URDU CONSONANTAL AND VOCALIC SOUNDS

ABDUL MANNAN SALEEM, HASAN KABIR, MUHAMMAD KHURRAM RIAZ, MUHAMMAD MUSTAFA RAFIQUE, NAUMAN KHALID, SYED RAZA SHAHID

ABSTRACT

It is a fact that no scientific speech processing research has been done so far, that can be the basis for improved applications and further research in Pakistan. One of the primary reasons is the absence of any core material related to the phonetic inventory of Urdu. This paper addresses this matter by attempting to provide a listing of all possible sounds that are present in Urdu language, and also attempts at justifying their presence.

1. INTRODUCTION

Urdu, the national language of Pakistan, is partially spoken and more than partially understood in South Asian countries like India, Bangladesh and UAE. Hindi seems to be phonetically similar to Urdu, but it differs in its script and historical characteristics. The pronunciation of Urdu varies from region to region due to different characteristics of their regions.

The word Urdu has a Turkish origin, meaning 'camp or army with its follower'. It is popularly regarded as an offspring of Persian. It borrows words from different languages to expand its vocabulary. Major languages participating in the camp of Urdu are: Persian, Arabic, Portuguese and English (Saksena).

It is a fact that no scientific speech processing research has been done so far, that can be the basis for improved applications and further research in Pakistan. One of the primary reasons is the absence of any core material related to the phonetic inventory of Urdu. This paper addresses this matter by attempting to provide a listing (inventory) of all possible sounds that are present in Urdu language, and also attempts at justifying their presence. Some comparison with other published lists is also done as a literature review. This listing (inventory) is the basis for any research to be done on Urdu language speech processing.

2. LITERATURE REVIEW

According to Kachru (1990), there are seven long oral vowels, and three short oral vowels, and according to Bokhari (1985; 1991) there are seven long oral vowels, but seven short oral vowels. Bokhari (1985; 1991) contains many allophones of the corresponding long vowels as discussed by Kachru (1990). Kachru (1990) maintains that the front low cardinal vowel [æ] exists as front middle low vowel [ɛ] in Urdu. As a result the back low cardinal vowel [ɑ] is shifted to the low center, making it [a]. Khan (1997) also agrees with the long and short vowel distribution of Kachru (1990).

Bokhari (1985; 1991) and Khan (1997) list ten nasalized vowels including five short and five long nasalized vowels. Kachru (1990), on the other hand, has not listed any nasalized vowel, but mentions in the text that oral and nasal vowels contrast, and that nasalization is distinctive (Kachru, 1991, p. 55).

Collectively Kachru (1990), Bokhari (1985; 1991), Khan (1997), and Hussain (1997) have listed forty-three (43) consonantal sounds of Urdu (see TABLE 1), out of which twenty-eight (28) sounds are agreed upon by all the above authors. Kachru (1990) lists 37 consonants and has missed [ʔ, h, rʰ, nʰ, mʰ, lʰ]. Hussain (1997) lists 36 consonants and has missed [ŋ, ʈʰ, rʰ, nʰ, mʰ, lʰ, q]. Bokhari (1985; 1991) lists 36 consonants and he has missed [f, ʃ, ʒ, z, q, x, r]. Bokhari misses interestingly many basic sounds, which are listed by Kachru and Hussain. Khan (1997) lists, most of all, 42 consonants and has missed only one consonantal sound [ŋ].

Overall, the controversial consonantal sounds are [ʔ, ʈʰ, rʰ, nʰ, mʰ, lʰ, ŋ]. This paper will try to discuss these cases in more detail.

3. METHODOLOGY

3.1 Subjects

For the purpose of finding Urdu consonantal and vocalic sounds, a group of eight native Urdu speakers were surveyed, and consulted. Furthermore, to confirm data, some established and well-known dictionaries (see section 0 below) and books of Urdu language were studied. Many published texts were consulted which are listed in the references section.

3.2 Data Recording and Processing

All acoustic analysis of the speakers was carried out on a speech processing software Xwaves® 5.3 by Entropic®, and Speech Analyzer® by SIL®. The equipment consisted of a high fidelity (Hi-Fi) microphone, a Teac integrated stereo amplifier and two high quality speakers with 8-ohm impedance.

3.3 Experimental Conditions

All subjects were required to speak a list of Urdu words collectively containing all the relevant vocalic and consonantal sounds. Due to the unavailability of a sound proof room, the datum thus collected was vigorously screened for errors.

4. RESULTS

4.1 Consonants

The inventory of all consonants is given in TABLE 1.

