A Speech-Based Conversation System for Accessing Agriculture Commodity Prices in Indian Languages

Gautam Varma Mantena, S. Rajendran, B. Rambabu, Suryakanth V. Gangashetty,
B. Yegnanarayana, Kishore Prahallad

International Institute of Information Technology, Hyderabad, India

`gautam.mantena@research.iiit.ac.in, su.rajendran,rambabu03@gmail.com, {svg,yegna,kishore}@iiit.ac.in`

1. Introduction

Human-computer interaction plays a significant role for literate/illiterate and visually challenged users to access information. The mode of human-computer interaction could be speech, text, gestures, facial expressions, symbols etc., or a combination of these. An interaction could be in the form of a conversation including statements, questions, answers and expressions. A (uni-modal or multi-modal) conversational type of human-computer interaction is often referred to as a conversational system.

A conversational system with speech as an input mode assumes significance as speech is a natural means of communication for human-beings. The goal of a speech-based conversation (SBC) system is to provide information by conversing with a human-being in a natural fashion. Our objective is to develop speech based conversational systems for information access in an Indian language.

As shown in Fig. 1, a simplistic view of a speech-based conversational system consists of: automatic speech recognition (ASR) which converts speech to text, natural language understanding (NLU), dialog manager (DM), natural language generation (NLG) and text-to-speech (TTS). When a user utters a query to a SBC system, the speech is converted to text by ASR and this text is parsed by NLU module to extract the relevant information or concepts. DM is a core component of the SBC system. It determines what are the necessary actions to be performed and response to be given to the user. The required response is provided to the user via NLG in generating the appropriate sentences which are then synthesized by a TTS module. In this work, a preliminary version of speech-based conversational system is demonstrated for accessing price of agricultural commodities by farmers in India. We refer to this conversational system as Mandi information system.

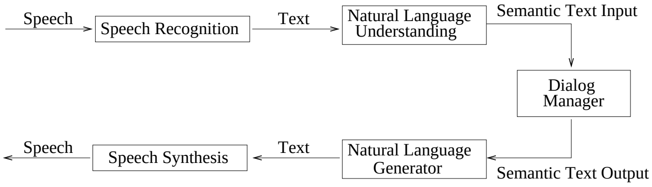

Figure 1: *Architecture of a Speech Based Conversation System.*

2. Mandi Information System (MIS)

Mandi information system (MIS) is built for farmers in rural and semi-urban areas to obtain price of commodities (vegetables, fruits, pulses, spices) that are being sold in the markets across state of Andhra Pradesh in India. MIS is a telephone/mobile based conversation system, as these are the most commonly available communication services. MIS provides price information in Telugu language. Table 1 gives the vocabulary size for Mandi information system. The price information is obtained from `http://agmarknet.nic.in/` on a daily basis and is provided by Ministry of Agriculture, Government of India. A few major issues in the development of the Mandi information system are the following:

- *Noisy environment:* The target audience of MIS are farmers in rural and semi-urban areas. The farmers call MIS through their mobile phones or landline. The quality of speech signal is affected by distance of microphone, mobile/telephone handsets, speech codecs and communication channel. The environment in which a call to MIS is made could also be noisy including vehicle/fan/background noise and background speech.

Table 1: *Vocabulary size used in Mandi Information System.*

Word Category	Vocabulary Size
Commodity	72
Markets	348
Districts	23

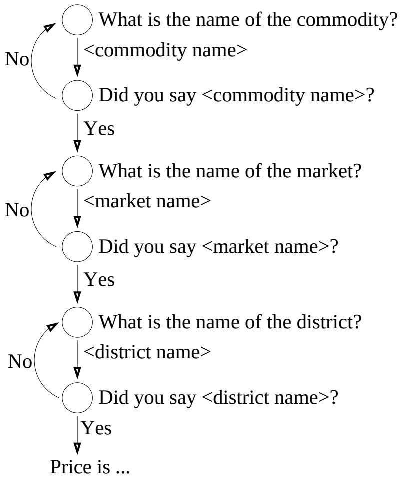
What is the name of the commodity?
<commodity name>
No
Did you say <commodity name>?
Yes
What is the name of the market?
<market name>
No
Did you say <market name>?
Yes
What is the name of the district?
<district name>
No
Did you say <district name>?
Yes
Price is ...

Figure 2: *State diagram for Mandi Information System.*

- *Dialect/Pronunciation variation:* Though the MIS is targeted for a particular language such as Telugu, the dialectal variations are large in India. It is often observed that the dialectal variations in Indian languages are hard to be quantized into a specific number. These variations tend to be a continuum in the linguistic space. A farmer could use a different name or a pronunciation variation for an agricultural commodity or converse/query the MIS in a casual style.

- *Unstructured conversation:* The target audience of the MIS may not have interacted with a computer based information access system. Hence, the conversation is typically unstructured and will be filled with disfluencies including repeats and false starts. Hence, it is a challenge to provide information for such users. Another relevant issue is eliciting the speech data from the farmers in order to capture the acoustic and pronunciation variations for building an ASR.

