EasyChair Preprint

№ 4484

Research Insight of Indian Tonal Languages: A Review

Jaspreet Kaur Sandhu and Amitoj Singh

October 30, 2020

Research Insight of Indian Tonal Languages: A Review

Jaspreet Kaur Sandhu[1], Dr. Amitoj Singh[2]

mca.jaspreetsandhu@gmail.com,amitojsingh@mrsptu.ac.in
Department of Computational Sciences,Maharaja Ranjit Singh Punjab Technical University, Bathinda,Punjab,India

Abstract

The most challenging research areas are the man-machine interaction and natural language. Speech signal conversion into text with the computer is called Automatic Speech Recognition (ASR). In the field of research, ASR is noted as a most important research domain in speech technology. This paper presents the research done in previous years related to ASR with respect to Indian Tonal Languages like Punjabi, Bodo, Mizo.It is observed that less research is carried out for Indian Tonal languages ASR.

Keywords: Speech, Automatic Speech Recognition, Tonal, Punjabi,Mizo.

1.Introduction

The most powerful method of communication is speech. Encoded information is contained in the speech and without any efforts; human can decode this encoded information. So researchers are interested in the development of a system that can extract and process huge quantities of information from speech automatically. They also want the speech to be recognized by computers, the recognition can be voice commands or it can be in the form of text as speech result. ASR is the system where the conversion of speech signals with respect to their respective matching text is done with the aid of computers. These days ASR has become famous as well as potential. The interaction between human and machine is growing faster in everyday applications due to the developing of computer science. ASR systems for Indian tonal languages have been presented in this paper. The section 2 of the paper discusses about Indian tonal languages and section 3 about the work done in India Tonal languages ASR. In section4 analysis of the work is presented and section 5 concludes the paper.

2. Tonal Languages

When in a language the pitch of a word result in the change of meaning of words then the language is said to be tonal language. The tone is associated with every syllable of the language. Whenever a change in pitch is observed, then it is called tone. The tonal languages contained the following

Detection of tone
Consonants of syllables
Vowels of syllables

The phonological characteristic of Indian tonal languages is represented in table1.

Table 1: Indian tonal langauges phnological features

Indian tonal languages	Langauge classification family	Tones & Types
Bodo	Sino-Tibetan	2 -high, low
Mizo	Sino-Tibetan	4-high,low,rising,falling
Punajbi	Indo-European	3-high, mid, low

3. Research Work Done on Indian Tonal languages

Punjabi

Kumar (2010) done the comparison of Hidden Markov Model(HMM) and Dynamic Time Wrap(DTW) for Recognition of isolated Punjabi words. 91.3% accuracy obtained by HMM and 94% by DTW.I in the year 2012 Dua et al. (2012b) implemented Punjabi Speech To Text (STT) system in the classroom as well as in an open environment. Word Error Rate (WER) noted for the classroom and open environment was 5.9% and 8.35%, respectively. In the same year Ghai and Singh (2012) addressed the efforts done for ASR for Assamese, Oriya, Hindi, Bengali, Marathi, Sinhala , Punjabi, Urdu by many researchers. Lata and Arora (2012) reviewed Punjabi tones related with orthographic characters and found that starting as well as middle position result in falling tones. Ghai and Singh(2013a) done work for Punjabi continuous speech with HMM. 94.32% word recognition accuracy was obtained and 82.18% for sentence. /h/ sound of Punjabi was analyzed by Lata and Arora (2013) with PRAAT and MATLAB on Malawi dialect .Kaur and Singh(2016a) designed Punjabi ASR based on connected words using Power Normalized Cepstral Coefficients (PNCC) and HMM. Noise free environment gave WER of 16.28% and noisy environment provided WER of 32.08%.Punjabi ASR with connected words was built by Kaur and Singh (2016b) with Mel–Frequency Cepstral Coefficient(MFCC), Perceptual Linear Prediction(PLP) and PNCC.13.19% WER was obtained with MFCC and 16.28% with PLP whereas PNCC obtained 16.28%.Automatic Spontaneous Punjabi speech model was build by Kumar and Singh(2016b) with the help of Sphinx tool. Punjabi words achieved 85.38% accuracy and Punjabi sentences noted 91.17% accuracy. Kadyan et al. (2017a) presented heterogeneous feature extraction on speech signal to neutralize effect of many conditions of the non-ideal environment.HMM with Genetic Algorithm(GA),HMM with Differential Evolution(DE) was used. The highest accuracy was obtained by HMM+DE. Punjabi ASR system performance was evaluated by Kadyan et al. (2017b) .The Word Accuracy (WA) on Malawi dialect by DE-HMM was 84.96%, 83.26% for GA-HMM and 78.44% for HMM. Kumar and Singh (2017) designed Punjabi automatic speaker independent speech model to recognize Punjabi speech.Guglani and Mishra (2018) observed that less WER is obtained with MFCC than PLP on Continuous Punjabi speech. Acoustic mismatched was reduced between training as well as testing conditions with the use of the Deep Neural Network(DNN)-HMM was stated by Kadyan et al. (2018).Pitch, fundamental frequency, intensity on Punjabi dialects with their effects was studied by Arora et al.(2019).

