Acoustic Analysis of Phonetics of Arabic Script Sindhi Language to evaluate Vowel-Consonant Segmentation

Muhammad Asif Khawaja and Dr. Najmi G. Haider
SZABIST
Karachi, Pakistan

Abstract:
This paper proposes an efficient speech recognition method for any spoken language of the world in general and Arabic script languages, including Arabic, Urdu, and Sindhi etc, in particular.

For the purpose, Sindhi has been selected as an example language, since it has a superset of all other Arabic script languages' phonemes and research has been conducted in two major areas including the definition and refinement of standard phonemes for Sindhi language comprising of vowels, semi-vowels, diphthongs, and consonants as defined by International Phonetic Association (IPA) and acoustic analysis of phonetics for Sindhi which includes analysis of waveforms, Linear Predictive Coefficient (LPC), and spectrographic characterizations, especially formants, of some of the phonemes, to identify the categorical properties of these phonemes and their boundary detection in an utterance. The objective is to provide a guideline and solid foundation for development of efficient speech recognition systems for Sindhi language in particular and all Arabic script languages in general.

1. INTRODUCTION

Sindhi is an Indo-Aryan language and is one of the major languages of Pakistan, spoken by approximately 40 million people in the country. It is one of the oldest languages of the sub-continent with a rich culture, vast folklore and extensive literature.

Sindhi is also a recognized official language of India, where it is spoken by approximately 1.2 million members of an ethnic group which migrated from the province of Sindh, Pakistan during the partition of British India in 1947 and settled in the central and western parts of India. Besides Pakistan and India, it is also spoken by approximately 4,00,000 people around the world.

Despite its importance, Sindhi language is still lacking robust implementations in the field of Information Technology especially in the area of speech recognition. The implementation of Sindhi language in Information Technology can be pursued in three major areas of Optical Character Recognition (OCR) for reading, Fonts and Text Editors for writing and Speech Recognition for speaking and listening.

Most of the work has been conducted in only the fonts and text editor development with support of True Type and Unicode character sets. OCR and Speech Recognition still need to be implemented. According to Sindhi Language Authority, Hyderabad, Sindh, no significant and documented work has been carried out in these two areas especially in Sindhi speech recognition.

2. THE SINDHI LANGUAGE

Sindhi is an Indo-Aryan language and is one of the major languages of Pakistan, spoken by approximately 40 million people in the province of Sindh and Lasbela (Baluchistan) regions of Pakistan [1]. It is one of the oldest languages of the sub-continent with a rich culture, vast folklore and extensive literature.

The evolution of Sindhi language is stretched to a period of over 2400 years, with 8 stages of migration of Scythians, people from Southern Iran. The language of the people of Sindh, after coming in contact with the Aryan, became Indo-Aryan (Prakrit). Sindhi language, therefore, has a solid base of Prakrit as well as Sanskrit, the language of India, with vocabulary from Arabic, Persian, and some Dravidian, descendants from Mediterranean sub-continent, also known as Moen-jo-Daro civilization. The script that is predominantly used in Sindh as well as in many states in India and elsewhere, where the migrant Sindhis have settled is in Arabic Nask, having 52 alphabets. However, in some of the circles in India, Devanagri, the Hindi script, has also been used as a script for writing Sindhi, although the vocal and oral style of speech remains same as in Sindh itself. [2]

Sindhi language has widened its boundaries beyond the Sindh province. In Northern Sindh it runs over the North West into the province of Baluchistan, to the Punjab and the former Bahawalpur state, on the west it is bounded by the mountain range separating Sindh from Baluchistan [1]. It has extended its influence still further towards the Persian Gulf, Maskat, Abu Dhabi, Kachh, Gujrat, Kaathiawaar, Maarwaar, Jaisalmir in India.

Sindhi is also one of the recognized official languages of India, where it is spoken by about 1.2 Million people majority of whom migrated from the province of Sindh (Pakistan), during the partition of British India in 1947 and settled in the central and western parts of India. Sindhi is also spoken by around 4,00,000 people as their first language, in Canada, U.S.A, U.K, East Africa, South Africa, Congo, Uganda, Madaagascar, Kenya, and Tanzaania, and by those who have migrated from Sindh and settled there. It is also spoken in Sri Lanka, Thailand, Singapore, and Hong Kong and in some other countries in Far East and South East Asia.

2.1 Sindhi Alphabet

The Sindhi alphabet is a super set of Urdu, Persian, and Arabic languages with 52 alphabets in total as shown in Table 1. Additionally, a part from the basic punctuation characters and numbers, it has some special characters like ۽ "and" and ۾ "in". The graphic writing representation of each alphabet has more than one form depending on its position. In general each letter has four forms: beginning, middle, final, and standalone.

2.2 Institutions Promoting Sindhi Language

There are several institutions that are promoting Sindhi language and cultural heritage in Indo-Pak including Institute of Sindhology, Jamshoro, Sindh, Pakistan [3], The Indian Institute of Sindhology, Adipur, India [4], and Sindhi Language Authority, Hyderabad, Sindh, Pakistan [1].

2.3 Sindhi Language and Information Technology

The implementation of Sindhi language in Information Technology can be pursued in three major areas of Optical Character Recognition (OCR) for reading, Fonts and Text Editors for writing and Speech Recognition for speaking and listening.

Out of these three areas most of the work has been conducted in only the fonts and text- editor development with support of True Type and Unicode character sets. OCR and Speech Recognition still need to be implemented. According to Sindhi Language Authority, Hyderabad, Sindh, no significant and documented work has been carried out in these two areas especially in Sindhi speech recognition [5].

However, there has been a lot of work done in Sindhi computing which ranges from keyboard and font standardization to utility software development, including text editing, database management, web site development, emailing, chatting, text compression, text editors, dictionaries, newspaper composing, and agro-MIS systems etc. [1], [6], [7], [8], and [9].

3 PHONETICS OF SINDHI LANGUAGE

3.1 Phonetics and Phonology

Phonetics is the study of speech sounds. It is concerned with the actual nature of the sounds and their production i.e. how speech sounds are actually made, transmitted, and received, while phonology operates at the level of sound systems and linguistic units called phonemes. Phonology, in fact, is a sub-category of phonetics. Phonetics was studied as early as 2500 years ago in ancient India. [10]

Phonetics has three main branches [10]:

- Articulatory phonetics is concerned with the positions and movements of the lips, tongue, and other speech organs in producing speech.
- Acoustic phonetics is concerned with the properties of the sound waves.
- Auditory phonetics, concerned with speech perception.

3.2 Acoustic Phonetics of Sindhi

Most languages, including Sindhi, can be described in terms of a set of distinctive sounds, or phonemes. In particular, for Sindhi language, there are about 50 phonemes including 38 consonants, 3 semi-vowels, 8 vowels, and one diphthong as shown in Table 3.