Type	Place	Manner	Sound Symbol	Minimal Pairs
Stops	Bilabial	Voiceless	p	pɑp, bɑp, bʰɑp
		Voiced	b	pəlnɑ, pʰəlnɑ
		Aspirated Voiceless	pʰ	pɑn, mɑn
		Aspirated Voiced	bʰ	bəndɑ, pʰəndɑ
		Nasalized Voiced	m	əbər, əmər
				pʰɑɽɑ, bʰɑɽɑ
				pʰəlnɑ, məlnɑ
				bʰərɑ, mərɑ
		Aspirated Nasalized Voiced	*mʰ*	t̪ʊmʰẽ
	Dental	Voiceless	t̪	t̪ɑl, t̪ʰɑl, d̪ɑl
		Voiced	d̪	t̪ɑli, t̪ʰɑli
		Aspirated Voiceless	t̪ʰ	t̪ɑr , d̪ʰɑr
		Aspirated Voiced	d̪ʰ	t̪ɑn, t̪ʰɑn, d̪ʰɑn, nɑn
		Nasalized Voiced	n	d̪ɑɽ, d̪ʰɑɽ
				d̪ɑm, nɑm
				t̪ʰən, d̪ʰən
		Aspirated Nasalized Voiced	*nʰ*	
	Alveolar	Voiceless	ʈ	ʈɑlnɑ, ɖɑlnɑ
		Voiced	ɖ	ʈɑʈ, t̪ʰɑʈ
		Aspirated Voiceless	ʈʰ	ʈɑl, ɖʰɑl
		Aspirated Voiced	ɖʰ	ɖəndɑ, ʈʰəndɑ
				ɖɑl, d̪ʰɑl, t̪ʰɑl
	Velar	Voiceless	k	kɑli, gɑli
		Voiced	g	kɑnɑ, kʰɑnɑ
		Aspirated Voiceless	kʰ	kɑʈi, gʰɑʈi

		Aspirated Voiced	g^h	gɑnɑ, kʰɑnɑ
		Nasalized Voiced	ŋ	gɑṭi, gʰɑṭi
				kʰəɽɑ, gʰəɽɑ
				səŋ, sən
Fricatives	Uvular	Voiceless	q	qələm, ʔələm
	Glottal	Voiceless	ʔ	ʔəlim, qəlim
	Labio-Dental	Voiceless	f	vərk, fərk
		Voiced	w (v)	faʃ, laʃ
	Alveolar	Voiceless	s	ser, zer
		Voiced	z	
	Palatal	Voiceless	ʃ	ʃɑɖɪ, sɑɖɪ
		Voiced	ʒ	ʒɑlɑ, ʃɑlɑ
	Velar	Voiced	ɣ	ɣərib, qərib
	Uvular	Voiceless	x	xɑnɑ, gɑnɑ
	Glottal	Breathy Voiceless	h	hɑmɪ, nɑmɪ
Affricates	Alveolar	Voiceless	tʃ	tʃəlɑ, dʒəlɑ
		Voiced	dʒ	tʃʰal, tʃɑl
		Aspirated Voiceless	$tʃ^h$	dʒʰɑg, dʒɑg
		Aspirated Voiced	$dʒ^h$	
Trills	Palatal	Voiced	r	mʊqərər, mʊqərəb
		Aspirated Voiced	r^h	
Flap	Palatal	Voiced	ɽ	bɑɽ, bɑl
		Aspirated Voiced	$ɽ^h$	
Approximants	Back	Central	j	lɑɖ, jɑɖ
	Middle	Lateral	l	gɑjɑ, gɑnɑ
		Aspirated Lateral	l^h	

TABLE 1 *Urdu Consonantal Sounds, tabulated by survey of native speakers, and some established dictionaries [2,3,4]*

4.2 Vowels

The vowels present in Urdu language are tabulated in TABLE 2

Type	Position		Sound	Minimal Pairs
	Front/Back	High/Low	Symbol	
Long	Front	High	i	ɖin, ɖən, ɖɪn
		Middle	e	bæl, bel
		Low	æ	
	Back	High	u	sunɑ, sʊnɑ, sonɑ
		High-Middle	o	bɔnɑ, bonɑ
		Low-Middle	ɔ	kɑn, kɔn
		Low	ɑ	
Short	Front	High	ɪ	bɪn, bən

	Middle	ɛ	sɛhər, səhər
Back	High	ʊ	sʊn, sən
Middle	Middle	ə	kəlɪ, kʊlɪ
Nasalized Long	Front	High	ĩ	pɛhnĩ, pɛhni
		Middle	ẽ	kəhẽ, kəhe
		Low	æ̃	hæ̃, hæ
	Back	High	ũ	hũ, hu
		High-Middle	õ	dʒanõ, dʒano
		Low	ɑ̃	kəhɑ̃, kəhɑ

TABLE 2 Urdu Vocalic Sounds, tabulated by survey of native speakers, and some established dictionaries [2,3,4]

