

Phonemic Inventory of Pashto

Abstract: *Pashto is one of the East Iranian groups of languages, which include, for example, Ossete (North Ossetian, south Ossetian, and Caucasus Soviet Socialist Republic) and Yaghnobi (Tajikistan) (www.afghan-network.com). In this paper its phonemic inventory is discussed. Dialect spoken in Peshawar region (Yusufzai) is considered. Documented inventories were considered as basis of the analysis and then further experiments are conducted for investigation of status of the phonemes particular to the dialect.*

Keywords: *Pashto phonemes, Yusufzai dialect, dental affricates, retroflex nasal, retroflex oral stops, retroflex palatal fricatives.*

1. INTRODUCTION

Pashto is spoken in NWFP in Pakistan including Swat and Kaghan valley, in eastern and southern regions of Afghanistan where it is the national language, second is Dari. Spoken by 9,585,000 people in Pakistan or 8.47% of population (1993 estimate); 100,000 in United Arab Emirates (1986 estimate); 14,161 in India (1994 IMA). All Pashto speaking in Pakistan are 13.2% of the population (1981 Census). It has many dialects like Kohat (Khatak), Yusufzai (Peshawar), Afridi, Shinwari, Mohmand and Shilmani but major are Khatak also called Kandahari or soft dialect and Yusufzai or hard dialect which is the literary dialect, used in schools and media in NWFP and adjacent tribal territories (Baart, 2001, p.69).

Pashto is an ancient language that is written in Perso-Arabic script. Its vocabulary contains words borrowed from Ossete, Persian, Sanskrit, Hindi, Urdu and other regional languages of Pakistan, also some Indo-Aryan languages. It is considered to be in close relation with Persian but there are certain features in Pashto that are not found in Persian e.g. there are certain consonants and vowels in Pashto that are not found in Persian like retroflex oral stops [ʈ] and [ɖ], retroflex flap [ɳ], retroflex nasal [ɳ̌] and retroflex fricatives [ʂ] and [ʐ]. Secondly in Persian, there is no gender and noun case, nouns have only categories of definiteness and number but in Pashto there is. Stress pattern is also different, in Pashto the emphasis, again unlike Persian, is not on the last syllable, but can vary. This freedom plays an important grammatical role in Pashto and is used to give different meanings to same words (www.iranianlanguages.com).

Due to these facts many researchers have suggested that origin of Pashto is not Persian rather it is either Ossete or a language from which Ossete originated (Abdul Majeed, 1992, p. 398). Ossete is also an East Iranian language that is nowadays written in

Cyrillic scripts but in past it was written in Perso-Arabic script (www.ethnologue.com).

2. LITERATURE REVIEW & PROBLEM STATEMENT

The phonemic inventory of Pashto is discussed in some books, but most rather all of the work is done on the dialects spoken in Afghanistan. The term "Pashto" actually refers to the so-called soft dialect of Afghanistan that preserves the ancient [ʈ] and [ɖ] sounds. For those parts of Pakistan where the "hard" [ʈ] and [ɖ] prevail, the language is generally referred to as Pakhto.

This paper presents the phonemic inventory of the Yusufzai dialect spoken in Peshawar and its surroundings. After collecting phonemic inventory from different sources, the aim was to verify them for Yusufzai dialect.

One of the problems was to check for the retroflex nasal [ɳ̌], dental affricates [tʃ] and [dʒ], retroflex oral stops [ʈ] and [ɖ], retroflex palatal fricatives [ʂ] and [ʐ], there were references to existence of these in various inventories of Pashto but nothing was mentioned about the dialect, (Mackenzie, 1987, p. 135). Also there was controversy about existence of [w] and [v].

Other problem was to check for the vowel system as it is said that Pashto follows seven-vowel system (www.iranianlanguages.com), also the presence of long vowels was questionable. Another one was to compare the phonemic inventory of Pashto with Urdu.

3. METHODOLOGY

3.1 Subject

First those sounds were recognized that were considered to be either absent or different from those found in other languages especially Urdu. Data was obtained from the native speakers of Pashto of different ages and also from the dictionary (Khan, 1990) of Pashto.

