



# Challenges of Accent and vowels for Sindhi Speech Recognition System

Fida Hussain Khoso<sup>a,b</sup>, Dil Nawaz Hakro<sup>c</sup>, Syed Zafar Nasir<sup>a</sup>

<sup>a</sup>Indus University Karachi Pakistan, <sup>b</sup>Dawood University of Engineering and Technology Karachi Pakistan

<sup>c</sup>Institute of Information and Communication Technology, University of Sindh, Jamshoro

## ABSTRACT

While talking and writing in Sindhi language, many challenges are faced because of the large number of 52 characters or alphabets. Vowels and the accent keep changing in fluency of speaking and writing. Due to the different varieties of languages in the world and the dearth of computer scientists in the field of Speech Recognition, it is considered difficult area of study and is the least advanced field of Artificial Intelligence. More specifically, the difficulties are faced in the speech recognition for languages like Arabic and its adapting languages such as Sindhi, Pashto, Urdu, and others. The script and sounds in every language are directly proportional to each other i.e. the shorter script has less sounds while the longer script has more sounds. We developed a system for speech to text recognition system for Sindhi language with the help of Sphinx model. We have also tested the different datasets through the input in various phases and compare the results and accuracy of the vowels and accents through the proposed system.

**Key words:** Accent, Vowels, alphabets, consonants, phonemes

## 1. INTRODUCTION

When we determine the acoustic features of the language, it helps to bring the sounds of similar acoustic properties with accent, vowels, phoneme, and meaningful natural class. One of the important goals of this research study is by learning acoustic phonetic features for the compilation of language phoneme and accent. There are twelve sounds of Sindhi, not the part of language alphabet, are hereby studied. The formants and fundamental frequency in phonological structure of Sindhi and they are main acoustic features. The incentive for such research comes from the detailed knowledge of the sound characteristic of language elements, which has wide variety of applications. To develop and accurate speech recognition system and speech recognizer of speaker independent are required [1].

To develop a speaker independent system, we have to require the accent variation of speaker calling from all over Sindhi.

The major languages spoken in Sindh are Sindhi, Urdu, Punjabi, Pashto, Balochi, and Seraiki having different accents. Accent of each language's speaker is quite different from others. The complexity to understand the various accents in Sindhi, we have evaluated and implemented some various techniques. To work in mild noise environment the system is needed to be robust enough. By the preprocessing of speech, input signal the goal is achieved. To evaluate the output of the system the strict field-testing of speech recognition system is required in real world environment [2].

Any suitable design of Sindhi ASR system the acoustics features of language sounds are required. A speech database comprised of the voice samples of native Sindhi speakers is evolved to understand the particular acoustic variation of basic sounds of Sindhi. To design a state of the art prototype the basic building block of feature are created by this research study. These are environment and genders free which continuous as conversational ASR for Sindhi language. A speech processing is a very wide area covering the applications and different disciplines which include the sophisticated speech designing for state-of-the-art speech recognition system. The history of research and development process in this field varies a period of many years. In the field of speech processing the challenges are coming in the efficient and accurate designing of a speech recognition system [3].

Usually the sequence of phonemes are represented by spoken sentence. They are defined in a language based on their linguistic relevance. The acoustically homogeneous units are needed by a phone. With reference to machine identification speech auditory compactness of the units are important because the spectral similarity of various consideration of a unit enables to identify a recognition system. The identification and recognition of sound and acoustics are very important and essential for the system. So that we do implement the models for acoustic and sound to comply with the required results. The units should be fairly various and acoustically homogeneous from each other. With the phonemes, the units need not to be having 1 to 1 correspondence. It is required that between these units and the phonemes of the language there is a fair degree of association. Furthermore, it is worthwhile in acoustic phonetic of any language are to be developed. All these exercises and

technical practices are to be elaborated in accordance to the basic rules and principles of computer technology and sound system [4].

