I J C T A, 9(20), 2016, pp. 259-266

An Automatic Spontaneous Live Speech Recognition System for Punjabi Language Corpus

Yogesh Kumar* and Navdeep Singh**

ABSTRACT

In spontaneous Punjabi speech model, the speech is basically non-planed and non designed, there are generally depicted by repetitions, preservation, wrong start, half-spoken words and non-planned words, silence gap etc. In a system of Punjabi speech detection including vocabulary, the identification needs the evaluation among the audio signal of the utterance and the variety of utterances of the vocabulary. The primary purpose of the research paper is to build the automatic spontaneous speech recognition system for the Punjabi language. The Punjabi speech system is trained and tested by the voice of different male and female speakers and GUI for spontaneous Punjabi live speech system also has been developed. The performance analysis of the speech system is measured by using live speech method. The recognition accuracy for Punjabi speech sentences is 91.17% and 85.38% for different Punjabi words. The sphinx framework and java programming language are used to develop an automatic spontaneous Punjabi live speech model.

Keywords: Acoustic Model, ASR, Filler, Dictionary File, Phone File, Transcription, Field File.

1. INTRODUCTION

Automatic speech recognition (ASR) is the method of mapping a sound waveform into a text and situates of words which is alike to the sequence being communicated by the verbal words. [1] Presently, there have been numerous of ASR products available. However, a lot of problems still present in real-world ASR applications. The identification correctness of a machine is, in the majority, distant from that of a human listener, and its presentation would debase noticeably with little variation of dialogue indications or talking surroundings. Some of the works have been performed in the area of isolated, connected word, continuous

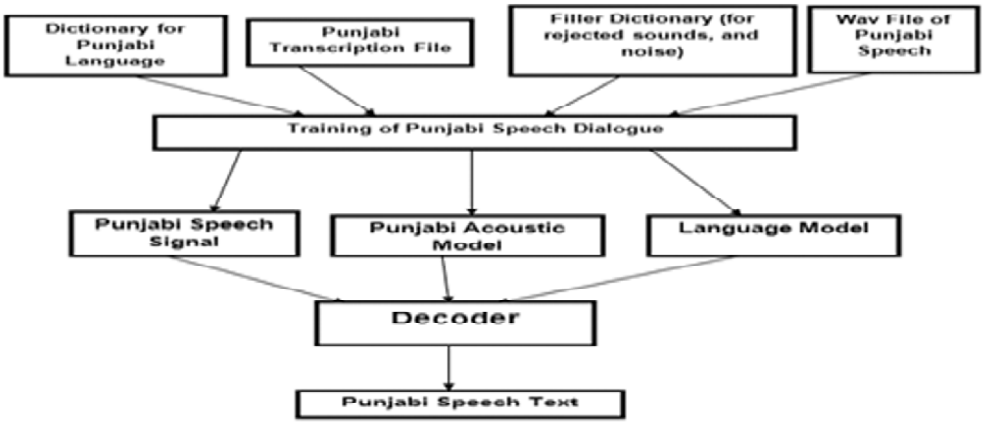


Figure 1: Automatic Speech Recognition [6] System for Punjabi Language

* Research Scholar, Punjabi University, Patiala-147001, India, *Email: Yogesh.arora10744@gmail.com*

** Associate Professor, Mata Gujri College, Sri Fatehgarh Sahib-140406, India, *Email: navdeep_jaggi@yahoo.com*

speech recognition system for the Punjabi language. But spontaneous speech recognition for the Punjabi language [6, 7] is still a challenge for the researcher due to the recognition accuracy of the speech.

To construct a general purpose Spontaneous speech recognizer for the Punjabi Language, a large quantity of Punjabi data is required. So before starting to build an acoustic model for spontaneous Punjabi speech, we need to prepare the Punjabi speech corpus.

1.1. Feature Extraction Techniques

In automatic spontaneous Punjabi speech recognition system, it is general to extort a set of features from dialogue signal. Categorization is accepted out on the set of features as an alternative of the language indicators themselves. The feature extraction phase seeks out to offer a dense demonstration of the speech waveform. Feature extraction, by MFCC, is the well-known method used for feature extraction. CMUSphinx utilizes Mel-cepstrum MFCC features through sound following and spectral subtraction for noise lessening. There is a variety of MFCC [4, 5] which is different by numeral factors, but not actually dissimilar for correctness. The selection of the numeral of MFCCs to comprise in automatic speech recognition system is widely experimental. In practice, the best possible amount of coefficients depends on the number of the training corpus, the particulars of the training method, the numeral of Gaussian mixtures in the HMMs, the speaker and surroundings disturbance distinctiveness, and every so often accessible computing assets.