3.2 Preparation

To identify the acoustic features of those sounds, recordings were done in VCV sequence with combination of long and short vowels consonant initially and finally in order to have better understanding. A computer system was setup that was able to record sounds through a high quality microphone, and an amplifier. For sound recordings and analysis digital speech signal processing analysis software Praat and Winsnoori were used. This system was tested so that it gave minimal error on the data input, and was reliable throughout the experiment.

3.3 Data Recording and Processing

The subjects were asked to speak each consonant four times with different vowels word initially and finally, in their natural style. The recordings were done at 22.0 KHz frequencies, and single channel. These recordings were saved as raw sampled data, in separate files. The recordings were then processed to get spectrograms of the speech.

4. RESULTS

Some sounds were not recorded to analyze their acoustic features, as their features were obvious from their articulation. For example in many books it was described that retroflexed stops both unvoiced [ʈ] and voiced [ɖ] i.e. oral stops made by curling tongue backward, exist in Pashto. But it was found that these were alveolar stops [t] and [d] as the native speakers reported that their tongue tip touched alveolar region and it wasn't at all curled back. Existence of dental affricates [ts] and [dz] was examined.

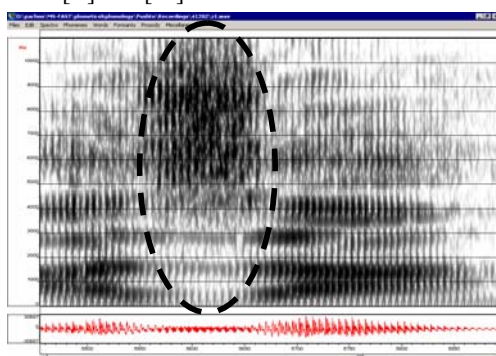


Figure 1 Spectrogram analysis of [aza:]

It is clear from the spectrogram of Figure 1 that it is not an affricate as affricates are stops in which the release of the constriction is modified in such a way as to produce a more prolonged period of frication after the release and in the above case there is no closure and release (Ladefoged, 1996). It is strong above 4kHz so is [z] (Pickett, 1999). In the same way [ts] turned out to be voiceless alveolar fricative [s].

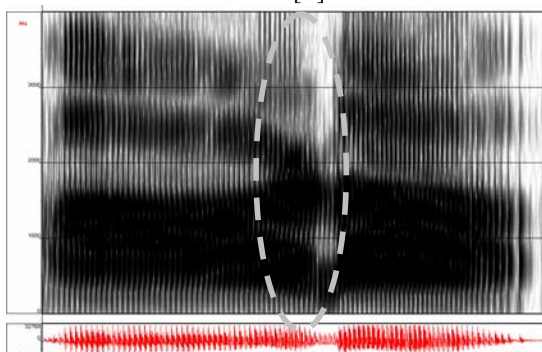


Figure 2 Spectrogram analysis of [a a:]

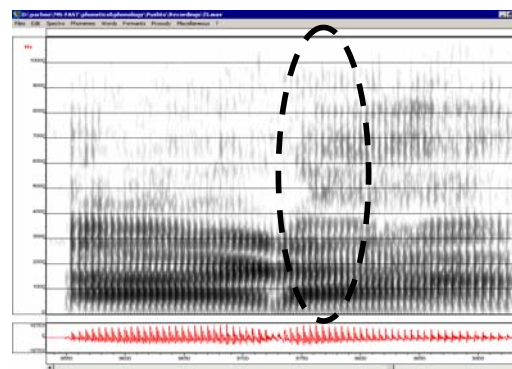


Figure 3 Spectrogram analysis of [a a:].

Existence of [ʈ] is obvious from the fact that native speakers reported that it was made with tongue curled back ahead of hard palate but well behind alveolar ridge and airflow through the nasal cavity as velum is lowered during it. It is also evident from the spectrogram of Figure 3, as it clearly shows sign of nasality; also F3 comes down and comes close to F2, which shows retroflex modification of palatal sounds so it is clearly a retroflex nasal (Ladefoged, 1996, p. 28). The following vowel is also nasalized coming out of consonant.