Mizo

Pitch slope as well as duration to synthesize Mizo tones was studied by Govind et al. (2012) The correctness score for falling tone was 84% and for low tone it was 63% and for rising tone it was 98%.Sarma et al.(2015) developed Mizo automatic method for detection of Mizo level tone. Classification method was on the basis of values of pitch and parameters of height that were obtained from the values of pitch fitted with tone.Dey et al.(2017) used DNN as well as Subspace Gaussian Mixture Model(SGMM) to build phone recognition system of Mizo.DNN gave 23.2% of Phone Error Rate (PER) and SGMM provided 10.3%.

Bodo

Thakuria and Das (2013) used HMM to build ASR for Bodo.The summary of the important work related to ASR of Indian tonal languages is represented in table2.

Table 2: Work on ASR of Indian Tonal Languages

Author	Language	Dataset	Feature extraction methods	Classifier	Result
Govind et al.(2012)	Mizo	1speaker, 40 min	-	-	Correctness (%) 84.6: Falling tone: 63: Low tone 98: Rising tone:
Sarma et al.(2015)	Mizo	5 speakers Syllable contained 672high tones 2254lowtones 808falling tones 650rising tones			DR(%) Without CE DR(%) with CE 50.14 35.26 81.28 56.78 75.23 78.92 52.47 74.62
Dey et al.(2017)	Mizo	Training 62 speakers, 7394words, 5 hrs 5 min Testing 19speakers, 1568 words	MFCC	SGMM DNN	PER (%) 10.3 23.2
Dua et al.(2012a)	Punjabi	30-35 words	MFCC	HMM	Accuracy (%) 95.63(classroom environment) 94.08(open environment)
Dua et al.(2012b)	Punjabi	30-35 words	MFCC	HMM	WER (%) 5.9 (classroom environment) 8.35(open environment)
Ghai and Singh(2013b)	Punjabi	3Speakers, 200 words	MFCC	HMM	Word Recognition Accuracy(%) Connected 87.75 91.62
Kaur and Singh(2016a)	Punjabi	3speakers,43 words (noise free) 3speakers,53 words(noisy)	PNCC	HMM	WER (%) 16.28 (Noisy free) 32.08 (Noisy)
Kumar and Singh(2016a)	Punjabi	Punjabi live speech 462sentences, 1213 words	MFCC	HMM	Recognition accuracy (%) 98.6(sentences) 98.8(words)

Kadyan et al.(2017a)	Punjabi	25speakers, 5000words, 32 hrs	MFCC	HMM GA-HMM DE-HMM HMM GA-HMM DE+HMM	Accuracy (%) Speaker Dependent 60.97 70.42 74.59 Speaker Independent 58.48 67.36 72.47
Kadyan et al.(2017b)	Punjabi	Training 24speakers ,58700words,46 hrs 40 min Testing 13 speakers, 6100 words, 5hrs30 min	MFCC	DE-HMM GA-HMM HMM DE-HMM GA-HMM HMM	WA (%) Noisy Environmnet 84.96(Malwai) 83.26(Malwai) 78.44(Malwai) Real Environment 82.26(Malwai) 78.47(Majhi) 80.49(Malwai)
Guglani and Mishra(2018)	Punjabi	24 speakers, 0-9 digits in Malwai dialect,240 hrs	MFCC PLP	GMM HMM	WER(%) 21.2 22.7
Kadyan et al.(2018)	Punjabi	Continuous speech corpus 3611sentences, 13 speakers	MFCC GFCC	DNN-HMM	WER(%) 5.22 24.67

4. Synthesis Analysis of Work Done for Indian Tonal Languages ASR

Mizo and Bodo

- There is lack in research in the area of Mizo continuous ASR.
- Non-availability of standard corpus .

Punjabi

- Need to develop an ASR with advanced user interface.
- More work to be done on continuous as well as on spontaneous speech
- Standard corpus is not available
- Work to be done on large size vocabulary

5. Conclusion

This paper discussed the classifiers and feature extraction techniques used for Indian tonal languages ASR .The paper suggest that a very much research has to be carried out in Indian tonal languages ASR. In future ASR system for other tonal languages of globe will be reviewed

Appendix:

ASR: Automatic Speech Recognition
CE: Consonantal Effects
DE: Differential Evolution
DNN:Deep Neural Network

DR: Detection Rate
DTW: Dynamic Time Wrap
GA: Genetic Algorithm
GFCC: Gammatone Frequency Cepstral Coefficient
GMM: Gaussian Mixture Model
HMM: Hidden Markov Model
MFCC: Mel–Frequency Cepstral Coefficient
PER: Phone Error Rate
PLP:Perceptual Linear Prediction
PNCC: Power Normalized Cepstral Coefficients
SGMM: Subspace Gaussian Mixture Model
STT: Speech-To-Text
WA: Word Accuracy
WER: Word Error Rate

References:

Arora, A., Kadyan, V. and Singh, A., 2019. Effect of tonal features on various dialectal variations of Punjabi language. In Advances in Signal Processing and Communication (pp. 467-475). Springer, Singapore.

