An Automatic Spontaneous Speech Recognition System for Punjabi Language

Yogesh Kumar and Navdeep Singh

Abstract Punjabi is a very tonal language, making employ of a range of tones to distinguish words that would otherwise be alike. Three main tones can be recognized: high-rising-falling, mid-rising-falling, and low rising. Some work has been done in the field of isolated word, connected word, and continuous speech recognition system for Punjabi language. Spontaneous speech recognition is one area where no work has been done so far for Punjabi language. Spontaneous speech and speech from written language are exceptionally dissimilar both acoustically and linguistically. Spontaneous speech contains crammed silence, preservation, faltering, duplications, incomplete vocabulary, and stuttering. In this paper, an effort has been made to build an automatic spontaneous speech recognizer to recognize Punjabi live speech by using speech recognition model using sphinx toolkit.

Keywords Acoustic model · Feature vector · Sphinx · Decoder · Phones Transcript · Filler dictionary

1 Introduction

Dealing with unplanned speech [1] is one of the numerous challenges that Automatic Speech Recognition (ASR) systems for Punjabi language have to compact with. The primary indications describing spontaneous speech are hesitating like packed pause, repetition, repair and false start and many learning have paying attention on the recognition and improvement of these hesitations [2]. Therefore, identification of spontaneous speech will need a standard move from speech to accepting where original messages of the speaker are removed, as a

Y. Kumar (✉)
Department of Computer Engineering, Punjabi University Patiala, Patiala, Punjab, India
e-mail: Yogesh.arora10744@gmail.com

N. Singh
Department of Computer Science, Mata-Gujri College, Sri Fatehgarh Sahib, Punjab, India
e-mail: navdeep_jaggi@yahoo.com

S. S. Agrawal et al. (eds.), *Speech and Language Processing for Human-Machine Communications*, Advances in Intelligent Systems and Computing 664,
https://doi.org/10.1007/978-981-10-6626-9_7

substitute of transcribing every vocal words. Spontaneous speech, as evaluated to designed speech, is a more natural way in which people communicate with each other. However, the recognition of spontaneous speech is facing numerous challenging by the rigorous articulation alternatives and changeable silence gaps or amusement in between words. Presently, a variety of novel applications of LVCSR (large vocabulary continuous speech recognition) systems, such as automatic closed captioning, making minutes of meetings, conferences, and summarizing and indexing of speech documents for information retrieval, are dynamically being explored.

2 Automatic Spontaneous Speech Recognition System for Punjabi

Speech recognition [3] is a complicated task and states of the ability recognition systems are very complex. Automatic spontaneous speech has many prospective purposes including rule and organize, transcription of confirmed dialogue, live speech, and interactive vocal conversations (Fig. 1).

The primary phase [4] of speech identification is to reduce the speech signals into flows of acoustic feature vectors, called as *observations*. The key chore [5] of the speech system is to obtain an audio signal as input and fabricate a sequence of words as output. The acoustic model begins a mapping among phonemes and their potential acoustic demonstrations, i.e., the phones. The prior probability is computed using the language model. Usually trigram or even 4-g supported language models are utilized in recent speech systems. The decoding method [6] in a speech recognizer's procedure is to discover a string of words whose consequent acoustic and language models finest equivalent the input feature vector string. For that reason, the procedure of such a decoding process with trained audio and language models is often submitted to as a explore method.

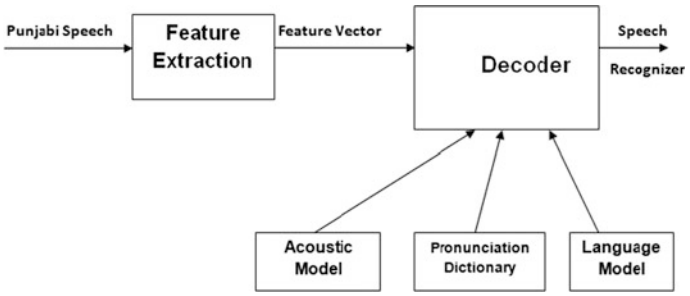

Fig. 1 Automatic speech recognition system for Punjabi speech

3 Building an Acoustic Model for Spontaneous Punjabi Speech

In order to build an acoustic model for spontaneous Punjabi speech, it is required to train the system with word level. But the single word wav file has small in size and silence gap is more therefore even for training single word, we need sentences. For this purpose, we trained the Punjabi spontaneous speech system with multiple words and sentences with variable silence gap.

