



A Method for Increasing User Engagement with Voice Assistant System

Daehee Park^(✉), Heesung Park, and Scott Song

Samsung Electronics, 56, Seongchon-gil, Seocho-gu, Seoul, Republic of Korea
daehee0.park@samsung.com

Abstract. Voice interaction is becoming one of the major methods of human interaction with computers. Several mobile service providers have introduced various types of voice assistant systems such as Bixby from Samsung, Siri from Apple, and Google Assistant from Google that provide information including the schedule for a day, the weather, or methods to control the device to perform a task such as playing music. Although the voice assistant system provides various types of functions, generally, the users do not understand what functions the system can support. We conducted a control task analysis based on expert interviews and found that the main bottleneck of using a voice assistant system is that the user cannot know all the commands. Thus, we believe that presenting recommendable commands is an effective way to increase user engagement. Through buzz data analysis, we discovered what functions could be used and determined the relevant usage context. Subsequently, we performed context modelling and designed the user interface (UI) of a voice assistant system and conducted a case study. Through this case study, we proved that presenting commands on a UI induced more user engagement and usability.

Keywords: Voice interaction · Voice assistant system · Human-computer interaction · User engagement

1 Introduction

1.1 The Trends of Voice Interaction

Recently, voice interaction has become one of the major methods for people to interact with computers. According to Fortune Business Insights, the global speech and voice assistant system market size was valued at USD 6.9 billion in 2018, and it is anticipated to reach USD 28.3 billion by 2026. It indicates that the compound annual growth rate for the forecast period will be 19.8% [1]. Several mobile service providers have introduced various types of voice assistant systems like Bixby from Samsung, Siri from Apple, and Google Assistant from Google that provide information including the schedule for a day, the weather, or methods to control the device such as playing music. Google announced in 2014 that among teens 13-18, 55% use voice assistant systems every day, and 56% of adults said that using these systems makes them “feel tech-savvy” [2]. Business Insider presented the shares of the market in voice assistants: Apple Siri possesses 45.6%, Google Assistant possesses 28.7%, Amazon Alexa possesses 13.2%, Samsung Bixby possesses 6.2%, and Microsoft Cortana possesses 4.9%

[3]. Kiseleva et al. (2016) indicated that smart voice assistants are increasingly becoming a part of the daily lives of users, in particular, on mobile devices [4]. They introduce a significant change in how information is accessed, not only by introducing voice control and touch gestures, but also by enabling dialogues where the context is preserved. Significant change in information access, not only by introducing voice control and touch gestures but also by enabling dialogues where the context is preserved.

However, voice systems cannot recognise many types of natural languages; currently, they only provide limited functions based on recognised user voice inputs. Generally, the users do not know all of the functions the voice assistant system can support. Additionally, the users give commands with short words or phrases to these systems as they do not know what the system will accept. The user will try to use the voice assistant system several times but will easily give up when they cannot fulfil their purpose. If the voice assistant system understands all commands naturally from the users and all functions of the mobile phone can be supported through voice, the system does not need a screen to present possible commands. However, current systems only provide limited functions to the users; thus, the voice assistant systems equipped on the mobile phones try to support the users by providing recommend voice commands on the screen.

The contents of the first screen of a system can present instructions on how to use it. Thus, we can consider that the first screen of the voice assistant system is regarded as a first step to use the voice assistant system, and it determines the user's engagement. Some voice assistant systems only provide a message such as "What can I help you with?" Whereas, some voice assistant systems present a home screen consisting of recommended commands.

To guide the users on how to use a voice assistant system, only presenting a few possible commands is not sufficient. We hypothesised that recommending commands based on context would induce higher user engagement for voice interaction. To design a context-based voice assistant system, we need to find out the purpose of using a voice assistant system, examine a variety of usage scenarios, and determine usage bottlenecks.

