

p-ISSN: 2348-6848 e-ISSN: 2348-795X Volume 03 Issue 11 July 2016

A Complete Voice Over Control of Mobile in Native Language for Blind

K.Soumya¹,Ambica Venna², B.Manjulatha³ ¹Assistant Professor in IT Dept.,VBIT,Telangana,India. *kothapallisowmya@gmail.com* ²Assistant Professor in IT Dept.,VBIT,Telangana,India. *venna.ambica@gmail.com* ³Assistant Professor in CSE Dept.,VBIT,Telangana,India. *manjulathareddy86@gmail.com*

Abstract : Visually impaired individuals are a growing segment of our population. However, social constructs are not always designed with this group of people in mind, making the development of better electronic accessibility tools essential in order to fulfill their daily needs. Technological improvements have given more independence to the visually impaired population. The past few years have witnessed an exponential growth in the computing capabilities and onboard sensing capabilities of mobile phones making them an ideal candidate for building powerful and diverse applications. Many mobile assistive applications have come out in this regard based on voice over i.e where the blind people can hear the voice output for navigating through the menu's or options in the mobile with the help of tethering and also can give some voice input for some applications like email or SMS. But most of these assistive applications are based on international language that is English or in Indian English. So for a person with visual impairment and who have a language barrier too is difficult to hear and use the mobile with voice over in English. In this paper we propose a solution to break this barrier by controlling the mobile through voice input in telugu and as a response it should produce the voice output also in telugu language.

Keywords: Voice translation, automatic speech recognition, native language, visually impaired.

1. INTRODUCTION

Language has always been the basis of any form of written or speech communication and for a person with vision impairment it is must to communicate with them in speech (in their native language). There are various applications in mobiles to assist the persons with visual impairment like in reading text, moving around in the traffic or remote places too, identifying the color, object etc. All these applications will produce voice over in international language that a blind may or may not be able to understand if he had a language barrier. So In this paper we use some "acoustic models" to convert telugu voice input to system understandable text commands operate the mobile and with TTS using HTK methods we convert the generated output into native voice again for the blind.



2. SYSTEM MODEL



Figure 1

As shown in the system model the major component kaldi performs the required task in six sequential phases as discussed in next steps.

3. Kaldi Speech Recognition Toolkit

Kaldil is an open-source toolkit for speech recognition written in C++ and licensed under the Apache License v2.0.The goal of Kaldi is to have modern and flexible code that is easy to understand, modify and extend. Kaldi is available on SourceForge (see http://kaldi.sf.net/). The tools compile on the commonly used Unix-like systems and on Microsoft Windows.

3.1 Overview of toolkit

We give a schematic overview of the Kaldi toolkit in figure 1. The toolkit depends on two external libraries that are also freely available: one is OpenFst for the finite-state framework, and the other is numerical algebra libraries. We use the standard "Basic Linear Algebra Subroutines" (BLAS) and "Linear Algebra PACKage" (LAPACK) routines for the latter. The library modules can be grouped into



two distinct halves, each depending on only one of the external libraries (c.f. Figure 1). A single module, the Decodable Interface bridges these two halves. Access to the library functionalities is provided through command-line tools written in C++, which are then called from a scripting language for building and running a speech recognizer. Each tool has very specific functionality with a small set of command line arguments: for example, there are separate executables for accumulating statistics, summing accumulators, and updating a GMM-based acoustic model using maximum likelihood estimation.



Figure 2

A simplified view of the different components of Kaldi. The library modules can be grouped into those that depend on linear algebra libraries and those that depend on OpenFst. The decodable class bridges these two halves. Modules that are lower down in the schematic depend on one or more modules that are higher up.

Moreover, all the tools can read from and write to pipes which make it easy to chain together different tools. To avoid "code rot", we have tried to structure the toolkit in such a way that implementing a new feature will generally involve adding new code and command-line tools rather than modifying existing ones.

3.2 Voice control system architecture

The Voice control system has a speech to speech user interface consisting of six major components.