A. Steps for training the acoustic model for Punjabi corpus

To train the system for Punjabi Language, we need following configuration files:

1. **Dic (Independent words are store in it)**:

The main purpose of the dictionary file is to map Punjabi stored words with the every recorded Punjabi sound unit associated with each sounds. Two types of the dictionaries are present, first type is used in which reasonable words in the language are planned progressions of sound units, and second type of dictionary in which non-vocalizations sounds are planned to corresponding non-vocalizations or speech-like sound units is also created. The training data which we are giving as input to our system are shown in the given figure [7, 8] (Fig. 2).

The dictionary file (Punjabi.dic file) will look like as shown in Fig. 3:

2. **Filler and noise**: It is also type of dictionary in which rejected noise is stored [2]. For example:

<s>	SIL
</s>	SIL
<sil>	SIL

3. **Phone**: Phone file [9] is a record of individual sound unit that needs to make a word. The various phone files are shown in the Table 1.

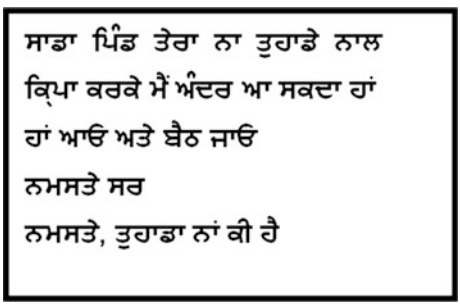

Fig. 2 Training data of Punjabi language

ਸਾਡਾ	ਸ ਾ ਡ ਾ
ਪਿੰਡ	ਪਿੰਡ
ਤੇਰਾ	ਤ ੇ ਰ ਾ
ਨਾ	ਨ ਾ
ਤੁਹਾਡੇ	ਤ ੁ ਹ ਾ ਡ ੇ
ਨਾਲ	ਨ ਾ ਲ
ਕ੍ਰਿਪਾ	ਕ੍ਰਿਪ ਾ
ਕਰਕੇ	ਕ ਰ ਕ ੇ
ਮੈਂ	ਮੈ ਂ
ਅੰਦਰ	ਅੰਦਰ
ਆ	ਆ
ਸਕਦਾ	ਸਕਦ ਾ
ਹਾਂ	ਹਾ ਂ
ਸਰ	ਸਰ
ਆਓ	ਆਓ
ਅਤੇ	ਅਤ ੇ
ਬੈਠ	ਬ ੈ ਠ
ਜਾਓ	ਜ ਾ ਓ
ਨਮਸਤੇ	ਨਮਸਤ ੇ
ਤੁਹਾਡਾ	ਤ ੁ ਹ ਾ ਡ ਾ
ਨਾਂ	ਨਾ ਂ
ਕੀ	ਕ ੀ
ਹੈ	ਹ ੈ

Fig. 3 Dictionary files of Punjabi corpus

Table 1 Phone files of Punjabi language

ਸ	ਡ	ਤ	ਰ	ਨ	ਹ	ਲ	ਕ
ਪ	ਜ	ਦ	ਠ	ਅੰ	ਆ	ਪਿੰ	ਕ੍ਰਿ
ਮੈ	ਹਾ	ਓ	ਅ	ਬ	ਮ	ਨਾ	ੁ
ੈ	ਾ	ਂ	ੈ	ੀ			

4. **Transcript** (path of wav files) **and Fields** (conversation of wav File):

Transcription file is a listing the dictation for each acoustic file. For example, in our Punjabi corpus, the Table 2 shows the transcription file for test audio:

It is essential that each line of Punjabi text begins by <s> and finishes by </s> followed by id in parentheses. Also note that parenthesis includes only the file, exclusive of speaker_n directory. It is vital to have correct match among fields file and the transcription file.

