

Available online at <http://www.mecspress.net/ijeme>

Automatic Spontaneous Speech Recognition for Punjabi Language Interview Speech Corpus

Yogesh Kumar ^a, Dr. Navdeep Singh ^b

^a *Research Scholar, Punjabi University, Patiala and 147001, India*

^b *Associate Professor, Mata Gujri College, Fatehgarh Sahib and 140406, India*

Abstract

Automatic Speech Recognition presents natural phenomena for the communication among man and machine. The purpose of Speech Recognition speech system is to convert the sequence of sound units in the form of text description. The main objective of the research work is to develop the automatic spontaneous speech model for the Punjabi language. Punjabi is categorized as a constituent of the Indo-Aryan subgroup of the Indo-European family of languages. So far no work has been done in the field of spontaneous Punjabi speech recognition system. In spontaneous speech system, the sounds are usually unprompted and non- designed and are commonly described by repetitions, repairs, false start, partial words and non-planned words, silence gap etc. In this paper, the focus is on the development of the spontaneous speech model for the recognition of the Punjabi language. The GUI for Punjabi speech model also has been created and tested for the live Punjabi interview speech corpus. The recognition accuracy is 98.6% for Punjabi sentences and 98.8% for Punjabi words. The sphinx toolkit and java programming are used to build a spontaneous speech model for Punjabi live speech.

Index Terms: Sphinx, Acoustic model, Feature vector, Transcription, Field file, Dictionary file, Phone file.

© 2016 Published by MECS Publisher. Selection and/or peer review under responsibility of the Research Association of Modern Education and Computer Science.

1. Introduction

Automatic Speech Recognition offers a medium used for natural communication among man and machine. A Spontaneous speech recognition scheme for Punjabi language consists of a microphone for the human being to articulate into, vocalizations identification method, a computer to obtain and understand the speech, a superior quality soundcard for input and output and an accurate and fine articulation. Spontaneous speech, still, is extremely changeable and hardly ever corroborates to proficiently described articulation policy. Consequently, recognition of spontaneous speech [1] will require a recording every spoken words. Since users' sounds are usually natural and non-planned, they are usually illustrated by repetition, artificial initiate, partial

* Corresponding author.
E-mail address:

words, discontinue in the core and re-starts, or additional linguistic occurrence such as cough etc.. The speech detection [2] section should be capable to extort away of the verbal communication indication, a speech progression permitting the semantic analyzer to infer the significance of the user's speech. The spontaneous speech recognition is mainly complicated due to representation divergence, for example when a person talks to another person as in conventions or discussions, interview or over telephone conversations.

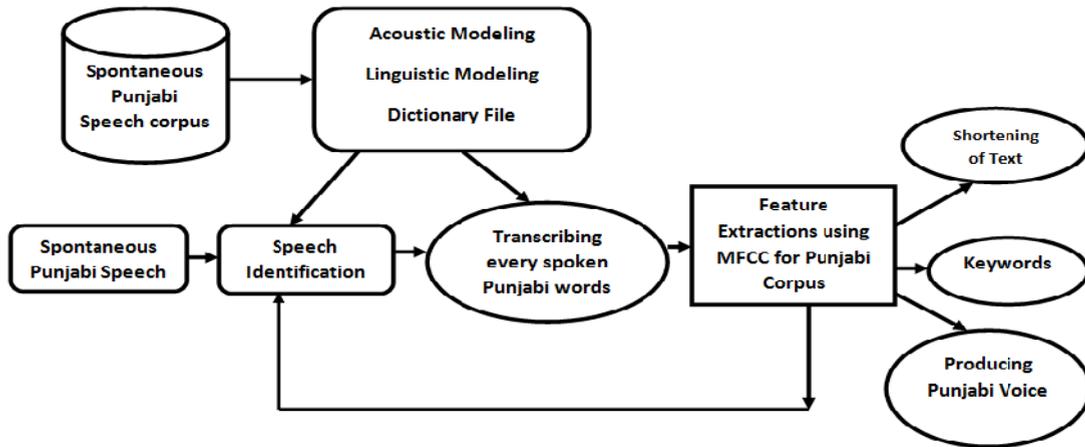


Fig.1. Spontaneous Speech Processing For Punjabi Language

1.1. Acoustic Modeling

An acoustic model is an organizer that encloses arithmetical demonstrations of every discrete resonances that build up an utterance. Each of these arithmetical demonstrations is allocated a tag described as a phoneme. Sounds of lexis in a speech are usually collected from a set of sounds, i.e. phones, which might be measured as sub-word units. A sound model is produced by captivating a huge corpus of words and using particular training algorithms to generate arithmetical depictions for every phoneme in a verbal communication.