However, we could not collect data, which include an individual user's voice assistant system usage due to limited resources and privacy issues. Thus, we analysed buzz data regarding two current voice assistant systems, Google Assistant and Apple Siri, collected during a specific period (90 days prior from 9 December 2019). In addition, we conducted cognitive work analysis based on the result of buzz analysis, to analyse the tasks of the users using these voice assistant systems in detail. Then, determined a data index for context-awareness through analysis of the purpose of using voice assistant systems, usage scenarios, and usage bottlenecks. The home screen had to present example context-based options that were consistent with the purpose required by the users to increase user engagement. After developing the data index for context-awareness, we developed a prototype that applied the context-aware voice assistant. Next, we measured the user engagement of each voice recognition system through a modified user engagement scale (UES) [5]. We validated the effectiveness of the context-aware voice assistant system by measuring the UES.

Although there is some research regarding user satisfaction while using voice assistants, there is still a research gap because there is little research regarding user engagement using voice assistants. In addition, designing a context-aware voice assistant has not yet been researched. Thus, we propose three main objectives of this paper. The first objective is to provide a comprehension of user behaviours based on the buzz analysis, i.e., how the users use the voice assistant system. Then, we provide context modelling for easy use of the voice assistant system. Finally, we validate the effectiveness of the context-aware assistant system by measuring the UES.

2 Related Works

According to Burbach et al. (2019), voice assistant systems have been one of the major changes in user interaction and user experience related to computers recently [6]. They already support many tasks such as asking for information, turning off the lights, and playing music, and are still learning with every interaction made with the users.

The study completed by Burbach et al. (2019) presented acceptance of relevant factors of virtual voice-assistants. They insisted that although individual users frequently use voice assistant systems in their everyday lives, their use is currently still limited. Currently, voice assistant systems are mainly used to call people, ask for directions, or search the Internet for information. One of the reasons for such limited use is that automated speech recognition can cause users to be dissatisfied if errors occur. Another reason is interactions between users and voice assistant systems are more complex than web searches. For example, the voice assistant systems must be able to comprehend the user's intention and context so that they can choose the proper action or provide a proper answer. Burbach et al. (2019) conducted a choice-based-conjoint analysis to find out factors, which influence acceptance of voice assistant systems. The method aimed to study consumer choices or preferences for complex products. Their study presented that "privacy" was the most important factor in the acceptance of voice assistant systems rather than "natural language processing-performance" and "price." It means that the participants did not want the assistant always to be online; it would be expected that they would rather reject this option.

In this study, we believe that the first screen design of the voice assistant systems influences the user engagement. One of the first screens of the voice assistant system should consist of recommendable commands, and we believe that if it can become one of the factors for acceptance of voice assistant system, then it may influence the level of engagement.

3 An Analysis of Using Voice Assistant Systems

3.1 Buzz Data Analysis

We hypothesised one of the important bottlenecks of using voice assistant system is that the user does not know the supported functions and command words because the voice assistant system can neither recognise all natural languages nor provide all

functions of the mobile phone via voice. This indicates that learnability is one of the improvement points in the voice assistant system. We tried to analyse the usability issues of voice assistant systems. Rather than collecting user data or performing a user survey, we used big data to find UX insights in the voice assistant system as well as to discover users' natural opinions regarding the voice assistant systems. Various kinds of data are currently collected to find out new trends in the market. According to Lavalle et al. (2011), recent studies of predictive analytics are not limited to specific areas, such as marketing or customer management, due to the growth of various analytical data [7]. Data analysis performed online has expanded to financial management and budgeting, operations and production, and customer service.

Samsung Electronics has a big data system called BDP; it has collected a variety of buzz data from social network systems such as Twitter, Facebook, Instagram, news, and other forums. Through the BDP, we can collect English written buzz data about Google Assistant and Apple Siri from the 9th of December 2019 to 90 days prior. The system can categorise whether buzz data is positive, neutral, or negative. Table 1 indicates that negative posts comprised 17% of buzz data regarding Google Assistant. Table 2 indicates that negative posts comprised 38% of buzz data regarding Apple Siri. We hypothesised that presenting only "What can I help you with?" is not sufficient to guide users; thus, there were more negative posts about using Apple Siri. We will discuss the context of using voice assistant systems through work domain analysis (WDA) and the voice assistant system users' decision-making style through Decision Ladder based on some expert interviews in Sect. 3.3.

