

Speech2Spell: A Voice-Enabled Spell-Out Game using Voice Recognition Algorithm

Teodoro F. Revano Jr.^{#1}, Bryan Dadiz^{#2}, Jasmin D. Niguidula^{#3},
Maribeth B. Fernandez^{#4}, Rejan Tadeo^{*5}

[#]*College of Information Technology Education, Technological Institute of the Philippines Manila, Philippines*

¹tfrevano@yahoo.com

²bryan.dadiz@tip.edu.ph

³jasniguidula@yahoo.com

⁴bethbonzon@yahoo.com

^{*}*College of Computer Studies Lyceum of the Philippines University Manila, Philippines*

⁵rltadeo@yahoo.com

Abstract - Speech2Spell: A Voice-Enabled Spell-Out Game Using Voice Recognition Algorithm using Android platform is a mobile game using voice recognition. Voice recognition are used in security reasons to protect or secure one's property, some are used in voice command machines like televisions, lights, air conditioners and other appliances even in small devices help the users in asking directions and many other things that you didn't know it can be done by voice recognition. The game speech2spell is a voice recognition game that simply recognizes all single word that is being spoken by the user, the game is played only by a single player; also the game must be played at a quiet environment or user can also use their headphone to limit the noise that has been absorbed or heard by the mobile. The game can help users to improve their knowledge in speaking the right word. The game has a custom mode setting for the user. On custom mode game, the user can take a picture on the object and can be used in game system. The captured images taken from the custom mode is only exclusive at the custom mode. The game also has classic mode, it has a default images that flash on the screen of the mobile for the user can guess what the system shows.

Keywords- Voice Enabled; Voice Recognition; Mobile Game and Android Platform.

I. INTRODUCTION

Technology had emerged as a powerful tool in life. In early days, books, newspapers and magazines are the only sources of information but now because of the advent of the technology, can quickly browse it on the Internet. These devices offer not only as a gaming platform but also with numerous features its users could acquire, like for instance the capability to connect to the World Wide Web, and play videos and music. Search engines were full-developed for the user to use efficiently it. From a keyword- based query, it was produced by an input voice-based query. Accuracy in search engines became an issue because of the methods used for input voice segmentation and input voice matching process in the voice of entry and the list of words on the web. Sometimes, the retrieved word(s) are not relevant to the word(s) query and input

voice must be clear and must be pronounced in right English language way and a good accent of the word(s). Most of the search engines were conducted in desktop PCs or laptops. Smart Mobile devices were rarely used as a medium of calling texting for the relevant web pages of the queries. Nowadays, mobile devices play a significant role in our daily life. It is very much handy and easy to use. A search engine in mobile devices could help people in searching for the relevant web pages to their query anytime and anywhere as long as you are connected to the Internet.

The Speech2Spell: A Voice-Enabled Spell-Out Game Using Voice Recognition Algorithm using Android platform involves the use of additional devices like headphones as well as microphones for the enhancement of game play. Intended users are kids and kids at heart. In addition, to that, it can be used for educational purposes for the students to enhance their skills in vocabulary.

II. RELATED WORKS

According to [1] in their study entitled "Multi-Modal Emotion Recognition from Speech and Text" "they are stated that the emotion recognition from speech signals and textual content. In the analysis of speech signals, thirty-three acoustic features are extracted from the speech input. After Principle Component Analysis (PCA) is performed, 14 principle components are selected for discriminative representation. In this representation, each principle component is the combination of the 33 original acoustic features and forms a feature subspace. Support Vector Machines (SVMs) are adopted to classify the emotional states. In text analysis, all emotional keywords and emotion modification words are manually defined. The emotion intensity levels of emotional keywords and emotion modification words are estimated based on a collected emotion corpus. The final emotional state is determined based on the emotion outputs from the acoustic and textual analyses.