The basic form of conversation between human beings is speech. Speech sound can be converted into correspondence text and this process is called speech recognition. During last few decades much work has been done and developed for speech recognition technology. Apart from all development and researchers, still there is a need to resolve many important challenges. In the field of signal processing the speech, processing is one of the major fields. When we put human speech into a machine there is a need to develop area of techniques in speech recognition. When the computer device evolved and developed, its scope and power was very limited. With the passage of time, the computer technology grew and flourished very fast then this revolution led towards automatic speech recognition system [5].

The people motivate desire to build mechanical models using research in speech processing and communication to emulate human verbal communication capabilities. The most natural form of human communication and speech processing is speech which has been one of the most exciting areas of the signal processing. The computer to follow human voice commands and understand human languages have made the speech recognition technology possible. To develop systems and techniques for speech input to machine the main aim is the speech recognition area. The spoken words or samples are needed to be collected from various regions and areas so that the different accents can be used on the basis of environment [6].

## 2. MATERIALS AND METHODS

Although the Persian, an Arabic script are the bases of writing system in Sindhi language, but many sounds have been symbolized with modification in Sindhi language. The phonological systems of other Indo-Aryan languages are resembled mostly. In Sindhi language, there are 10 vowels and 43 consonants phonemes are unique [6]. When a person speaks in microphone are on telephone, the speech acquisition starts. The acoustic signal is fabricated when the sound pressure wave is produced by the act of speaking. An electronic device understands the analog signal, which is received by telephone or microphone which convert it into analog signal. To store the analog signal on a computer lately it must be converted to a digital signal [7].

### 2.1 Compliance for Sindhi Speech Recognition System

It is known to every person of Sindh that when, how and by whom the current Sindhi alphabets were developed, and what improvements were suggested by the committee under the chair of Mr. Elles during the British rule, when it was given the status of national language. Later in 1913, Sindhi alphabets were made qualitative and easy after deliberations on those aspects, which were not considered by Mr. Elles committee in a gathering chaired by Late Mirza Qaleech Baig. Despite that, the Sindhi Alphabets witnessed many issues with the passage of time and addition of new words in the language as it was influenced by the modern trends. Some of the issues have been discussed below [8].

### 2.2 Accent

Accent is dependent on the language of each region and area. If a person is from Thar, then his accent will be like ‘Tharis’ and his language will be different from common interior Sindh. The different accents of Sindhi language are according to different regions and some of them are Utradi, Lari Thattvi and others. These are all accents of Sindhi language identified as the words of Sindhi language by regions. Analysis of existing sindhi sounds accents and dialects Such as dhal dhobhi , daal dedar , ghay ghero, ngi singn, ray mari one word used for many other languages such as I used as toon tawahaan and tawaheen [9].

### 2.3 Vowels

In Sindhi language, there are 10 vowels and 43 consonants phonemes are unique. It is necessary that to captured and understand the systematic variation which occur in natural speech are useful in acoustic phonetic study of sounds. For the synthesis of high-quality speech there should be better modeling of identification units which lead to better speech. It is a fact in language writing that vowels occur more frequently than consonants. So practically the meaningful study of vowels for moderate size database is more useful and meaningful that of consonants [10], [11].

### 2.4 Alphabets

If comparison is made between the numbers of alphabets of major languages of the world, it is revealed that the English Language consists of 26 alphabets, Arabic has 28, Urdu has 37, Persian has 32, and Sindhi has 52 alphabets. As Sindhi has the greatest number of alphabets, different issues are being faced in forming words and even the letters like using two letters for a single sound, which means that there may be differences in spoken and written scripts. Example includes bh‘پ’, th‘ٹ’, chh‘چ’, ch‘چ’ etc[13].

### 2.5 Similar Sounds

There are some specific words in sindhi language for which the direct words are not used but some supporting words can be used for example: the ‘چ’ chh ‘چٹی’ means umbrella and ‘ش’ sh ‘شینه’ it means lion and ‘ر’ and ‘ڑ’ and ‘ماڑی’ there is little difference between them. “ٹ، ج، گ، پ”.

