HUMAN SPEECH RECOGNITION FOR SYNCHRONOUS IDE

Abhijeet Panhale[1], Prof. Monali Deshmukh[2], Ajinkyaraj Kadam [3], Pramod Patil [4], Chinmay Shelar[5]
Department of Information Technology
[1,2,3,4,5] JSPM's Rajarshi Shahu College Of Engineering

Abstract- *Speech Recognition is the process of automatically recognizing a certain word spoken by a particular speaker based on individual information included in speech waves. This technique makes it possible to use the speaker's voice to verify his/her identity and provide controlled access to services like voice based biometrics, database access services, voice based dialing, voice mail and remote access to computers. Voice can be a powerful tool for use in human computer interaction because it is the fundamental means of human communication. With the rapid growth of wireless communications, the need for voice recognition techniques has increased greatly. Portability and wear ability, which are necessary items for being computationally powerful computer devices, will be reinforced by attaching voice applications, since voice can support invisible communication with a computer device as a natural way of communicating.*

Signal processing front end for extracting the feature set is an important stage in any speech recognition system. The optimum feature set is still not yet decided though the vast efforts of researchers. There are many types of features, which are derived differently and have good impact on the recognition rate. This project presents one of the techniques to extract the feature set from a speech signal, which can be used in speech recognition systems.

Keywords- Multi-label classification, Speaker recognition, Speaker traits and states, Voice casting, Voice similarity. Voice applications, Voice-based interfaces

I. INTRODUCTION

With advances in new technologies, computer devices have grown in popularity to become one of the most common consumer devices. Even as these devices are shrinking in size, however, their capability and content are changing into more complex and diverse functionalities to meet user requests. Now, it is common for many computer devices to include a phone, a personal directory, a memo capability, an alarm clock, a scheduler, a camera, games, and several applications which were working in Personal Digital Assistants (PDAs) before, so there is no boundary between computer devices and PDAs. However computer devices in which designers have worked more and more to decrease their size are likely to have small keypads and screens, whereas they should have more complex and diverse functions for users. Their functionality and ease- of use are greatly limited, and thus many researchers look to find alternative communication channels when interacting with these devices. In recent years, many researchers in the area of human computer interaction (HCI) have attempted to enhance the effectiveness and efficiency with which work and other activities are performed using voice based interfaces. Even if voice technology has been explored for use in desktop com-putter and telephone information system, the role of voice in interfaces has received little attention because of its difficulty of use and tiresomeness of recognition. Actually, in the past, the accuracy of voice recognition was unacceptably low, and it's a role in a system was questionable because of ambiguity and error. However, voice technology has reached the point of commercial viability and reliability now, and also many computer devices adapt voice applications for providing better services to users. Using voice allows the interface size to be scaled down because voice interaction requires only audio I/O devices such as a microphone and speaker, which are already quite small and inexpensive. Currently, in a computer device voice interfaces need only small space and power consumption, but are able to provide every user with a friendly interface by adding a feeling of natural interaction. Voice interfaces are sufficient to replace graphical user interfaces for accessing all information and content without using keyboards, buttons, and touch screens, since voice is the fundamental means of human communication.

II. LITERATURE REVIEW

Voice technology has been explored for use in desktop computers and telephone information systems, so that multiple studies have been focused on voice recognition systems or applications for general computer systems. Rona Rosenfeld and others considered how to build voice application interfaces, especially achieving reliable and accurate speech recognition, and presented their thoughts about the future of speech-based interaction with at least three fundamental advantages for speech:

Speech is an ambient medium rather than an intentional one. Visual activity requires our focused attention while speech allows us to do something else.

Speech is descriptive rather than referential. When we speak we describe objects in terms of their roles and attributes. In visual situations we point to or grasp the objects of interest. For this reason, speech and pointing are to a large extent complementary, and can often be combined to great effect.

Speech requires more modest physical resources. Speech- based interaction can be scaled down to much smaller and much cheaper form-factors than visual or manual modalities.

