An Experimental Investigation of Automatic Speech to Multilingual Text Translation(S-MLT^2) System

Amitoj Singh[1], *Virender Kadyan*[2], *Munish Kumar*[3]

[1,3]Assistant Professor, MRSPTU, Bathinda
[2] Assistant Professor-Research,CURIN, Chitkara University, Punjab, India

Abstract

On the path of meaningful translation that can be well versed and experienced, the motivation has been Multi-Lingual Speech translation for differently-abled users of an under-resourced Language such as Punjabi. In order to process the system, input from various speakers(20 males and 15 females of each dialect i.e. powadi, majhi Doabi and malwai regions) have been processed to minimize the tonal and dialectal differences. Collected corpus for idioms is processed with the help of mfc technique, the language model consisting of dictionary and phones of Punjabi language components, while decoding the speech signal, it is processed into text that act as an input for 'Grammatical Framework'. The recognized text output works as abstract syntax for Grammatical Framework, which undergoes its concrete matching to give text in meaning and sentence form using three different Indian languages(Hindi, Punjabi and English). With this paper, we tackle various issues faced during training and testing, and focus mainly on four main factors-accurate detection of spoken utterance, logical translation of idioms(because idioms don't have literal translation from one language to another), an interactive GUI and expansion of resources for this under-resourced language.

Index Terms: Automatic Punjabi speech recognition, HMM, Text to text translator, hidden Markov model

1. Introduction

Speech is the vocalized form of communication and probably the main mode of communication. But due to the evolution of numerous languages across the world, there are problems faced by people who speak different languages. This translator works for the translation of Punjabi language (Punjabi being the 10th most spoken language in the world) into other Indian Languages(Punjabi, Hindi and English). The Automatic SMLT^2 system is built with the help of CMU Sphinx toolkit for generation of Speech to text output, Grammatical Framework for Punjabi text to other language Text Translation. Despite of the numerous sources available[1], people face the problem of meaningful translation of Punjabi language Idiom into others.

2. Motivation

Punjabi being an under-resourced language[2], it came into consideration and became the main motive of this research. Moreover, in order to connect some of the disabled persons (ones with hearing impairment) to normal education and provide them with proper educational facilities, we thought of designing a system which would convert speech/text into further multilingual text as required.

3. Corpus for Speech

The basic input required in ASR (Automatic Speech Recognition) is speech that is converted into text and the text output is further translated into different languages(Hindi and English). This input acts as a base for training and testing process required in any automatic speech pattern recognition system. The corpus is built with the help of 20 male and 15 female speakers. The database is application specific based on Punjabi idioms. The collected corpus covers various tonal and dialectal information of Punjabi language. Also, the database/corpus is collected mainly from the native speakers of Punjabi in 16Khz frequency using Shure SM-10 microphone with Sound forge software. The speech data is recorded in in-house studio in noise free environment. Th transcription of data is done by the native speaker of Punjab that have knowledge of tonal word spoken in particular dialect[3][4].

4. Proposed Architecture of S-MLT^2 (Speech to Multilingual Text Translation) System

The system on the whole consists of two module– Speech to text recognizer module and Text to multi-language text conversion module.

During first phase, the speech input is recorded that covers tonal information of a particular dialect of Punjabi Language which is further processed by Sphinx toolkit, from which the input is matched with the ones in the language model. The language model consists of the dictionaries and phones of the language components. It is then followed by an acoustic model which is the digitalized form of the audio. Then, the decoding model recognizes the complete speech input and gives the final output as text.

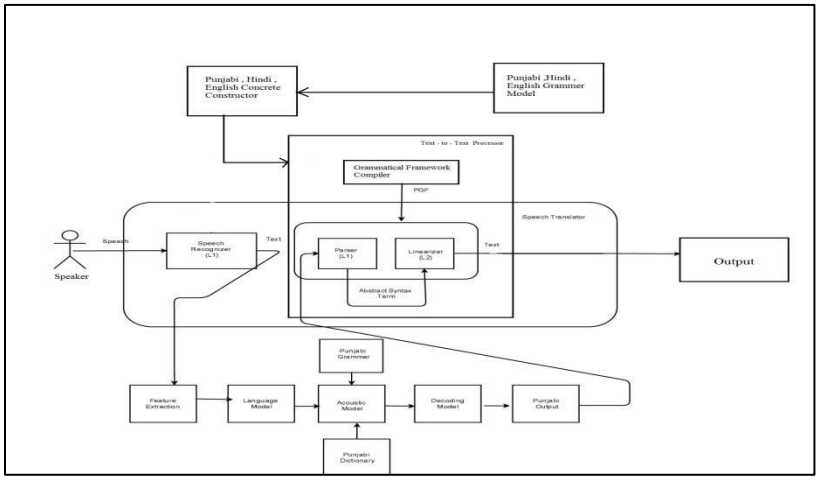

Fig 1.1 Architecture of S-MLTT (Speech to Multilingual Text Translation)

During second phase, this recognized text is fed into the system of Grammatical Framework which makes use of its Multi-language Grammar model to decode the abstract syntax of the given input. The text is then parsed to form trees and further the process of linearization follows. During this, the specific concrete syntax's of the aforementioned abstract are deployed for giving the output in different languages. The final output is then displayed on the terminal device.