### 2.6 Dots

The letters in Sindhi alphabets can be bifurcated with respect to dots on each letter and their position. It may be noted that there are 18 letter that do not contain any dots, there are 12 letters with only one dot, there are 11 letters that contain two dots, while there are 6 and 5 letters with three and four dots respectively. This bifurcation of Sindhi alphabets is depicted in the table:1 shows the bifurcation of Sindhi alphabets[14],[15].

**Table: 1** Shows Glyph fluctuation in Eastern Arabic Indic Digits

Code Point	Digit	Persian	Sindh i	Urdu
U+06F4	4	۴	۴	۴
U+06F5	5	۵	۵	۵
U+06F6	6	۶	۶	۶
U+06F7	7	۷	<	<

**Table: 2** Shows Sindhi numbers from 0 to 10

۰	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
ڏھ	نو	اٺ	ست	ڇھ	پنج	چار	ٽي	ٻ	ھڪ	ٻڙي

**3. Dialects**

The research on Sindhi Language and its dialects reveal that the melody and intonation of dialects are different from each other with respect to the phonology. This difference not only hinders the sound development of various components of Sindhi speech processing applications but also creates complications in understanding, recognition, and translation of Sindhi language [9]. Some dialect of Sindhi language are following as, Utradi, Lasi, Lari, Thareli and wicholy are the different dialects and accents which are the basis on their region.

**3.1 Noise**

The noisy environment also poses another challenge that includes the recognition of keywords from the sentences mixed with the noise background. A noise suppression approach has been proposed for the speech controllers so that the accuracy can be improved. Various noise suppression filters work to integrate and process signal in a better way. The environment must be free from any interruption inside or outside noises and the recording must be very clear and fair [12],[16].

**4.1 samples collection for check Accents and Vowels from different districts of Sindh**

We have collected the different voices of male and female from different districts of Sindh for the purpose to check the accuracy of accents and vowels of the people from region wise. We have selected ten districts of Sindh for data sampling through voice recorder; microphone and other resources have been used for check the accuracy of voice. We know very that in Sindh, various types of people are living with different dialects and regions their living style and speaking style are very different. In below two tables have been given for district wise and different dialect which are used in region wise.

**Table:3** Shows List of the Selected Districts

S.No	The list of ten districts of Sindh	سنڌ جي 10 ضلعن جا نالا
1	Jacobabad (Northern Sindh)	جيڪب آباد (اتر سنڌ)
2	Shikarpur (Northern Sindh)	شڪارپور (اتر سنڌ)
3	Larkana (Northern Sindh)	لاڙڪاڻو (اتر سنڌ)
4	Dadu (Lower North)	دادو (هٿيون اتر)
5	Naushararo Firoze (Central Sindh)	نوشهروفيروز (وچولو سنڌ)
6	Umarkot (East Sindh)	عمر ڪوٽ (اوڀر سنڌ)
7	Tharparkar (East Sindh)	ٿرپارڪر (اوڀر سنڌ)
8	Badin (Southern Sindh)	بدين (ڏکڻ سنڌ)
9	Thatta (Southern Sindh)	ٿٽا (ڏکڻ سنڌ)

	Sindh)	
10	Karachi (west Karachi)	ڪراچي (اولهه)

**Table4:** Dialects region wise

DIALECTS	LOCATION	NO OF SPEAKERS
Utradi	Larkana	05
	Shikarpur	04
	Pano Aqil	03
Lari	Badin	04
	Sujawal	04
	Thatta	03
Middle	Nawabshah	05
	Jamshoro	03
	Hyderabad	04
Thari	Islamkot	03
	Mithi	03

**4.2 Experimentation Setup for Accent**

We have collected speech samples and training data from direct observations. In the first phase, the three respondents were collected which were dependent on random sampling and the it was to be pronounced as five words sequence which were used as code of training. To train the Neural Network these responses will be used which were recorded for the production of training dataset. The training data was used to train the network. To train Artificial Neural Network in three different ways, the speech data were used. Once the training of model was complete, the proper testing data was tested. For the addition with dataset and training the sample of 7 testing data respondents were achieved. The responses were testified with than 10 different sequences of the Sindhi words. Excluding some of non-native speakers all the despondent were native. It was obvious that the recorded data transcription was self-created. Having the heterogeneity of accent which was collected from test data of the speech of native speaker was included from male and female non-Sindhi speakers.