Dey, A., Lalhminghlui, W., Sarmah, P., Samudravijaya, K., Prasarma, S.M., Sinha, R. and Nirrnala, S.R., 2017, December. Mizo Phone Recognition System. In 2017 14th IEEE India Council International Conference (INDICON) (pp. 1-5). IEEE.

Dua, M., Aggarwal, R.K., Kadyan, V. and Dua, S., 2012. Punjabi Automatic Speech Recognition using HTK. International Journal of Computer Science Issues (IJCSI), 9(4), p.359.

Dua, M., Aggarwal, R.K., Kadyan, V. and Dua, S., 2012a. Punjabi Automatic Speech Recognition using HTK. International Journal of Computer Science Issues (IJCSI), 9(4), p.359.

Dua, M., Aggarwal, R.K., Kadyan, V. and Dua, S., 2012b. Punjabi Speech to Text System for Connected Words. *In 4th International Conference on Advances in Recent Technologies in Communication and Computing*, 206-209.

Ghai, W. and Singh, N., 2012. Analysis of Automatic Speech Recognition Systems for Indo-Aryan Languages: Punjabi A Case Study. International Journal of Soft Computing and Engineering (IJSCE), 2(1), pp.379-385.

Ghai, W. and Singh, N., 2013. Phone Based Acoustic modeling for Automatic Speech Recognition for Punjabi Language. Journal of Speech Sciences, 1(3), pp.69-83

Ghai, W. and Singh, N., 2013a. Continuous Speech Recognition for Punjabi Language. International Journal of Computer Applications, 72(14).

Govind, D., Sarmah, P. and Prasanna, S.M., 2012. Role of Pitch Slope and Duration in Synthesized Mizo Tones. In Speech Prosody 2012.

Guglani, J. and Mishra, A.N., 2018. Continuous Punjabi Speech Recognition Model based on Kaldi ASR Toolkit. International Journal of Speech Technology, 21(2), pp.211-216.

Kadyan, V., Mantri, A. and Aggarwal, R.K., 2017a. A Heterogeneous Speech Feature Vectors Generation Approach with Hybrid HMM Classifiers. International Journal of Speech Technology, 20(4), pp.761-769.

Kadyan, V., Mantri, A. and Aggarwal, R.K., 2017b. Refinement of HMM Model Parameters for Punjabi Automatic Speech Recognition (PASR) System. IETE Journal of Research, 64(5), pp.673-688.

Kadyan, V., Mantri, A., Aggarwal, R.K. and Singh, A., 2019. A Comparative Study of Deep Neural Network based Punjabi-ASR System. International Journal of Speech Technology, 22(1), pp.111-119.

Kaur, A. and Singh, A., 2016a, July. Power-Normalized Cepstral Coefficients (PNCC) for Punjabi Automatic Speech Recognition using Phone based Modelling in HTK. In 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT) (pp. 372-375). IEEE.

Kaur, A. and Singh, A., 2016b, September. Optimizing Feature Extraction Techniques Constituting Phone based Modelling on Connected Words for Punjabi Automatic Speech Recognition. In 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI) (pp. 2104-2108). IEEE.

Kumar, Y. and Singh, N., 2016a. Automatic Spontaneous Speech Recognition for Punjabi Language Interview Speech Corpus. International Journal of Education and Management Engineering (IJEME), 6(6), pp.64-73..

Kumar, Y. and Singh, N., 2016b. Automatic Spontaneous Speech Recognition for Punjabi Language Interview Speech Corpus. International Journal of Education and Management Engineering (IJEME), 6(6), pp.64-73.

Kumar, Y. and Singh, N., 2017. An Automatic Speech Recognition System for Spontaneous Punjabi Speech Corpus. International Journal of Speech Technology, 20(2), pp.297-303.

Lata, S. and Arora, S., 2012, May. Exploratory Analysis of Punjabi Tones in Relation to Orthographic Characters: A Case Study. In Workshop on Indian Language and Data: Resources and Evaluation Workshop Programme (p. 76).

Lata, S. and Arora, S., 2013, August. Laryngeal Tonal Characteristics of Punjabi—An Experimental Study. In 2013 International Conference on Human Computer Interactions (ICHCI) (pp. 1-6). IEEE.

Ravinder, K., 2010, November. Comparison of HMM and DTW for Isolated Word Recognition System of Punjabi Language . In Iberoamerican Congress on Pattern Recognition (pp. 244-252). Springer, Berlin, Heidelberg.

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

Thakuria, L.K., Acharjee, P., Das, A. and Talukdar, P.H., 2013. BODO Speech Recognition Based on Hidden Markov Model Toolkit (HTK). International Journal of Scientific and Engineering Research, 4(12), pp.2309-2313.