RESEARCH ARTICLE OPEN ACCESS

Wake-up-word speech recognition using GPS on smart phone

Humaid Alshamsi, Veton Këpuska, Hazza Alshamsi
Electrical &Computer Engineering Department Florida Institute of Technology, Melbourne

ABSTRACT
Wake-Up-Word (WUW) is a new prototype of speech recognition not widely recognized. Lately, the use of GPS is widely increased in everyday life that means that our necessities have changed. We can use a new paradigm in controlling the voice of a map in the digital era. This would bring benefit for people while driving a car. In this paper we present a set of voice commands to integrate within the map and navigation voice control. Using a voice control for Global Positioning System (GPS) helps to determine and track the precise location using a technology called Google API. The benefit of this application would be avoiding car accidents using speech command instead of typing.
Keywords: Wake-Up-Word, Speech recognition, GPS, Voice command, mobile computing.

I. INTRODUCTION

Using the wake-up-word (WUW) recognition Android application, the user could search things via human voice but within a defined and complex environment. Moreover, the use of voice is a characteristic easily reproducing by humans. Today people love mobile phones, not only for staying in touch with others and talking, but also for emails, texts, and so on. We are going at the same pace with technology and for this reason, more users mean also more facilities.

Nowadays smart phones have become an important part of our daily life, also related to our needs such as a camera, Music player, Tablet PC, T.V, Web browser etc. New application and operating systems are required with the new technologies. In recent years, smart phones have placed an increasing emphasis on bringing speech technologies into limelight usage. This focus has led to products such as Speech server. However, now we need to focus our attention towards voice message system. It is a service component of the phone that uses standardized communications protocols.

As we have previously said, mobile phones are an important part of modern life, for instance, we need to make an urgent call or send a message at anytime from anywhere. Unfortunately, sometimes we can lose our attention doing these actions and that could cause serious problems, for instance when we're driving or cooking, or doing activities that actually required a high level of attention. In these situations, a voice recognition application for mobile phones could be really useful. First of all, let's recap what an Android operating system is. It is an open source OS that is used to develop an application for mobile users.Going back to the speech recognition application, it was also a part a 1950's research, but it has been not so popular until the mid-2000s. Nowadays, speech recognition technologies have been rapidly evolving thanks to the proliferation of portable computing terminals interconnected with the expansion of the cloud infrastructure. About the mobile voice interface, we could quote Siri, the more recent and famous iPhone, that has also created a voice-activated personal assistant. Moreover, Android, Windows Phone, and other mobile systems have voice functionality and applications. While these interfaces still have a considerable constraint, we are inching closer to machine interfaces we can actually talk to.

II. RELATED WORK

Hae-Duck J. Jeong, Sang-Kung Ye, Jiyoung Lim, Ilsun You and Woo Seok Hyun[1] had proposed a computer remote control system using voice recognition technologies of mobile devices and wireless communication technologies for the driver and physically disabled population as assistive technology.Using speech as the interface has many pros over the traditional tools as a GUI with mouse and keyboard, because speech represents an extension of the human being, that does not require any training and gives the chance of being multitasking and in a faster way. Speech Recognition (SR) represents a perfect interface for the human needs, that could be able to achieve the tasks [2,3,4]. In these cases, people could do a lot of things with computer assistance.To close the gap between natural languages and recognition tasks [7] there is the Novel SR technology named Wake-Up-Word (WUW) [5, 6]. While rejecting the "noise" such as other words, sounds, and phrases WUW SR detects with high efficiency and 100 % accuracy a single word or phrase spoken during this alerting, so called WUW context. WUW speech recognition works like the Key-Word spotting but is able to discriminate the word or phrase during the alerting context. For example, in the phrase "Computer, start PowerPoint presentation", the word "Computer" is used in an alert context. But if we say ''my

computer works with a dual Intel 64 bit processors each with quad cores'' the word computer is used in a not alerting context. Traditional keyword spotters will not be able to discriminate between the two cases. The discrimination will be only possible by deploying higher level natural language processing subsystem in order to discriminate between the two. However, for applications deploying such solutions is very difficult to determine in real time if the user is speaking to the computer or about the computer. Traditional approaches to keyword spotting are usually based on a large vocabulary word recognition [9], phone recognizer [9], or whole-word recognizer that either use HMMs or word templates [10]. Word recognition requires tens of hours of word-level transcriptions as well as a pronunciation dictionary [11].

Usually, recognizers need transcription but on a global scale word markings for the keywords are fundamental. If we choose to configure a system, firstly we need that the tool (i.e. the smart phone) and a Google server are connected. Secondly, user can give command via voice (searching on internet, writing a message, etc.) and at this point, the instructions have been followed. Moreover, this system can also help people with disabling health conditions thanks to a particular function using a TTS procedure (Text-to-Voice) linked to a Google server. Halimah, B.Z. Azlina, A. Behrang, P. Choo, W.O. [12] have proposed a system named Mg Sys Visi that allows to surf the internet and doing many activities via voice command. This system is also thought to help people with disabilities, in fact, it gives the possibility to translate different codes: HTML codes to voice, voice to Braille and then to text again.