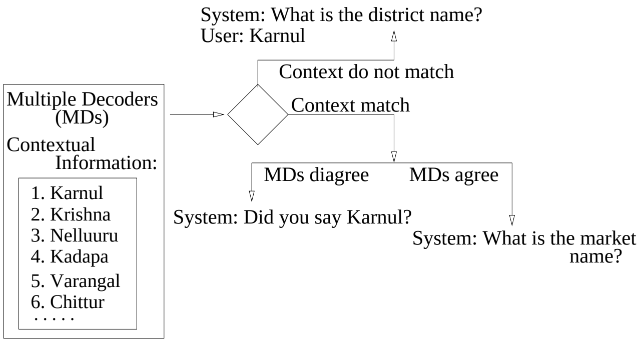

Figure 3: *Use of Multiple Decoders and Contextual Information in determining the dialog flow.*

3. Current status of the MIS

Keeping in view of the multiple challenges involved, a prototype of the MIS is built in Telugu language. Speech data was initially collected via telephone medium (digital line), wherein users were requested to call from a mobile phone. Users were given the list of words of commodity, market and district names to be uttered for recording. For the commodity names, users were given only pictures and were asked to say the name of that commodity in (their native) Telugu language. Goal for such an exercise is to capture different pronunciation variations and also to cover the commodity names in various regions. Data was collected for 96 speakers consisting of 17 hours of data. A total of 500 words were collected from each of the speakers.

Approximately 15 hours of recorded speech data was taken and used to building the acoustic models of ASR. These are context dependent tri-phone HMM models built with 8 Gaussian mixtures per state using Sphinx recognition system [1].With appropriate NLU, DM, NLG and TTS, a baseline version of MIS (MIS_V1) is built. MIS requires three concepts or inputs from the user, which are commodity, district and market names. A typical dialog state diagram is as shown in Fig. 2. When a user provides some information to MIS, the system would ask for an explicit confirmation. The user is required to respond either yes/no. This is for MIS to make sure that the input query is right as recognition is error prone.

An ideal SBC system provides accurate information to a user in less number of turns (or interactions). Speech recognition being error prone, confirmations from users cannot be avoided. But the goal would be how to limit those confirmations. An approach would be to associate a confidence score to the recognition output of an ASR using confidence scoring techniques like normalized likelihood scores, counts from N-best hypothesis, language model scores, parsing related etc [2]. System can also use features from various levels of dialogue system like decoding, parsing and dialogue features [3] or using semantic and pragmatic features on the N-best list to measure the confidence score [4, 5, 6].

We have incorporated a confidence measure into the system and the modifications done to the MIS_V1 is as shown in Fig. 3, where we use multiple decoders (MDs) and contextual information to determine the dialog flow. Table 2 shows a sample conversation recorded by MIS.

Experiments were conducted with base line system (MIS_V1) and with improved MIS (MDS and contextual information included and is called MIS_V2). It was observed that using MIS_V2, the users were able to retrieve the commodity price accurately in 31 out of 40 trials. While using MIS_V1, the users were able to retrieve the commodity price accurately only in 23 out of 40 trials. At the workshop, a live demonstration of MIS_V2 will be given.

4. Conclusions

We demonstrate a speech based conversation system under development for information access by farmers in rural and semi-urban areas of India. The challenges are that the system should take care of the significant variations in the pronunciation and also the highly natural and hence unstructured dialog in the usage of the system. The focus of this study is to develop a conversational system which is adaptable to the users over a period of time, in the sense that fewer interactions with the system to get the required information. Some other novel features of the system include multiple decoding schemes and accountability of the wide variations in dialog, pronunciation and environment. A video demonstrating the Mandi information system is available at http://speech.iiit.ac.in/index.php/demos.html

Table 2: *Recorded MIS conversation with a user*

System:	What is the commodity name?
User:	*Red Grams*
(Recognition Model 1,3:)	*Red Grams*
System:	*Red Grams* is being sold in 9 districts.
System:	What is the name of the district?
User:	*Karnul*
(Contextual Information:	*Karnul, Krishna, Nelluru, ...*)
(Recognition Model 1,3:)	*Karnul*
System:	Dal is sold in 6 markets in *Karnul*
System:	What is the market name?
User:	*Dhoni*
(Contextual Information:	*Adoni, Karnul, Dhoni, ...*)
(Recognition Model 1:)	*Dhoni*
(Recognition Model 3:)	*Guti*
System:	Did you say *Dhoni*?
User:	Yes

5. References

[1] "CMU Sphinx, The Carnegie Mellon Sphinx Project," http://cmusphinx.sourceforge.net.

[2] H. Jiang, "Confidence measures for speech recognition: A survey," *Speech Communication*, vol. 45, no. 4, pp. 455–470, Apr. 2005.

[3] Paul Carpenter, Chun Jin, Daniel Wilson, Rong Zhang, Dan Bohus, and Alexander I. Rudnicky, "Is this conversation on track," in *Proc. Eurospeech*, 2001.

[4] Joseph Polifroni Timothy J. Hazen, Theresa Burianek and Stephanie Seneff, "Integrating recognition confidence scoring with language understanding and dialogue modeling," in *Proc. ICSLP*, 2000, p. 2000.

[5] Malte Gabsdil and Oliver Lemon, "Combining acoustic and pragmatic features to predict recognition performance in spoken dialogue systems," in *Proc. ACL*, 2004, pp. 344–351.

[6] B. Thomson, K. Yu, M. Gasic, S. Keizer, F. Mairesse, J. Schatzmann, and S. Young, "Evaluating semantic-level confidence scores with multiple hypotheses," in *Proc. Interspeech*, 2008.