**Table5:** Shows collected samples for check accuracy

S.No:	Speakers	Sample Count		Mother Tongue	Gender
		Training	Testing		
1.	Hassan	50	10	Urdu	Male
2.	Saima	100	10	Sindhi	Female
3.	Ahmed	--	10	Sindhi	Male
4.	Iqra	--	10	Sindhi	Female
5.	Kashif	50	10	Sindhi	Male
6.	Fida	--	10	Urdu	Male
7.	Dil Nawaz	--	10	Sindhi	Male

## 5.1 Results and Discussion

### 5.2 First Test

The Sphinx was proved for all the test sets the net output was correct. From Ahmed three out of five sample sequences were correctly recognize and two words were identified in correctly in a sequence. The word recognition ratio accuracy is 60. After the other entire sample, sequences of speakers gave poor results than Iqra. Few of the words were recognized correctly in the sequences. However, in the sequence few of the words were accurately identified.

Number of Speakers=1

Error Rate 40%

Accuracy 60%

Total Words = 5

Detected Words = 3

Sphinx Results

Aligning results to find error rate

Errors per sentence: 0.0% (0/5) errors per word: 0.0% (0/5)

```
INFO: continuous.c(275): Ready...
INFO: continuous.c(261): Listening...
INFO: cmn_live.c(120): Update from < 48.02 23.27 -8.58 -14.10 3.12 10.34 -4.10 -12.80 0.55 -2.02 >
INFO: cmn_live.c(138): Update to < 47.66 23.05 -7.75 -13.40 3.13 10.00 -3.65 -12.61 0.17 -2.29 >
INFO: ngram_search_fwdtree.c(1550): 207 words recognized (2/fr)
INFO: ngram_search_fwdtree.c(1552): 522 senones evaluated (6/fr)
INFO: ngram_search_fwdtree.c(1556): 280 channels searched (3/fr), 0 1st, 280 last
INFO: ngram_search_fwdtree.c(1559): 280 words for which last channels evaluated (3/fr)
INFO: ngram_search_fwdtree.c(1561): 0 candidate words for entering last phone (0/fr)
INFO: ngram_search_fwdtree.c(1564): fwdtree 0.00 CPU 0.000 xRT
INFO: ngram_search_fwdtree.c(1567): fwdtree 1.84 wall 2.065 xRT
INFO: ngram_search_fwdflat.c(302): Utterance vocabulary contains 4 words
INFO: ngram_search_fwdflat.c(948): 254 words recognized (3/fr)
INFO: ngram_search_fwdflat.c(950): 690 senones evaluated (8/fr)
INFO: ngram_search_fwdflat.c(952): 400 channels searched (4/fr)
INFO: ngram_search_fwdflat.c(954): 400 words searched (4/fr)
INFO: ngram_search_fwdflat.c(957): 157 word transitions (1/fr)
INFO: ngram_search_fwdflat.c(960): fwdflat 0.00 CPU 0.000 xRT
INFO: ngram_search_fwdflat.c(963): fwdflat 0.00 wall 0.001 xRT
INFO: ngram_search.c(1250): lattice start node <s>.0 end node </s>.81
INFO: ngram_search.c(1276): Eliminated 0 nodes before end node
INFO: ngram_search.c(1381): Lattice has 18 nodes, 26 links
INFO: ps_lattice.c(1376): Bestpath score: -1905
INFO: ps_lattice.c(1380): Normalizer P(0) = alpha(</s>:81:87) = -118519
INFO: ps_lattice.c(1437): Joint P(0,S) = -130865 P(S|0) = -12346
INFO: ngram_search.c(872): bestpath 0.00 CPU 0.000 xRT
INFO: ngram_search.c(875): bestpath 0.00 wall 0.003 xRT
توهان
```