Table 1. Buzz data analysis for Google Assistant (last 90 days from 9th/Dec/2019).

	Number of buzz
Total	54790
Positive posts	8088 (15%)
Neutral posts	37387 (68%)
Negative posts	9315 (17%)

Table 2. Buzz data analysis for Apple Siri (last 90 days from 9th/Dec/2019).

	Number of buzz
Total	100000
Positive posts	23383 (23%)
Neutral posts	38346 (38%)
Negative posts	38271 (38%)

3.2 Work Domain Analysis Based on Buzz Data

We analysed buzz data specifically pertaining to Google Assistant and Apple Siri to determine various contexts of using voice assistant systems. The result of WDA indicates that the users normally use these voice assistant systems for several different events as detailed in Fig. 1. In addition, their environment can influence events.

According to the buzz data, the users use the voice assistant systems to control light, air conditioners, or even set an alarm at night. The users regularly send e-mails, set a schedule, and receive weather information and news through their voice assistant systems. While driving, users use it for calling, playing music, finding information about songs, controlling functions of the vehicle, and setting a destination on their navigation system. Through the WDA, we can assume that time and location information can be an important factor to recognise the context of using the voice assistant system. Additionally, GPS data can also be regarded as a significant factor because it can recognise where the user is driving.

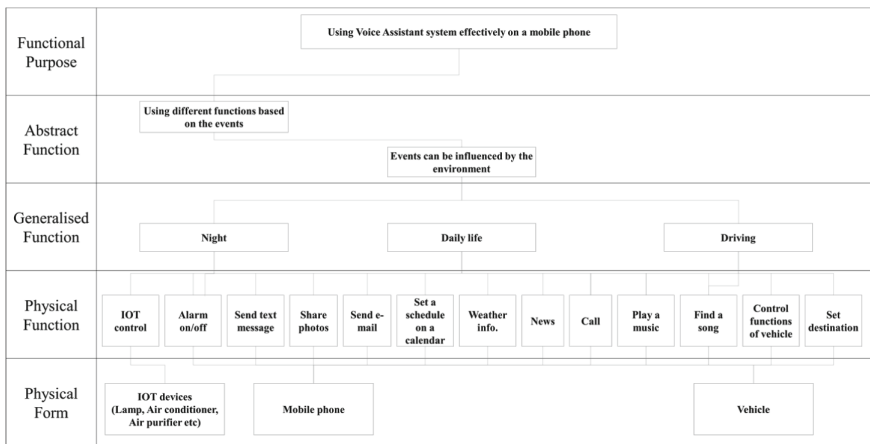


Fig. 1. Work domain analysis of the contexts of using voice assistant systems.

3.3 Control Task Analysis Based on Expert Interview

We recruited four experts in voice assistant systems, and then we conducted in-depth interviews to analyse how users control their voice assistant system. Based on the results of these in-depth interviews, most experts indicated that many potential voice assistant systems users could not overcome the first stage of system use; the user thinks about how to say the command and cannot get any further. In the early days, when voice recognition came on the market, the main issue of voice assistant systems was the recognition rate. Although the recognition rate has improved, the main bottleneck of using the voice assistant system is still that the users do not know what functions the system can support because the voice assistant systems only present one sentence; “What can I help you with?” (Fig. 2).