The table shows how the sounds of Sindhi are broken into phoneme categories. The four broad categories of sounds are vowels, diphthongs, semivowels, and consonants. Each of these classes can be further broken down into sub-categories which are related to manner, and place of articulation of the sound within the vocal tract.

3.3 Phonetics of Sindhi Language by IPA

The aim of the International Phonetic Association (IPA) is to promote the study of the science of phonetics and the various practical applications of that science. For both these it is desirable to have a consistent way of representing the sounds of language in written form. From its foundation in 1886 the Association has been concerned to develop a set of symbols which would be convenient to use, but comprehensive enough to cope with the wide variety of sounds found in the languages of the world and to encourage the use of this notation as widely as possible among those concerned with language. The system is generally known as the International Phonetic Alphabet, a notational standard for the phonetic representation of all languages [11].

3.3.1 Classification of Consonant Phonemes

IPA has classified phonetic symbols for Sindhi consonant system which consists of 12 stops or plosives (including 4 implosive stops), 8 aspirates, 5 nasals, 6 fricatives, 2 affricates, 2 retroflex, 1 lateral, and 2 semivowels. [11] Table 4, presents the author's reformatted version of these symbols along with the corresponding Sindhi sounds. The row highlighted in yellow shows the increment made by author in [11]'s work which will be discussed in following sections. Table 2 lists some of the examples of consonant phonemes by IPA.

3.3.2 Classification of Vowel Phonemes

IPA has also classified phonetic symbols for eight-vowel system of Sindhi, showing three-fold contrast in the tongue-position; front, central and back; and four-fold contrast in the tongue-height; high, lower-high, mid and lower-mid. See Table 5. Additionally, two diphthongs, which combine sounds of two vowels, have also been defined and are shown in Table 6.

The two diphthongs generate a sound which starts with one vowel and end at another, as /əɛ/ and /əʊ/. Table 7, exemplifies the IPA symbols for 8 vowels and 2 diphthongs with some Sindhi words. For each vowel in Sindhi, a corresponding nasalized version of vowel also exists.

3.4 Refinement to Phonetics of Sindhi Language

Although the phonetics defined by IPA is covering all the aspects of phonetics of Sindhi language but based on certain observations, author is suggesting some enhancements to it for two sounds of Sindhi language that IPA has not covered, perhaps because the speech samples that IPA recorded of a Sindhi speaker, Paroo Nihalani, who grew up in Sindh but moved to India in 1947 [12], had no such sounds in them. In fact, these two sounds are variations of two of the phonemes that IPA has already defined.

For these sounds, the same Sindhi alphabets are used in writing but the sounds are totally different and seem like a mix of plosives and retroflex. Following table shows the examples of these two sounds and their comparison with IPA corresponding phonemes.

Table: Two new consonant phonemes suggested by author

IPA Symbol	Sindhi Alphabet	Example Word	IPA Transcription	English Meaning
ʈ	ٽ	پَٽُ	patu	floor
-	ٽ	پَٽُ	-	(metallic strip)
ɖ	ڊ	ڊَپُ	dapu	fear
-	ڊ	ڊَبُ	-	bush

For the purpose of verification of these sounds, author recorded several speech samples of different people which contained these sounds.

The place and manner of articulation for these two phonemes are discussed in following sections. Table 4 is the classification of Sindhi consonant phonemes as compiled by the author and refinement highlighted in yellow.

3.5 Articulation of Sindhi Phonemes

Sindhi language has the most comprehensive stop system of any of the Indo-Aryan languages. The stop series has got the contrast between voicing and un-voicing, aspiration and pressure, and suction. It has a series of four implosive stops, ٻ (/ɓ/), ڏ (/ɗ/), ڄ (/ʄ/), and ڳ (/ɠ/); in sounding them breath is drawn in instead of being expelled as in ب (/b/), ڊ (/d/), ج (/ɟ/), and گ (/g/) which is a striking characteristic of Sindhi phonology. Table 10 describes the place of articulation for consonants along with the method of their speech production.

In Sindhi, و (/ʋ/), ي (/j/), and ح (/h/) function similarly to consonants in initial and certain medial positions. But in final positions and also medially when preceding or following a consonant, these occur as vocalic glides; thus forming diphthongs with preceding or following vowels; these are classified as semivowels. Table 11 describes ten different manners of articulation for all consonants (including the refined ones) and semivowels along with the level and location of obstruction of the air-stream required for each phoneme.

4 ACOUSTIC ANALYSIS OF SINDHI PHONETICS

4.1 Selection of Sindhi Speech Sounds

Sindhi language has one of the richest collections of sounds in all Arabic script languages of the world. Since the major concentration of this study was on the analysis of Sindhi vowels and their characteristics, for their identification and boundary detection in a spoken word, it covers only vowels, and not consonants.

Although the study discusses vowels in general, but the special attention has been given to the analysis of the vowel /a/ because it is different from all English vowels and one of the most frequently used vowels in Sindhi language. Table 8 provides the list of Sindhi words selected for this study along with the vowels that they contain, their pronunciations, and their English translations.

4.2 Collection of Speech Samples

Several Sindhi language words with specific vowels were selected as listed in Table 8.

4.2.1 Speech Sample Format

The words were recorded using Microsoft ® Sound Recorder Version 5.0 in Microsoft PCM format with 1 channel (mono), a sampling frequency of 22KHz (22050 samples per second) with 16 bits per sample, and a bit rate of 43Kbytes (44100 bytes per second). The operating system used was Microsoft ® Windows 2000.

4.2.2 Speakers

The speech samples were recorded from four people, 2 males and 2 females so that the detailed analysis of speech sounds of different people could be performed. The male people included author himself (MAK) and one of his male colleagues at SZABIST (APM). The female people included author's wife (SN) and one of author's female colleagues at SZABIST (FN).

4.2.3 Environment

All the samples were recorded in a quite office environment with a minor background noise of air conditioner installed in the room.

4.3 Acoustic Analysis of Speech Samples

4.3.1 The Main Idea

As mentioned earlier that each phoneme of any speech utterance has unique formant frequency positions and can be isolated and hence identified by looking at the formants positions and behaviors. But as mentioned earlier, it is difficult to detect the boundaries of different phonemes in a speech signal that is changing smoothly over time and not abruptly, and hence those phonemes can not be recognized. This is the reason that most speech recognition systems, specially isolated word recognizers, recognize speech by comparing the whole utterances (words) with the already stored templates generated through training, which is a very time consuming process.