We have two kinds of transcript and field files:

- **For training purpose** (Punjabi_parpare.trans and Punjabi_parpare.fileds)
- **For testing purpose** (Punjabi_check.trans and Punjabi_check.fileds)

Table 2 Transcript file

<s> ਸਾਡਾ ਪਿੰਡ ਤੇਰਾ ਨਾ ਤੁਹਾਡੇ ਨਾਲ</s>	(test1.wav)
<s>ਤੇਰਾ ਨਾ ਤੁਹਾਡੇ ਨਾਲ ਸਾਡਾ ਪਿੰਡ</s>	(test2.wav)
<s>ਪਿੰਡ ਤੇਰਾ ਨਾ ਤੁਹਾਡੇ ਨਾਲ ਸਾਡਾ</s>	(test3.wav)
<s>ਕ੍ਰਿਪਾ ਕਰਕੇ ਮੈਂ</s>	(test4.wav)
<s>ਅੰਦਰ ਆ ਸਕਦਾ ਹਾਂ</s>	(test5.wav)
<s>ਅੰਦਰ ਆ</s>	(test6.wav)
<s>ਸਕਦਾ ਹਾਂ ਸਰ</s>	(test7.wav)
<s>ਅੰਦਰ</s>	(test8.wav)
<s>ਕ੍ਰਿਪਾ ਕਰਕੇ ਮੈਂ ਅੰਦਰ ਆ ਸਕਦਾ ਹਾਂ</s>	(test9.wav)
<s>ਕ੍ਰਿਪਾ ਕਰਕੇ ਮੈਂ ਅੰਦਰ ਆ ਸਕਦਾ ਹਾਂ</s>	(test10.wav)
<s>ਮੈਂ ਅੰਦਰ ਆ ਸਕਦਾ ਹਾਂ</s>	(test11.wav)
<s>ਕ੍ਰਿਪਾ ਕਰਕੇ ਮੈਂ ਅੰਦਰ ਆ ਸਕਦਾ ਹਾਂ</s>	(test12.wav)
s>ਕ੍ਰਿਪਾ ਕਰਕੇ ਮੈਂ ਅੰਦਰ ਆ ਸਕਦਾ ਹਾਂ</s>	(test13.wav)
<s>ਆਓ ਅਤੇ ਬੈਠ ਜਾਓ</s>	(test14.wav)
<s>ਹਾਂ</s>	(test15.wav)
<s>ਆਓ ਅਤੇ ਬੈਠ ਜਾਓ</s>	(test16.wav)
<s>ਹਾਂ ਆਓ ਅਤੇ ਬੈਠ ਜਾਓ</s>	(test17.wav)
<s>ਨਮਸਤੇ ਸਰ</s>	(test18.wav)
<s>ਨਮਸਤੇ ਤੁਹਾਡਾ ਨਾਂ ਕੀ ਹੈ</s>	(test19.wav)

Training files are used to create feature vector which will be used later for recognition. Testing files are used by decoder to check the recognition. **Sphinx_train.test file**: This is the configuration file where configuring the path for all required files (for field, transcript, etc.).

4 Steps of Creating the Language Model for Punjabi Corpus

Language model is used for decoding purpose. The language model gives framework to differentiate between words and expression that sounds alike. There are two forms of language models [10] that illustrate language—grammars and statistical language models [11, 12]. Grammars portray very simple forms of languages for grasp and organize, and they are usually written manually or produced mechanically with plain code [13, 14]. Steps for creating language model are:

Step1: During compilation, first we input given text file as shown in Fig. 4.
Step2: Execute cmu command and create vocab file (Fig. 5).
Step3: Finally, language model is created with extension lm.DMP, which is used for training purpose. While training it use decoder to test the training and generate log files of decoding.

Figure 6 clearly shows that while compiling the Punjabi acoustic model for spontaneous speech, out of 128 lines and 390 words, only 2 lines and 1 word are failed. So the sentence error rate is 1.6% and word error rate is 0.5%.

5 Graphical User Interface for Automatic Spontaneous Speech System for Punjabi Language

Language model and training data are both compiled in final jar file which is used for recognition. For live testing of speech, we have created the java based GUI for spontaneous Punjabi speech (Fig. 7).

It has an option of live speech test and speech recognition for already recorded wav files.