Fig:1 shows the tested word in Sindhi” توهان “

```
INFO: continuous.c(275): Ready...
INFO: continuous.c(261): Listening...
INFO: cmn_live.c(88): Update from < 49.49 24.62 -5.07 -9.86 0.93 10.18 -1.00 -14.89 -1.54 -5.29 -1.62 -9.54 -0.42 >
INFO: cmn_live.c(105): Update to < 49.36 24.25 -4.88 -9.25 1.12 10.20 -0.57 -15.35 -1.86 -5.16 -1.49 -9.12 -0.83 >
INFO: cmn_live.c(120): Update from < 49.36 24.25 -4.88 -9.25 1.12 10.20 -0.57 -15.35 -1.86 -5.16 -1.49 -9.12 -0.83 >
INFO: cmn_live.c(138): Update to < 48.55 23.62 -4.76 -9.13 1.54 10.08 -0.84 -14.82 -1.51 -4.69 -1.36 -8.81 -0.71 >
INFO: ngram_search_fwdtree.c(1550): 254 words recognized (3/fr)
INFO: ngram_search_fwdtree.c(1552): 657 senones evaluated (8/fr)
INFO: ngram_search_fwdtree.c(1556): 341 channels searched (4/fr), 0 1st, 341 last
INFO: ngram_search_fwdtree.c(1559): 341 words for which last channels evaluated (4/fr)
INFO: ngram_search_fwdtree.c(1561): 0 candidate words for entering last phone (0/fr)
INFO: ngram_search_fwdtree.c(1564): fwdtree 0.02 CPU 0.019 xRT
INFO: ngram_search_fwdtree.c(1567): fwdtree 2.35 wall 2.871 xRT
INFO: ngram_search_fwdflat.c(302): Utterance vocabulary contains 4 words
INFO: ngram_search_fwdflat.c(948): 296 words recognized (4/fr)
INFO: ngram_search_fwdflat.c(950): 708 senones evaluated (9/fr)
INFO: ngram_search_fwdflat.c(952): 392 channels searched (4/fr)
INFO: ngram_search_fwdflat.c(954): 392 words searched (4/fr)
INFO: ngram_search_fwdflat.c(957): 171 word transitions (2/fr)
INFO: ngram_search_fwdflat.c(960): fwdflat 0.00 CPU 0.000 xRT
INFO: ngram_search_fwdflat.c(963): fwdflat 0.00 wall 0.002 xRT
INFO: ngram_search.c(1250): lattice start node <s>.0 end node </s>.41
INFO: ngram_search.c(1276): Eliminated 0 nodes before end node
INFO: ngram_search.c(1381): Lattice has 10 nodes, 8 links
INFO: ps_lattice.c(1376): Bestpath score: -1238
INFO: ps_lattice.c(1380): Normalizer P(0) = alpha(</s>:41:80) = -99482
INFO: ps_lattice.c(1437): Joint P(0,S) = -112382 P(S|0) = -12900
INFO: ngram_search.c(872): bestpath 0.02 CPU 0.019 xRT
INFO: ngram_search.c(875): bestpath 0.00 wall 0.005 xRT
نالو
```