As vowels can be easily identified by looking at the positions and values of the formants, as will be demonstrated during the analysis of vowels in the forth coming sections, their boundary detection and identification in an utterance can help in identifying other parts of the speech, that is, the consonants and can provide a way to identify them as well to some extent and hence speed up the performance of the recognition system.

This can be achieved initially by converting the utterance into a string of CVC… (for Consonant Vowel Consonant) by detecting the boundaries of the phonemes using vowels and their formant frequencies. Next, using the same formant frequencies the vowels can be identified (as they are easier to identify). Once vowels are identified and isolated, the consonants in the utterance will be identified using formants and other features. If all the CVC combinations in an utterance are recognized, an output in the form of written word or some process execution will be generated. On the other hand, if some of the consonant parts of the utterance are not recognized, then the template library will be searched for only those templates which have the CVC combination and the utterance will be matched with the required template to recognize the word. The author terms this process of recognizing an utterance as 'divide-and-conquer recognizer' because it divides the whole utterance into several smaller parts of CVC and then individually tries to identify each part and one which is not recognized is located from template library. This speech recognition process will boost up the performance of any speech recognition system drastically.

Although, author has suggested a method to implement above recognition process for Sindhi language in the last section, the study's focus is on the boundary detection and identification of only the vowel phonemes, and not consonants, for particular speakers only (i.e. speaker dependent).

4.3.2 Formants Data Generation

The basis of the acoustic analysis of Sindhi speech samples in this study, is the formants data which is the values of first three formant frequencies generated over time after every 20 milliseconds.

Colea, a tool for Matlab [13-15] was used to generate this formant data. Following is the process performed to generate the formant data of all speech samples collected for this study. The process shows formant data generation for only one speech sample, "ٻارَ" ("barə") meaning "children", spoken by the speaker MAK.

- Start the Matlab application and run the Colea software in it.
- Load the .wav file with the speech sample.
- Click on the menu item "Display" and select "Formant track". A window titled "Formant Tracks" will appear showing a track of the first three formant frequencies (in Hz) over time (in msecs).
- From the 'Formants Tracks' window select 'Save Formants' menu option. This will enable Colea to save all formant data of first three formants for this speech sample to be saved in a file with extension of .frm. The saved file contains a table with three columns, t(msec), F1(Hz), F2(Hz), and F3(Hz). The values have been calculated after every 20 milliseconds. Table 9 illustrates the contents of the saved .frm file.

4.3.3 Identification of Formant Ranges and Boundary Detection for Selected Vowels

4.3.3.1 Same Vowels, Same Words

To start with the analysis of Sindhi vowel phonemes and to identify their formant ranges, author selected one word "سارَ" ("sarə") meaning "care" with selected vowels "آ" /a/ and "اَ" /ə/ and recorded its sample three times from the four different speakers, MAK, APM, FN, and SN, as mentioned in Section 4.2.2. The emphasis was on the formant ranges for individual speakers (i.e. speaker dependent).

Firstly, MAK's speech sample was evaluated. Figure 1 shows the spectrogram of the first utterance of the selected sample "sarə", LPC spectra of the vowel phoneme /a/, and the formant track for the utterance.

By evaluating the three .frm files of the three samples of same word, from the same speaker (MAK), the ranges of the three formants for the vowel /a/ were generated as illustrated in Table 12(a). Note that the ranges of the three formants are almost same. Table 12(b) shows the optimum ranges and average values of the three formants for the vowel /a/ and for the speaker MAK. Using Table 12(b) the boundaries of vowel /a/ in a speech sample of MAK can be detected easily.

Similarly, the ranges of the three formants for the vowel /ə/ were generated as illustrated in Table 13(a). Note that the ranges of the three formants are again almost same. Table 13(b) shows the optimum ranges and average values of the three formants for the vowel /ə/ and for the speaker MAK. Using Table 13(b) the boundaries of vowel /ə/ in a speech sample of MAK can also be detected easily.

Second, speaker APM's speech sample was evaluated. As with MAK's sample, spectrogram, LPC spectra, formant track, and .frm files were generated with APM's speech sample also.

By evaluating the three .frm files of the three samples of same word, from the same speaker (APM), the ranges of the three formants for the vowel /a/ were generated as illustrated in Table 14(a). Note that the ranges of the three formants are almost same. Table 14(b) shows the optimum ranges and average values of the three formants for the vowel /a/ and for the speaker APM. Also note that these ranges are different from the ones that were generated for MAK and show an overall shift in values. Using Table 14(b) the boundaries of vowel /a/ in a speech sample of APM can be detected easily.

Similarly, the ranges of the three formants for the vowel /ə/ were generated as illustrated in Table 15(a). Note that the ranges of the three formants are again almost same. Table 15(b) shows the optimum ranges and average values of the three formants for the vowel /ə/ and for the speaker APM. Using Table 15(b) the boundaries of vowel /ə/ in a speech sample of APM can also be detected easily.

Next, FN's (one of the female speakers) speech sample was evaluated. Figure 2 shows the spectrogram of the first utterance of the selected sample "sarə", LPC spectra of the vowel phoneme /a/, and the formant track for the utterance.

By evaluating the three .frm files of the three samples of same word, from the same speaker (FN), the ranges of the three formants for the vowel /a/ were generated as illustrated in Table 16(a). Note that the ranges of the three formants are almost same. Table 16(b) shows the optimum ranges and average values of the three formants for the vowel /a/ and for the speaker FN. Also note that the formant ranges for female speaker are bit higher that male speakers, specially the first formant F1. Using Table 16(b) the boundaries of vowel /a/ in a speech sample of FN can be detected easily.

And lastly, SN's (another female speaker) speech sample was evaluated. As with FN's sample, spectrogram, LPC spectra, formant track, and .frm files were generated with SN's speech sample also.

By evaluating the three .frm files of the three samples of same word, from the same speaker (SN), the ranges of the three formants for the vowel /a/ were generated as illustrated in Table 17(a). Note that the ranges of the three formants are almost same. Table 17(b) shows the optimum ranges and average values of the three formants for the vowel /a/ and for the speaker SN. Also note that these ranges are different from the ones that were generated for FN and show an overall shift in values but again are higher than both male speaker formants. Using Table 17(b) the boundaries of vowel /a/ in a speech sample of SN can be detected easily.

It is very much clear from the above analysis of formants for the speech sample "sarə" for four different speakers that the formants ranges and averages can be comfortably used for the identification and boundary detection of Sindhi vowel phonemes and CVC segmentation is possible for specific speakers only.

4.3.3.2 Same Vowel, Different Words

In the next phase of analysis, eight different speech samples with same vowel /a/ from the four speakers were recorded. Table 18 lists those eight speech samples.