1.2. Phonetic list

Punjabi is a tonal speech and its phoneme list (5, 15) contains 10 vowels, 25 consonants, 7 diphthongs and three tenor whose construction has neither friction nor strike of a manner in the entrance.

1.3. Linguistic models

A language model (LM) is a group of past information concerning a speech. This information is autonomous of a speech to be accepted. It consequently signifies prior knowledge about speech and the prospects at sounds. Information regarding a speech can be uttered in terms of which terms or utterance sequences are probable or how commonly they arise.

1.4. Feature extraction mechanism for Spontaneous Punjabi Speech

The speech feature extraction [3, 6] in a classification is concerned with dropping the dimensionality of the input vector while sustaining the discriminating control of the signal. Feature extraction (Fig. 2) needs more

consideration as detection performance which depends greatly on the feature extraction stage. LPC, MFCC, AMFCC, RAS, DAS, MFCC, Higher lag autocorrelation coefficients, PLP, MF-PLP, BFCC, RPLP are the poles apart methods for feature extraction.

But extracted feature must assemble various criteria whereas selling by way of the speech indication such as:

- Effortless to compute extort speech features
- It must not be disposed to mimicry
- It must explain slight variation from one speaking situation to another
- It must be steady above time
- It must happen recurrently and obviously in speech

1.5. Mel Frequency Cepstral Coefficient (MFCC)

Mel-frequency cepstral coefficients (MFCCs) [7, 11] are coefficients that commonly create up and concerning an MFC. The distinction among the cepstrum and the Mel-frequency cepstrum is that in MFC the frequency bands are evenly spaced on the Mel scale, which reasonably exacts the human auditory system's response. MFCCs are normally used as features in speech recognition system. Sphinx-4 is usually configured with a Front End that creates Mel-Frequency Cepstral. The idea of the FrontEnd is to parameterize a contribution indication (e.g., audio) into a series of productivity Features.

2. Steps to Design, Development and Adapting the Punjabi Speech corpus for spontaneous Speech Recognition

The main aim of the research is to develop the Punjabi speech corpus for Spontaneous Punjabi speech. In regulate to attain these objectives the effort separated into various parts.

2.1. Design the Punjabi Speech Corpus

- For Spontaneous Punjabi Speech, Punjabi text corpus has to be invented that can be interpreted and verified.
- For Spontaneous Punjabi Speech interviews have to be planned to confine spontaneous conversation supported dialogue
- Sphinx toolkit and Standard Programming language are used to developed and evaluated the Punjabi Speech Model

2.2. Development of the Punjabi Speech Corpus

- The Punjabi Speech corpus [4,5] has to be interpreted and captured in controlled environmental stipulation
- The Punjabi spontaneous speech corpus has to be produced from interviews and question and answer sessions
- Analyze the Punjabi speech corpus into different frequency
- The utterances of the spontaneous Speech corpus have to be transcribed in Punjabi and also using phonemic transcription

2.3. Adaptation of the Punjabi Speech corpus to the automatic spontaneous speech recognition system

- The first step for adaption [4,5] involves dealing out the dialogue corpora to match to the input levels of the sphinx speech detection scheme
- It involves the improvement of the compiler like tool which creates a recorded Punjabi database as input and generates all the useful information needed by the sphinx Speech Recognition scheme as output

3. Steps to Build-up the Spontaneous Punjabi Speech Acoustic Model

To construct an acoustic model for spontaneous Punjabi corpus, it is mandatory to prepare the system with various Punjabi word level training. But the single utterance wav file is small in dimension and silence gap is supplementary. Consequently even for preparing one word, we need to require multiple sentences. For this reason, we qualified the Punjabi spontaneous speech method with numerous words and sentences of Punjabi Speech with variable silence gap.