The system is composed of 5 modules: Automatic Speech Recognition (ASR), Text-to-Speech (TTS), Search engine, Print (Text-Braille) and Translator (Text-to-Braille and Braille-to -Text). The first testing's results were positive. Moreover, Md. Sipon Miah and Tapan Kumar Godder [13] proposed a voice Control Keyboard Systems which runs from a computer and shows the output on the device's display. In this way also people with a lower knowledge about computer system can use it. But there is also an additional implementation of this system that consists into applying the voice control to the car system.

III. SYSTEM DESIGN

Android App which is going to be designed will have these functionalities: updates and shows the current location with weather status and keeps listening to call any destination that you need to go and do a beep sound every 8 seconds.

The Incremental Model will help us to better accommodate the android app, considering possible future changes. Even if a lot of commercial software manufacturer use the popular model software. There are two conditions in which we can apply the Incremental Model:

1. In the first case you need clear software requirements are clear defined, but the realization can be done later;
2. The basic software functionality is essential from the first moment.

It's important to note that at the beginning we can find software requirements divided into multiple models, outlined according to their functionality. These modules can work alone, but also merging with other modules that have different functionalities. We can also observe that this Model is the most required in a great number of projects, in fact, it makes possible to implement individual functions, but also can give the chance to add stand-alone models.In conclusion, we need to outline three fundamental phases that each increment presents: design, implementation, and analysis. The first one is useful to select which functionality takes priority; during the second phase the implementation of design and the testing are done and in the last phase the functional capability of the product is analyzed. This process is valid for all the functions and it is repeated until the implementation of all the functions.

IV. IMPLEMENTATION

The starting point of the implementation of the software is the user's voice recognition as input. It can be done using the voice command "COMMAND or GO TO" within some limitations of recognition. This command will be translated in a text that activates the GPS system that allows track the user's location and the nearby public spaces such as restaurant, libraries, schools, etc.

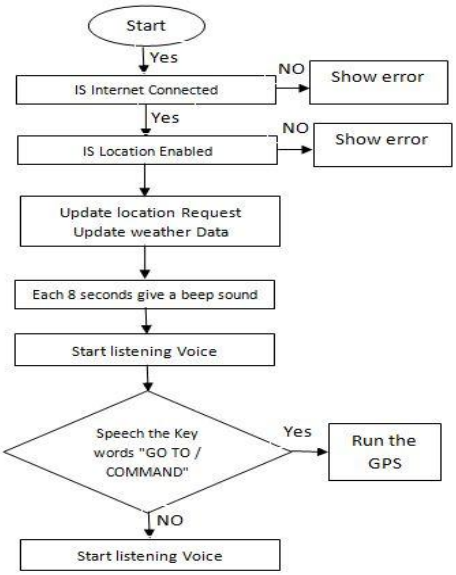

Figure 1: Flowchart – WUW speech recognation using GPS for smart phone.

This is possible only with an Internet connection available, otherwise it gives an error. Another "error condition" can be a wrong command from the user. In this last case, the process continues to listening because doesn't recognize the command. Moreover, the beep-sound every 8 seconds indicates when the user starts a new research or refreshes the current location.

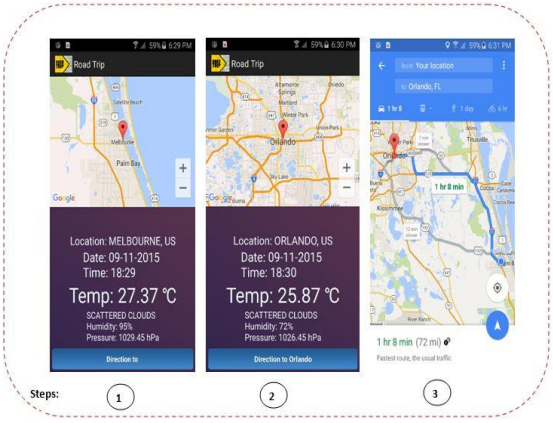

Figure 2: An Overview of the system

The above figure shows how the system runs in three steps. When the system starts to run it will check if there is an internet connection, then the system starts to update and shows the current location and the weather as in step number 1. Then the system keeps listening until the user says the command keyword which is "Go to …name of destination ". For example "Go to Orlando" it will show the location of Orlando and weather status as it shows in step number 2, and finally step number 3 shows three different routes and the user can choose the fastest one.

V. TESTING AND FEATURE

A. Testing summery

The final step for this paper is to assess and evaluate the project performance; to measure how many of the requirements for WUW speech recognition system using GPS in a mobile phone can be achieved. Actually, testing has been continuously addressed from the early implementation stage until the final stage.

Firstly, the testing of each function is carried out individually. It is tested to ensure that the algorithm and each line code works correctly. Sometimes, I run the application in a different phone so to make sure that is running same as expected. Secondly, after completing a certain stage, the performance of that stage is tested. Furthermore, after integrating the

system stages, the overall system performance was tested. In these phases, sometimes an implement and use Google API that is useful for this project is discovered. The problem was with the huge number of multi-class features that need to be trained. To solve this problem, attention was turned to the Android platform tools that can be used with the project data. The Android platform was used to program the application and test the application. Eventually, after many attempts, the optimal solution was found.