III. EXISTING SYSTEM

A voice application based on voice interfaces is also useful as a form of input especially when someone's hands, eyes, or

Ear use computer devices. Voice interfaces and voice recognition technology allow people working in active environments to use them without any holding or touching devices. According to Rick Beasley and others, voice user interfaces (VUIs) are a new concept to many who now have the task of doing everything it takes to develop a voice XML application. Major differences between VUIs Other studies have focused on voice applications especially a speech interface for handheld devices that allow user to capture and randomly access voice notes, which are segments of digitized speech containing thoughts and ideas. To improve user friendliness and dialogue success rate for third generation computer communication systems, multimodal interfaces based on speech displayed a platform which has restricted the functionality to speech centric multimodal interfaces with two input modes: speech and touch, and two output modes: audio and vision. A voice assisted simulation animation architecture (Asarco) described voice assist technology by providing software architecture integrating with speech input and output. It supports simulation- animation environments by providing input through spoken commands, mouse manipulation, and keyboard entry so that it can provide more user friendliness to end users.

IV. PROPOSED SYSTEM

According to B. Tognazzini, voice can be used three ways: to command the computer, to enter information, and to communicate with other people. In this part, we discuss the general components of building a voice application. As seen in Figure 1, a fundamental voice application consists of four basic parts: end user, front-end interfaces, voice recognition system, and dictionary-and-text file database. Each component is explained as follows:

End Users:-

Generally end users mean device users. They can use devices to communicate and make voice feedback with the application, and especially end users are the users who currently use computer devices.

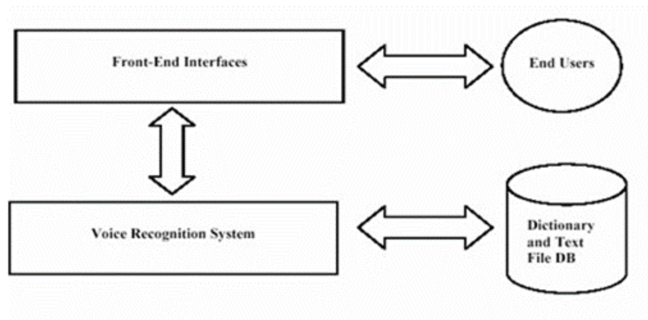

Fig. 1. General Voice Application

V. ALGORITHM INTERVAL SEARCH ALGORITHM

We now introduce our new Interval Search (ISA) algorithm for solving the interval set intersection problem. This algorithm uses two binary searches to identify interval intersections

While avoiding the aforementioned complexities caused by contained intervals. The key observation underlying ISA is that the size of the intersection between two sets can be determined without enumerating each intersection. For each interval in the query set, two binary searches are performed to determine the number of intervals in the database that intersect the query interval. Each pair of searches is independent of all others, and thus all searches can be performed in parallel.

Existing methods define the intersection set based on inclusion, that is, the set of intervals in the interval database B that end after the query interval ai begins, and which begin before ai ends. However, we have seen that contained intervals make it difficult to find this set directly with a single binary search. Our algorithm uses a different, but equivalent, definition of interval intersection based on exclusion, that is, by identifying the set of intervals in B that cannot intersect ai , we can infer how many intervals must intersect ai . Formally, we define the set of intervals that intersect query interval to be the intervals in B that are neither in the set of intervals ending before ('left of', set below) ai begins nor in the set of intervals starting after ('right of', set below) ai ends. That is:

Algorithm 1 : Single interval intersection counter

Input : Sorted interval starts and ends BS and BE, query interval a

Output : Number of intervals c intersecting a

```
Function ICOUNT(BS,BE,a)begin
    first- BINARYSEARCH(BS,a.end)
    last- BINARYSEARCH(BS,a.start)
    c- first-last   /*=|B | - (last+(|B|-fisrt)) */
    return c
```

VI. SYSTEM OVERVIEW

The voice recognition program installed on the computer interprets the user's voice commands and translates them into commands such as "my computer", "opera" or "Notepad ". A microphone system transmits the voice commands from the user to the computer sound input.

Our proposed system can serve as a general GUI For voice signal management. In terms of the scope of this proposal, the project will have as its target the U.S. English language model. The voice interface is to feature a speaker independent speech recognition engine backed by a speaker dependent like functionality. The project will be a collaborative development effort between user and computer system, and Mic system will provide a voice engine, knowledge transfer, and Palm-specific API command set suitable for the device, and Mic will start the signal accepting process and send toward computer OS and creating GUI for the application of project. As shown in Figure 3, the application consists of three parts: interfaces, voice engine, and data storage.

