VISQUIT - Voice IS a Quite Universal Interactive Tool (Kiosk with Vioce Recognition)

Kim Jung Mo
2014005414
Dept. of Information System
College of Engineering,
Hanyang University
Seoul, Rep. of Korea
Email: cadenzah93@gmail.com

Lee Hyo Sik
2014026208
Dept. of Business Management
School of Business,
Hanyang University
Seoul, Rep. of Korea
Email: lhslhs9394@gmail.com

Lee Ha Min
2017029134
Dept. of Information System
College of Engineering,
Hanyang University
Seoul, Rep. of Korea
Email: ggamini7@gmail.com

Hwang Sung Woo
2016026599
Dept. of Information System
College of Engineering,
Hanyang University
Seoul, Rep. of Korea
Email: hsw0194@gmail.com

Abstract—VISQUIT is a Kiosk service powered by voice recognition technology based on artificial intelligence. Nowadays, many stores adopt using Kiosk system, by which the cost of hiring clerks and the time spent to make an individual order reduce. People who are accustomed to digital-based system has no much difficulty using this type of new method in store, but for those people who are digitally illiterate, even ordering their meal in a restaurant is a hard task. To solve the problem with current Kiosk system's poor user experience, we suggest using voice recognition technology to compose the ordering system's UI. This can effectively replace the previous "select-based" ordering system with "speech-based". Based on this approach, we will develop and propose new paradigm of designing UI of Kiosk system.

1. Introduction

1.1. Motivation

Kiosks are unmanned machine that can take orders on behalf of a clerk. Recently, the installation and operation of kiosks is increasing, especially in franchise restaurants. This saves labor costs for the personnel required to place an order and allows more orders to be processed in unit time.

However, some people suffer from the appearance of kiosks. That people are who are not familiar with these Digital technology. In particular, for older people, they are very new to digital technology. According to the National Information Society Agency (NIA) '2018 Digital Information Gap Survey,' the overall 'digital information level' for the elderly aged 55 or older is only 63.1% of the general public. The contrary situation arises where the marginalized class exists due to the emergence of technology to make people convenient.

1.2. Problem Analysis

Among the problems of kiosks, we will focus on improving the ordering / payment software installed inside the kiosks. In the conventional method, as all the menus are displayed on the screen, the amount of information increases, which causes inconvenience to those who are not familiar with the UI. In addition, in order to maximize the amount of information displayed, the size of text and images is small, and the screen resolution is limited, so the display is not clear. In addition, the sensitivity of the touch screen is not as accurate as that of the physical buttons, providing a confusing UX to the user.

1.3. Solution

Using the artificial intelligence speech recognition technology of SKTelecom's NUGU Platform, the user's speech is recognized and ordered based on it. This provides a familiar interface to users who are not familiar with digital technology, as if they are talking to a real person. It can also handle multiple orders without the need for a complex UI.

In addition, the UI is configured to maximize the clarity of the content by using a large image and text, and to better understand the content by users with low vision. As the size

of the content increases, the amount of information that can be contained on one screen decreases, so the UI is designed to implement the same business logic as the existing despite the reduced amount of information.

3. Development Environment

3.1. Software Development Platforms

We chose Web environment to develop our project. Many applications can be run on the Web. Web technol-ogy(Node.js) will be used to show that voice kiosks work properly. In addition, we will use RDS to apply what we have learned in class, and additionally use AWS commercial cloud platforms such as EC2, S3, etc. SKT's NUGU API will be used to analyze users' voice commands and translate them into text.

3.1.1. Node.js.

Node.js is an open-source, cross-platform, JavaScript runtime environment that executes JavaScript code outside of a browser. Node.js lets developers use JavaScript to write command line tools and for server-side scripting—running scripts server-side to produce dynamic web page content before the page is sent to the user's web browser. Consequently, Node.js represents a "JavaScript ev-erywhere" paradigm, unifying web application development around a single programming language, rather than different languages for server- and client-side scripts.

3.1.2. React.js.

React (also known as React.js or ReactJS) is a JavaScript library for building user interfaces. It is maintained by Facebook and a community of individual developers and companies. React can be used as a base in the development of single-page or mobile applications, as it is optimal for fetching rapidly changing data that needs to be recorded.

3.1.3. Amazon Web Service EC2.

Configure an environment where applications can be developed and distributed quickly without the need for physical hardware equipment. Deploy as many virtual servers as you want, configure your security, network, and manage your storage. A sud-den change in server spacs can quickly scale up or down depending on the changes, reducing the need for forecasting server traffic.

3.1.4. Amazon Web Service S3. Amazon S3 or Amazon Simple Storage Service is a service offered by Amazon Web Services (AWS) that provides object storage through a web service interface. Amazon S3 uses the same scalable storage infrastructure that Amazon.com uses to run its global e-commerce network. Amazon S3 can be employed to store any type of object which allows for uses like storage for Internet applications, backup and recovery, disaster recovery, data archives, data lakes for analytics, and hybrid cloud stor-age. In its service-level agreement, Amazon S3 guarantees 99.9% uptime, which works out to less than 43 minutes of downtime per month.