Firstly, MAK's speech samples were evaluated. For all the eight samples, author generated spectrograms, LPC spectra, formant tracks, and .frm files using Colea. By evaluating the eight .frm files of the eight samples of different words but with same vowel /a/, from the same speaker MAK, the ranges of the eight formants for the vowel /a/ were generated as illustrated in Table 19(a). Note that the ranges of the three formants are almost same for all the samples. Table 19(b) shows the optimum ranges and average values of the three formants for the vowel /a/ and for the speaker MAK. Also note that these ranges and averages are almost same as for previous MAK sample ranges and averages. Again using Table 19(b) the boundaries of vowel /a/ in a speech sample of MAK can be detected easily.

Secondly, APM's speech samples were evaluated. For all the eight samples, author generated spectrograms, LPC spectra, formant tracks, and .frm files using Colea. By evaluating the eight .frm files of the eight samples of different words but with same vowel /a/, from the same speaker APM, the ranges of the eight formants for the vowel /a/ were generated as illustrated in Table 20(a). Note that the ranges of the three formants are almost same for all the samples. Table 20(b) shows the optimum ranges and average values of the three formants for the vowel /a/ and for the speaker APM. Also note that these ranges and averages are almost same as for previous APM sample ranges and averages and show an overall shift in values as compared to MAK's ranges and averages. Again using Table 20(b) the boundaries of vowel /a/ in a speech sample of MAK can be detected easily.

Thirdly, FN's (female speaker) speech samples were evaluated. For all the eight samples, author generated spectrograms, LPC spectra, formant tracks, and .frm files using Colea. By evaluating the eight .frm files of the eight samples of different words but with same vowel /a/, from the same speaker FN, the ranges of the eight formants for the vowel /a/ were generated as illustrated in Table 21(a). Note that the ranges of the three formants are almost same for all the samples. Table 21(b) shows the optimum ranges and average values of the three formants for the vowel /a/ and for the speaker FN. Also note that these ranges and averages are almost same as for previous FN sample ranges and averages. Again using Table 21(b) the boundaries of vowel /a/ in a speech sample of FN can be detected easily.

And lastly, SN's (another female speaker) speech samples were evaluated. For all the eight samples, author generated spectrograms, LPC spectra, formant tracks, and .frm files using Colea. By evaluating the eight .frm files of the eight samples of different words but with same vowel /a/, from the same speaker SN, the ranges of the eight formants for the vowel /a/ were generated as illustrated in Table 22(a). Note that the ranges of the three formants are almost same for all the samples. Table 22(b) shows the optimum ranges and average values of the three formants for the vowel /a/ and for the speaker SN. Also note that these ranges and averages are almost same as for previous SN sample ranges and averages and show an overall shift in values as compared to FN's ranges and averages. Again using Table 22(b) the boundaries of vowel /a/ in a speech sample of SN can be detected easily.

Now it is very much confirmed from the above analysis of formants for the eight different speech samples for four different speakers that the formants ranges and averages can be comfortably used for the identification and boundary detection of Sindhi vowel phonemes and C V C segmentation is possible for specific speakers only.

4.4 The Vowel Pyramid

From previous sections it is already clear that using formant frequencies of any speech signal for a particular speaker, the vowels' boundary detection can be easily performed. Based on these results, author has defined formant ranges for all the eight vowel phonemes of Sindhi language for a particular speaker, that is, MAK in this case. For the purpose, author recorded eight different sound samples with different words and different vowels twice. Twice because the author wanted to make sure that the formants are same. Table 23 lists those eight speech sample words.

The author then evaluated all the eight samples twice in the same way as the previous samples were done, and spectrograms, LPC spectra, formant tracks, and .frm files were generated for them using Colea. By evaluating the eight .frm files of the eight samples of different words with different vowels twice, from the same speaker MAK, the formant ranges for those vowels were generated as illustrated in Tables 24(a) and 24(b). Table 25 shows the average values of the eight formants for the eight different vowels of Sindhi language for the speaker MAK, generated from Tables 24(a) and 24(b).

Based on the average formant frequencies shown in above table, author has developed a 3D plot, termed as "The Vowel Pyramid" and shown is Figure 3, of the first, second, and third formant frequencies on x, y, and z-axes respectively, for the Sindhi vowel phonemes.

At the upper left hand corner of the pyramid is the vowel /i/ with a low first formant and high second formants. At the lower left hand corner is the vowel /u/ with low first and second formants. The third corner of the pyramid is the vowel /a/ with a high first formant and a low second formant. All the other vowels' positions are also clearly visible from the pyramid.

5 CONCLUSION

Efforts have been put forward to explore the phonetics of Sindhi language defined by International Phonetic Association (IPA) and to suggest certain refinements for enhancing its scope. The aim was to propose a complete phonetic system for Sindhi language which could be used by different organizations working in the domain e.g. those publishing Sindhi-to-other-languages dictionaries for defining the pronunciations of the words. The author hopes that the phonetic system of Sindhi language defined here will be fully utilized by organizations and individuals to make Sindhi language read and write in a better way.

Besides defining Sindhi phonetics, a detailed acoustic analysis of the phonetics was also performed. The areas of acoustic analysis, speech sample collection, analysis of phonetic features including formants, formants data generation, identification of formant ranges and certain formant behaviors, and vowel boundary detection have been covered in the course of study. It is concluded that all the formant frequencies of the Sindhi vowels, spoken by a particular speaker, will always fall around the vowel pyramid boundaries defined by author resulting in easy identification of a vowel in an utterance.

From the detailed analysis performed in the study, it is evident that this method of identification and boundary detection works for the specific speakers individually. Based on this method a speaker dependent speech recognition system can be designed to perform Sindhi vowel identification.

Additionally, the ranges and average values of the formants can easily help in finding the vowels' starting and ending positions in a speech signal for boundary detection, and finally for building CVC strings, as suggested in the earlier chapters of the study, to identify the structure of the speech signal, and ultimately, recognize it.

6 FUTURE WORK

One important aspect of speech recognition that has not been addressed in this study is the noise factor. Noise issue has always been a major obstruction in the development of speech recognition technology but author presumes that this important issue could also be solved through formant frequency analysis.

Additionally there is a much need to work on following aspects of speech recognition for Sindhi language.

6.1 Speaker Independence

It has not been studied that how the vowel identification and boundary detection system that works for specific speakers will work in any speaker independent environment. One approach to achieve it could be to normalize the formant data to some optimum level before deciding about the vowel and its boundaries. Another could be, through behavioral analysis of the formant frequencies which involves the investigation that how the formant values change from higher to lower positions and vice versa rather than the formant values itself.

6.2 Male and Female Identification

Based on the formants ranges, their normalization, and behavioral analysis, male and female speaker identification can also be performed easily.