1.2. Building a Spontaneous Speech Modeling for Punjabi language

The objective of the research is to fully develop the Punjabi speech database for Spontaneous Punjabi speech recognition system. In order to perform these objectives, the work distributed into several phases.

1.2.1. Training Requirements for Spontaneous Punjabi Speech

Speech identification [1] performs on the basis of anyone speech reveals distinctiveness are exclusive to a dissimilar narrator. The signal through training data and testing data session can be deeply dissimilar due to many factors such as person accent vary with time, due to the health stipulation (e.g. the speaker has a cough), different rate of speaking and also acoustically and linguistically disturbance and variation recording surroundings via microphone or telephone.

Table 1
Punjabi Speech Training Requirement process

Sr. No	Process	Description
1.	Type of Speaker	25 Male Voice 5 Female Voice
2.	Recording Tools	Microphone Smart Voice Recorder Wave Recoding Software
3.	Sampling rate of the audio	16 kHz
4.	Bit Rate(number of bits per second)	16
5.	Channel	Mono
6.	Feature Computational	39 double delta MFC coefficient

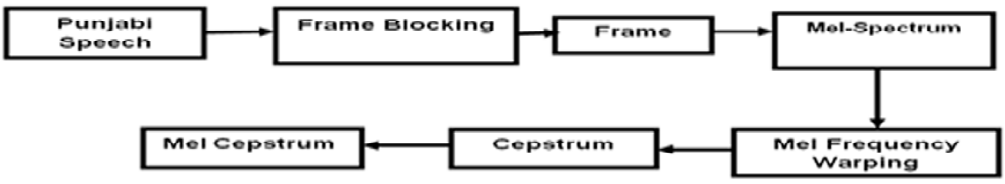


Figure 2: Feature extraction Scheme [12]

The list of configurations files is required to build for building an Acoustic model [8] for Spontaneous Punjabi speech. First, we need to train all the below-mentioned files for the training and decoding purpose. For building acoustic model for the Punjabi language, the designing of etc folder and sphinx configuration files, and for this purpose, we need to create following files

- PunjabiSpeech_db.dictionary – for Punjabi language Phonetic lexicon
- PunjabiSpeech_database.phone – Phone set file (individual sound units)
- PunjabiSpeech_database.lm.DMP - Language model for Decoding purpose
- PunjabiSpeech_database.filler - List of fillers (for rejected sound and silence gap)
- PunjabiSpeech_database_training.fileids - List of training files (path of wav file)
- PunjabiSpeech_database_training.transcription - Transcription for training (text conversion of the wav file)
- PunjabiSpeech_database_testing.fileids - List of files for testing (wav File)
- PunjabiSpeech_database_testing.transcription - Transcription for testing

— **Wav for Recording**

Male_Speaker_1.wav

Test_1.wav - Recording of Punjabi speech sound

Female_Speaker_2.wav

Test_2.wav - Recording of Punjabi speech sound

Speaker_n.wav

Test_n.wav - Recording of Punjabi speech sound

1.2.2. Structure of Punjabi Phone File

Sounds of lexis in a speech [13] are typically collected at a place of resonances, i.e. phones, which might be measured as sub-word units. The Punjabi phone file depicts each individual resonance unit which is required in the Punjabi dictionary for creating of words and sentences for the purpose of speech.