The study of [2] in their study entitled "Apparatus for converting speech to text" shows that the speech to text

conversion system which comprises a plurality of user terminals for recording speeches, at least, one automatic speech recognition processor to generate text from recorded speeches, and communication links to return corresponding documents to respective users. At least one automatic speech recognition processor improves recognition accuracy using data from the recorded speeches and the relevant texts, the data being selected dependent upon subject matter area.

Based on the study of [3] entitled "Fundamentals of Speech Recognition" Their study shows that the mobile telephone was having speech recognition and speech synthesis functionality. The telephone has a memory for storing a set of speech recognition templates corresponding to a set of relevant spoken commands and a transducer for converting a spoken command into an electrical signal. Signal processing means provided for analyzing a converted spoken command together with templates stored in the memory to identify whether or not the converted spoken commands corresponds to one of the sets of spoken commands. The phone user can select to download, into the phone's memory and a set of templates for a chosen language, from a control station via the wireless transmission channel.

A method of operating a mobile communication device having speech recognition functionally and comprising a memory for storing a set of speech recognition templates, each template corresponding to a human language, transcoder means for converting a command spoken by a user into an electrical signal, and signal processing means for analyzing a converted spoken.

Control together with templates stored in memory to identify whether or not the converted spoken command corresponds with one of said set of spoken commands, the method comprising the step of receiving from a central station via a wireless transmission channel set of speech recognition templates and storing the received set of templates in the memory of the device.

According to Imre [4] entitled "Speech Recognition in Mobile Phones". Their study aimed to propose a speech input implemented in voice user interface (voice UI) plays a significant role in enhancing the usability of small portable devices, such as mobile phones. In these methods more traditional ways of interaction (e.g. keyboard and display) are limited by small size, battery life and cost. Speech is considered as a natural way of interaction for man-machine interfaces. After decades of research and development, voice UIs are becoming widely deployed and accepted in commercial applications. It is expected that the global proliferation of embedded devices will further strengthen this trend in the coming years. A core technology enabler of voice UIs is automatic speech recognition (ASR). Example applications in mobile phones relying on embedded ASR are name dialing, phonebook search, command-and-control and more recently large vocabulary dictation. In the mobile context, several technological challenges have to be overcome

concerning ambient noise in the environment, constraints of available hardware platforms and cost limitations, and the necessity for extensive language coverage. Also, mobile ASR systems need to achieve a virtually perfect performance level for user acceptance. This chapter reviews the application of embedded ASR in mobile phones, and describes specific issues related to language development, noise robustness and embedded implementation and platforms. Several practical solutions are presented throughout the chapter with supporting experimental results. However, the proposed study of speech recognition in mobile phones are related in our proposed study but there are some problems with another functionality is not primary to mobile phones originally.

Based on the gathered related studies, the researchers have come up with the idea of developing Speech2Spell: A Voice-Enabled Spell-Out Game Using Voice Recognition Algorithm using Android platform. With the studies mentioned, the researchers agreed that they can use and apply the voice recognition, spell out a game in the mobile phones.

III. OBJECTIVES OF THE STUDY

The objective of the researchers is to develop a mobile game application implementing the voice recognition algorithm using the mobile phone. The study aims specifically: to develop a mobile game application that recognized voice from the user using voice recognition; to apply the capabilities of mobile phones as a medium of Game engines using speech recognition; to use captured image of a camera- equipped mobile phones as an item in-game by providing an image selection from the match.

IV. SYSTEM ARCHITECTURE

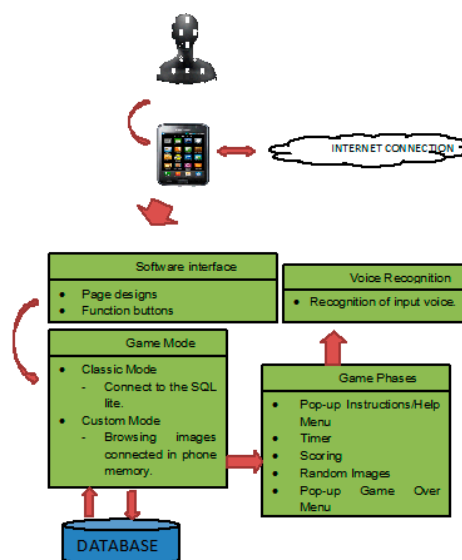


Fig 1. System Architecture

Figure 1 shows the connectivity of Internet Connection. Since the system is Web base, it is a must to ensure that the mobile phone has the capability of

connecting to Mobile Web. The Mobile Web refers to the Internet-Connected Applications. At present, wireless connectivity became a powerful tool in different conditions.