6.3 Implementation Model for Sindhi Vowel-Consonant Segmentation and Recognition

The target of efficient Sindhi (or any other Arabic script language) speaker independent speech recognition can be accomplished by presenting some implementation model of vowel consonant segmentation presented in this study.

Model suggested by author is outlined as follows:

- Speech Signal Capture
- Formant Data Generation
- Formant Normalization
- Formant Template Generation
- Formant Template Analysis
- CVC Boundary Detection
- CVC String Generation
- Vowel Identification
- Consonant Identification
- Intelligent Pattern Matching
- Speech Recognition

REFERENCES

[1] Sindhi Language Authority, Hyderabad, Sindh, Pakistan, http://www.sindhila.com.pk

[2] Ahmed H Makhdoom, Makhdoom International Management Centre (MIMC), Singapore, http://pachome1.pacific.net.sg/~makhdoom/

[3] Sindh Today, The Internet Newspaper, Issue 58, May 2004, http://www.sindhtoday.net/institutions.htm

[4] Indian Institute of Sindhology, Adipur 370 205 (Kutch), India, www.sindhology.com

[5] Kashif Kazi, Audio Visual Officer, Sindhi Language Authority (SLA), Hyderabad, Sindh, Pakistan, January 2004.

[6] Abdul Majid Bhurgri, majidbhurgri@yahoo.com, a US based renowned Software Professional, voluntarily working for Sindhi Computing and its standardization and runs MSN Group on Sindhi Computing, URL: http://groups.msn.com/SindhiComputing/home.htm.

[7] Dr. Mohammad Qasim Bughio, *Two Years' Progress Report (August 2001 to July 2003) of Sindhi Language Authority*, 2003.

[8] SZABIST (www.szabist.edu.pk), A proposal prepared on *Standardization and Bidirectional Translation of Pakistan Languages*, submitted to Ministry of Science and Technology (MoST), January 2001.

[9] Aslam Parvez Memon and Ayaz Uqaily, *Study of Unicode Specifications and their Implementation in Arabic Script Langauges by designing a Multi-lingual Unicode Editor*, Proceedings of International Multi-topic Conference (INMIC), 2001. p 229-233., LUMS, Lahore Pakistan.

[10] Wikipedia, The Free Encyclopedia, *Phonetics*, http://wikipedia.org

[11] *Handbook of the International Phonetic Association, A Guide to the Use of the International Phonetic Alphabet*, Cambridge University Press, 1999; reprinted in 2002.

[12] *Handbook of the International Phonetic Association, A Guide to the Use of the International Phonetic Alphabet*, Cambridge University Press, 1999; reprinted in 2002; page 131

[13] Philip Loizou, *COLEA: A MATLAB software tool for speech analysis*, Dept. of Electrical Engineering, University of Texas at Dallas, 1989-99 http://www.utdallas.edu/~loizou/speech

[14] http://www.utdallas.edu/~loizou/cimplants/

[15] http://www.utdallas.edu/~loizou/speech/colea.htm

APPENDIX

Table 1: Sindhi Alphabet

ٿ	ت	ڀ	ٻ	ب	ا
ج	ڦ	پ	ث	ٺ	ٽ
ح	ڇ	چ	ڃ	جھ	ڄ
ڏ	ڊ	ڍ	ڌ	د	خ
ش	س	ز	ڙ	ر	ذ
غ	ع	ظ	ط	ض	ص
ڳ	گ	ک	ڪ	ق	ف
ڻ	ن	م	ل	ڱ	گھ
		ي	ء	ه	و

Table 2: Examples of some of the Sindhi Consonants

IPA Symbol	Sindhi Alphabet	Example Word	Transcription	English Meaning
ɳ	ڻ	وَڻُ	waɳu	tree
r,	ر	رولُو	roloo	wanderer
ɽ	ڙ	ساڙُ	saaɽu	jealousy
l	ل	ليمو	leemo	lemon
ʄ	ڄ	ڄارو	ʄaaro	cobweb
ʂ	ش	شيء	ʂey	thing

Table 3: Phonemes in Sindhi Language

Vowels			Diphthongs	Semi-vowels	Consonants							
Front	Mid	Back			Aspirates	Plosives	Fricatives	Affricates	Nasals	Lateral	Retroflex	Implosive Stops
اِي (ee)	اَ (a)	اُو (oo)	ؤ (au)	و (w)	ڀ (bh)	ب (b)	ف (f)	چ (ch)	م (m)	ل (l)	ر (r)	ڳ (gg)
اِ (e)		اُ (u)	يء (ay)	ي (y)	ڦ (ph)	پ (p)	خ (kh)	ج (j)	ن (n)		ڙ (rh)	ڄ (jj)
اَي (ey)		اَو (o)		ح, ھ (h)	ڍ (dhh)	د (dh)	غ (gh)	جھ (jh)	ڻ (nr)			ڍ (dd)
	آ (aa)				ٿ (thh)	ت, ط (th)	ث, س, ص (s)	ڇ (chh)	ڃ (nj)			ٻ (bb)
					ڏ (dh)	ڊ (d)			ڱ (ng)			
					ٺ (tth)	ٽ (t)	ظ, ذ, ز ض (z)					
					گھ (ghh)	گ (g)						
					ک (khh)	ق,ڪ (k)	ش (sh)					

Table 4: Classification of Consonants refined by the author

Manner of Articulation	Voicing	Place of Articulation											
		Bilabial		Labiodental		Interdental		Alveolar		Alveopalatal		Velar	
Aspirates	Voiced	b^{ɦ}	ڀ			D^{ɦ}	ڍ	ɖ^{ɦ}	ڏ			g^{ɦ}	گھ
	Un-voiced	p^{h}	ڦ			t^{h}	ٿ	ʈ^{h}	ٺ			k^{h}	ک
Plosives	Voiced	b	ب			D	د	ɖ	ڊ			g	گ
	Un-voiced	p	پ			T	ت, ط	ʈ	ٽ			k	ق, ڪ
Implosive Stops		ɓ	ٻ					ɗ	ڍ	ʄ	ڄ	ɠ	ڳ
Fricatives	Voiced							z	ظ,ض, ذ ,ز			ɣ	غ
	Un-voiced			f	ف			s	ص ,ث ,س	ʂ	ش	x	خ
Affricates	Voiced									ɟ, ɟ^{ɦ}	جھ , ج		
	Un-voiced									c, c^{h}	ڇ , چ		
Nasals		m	م					n, ɳ	ڻ , ن	ɲ	ڃ	ŋ	ڱ
Retroflex	Voiced							r, ɽ	ڙ , ر				
Retroflex Plosives	Voiced							ɖ-	ڊ_				
	Un-voiced							ʈ-	ٽ_				
Laterals								l	ل				
Semivowels				ʋ	و					j	ي		

ournal of Independent Studies and Research (JISR)