For training the Punjabi acoustic model following configurations files are required:

3.1. Dictionary Files

The *Dictionary* [12, 13] offers articulations for Punjabi words initiate in the Language Model. The pronunciations break words into sequences of sub-word units found in the Acoustic Model. The Dictionary File creates the Punjabi word to articulate dictation mappings. The Dictionary file will look like as:

Table 1. Dictionary File [14] for Punjabi Corpus

Punjabi Words/Sentences	Dictionary Entry of Punjabi Words	Punjabi Words/Sentences	Dictionary Entry of Punjabi Words
ਮੇਰਾ	ਮ ੇ ਰ ਾ	ਕੀਤੇ	ਕ ੀ ਤ ੇ
ਨਾਮ	ਨ ਾ ਮ	ਤੱਕ	ਤ ੱ ਕ
ਯੋਗੇਸ਼	ਯ ੋ ਗ ੇ ਸ਼	ਤੁਸੀਂ	ਤ ੁ ਸੀ
ਕੁਮਾਰ	ਕ ੁ ਮ ਾ ਰ	ਕਿੰਨੀ	ਕਿੰ ਨ ੀ
ਤੁਹਾਡੀ	ਤ ੁ ਹ ਾ ਡ ੀ	ਤਨਖਾਹ	ਤ ਨ ਖ ਾ ਹ
ਸਿੱਖਿਅਕ	ਸਿੱ ਖਿ ਅ ਕ	ਦੀ	ਦ ੀ
ਯੋਗਤਾ	ਯ ੋ ਗ ਤ ਾ	ਉਮੀਦ	ਉ ਮ ੀ ਦ
ਕੰਪਿਊਟਰ	ਕੰ ਪਿ ਊ ਤ ਰ	ਰੱਖਦੇ	ਰ ੱ ਖ ਦ ੇ
ਸਾਇੰਸ	ਸ ਾ ਇੰ ਸ	ਹੈ	ਹ ੈ
ਇਨਜੀਨੀਅਰਨਿੰਗ	ਇ ਨ ਜ ੀ ਨ ੀ ਅ ਰ ਨਿੰ ਗ	ਘੱਟੇ	ਘ ੱ ਟ ੇ
ਵਿਚ	ਵਿ ਚ	ਘੱਟ	ਘ ੱ ਟ
ਐਮ. ਤੈਕ	ਐ ਮ . ਤ ੈ ਕ	35000	35000
ਕੀਤੀ	ਕ ੀ ਤ ੀ	ਰੁ.	ਰੁ.
ਹੋਈ	ਹ ੋ ਈ	ਰੱਖਦਾ	ਰ ੱ ਖ ਦ ਾ
ਪ੍ਰਤੀਸ਼ਟ	ਪ੍ਰ ਤ ੀ ਸ਼ ਟ	ਅਸੀਂ	ਅ ਸੀ
ਨੰਬਰ	ਨ ੰ ਬ ਰ	ਤੁਹਾਨੂੰ	ਤ ੁ ਹ ਾ ਨੂੰ
ਕਿੰਨੇ	ਕਿੰ ਨ ੇ	ਥੋੜ੍ਹੇ	ਥ ੋ ਝੁ
ਹਨ	ਹ ਨ	ਦਿਨਾਂ	ਦਿ ਨ ਾਂ
77%	77%	ਸੂਚਿਤ	ਸ ੂ ਚਿ ਤ
ਅੱਕ	ਅ ੱ ਕ	ਕਰਾਂਗੇ	ਕ ਰ ਾਂ ਗ ੇ
ਪ੍ਰਾਪਤ	ਪ੍ਰ ਾ ਪ ਤ	ਧੰਨਵਾਦ	ਧੰ ਨ ਵ ਾ ਦ

3.2. Phone File:

The Punjabi phone [12] file describes every individual sound unit which is used in the Punjabi dictionary for making of words and sentences. The phone file is the most difficult part in the acoustic modeling because it requires depth knowledge of the Punjabi language. The structure of Punjabi phone file is shown in the Table 2.