3.1.5. Amazon Web Service RDS.

Amazon Relational Database Service (or Amazon RDS) is a distributed rela-tional database service by Amazon Web Services (AWS). It is a web service running "in the cloud" designed to simplify the setup, operation, and scaling of a relational database for use in applications. Administration processes like patching the database software, backing up databases and enabling point-in-time recovery are managed automatically. Scaling storage and compute resources can be performed by a single API call as AWS does not offer an ssh connection to RDS instances.

3.1.6. SKT NUGU SDK.

Based on SK Telecom's technical skills such as voice recognition, voice synthe-sis, and understanding of natural language through NUGU developers, the company can develop new functions and provide NUGU's various functions through voice command in devices or applications owned by its affiliates. We will recognize and categorize the user's voice commands through the NUGU SDK and send output results to the user via voice.

3.1.7. Docker.

Docker is a set of platform-as-a-service (PaaS) products that use OS-level virtualization to deliver software in packages called containers. Containers are iso-lated from one another and bundle their own software, libraries and configuration files; they can communicate with each other through well-defined channels. All containers are run by a single operating-system kernel and are thus more lightweight than virtual machines. The service has both free and premium tiers. The software that hosts the containers is called Docker Engine. It was first started in 2013 and is developed by Docker, Inc.

3.2. Programming Languages

3.2.1. Javascript.

Javascript is a high-level, interpreted scripting language that conforms to the ECMAScript spec-ification. Javascript has flexible grammars: freedom from indentation, loose type checks. Also, it adopts modern progamming padigms and has convenient and great features: function programming, reactive programming. By using this language we can learn various modern progamming paradigms. Javascript is used in web browsers, which means it does not require any special working environment to run program written by Javascript.

3.3. Cost Estimation

This project heavily rely on Amazon Web Service. The cost estimation is in Table 1 . This is calculated by Amazon Web Service Cost Calulator.

Service	Region	Cost(Monthly)
Amazon EC2	Asia Pacific (Seoul)	USD($) 11.65
Amazon RDS	Asia Pacific (Seoul)	USD($) 19.64

TABLE 1. COST ESTIMATION FOR THE PROJECT

3.4. Development Environment Description

Used development environment tools information is de-scribed in Table 2 .

Name	Version	Description
Windows	10 Pro	Operating System made by Microsoft
macOS	Catalina(10.15)	Operating System made by Apple, used in Macbook
Ubuntu	16.04.6	Operating System developed by Linux Foundation
Visual Studio Code	1.39.1	Text editor and Integrated Development Editor made by Microsoft
Atom.io	1.40.1	Text editor and Integrated Development Editor made by GitHub
iTerm2	3.3.6	Terminal Emulator used in macOS
Z shell(zsh)	5.7.1	Unix shell for CLI

TABLE 2. ROLE ASSIGNMENT FOR EACH PARTICIPANTS OF THIS PROJECT

3.5. Market Research & Software in Use

3.5.1. Kiosk Market in Korea.
Unmanned stores are spreading all over the world. The unmanned store are ac-tively being made due to artificial intelligence, the Internet of Things, the development of sensor technology, the lack of labor force and the minimum wage hike. South Korea has been burdened with labor costs following a 10-percent minimum wage hike for two consecutive years, and the unmanned era is in full swing due to the spread of an ”un-tact” culture that is reluctant to contact. In restaurants, cus-tomers order on their own through KIOSK. And unmanned convenience stores have begun to sprout up everywhere. Unmanned store is spreading not only in the retail industry but in almost all areas, regardless of industry. Demand for kiosks is also on the rise as unmanned stores have emerged. A market for unmanned payments using kiosks is emerging. As of 2017, the size of South Korea’s kiosk market is estimated to be around 250 billion won, with a high annual growth rate of 14 percent. The market, which stood at only 60 billion won in 2006, rose to 180 billion won in 2013 and rose to 250 billion won in 2017.

3.5.2. Voice Recognition ARS.
The ARS allowed us to perform the tasks we needed without waiting time. How-ever, the initial ARS was a one-sided way of listening and pressing buttons. As people become accustomed to ARS, they can’t wait to get the number of services they need. In addition, the service used was mostly concentrated in some services. Based on this, ARS through voice recognition was able to connect the necessary services quickly by listening to the voice or words. Considering these voice recognition ARS, it is thought that it will help us solve the problem.

3.5.3. Voice Recognition AI.
There are services in the mar-ket called artificial intelligence voice assistant such as Bixby, Siri, Nugu and Gigagenie. Despite providing convenient services, many people are reluctant to use these services. Understanding these reasons is thought to help us envision the services we will provide in the future.