Table 5: Classification of Vowels by IPA

Tongue Position	Front Position (unrounded lips)		Center Position (unrounded lips)		Back Position (rounded lips)	
High	اِي	i			اُو	u
Lower High	اِ	ɪ			اُ	ʊ
Mid	اَي	e	اَ	ə	اَو	o
Lower Mid			آ		A	

Table 6: Diphthongs of Sindhi as defined by IPA

يء	ɛ
ؤ	ɔ

Table 7: Examples of IPA symbols for Sindhi vowels and diphthongs

IPA Symbol	Sindhi Vowel	Example Word	IPA Transcription	English Meaning
i (see)	اِي	سِيرَ	sirə	midstream
ɪ (it)	اِ	سِرَ	sɪrə	brick
e (pen)	اَي	سَيرَ	serə	(a measure of weight)
ɛ (ai)	يء	سَيرَ	sɛrə	walk
a (aa)	آ	سارَ	sarə	care
ə (cut)	اَ	سَرَ	sərə	funeral
ɔ (au)	ؤ	ڀؤ	bɦɔ	fear
o (core)	اَو	سوڙھ	sorə	congestion
ʊ (put)	اُ	سُرَ	sʊrə	tunes
u (boot)	اُو	سُورَ	surə	aches and pains

Table 8: List of selected Sindhi words

Word	Pronunciation	Vowel	Translation
جانِ	Janɪ	a	Life
ٻارَ	barə	a	Children
يارَ	jarə	a	Friends
وارَ	varə	a	Hair
ساڙُ	sarʊ	a	Jealousy
ھاٿِي	$hat^{h}i$	a	Elephant
ڄارو	Jaro	a	Cobweb
ڏارُ	darʊ	a	Crack
ڪارا	kara	a	Black

Table 9: Contents of the saved .frm file with formant values for the sample "ٻارَ".

t(msec)	F1(Hz)	F2(Hz)	F3(Hz)
1	177.536	2917.642	3340.985
21	166.18	3071.857	3340.985
41	147.525	800.773	2910.945
61	226.614	740.295	2834.678
81	408.46	932.204	2822.556
101	455.87	940.442	2910.234
121	500.96	953.779	2894.446
141	499.242	933.014	2861.765
161	461.957	896.605	2849.318
181	464.043	910.622	2813.987
201	444.79	890.793	2860.92
221	406.823	885.39	2848.976
241	416.709	899.295	2787.603
261	453.309	1017.487	2773.829
281	435.26	1041.181	2682.918
301	407.805	1074.216	2406.911
321	408.193	1355.606	3320.539
341	166.214	1551.403	3339.257
361	279.377	1493.917	3383.053
381	388.698	1564.846	3410.328
401	379.821	1385.544	2765.627
421	421.03	1294.764	2596.156
441	443.331	1368.775	2675.015
461	359.329	1211.262	2546.476
481	233.914	1278.289	2750.218
501	167.719	1578.415	3340.985

Table 10: Place of Articulation of Sindhi Consonant Phonemes

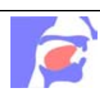
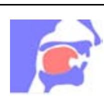
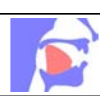
Articulation	Place of Articulators	Description	Classification
Bilabial		Speech Production Place of articulation	Both lips (bi ≡ both, labial ≡ lips) Lips come together and touch momentarily thereby obstructing the air stream from the lungs.
Labiodental		Speech Production Place of articulation	Lower lip and upper teeth. (labial ≡ lips, dental ≡ teeth) Bottom lip and the top teeth touch again obstructing the air stream from the lungs.
Interdental		Speech Production Place of articulation	Tip of the tongue and both upper and lower teeth (inter ≡ between, dental ≡ teeth) Air stream obstructed due to the tip of the tongue located between or slightly behind the teeth.
Alveolar		Speech Production Place of articulation	Tip of tongue and roof of the mouth (alveolar ≡ tooth ridge behind teeth) Air stream obstructed due to the tip of the tongue approaching or touching the alveolar ridge located on the roof of the mouth slightly behind the teeth.
Alveopalatal		Speech Production Place of articulation	Blade of the tongue and the hard palate slightly behind the tooth ridge. Lips rounded. (alveo ≡ ridge, palatal ≡ hard palate) Air stream obstructed due to the blade of the tongue approaching the hard palate on the roof of the mouth slightly behind the alveo ridge.
Velar		Speech Production Place of articulation	Back of tongue and the soft palate. (velar ≡ soft palate, back-roof of mouth) Air stream obstructed due to the back of the tongue rising to touch the soft palate (or velar) on the back of the roof of the mouth.

Table 11: Manners of Articulation for all Sindhi Consonant and Semi-vowel Phonemes

Articulation	Description	Classification
Aspirates	Level of Obstruction Phonemes: /ڀ/, /ڦ/ /ٺ/, /ڍ/, /ک/, /گھ/ /ڌ/, /ٿ/	Complete obstruction of the air stream followed by a plosive air blow. Location of obstruction: lips (bilabial) tongue & tooth ridge (alveolar) back of tongue and soft palate (velor) tip of tongue and teeth (interdental)
Plosives	Level of Obstruction Phonemes: /پ/, /ب/ /ٽ/, /ڊ/ /گ/, /ڪ/, /ق/ /د/, /ت/, /ط/	Complete obstruction of the air stream Location of obstruction: lips (bilabial) tongue & tooth ridge (alveolar) back of tongue and soft palate (velor) tip of tongue and teeth (interdental)
Implosive Stops	Level of Obstruction Phonemes: /ڳ/ /ڄ/ /ڏ/ /ٻ/	Complete obstruction of air stream. Unlike stops where air is expelled, breath is drawn in. Location of obstruction: back of tongue and soft palate (velor) back of tongue & hard palate (alveo-palatal) tip of tongue and roof of the mouth (alveolar) lips (bilabalial)
Fricatives	Level of Obstruction Phonemes: /ف/ /ث/, /س/, /ذ/, /ز/, /ص/, /ض/, /ظ/, /ش/ /غ/, /خ/	Partial obstruction of the air stream Location of obstruction: lower lip and upper teeth (labiodental) tongue approach tooth ridge (alveolar) tongue and hard palate (alveo-palatal) back of tongue, soft palate (velar)
Affricates	Level of Obstruction Phonemes: /ج/, /چ/ /ڇ/, /جھ/	Combination of a stop followed directly by a fricative obstruction. Complex sound in which the tip of the tongue makes contact with the roof of the mouth and then separates slightly for the fricative. Location of obstruction: hard palate (alveo-palatal) hard palate (alveo-palatal and aspirates)
Nasals	Level of Obstruction Phonemes: /م/ /ن/ /ڻ/ /ڱ/ /ڃ/	Complete obstruction of the air stream through the mouth but lowering of the soft palate to allow the air to escape through the nose. Same obstruction as for stop consonants. Location of obstruction: lips (bilabial) tongue & tooth ridge (alveolar) tongue curled near tooth ridge (alveolar) back of tongue, soft palate (velar) back of tongue, hard palate (alveo-palatal)
Lateral	Level of Obstruction Phonemes: /ل/	Little obstruction of the air stream. The tip of the tongue touches the tooth ridge however air is allowed to pass over the sides of the tongue to reduce turbulence. Location of obstruction: tongue and tooth ridge (alveolar)
Retroflex	Level of Obstruction Phonemes: /ر/ /ڙ/	Little obstruction of the air stream. Location of obstruction: tongue curled near tooth ridge (alveolar) tongue and tooth ridge (alveolar)
Retroflex Plosives	Level of Obstruction Phonemes: /ٽر/, /ڊر/	Complete obstruction of air initially but later on little obstruction. Location of obstruction: tongue curled near tooth ridge (alveolar)
Semi-Vowels	Level of Obstruction Phonemes: /و/ /ي/ /ح/, /ھ/	Little obstruction of the air stream. The back of the tongue approached the soft palate or the blade of the tongue approaches the hard palate. Location of obstruction: back of tongue near soft palate (velum) with lips slightly together and rounded blade of tongue near hard palate minor obstruction initially but later on no obstruction of air stream.