Fig. 4 Input Punjabi text file

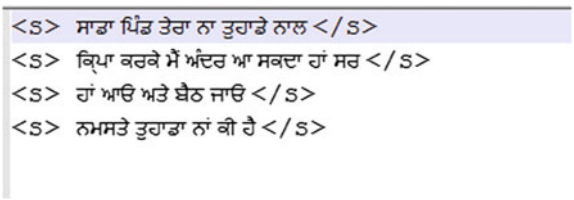

Fig. 5 1-, 2-, and 3-g after compiling vocab file

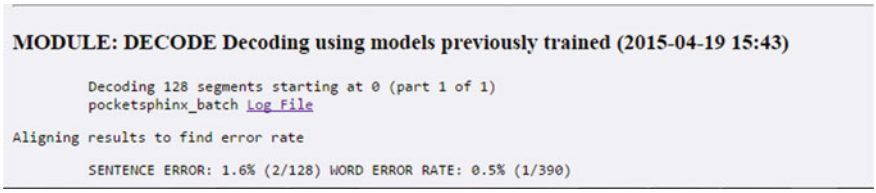

Fig. 6 Output of the decoder for Punjabi corpus

Figure 8 clearly shows that the output of the live speech testing for spontaneous Punjabi speech.

6 Performance Evaluation

The performance of the research work is evaluated by comparing it with previous work done for small vocabulary system [5]. In the previous research, the total numbers of sentences were taken 7 and words were 42 of Punjabi language [15, 16, 17]. The present work has total 128 sentences and 390 words. Table 3 shows the comparison between the previous and present work on the basis of sentences error and word error rate.

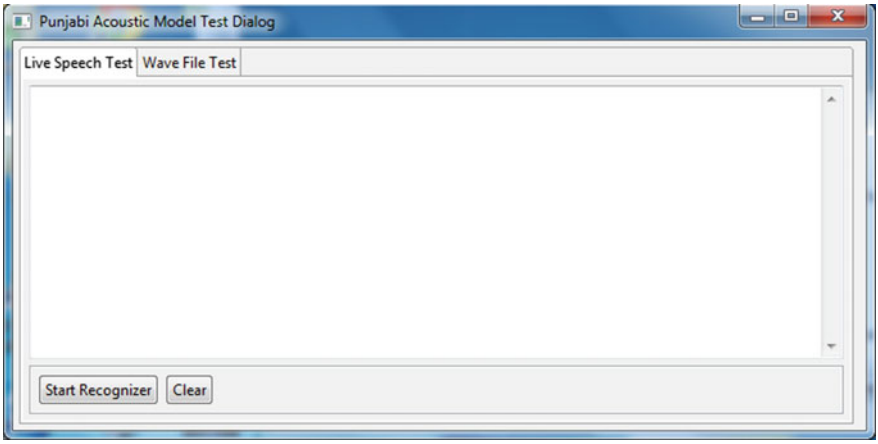

Fig. 7 GUI for spontaneous Punjabi speech recognition

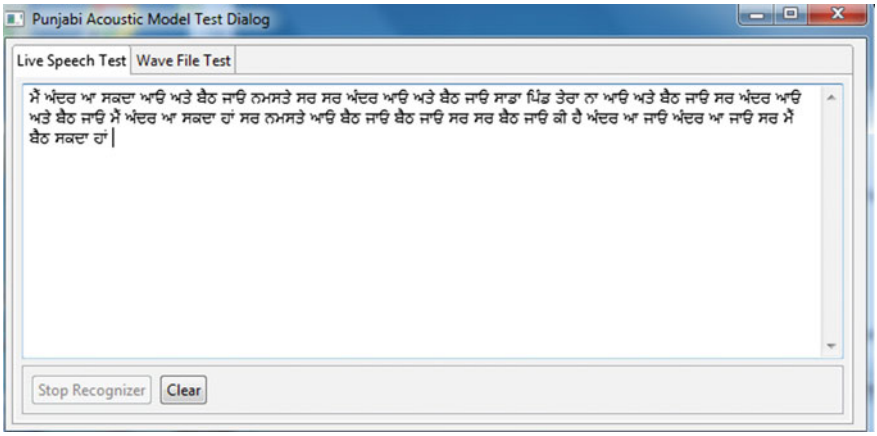

Fig. 8 Output of the Punjabi spontaneous speech recognition model

Table 3 Result comparison

Previous work		Present work	
Total number of sentences 7	Total number of words 42	Total number of sentences 128	Total number of words 390
Sentence error rate 28.6%	Word error rate 4.8%	Sentence error rate 1.6%	Word error rate 0.5%