Table 2
Structure of the Phone

ਸ	ਡ	ਤ	ਰ	ਨ	ਹ	ਲ	ਕ	ਪ	ਜ
ਦ	ਠ	ਅੰ	ਆ	ਪਿੰ	ਕ੍	ਮੈ	ਹਾ	ਓ	ਅ
ਬ	ਮ	ਨਾ	ਯ	ਗ	ਸ	ਸਿੱ	ਖਿ	ਕੰ	ਪਿ
ਊ	ਇੰ	ਨਿੰ	ਐ	ਈ	ਵਿ	ਚ	ਟ	ਕਿੰ	ਪ੍
ਸੀਂ	ਖ	ਉ	ਘ	ਰੁ.	ਨੂੰ	ਥ	ਤੂੰ	ਦਿ	ਚਿ
ਰਾਂ	ਢ	ਧੰ	ਸਿ	ਫ	ਧ	ਣ	ਫ਼	ਹੀਂ	ਤੀਂ
ਕੈ	ਜ਼	ਮਾਂ	ਉਂ	ਨਿ	ਲਿ	ਕਿ	ਜਿ	ਇ	ਤਿ
ਇੱ	ਹਿੱ	ਚਿ	ਮਿ	ਭ	ਹਿ	ਲੈਂ	ੁ	ੈ	ਾ
ਂ	ੈ	ੀ	ੈ	.	ੰ	ੱ	ੂ	।	

1.2.3. Structure of Dictionary file for Punjabi Corpus

A Punjabi language [15] dictionary file, in which every word plans to a series of audio units to obtain the series of audio units connected with every signal. The Punjabi dictionary files always contain a distinctive entry and arranged consequently. The Punjabi dictionary offers articulations for lexis originate in the Language Model. The articulations shatter words into series of sub-word components found in the Punjabi acoustic model. Sphinx-4 presently offers developments of the dictionary interface to maintain the CMU speaking dictionary [5]. The Punjabi words dictionary looks like as:

