Phonology for Sindhi Letter-to-Sound Conversion

Javed Ahmed Mahar[*]
Department of Computer Science,
Shah Abdul Latif University, Khairpur, Pakistan.

Ghulam Qadir Memon[*]
FEST, HIIT, Hamdard University, Karachi, Pakistan.

ABSTRACT

The Text to Speech (TTS) synthesis technology enables machines to convert text into audible speech and used throughout the world to enhance the accessibility of the information. Letter to sound (LTS) conversion is necessary component of any TTS system and phonological knowledge is essential for LTS conversion. This study deals with the conversion of Sindhi alphabet letters into their appropriate sounds. In this paper, phonology of Sindhi language is focused. For this purpose, some important areas of Sindhi phonology and writing system is reviewed and presented which can be used for Sindhi letter to sound conversion and also for the development of rule based Sindhi TTS synthesis system.

INSPEC Classification : C6150, C6170, C6180, C150, C7820.

Keywords : Text to speech, Letter to sound, Phonology, Phoneme, Diphthongs.

1. INTRODUCTION

Sindhi is an Indo-Aryan language of the Indo-European family, related to Hindi, Urdu and the languages of northwest Indian subcontinent. In Pakistan it is written using a modified form of the Perso-Arabic script with several additional letters to accommodate Sindhi implosive, retroflex and nasal sounds. It has many more consonants and vowels than Arabic. Sindhi occupies a prominent place among the languages of South Asia (Cole, 2005).

Sindhi is an earliest language of sub-continent. According to alphabet some languages like Urdu and Arabic are the sub-set of Sindhi language unfortunately it has not received the attention in computational language processing especially in terms of speech synthesis. In this paper phonology for Sindhi LTS conversion is focused because LTS conversion module is necessary component of Sindhi TTS system and phonological information is essential for LTS conversion.

The purpose of TTS synthesis is to convert input text to natural sounding speech as a result the information will transmit from a machine to a person. TTS systems provide voice output for all kinds of information such as phone numbers, addresses, navigation information,

and for reading books (Shah, 2004). TTS is divided into two stages. The first stage takes text input, processes it and converts it into precise phonetic string to be spoken. The second stage takes phonetic representation of speech and generates the digital signal.

LTS conversion is always based on some specific language rules. The two main justifications are conforms the need of LTS component. Firstly, there will always be genuinely new words in Sindhi language such as: glass, email., table created in the course of time or adopted in other languages and there are many words which may not be new, but were ignored when the system was originally built and have now become common enough to require proper pronunciation such as: bin laden, Obama. Secondly, LTS by rules can be used in cases where memory is limited.

Phonology is the study of the sound systems of languages. It is concerned with the linguistic patterning of sounds in human languages. Generally phonology is divided into two branches: (i) phonetics (ii) phonemics. In phonetics sounds of a language their types, pronunciation and segmentation are analyzed. The arrangement of phonetic sounds and their linguistically use is study in phonemics.

2. RELATED WORK

European scholars were the first to attempt a phonological and grammatical analysis of Sindhi. Their attention was drawn especially to the implosive stops which are unique characteristics of Sindhi and a few other Indo-Aryan languages. The four implosive stops in Sindhi were first described by George Stack in 1853. From that time to the present linguists have, with varying degree of clarity, attempted to describe these sounds. However, two contemporary linguists Bordie (1958) and Khubchandani (1961) have applied modern linguistic methods in their analysis and description of Sindhi sounds.

In past, there have been many developments in the Sindhi language particularly in terms of phonology. The Sindhi phonology, its morphological structure and syntax is discussed in (Jatoi, 1968). Cole (2005 and 2006) discussed the chart of Sindhi vowel and consonant sounds with IPA symbols, Sindhi syntax with grammar, morphological sound structure and Sindhi phonology. Bugio (2001) and Pauline (1981) discussed the consonantal vowel sounds and its types and present Sindhi Letters and their sounds. TTS synthesis system for Urdu and Sindhi is designed and developed by Shah et al. (2004) using knowledge based and hybrid rule based approach. Concatenative synthesis method is selected for this TTS in which actual snippets of recorded speech is used that were cut from recordings and stored in voice database. They also presented the phonemes of Sindhi and Urdu. Bird (1991) investigates Arabic verb morphology, Arabic syllable structure, phonological constraints and present theory of phonology. Sarfraz et al. (2003) discussed the writing forms of the Arabic alphabet.