Table 2. Phone Files [12] of Punjabi Language

ਸ	ਙ	ਤ	ਰ	ਨ	ਹ	ਲ	ਕ
ਪ	ਜ	ਦ	ਠ	ਅੰ	ਆ	ਪਿੰ	ਕਿ
ਮੈ	ਹਾ	ਓ	ਅ	ਬ	ਮ	ਨਾ	ਯ
ਗ	ਸ਼	ਸਿੰ	ਖਿ	ਕੰ	ਪਿ	ਊ	ਇੰ
ਨਿੰ	ਐ	ਈ	ਵਿ	ਚ	ਟ	ਕਿੰ	ਪ੍ਰ
ਸੀਂ	ਖ	ਉ	ਘ	ਰੁ.	ਨੂੰ	ਥ	ਝੇ
ਦਿ	ਚਿ	ਰਾਂ	ਵ	ਧੰ	ੜ	ੇ	ਾਂ
ਂ	ੈ	ੀ	ੇ	ੰ	ੌ	ੂ	!

3.3. Transcription (path of wav files) and Fields (conversation of Wav File):

A Punjabi transcript [14, 15] file is required to signify what exactly the presenters are saying in the acoustic file. So in a file the interview of the speaker renowned accurately the similar defined way it has been confirmed, with silence tags (beginning with <s>, ending tag </s>). Basically, there are two primary types of transcription file and field file. First one is used to training purpose and second one is used for testing purpose.

- For training purpose (Punjabi.train.transcript and Punjabi.train.field)
- For testing purpose (Punjabi.test.transcript and Punjabi.test.field)

Training files are basically required to generate the feature vector which later on will be used for Punjabi live speech recognition. And testing files are required in language model for decoding purpose to check the recognition performance of the Punjabi live speech. For example, in our Punjabi database, figure 2 shows the transcription file along with field file.

```

<s> ਕਿਪਾ ਕਰਕੇ ਮੈਂ ਅੰਦਰ ਆ ਸਕਦਾ ਹਾਂ </s> (test1.wav)
<s> ਮੈਂ ਅੰਦਰ ਆ ਸਕਦਾ ਹਾਂ </s> (test2.wav)
<s> ਆਓ ਅਤੇ ਬੈਠ ਜਾਓ </s> (test3.wav)
<s> ਆਓ ਅਤੇ ਬੈਠ ਜਾਓ </s> (test4.wav)
<s> ਹਾਂ ਆਓ ਅਤੇ ਬੈਠ ਜਾਓ </s> (test5.wav)
<s> ਨਮਸਤੇ ਸਰ </s> (test6.wav)
<s> ਤੁਹਾਡਾ ਨਾਂ ਕੀ ਹੈ </s> (test7.wav)
<s> ਨਮਸਤੇ ਤੁਹਾਡਾ ਨਾਂ ਕੀ ਹੈ </s> (test8.wav)
<s> ਮੇਰਾ ਨਾਮ ਯੋਗੇਸ਼ ਕੁਮਾਰ ਹੈ </s> (test9.wav)
<s> ਤੁਹਾਡੀ ਸਿੱਖਿਅਕ ਯੋਗਤਾ ਕੀ ਹੈ </s> (test10.wav)
<s> ਮੈਂ ਕੰਪਿਊਟਰ ਸਾਇੰਸ ਅਤੇ ਇੰਜੀਨੀਅਰਿੰਗ ਵਿਚ ਐਮ.ਏਕ ਕੀਤੀ ਹੋਈ ਹੈ </s> (test11.wav)
<s> ਐਮ.ਏਕ ਵਿਚ ਤੁਹਾਡੇ ਪ੍ਰਤੀਸ਼ਟ ਨੰਬਰ ਕਿੰਨੇ ਹਨ </s> (test12.wav)
<s> ਤੁਹਾਡੇ ਪ੍ਰਤੀਸ਼ਟ ਨੰਬਰ ਕਿੰਨੇ ਹਨ </s> (test12.wav)
<s> ਐਮ.ਏਕ ਵਿਚ ਤੁਹਾਡੇ ਪ੍ਰਤੀਸ਼ਟ ਨੰਬਰ ਕਿੰਨੇ ਹਨ </s> (test13.wav)
<s> ਸਰ ਮੈਂ ਐਮ.ਏਕ ਵਿਚ 77% ਤੱਕ ਅੰਕ ਪ੍ਰਾਪਤ ਕੀਤੇ ਹਨ </s> (test14.wav)
<s> ਤੁਸੀਂ ਕਿੰਨੀ ਤਨਖਾਹ ਦੀ ਉਮੀਦ ਰੱਖਦੇ ਹੋ </s> (test15.wav)
<s> ਮੈਂ ਘੱਟੋ ਘੱਟ 35000 ਰੁ. ਦੀ ਉਮੀਦ ਰੱਖਦਾ ਹਾਂ </s> (test16.wav)
<s> ਅਸੀਂ ਤੁਹਾਨੂੰ ਥੋੜ੍ਹੇ ਦਿਨਾਂ ਵਿਚ ਸੂਚਿਤ ਕਰਾਂਗੇ </s> (test17.wav)
<s> ਧੰਨਵਾਦ ਸਰ </s> (test18.wav)