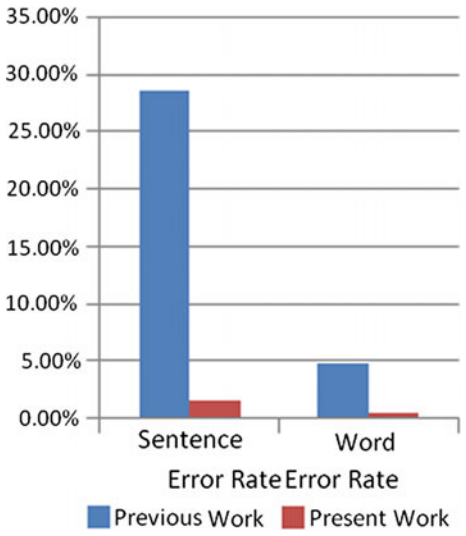

Fig. 9 Performance comparison

Graphical analysis shown in Fig. 9 represents drastic reduction in the word and sentence error rate with increase in vocabulary size in the previous and present work.

7 Conclusion and Future Work

In this paper, an effort has been made to develop an automatic spontaneous speech recognition system for Punjabi corpus using sphinx toolkit. The accomplishment of spontaneous speech detection system has considerably improved in provisions of sentence along with word error rate. GUI has been created to test the live Punjabi speech using java framework. In future, system will be trained for large vocabulary so that recognition rate can be improved for voice input taken from the different person. The Language model will also be improved in future work for fast decoding and recognition.

References

1. Atassi, H., Smékal, Z.: Emotion recognition from spontaneous slavic speech. In: 3rd IEEE International Conference on Cognitive Info Communications, 2–5 December 2012
2. Furui, S.: Spontaneous speech recognition and summarization. In: Proceedings IEEE Workshop on Spontaneous Speech Processing and Recognition (2010)

3. Singh, P., Dutta, K.: Formant analysis of punjabi non-nasalized vowel phonemes. In: The International Conference on Computational Intelligence and Communication Systems, pp. 375–380, Proceedings IEEE (2011)
4. Dua, M., Aggarwal, R.K.: Punjabi automatic speech recognition using HTK. IJCSI Int. J. Comput. Sci. Issues **9**(4), No 1 (2012)
5. Kumar, Y., Singh, N.: A first step towards an automatic spontaneous speech recognition system for Punjabi language. Int. J. Stat. Reliab. Eng. **2**(1), 81–93 (2015)
6. http://research.microsoft.com/pubs/118769/Book-Chap-HuangDeng2010.pdf
7. www.shabdkosh.com/pa/…/corpus/corpus-meaning-in-Punjabi-English
8. https://corplinguistics.wordpress.com/tag/punjabi/
9. http://cmusphinx.sourceforge.net/
10. www.speech.cs.cmu.edu/sphinx/doc/Sphinx.html
11. Sixtus, A., Molau, S., Kanthak, S.: Spontaneous speech characterization and detection in large audio database. In: SPECOM'2009, St. Petersburg, 21–25 June 2009
12. Izzad, M., Jamil, N.: Speech/non-speech detection in malay language spontaneous speech. In: The proceedings IEEE 2013, pp 219–224 (2013)
13. Shih, P.O., Chen, B.W.: Enhanced lengthening cancellation using bidirectional pitch similarity alignment for spontaneous speech. In: The international Symposium on Chinese Spoken Language Processing Proceedings (2012)
14. Ghai, W., Singh, N.: Analysis of automatic speech recognition systems for Indo-Aryan languages: Punjabi a case study. Int. J. Soft Comput. Eng. (IJSCE), **2**(1) March 2012. ISSN: 2231–2307
15. Ghai, W., Singh, N.: Tri-phone based acoustic modeling on continuous speech recognition for Punjabi language. IJCA, **72** (2013)
16. Hu, X., Wu, Y.: Collecting sentences from web resources for constructing spontaneous Chinese language model. In: The International Symposium on Chinese Spoken Language Processing Proceedings (2012)
17. Akita, Y., Kawahara, T.: Statistical transformation of language and pronunciation models for spontaneous speech recognition. In: The IEEE Transactions on Audio, Speech, and Language Processing, **18**(6) (2010)