Recently, many research efforts have been put into the field of natural language processing, including text to speech synthesis systems. The first task in the phonological processing is to convert the input text into a phonemic string using LTS rules. Hussain (2004) describe Urdu writing system its phonemic inventory, LTS rules and architecture of NLP for Urdu TTS. He also discusses Urdu consonantal and vocalic system. Zamirli (2007) proposed an algorithmic approach for the automatic generation of the stressing in Arabic language and represents the tonal rules which are employed in the phonetic module. They adapted, diagrams, generated for the text processing that acting on the size of the sentences to reading with intonative contours of natural speech. Muhtaseb et al. (2002) defines a set of Arabic diaphones/sub-syllables for concatenative Arabic TTS synthesis and proposed Arabic TTS diagram. They discussed speech segmentation rules, classification of Arabic consonants and types of syllables. Dakkak et al. (2005) introduced a work to incorporate emotions: anger, joy, sadness, fear and surprise, in an educational Arabic TTS system and they presents rules for emotion generation.

3. SINDHI WRITING SYSTEM

The Sindhi writing system, is based on Persian Arabic Script. Sindhi adds its own modifications in order to symbolize the many sounds not found in Arabic or Persian. For example, in the Sindhi alphabet, the original Arabic /t/, written ت, is extended to include /th/, /T/, and /Th/, written as ث, ٺ, and ٽ, respectively. All sounds not found in Arabic. This was done by taking the basic shape of the letter ت and adding or rearranging dots. In this way Sindhi has extended the 28 Arabic characters to 52 so that the sounds unique to Sindhi may be symbolized.

Because of the rich heredity of Sindhi in its Sanskrit origins, and the later additions of many Arabic and Persian words, the alphabet contains some sounds which are represented by more than one letter. Therefore only one sound is associated with any one letter among them. The letter used is determined by the origin of the words. This makes spelling more difficult although on the whole Sindhi is very phonetic in its spelling. The following are the sounds which may be represented by more than one letter:

/t/ ت, the common letter, and ط, in words of Arabic origin.
/s/ س, common, and ث, ص in words of Arabic origin. ص is also found in a few words of Persian origin.
/z/ ز, and ذ, common, ض, and ظ in words of Arabic origin.
/H/ ه, common, ح, found in words of Arabic origin.

Sindhi characters are written from right-to-left. This means that the first letter of a word appears at the right edge of the word, and the successive letters follow in a leftward direction. There are 52 distinct letters in the Sindhi alphabet and seven diacritic signs, but some of these like ا alifu and ء small alifu, represent a consonant sound.

The graphic representation of each alphabet of Sindhi, Arabic and Urdu languages has more than one form depending on its position. Most of the letters have four related forms (Beginning form BF, Middle form MF, End form and Isolated form). Four forms of Sindhi letters are described in Table 1. Some letters only connect on one side and are called "partially connecting" letters. They use just one shape for the initial and medial, and another shape for final and detached (Sarfraz, 2003) .

Table 1
Four forms of Sindhi letters

IF	BF	MF	EF
ع	ع	ع	ع
ڊ	ڊ	ڊ	ڊ
ش	ش	ش	ش
خ	خ	خ	خ

3.1 Basic Shape Groups

The 52 letters of Sindhi language are divisible into sixteen basic shape groups. Various letters may have the same basic shape, but are differentiated from each other within the group by the use of dots above, within or below the basic shape of the letter.

The four major shape groups are illustrated by these letters:

ب b   ج j   د d   ک g

Letter Group 1
This group contains only ا /A/. When found at the beginning of a word, the diacritic

"madd" will be written over ‌ا like "aana" /eggs/ ﺍﻧ . It is not usually found over ‌ا in the medial or final position. An important function of ‌ا is as a "carrier" of other vowels when a word begins with a vowel. The diacritical marks representing the short vowels must always be carried by ‌ا when at the beginning of a word. In other position in the word they are carried by the relevant consonant symbol.

