Possibility of Voice Recognition Control for substituting Smart Phone Applications

Jaekyu Park¹, Youngjae Im²
¹Home Appliance & Air Solution Company, LG Electronics, Changwon, Korea
²Division of Design Engineering, Dong-eui University, Busan, Korea

Abstract

The purpose of this study is to investigate the possibilities through voice recognition control for smart phone applications. Voice recognition on smart phones is currently implemented with some functions and is expected to be expanded in the future. Voice recognition control on a smart phone will gradually reduce the ratio of direct manipulation using a visual perception and finger. This study identified the possibility of voice recognition control for high frequency applications in smart phones based on Android and iOS. In addition, the steps of the voice recognition control for the applications are classified and it is determined whether the control is possible by only voice recognition. As results of the assessment, it was derived the level of voice recognition control (full application, partial application, not applicable) by the function of the applications. This study verified the possibility of substituting Smart Phone Applications through using back-end approach. This is an effective in determining the degree to which existing technologies can be replaced due to the adoption of new technologies. In addition, this study will be expected to provide a basic protocol of voice recognition control for smart phones.

Keywords: Voice recognition, Smart phone, Application

1. Introduction

The emergence of smart phones has enabled us to integrate what we had with our existing MP3 players, digital cameras, and other products into one. One of the main factors that make this possible integration is the role and benefits of the touch interface. In fact, the touch interface is suitable for manipulating various functions at the same time through an intuitive finger input, and the user's initial learning cost is also low [1].

However, the touch interface that is applied to the smart phone causes some problems. First, the size of the display on the smart phone is getting larger than in the past, but the amount of information that can be displayed on one screen is limited. Therefore, it is inevitable that the goal of tasks can be achieved through various stages of input process. In addition, this process requires not only the touch input but also the visual attention of the user, which may cause a safety problem particularly while using the smart phone on the move.

Voice user interface (VUI) is attracting attention as a way to address the touch interface issues mentioned above. The VUI is the most natural and comfortable interaction method by communicating information based on words and conversations that people use in ordinary day [2, 3]. The full-scale commercialization of the voice recognition interface can be seen after the ‘SIRI’ service provided by Apple in 2011 with the release of i-phone 4S. These voice recognition technologies are not limited to devices such as smart phones, but are gradually being expanded and applied to AI speakers and smart TVs [4].

Meanwhile, companies manufacturing voice interface products are applying their technology in their own way without standardized standards at the moment. Therefore, how to process and search the functions of voice interface products is the key success factor of each manufacturer.

As a result of the related study [5], it has been found that the use rate of voice interface through smartphone is somewhat low. However, if the related functions are improved in the future, they will be used more. Koh [6] analyzed the error type and speech pattern for users of smart phone voice interface. The results showed that the satisfaction with voice interface was correlated with the number of errors. Typical user error types are 'input time-out', 'false voice recognition', and 'incorrect result of system'.

Recently, smart devices based on voice recognition are used as a means of directly or indirectly controlling smart phones. At this time, the user's operation is subjected to an action-based procedure based on voice command [7]. That is, it is an important issue to accurately transmit the intention of the user to the system, and accordingly, whether the device is normally working. Therefore, it is necessary to make efforts to improve the accuracy of speech recognition, but it is also important to consider the control method that the user thinks.
Applying voice recognition on smart phones will gradually reduce the ratio of visual perception and direct manipulation with the finger. This interface design direction can be an input method to replace the application itself in the future. However, through the literature review, it seems that the overall usability improvement of the voice interface and the design considering the user experience are still necessary. The purpose of this study is to evaluate the possibility of voice recognition control based on the smart phone applications that users often use. In addition, the control method similar to the user’s mental model is investigated to utilize it for interface design for more natural function implementation.

2. Method

2.1 Research overview

We conducted a questionnaire survey to identify how users use the voice interface in smartphones and how to control them. This experiment was conducted through a three-step survey.

Google Forms platform was utilized to conduct surveys based on mobile responses as illustrated in Figure 1.

2.2 Participants

A total of 107 survey participants (85 males and 22 females) including university students, graduate students, and the general public participated in the questionnaires. The mean of participants’ age was 24.42 years.

All subjects were selected at people with experience using smart phones. The average using experience was 6.22 years. Before participating in the questionnaire, we checked whether there was any auditory or speech impairment associated with using the voice interface.

3. Results

The type of operating system (OS) in smart phone selected by survey participants was 62.6% (n=67) based on android, and 37.4% (n=40) based on i-Os. It seems that a relatively large number of smartphones equipped with android-based operating systems are being adopted.

The first survey results showed that mobile messenger, phone, web-browser, music, and social network service were those of frequently used applications.

As a result of the second questionnaire, functions that can be preferentially controlled by voice recognition were derived from the applications in the previous survey. The ratio of respondents’ priority for ‘texting’, ‘phone call’, ‘message reading’ and ‘internet browsing’ were 36%, 36%, 10%, and 6%, respectively. Figure 2 indicates the results of second survey.

In the third questionnaire, possible control procedures were analyzed based on the user expected functions through voice recognition. Table 1 shows possible voice recognition control procedures for ‘Texting’, ‘Phone Call’, ‘Message Reading’, and ‘Internet Browsing’ functions.
As a result of the possible control procedures analysis, we found that the pattern of the voice recognition control that the user expects is various in the case of 'Texting' and 'Phone Call'. On the other hand, 'Message Reading' and 'Internet Browsing' showed relatively few types of speech recognition control patterns. In other words, depending on the functional characteristics of the application, it can be seen that the control pattern expected by the user is different.