Fig: 2 shows the tested word in Sindhi ” نالو “

```
INFO: continuous.c(275): Ready...
INFO: continuous.c(261): Listening...
INFO: cmn_live.c(120): Update from < 49.78 24.93 -12.12 -19.04 1.77 13.52 -5.73 -13.72 1.38 -1.92 0.74 -5. -5. >
INFO: cmn_live.c(138): Update to < 49.89 24.53 -12.87 -17.85 2.47 12.54 -5.85 -13.53 0.82 -1.68 0.85 -6. -6. >
INFO: ngram_search_fwdtree.c(1550): 188 words recognized (2/fr)
INFO: ngram_search_fwdtree.c(1552): 483 senones evaluated (5/fr)
INFO: ngram_search_fwdtree.c(1556): 270 channels searched (3/fr), 0 1st, 270 last
INFO: ngram_search_fwdtree.c(1559): 270 words for which last channels evaluated (3/fr)
INFO: ngram_search_fwdtree.c(1561): 0 candidate words for entering last phone (0/fr)
INFO: ngram_search_fwdtree.c(1564): fwdtree 0.00 CPU 0.000 xRT
INFO: ngram_search_fwdtree.c(1567): fwdtree 2.44 wall 2.745 xRT
INFO: ngram_search_fwdflat.c(302): Utterance vocabulary contains 3 words
INFO: ngram_search_fwdflat.c(948): 247 words recognized (3/fr)
INFO: ngram_search_fwdflat.c(950): 408 senones evaluated (5/fr)
INFO: ngram_search_fwdflat.c(952): 306 channels searched (3/fr)
INFO: ngram_search_fwdflat.c(954): 306 words searched (3/fr)
INFO: ngram_search_fwdflat.c(957): 124 word transitions (1/fr)
INFO: ngram_search_fwdflat.c(960): fwdflat 0.00 CPU 0.000 xRT
INFO: ngram_search_fwdflat.c(963): fwdflat 0.00 wall 0.001 xRT
INFO: ngram_search.c(1250): lattice start node <s>.0 end node </s>.49
INFO: ngram_search.c(1276): Eliminated 0 nodes before end node
INFO: ngram_search.c(1381): Lattice has 12 nodes, 13 links
INFO: ps_lattice.c(1376): Bestpath score: -1137
INFO: ps_lattice.c(1380): Normalizer P(0) = alpha(</s>:49:87) = -82882
INFO: ps_lattice.c(1437): Joint P(0,S) = -94258 P(S|0) = -11376
INFO: ngram_search.c(872): bestpath 0.00 CPU 0.000 xRT
INFO: ngram_search.c(875): bestpath 0.00 wall 0.005 xRT
اھي
```

Fig: 3 shows the tested word in Sindhi “ اھي “

### 5.3 Second Test

The recording of five speakers were made in the second test. The names of various people and few of them were scrutinized. The somehow mix type net result was achieved. The word recognition accuracy ratio was 12.5. 87.5% is error rate. Any how some of the words were correctly recognized in the sequence.

Number of Speakers = 5

Error Rate 87.5%

Accuracy 12.5%

Total Words = 24

Detected Words = 3

Sphinx Results

Aligning results to find error rate

SENTENCE ERROR: 11.7% (14/120) WORD ERROR

RATE: 11.7% (14/120)

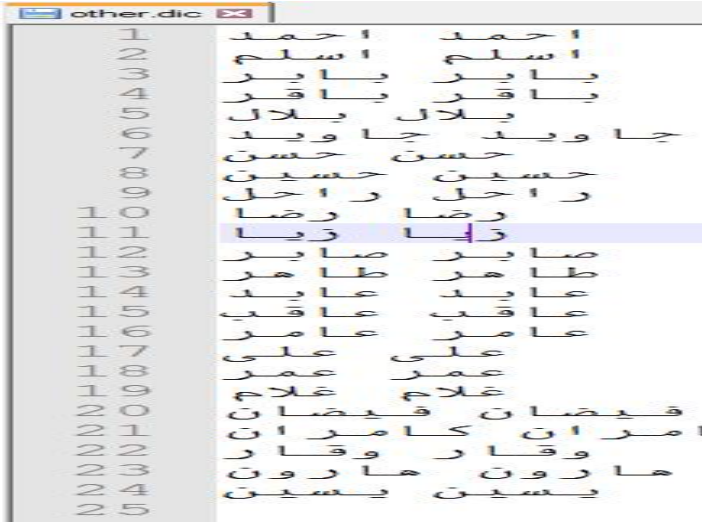


Fig: 4 Shows the list of tested word

### 5.4 Third Test

In the third test the recording of 5 presenters was done. The verified numbers in Sindhi were from 1 to 10 and then out of ten one. The somehow mix type net result was achieved. The numbers recognition accuracy ratio was 10.0. 90% ratio calculated as error rate.