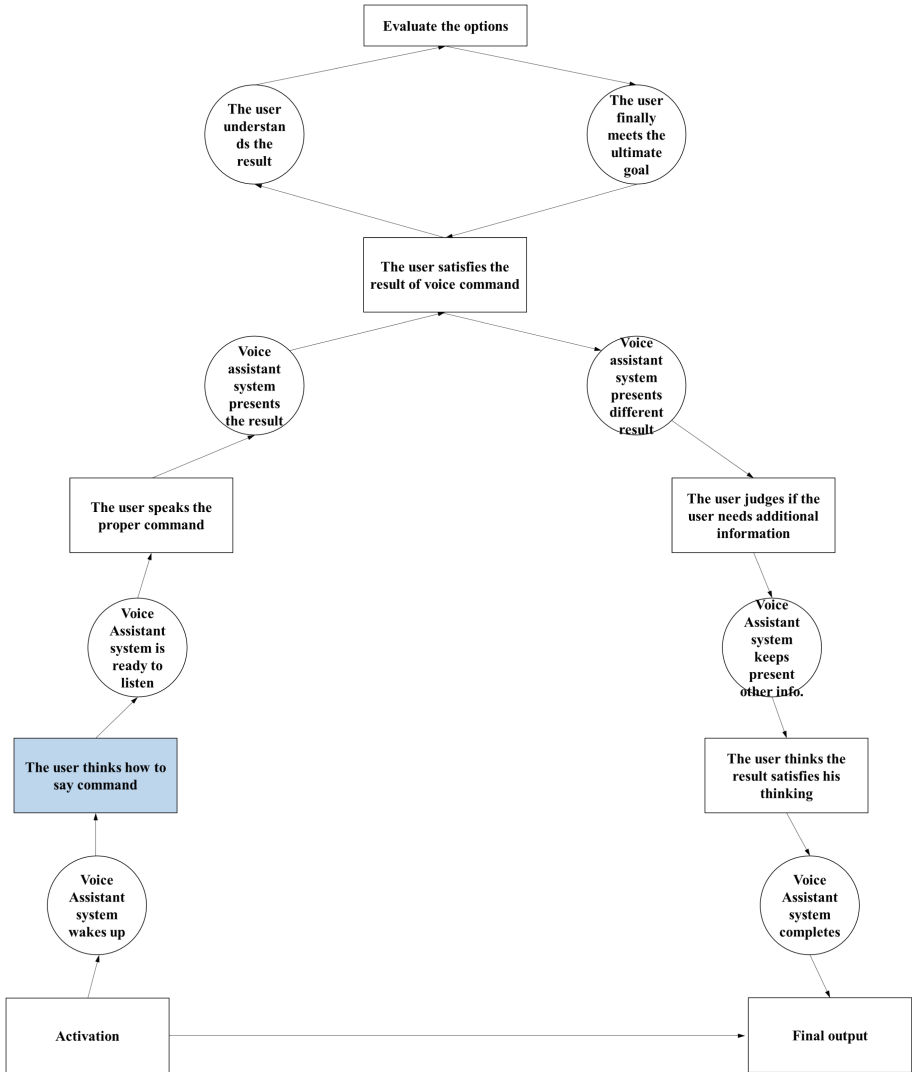


Fig. 2. Decision ladder of using voice assistant systems.

Thus, the experts argued that only presenting one sentence is not a sufficient guideline for users because they could not know the functions of voice assistant systems without being given more information. They try to use their voice assistant several times, and unsuccessful trials cannot lead to continuous usage.

4 A Suggestion of Context-Based Design

4.1 Context Modelling for the Voice Assistant System

Dey and Abowd (1999) insisted that context must be regarded as an important factor for interactive applications [8]. In the field of mobile computing, some information such as user's location, nearby users and their devices, time of day, and user activity can be manipulated to improve latency time or boost the communication signal by using bandwidth from nearby devices (Chen and Kotz, 2000) [9]. In this study, we suggest user context modelling of using the voice assistant system based on the buzz analysis.

The buzz data analysis identifies that the users use the voice assistant system differently according to the time of day. In general, at home, the users want to start the morning with useful information; thus, they want to listen to weather, news and get daily briefing information from the voice assistant system. On the other hand, at night, the users prepare to sleep. The relevant commands analysed include controlling IoT devices, setting the alarm, and setting a schedule on a calendar. During the daytime, the user uses a voice assistant system for application behaviours that operate via touch interaction. For example, they may send e-mail, share photos, send text messages, etc. When the user is walking or driving a car, the user's context changed. When the user is walking, opening the map application and voice memo applications generally are used to avoid uncomfortable touch interaction while walking. While driving a car, the user's vision should focus on scanning the driving environment, and the user's hands should hold the steering wheel; the user feels uncomfortable when they need to interact with a touch device due to driver distraction. The buzz data analysis described that calling, setting a destination, and playing music are regarded as the higher priority functions of the voice assistant system while driving. Furthermore, the voice assistant system can recognise the user context through the sound input. If the system determines that music is playing, it can provide song information. Alternatively, if the system determines that the user is in a meeting, it can suggest recording voices.