Also, the technology of Qt is used to design the GUI for this system to make it more user-friendly and interactive.

5. Discussion

Our MLT^2 system, will processed the idiom that covers more tonal word information i.e. the word that has phone whose position occur at initial, middle and end location in the word. Due to occurrence o different 5 phoneme at three different location, it will result in change of F0 value of spoken utterance of that word by the native speaker of Punjabi. As Punjabi contain three different tone low, high and mid[] due to which it is capable of carrying toneme information that is only spoken by native speaker of Punjab in different dialectal regions. Some words are totally different due to present of this tonal information.

This research work explain the importance of these words in the collected speech corpus and it impact while testing the system. A more likely information of tonal words are required while preparing an Punjabi ASR system otherwise it will degrade the performance of the system. More tonal work is done in other Indian and Non-Indian Language like Mandrian[5][6] and Mizo[7] etc. Our result here suggested the significance of inclusion of Tonal information and meaningful transformation of one language into other language to make resources for this less resource language to assist to needy person of society.

6. Results

We were able to collect voices from various speakers of different dialect of Punjabi . This paper summarizes some of patterns as mentioned in the tables below.

Table1 Testing of Speech to text utterance in noisy environment

S.No.	Speaker Type	No. of spoken idioms	No. of detected idioms	Percentage accuracy	Percentage correctness	Error rate
Speaker 1	Male	3	2	66.67	66.67	33.33
Speaker 2	Female	4	3	75	75	25
Speaker 3	Male	10	8	80	80	20
Speaker 4	Female	7	5	71.42	71.42	28.58
Speaker 5	Male	12	9	75	75	25

Table2 Testing of Speech to text utterance in different environments

Speaker	Environment	No. of spoken idioms	No. of detected idioms	Percentage accuracy	Percentage correctness	Error rate
Speaker 1	Open	3	1	33.33	33.33	66.67
Speaker 2	Closed	4	3	75	75	25
Speaker 3	Open	10	7	70	70	30
Speaker 4	Closed	7	5	71.42	71.42	28.58
Speaker 5	Open	12	8	66.67	66.67	33.33

Both the tables above (Table1 abd Table2) depict the accuracy obtained during the testing process of Speech to text utterance by native and non-native speaker of Punjabi language. We procure different values of the correctness of detection system, resulting into different values of the error rates corresponding to each speaker. An important perception of these results is that the accuracy of detection power of the MLT^2 system in a closed environment (noiseless) is much more than that of the system in an open environment (noisy).

Table3 Testing of Speech to multi-language Text conversion in different environment

Speaker	**No. of spoken idioms**	**Correctly translated idioms in Hindi**	**Correctly translated idioms in English**	**Correctly translated idioms in Punjabi**
Speaker 1	5	4	3	4
Speaker 2	12	10	10	11
Speaker 3	7	6	5	7

As told before, Grammatical Framework does the work of multilingual translation. Therefore, as per table 3, the results obtained on testing the accuracy of Speech to multi-language Text conversion in different environment, we can see the number of idioms which have been correctly translated into Hindi, English and Punjabi respectively. This also depends on the speaker because the translation process follows speech recognition which gives a different result for a different speaker.

7. Future Prospects

The future prospect of this research is basically covering up the complete Punjabi language including all its dialects. Our target is to make a spontaneous multi-language translator not only for idioms but for each and every kind of sentence, phrase, etc. The system would prove to be of great benefit to school students studying in government funded schools with minimal educational facilities. It would also help people with hearing impairments as they would be able to understand and learn a language merely through the display of content in an application.

8. Conclusion

Although there are numerous translators available today in the market, but all of them perform only word-to-word translation, which mostly does not result into a meaningful sentence or conversion. Here, we build a application based small system with meaningful translation and also try to achieve maximum accuracy with the speech to text translation section. The resultant system is achieved in the form of an application which is compact in size, requires no internet connection and gives control on quality. The system gain accuracy of 94.3% for 1000 Punjabi Idioms trained by four dialectal region speakers of Punjab, that covers 400 most spoken tonal words and tested by 20 native and non-native speakers of 15 to 30 years age group, finally it induces an error rate of 5.7%.

9. References

1. Karamat, N., (2001). Phonemic inventory of Punjabi. *Annual student report*, pp.179-83.

2. Lata, S., & Arora, S. (2013). Laryngeal tonal characteristics of Punjabi—An experimental study. In *Human Computer Interactions (ICHCI), 2013 International Conference* pp. 1-6, IEEE.

3. Swaran, L. (2011). Challenges for Design of Pronunciation Lexicon Specification (PLS) for Punjabi Language pp.56-58.

4. Singh, H. (1991). Prominent features of Punjabi language. *Patiala: Publication Bureau, Punjabi University*.

5. Y.Xu (1997), Contextual tonal variations in Mandarin, Journal of Phonetics, vol. 25, pp. 61-83.

6. Huang, H., & Jie, Z. H. U. (2007). Discriminative tonal feature extraction method in mandarin speech recognition. *The Journal of China Universities of Posts and Telecommunications*, *pp.* 126-130.

7. Sarma, B. D., Sarmah, P., Lalhminghlui, W., & Prasanna, S. M.(2015) Detection of Mizo Tones. In *Sixteenth Annual Conference of the International Speech Communication Association*.