Analysis of User's Speech Behavior Pattern after Correction: focusing on Smartphone Voice User Interface

Youngin Koh (theophila@postech.ac.kr), Jiyoung Kwahk (kgy@postech.ac.kr)

Department of Industrial and Management Engineering, POSTECH, Pohang, 37673

ABSTRACT

Voice interaction between users and IT devices has become possible through voice user interface. Since VUI is based on vocabulary language, the most basic form of human communication, it lowers the entry barrier for those who experience difficulty using graphic user interface (GUI) or situations where using IoT devices with tiny screen or no screen at all. User experience can be a critical marker for the development of technology, however, few academic researchers have attempted to investigate the users' speech behavior pattern when they use VUI. Therefore, this study aims to provide insight about guidelines and design feedback for the novice users through observation in their adaption process of VUI. The author investigated speech patterns of error correction through observing 'iPhone Siri' users who have less experience on VUI. Through the 7 days Siri usage history, the speech behavior patterns of error correction were analyzed in tone (amplitude, pitch, and duration), and in sentences (sentence elements or structure) based on the Communication Accommodation Theory. The results of this study can be summarized as follow; 1) After the first error, users change their speech tone and sentences at a similar rate, 2) but, when the second error occurred, they change sentences more, 3) the speech behavior patterns after error were different depending on the types of errors and 4) the users' speech behavior patterns were changed depending on time period of use. This study has primarily focused on variation in use behavior by observing real user experience. Through demonstrating how users respond to the situations of errors, this research has practical implications for the industry as it can positively contribute to the development of VUI voice feedback design and improve user experience on VUI.

Keywords: Voice User Interface, Communication Accommodation Theory, User's Speech Behavior Pattern after Correction, Siri

1. Introduction

As the technology of speech recognition and AI developed, voice interaction between user and IT device has become possible through voice user interface (VUI) (Marshall et al., 2013; Sharma et al., 2001; Kortum and Philip., 2008). Since VUI is based on vocabulary language, the most basic form of human communication, it lowers the entry barrier for those who experience difficulty using graphic user interface (GUI) or situations where using IoT devices with tiny screen or no screen at all (Sears et al., 2003; Macek et al., 2013). However, VUI users are constantly complaining about its voice recognition error, response delay, etc. These complaints make users to stop use VUI (Munteanu et al., 2014; Schaffer et al., 2012). By the way, when the users encounter an error condition, they change their speech pattern. For example, the users speak only keywords, change the order of words, and use other commands with similar meanings. Moreover, the users change their tone like speak slowly or higher their pitch.

Since user experience is a critical marker for the development of technology (Damopoulos et al., 2013; Unger et al., 2012), it is also important to study VUI users' speech pattern changes before and after errors. Previous studies, however, have mostly focused on studying a specific group of users such as the visually impaired or the elderly (Schlögl et al., 2013; Zhong et al., 2014), or short-term studies that only conducted interviews or surveys for a specific condition of using the VUI (Peissner and Matthias, 2002; Macek et al., 2013). Only few academic researchers have attempted to investigate the users' speech behavior pattern when they use VUI. Therefore, this study aims to provide insight about guidelines and design feedback for the novice users through observation in their adaption process of VUI. The following research questions were set up for this study.

1. *How VUI users' speech pattern changes before and after errors?*
2. *Does the user's speech pattern change differently depending on type of error?*
3. *Does the user's speech pattern change over time?*

2. Related Study

2.1 Communication Accommodation Theory

Adults who communicate with children speak like a child. Burnham et al. (2002) studied this phenomenon and figure out that when people talk to children, they simplify speech, grammar, and rhyme, unlike talking to other adults (Uther et al., 2007). A similar phenomenon can be observed in the situation of conversation with a foreigner using native language. When a foreigner is immature in native speech, the native speaker speaks more clearly and speaks more slowly than usual (Scarborough et al., 2007). This phenomenon can be explained by the Communication Accommodation Theory(CAT) (Giles et al., 1991). According to the theory, a speaker adjusts their language pattern to improve the efficiency of conversation according to the potential language ability of the other. However, the phenomenon not only occurs in the situation of conversation between human, but also when the voice command is given to the machine (Burnham et al., 2010). These shows that the communication accommodation can be observed in interaction with non-human subjects. Therefore, I conduct the VUI user's speech pattern after error correction based on CAT.