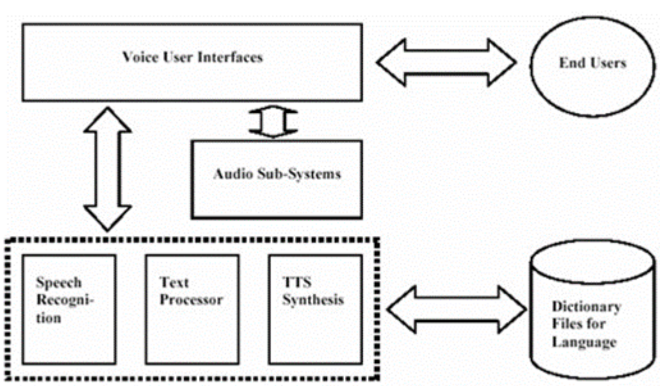

Fig. 2. Proposed System Functional Design Diagram

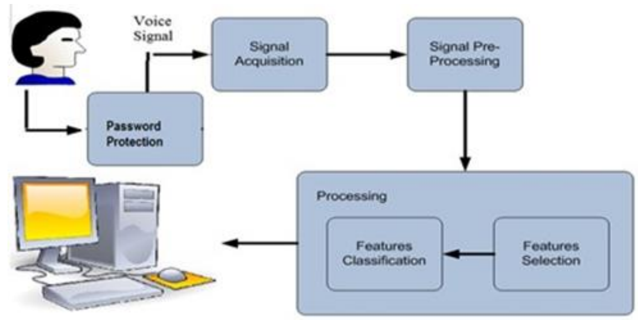

Fig. 3. System Architecture

VII. MODULE AUTHENTICATION MODULE

System authenticates the user by entering the correct pass- word. By chance if the password is wrong then the user has to contact the software developer. The developer reset the system and gives the new password to user. The graphical user interface contains a login tab with enter password field, login tab, change password.

Database Design Module:-

1. A database is a collection of information that is organized so that it can easily be accessed, managed, and updated. In one view, databases can be classified according to types of content: bibliographic, full text, numeric, and images.

2. Database design is the process of producing a detailed data model of a database. This data model contains all the needed logical and physical design choices and physical storage parameters needed to generate a design in a data definition language, which can then be used to create.

VIII. TECHNOLOGY USED

[1] JAVA:

Java is a programming language expressly designed for use in the distributed environment of the Internet. Java is used to create complete applications that may run on a single computer or be distributed among servers and clients in a network. It can also be used to build a small application module or applet for use as part of a Web page. Applets make it possible for a Web page user to interact with the page. It is used to build

Application module for use as a part of Web page and provides interaction between the pages.

[2] HTTP:

Hypertext Transfer Protocol is a transaction or oriented client/server protocol between web browser and a Web Server. It provides communication between client and server.

[3] Servlet :

A servlet is a small program that runs on a server. The term was coined in the context of the Java applet, a small program that is sent as a separate along with a Web (HTML) page. Java applets, usually intended for running on a client, can result in such services as performing a calculation for a user or positioning an image based on user interaction. It is used to extend capabilities of server.

[4] JSP :

The web server needs a JSP engine, i.e., a container to process JSP pages. The JSP container is responsible for intercepting requests for JSP pages. This tutorial makes use of Apache which has built-in JSP container to support JSP pages development.

A JSP container works with the Web server to provide the runtime environment and other services a JSP needs. It knows how to understand the special elements that are part of JSPs.

[5] HTML:

HTML (Hypertext Markup Language) is the set of markup symbols or codes inserted in a file intended for display on a World Wide Web browser page. The markup tells the Web browser how to display a Web page's words and images for the user. Each individual markup code is referred to as an element (but many people also refer to it as a tag). Some elements come in pairs that indicate when some display effect is to begin and when it is to end. It is used to create static web page.

[6] CSS 3:

A cascading style sheet (CSS) is a Web page derived from multiple sources with a defined order of precedence where the definitions of any style element convict. The Cascading Style Sheet, level 1 (CSS1) recommendation from the World Wide Web Consortium (W3C), which is implemented in the latest versions of the Netscape and Microsoft Web browsers, specifies the possible style sheets or statements that may determine how a given element is presented in a Web page.