Table 12(a)

msec	F1 (Hz)	F2 (Hz)	F3 (Hz)
325-500	350-450	950-1050	2800-3000
350-500	340-480	1010-1100	2800-2950
350-500	370-520	980-1050	2800-3000

Table 12(b)

F1 (Hz)	F2 (Hz)	F3 (Hz)
340-520	950-1100	2800-3000
430	1025	2900

Table 13(a)

msec	F1 (Hz)	F2 (Hz)	F3 (Hz)
625-675	320-420	1440-1540	2880-3000
600-680	320-500	1510-1570	2850-3400
610-675	440-500	1450-1510	2850-3150

Table 13(b)

F1 (Hz)	F2 (Hz)	F3 (Hz)
320-500	1440-1570	2850-3400
410	1505	3125

Table 14(a)

msec	F1 (Hz)	F2 (Hz)	F3 (Hz)
350-550	450-560	940-1150	2800-3100
280-480	500-580	935-1080	2800-3000
300-450	550-580	950-1010	2900-3200

Table 14(b)

F1 (Hz)	F2 (Hz)	F3 (Hz)
450-580	935-1150	2800-3200
515	1040	3000

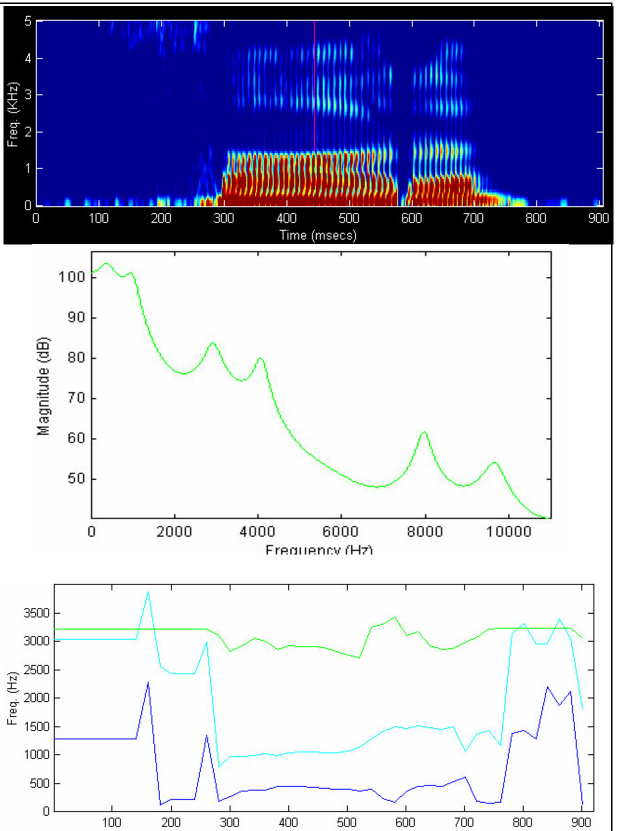


Fig 1: Spectrogram, LPC, and formant track for utterance “sarə”, MAK

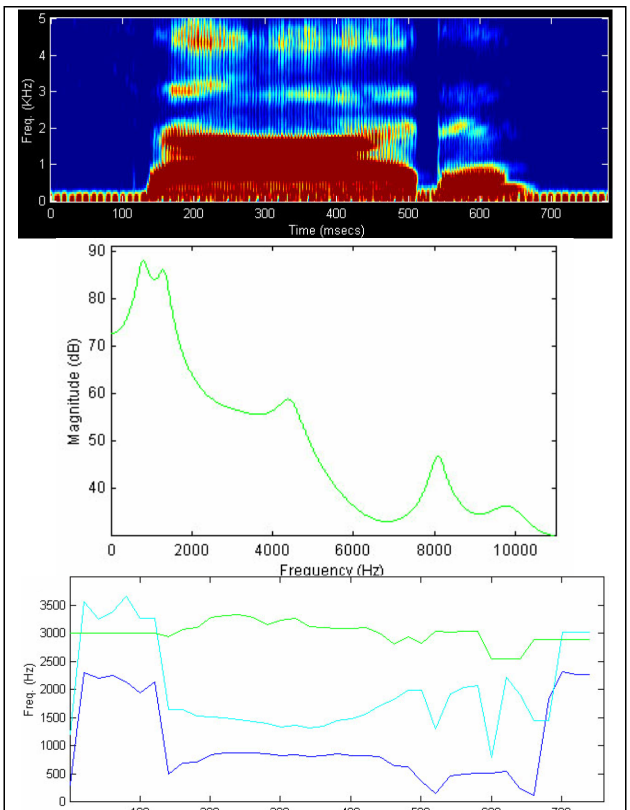


Fig 2: Spectrogram, LPC, and formant track for utterance “sarə”, FN

Table 15 (a)

msec	F1 (Hz)	F2 (Hz)	F3 (Hz)
620-700	350-450	1430-1550	2950-3350
550-650	320-500	1490-1570	3150-3550
550-630	300-450	1450-1600	3200-3600

Table 15 (b)