Existence of [w] and [v] was controversial.

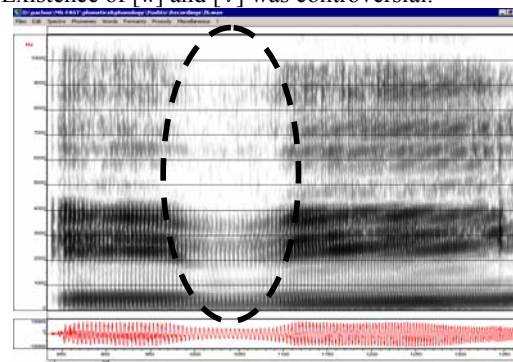


Figure 4 Spectrogram analysis of [awa:].

It is obvious from the spectrogram of Figure 4 that it is not fricative, it is the glide [w] as F1 and F2 are lowered and reduction of the frequency of F3 and higher formants (Pickett, 1999, p. 105) but F1 and F2 are not lowered as much as they should be as in case of [w] which shows that lip rounding is not that pronounced.

The conclusion drawn from the results of the experiments stated above are as follows.

Pashto system of sounds involves five distinctive tongue positions: labial, dental, post-alveolar, retroflex and velar. The retroflex position varies from person to person and may involve curling back of the tongue to make the contact with the underside of the tip, or merely retraction. Voiceless series of stops is /p/, /t/, /t/, /k/ and their corresponding voiced phonemes are /b/, /d/, /d/, /g/. There is a post alveolar affricate /t/ and its voiced version is /d/.

Voiced and voiceless aspirates series is absent in Pashto. Aspirates exist in written script but they are

present in those words that have been borrowed from other languages so in speech they are not spoken and whenever in written script there is an aspirate, its corresponding unaspirated sound is spoken (Bukhari, 1984, p. 16).

There are three nasals observed; bilabial /m/, alveolar /n/ and retroflex /ɳ/. /ɳ/ is a retroflex flap. There is an alveolar trill /r/ and an alveolar lateral /l/.

Fricatives include labiodental /f/ which is mostly replaced by /p/ while speaking (Bukhari, 1984, p. 18), alveolar /s/ and post alveolar /ʃ/. Their corresponding voiced version is only for alveolar i.e. /z/. Voiced fricative does not exist at labiodental and post alveolar places. Velar fricatives are /x/ and /ɣ/.

Glottal fricative [h] exists in Pashto. It occurs frequently in written script but is only pronounced word-initially and sometimes not spoken even word initially, word medially and finally it is normally spoken as vowel /a/ (Bukhari, 1984, p. 18). Approximates /w/ and /j/ are also in phonemic inventory of Pashto. /w/ is sometimes replaced by /b/ in speech.

Complete consonant inventory is given in the Table 1. (Appendix B). List of minimal pairs to prove the existence of these consonants is given in Appendix C.

Now the vowel system in Pashto it was found that there are seven short vowels in Pashto /a/, /e/, /i/, /ə/, /ɔ/, /o/, /u/ and three long vowels /a:/, /u:/, /i:/. Also there are four diphthongs /ey/, /ay/, /oy/ and /aw/ (<http://lrc.lib.umn.edu/pushto.htm>).

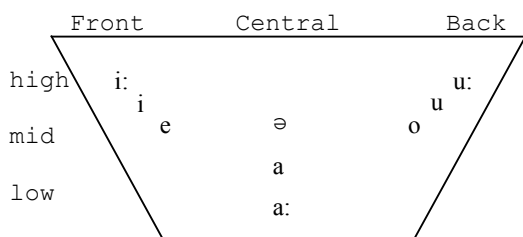


Figure 4 Vowels of Pashto

/i:/ is a high front un-rounded vowel, it is mostly found in loan words. /i/ is a middle high front vowel. It is found infrequently in Pashto and mostly in those words that are borrowed from other languages. /e/ is a mid front vowel. It is found word initially in those words that are from other languages.