It describes the presentation of Web pages, including colors, layout, and fonts.

[7] MySQL:

MySQL is a freely available open source Relational Database Management System (RDBMS) that uses Structured Query Language (SQL). SQL is the most popular language for adding, accessing and managing content in a database. It is most noted for its quick processing, proven reliability, ease and exibility of use.

IX. CONCLUSION

Speech Recognition is the process of automatically recognizing a certain word spoken by a particular speaker based on individual information included in speech waves. This technique makes it possible to use the speaker's voice to verify

His/her identity and provide controlled access to services like voice based biometrics, computers. Voice can be a powerful tool for use in human computer interaction because it is the fundamental means of human communication. With the rapid growth of wireless communications, the need for voice recognition techniques has increased greatly. Portability and wear ability, which are necessary items for being computationally powerful computer devices, will be reinforced by attaching voice applications, since voice can support invisible communication with a computer device as a natural way of communicating.

Signal processing front end for extracting the feature set is an important stage in any speech recognition system. The optimum feature set is still not yet decided though the vast efforts of researchers. There are many types of features, which are derived differently and have good impact on the recognition rate. This project presents one of the techniques to extract the feature set from a speech signal, which can be used in speech recognition systems.

For fast program execution and construction process. In Existing system there is a problem of time consumption. To overcome this problem we use signals and system techniques. And we also use voice commands.

REFERENCES

[1] Zhehuai Chen,Yimeng Zhuang, YanminQian, And Kai Yu, "Phone Synchronous Speech Recognition With CTC Lattices "IEEE/Acm Transactions On Audio, Speech, And Language Processing, Vol. 25, No. 1, January 2017.

[2] Nicolas Obin,Axel Roebel, "Similarity Search Of Acted Voices For Automatic Voice Casting"IEEE/Acm

Transactions On Audio, Speech, And Language Processing, Vol. 24, No. 9, September 2016.
[3] H.Kido and H. Kasuya, "Everyday expressions associated with voice quality of normal utteranceExtraction by perceptual evaluation ,"J. Acoust. Soc. Jpn., vol. 57, no. 5, pp. 337344, 2001.
[4] F. Nolan, P. French, K. McDougall, L. Stevens, and T. Hudson, "The role of voice quality settings in perceived voice similarity ,"presented at the International Association Forensic Phonetics Acoustics, Vienna, Austria, 2011.
[5] Z. Karam, W. M. Campbell, and N. Dehak, " Graph relational features forspeaker recognition and mining,"in Proc. IEEE Statist. Signal Process. Workshop, 2011, pp. 525528.
[6] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker verification using adapted Gaussian mixture models,"Digital Signal Process., vol. 10, nos. 13, pp. 1941, 2000.
[7] W. M. Campbell, D. E. Sturim, and D. A. Reynolds, "Support vector machines using GMMsupervectors for speaker verification,"IEEE Signal Process. Lett., vol. 13, no. 5, pp. 308311, May 2006.
[8] N. Dehak, P. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet, "Front-end factor analysis for speaker verifica- tion,"IEEE Trans. Audio, Speech, Lang. Process., vol. 19, no. 4, pp. 788798, May 2011.
[9] B. Schuller et al., " The INTERSPEECH 2012 speaker trait challenge,"presented at the Interspeech, Portland, Oregon, USA, 2012.
[10] N. Obin, A. Roebel, and G. Bachman, "On auto- matic voice casting for expressive speech: Speaker recog- nition vs. speech classification ,"presented at the Interna- tional Conf.Acoustics, Speech, Signal Processing., Florence, Italy,2014.
[11] X. Huang, A. Acero, and H.-W. Hon, "Spoken Language Processing: A Guide to Theory, Algorithm, and System Devel- opment ", 1st ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2001.
[12] F. Jelinek, "Statistical Methods for Speech Recognition". Cambridge, MA,USA: MIT Press, 1997.
[13] T. Hori and A. Nakamura,"Speech recognition al- gorithms using weighted finite-state transducers", Synthesis Lectures Speech Audio Process., vol. 9,no. 1, pp. 1162, 2013.