Table 3
Structure of the Dictionary File

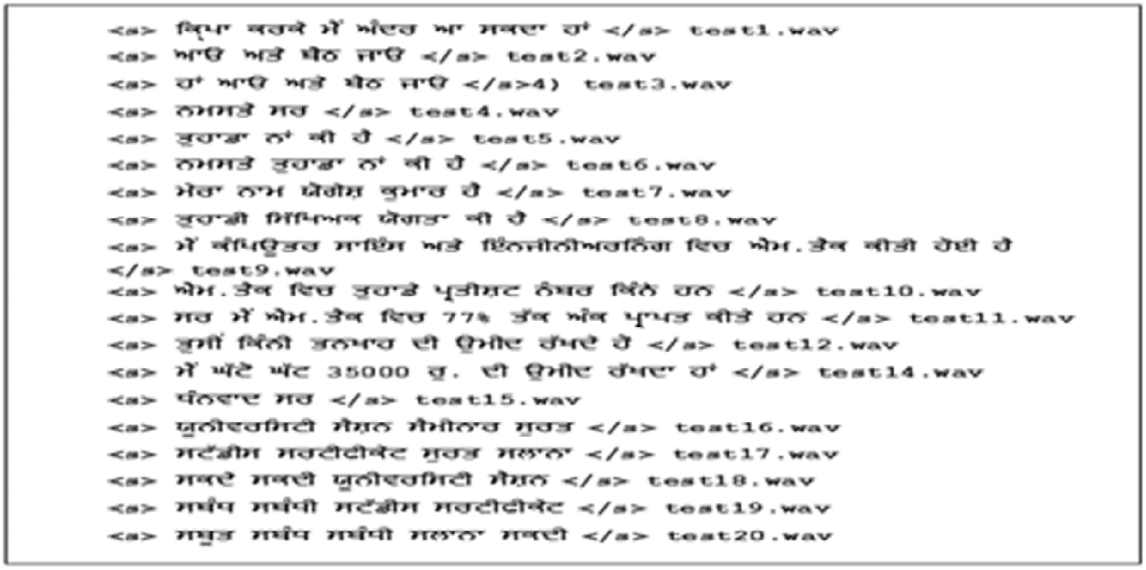
ਕਿਪਾ	ਕਰਕੇ	ਮੈਂ	ਅੰਦਰ	ਆ	ਸਕਦਾ	ਹਾਂ
ਸਰ	ਹਾਂ	ਆਓ	ਅੰਤ	ਬੋਣ	ਜਾਓ	।
ਨਮਸਤੇ	ਸਰ	ਨਮਸਤੇ	ਤੁਹਾਡਾ	ਨਾਂ	ਕੀ	ਹੈ?
ਮੇਰਾ	ਨਾਮ	ਯੋਗੇਸ਼	ਕੁਮਾਰ	ਹੈ।	ਤੁਹਾਡੀ	ਸਿੱਖਿਅਕ
ਯੋਗਤਾ	ਕੀ	ਹੈ	ਮੈਂ	ਕੰਪਿਊਟਰ	ਸਾਇੰਸ	ਅੰਤ
ਇੰਜੀਨੀਅਰਿੰਗ	ਵਿਚ	ਐਮ	.ਟੇਕ	ਕੀਤੀ	ਹੋਈ	ਹੈ
ਤੁਸੀਂ	ਕਿੰਨੀ	ਤਨਖਾਹ	ਦੀ	ਉਮੀਦ	ਰੱਖਦੇ	ਹੋ
ਮੈਂ	ਘੱਟੋ	ਘੱਟ	35000	ਤੋ	ਤਨਖਾਹ	ਦੀ
ਉਮੀਦ	ਰੱਖਦਾ	ਹਾਂ	ਅਸੀਂ	ਤੁਹਾਨੂੰ	ਥੋੜ੍ਹੇ	ਦਿਨਾਂ
ਵਿਚ	ਸੂਚਿਤ	ਕਰਾਂਗੇ	ਧੰਨਵਾਦ	ਸਰ	ਨਮਸਤੇ	ਯੂਨੀਵਰਸਿਟੀ
ਸੈਸ਼ਨ	ਸੈਮੀਨਾਰ	ਸੁਰਤ	ਸਟੱਡੀਸ	ਸਰਟੀਫੀਕੇਟ	ਸਲਾਨਾ	ਸਬਦੇ
ਸਬਦੀ	ਸਬੰਧ	ਸਬੰਧੀ	ਸਬੂਤ	ਸਤੰਬਰ	ਸਮਾਣ	ਸੋਮਵਾਰ
ਫ਼ੀਸ	ਨੇ	ਲੇਟ	ਲਈ	ਨਹੀਂ	ਨਾ	ਨਾਲ
ਨਵੰਬਰ	ਲੇਡੀਦੇ	ਨੰਬੀ	ਨੰਬਰ	ਫਬੈਲਟੀ	ਫਰਵਰੀ	ਦੇ
ਦਸਤਾਵੇਜ਼	ਦੀ	ਦਾ	ਦੇਰਨ	ਕੇ	ਕੈਂਸਲ	ਕੀਤੀ
ਕੇਟੀ	ਜਾਪੈ	ਜਰ	ਜਰਨ	ਜਰਨੇ	ਜਰਨਾ	ਜਰਕੇ
ਜਰਵਾਉਣ	ਨਿਯਮਾਂ	ਨਿਬੰਧ	ਨਿਮਨਲਿਖਤ	ਲਿਖ	ਲਿਖੇ	ਨਿੰਬਾਰਨ
ਦਿਸੰਬਰ	ਦਿਤਾ	ਜਿਤੇ	ਸਿਸ	ਇਸ	ਇਹ	ਇਮਤਿਹਨ
ਇੱਕ	ਚਿੱਤ	ਰਿਸਕਤ	ਰਿਪੋਰਟ	ਵਿਖੇ	ਮਿਤੀ	ਭੇਜ
ਭੇਜੇ	ਪੋਸਟ	ਪੈਰੇ	ਭੁਜਦੇ	ਪੋਫੈਸਰ	ਪਹਿਲੇ	ਪਹਿਲਾਂ
ਪੀ.ਐਚ.ਡੀ.	ਪਾਸ	ਭਾਲਣ	ਪ੍ਰੀਖਿਆ	ਪ੍ਰਗਤੀ	ਪ੍ਰਧਾਨ	ਪ੍ਰਧਾਨਗੀ
ਪੰਜਾਬੀ	ਪੰਤਰ	ਸੇ	ਸੋਕਰ	ਚੁਲਾਈ	ਜੂਨ	ਸਨਬਰੰ
ਸਾਣ	ਸਾਣਕਾਰੀ	ਸਾਰੀ	ਸਾਵੇ	ਸਾਵੇਗੀ		