Letter Group 2
This group contains ﺏ/b/, ﭖ/bb/, ﮀ/bh/, ﺕ/t/, ﺙ/th/, ﺕ/T/, ﺙ/Th/, ﺙ/s/, ﭖ/p/, an partially ﻥ/n/, ﮌ/ $\bar{R}$ /. The letter ﺙ is an uncommon Arabic consonant that is; it is not frequently used in Sindhi. The letters ﻥ/n/ and ﮌ / $\bar{R}$ / differ somewhat from others. The forms of ﻥ and ﮌ are more rounded than the others also they drop below the main lines of writing. The letter ﻯ also has special forms, initial ﻯ stands only for the consonant sound /y/, /I/, /E/, or /ai/ is symbolized by ﻯ plus ‌ا . For example, "eiman" /faith/ ﺍﻳﻤﺎﻥ . Note that the only difference between /I/ and /E/ sound as symbolized is the inclusion of the diacritic "zer", with ﻯ thus ﻯ .

Letter Group 3
This group includes ﺝ/j/, ﺟ/jj/, ﺟﻬ/jh/, ﺝ/ $\bar{N}$ /, ﭺ/c/, ﭺ/ch/, ﺡ/H/, ﺥ/K/. The letter ﺡ/H/ occurs only in words of Arabic origin.

Letter Group 4
This group contains ﺩ/d/, ﺫ/dh/, ﮈ/D/, ﮉ/Dh/, ﮃ/dd/ and ﺫ/z/. The letter ﺫ is an uncommon Arabic consonant. Thus it is not found frequently in Sindhi.

Letter Group 5
This group contains ﺭ/r/, ﮌ/R/, and ﺯ/z/. The letter ﺯ is the most common representation of /z/ in Sindhi. Notice the difference in the shape of the ﺭ group and that of the ﺯ group. Peoples sometimes confuse the two in their writing. The ﺯ is written with a relatively closed angle. Also, the ﺭ drops down below the line of writing and the ﺯ does not.

Letter Group 6
This group includes ﺱ/s/ and ﺵ/S/.

Letter Group 7
This group includes ﺹ/s/ and ﺽ/z/. These letters are found only in loan words of Arabic origin.

Letter Group 8
This group includes ﻁ/t/ and ﻅ/z/.

Letter Group 9
This group contains ﻉ /!/ and ﻍ /G/. The ﻉ has no easily assignable phonemic value in Sindhi. It occurs only in very literary pronunciations of Arabic loan words. Sindhi speakers usually omit the pronunciation entirely.

Letter Group 10
This group contains ﻒ /ph/, ﻒ/f/ and ﻕ/q/.

Letter Group 11
This group contains only ﻙ/k/.

Letter Group 12
This group includes ﻚ/kh/, ﮒ/g/, ﮛ/gg/, ﻍ/gh/, ﮎ/ g /. Before ‌ا /A/ and ﻝ/l/, special initial and medial forms are found. Like "khadho" /food/ﻛﮬﺎ, "bhaggalu" /broken/ﮐ etc. Notice the extra stroke that distinguishes the voiced velar stops from the voiceless ﻙ .

Letter Group 13
The only member of this group is ﻝ /l/.

Letter Group 14
The only member of this group is ﻡ /m/.

Letter Group 15
This group contains only ﻭ /v/. When ﻭ is found at the beginning of a word, it stands for the consonant sound /v/. When it is used to represent an initial vowel sound, /O/, /U/, or /ao/ it is found with ﺍ /A/. For example, "ocito" /sudden/ ﻭﻗﺘﻮ . In the medial and final position, ﻭ may represent either the consonant sounds or any of the vowel sounds. The only difference in representation between /O/ and /U/ and the diphthong /ao/ is the presence or absence of the diacritical marks.

Letter Group 16
The only member of this group is ﻩ /H/, like "hath" /hand/ ﻫﺚ. This letter also functions as the symbol for aspiration in ﺟﻬ /jh/, and ﮔﻬ /gh/.

/**hamzo**/ ء When within a word any syllable ends with a vowel (long or short) and the following syllable begins with one, the two vowels are separated by /hamzo/ء. It serves the same purpose that a hyphen does in English, that is to separate two syllables. In the initial and medial forms /hamzo/ ء must be written over a "carrier" which is the same basic shape as letter group 2.