In order to understand the reason of this result, the level of the user requested function was evaluated with respect to the various voice recognition devices currently being released. In the case of 'Texting' and 'Phone Call', the implementation of speech recognition is high, but it is still applied only to some manufacturers' products. Therefore, it can be seen that the standardized control pattern of users is not yet constructed.

However, in the case of 'Internet Browsing', most of the products can be controlled by voice recognition and there is no significant difference in implementation degree. Therefore, it has been confirmed that most of users' control methods have a similar pattern. On the other hand, the 'Message Reading' function has a tendency of unified control pattern because the step that is expected from users' mental model is simple.

This study seeks to identify the control method which better match the users’ mental model through an example of the ‘Texting’ function with various control procedures of voice recognition. For this purpose, three representative control procedures were derived based on the frequency ranking from the data of the total survey respondents. Table 2 summarizes representative control procedures and detailed voice commands.

**4. Discussion and Conclusion**

Smart phones are still the key mobile devices that enable diverse user experiences. As the voice recognition interface is commercialized, the experience of recognizing and controlling information is evolving. In this technical trend, the possibility of substituting the voice recognition function through this study is the main contribution to related industries. As a result of the study, the user's control pattern varies in spite of the expeditiousness and familiarity provided by the voice interface.

In other words, there is clearly a limit to how all functions of the product can be controlled by voice recognition. Speech recognition is attractive to users, but the clarity of the actions and the feedback aspects of the response provided by the visual and touch interface are still forceful. Therefore, direction of future user interface design for smart device is expected to be realized by mixing touch and voice recognition.

When performing control through only voice recognition, only a simple answer can be provided from the device. However, providing both a touch and voice interface yields additional information and detailed feedback. Future smart phone should improve the overall quality of voice recognition and reflect the optimal level of control pattern desired by the user. Even AI speakers such as the Amazon Echo Show will lead a growing number of adoption of touch displays that allow users to view and control information. The

### Table 1. Possible voice recognition control procedures for various functions

| Requirement function of user | Procedure | Texting |
|-----------------------------|-----------|---------|
|                            | Inputting text → Identify(Visual) → Voice command(Send) | Calling → Voice command(Call) → Identify(Visual) → Voice command(Send) → Unlock(Touch) → Unlock(Pattern) → Voice command(Send) → Calling → Inputting text → Voice command(Send) |
| Phone call                  | Calling → Voice command(Search) → Identify(Visual) → Voice command(Call) → Calling → Voice command(Search) → Identify(Visual) → Voice command(Call) | Calling → Voice command(Search) → Calling → Voice command(Search) → Call the voice interface. → Ask voice interface to find 'Hong Kil Dong' on the contact list of the phone. → Tell the voice interface what message you want to send. → Confirming that the message is correct. → Ask voice interface to send the message. |
| Message reading             | Calling → Voice command(Message reading) | Call the messenger(ex. Kakao Talk) → Say ‘Send a message to a friend’ |
| Internet browsing           | Voice command(Search) → Identify(Visual) → Voice command(Select area) → Identify(Visual) → Voice command(Search) → Calling(Touch) → Voice command(Search) → Identify(Auditory) | Input text → Identify(Visual) → Input text using voice interface. → Identify inputting text by eyes. → Command to send message. |

### Table 2. Representative control procedures and detailed voice commands
boundary and distinction between the existing smartphone and AI speaker may become increasingly vague in terms of user experience.

This study determines the possibility of voice recognition control for smartphone applications through the back-end approach. 'Texting', 'Phone Call', 'Message Reading' and 'Internet Browsing' were selected as the voice control functions that can be preferentially applied based on the users' responses. In addition, the results of this study confirm that the control steps and procedures of the interface that users expect for the same function can be varied. The implementation method of the voice user interface for the same function is different according to the operating system provided by the manufacturer [8]. Therefore, it is necessary to develop a certain level of patterns or rules even though not one uniform control procedure exists in the nature of the voice interface.

In future research, it is required to find functional control patterns in accordance with user's mental model and to analyze user errors related to voice control. Moreover, it needs to develop a standardized control protocol for designing voice recognition interface products. These efforts expect to derive guidelines that are more user-friendly and can enrich the user experience.

References
[1] Jung, E.S., and Im, Y.: Touchable area: An empirical study on design approach considering perception size and touch input behavior, International Journal of Industrial Ergonomics, 49: 21-30, 2015
[2] Jeong, J., Jeong, S., Cho, H., and Shin, D.H.: A Study of Smartphone Voice-User Interfaces (VUI) for Non-Verbal Communication, proc. The HCI Society of Korea Conference, 2014
[3] Kang, S.S.: AI Assistant Apple’s Siri and Wolfram Alpha, Communications of the Korean Institute of Information Scientists and Engineers, 35(8): 42-50, 2017
[4] Yang, H., and Kim, D.: Intelligent personal secretary market trend and domestic industry influence forecast, Trend and Issues, 35: 1-30, 2017
[5] Choe, J. and Kim, H.: Survey Study on the Utilization Status and User Perception of the VUI of Smartphones, Journal of Society for e-Business Studies, 21(4): 29-40, 2017
[6] Koh, Y.: Analysis on Smartphone Voice User Interface Error Type and User’s Speech Behavior Pattern after Correction: Focusing on iPhone Siri Users, Master’s Thesis, Graduate School of Convergence Science and Technology in Seoul National University, 2015
[7] Roh, J.S.: A study on the device control using voice recognition function, proc. KICS Summer Conference, 1280-1281, 2017
[8] Lee, H.M., and Kim, S.I.: Comparative Study on the usability of mobile intelligent personal assistance service based on voice recognition technology: Focused on Samsung S Voice and Apple Siri, Journal of Digital Design, 14(1): 231-240, 2014