The quantity of buzz data has arranged the order of commands in each context. Based the context modelling from the buzz data, we designed a user interface (UI) for the first screen of the voice assistant system. The UI presents several recommendable commands, which are consistent with the user's context and are easily accessible. In the next section, we will evaluate how the suggested UI influences the user engagement of the voice assistant system (Fig. 3).

Category	Home			Outside		If input data available
	Morning	Day-time	Night	Walking	Driving	Sound
Recommended functions	Weather info	Send e-mail	IOT control	Open map	Call	Find a song
	News	Share photos	Alarm on/off	Voice Memo	Set a destination	Voice Memo
	Daily Briefing	Send text messages	Set a schedule on a calendar		Play a music	
		Set a schedule on a calendar	Call		Find a song	
		Call	Send text messages		Control functions of vehicle	
			Send e-mail		Voice memo	
			Share photos			

Fig. 3. Context modelling based on the buzz data analysis

5 Is Presenting Recommendable Commands Better?

To design a more effective voice assistant system, we conducted expert interviews to determine the bottleneck of using a voice assistant system. Previously, we discussed expert interviews; the result indicated that the users do not properly use their voice assistant systems because they do not know the functions of the system. Thus, we designed the UI of the proposed voice assistant system consisting of recommendable commands. For suggesting recommendable commands, we developed context modelling based on the buzz data analysis. We believe that a voice assistant system, which involves recommendable commands can create more user engagement than current voice assistant systems, which only involve a guide sentence. Thus, through the experiment, we tried to compare two different UI screens in the level of engagement; the first one presents recommendable screen, the second one only presents “What can I help you with?” (Fig. 4)

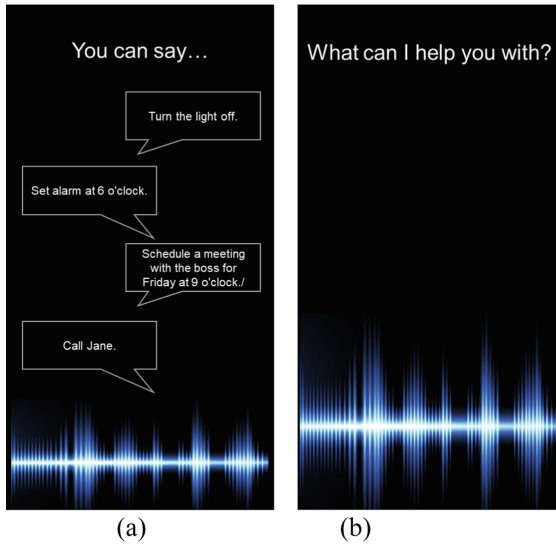


Fig. 4. First screen of each type of voice assistant system: (a) Recommending commands; (b) Presenting a guide sentence

5.1 Method

We measured user engagement regarding the two different types of voice assistant systems. According to O'Brien (2016a), user engagement is regarded as a quality of user experience characterised by the depth of user's cognitive, temporal, affective, and behavioural investment when interacting with a specific system [10]. In addition, user engagement is more than user satisfaction because the ability to engage and sustain engagement can result in outcomes that are more positive. To measure user

- FA-S.1 I lost myself in this experience.
- FA-S.2 The time I spent using this voice assistant system just slipped away.
- FA-S.3 I was absorbed in this experience.
- PU-S.1 I felt frustrated while using this voice assistant system .
- PU-S.2 I found this Voice Assistant System confusing to use.
- PU-S.3 Using this Voice Assistant System was taxing.
- AE-S.1 This Voice Assistant System was attractive.
- AE-S.2 This Voice Assistant System was aesthetically appealing.
- AE-S.3 This Voice Assistant System appealed to my senses.
- RW-S.1 Using Voice Assistant System was worthwhile.
- RW-S.2 My experience was rewarding.
- RW-S.3 I felt interested in this experience.