4. PHONOLOGICAL ANALYSIS

The phonological systems of Sindhi in most respects resemble that of other Indo-Aryan languages. Sindhi has a very rich sound inventory. It has 43 distinctive consonant phonemes and 10 vowels. The phoneme is usually pronounced as an alveolar tap, though occasionally reminiscent of a trill with two or more contacts. There are three short vowels [a, i, u] and five long vowels [aa, ii, uu, e, o]. There are also two diphthongs [ai, au], but these are infrequent and many dialects pronounce these the same as [e, o].

Among the fifty two characters and seven diacritic signs, twenty nine characters are adopted from Arabic script. Three modified characters adopted from the Persian script: پ , چ , گ.

Twenty modified characters to represent Sindhi sounds:

Retroflex sounds:
ڙ, ت, ڍ, ڊ, ڻ, ٽ

Rest: Voiceless Aspirates:
ڪ, چ, ٿ, ڦ

Voiced Aspirates:
ڀ, ڌ, ڌ, ڱ

Implosive:
ٻ, ڄ, ڳ

Nasal:
ڃ, ڱ

4.1 Consonant sounds of Sindhi

There are 50 letters in the Sindhi alphabet that stand for consonant sounds. As some letters

represents the same sound discussed above so that total of 43 letters are symbolize the consonant sounds. Each letter always represents the same sound in Sindhi alphabet respectively, which makes it very easy to read and sound out new words.

There is no word in Sindhi language in which two or more consonants used in any portion. For example in English the word "structure" has three consonants in starting position. Even two consonants are not found in any word of this language. In reference to Sindhi language word structure, vowels are used mostly after each consonant. Similarly no word is found ending consonants.

The following are the consonants sounds of Sindhi language (Pauline, 1981).

b (ب): Voiced unaspirated bilabial stop, as in "baby"
bb (ٻ): Voiced bilabial implosive stop
bh (ڀ): Voiced aspirated bilabial stop
t (ٽ): Voiceless unaspirated retroflex stop.
th (ٺ): Voiceless aspirated retroflex stop.
T (ت): Voiceless unaspirated dental stop; not the English "t" sound, which is alveolar.
Th (ٿ): Voiceless aspirated dental stop
P (پ): Voiceless unaspirated bilabial stop.
j (ج): Voiced unaspirated palato-alveolar affricate, as in "joy".
jj (ڄ): Voiced palatal implosive stop.
jh (جھ): Voiced aspirated palato-alveolar affricate, as in "judge"
N (ڃ): Voiced palatal nasal.
c (چ): Voiceless unaspirated palato-alveolar affricate similar to the sound in "cheese" if pronounced without aspiration.
ch (ڇ): Voiceless aspirated palato-alveolar affricate, as in "choo-choo".
k (خ): Voiceless velar fricative, similar to the German "ach" and the Scottish "loch"
d (ڊ): Voiced unaspirated retrofrlx stop.
dh (ڍ): Voiced aspirated retroflex stop.
dd (ڏ): Voiced alveolar implosive stop
D (د): Voiced unaspirated dental stop; not the English "d" sound, which is alveolar
Dh (ڌ): Voiced aspirated dental
r (ر): Voiced alveolar trill or flap; similar to the Spanish trilled "r."
R (ڙ): Voiced retroflex flap
z (ز): Voiced alveolar fricative, as in "zebra".
s (س): Voiceless alveolar fricative, as in "see"
S (ش): Voiceless palato-alveolar fricative, as in "sheep".
! (ع): Glottal stop. This is the "ain" of classical Arabic and in Sindhi it occurs in loan words from Arabic. It is not emphatically pronounced as a glottal stop in Sindhi.
G (غ): Voiced velar fricative.
f (ف): Voiceless labiodental fricative, as in fish.
ph (ڦ): Voiceless aspirated bilabial stop
q (ق): Voiceless unaspirated uvular stop. This is the "qui" of classical Arabic and in Sindhi it occurs in loan words from Arabic. In ordinary Sindhi pronunciation it becomes /k/.
K (ڪ): Voiceless unaspirated velar stop, as in "school"
kh (ک): Voiceless aspirated velar stop, as in "kin".
g (گ): Voiced unaspirated velar stop, as in "go"
gg (ڳ): Voiced velar implosive stop.
gh (ڳھ): Voiced aspirated velar stop
g (ڱ): Voiced velar nasal. This is one consonant sound, not two as in "bingo". It is similar to the sounds in "singing".
L (ل): Voiced dental lateral; similar to the sound of "l" in "lean" "feel" when the tongue is behind the upper front teeth.
M (م): Voiced bilabial nasal, as in "man"

N ( ): Voiced dental nasal, some what similar to the English "n", but the tip of the tongue is behind the upper teeth.