The second stage is the entering the game mode selection. When the user picks the classic game, it will first show a quick reminder on how to play; having a play button will bring you to the game simulation. On game screen show an image and timer also a speak button, all images are randomly demonstrated by the match. When the game starts the timer will automatically countdown, for the user will hurry to guess what image is being shown by the application, it will be ended after 25 seconds. When the user picks the custom mode, just like the classic mode it shows a quick reminder on how to play the game, it has the play button to start the game mode. At the page, shows you selections the view image and take the picture, at the view image button it will bring you to the gallery of phone memory then you can pick a picture to use it in the game. The take picture button it allows you to get access to the mobile camera to take an image on a chosen item and bring it to the game. The next stage is the voice matching when the user taps the speak button, and then user guesses the image showed by the application. The Input word given by the user converts into the file and send it to the Google server; the Google server will send back to the mobile all the related words. Then the game database will query and matched the word in the required shown image. When the input voice or word is correct, it will give a score and then brings the user to the next image. The game will continue loop and randomly shows the images until the user game over. At the end of the game, it will show you the total score that he/she gets from the game.

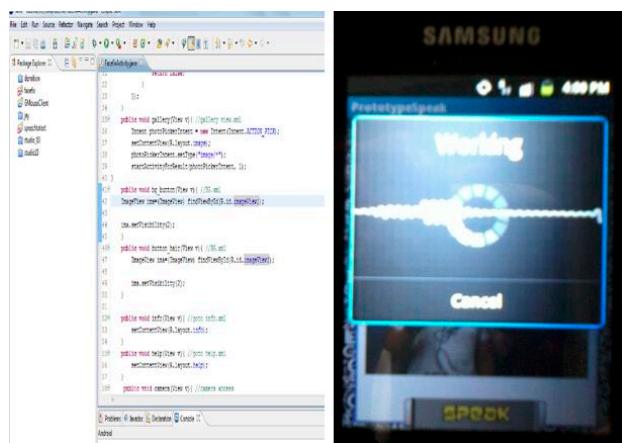


Fig 2. Voice Recognition

Figure 2 shows the design of the application with the use of speech to spell application. In this illustration, you can see the graphical user interface (GUI) of the application. The researchers applied the voice recognition algorithm suitable for the proposed application. It shows how the mobile gets the input voice of the user and process to be recognized by the application.



Fig 3. Speech2Spell Main Interface and Options

Figure 3 shows the main screen of the proposed application. It has three buttons: the start the Help and Exit buttons. The user chooses on what they want to do. The user can select start, help or exit as the user navigate the menu selections, for the user have an option on what to do. The Custom interface shows a variety of browse and the play button. This page lets the user browse for the image from the memory of the mobile phone and rename the image to use in a game. The database of the custom mode is different to the classic style. In custom mode, the user can select all the images that he/she want to put in the game. In naming image, the user must set the right spelling so that the picture can be identified in a game. The Browse button will allow the user to pick an image from the memory of the phone. After the user picks the image, the image must be named, for the system will know what image has been choose by the user. In calling the image must be in a correct spelling and must be the simple word of text because the words are limited to the Google server.