- 1. Voice Activity Detection (VAD)
- 2. Automatic Speech Recognition (ASR)
- 3. Spoken Language Understanding (SLU)
- 4. Command Language Generation (CLG)

5. Command Manager (CM)

6. Text to Speech (TTS)

The system interacts with a user in turns. The schema in Figure 3 illustrates how the user's input is processed in single turn. The spoken input is passed to ASR component which generates corresponding textual representation. SLU extracts semantic meaning from the text and The CLG component generates textual form of a command representation and CM decides which response to present and finally the TTS read the text with human voice. Each of the component runs in separate process in order parallelize the input data processing and output data generation. The components communicate among themselves through systempipes.

3.2.1. Voice Activity Detection (VAD)



The user (blind) is freely allowed to speak in his/her native language without the knowledge about processing of their input inside. The voice input then detected for actual activity which is initiated by the user



Figure 3 Single turn in Voice control system

3.2.2. Automatic Speech Recognition (ASR)

The goal of statistical ASR is to decode the most likely word sequence given speech. In speech recognition it is equivalent to recognizing the word sequence from the speech. Formally, we search for the most probable sequence of words w^* given the acoustic observations as described in following equation. The best word sequence w^* does not depend on probability of the acoustic features P (a) so it can be eliminated as shown on the second row of the equation.

$$w^* = \operatorname{argmax}_w \{P (w \mid a)\} = \operatorname{argmax}_w \{$$

The task of acoustic modeling is to estimate the parameters θ of a model so the probability P (a | w; θ) is as accurate as possible. Similarly, the LM represents the probability P (w).

3.2.3. Spoken Language Understanding (SLU)

SLU extracts semantic meaning from the text recognized by ASR. Which means understanding the actual action needed by the user?

3.2.4. Command Language Generation (CLG)

The CLG component generates textual form of a command representation.

3.2.5 & 3.2.6. Command Manager (CM) & Text to Speech (TTS)

 $P (aM_w \text{decides})$ which response to present and finally the TTS reads the text with human voice in native language which a blind could easily understand.

P (a)

4. APPLICATIONS

The most common application of voice translation is in the field of telephonic communication. By eliminating the dependence on languages, real-time communication can be performed without any common language backbone. And for the people with visual impairment it is more helpful. Other domains where voice translation can be used are education, entertainment, call center services and broadcast channels.



5. CONCLUSIONS

Spoken dialogue is the most intuitive form of communication among any kind of people especially for visually challenged will rely completely on language. Language has always been a barrier to effective communication. While considerable progress has been done in this direction, more efforts need to be taken in order to make the latest technologies more useful for the persons with disabilities. With this paper, we propose a new system model to ensure effective real- time communication between a blind user who have a visual impairment and language barrier too.

REFERENCES

[1]. Master thesis by Ondrej Platek on KALDI in Charles University in Prague in 2014.

[2]. S.Young, G.Evermann, M.Gales, T.Hain,D. Kershaw, X.Liu,G.Moore, J.Odell,D.Ollason, Povey,V. Valtchev, and P. Woodland, The HTK Book (for version 3.4). Cambridge University Engineering Department, 2009.

[3]. A. Lee, T. Kawahara, and K. Shikano, "Julius – an open source realtime

large vocabulary recognition engine," in EUROSPEECH, 2001, pp.1691–1694.

[4]. The Kaldi Speech Recognition Toolkit, Povey, Daniel, Ghoshal, Arnab, Boulianne, Gilles, Burget,

Lukas, Glembek, Ondrej, Goel, Nagendra, Hannemann, Mirko, Motlicek, Petr<u>Qian</u>, Yanmin,

Schwarz, Petr, Silovsky, Jan, Stemmer, Georg and Vesely, Karel, Idiap-RR-04-2012.

[5]. "A complete Kaldi recipe for building Arabic speech recognition systems" by Ahmed Ali,

Yifan Zhang, Patrick Cardinal, Najim Dahak in 2014.