$\mathcal{R}$**( ):** Voiced retroflex nasal flap. Curl the tip of the tongue up to the back of the alveolar ridge, and make as "n" sound as you flap the tongue against the back of the ridge as it returns to its position behind the lower front teeth.

v ( ): Voiced labio dental fricative, similar to the sound in "vine" but the friction is weaker.

H ( ): A stream of air passed through the vocal cords. Position of the tongue is determined by the vowel that follows it.

Y ( ): Voiced palatal approximant, as in "yes".

4.2 Sounds of Sindhi Vowels and Diphthongs

Vowels in any language are more difficult than consonants to describe because they vary from person to person. Overlooking minor variations, then, it is possible to distinguish eight vowels and two diphthongs.

Vowels are distinguished from each other by the position of the tongue. The Sindhi vowels range from the tongue high in the front of the mouth for /I/ to the tongue high in the back of the mouth for /U/, with the other vowels falling somewhere in between and lower. Linguists sometimes plot the location of these vowels on a chart. As they are pronounced in the order of /I/, /i/, /E/, /ai/, /a/, /A/, /ao/, /O/, /u/, and /U/, you will notice that the jaw is almost closed for /I, that it opens as we move down the chart to /E/; then it opens more jaw move down the chart to /A/; then it closes again as we move up to /U/. The Sindhi vowels are described in Table 2 that describes the movement of the jaw and tongue as Sindhi vowels are formed. The tongue, too, drops progressively lower from a high front position at the beginning until it reaches a mid-position, then rises progressively at the back (Jatoi,1968).

Table 2.
The Chart of Sindhi Vowels

	Front	Central	Back
High	I		U
Lower-high	i		U
Higher-mid	E		O
Mid		a	
Lower-mid	ai		Ao
Low		A	

4.3 Sindhi Phonemes

Sindhi may be divided into six major dialects: (1) Siro or Siraiki, spoken in the northern part of Sindh. (2) vicholi, spoken in the central part of Sindh. (3) Lari, spoken in the southern part of Sindh. (4) Sasi, spoken in lasbela and the khairthar range on the western border of Sindh. (5) Thari, spoken in the eastern part of Sindh and the Sindh-Rajastan border. (6) Kachi, spoken in the Kutch region of Gujrat on the southern border of Sindh (Cole, 2006). The totals of 520 phonemes of Sindhi language are used for all dialects. There are 52 letters in Sindhi language, each letter have 10 different sounds by using the different diacritics and two letters ? and ?. 520x10 makes 520 phonemes. The phonemes of Sindhi language are described in Table 3.