```

Fig.2. Transcript File for Punjabi Corpus

4. Training Steps to build Language Model for Punjabi Dataset

The Language Model [10] section of the Linguist presents word-level verbal communication construction, which can be characterized by any amount of pluggable completions. The primary purpose of language model is to decoding. A language model is to map the series of acoustic units on behalf of articulation. Steps to follow to create the Language model for Punjabi Language model are:

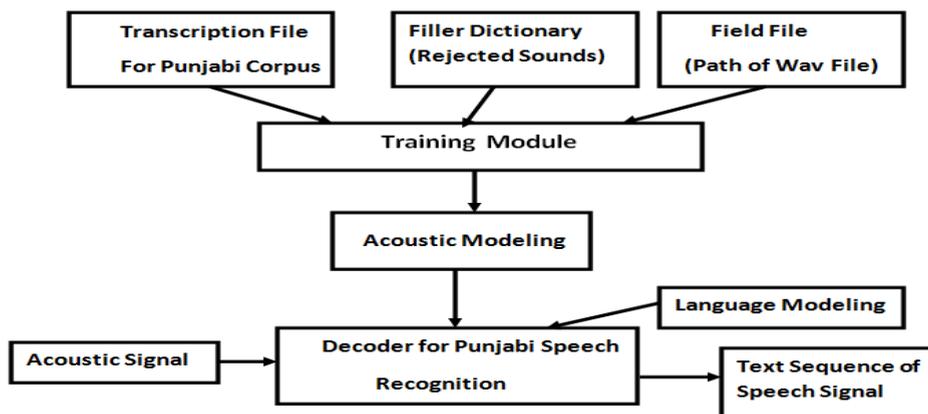


Fig.3. Language Modeling [8] for Spontaneous Punjabi Speech

Step I: First to input interview dataset of Punjabi sentences.

Step II: To execute the cmu command and creating the vocab file for Punjabi dataset.

Step III: At last language model will be created with extension lm.dmp, which is basically used for training purpose. Fig. 3 shows the working of language model clearly.

MODULE: DECODE Decoding using models previously trained (2016-01-03 21:14)

```
Decoding 461 segments starting at 0 (part 1 of 1)
pocketsphinx_batch Log File
```

Aligning results to find error rate

```
SENTENCE ERROR: 1.3% (6/461) WORD ERROR RATE: 1.2% (14/1227)
```

Fig.4. Output after Decoding of Punjabi Interview Dataset

Fig.4, perfectly illustrates that after compiling the Punjabi spontaneous speech modeling, the word error rate and sentence error rate of the given Punjabi dataset. Out of 461 lines, 6 lines and out of 1227 words, 14 words are failed during recognition. So sentence error rate is 1.3% and word error rate is 1.2%.

5. Graphical User Interface for Live Speech Recognition Model of Punjabi Language

Finally, the language model and training set of Punjabi corpus are used to make the final GUI for spontaneous speech model. The GUI of Spontaneous Punjabi speech has two different methods for recognition of Punjabi speech. At first method, we can use live speech test option and secondly we have wav file test method, in which we can browse the wav file for speech and then recognize it.

