Interspeech 2018
2-6 September 2018, Hyderabad

AGROASSAM: A Web Based Assamese Speech Recognition Application for Retrieving Agricultural Commodity Price and Weather Information

Abhishek Dey[1], *Abhash Deka*[2], *Siddika Imani*[1], *Barsha Deka*[1], *Rohit Sinha*[2], *S.R.M. Prasanna*[2,3], *Priyankoo Sarmah*[2], *K. Samudravijaya*[2] *and S.R. Nirmala*[1]

[1]GUIST, Gauhati University, Guwahati-781014, India
[2]Indian Institute of Technology Guwahati, Guwahati-781039, India
[3]Indian Institute of Technology Dharwad, Dharwad-580011, India

{abhishekdey.gu, abhashdeka1, siddika.115, barsha.deka4, nirmalasr3}@gmail.com,
{rsinha, prasanna, priyankoo, samudravijaya}@iitg.ac.in

Abstract

This paper presents a speech-based web application developed for retrieving the price of agricultural commodities and weather related information in Assamese language. The price of agricultural commodities are retrieved from AGMARKNET website while the weather related informations are extracted from the IMD website. Both these websites are updated on a daily basis by the Government of India. The back-end of the application consists of automatic speech recognition (ASR) modules developed using state-of-the-art acoustic modeling approaches. Word error rates (WERs) of 7.79% and 4.98% are achieved for commodity and district names respectively.

Index Terms: speech application, Assamese language, ASR

1. Introduction

Voice query mediated agriculture commodity price and weather information systems are of utmost importance to farmers in an agrarian economy based country like India. Hence, in the recent past attempts were made to implement such systems through toll free phone numbers, where, the information is provided to a user in his or her native language [1]. However, in the last few years, increase in the use of smart phones, tabs and computers by the common people of India demanded that some of such voice query based systems are available for public dissemination. At the same time with the emphasis on e-governance and digital India initiatives, information kiosks are also set up in public areas. Considering these developments, there is a need to adapt the systems, originally designed for over the phone line applications, to web based applications. At the same time, such adaptation should not incur any new expenses in terms of new system building.

In this work we demonstrate a web based interface developed for a voice query based information retrieval system in Assamese, built on a system originally for use over phone lines. We demonstrate that the voice query systems intended for over the phone usage can be seamlessly integrated into a web based platform. While this system is integrated into a voice over internet protocol system, potential issues related to channel mismatch, unfamiliar noise profile etc. are handled effectively and they bear no negative consequences to system performance.

The rest of the work is organized as follows. Section 2 reports the experimental setup and results, Section 3 details the application design and Section 4 concludes the work.

2. Experimental Setup and Results

This section discusses the experimental setup used for developing the back-end ASR systems for recognizing the names of agricultural commodities and districts of Assam. All the experimental evaluations are performed using Kaldi speech recognition toolkit [2].

The speech corpus used in this work is collected over mobile channel from native Assamese speakers in actual field condition. In order to capture the dialectal variations, the speech data is collected from different districts of Assam. The data collection activity can be segmented into 3 different components. The first component involves collection of 30 hours of isolated commodity and district names from 885 speakers [1]. The vocabulary size of commodity names is 109 while that of district names is 27. The second component involves collection of 3 hours of phonetically balanced sentences spoken by 25 speakers [3]. The third component involves collection of additional 5658 speech files spoken in continuous manner from 27 speakers [4]. Since the speech files are recorded in actual field conditions, some interfering background noise is also recorded in addition to the desired foreground speech. In order to separate the foreground speech from the background noise, we have employed the zero frequency filtered signal (ZFFS) based technique as reported in [5]. The foreground segmented speech signal is analyzed using a Hamming window of 25 ms frame-size and 10 ms frame-shift with a pre-emphasis factor of 0.97. Using 21-channel filter bank, 13 dimensional Mel frequency ceptral coefficients (MFCC) base features ($C0$-$C12$) are computed. In order to capture the dynamic characteristics of the speech signal, the static MFCC feature vectors are spliced in time with 4 frames to the right and 4 frames to the left of the central frame thereby making the feature vector dimension of 117 (13×9). Using linear discriminant analysis (LDA), the 117 dimensional feature vector is reduced to 40 which are further decorrelated using maximum linear likelihood linear transformations (MLLT).

