
		3	274	674	2313	3253
		4	272	605	1125	2243
		5	308	660	2483	3518
		6	265	686	2387	3542
		7	296	709	2637	3670
		8	312	736	2672	3384
		9	285	759	2385	3687
		10	251	549	1965	2547
		11	294	510	1308	2424
		12	314	590	2750	3289
		13	285	713	2260	3465
		14	283	683	2662	3738
		15	272	604	1713	2705