3.5.4. Delivery Application.
We think the delivery app is the same situation as the kiosk we are currently thinking about. Orders, which were previously made through voice-to-speech conversations over the phone, have been made possible through the application with several touch points. There are many advantages, such as the availability of personalized orders through delivery apps, but it is still a far from those who are not familiar with digital culture. If Kiosk’s problem is solved, it is believed that the delivery app will be able to solve the problem in the same way later.

3.5.5. Starbucks Siren Order with Voice.
T-Map x NUGU introduced a voice service for Starbucks’ ”Siren Order”. The combination of ”T-Map x Nugu” and ”Siren Order” makes it possible for users to order for various Starbucks products with voice, even while driving. Orders are based on voice while driving, followed by beverage selection, neighborhood store selection, order product confirmation, order receipt and payment. In addition, linking T-map, by forwarding the order to the store within five minutes of arrival at the designated store, users were able to improve the inconvenience of waiting for the ordered product at the store for a long period of time, while considering the freshness of the product.

3.6 Task Distribution

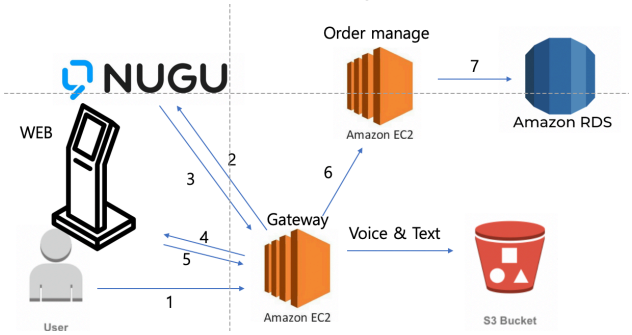
Task Distribution is shown in Table 3 . Note that each project participants periodically switched their roles with the others to apply different perspectives and improve their expriences.

Task	Assignee	Description
Project Manager	Kim Jung Mo	Schedule overall development plan and assign proper jobs to team members
Consumer	Lee Ha Min	Test and try out a prototype application, gather the potential improvement
User	Lee Hyo Sik	Gather basic feature requirement for this project by asking qeustioaires the eldery near the university.
Developer	Hwang Sung Woo	Prepare a development environment for this project and build up the application

TABLE 3. TASK DISTRIBUTION FOR EACH PARTICIPANTS OF THIS PROJECT

4. Specification

Our service helps people who are not familiar with KIOSK to place voice-based orders. When the user presses the Speech Recognition button for an order and speaks, the voice file is passed to the server that was built with AWS. The server delivers the voice recognition file to the NUGU platform, which converts the voice file to text and returns the result to the server. (Results consist of text.) The server receives confirmation from the user about the order and then places the actual order within the server based on the user-requests.



4.1. Speech To Order(ID: 110)

4.1.1. Receiving Voice Input from a User.

1) Receive raw voice input from users
2) Send it to the server

4.1.2. Voice Preprocessing and Noise Removal(ID: 120).

1) Receive voice from KIOSK
2) Noise rejection for better voice processing
3) Send noise-free voice to the NUGU server

4.1.3. Convert preprocessed speech into text data(Speech-To-Text)(ID: 130).

1) Receive noise-free voice
2) Convert the voice to text
3) Send the voice back to the server

4.1.4. Extract order data from text data(ID: 140).

1) Receive converted text data from NUGU server
2) Extract only significant data relevant to the order
3) From extracted information, create mapping order content(menu, quantitiy, etc)
4) Send the information to KIOSK

4.1.5. Validate Extracted Order Data(ID: 150).

1) Receive the interpreted order information from the server
2) Diplay the order information on the screen
3) The user checks whether the order information given is correct or not
   If the order is correct, add the order item to the order list
   If the order is wrong, delete the order item and repeat from the beginning
4) If the user has more to order, repeat from the beginning

4.2. PoS: Point of Sales System(ID: 210)

1) Show all menu list
2) Add an order item to the total order list
3) Show a summary of an entire order list on the screen
4) Payment Process (Not covered in this project)
5) Send an order list to the server

4.3. Specification

4.3.1. ID: 110.

1) The length of voice received from the client is limited to up to 30 seconds.
2) Voice should be mentioned in order of menu and quantity.
   ex) Hamburger hana(One), Cola du(Two) gae

4.3.2. ID: 130.

1) Use NUGU SDK

4.3.3. ID: 140.

1) Menu word, quantity word classified as signal data
2) Match each menu and quantity
3) When a word regarding quantity is encountered, it is recognized as the end of the corresponding menu-related order.

4.3.4. ID: 210.

1) To show the behavior of KIOSK, show the same menu as the menu of any store.

5. Conclusion

Not added yet :)

Acknowledgments

The authors would like to thank...

References

[1] H. Kopka and P. W. Daly, A Guide to LaTeX, 3rd ed. Harlow, England: Addison-Wesley, 1999.