The current work explores two different acoustic modeling techniques. The first technique involves the conventional context dependent Gaussian mixture model (GMM) based system while the second one involves Karel's implementation of deep neural network (DNN). In order to increase the flexibility of the ASR system, we have employed the task specific modified language model as reported in our previous work [1]. The developed models are evaluated on two test sets : commodity test set and district test set. The commodity test set comprises of 2552 utterances of isolated commodity names collected from

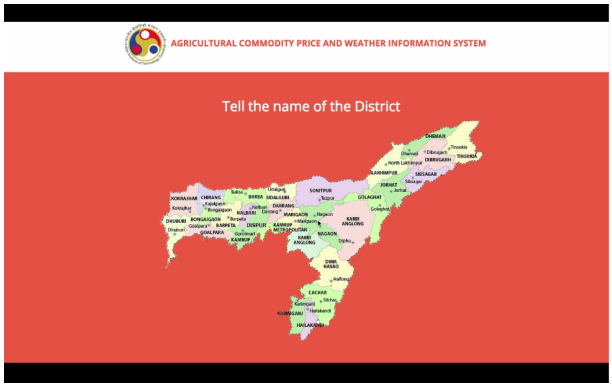

Figure 1: *District name recognition*

Figure 2: *Commodity name recognition*

275 speakers while the district test set consists of 1125 utterances of isolated district names. The results obtained in the experimental evaluations are shown in Table 1. It is observed that relative improvements of 24.44% and 3.48% in WER are obtained in commodity and district name recognition respectively in the DNN based system as compared to the GMM based system.

Table 1: *% WER for the two acoustic modeling approaches employed in the development of the ASR system*

Acoustic Modeling Technique	% WER	
	Commodity	District
GMM-HMM	10.31	5.16
DNN-HMM	7.79	4.98

3. Application Design

The web based speech application is designed to retrieve the price of 109 agricultural commodity names and 27 districts of Assam. On loading the URL of the application, a welcome prompt greets the user. The system first prompts the user to tell the name of the desired district. On successful recognition of the district name, the system prompts the user to make the decision for commodity price information or weather related information. If the user wishes to know the price of agricultural commodities, the system prompts the user to tell the name of the desired commodity name. On successful recognition of the commodity name, the system checks the district-commodity combination in the database and displays the available modal price of the desired commodity in the desired district. On the other hand, if the user wishes to know weather related information, the system checks for current day and next 4 days weather related information in the database and displays the available information for the desired district. The price information of the agricultural commodities are crawled from AGMARKNET website while the weather related information are crawled from the IMD website. Snapshots of district name recognition and commodity name recognitions are shown in Figure 1 and Figure 2 respectively.

4. Conclusion

This work demonstrates that the existing speech recognition systems for voice query based information retrieval, designed to be used over phone lines, can be seamlessly integrated into a web interface. Even the same system can be simultaneously used by a phone line and a web interface with affecting its performance. The demonstrated web based systems can be implemented on public kiosks for users to retrieve information in their native languages.

5. Acknowledgement

This work is part of the ongoing consortium project on "Speech-based Access of Agricultural Commodity Prices and Weather Information in 11 Indian Languages / Dialects", funded by the Technology Development for Indian Languages (TDIL) programme initiated by the Department of Electronics & Information Technology (DeitY), Ministry of Communication & Information Technology (MCIT), Govt. of India. The authors would like to thank the consortium leader Prof. S. Umesh and other consortium members for their valuable inputs and suggestions.

Part of this work is also supported by another consortium project titled "Development of Speech Interface for Form-filling application (SiFA) in Five Indian languages", generously funded by the Ministry of Electronics and Information Technology and the Ministry of Human Resources Development, Government of India.

6. References

[1] A. Dey, S. Shahnawazuddin, D. K. T., S. Imani, S. R. M. Prasanna, and R. Sinha, "Enhancements in assamese spoken query system: Enabling background noise suppression and flexible queries," in *2016 Twenty Second National Conference on Communication (NCC)*, March 2016, pp. 1–6.

[2] D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlicek, Y. Qian, P. Schwarz, J. Silovsky, G. Stemmer, and K. Vesely, "The Kaldi Speech Recognition Toolkit," in *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding*. IEEE Signal Processing Society, Dec. 2011.

[3] S. Shahnawazuddin, K. Deepak, B. Sarma, A. Deka, S. R. M. Prasanna, and R. Sinha, "Low complexity on-line adaptation techniques in context of assamese spoken query system," vol. 81, 05 2014.

[4] B. Deka, J. Chakraborty, A. Dey, S. Nath, P. Sarmah, S. Nirmala, and S. Vijaya, "Speech corpora of under resourced languages of north-east india," in *Oriental COCOSDA*, May 2018.

[5] K. T. Deepak and S. R. M. Prasanna, "Foreground speech segmentation and enhancement using glottal closure instants and mel cepstral coefficients," *IEEE/ACM Trans. Audio, Speech and Lang. Proc.*, vol. 24, no. 7, pp. 1204–1218, Jul. 2016.