Table 1. Speech behavior pattern after error correction

Tone and Sentence Maintenance	No change on Tone or Sentence
Change of Tone	Amplitude
	Pitch
	Duration
Change of Sentence	Sentence Element
	Sentence Structure

According to CAT and related studies, I defined the speech behavior pattern after error correction in three levels: 1) the same voice and sentence, 2) the change of the tone, and 3) the change of the sentence, as in the table 1. The change of tone is defined as the change of amplitude, pitch and duration of voice. The change of sentence is defined as the change of element or structure of sentence. Also, the same tone and sentence, means that the user speaks the same sentence after error correction and the tone change is within 5%.

2.2 The types of errors

Schelle-Walka and Dirk (2010) explained that the cause of errors in voice interaction is caused by speech recognition. However, the errors, caused by voice recognition, are not sufficient to explain all the errors that occur when using the VUI. On the other hand, Duff et al. (1996) classifies the types of errors that can occur in voice interaction from level 0 to 7, which is more appropriate to explain the error of VUI than Schelle-Walka's study. However, there was a limit that each level of error occurs at same state, so the meaning of level is ambiguous. Therefore, in this study, I distinguish the states of error occurrence in VUI. In human computer interaction, interaction states were defined into 4 states, which are articulation, performance, presentation, and observation (Dix, 2009). From these states, I have modified it to fit the VUI. And those new states are as follows: 1) the input state in which the user enters a voice command; 2) the recognition state in which the voice command is recognized; and 3) the result state in which the result of voice command is shown.

3. Method

13 iPhone users who had little experience using smartphone VUI were recruited for 3 weeks (April - May, 2015) as the subjects. In addition, they are aged 20 to 30, graduate students using the standard language, and live in Seoul and Gyeonggi-do. Experimental data was collected by video recording the smartphone usage screen and recording the user's voice. The subjects performed 5 tasks each day. The 5 tasks are: a) asking for tomorrow's weather, b) setting an alarm for tomorrow, c) entering an appointment on a calendar, d) asking for directions to a nearest restaurant e) sending a text message.

To prevent data contamination due to the external environment, it was necessary to create controlled and similar environment. Therefore, the subjects performed the task with the researcher in laboratory on the 1st, 4th, and 7th day of experiment. Except the 3 days, they performed the tasks freely in daily life. The subjects sent text message to the researcher in the last task, confirming that he performed the daily task. Also, they recorded their task usage screen by themselves.

4. Results

91 video data were collected. The user's speech input, and result of the input recognition were confirmed through the video. In addition, 39 experimental voice data of the subjects were collected in the laboratory for 3 days. I analyzed the error type and the change of user's speech behavior through video and audio as shown in Figure 1.

Mission	Date	Screen result	Usage Result			error type	Diff
			input	recognition	result		
b	Set the alarm tomorrow	1st	“내일 아침 열시 알람” (Tomorrow morning alarm at ten o'clock)	“내일 아침 10시 알람” (Tomorrow morning alarm at 10 o'clock)	“오전에 설정된 알람이 없습니다” (No alarm set in the morning)	input 1 recognition2 result 1	Sentences

Figure 1. The example of analyzed video and audio data

4.1 Analysis of user's speech behavior pattern after error correction between 1st and 2nd error occur

Figure 2. User's speech behavior pattern after error correction between 1st and 2nd error occur

As Figure 2, after the first error, 42.9% of the VUI users change the tone, 55.7% change the sentence. The ratio of maintaining tone and sentence is very low, at 1.4%. The probability of occurring the second error is 16.7% when the tone is changed, and 53.8% when the sentence is changed. The error occurs at a higher rate when the sentence is changed. After the second error, 80% of the subjects changed the sentence, and it is higher rate than those after the first error. Although the error occurs at higher rate when they change sentences, they change sentence more after 2nd error occur.

4.2 Comparison of speech behavior according to error types

Among the 802 errors of the VUI during the experiment, the three most frequently experienced errors were summarized as type 1, 2, and 3. Type1 error is an error that the user can not recognize the correctness of the input command because the command is not recognized. Type2 error is the case where the speech recognition is wrong, and type3 error is the case where the speech recognition is correct but the result is wrong.

Figure 3. User's speech behavior pattern after error correction between error types

As figure 3, after the error of type 1, the subject showed a tendency to change the tone while maintaining the sentence. On the other hand, after type 2 and 3 error, subjects changed sentences mainly. This is because the user recognizes that the vocabulary that the user uttered is difficult to recognize by Siri or Siri cannot provide the desired result from the command.