Table 3.
The Phonemes List of Sindhi language

او	او	اي	اي	او	ا	اي	ا	ا	ا
بو	بو	بي	بي	بو	ب	بي	ب	با	ب
بو	بو	بي	بي	بو	ٻ	بي	ٻ	با	ٻ
ڀو	ڀو	ڀي	ڀي	ڀو	ڀ	ڀي	ڀ	ڀا	ڀ
تو	تو	تي	تي	تو	ت	تي	ت	تا	ت
ٿو	ٿو	ٿي	ٿي	ٿو	ٿ	ٿي	ٿ	ٿا	ٿ
ٽو	ٽو	ٽي	ٽي	ٽو	ٽ	ٽي	ٽ	ٽا	ٽ
ٺو	ٺو	ٺي	ٺي	ٺو	ٺ	ٺي	ٺ	ٺا	ٺ
پو	پو	پي	پي	پو	پ	پي	پ	پا	پ
جو	جو	جي	جي	جو	ج	جي	ج	جا	ج
جو	جو	جي	جي	جو	ڄ	جي	ڄ	ڄا	ڄ
ڄو	ڄو	ڄي	ڄي	ڄو	ڃ	ڄي	ڃ	ڃا	ڃ
جهو	جهو	جهي	جهي	جهو	جه	جهي	جه	جها	جه
چو	چو	چي	چي	چو	چ	چي	چ	چا	چ
چهو	چهو	چهي	چهي	چهو	ڇ	چهي	ڇ	ڇا	ڇ
حو	حو	حي	حي	حو	ح	حي	ح	حا	ح
خو	خو	خي	خي	خو	خ	خي	خ	خا	خ
دو	دو	دي	دي	دو	د	دي	د	دا	د
ڏو	ڏو	ڏي	ڏي	ڏو	ڏ	ڏي	ڏ	ڏا	ڏ
ڊو	ڊو	ڊي	ڊي	ڊو	ڊ	ڊي	ڊ	ڊا	ڊ
ڌو	ڌو	ڌي	ڌي	ڌو	ڌ	ڌي	ڌ	ڌا	ڌ
ڍو	ڍو	ڍي	ڍي	ڍو	ڍ	ڍي	ڍ	ڍا	ڍ
ڎو	ڎو	ڎي	ڎي	ڎو	ذ	ڎي	ذ	ذا	ذ
رو	رو	ري	ري	رو	ر	ري	ر	را	ر
ڙو	ڙو	ڙي	ڙي	ڙو	ڙ	ڙي	ڙ	ڙا	ڙ
زو	زو	زي	زي	زو	ز	زي	ز	زا	ز
سو	سو	سي	سي	سو	س	سي	س	سا	س
شو	شو	شي	شي	شو	ش	شي	ش	شا	ش
صو	صو	صي	صي	صو	ص	صي	ص	صا	ص
ضو	ضو	ضي	ضي	ضو	ض	ضي	ض	ضا	ض
طو	طو	طي	طي	طو	ط	طي	ط	طا	ط
ظو	ظو	ظي	ظي	ظو	ظ	ظي	ظ	ظا	ظ
عو	عو	عي	عي	عو	ع	عي	ع	عا	ع
غو	غو	غي	غي	غو	غ	غي	غ	غا	غ
فو	فو	في	في	فو	ف	في	ف	فا	ف
ڦو	ڦو	ڦي	ڦي	ڦو	ڦ	ڦي	ڦ	ڦا	ڦ
قو	قو	قي	قي	قو	ق	قي	ق	قا	ق
ڪو	ڪو	ڪي	ڪو	ڪو	ڪ	ڪي	ڪ	ڪا	ڪ
كو	كو	كي	كو	كو	ك	كي	ك	كا	ك
گو	گو	گي	گو	گو	گ	گي	گ	گا	گ
ڳو	ڳو	ڳي	ڳو	ڳو	ڳ	ڳي	ڳ	ڳا	ڳ
ڳهو	ڳهو	ڳهي	ڳهو	ڳهو	ڱ	ڳهي	ڱ	ڱا	ڱ
کو	کو	کي	کو	کو	ک	کي	ک	کا	ک
لو	لو	لي	لي	لو	ل	لي	ل	لا	ل
مو	مو	مي	مي	مو	م	مي	م	ما	م
نو	نو	ني	ني	نو	ن	ني	ن	نا	ن
ڻو	ڻو	ڻي	ڻي	ڻو	ڻ	ڻي	ڻ	ڻا	ڻ
وو	وو	وي	وي	وو	و	وي	و	وا	و
هو	هو	هي	هي	هو	ه	هي	ه	ها	ه
ءو	ءو	ءي	ءي	ءو	ء	ءي	ء	ءا	ء
يو	يو	يي	يي	يو	ي	يي	ي	يا	ي

4.4 Syllables

The concatenation of such letters make syllables and the concatenation of syllables make words. There is differ of opinion of linguistics for the definition of syllables, all are agreed that the syllables are exists in words. Every intellectual can differentiate the syllables of words because his ears are habitual for listening of his language sounds. Ears play important role for segmentation of words into syllables. Syllabification is not measured in written form. The concatenation of letters makes syllables only for the sonority of sound.

Syllable division in a word is predictable in Sindhi. Sindhi is primarily an open syllable language, i.e., syllables mostly end with a vowel or semivowel. Words in Sindhi mostly have vocalic ending and the occurrence of consonant cluster is also irregular in the language.

