Voice Control Mode of Operation Used on Mobile Devices

Cai-Yi Jheng, Chung-Hua Chu

National Taichung University of Science and Technology, Taiwan

0299

The Asian Conference on Technology in the Classroom 2013

Official Conference Proceedings 2013

Abstract

This paper surveys a novel speech application on the Mobile Devices. Traditional speech interfaces such as Apple Siri and Google Voice Search cannot directly handle system setting on the mobile devices. Therefore, this thesis presents a friendly user interface based on a voice-controlled API.

I. INTRODUCTION

In the era of updated technology, mobile devices are modern essentials for everybody. Mobile device is a portable and convenient electronic equipment of multiple functions including display and touch control as well as powerful computing capability. The functions of the mobile device are diversified and it has the basic operating capabilities of conventional PC. Hence, due to personal use and rapid information flow, the use of mobile devices has been growing by years. As suggested by the surveys of ACI-FIND in previous years, in terms of "Mobile Internet Penetration Rate" growth trends (Figure 1), the mobile Internet penetration rate of Taiwan reached 10% in Q1 of 2009. In recent years, due to the gradual maturity of mobile communication technology, mobile Internet devices have become the most popular 3C consumer products. As the information of the following figure has suggested, the Mobile Internet penetration rate of Taiwan in 2012 ha reached 31.9%.

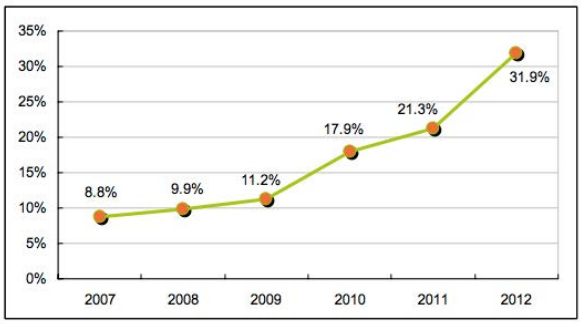

Figure 1 Changes in the mobile Internet penetration rate of Taiwan in 2007-2012

According to the above information, the number of users has increased year by year. The impact of the mobile device penetration is that the contents of mobile device have been making constant progress, providing users more conveniences in use, search, and ready access and reminding. Voice control can be applied in mobile device applications.

II. RELATED WORKS

A. Voice Application Development

The most well-known voice applications at present are Apple Siri intelligent personal assistance and Google Voice Search. Due to the unique intuition without spatial limitation of the voice technology, it has gradually become a new interface of mobile devices including applications such as voice control, voice input method, voiceprint identification. The core of the voice technology is the speech recognition technology (SR) that converts voice into text. The recognition capability of SR engine landed in a development bottle when it achieved 80% accuracy in early 21st century. However, after the Cloud computing of voice platform, SR engine has made major breakthrough in recognition capability after huge amounts of modification and learning of voice samples. The major voice technologies for mobile devices (Figure 2) and relevant companies are: [1].

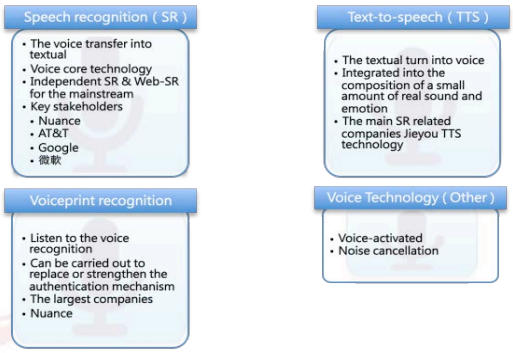

Figure 2 Major mobile device voice technologies

B. *Speech recognition tools*

CMU Sphinx is an open tool speech recognition engine released in 1999 as the open-source code. It is mainly established in ASR (Automatic Speech Recognition) applications and supports multi-platform development. During the research process, the SR development tool of PocketSphinx is used for the offline speech recognition on the Android platform. [2][3]

Android Speech is a Google speech recognition, which is realized by the Server by the division of the API of Android. Speech into application end API and Server end API. It can establish and identify the engine Server; however, the identification results should be fixed. If using Google Server as the engine, it requires Internet connection to conduct the recognition. Meanwhile, implemented in the Server, it is not open in relevant information such as the source code. (Figure 3) is the Google Recognition Server structural diagram[4][5] divided into two parts of the applications and the framework; in the application part, speak now dialog is used by means of Use Google activity of SEMC Car APK on GUI (Graphical User Interface). The means of Use Google Service is to use Recognizer Service to realize Speech Reognizer.startListening (intent) before executing the VR service interface in the Android Voice framework in VR (Voice Recognition) Service. However, the Android interface part is of the open source code.