Fig. 5. Questions of UES from O'Brien's study (2018)

engagement, the UES, specifically UES-SF provided by O’Brien et al. (2018) was used in this experiment. Questions of UES consist of several dimensions, focused attention (FA), perceived usability (PU), aesthetic appeal (AE), and rewarding (RW) (Fig. 5) [5].

5.2 Experimental Design and Participants

The within-subject design was used to measure two kinds of voice assistant systems. The order of the voice assistant systems presented was random. The independent variable was each UI screen of the voice assistant system. The dependent variable was the score of UES. A total of 24 participants were recruited, ranging in age from 26 to 46 years (mean age of 34.4 years). They were instructed about the experiment first; then they answered the UES questions about each voice assistant system.

5.3 Results

Figure 6 presents the average UES score of each voice assistant system. The total UES score of the “Presenting recommendable commands” voice assistant system was 46.09, out of a maximum score of 60. The total UES score of the “Presenting guide sentence” voice assistant system was 26.09. Through statistical analysis, there was a significant difference between the two systems ($p < 0.05$). In the aspect of dimension, the average UES score of the “Presenting recommendable commands” was significantly higher than “Presenting guide sentence.” It describes that “Presenting recommendable commands” voice assistant system induces more user attention, usability, aesthetic appeal, and rewarding than the voice assistant system which only presents a guide sentence.

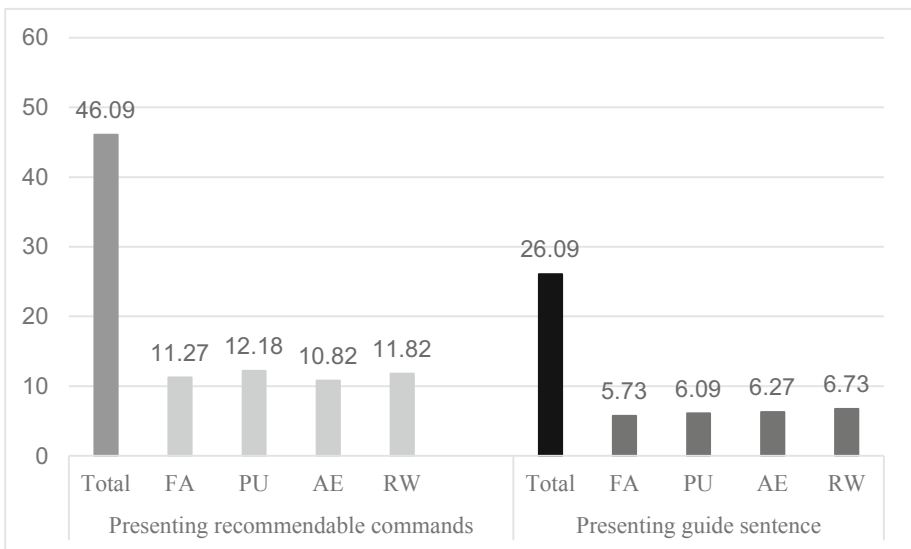


Fig. 6. Bar graph of user engagement score for each voice assistant system

6 Discussion, Limitations, and Future Work

The result indicated that most participants chose the version of the voice assistant system, which can present recommended commands. It means that the voice assistant system presenting recommended commands led to more user engagement than the voice assistant system that only includes a guide sentence. In the aspect of dimensions, the recommended commands UI had the greatest effect on usability. The score of perceived usability was relatively higher than in other dimensions. However, the score of perceived usability in other systems was not higher than other dimensions. Presenting recommended commands led to a higher level of usability, and it induced more user engagement. Therefore, the result described that presenting commands helped the users to learn how to use the voice assistant system, and it led to higher user engagement.