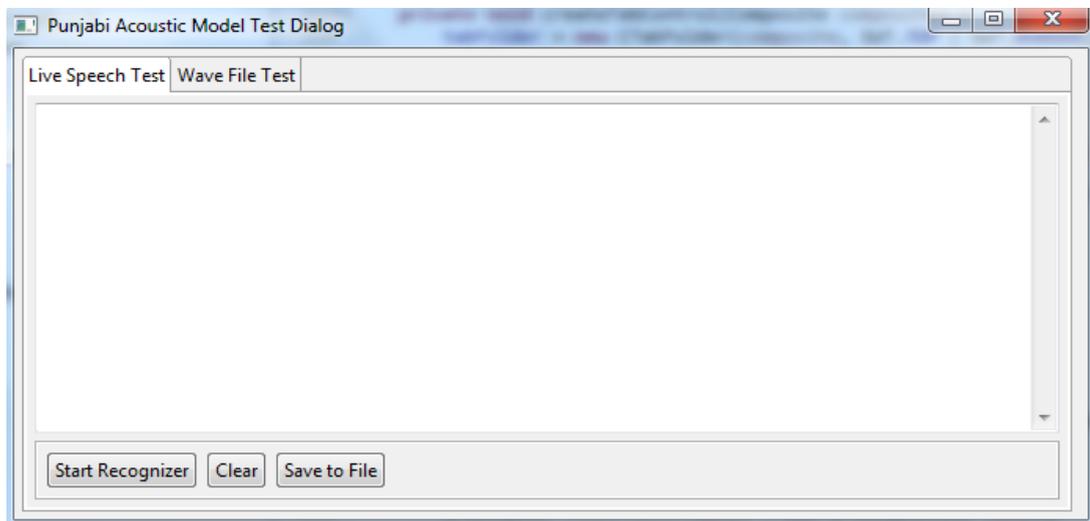


Fig.5. First Screen for the GUI of Punjabi Speech Model

Above fig. 5, clearly shows that starting screen of the Punjabi speech model. We also have the option to save the live speech in the text file. So that, we can use it further.

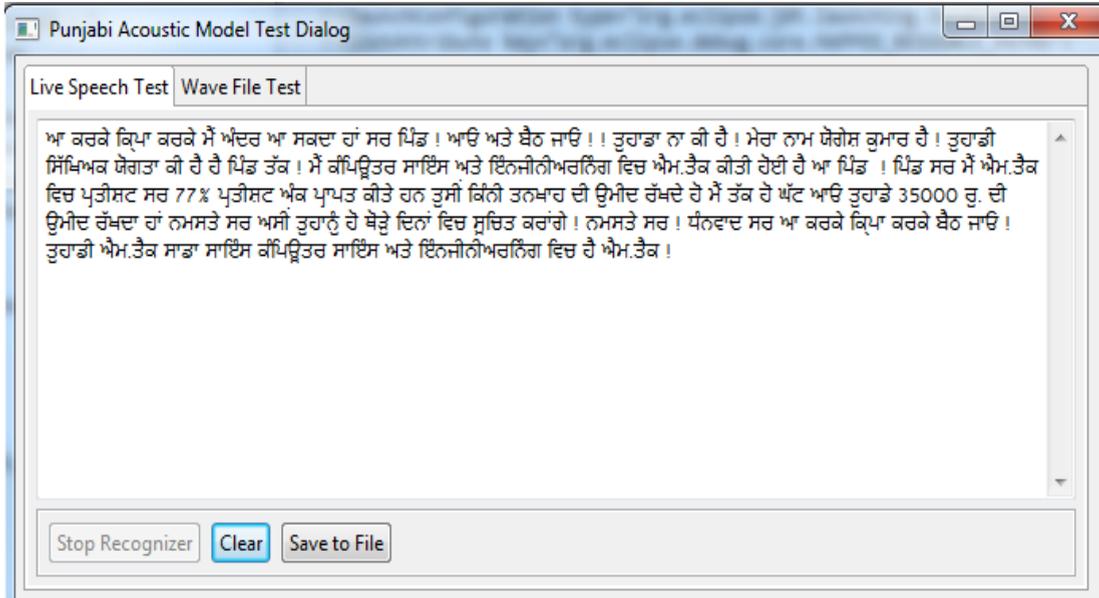


Fig.6. Output of the Spontaneous Punjabi Speech Model

Fig. 6, clearly depict that the output of the spontaneous Punjabi live speech model.

6. Performance Analysis

The performance analysis of the research work is considered while decoding of the Punjabi interview speech corpus. Table 3 clearly depicts the recognition accuracy rate of the Punjabi interview speech corpus. Presently, we have recognized the 461 sentences and 1227 Punjabi words. Out of 461 sentences, 6 sentences and out of 1227 Punjabi words, 14 words are failed while recognition.