4.3 Counts

Stops	23	Fricatives	9	Vowels	17
Bilabial	6	Labio-Dental	2	Long	7
Dental	6	Alveolar	2	Short	4
Alveolar	4	Palatal	2	Nasalized Long	6
Velar	5	Velar	1	Front	8
Uvular	1	Uvular	1	Back	8
Glottal	1	Glottal	1	Middle	1
Affricates	4	Trill	2	**VOWELS**	**17**
Alveolar	4	Palatal	2		
Approximants	3	Flaps	2	**CONSONANTS**	**43**
Back	1	Palatal	2		
Middle	2				

TABLE 3 Summary of the number of different sounds

5. DISCUSSIONS

Distinction Between [v] and [w]

Kachru explains that the phoneme [v] was originally part of Urdu (called Highly Persianised Urdu by him, and Khari Boli by some others). At that stage Urdu was part of Old Indo-Aryan languages. During the transition of Urdu from Old Indo-Aryan to New Indo-Aryan languages, the membership of [v] became questionable, and many linguists claimed of [w] being its replacement. A confusion whether [v] or [w] is now present in Urdu was created and is present even now (Khalid, 2002).

Khalid (2002) discusses the existence and usage of [v] and [w] in much detail in his paper. The experiment and its analysis revealed that the native speakers of Urdu do not distinguish between the use of [v] or [w] in words. The message is conveyed equally well in both cases. Also enough data is present to suggest that being given freedom to pronounce anything from [v] to [w], the speakers generally tend towards a more loose pronunciation, somewhere between the

approximant [w], the fricative [v], and the vowel [u].

The discussion in (Khalid, 2002) also reveals that the phoneme [w] is a more appropriate member of the consonantal inventory for Urdu.

Aspirated Nasalized Stops

A study conducted on nasal aspirants in Urdu revealed that [nʰ] and [mʰ] do not occur in word initial and word final positions (Aziz, 2002).

A word medial aspirated [m] is seen to occur when its following segment is a non-vowel. One thing is certain: [mʰ] occurs word medially. However, no case for existence of [nʰ] could be found (Aziz, 2002).

Although this suggests that [mʰ] is more often broken down in two separate phonemes [m], and [h], than uttered as [mʰ], but there are some cases described by Aziz (2002) that confirms the existence of [mʰ]. Also the existence is highly speaker dependent (Aziz, 2002).

Possible Vowel (Front, Low-Middle) in Urdu

Another possible vowel was encountered in the study. This vowel [ɛ] was found in words, such as [kɛhər]. The spectrogram of this word is shown in FIGURE 1, and it is quite clear that the [ɛ] vowel is spoken while pronouncing these words of Urdu language.
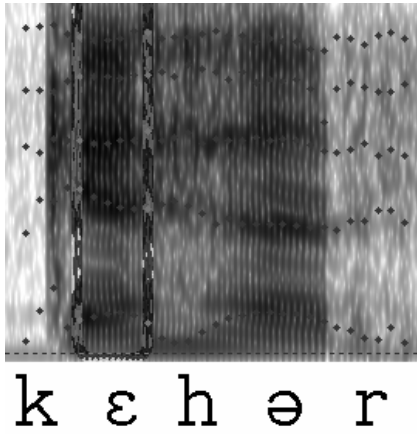

FIGURE 1 *Spectrogram Analysis of [kɛhər]*

The above example for the existence of [ɛ] gives a high degree of chance that [æ] and [ɛ] are allophones of each other. However, this controversy has to be examined in greater depth, and is not discussed in this paper.

Other controversies of [lʰ], [rʰ], and [ɽʰ] are also not discussed here, but they should be explored and corrected.

6. REFERENCES

Aziz, Omar 2002. *Nasal Aspirates in Urdu*. Center for Research in Urdu Language Processing (CRULP), National University of Computer and Emerging Sciences, Lahore, Pakistan.

Bokhari, Sohail 1985. *Phonology of Urdu Language*.

Bokhari, Sohail 1991. *Urdu Zubaan ka Soti Nizaam*.

Farhang-e-Talffuz by Shaan-ul-Haq Haqqi, Muqtadra Qaumi Zubaan. ISBN 969-474-153-X.

Feroz-ul-Lughat Urdu, ISBN 969-000-514-6

Hussain, Sarmad 1997. *Phonetic Correlates of Lexical Stress in Urdu*.

Jadeed Urdu Lughat—Tulba ke Liye, ISBN 969-474-051-7.

Kachru, Yamuna 1990. *Hindi-Urdu* in *The Major Languages of South Asia, The Middle East and Africa*, edited by Bernard Comrie.

Khan, Mahboob Alam 1997. *Urdu ka Soti Nizaam*.

Khalid, Nauman 2002. *Existence of [v] and [w] in Urdu Language.* Center for Research in Urdu Language Processing (CRULP), National University of Computer and Emerging Sciences, Lahore, Pakistan.

Saksena, Ram Babu *A History of Urdu.*

Shah, Mustafa 2002. *Urdu Nasal Consonants and Their Phonological Behavior*. Center for Research in Urdu Language Processing (CRULP), National University of Computer and Emerging Sciences, Lahore, Pakistan.