/ɔ/ is a vowel that isn't found in other languages belonging to this group. It is a middle high central vowel. It is articulated in the same manner as /ə/ but there is no major lip movement. It is found mostly word finally in words like [z] meaning "heart", [m] meaning "oh man". To prove its existence there is a minimal pair

/z / Me

/za/ A term related with animals.

/ə/ is a mid central vowel. It occurs rarely in word initial position. /a/ is a low central vowel. /a:/ is also a low central vowel just like /a/ but it is considerably longer in

duration. /u:/ is a high back rounded vowel and /u/ is a mid-high back rounded vowel. /o/ is a mid back rounded vowel. Occurs very rarely in final position.

For long vowels it is noticeable that they occur mostly in those words that are either Arabic or of Persian origin e.g. /i:man/ meaning faith, /i:mam/ meaning "priest" etc. There are very few words that contain long vowels and are of Pashto originally.

The diphthong /ey/ is quite closer to /e/ except that it is longer and has a characteristic of /j/ off-glide. It is found mostly in loan words from Arabic or Persian like /beygar/ meaning "work done forcefully" etc (Khan, 1990, p. 331).

/ay/ is found rarely, /oy/ isn't found word initially and /aw/ is found rarely in word initial position,

[paw]	intelligent
[pay]	milk
[pey]	possession
[poy]	wise

All these vowels are given in the Table 2 (Appendix D), words in which they occur initially, finally or medially along with their meanings. I have tried to list those words that aren't borrowed from other languages rather than are considered to be purely Pashto in nature.

Three degrees of stress are found in Pashto: strong, medium and weak. Strong stress is comparatively free, in that it can occur on any syllable of a word, but it is mainly restricted to the first, last or penultimate syllables. It can also, particularly in verbal reflection, be mobile, though the shifts involved follow regular patterns, e.g.

[prewaθəˈl]	to fell
[preˈwaθəl]	they fell

Occasionally lexical items may be distinguished solely by stress e.g.

[aˈspa]	Mare
[aspaˈ]	a disease

Similarly it is also used to change tense, gender and number (Comrie, 1987, p. 137).

If phonemic inventory of Pashto (Yusufzai dialect) is compared with that of Urdu, we find that both of them share more or less same set of consonants, retroflex nasal /ɻ/ is missing in Urdu, although it is found in other regional languages like Punjabi, Saraiki and Sindhi while voiced post-alveolar fricative /ʒ/ is missing in Pashto. Also there are no aspirates in Pashto while aspirated stops exist in Urdu.

In case of vowels, Urdu has both long and short vowels while Pashto mostly uses short vowels and those words that contain long vowels and diphthongs are of either Arabic or Persian origin. /ɔ/ is found in Pashto and it is absent in Urdu.

4 DISCUSSION

There were lots of controversies among the native speakers while working on the place of articulation and articulators regarding consonants. These have been resolved with the acoustic analysis of the sounds and in

case of /w/, degree of lip rounding varied from person to person which effects the formant transitions.

Similar was the case with vowels, especially when those speakers who were very well aware of Urdu were asked to compare those vowels with those found in Urdu, many controversies arose then. The acoustic analysis of vowels has not been done so all the results that have been established regarding them are open for discussion and further experimentation.

5 SUMMARY

The work that has been already done was on Khatak dialect of Pashto that included retroflex stops, retroflex fricatives, dental affricates and retroflex nasals. Most of these are absent in Yusufzai dialect of Pashto and it seems that Yusufzai dialect under Indo-Aryan influence has lost those sounds; it has just retroflex nasal and retroflex flap. Also there is a tendency to replace /f/ with /p/ and /w/ with /b/ in speech, /h/ is also avoided and mostly replaced with vowel.

Regarding vowels, it makes use of mostly short vowels and it has a central high vowel / / that is not found in other languages of this family. Long vowels and diphthongs are mostly found in those words that are borrowed from other languages mostly Arabic and Persian.

Stress is used to distinguish similar words having different meanings and is mostly lexical.