A syllable in Sindhi consists of at least one vowel or at most five sounds units, in which one is a vowel and others are non-vocalic sounds (consonants or semivowels preceding or following the vowel). Types of syllables in Sindhi are (Jatoi, 1968):

V	Vowel
$\overline{V}$	Long Vowel
C	Consonant

- $\overline{V}$: Like   او با
- CV: Like   تُه تَه
- C $\overline{V}$: Like   ثوت نام
- C $\overline{V}$V: Like   ماء هوء
- CC$\overline{V}$: Like   هيا ثو
- CC $\overline{V}$C: Like   هوريث
- C$\overline{V}$C: Like   ناغ
- CVCC: Like   بحرمد جرمد

4.5 Stress

The accent of words is usually change in different languages. Some words start with stress and few without stress. The languages in which meaningful difference is occur; we say that these languages have phonemic importance for sound stress. For example in English, word permit has two syllables (per.mit). If we stress first syllable then the word permit is considered as a noun means (اجازتنامون) but if we stress on second syllable then word permit is considered as a verb means (اجازت ڏيڻ).

In Sindhi, stress has only a limited use of demarcating words and putting emphasis on a particular word in an utterance. There are three main stresses: word stress, emphatic stress and drawled stress.

5. CONCLUSION

Letter to sound conversion is a central component of rule based Sindhi TTS synthesis system. The phonological knowledge is essential for LTS conversion. In this paper, we have reviewed and formulated the phonology of Sindhi language. This could be support for building automatic LTS conversion component. For this purpose, we have explained the details of Sindhi writing system and letters shape groups. Different areas of Sindhi phonology like: consonant and vowels sounds, Sindhi phonemes, the syllable structure and importance of stress ness of the language pronunciation are discussed and presented. This phonological analysis will also support for many other languages like Urdu, Arabic and Persian.

REFERENCE

Bird, S., (1991), "A Logical Approach To Arabic Phonology", Proceedings of the 5th Conference on European Chapter of the Association for Computational Linguistics, pp. 89-94.

Bordie, John, (1958), "The Phonology of Sindhi", PhD Theis, University of Texas.

Bugio, M.Q., (2001), "Sociolinguistics of Sindh", LINCOM EUROPA? PhD Thesis.

Cole, Jennifer., (2005), "Sindhi", In Strazny, Philipp(ed) Encyclopedia of Linguistic. New Yark: Routledge.

Cole, Jennifer, (2006), "The Sindhi Language", In K.Brow(ed) Encyclopedia of Language and Linguistics, 2nd Edition, v.11: pp. 384-386. Oxford:Flsevier.

Dakkak, O.; Ghneim, N. Abou Zliekha, M.; moubayed, S., (2005), "Emotion Inclusion in an Arabic Text-to-Speech", 13th European Signal Processing Conference.

Hussain, S., (2004), "Letter-to-Sound Conversion for Urdu Text-to-Speech System", COLING 2004 Computational Approaches to Arabic Script-based Languages, pp. 74-79.

Jatoi Ali Nawaz, (1968), "Ilm Lisan Ain Sindhi Zaban", Institute of Sindhalogy, Hyderabad.

Khubchandani, L.M., (1961), "The Phonology and Morphonemics of Sindhi", M.A. Thesis, University of Pannyslvania.

Muhtaseb, H.; Elshafei1, M; Ghamdi, M., (2002), "Techniques for High Quality Arabic Speech Synthesis", Informatics and Computer Science: An International Journal, Vol.140, pp. 255-267.

Pauline. A. Brown, (1981), "Functional Sindhi", Thesis.

Shah A. A.; Ansari, A. W.; Das, L., (2004), "Bi-Lingual Text to Speech Synthesis System for Urdu and Sindhi", National Conference on Emerging Technology, pp. 126-130.

Sarfraz, M.; Nawaz, S. N.; Khuraidly, A.A, (2003), "Offline Arabic text recognition system", Proc. of the Int. comference on geometric modeling and graphics.

Zemirli, Z.; Khabet, S; Mosteghanem, M., (2007), "An effective model of streesing in an Arabic Text to Speech System", IEEE AICCSA, pp. 700-707.