Several studies are needed for further research. First, in this study, we developed context modelling based on the buzz data as we could not collect personal data. The buzz data consists of user's preference functions. The context modelling was developed based on events and functions that were generally more useful for many users rather than individual data analysis. Defining individual context and recommending function is the first challenge. The voice assistant system cannot listen the user's voice every time due to privacy issues. Burbach et al. (2019) suggested that "privacy" is the most important factor in using a voice assistant system. It means that voice assistant system users do not want the system to continue to listen to their voices. Thus, the system only can check voice data when it is activated through the user's intention such as pushing the activation button or saying an activation command. At that time, the problem is that the system cannot analyse a novice user's context. In addition, if the period of using the system is short, the data must be sufficiently cumulated. We believe that if the user tries to use the voice assistant system for the first time, our context modelling based on the buzz data might be effective. However, for experienced users, the voice assistant system should sufficiently possess enough individual data to present recommended commands based on user context. The second challenge is presenting hidden functions to the users. The voice assistant system cannot present all functions on the first screen. Only two or three functions can be presented to avoid the complexity of the UI. This limitation requires further study in the future.

7 Conclusion

In recent years, voice assistant systems have changed user interaction and experience with computers. Several mobile service providers have introduced various types of voice assistant systems like Bixby from Samsung, Siri from Apple, and Google Assistant from Google that provide information such as the schedule for a day, the weather, or methods to control the device such as playing music. Although the voice assistant systems provide various kinds of functions, generally, the users do not know what functions the voice assistant system can support. Rather than collecting user data or performing a user survey, we used big data to find UX insights in the voice assistant system as well as to discover users' natural opinions regarding voice assistant systems.

Through the Samsung Big Data Portal system, we collected English buzz data about Google Assistant and Apple Siri. The system categorised whether the data is positive, neutral, or negative. The proportion of negative opinion of Apple Siri is higher than that of Google Assistant. We hypothesised presenting only “What can I help you with?” is not a sufficient guide for users. We discussed the contexts of using voice assistant systems through WDA and the voice assistant system users’ decision-making style through a control task analysis based on some expert interviews. The results of the control task analysis described that the main bottleneck of using the voice assistant system is the user cannot know all of the useful commands. Thus, we believed that presenting recommended commands is an effective way to increase user engagement. Through the buzz data analysis, we could discover what functions can be used and the usage contexts. Hence, we performed context modelling, designed the UI of a prototype voice assistant system, and conducted a case study. Through the case study, we proved that presenting commands on the UI induced more user engagement and usability. However, a method of finding out user context based on individual data and presenting hidden functions will be the subject of future work.

References

1. Fortune Business Insights. <https://www.fortunebusinessinsights.com/industry-reports/speech-and-voice-recognition-market-101382>. Accessed 09 Jan 2020
2. Google Inc. <https://www.prnewswire.com/news-releases/teens-use-voice-search-most-even-in-bathroom-googles-mobile-voice-study-finds-279106351.html>. Accessed 09 Jan 2020
3. Business Insider. <https://www.businessinsider.com/siri-google-assistant-voice-market-share-charts-2018-6>. Accessed 09 Jan 2020
4. Kiseleva, J., et al.: Understanding user satisfaction with intelligent assistants. In: Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval, pp. 121–130. ACM (2016)
5. O’Brien, H.L., Cairns, P., Hall, M.: A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. *Int. J. Hum Comput Stud.* **112**, 28–39 (2018)
6. Burbach, L., Halbach, P., Plettenberg, N., Nakayama, J., Ziefle, M., Valdez, A.C.: “Hey, Siri”, “Ok, Google”, “Alexa”. Acceptance-relevant factors of virtual voice-assistants. In: 2019 IEEE International Professional Communication Conference. IEEE (2019)
7. LaValle, S., Lesser, E., Shockley, R., Hopkins, M.S., Kruschwitz, N.: Big data, analytics and the path from insights to value. *MIT Sloan Manag. Rev.* **52**(2), 21–32 (2011)
8. Dey, A.K.: Understanding and using context. *Pers. Ubiquit. Comput.* **5**(1), 4–7 (2001). <https://doi.org/10.1007/s007790170019>
9. Chen, G., Kotz, D.: A survey of context-aware mobile computing research. Dartmouth Computer Science Technical report TR2000-381 (2000)
10. O’Brien, H.: Theoretical perspectives on user engagement. In: O’Brien, H., Cairns, P. (eds.) *Why Engagement Matters*, pp. 1–26. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-27446-1_1