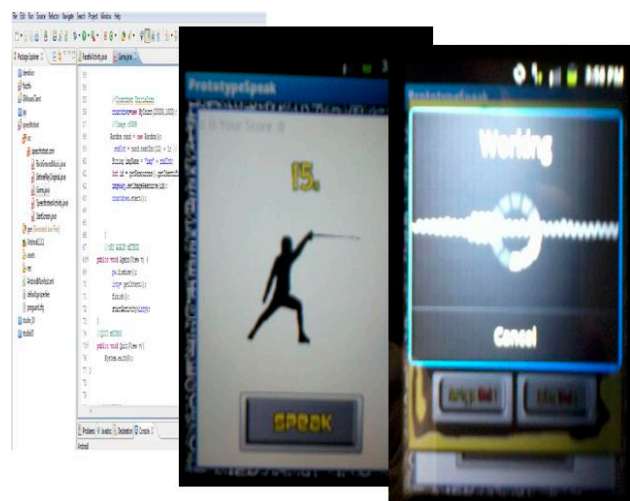


Fig 4. Game engines using speech recognition.

Figure 4 shows on how the mobile game application is simulated and being done with the voice recognition. This application shows random images for the user to guess on and also, a timer to limit and to gives a thrill of the of the, it also gives points for every correct answer on the game, and the most important recognize the input voice of the user.

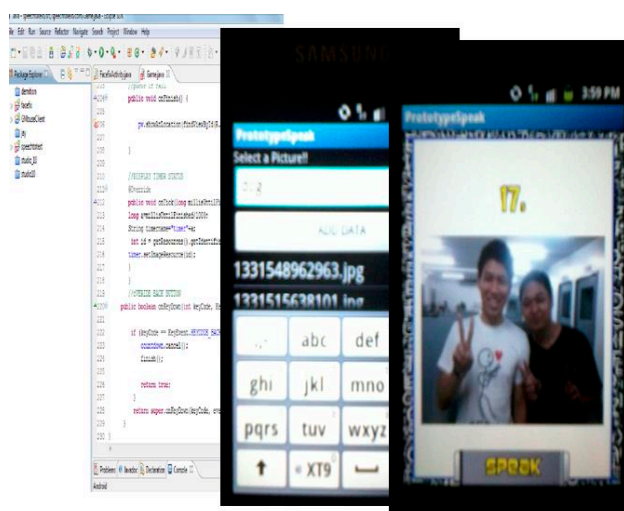


Fig 5. Capturing Image

Figure 5 shows the browsing mode of the game and how the users pick the image from the memory of the phone and uses it to the game. The application shows all the images from the memory, after picking an image the user have to name it and be used in the game.

IV. CONCLUSIONS

The study suggests a mobile game application that recognized the voice from the user. Using recognized intent of the Android OS, it recognized the input word given by the user and also using of the internet needed by the proposed system. The system sends the data from the input voice of the user in change the input voice into a file then send to Google sever; Google sever will send back all the possible verified words. The system will query from our database all the data given by the Google server. The capabilities of mobile phones as a medium of game engines. The proposed application successfully made a game module that gives a selection of game setting or game mode for the user to choose whether the "Classic" or the "Custom" game. Also, it shows help menu for the user. It has a countdown timer to limit the gameplay. It shows random images that needed to the game application. Capturing of an image using camera- equipped mobile phones as an item in a game. Another objective of the game that the user uses, the images from the mobile phone memory, to be able to use in the game. At the game mode selection, the user chooses the custom mode having a button the user browse all.

REFERENCES

- [1] C. Ze Jing and W. Chung Hsien, "Multi-Modal Emotion Recognition from Speech and Text " *Computational Linguistics and Chinese Language Processing*, vol. 9, no. 2, pp. 45–62, Aug. 2004.
- [2] B. David and T. Henry Hyde, "Apparatus for converting speech to text," 6366882, 04–02-2002.
- [3] R. Lawrence and J. Biing-Hwang, *Fundamentals of Speech Recognition*, 1 edition. Englewood Cliffs, N.J: Prentice Hall, 1993.
- [4] I. Varga and K. Imre, "Speech Recognition in Mobile Phones," in *Automatic Speech Recognition on Mobile Devices and over Communication Networks*, Springer London, 2008, pp. 301–325.