Number of Speakers = 5

Error Rate 90%

Accuracy 10%

Total Words = 10

Detected Words = 1

Sphinx Results

Aligning results to find error rate

Errors per sentence: 0.0% (0/50) errors per word: 0.0% (0/50)



Fig: 5 Shows the tested names

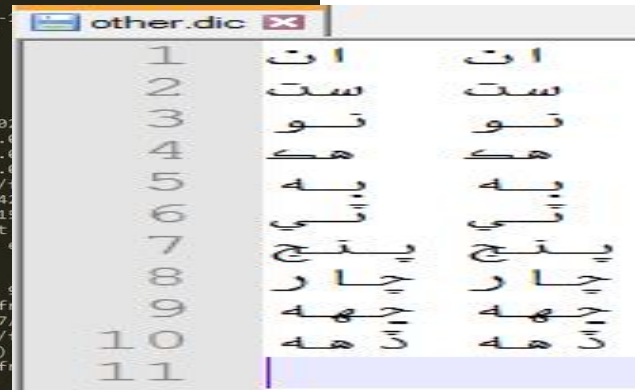


Fig: 7 Shows the numbers in Sindhi

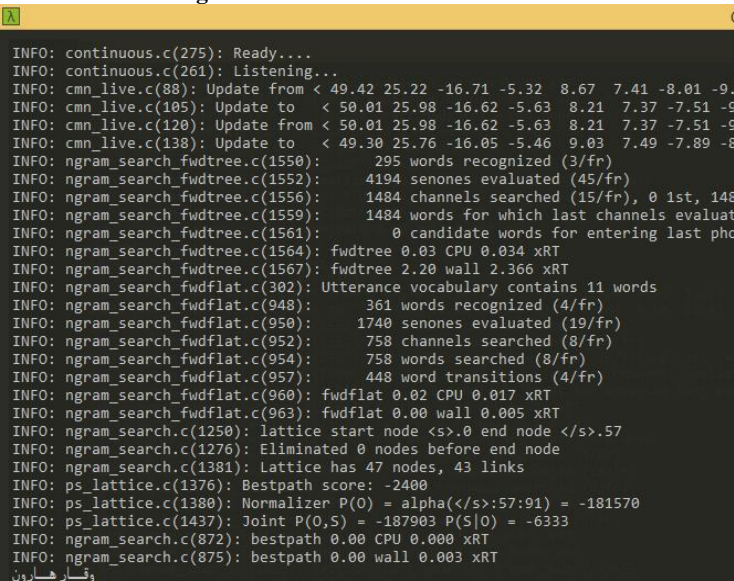


Fig: 6 Shows the tested different names

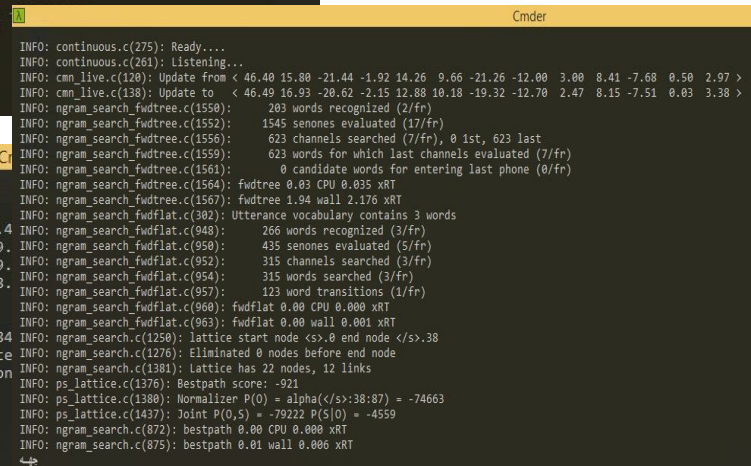


Fig: 8 The detected number six

## 6. CONCLUSION

In this present research paper, the complications and general complexities have been mentioned for Sindhi speech recognition system. This current research will provide and give helpful guidance for empirical approach of Sindhi speech recognition system. To frame the Sindhi speech recognition system, a comprehensive Sindhi speech database

and acoustic models are required as we have compared and reviewed that the available databases cannot fulfill the needs of Sindhi language speech recognition. We are trying our hard to develop and create such speech recognition system having different dialect and accents of the province of Sindh. In future, we will be thoroughly comparing and reviewing the speech recognition system with huge databases of Arabic, Urdu, English, and Persian languages. This task is very time consuming and concentration so we are facing so many problems and difficulties.