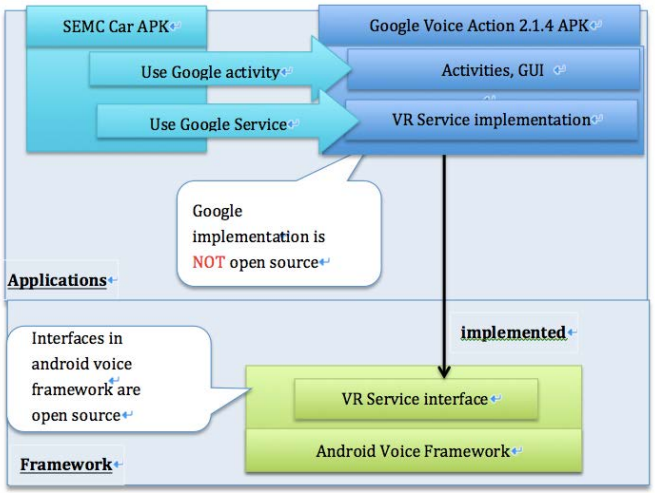

Figure 3 Google Recognition Server structural diagram

Dragon Mobile SDK Reference is the SR and text-voice service that can be built in applications. The database provides Server with voice processing components through a clean and non-synchronized network service API. API speech database mainly provides quick voice search, dictation, high quality and multi-language text-voice conversion function [6].

C. *Voice control and set the system user interface menu*

Apple Siri is the most well-known intelligent personal voice assistant of voice control, which uses the speech recognition technology developed by Nuance Company. It has certain capabilities of understanding the speech and recognizing the intent of the users. It can be used to kill the time and its main function is to help the user in handling some matters such as search, checking the weather, ordering a meal or a cab.

The aforementioned subsystems are executed in order in analyzing the event input by the user to understand the true intent of the user and provide useful services accordingly [7].

Input system supports multi-module status input including not only the voice recognition but also the text input, GUI operation and event trigger. Ambiguous meanings can be eliminated from the early input by the language interpreter in addition to the multi-module status input. In addition, the user input can be consciously guided, so that the user input can be truly reflected to Siri as possible to provide services.

Active body is an important concept in Siri. It can be regarded as the place for the specific execution environment of Siri. The execution system uses all the system data, dictionaries, models and programs for the parsing of user input in the “active body”, and interprets the text message as the true intent of the user before using the external services accordingly.

The execution system is the most technical part of the Siri system. The aforementioned “active body” is the place for the concrete processing of all dictionary resources and modeling resources. The actual processing is implemented by the execution system. The execution system does not only parse the original text input of the user by the internal semantics, it also determines the content of the next sentence that Siri should say in the interactive process between the user and Siri.

The service system is the service-oriented user intent recognition system. Regardless of dialogue control and task control, the fundamental purpose is to guide the user to Siri before providing a specific service to achieve the purpose of accomplishing certain task or solving a certain problem.

The output system is ultimately to provide the service results or content of dialogue process to the user. It does not only support voice, E-mail, text and other multi-module outputs, but also the personalized functions such as interface setting as well.

Menu (or functional list) is a set of limited options displayed on the interface. The menu options are usually simple textual descriptions including images, buttons or checkboxes. For the efficient use of menu, the option description, image content or button description should be clear and easily understandable to the user. It is more difficult for people to recall than recognize as the short term memory of mankind is

quite limited. Therefore, the advantage of menu is that it does not require too much training and memory of answer options, thereby reducing the memory load of the user and making it easier for the user to learn. However, the disadvantage of menu is that too many options may result in longer operation time and thus become slower to skilled users [8] (Tseng, 2008).