REFERENCES

- Baart, Joan L. G. and Baart, Esther L. Bibliography of languages of northern Pakistan. NIPS/SIL. 2001.
- Bukhari, Khayal. Ataleeq Pashto. De chapazay, Peshawar. 1984.
- Comrie, Bernard. *The major languages of the middle east and Africa*. Rontledge, London 1990.
- Khatak, Khan. Pashto Urdu Lughat. Pashto Academy University of Peshawar, 1990.
- Ladefoged, Peter Ladefoged and Maddieson, Ian. *The sounds of the world languages*. Blackwell publishers Inc, Massachusetts, USA. 1996.
- Majeed, Dr. Memon Abdul Sindhi. *Lisaniyat-e-Pakistan*. Muqtadra qaumi zuban, Islamabad. 1992.
- Massica, Collin P. *The Indo-Aryan Languages*. Cambridge University Press, 1991.
- Pickett, J. M.. *The Acoustics of Speech Communication*. Allyn & Bacon. 1999.
- Sabiri, Makhdoom and Raza, Afzal. *Pashto Urdu bool chaal ki aasan kitab*. Malik book depot, Lahore. 1992.
- <http://lrc.lib.umn.edu/pushto.htm>
- www.iranianlanguages.com
- www.afghan-network.com
- www.ethnologue.com

APPENDIX A

INDO-IRANIAN	WEST	OLD PERSIAN	PARTHIAN MIDDLE PERSIAN	PERSIAN KURDISH BALUCHI TAJIKI
	IRANIAN			
	EAST	<u>Avestan</u>	<u>Sogdian</u> <u>Khotanese</u> <u>Khwarezmian</u> <u>Bactrian</u> <u>Old Ossetic</u>	<u>Pashto</u> <u>Ossetic</u>
	Dard			Shina, Khowar, Kafiri, Kashmiri
		Sanskrit, Pali <i>Prakrits</i>	<i>Prakrits</i>	Lahnda Sindhi Panjabi Rajasthani Gujrati Marathi Konkani Oriya Bengali Assamese Bihari Hindi Urdu Nepali Sinhalese Romany
	Indic Sanskritic			

Figure A.1 Hierarchy chart for Pashto
Italics denote dead languages.

APPENDIX B

Table 1 Consonants of Pashto

	Bilabial	Labio-dental	Dental	Alveolar	Retroflex	Post-Alveolar	Velar	Glottal
Plosive	p b		t d	t d			k g	h
Nasal	m			n				
Affricate						t d		
Fricative	f			s z			x	
Trill				r				
Flap								
Approximate	W					j		

APPENDIX C

List of minimal pairs in order to prove phonemic existence of consonants given in the results.

1. /ba al / armpits
/paxal/ to cook
2. /dəp/ sound of a falling body
/təp crowd
3. /tat/ mat
/dat/ arch
4. /ka / mine
/ga / barren land
5. /ma:l/ goods
/na:l/ horse shoe
6. / tar / suggestion
/ dar / sacrifice
7. /sad/ noble
/zad/ provisions
8. / und/ blind
/ rund / flowing
9. / a / back
/ra/ to give.
10. /ba al / armpits
/paxal/ to cook
11. /sxra:/ rock
/sx a:/ extravagance

APPENDIX D

Table 2 Vowels of Pashto

Phonetic representation	Meaning
jəba	Tongue
jəm	I m
kəle	Town
ləs	Ten
pəg	Six
gənd e	Bald
bəs	Enough
medə	Grinded
pə	Night
nikə	Grandfather
pozə	Mat
spin	White
in	Green color
spe	Dogs
mami:	Auntie
mali:	Gardner

ni ta	No
peza:r	Shoe
wezər	Feather
z	Heart
m	Oh man
mami:	Auntie
gul	flower
u:gud	Tall
u:ləs	Masses
u:gdə	Tallness
Phonetic representation	Meaning
u: wa	Desire
la:s	Hand
awdes	Ablation
awratul	To fly
sile	Wind
aw	Calling
daway	Passage
yaw	One
paw	Intelligent
poy	Wise
poya	Land
poyədul	wise
loy	Old man
zoy	Boy
pey	Possession
pay	Milk
selane	Guide
winzul	To wash