## REFERENCES

- [1] Keerio, Ayaz, Naveed Channa, Bhargav Mitra, Rupert Young, and Chris Chatwin. "Acoustics of isolated vowel sounds of Sindhi." *Sindh University Research Journal-SURJ (Science Series)* 46, no. 2 (2014).
- [2] Qasim, Muhammad, Sohaib Nawaz, Sarmad Hussain, and Tania Habib. "Urdu speech recognition system for district names of Pakistan: Development, challenges and solutions." In 2016 Conference of the Oriental Chapter of International Committee for Coordination and Standardization of Speech Databases and Assessment Techniques (O-COCOSDA), pp. 28-32. IEEE, 2016.
- [3] Furui, Sadaoki. "50 years of progress in speech and speaker recognition research." *ECTI Transactions on Computer and Information Technology (ECTI-CIT)* 1, no. 2 (2005): 64-74.
- [4] Samudravijaya, K., P. V. S. Rao, and S. S. Agrawal. "Hindi speech database." In *Sixth International Conference on Spoken Language Processing*. 2000.
- [5] Shariah, Mohammad AM Abu, Raja N. Ainon, Roziati Zainuddin, and Othman O. Khalifa. "Human computer interaction using isolated-words speech recognition technology." In 2007 International Conference on Intelligent and Advanced Systems, pp. 1173-1178. IEEE, 2007.
- [6] Mahar, Javed Ahmed, and Ghulam Qadir Memon. "Sindhi part of speech tagging system using wordnet." *International Journal of Computer Theory and Engineering* 2, no. 4 (2010): 538.
- [7] El Choubassi, M. M., H. E. El Khoury, CE Jabra Alagha, J. A. Skaf, and M. A. Al-Alaoui. "Arabic speech recognition using recurrent neural networks." In Proceedings of the 3rd IEEE International Symposium on Signal Processing and Information Technology (IEEE Cat. No. 03EX795), pp. 543-547. IEEE, 2003.
- [8] Prof. Ijaz Qureshi, (2008). *Sindhi language past, present and future, Sindhi Adbi Sangat Sindhi*.
- [9] Jennifer, S. C. (2006). **The Sindhi language. In Brown (ed.) Encyclopedia of Language and Linguistics** 2ed., Vol. 11, 384-386). Oxford: Elsevier.
- [10] Hakro, D. N. I. A. Ismaili, A. Z. Talib. Z. Bhatti. And G. N. Mojai. (2014), **Issues and Challenges in Sindhi OCR**, Sindh University Research Journal (Science Series) 46(2), 143-152.
- [11] Raza, S., F. Z. Agha, and R. Usman, (2004). **Phonemic inventory of Sindhi and acoustic**
- [12] Rasipuram, R., & Doss, M. M. (2016), 'Articulatory feature based continuous speech recognition using probabilistic lexical modeling', *Computer speech & language* 36, 233 – 259.
- [13] Mahar, Javed Ahmed, and Ghulam Qadir Memon. "Phonology for Sindhi letter to sound conversion." *Journal of Information and Communication Technology* 3, no. 1(2009): 11-20.
- [14] Ashok Kumar, Vikas Mittal. (2019). **Speech Recognition: A Complete Perspective**, Vol-7(6).
- [15] Sodhar, Irum Naz, Akhtar Hussain Jalbani, and Muhammad Ibrahim Chena, (2019) "Identification of Issues and Challenges in Romanized Sindhi Text." Editorial Preface from the Desk of Managing Editor... 10, no. 9.
- [16] Ason, C. Andy, and Sandeep Kumar. 2020 "An Appraisal on Speech and Emotion Recognition Technologies based on Machine Learning." *language* 67: 68.