Table 3. Result Analysis for the Punjabi Interview Speech Corpus

Punjabi Interview Speech Corpus ↓	Correct	Error	Accuracy (Correct Percentage)
Total Number of Punjabi Sentences = 461	455	6	98.6 %
Total number of Punjabi words = 1227	1213	14	98.8 %

7. Conclusion and Future Work

The objective of the research paper was to develop a speaker specific spontaneous speech recognition system for Punjabi language. So the purpose was to emphasis on the actual work done in the area of spontaneous speech recognition system. For the development of the research work, the sphinx tool and java programming has been used. The objective of the research work is to reduce the error rate in the speech model. The GUI for the spontaneous speech model for Punjabi language also has been created to test the live Punjabi speech interview using java framework. In the future direction, the system will be trained more vocabulary of Punjabi corpus and also try to recognize the speech of different person voice. The language model will also be improved for proposed research work, so that the recognition rate will be improved and recognition becomes faster.

References

- [1] Yogesh Kumar and Navdeep Singh. A First Step towards an Automatic Spontaneous Speech Recognition System for Punjabi Language. *International Journal of Statistics and Reliability Engineering*, Vol. 2 (1), PP. 81- 93, 2015.
- [2] M.Izzad, Nursuriati Jamil. Speech/Non-Speech Detection in Malay Language Spontaneous Speech. *IEEE proceeding*, PP: 219-224, 2013.
- [3] Wiqas Ghai and Navdeep Singh. Analysis of Automatic Speech Recognition Systems for Indo-Aryan Languages: Punjabi a Case Study. *International Journal of Soft Computing and Engineering*, ISSN: 2231-2307, Volume-2, Issue-1, PP. 379-385, 2012.
- [4] Wiqas Ghai and Navdeep Singh. Tri-Phone Based Acoustic Modeling On Continuous Speech Recognition for Punjabi Language. *IJCA*, Vol-72, pp. 23-28, 2013.
- [5] Mohit Dua, R.K.Aggarwal. Punjabi Automatic Speech Recognition Using HTK. *IJCSI International Journal of Computer Science Issues*. Vol. 9, Issue 4, No. 1, PP.359-364, 2012.
- [6] Jeet Kumar, Om Prakash Prabhakar and Navneet Kumar Sahu. 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. 2014.
- [7] Chetana Prakash, Suryakanth V. Gangashetty. Fourier-Bessel Cepstral Coefficients for Robust Speech Recognition. *Proceeding of IEEE*, PP. 1-5, 2012.
- [8] Anupriya Sharma, Amanpreet Kaur. 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, June 2013.
- [9] Anusuya, M. A., Katti. Front end analysis of speech recognition: A review. *International Journal of Speech Technology*, Springer, Vol.14, PP. 99–145, 2011.
- [10] Kumar, R. Comparison of HMM and DTW for Isolated Word Recognition System for Punjabi Language. *International Journal of Soft Computing*, PP.88-92, 2010.
- [11] N.Uma Maheswari, A.P.Kabilan and R.Venkatesh. A Hybrid model of Neural Network Approach for Speaker independent Word Recognition. *International Journal of Computer Theory and Engineering*, Vol.2, No.6, PP.1793-8201, 2010.
- [12] www.shabdKosh.com/pa/.../corpus/corpus-meaning-in-Punjabi-English.
- [13] <http://cmusphinx.sourceforge.net/>
- [14] <https://corplinguistics.wordpress.com/tag/punjabi/>
- [15] <http://research.microsoft.com/pubs/118769/Book-Chap-HuangDeng2010.pdf>.

Authors' Profiles



Yogesh Kumar is a PhD Research Scholar in the field of automatic spontaneous speech recognition system for Punjabi language under Punjabi University, Patiala. And prior to this, he did his M.Tech CSE from the Punjabi University.



Navdeep Singh is working as Associate Professor at Mata Gujri College, Sri Fatehgarh Sahib. Prior to this, he did his MCA from Thapar University, Patiala. He is having teaching experience of 15 years. His research areas include Automatic Speech Recognition, Software Engineering and Mobile Computing.

How to cite this paper: Yogesh Kumar, Navdeep Singh, "Automatic Spontaneous Speech Recognition for Punjabi Language Interview Speech Corpus", International Journal of Education and Management Engineering(IJEME), Vol.6, No.6, pp.64-73, 2016.DOI: 10.5815/ijeme.2016.06.07