B. Advantages

The important advantage of the speech input is that user can do it easily and without specialized skills. Moreover, the command can be ordered even if the user is doing other activities. Automatic Speech Recognition could require Speaker Training, but it is not always essential; sometimes the program is set up during the system development with speech sample of an automatic collection of Speakers [14, 15].

VI. FUTURE WORK

We can improve the quality of navigation with increasing precision of GPS service in software way. Researching and implementing different mathematic algorithms can hide errors of the GPS locating. Theoretical researches in this theme are pending at the moment. Realizing the real-time route planning through the user interface of the phone is our target now as well. This way using the software could be detached from the PC so users don't have to plan the itinerary in advance and could get to immediate emerging targets. We are intent to make this application able to collaborate with map software to get more information's from the streets and manage the route planning if there isn't available Internet access. A map handling software must know public transport system as well to help people to use different vehicles. To realize it we have to contact a map developer firm, what specialized for mobile devices. Using functions of map software that knows the traffic rules could enable navigating in different vehicles in the future.

VII. CONCLUSION

A smart phone using a voice recognition system can work with simple commands and be implemented into a user-friendly device. Users can freely choose the device with the better qualities for their needs. This elaborate aimed to explain the importance of voice recognition software in the modern era and overall the importance for people with disabilities will gain more independence with a simple application, using only a voice control. In conclusion, we can affirm that this technology implementation could help the general population to execute simple daily commands via voice.

ACKNOWLEDGEMENTS

The work of Veton Këpuska was supported by Florida Institute of Technology which gave us the opportunity to work on this project. Also we need to thanks him which led us to do more research in such new topic to us.

REFERENCES

[1] Hae-Duck J. Jeong, Sang-Kug Ye, Jiyoung Lim, Ilsun You, and WooSeok Hyun ," A Computer Remote Control System Based on Speech Recognition Technologies of Mobile Devices and Wireless Communication Technologies", 'IEEE Conference Publication',2013,page no. 595-600 .

[2] Ron Cole, Joseph Mariani, Hans Uszkoreit, Giovanni Batista Varile, Annie Zaenen, Antonio Zampolli, Victor Zue (Eds.), Survey of the State of the Art in Human Language Technology, Cambridge University Press and Giardini, 1997.

[3] V. Këpuska, Wake-Up-Word Application for First Responder Communication Enhancement, SPIE, Orlando, 2006.

[4] T. Klein, Triple scoring of hidden markov models in wake-up-word speech recognition, Thesis, Florida Institute of Technology.

[5] V. Këpuska, Dynamic time warping (DTW) using frequency distributed distance measures, US Patent: 6983246, January 3, 2006.

[6] V.Këpuska, Scoring and rescoring dynamic time warping of speech, US Patent: 7085717, April 1, 2006.

[7] V.Këpuska, T. Klein, On Wake-Up-Word speech recognition task, technology, and evaluation results against HTK and Microsoft SDK 5.1, Invited Paper: World Congress on Nonlinear Analysts, Orlando 2008.

[8] V.Këpuska, D.S. Carstens, R. Wallace, Leading and trailing silence in Wake-Up-Word speech recognition, in: Proceedings of the International Conference: Industry, Engineering & Management Systems 2006, Cocoa Beach, FL., 259–266.

[9] J.R. Rohlicek, W. Russell, S. Roukos, H. Gish, Continuous hidden Markov modeling for speaker-independent word spotting, vol. 1, 23–26 May 1989, pp. 627–630.

[10] C. Myers, L. Rabiner, A. Rosenberg, An investigation of the use of dynamic time warping for word spotting and connected speech recognition, in: ICASSP '80. vol. 5, Apr 1980, pp. 173–177.

[11] A. Garcia, H. Gish, Keyword spotting of arbitrary words using minimal speech resources, in: ICASSP 2006, vol. 1, 14–19 May 2006, pp.

[12] Halimah, B.Z. Azlina, A. ; Behrang, P. ; Choo, W.O.,"Voice recognition system for the visually impaired: Virtual cognitive approach "IEEE Conference Publications Volume: 2 ,DOI: 10.1109/ITSIM.2008.4631738, Publication Year: 2008 , Page(s): 1 - 6 .

[13] Md. Sipon Miah, and Tapan Kumar Godder , "Design Voice Control Keyboard System using Speech Application Programming Interface "IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 6, November 2010 ISSN (Online): 1694-0814 www.IJCSI.org 269 To 277.

[14] Kenneth Thomas Schutte "Parts-based Models and Local Features for Automatic Speech Recognition" B.S., University of Illinois at Urbana-Champaign (2001) S.M.,V Massachusetts Institute of Technology (2003). Bain, K. Paez, D. Speech Recognition in Lecture.

[15] Fundamentals of Speech Recognition, L. R. Rabiner and B. H.Juang,Prentice Hall Inc., 1993.