4.3 Comparative analysis of speech behavior patterns by date of use

As Table 2, tone and sentences were maintained only after the first error of day 1. On the 1st and 4th day, the rate of changing the tone or sentence was similar, but on

Table 2. The result of comparison of speech behavior patterns by date of use

Speech Behavior Pattern	After 1 st Error		
	1 st Day	4 th Day	7 th Day
Tone and Sentence Maintenance	2.6%	0%	0%
Change of Tone	46.2%	45%	27.3%
Change of Sentence	51.3%	55%	72.7%

the 7th day, the rate of changing the sentence was increased to 72%.

5. Conclusion

The VUI user's speech patterns were analyzed by the number of errors, the types of errors, and the date of use. The results of the analyze can be summarized as follow; 1) After the first error, users change their speech tone and sentences at a similar rate, 2) but, when the second error occurred, they change sentences more, 3) the speech behavior patterns after error were different depending on the types of errors and 4) the users' speech behavior patterns were changed depending on time period of use. Although this study has limited data to generalize the result, it has practical implications for the industry. It can positively contribute to the development of VUI voice feedback design and improve user experience on VUI.

Acknowledgements

This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. NRF-2016R1A2B2011158)

References

- Marshall, Joe, and Paul Tennent. "Mobile interaction does not exist." CHI'13 Extended Abstracts on Human Factors in Computing Systems. ACM, 2013.
- Sharma, Chetan, and Jeff Kunins. Voicexml: Strategies and Techniques for Effective Voice Application Development with Voicexml 2.0 with Cdrom. John Wiley & Sons, Inc., 2001.
- Kortum, Philip. HCI beyond the GUI: Design for haptic, speech, olfactory, and other noadditional interfaces. Morgan Kaufmann, 2008.
- Sears, Andrew, et al. "Hands-free, speech-based navigation during dictation: difficulties, consequences, and solutions." Human-computer interaction 18.3 (2003): 229-257.
- Macek, Tomáš, et al. "Mostly passive information delivery in a car." Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications. ACM, 2013.
- Munteanu, Cosmin, et al. "Designing speech and language interactions." CHI'14 Extended Abstracts on Human Factors in Computing Systems. ACM, 2014.
- Schaffer, Stefan, and Michael Minge. "Error-prone voice and graphical user interfaces in a mobile application." Speech Communication; 10. ITG Symposium; Proceedings of. VDE, 2012.
- Damopoulos, Dimitrios, et al. "User privacy and modern mobile services: are they on the same path?." Personal and ubiquitous computing 17.7 (2013): 1437-1448.
- Unger, Russ, and Carolyn Chandler. A Project Guide to UX Design: For user experience designers in the field or in the making. New Riders, 2012.
- Schlögl, Stephan, et al. "Exploring voice user interfaces for seniors." Proceedings of the 6th International Conference on Pervasive Technologies Related to Assistive Environments. ACM, 2013.
- Zhong, Yu, et al. "JustSpeak: enabling universal voice control on Android." Proceedings of the 11th Web for All Conference. ACM, 2014.
- Peissner, Matthias. "What the relationship between correct recognition rates and usability measures can tell us about the quality of a speech application." Proceedings of the 6th International Scientific Conference on Work With Display Units. 2002.
- Macek, Tomáš, et al. "Mostly passive information delivery in a car." Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications. ACM, 2013.
- Burnham, Denis, Christine Kitamura, and Uté Vollmer-Conna. "What's new, pussycat? On talking to babies and animals." Science 296.5572 (2002): 1435-1435.
- Uther, Maria, Monja A. Knoll, and Denis Burnham. "Do you speak E-NG-LI-SH? A comparison of foreigner-and infant-directed speech." Speech Communication 49.1 (2007): 2-7.
- Scarborough, Rebecca, et al. "An acoustic study of real and imagined foreigner-directed speech." Journal of the Acoustical Society of America 121.5 (2007): 3044.
- Giles, Howard, Nikolas Coupland, and I. U. S. T. I. N. E. Coupland. "1. Accommodation theory: Communication, context, and." Contexts of accommodation: Developments in applied sociolinguistics 1 (1991).
- Burnham, Denis, Sebastian Jeffry, and Lauren Rice. "Computer-and human-directed speech before and after correction." space 6 (2010): 7.
- Schnelle-Walka, Dirk. "A pattern language for error management in voice user interfaces." Proceedings of the 15th European Conference on Pattern Languages of Programs. ACM, 2010.
- Duff, David, Barbara Gates, and Susann LuperFoy. "Spoken dialogue management." Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on. Vol. 2. IEEE, 1996.
- Dix, Alan. Human-computer interaction. Springer US, 2009.