1.2.4. Filler Dictionary

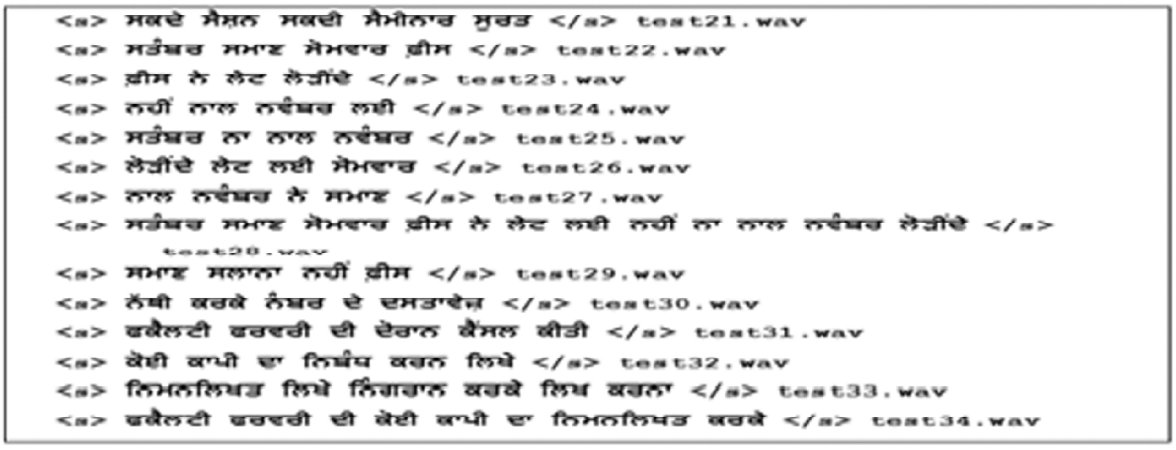
Filler dictionary (Punjabi Speech_database. filler) includes filler phones [14] (not covered by linguistic model non-linguistic sounds like breathe, hmm or laugh, Aaa). It can enclose presently silences.

1.2.5. Transcript and Fields File

A transcript [14] is required to signify what the presenters are saying in the acoustic file. So in a case, the conversation of the speaker renowned accurately the equal specific mode it has been traced, with silence tag (starting tag <s>, ending tag </s>), go after by the file id which signifies the speech. The transcription file is the text conversion of the wave file and Filed file is the path of the same wave file. The followings are the transcription and field file of the research work.



Figures 3: Transcription file-I



Figures 4: Transcription file-II

2. DECODING FOR THE SPONTANEOUS PUNJABI SPEECH MODEL

The language model is used for the decoding purpose. The language model [3] is usually a *N*-gram model in which the possibility of every word is trained only on its $N - 1$ predecessors. Language model allocates a possibility to a series of m words by means of a possibility allocation. To model it we can use a regular grammar. The language model, lmp extension file, depicts the possibility, likelihood in use when a series or compilation of words is observed [4]. This in order is used to restrain explore and as an effect considerably get better detection correctness. The language model file is simple text.

[Step 1] Input the Punjabi words data sets

[Step 2] To run the cmu command instruction and create the Punjabi vocab File

[Step 3] Finally language model will be used for decoding the given Punjabi dataset for recognition

Figure 5, completely demonstrate that once compilation the spontaneous Punjabi speech System, the recognition accuracy for words are 85.39% and sentences are 91.17% of the given Punjabi dataset. Out of 691 sentences, 61 sentences and out of 2115 words, 309 words are failed while recognition. So sentences error rate are 8.8% and words error rate are 14.6%.

```
MODULE: DECODE Decoding using models previously trained (2016-06-26 21:10)

        Decoding 691 segments starting at 0 (part 1 of 1)
        pocketsphinx_batch Log File

Aligning results to find error rate

        SENTENCE ERROR: 8.8% (61/691) WORD ERROR RATE: 14.6% (309/2115)
```

Figure 5: Results of Punjabi corpus decoding

3. GUI FOR SPONTANEOUS PUNJABI LIVE SPEECH MODEL

Finally, the GUI of spontaneous Punjabi speech model has been constructed using Java programming. The spontaneous Punjabi speech model has two distinct options for identification of Punjabi speech. At first, we have live speech test option and another we can use browse option to explore the path of the wave file.