Regarding the "setting system interface menu" for smart devices, due to increase in function, there are too many options to result in longer operational time. Therefore, it is not easy to use in operation due to the search for functions. The setting system interface menu is as shown in the following Figure 4.

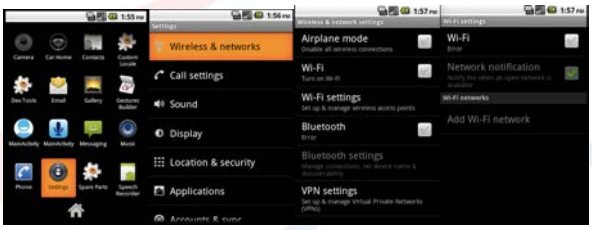

Figure 4 setting system interface menu

D. Case Study

This paper conducts the case study of Apple Siri and Google Voice Search in terms of function, interface and technology as shown in Table I.

TABLE I. **ANALYSIS OF APPLE SIRI AND GOOGLE VOICE SEARCH**

Application	Apple Siri	Google Voice Search
State of Use	online	online
Function	PDA (Remind the user of the desired information)	voice search (search for the user's desired information)
Technology	By the contextual dialogue control, the user is enabled to operate multiple applications by way of natural dialogue without operations according to the manual.	By Natural Language Processing (NLP) and question-answer dialogue system, the powerful search engine is transformed into an expert system to directly respond to the user's questions and needs.

III. USER INTERACTION DESIGN

The user cannot issue voice commands, this interface is not ideal on offline. The effectiveness of a proposed voice interaction mechanism. A yet no simplified the offline voice control interface design. Mainly for that system settings interface application is no longer single-use touch interface. Discrimination system for setting status through voice control, automatic switching action.

The following picture shows the identification return message (Fig 5), than whether its open the function, if so, the execution of its functions (Fig 6).

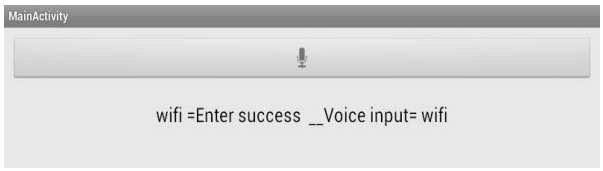

Fig 5 Display the comparison results

Fig 6 Implementation of Wi-Fi

Successful open Wi-Fi status display (Fig 7).

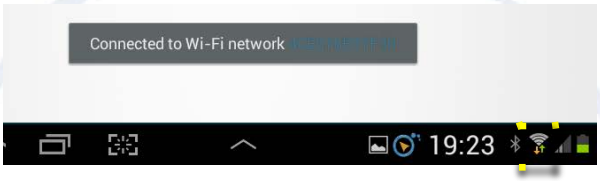

Fig 7 Turning Wi-Fi success

IV. CONCLUSION AND DISCUSSION

Voice control applications implemented in the operating mode, reached through voice control, a simple switch. Hope for the future can automatic switching on-line and off-line. Offline interface no of the simplifies operation. Online and offline can be done automatically detected Advanced system settings, complete voice input set, at this stage remains tactile interface. Advanced settings through voice control system is being developed

REFERENCES

[1] http://www.digitimes.com.tw/tw/rpt/rpt_show.asp?cnlid=3&v=20121212-689&n=1&wpidx=8

[2] http://cmusphinx.sourceforge.net/

[3] http://leiwuluan.iteye.com/blog/1287305

[4] http://mengxx.blog.51cto.com/2502718/767085

[5] Johan Schalkwyk, Doug Beeferman, Fran_coise Beaufays, Bill Byrne, Ciprian Chelba, Mike Cohen, Maryam Garret, Brian Strope, “Google Search by Voice: A case study”, 2012

[6] http://dragonmobile.nuancemobiledeveloper.com/public/Help/DragonMobileSDKReference_Android/SpeechKit_Guide/Basics.html

[7] http://www.infoq.com/cn/articles/zjl-siri

[8] Ching-Hsiang Tseng, Chih-Fu Wu, “A Study on the Graphic and Solid User Interface of Digital Camera Operation”, 2008

iafor