F1 (Hz)	F2 (Hz)	F3 (Hz)
300-500	1430-1600	2950-3550
400	1515	3250

Table 16 (a)

msec	F1 (Hz)	F2 (Hz)	F3 (Hz)
200-450	790-870	1330-1550	3100-3330
275-500	730-850	1370-1600	3000-3500
400-600	770-850	1300-1450	3070-3500

Table 16 (b)

F1 (Hz)	F2 (Hz)	F3 (Hz)
730-870	1300-1600	3000-3500
800	1450	3250

Table 17 (a)

msec	F1 (Hz)	F2 (Hz)	F3 (Hz)
125-325	630-720	1270-1400	3300-3700
300-600	610-700	1200-1500	2800-3600
150-300	670-750	1300-1500	2750-3750

Table 17 (b)

F1 (Hz)	F2 (Hz)	F3 (Hz)
610-750	1200-1500	2750-3750
680	1350	3250

Table 18: Eight selected words containing vowel /a/

Word	Pronunciation	Vowel	Translation
ٻارَ	barə	a	Children
يارَ	jarə	a	Friends
وارَ	varə	a	Hair
ساڙُ	saru	a	Jealousy
هاٿِي	hat[h]i	a	Elephant
چارو	Jaro	a	Cobweb
ڏارَ	daru	a	Crack
ڪارا	kara	a	Black

Table 19 (a)

msec	F1 (Hz)	F2 (Hz)	F3 (Hz)
300-450	420-480	890-1050	2750-2900
350-500	425-490	1050-1100	2650-2950
225-400	400-470	940-1080	2800-3050
300-450	360-460	920-980	2650-2950
300-400	300-420	920-1100	2750-3000
300-450	350-420	1000-1150	2600-2780
250-450	300-470	930-1150	2650-2860
150-300	430-530	970-1070	2690-2780

Table 19 (b)

F1 (Hz)	F2 (Hz)	F3 (Hz)
300-530	890-1150	2600-3050
415	1020	2825

Table 20 (a)

msec	F1 (Hz)	F2 (Hz)	F3 (Hz)
250-400	500-580	935-1080	2800-3000
300-450	450-580	1030-1150	2800-3000
300-450	450-560	970-1090	2750-2950
350-500	520-570	1025-1120	2890-3100
250-320	490-570	1050-1150	2750-3200
300-420	470-575	1100-1180	2800-2980
350-500	520-585	1040-1150	2800-3100
220-350	510-570	1000-1150	2700-3000

Table 20 (b)

F1 (Hz)	F2 (Hz)	F3 (Hz)
450-585	935-1180	2700-3200
515	1050	2950

Table 21 (a)

msec	F1 (Hz)	F2 (Hz)	F3 (Hz)
275-500	650-750	1150-1355	2700-2970
275-450	700-790	1360-1500	2770-2970
250-450	760-790	1250-1400	2950-3150
300-500	730-800	1240-1440	2900-3250
200-400	710-755	1200-1400	3000-3200
300-500	670-750	1240-1420	2950-3170
300-500	700-800	1330-1560	3000-3200
150-300	800-870	1200-1400	3000-3350

Table 21 (b)

F1 (Hz)	F2 (Hz)	F3 (Hz)
650-870	1150-1560	2700-3350
760	1355	3025

Table 22 (a)

msec	F1 (Hz)	F2 (Hz)	F3 (Hz)
125-325	630-720	1270-1400	3300-3700
300-550	660-740	1430-1560	3000-3800
300-500	650-675	1200-1350	2700-2830
300-600	610-700	1200-1500	2800-3600
200-350	640-665	1200-1350	2750-2930
250-500	590-680	1200-1550	2300-3900
250-500	650-735	1325-1520	3640-3850
150-300	670-800	1300-1500	2750-3750

Table 22 (b)

F1 (Hz)	F2 (Hz)	F3 (Hz)
590-740	1200-1560	2700-3900
665	1380	3300

Table 23

Word	Pronunciation	Vowel	Translation
سِيِرَ	Seer'a	i	Crevice
سِرَ	Sir'a	ɪ	Brick
سَيِرَ	Ser'a	e	Measuring unit
سَارَ	Saar'a	a	Care
سَرَ	Sar'a	ə	Funeral
سَوڙھ	Sor'ah	o	Congestion
سُرَ	Sur'a	ʊ	Tones
سُورَ	Soor'a	u	Pains

Table 24 (a)

Vowel	F1 (Hz)	F2 (Hz)	F3 (Hz)
i	210-250	2430-2580	3100-3600.
ɪ	300-330	2300-2385	3580-3700.
e	290-370	2400-2500	3300-3850.
a	300-450	1000-1100	2790-2950.
ə	390-450	1300-1460	2700-2930.
o	230-395	700-1050	2500-3260.
ʊ	320-380	1100-1200	2800-3350.
u	190-300	780-950	2800-3450.

Table 24 (b)

Vowel	F1 (Hz)	F2 (Hz)	F3 (Hz)
i	220-270	2380-2800	3200-3600
ɪ	270-300	2260-2400	3600-3800
e	300-340	2300-2450	3300-3800
a	420-500	980-1060	2800-3000
ə	400-440	1300-1420	2800-3000
o	200-350	770-1000	2500-2960
ʊ	250-330	1100-1200	2800-3250
u	190-260	700-850	3000-3700

Table 25: Average formant Frequencies for Sindhi Vowels

Vowel	F1 (Hz)	F2 (Hz)	F3 (Hz)
i	240	2590	3350
ɪ	300	2330	3650
e	330	2400	3575
a	450	1040	2900
ə	420	1380	2850
o	295	875	2880
ʊ	315	1150	3075
u	245	825	3250

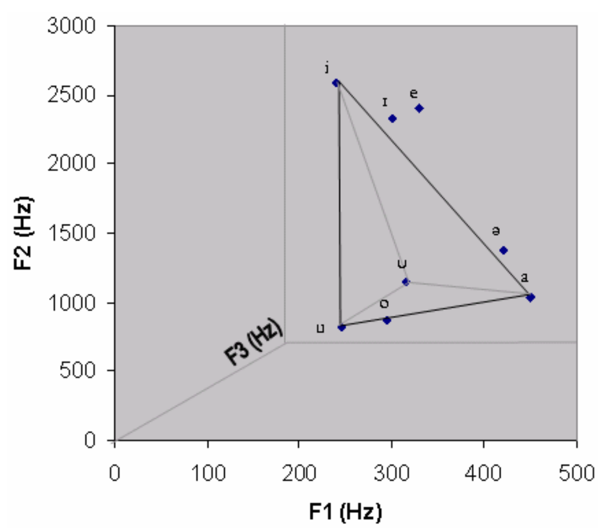


Figure 3: The Vowel Pyramid