Figure 6: GUI for spontaneous Punjabi speech model

Figure 7: Output of the Punjabi live speech model

Figure 8: Output of the speech model through browse button

4. PERFORMANCE MEASUREMENTS

The recognition accuracy [7] and the speed are the criteria for evaluating the optimality of the Punjabi live speech model. Recognition Accuracy is calculated by how many words and sentences are failed during decoding. And performance can be degraded by certain parameters such as:

a) Different speaking style

b) Different accents & Dialect

c) Poor speaker condition

d) Pronunciation issue

e) Noise while recording speech

On the basis of the Punjabi corpus, table 4 shows the performance analysis of the automatic Spontaneous Punjabi Speech model. Here we can see that the performance analysis on the basis of the recognition accuracy of the correct Punjabi sentences and words.

Table 4
Result Analysis

Spontaneous Punjabi Speech	Total	Correct Words/Sentences	Error	Recognition Accuracy (%)
Punjabi Sentences	691	630	61	91.17
Punjabi Words	2115	1806	309	85.38

5. CONCLUSION AND FUTURE WORK

The primary aim of the research work is to build a speaker independent automatic spontaneous speech recognition system for Punjabi language corpus. So the ultimate objective was to focus on the actual research work done in the area of spontaneous speech recognition system. For the implementation of the research paper, the sphinx framework and java programming language have been used. The purpose of the research work is to improve the recognition accuracy and to reduce the sentence and word error rate. The Punjabi speech model is trained and tested by different male and female speakers. The GUI also has been created for Spontaneous Punjabi live speech model for recognizing the speech at runtime environment. In the future, the Punjabi speech model will be trained by more speakers and more Punjabi vocabulary will be added. The focus will be shifting towards, for improve the recognition rate of the Punjabi speech and faster decoding of the speech.

REFERENCES

[1] Arun Narayanan, DeLiang Wang (2015). Improving Robustness of Deep Neural Network Acoustic Models via Speech Separation and Joint Adaptive Training. IEEE/ACM Transactions On Audio, Speech, And Language Processing, VOL. 23, NO. 1.

[2] Chandan Mittal, Vishal Goyal & Umrinderpal Singh(2015). HMM Chunker for Punjabi. Indian Journal of Science and Technology, *Vol 8(35).*

[3] AnchalKatyal, Amanpreet Kaur & Jasmeen Gill(2014). Punjabi Speech Recognition of Isolated Words Using Compound EEMD & Neural Network. International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-4, Issue-1.

[4] *Kun Han, Dong Yu, Ivan Tashev(2014).* Speech Emotion Recognition Using Deep Neural Network and Extreme Learning Machine. Interspeech , pp. 223-227.

[5] Jeet Kumar, Om Prakash Prabhakar and Navneet Kumar Sahu(2014). Comparative Analysis of Different Feature Extraction and Classifier Techniques for Speaker Identification Systems: A Review. International Journal of Innovative Research in Computer and Communication Engineering, Vol. 2, Issue 1, pp. 2760-2269.

[6] Wiqas Ghai, Navdeep Singh(2013). Continuous Speech Recognition for Punjabi Language. International Journal of Computer Applications (0975–8887) Volume 72– No. 14.

[7] Vivek Sharma, Meenakshi Sharma(2013). A Quantitative Study Of The Automatic Speech Recognition Technique. International Journal of Advances in Science and Technology (IJAST), Vol I, Issue I.

[8] Nermine Ahmed Hendy, Hania Farag(2013). Emotion Recognition Using Neural Network: A Comparative Study. International Scholarly and Scientific Research & Innovation 7(3), Vol 7.

[9] Mr. Kashyap Patel, Dr. R.K. Prasad(2013). Speech Recognition and Verification Using MFCC & VQ. International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 5.

[10] Anupriya Sharma, Amanpreet Kaur (2013). A Survey on Punjabi Speech Segmentation into Syllable-Like Units Using Group Delay. International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 6.

[11] Karen Livescu, Eric Fosler-Lussier & Florian Metze (2012). Sub-word modeling for automatic speech recognition. Signal Processing Magazine, Vol. Xx, No. Yy, Zzzz.

[12] Lindasalwa Muda, Mumtaj Begam and I. Elamvazuthi(2010). Voice Recognition Algorithms using Mel-Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques. Journalof Computing, Volume 2, Issue 3, ISSN 2151-9617.

[13] www.shabdkosh.com/pa/.../corpus/corpus-meaning-in-Punjabi-English.

[14] http://cmusphinx.sourceforge.net/

[15] http://www.scapepunjab.com/