TJE Thai Journal of Ergonomics

Thai Journal of Ergonomics: Volume 3 Issue 2 July – December 2020: p. 32-44
วารสารการยศาสตร์ไทย: ปีที่ 3 ฉบับที่ 2 กรกฎาคม–ธันวาคม 2563: หน้า 32-44

บทความปริทัศน์
(Review article)

การทบทวนวรรณกรรมการใช้งานของผู้ช่วยอัจฉริยะด้วยคำสั่งเสียงและความเป็นไปได้ในปัจจุบัน

Voice User Interface (VUI): A Review of Present and Potential Voice Assistant (VA) applications

กฤติยา แตงมณี* สกล ธีระวรัญญู และ ณัฐนิช บัวบาน

Krittiya Tangmanee*, Sakol Teeravarunyou and Nattanit Buaban

คณะสถาปัตยกรรมศาสตร์และการออกแบบ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี บางขุนเทียน กรุงเทพ 10150
School of Architecture and Design, Bangkhuntien, Bangkok 10150 Thailand (nardame@gmail.com)

*Corresponding author

Received: December 10, 2020/ Revised: December 28, 2020/ Accepted: December 29, 2020

บทคัดย่อ: *ประสบการณ์ในยูสเซอร์อินเตอร์เฟสของการใช้ผู้ช่วยอัจฉริยะด้วยคำสั่งเสียงในอนาคต สามารถนำมาทดแทนการใช้กราฟิกยูเซอร์อินเตอร์เฟสที่พบโดยทั่วไปบนหน้าจอโทรศัพท์ไร้สาย หรือ ในคอมพิวเตอร์ เหตุผลเนื่องจากผู้ใช้สื่อสารอย่างเป็นธรรมชาติด้วยเสียง มากกว่าการสัมผัสหน้าจอโทรศัพท์ไร้สาย แม้กระนั้นก็ตามตัวยูสเซอร์อินเตอร์เฟสแบบใช้เสียง มีองค์ประกอบของผู้ช่วยที่ยังไม่เป็นที่ยอมรับ และปัญหาหลายอย่าง ที่เกิดขึ้นจากการใช้งานจริง เช่น ความเป็นส่วนตัวของผู้ใช้งาน บุคลิกลักษณะผู้ช่วยเสียง ความแตกต่างด้านอายุของผู้ใช้ภาษาที่ใช้ในการสื่อสาร ลำเนียงในการสื่อสารภาษาใช้งาน ความเข้าใจถึงจุดประสงค์หลักของผู้ใช้งาน และการเชื่อมต่อและเข้าถึงอุปกรณ์ที่ต้องใช้งานร่วมกัน การศึกษาทบทวนวรรณกรรมจาก 30 บทความวิจัย จากฐานข้อมูลงานวิจัยทั้งการประชุมวิชาการและวารสาร นานาชาติ ได้มีข้อเสนอแนะของนักวิจัยในการวิเคราะห์และ ปรับปรุงการใช้งานด้านยูสเซอร์อินเตอร์เฟสแบบคำสั่งเสียง ทั้งในเรื่องของความหลากหลายของการผสม วิธีการป้อนข้อมูลและการแสดงผลข้อมูลก็ทำให้เกิดการใช้งานง่ายขึ้นในด้านประสบการณ์ผู้ใช้ เทคนิคของการสนทนา การจัดประโยครูปแบบคำพูด และการจำลองผู้ใช้ทั้งหมด ซึ่งจะทำให้ระบบสามารถเข้าใจบริบทของการใช้งานด้วยเสียงได้ดียิ่งขึ้น*

ABSTRACT: *Voice user interface (VUI) could come to replace the graphic user interface (i.e., mobile phone screen and computer). The reason is that users interact with VUI naturally more than touch screen interface. Nevertheless, the VUI with voice assistants still has problems in both acceptance and usability. The factors like privacy issue, voice assistants' personality, the differentiation of age, language impacts the usability. For this review, 30 papers from the database of conference and research are investigated. Many researchers recommended the multimodality of VUI both input and output make an interface ease of use. For user experience, the speech technique and reformulating query and modeling users is the technique that makes machines understand the context of use better.*

คำสำคัญ: *ผู้ช่วยอัจฉริยะด้วยคำสั่งเสียง ประสบการณ์ผู้ใช้ การใช้งานง่าย*

Keywords: *Voice assistant, user experience, ease of use*

1. INTRODUCTION

Voice user interface (VUI) is an interaction between human and computer by using spoken language. The technology includes speech recognition to understand spoken commands and answer questions, and using text to speech to reply to users. Radio Rex was the first voice-activated toy released in 1911 and it was developed to be the shoebox designed by IBM [1]. It can recognize 16 words spoken into its microphone and convert those sounds into electrical impulses. The mechanism behind the shoebox is the Automated Speech Recognition (ASR) and Natural Language Understanding (NLU). Later, three Bell Labs researchers, Stephen Balashek, R. Biddulph, and K. H. Davis built a system called "Audrey" for single-speaker digit recognition [2]. Their system located the formants in the power spectrum of each utterance.

Voice User Interface has the potential for ergonomics and human factors. For example, VUI can be used for special people such as the elderly and disable people to help them control home automation. In human factors, the interaction of VUI is different from the graphic user interface (GUI) on desktop computers or mobile phones. For example, the menu structure works for Graphic User Interface (GUI), but does not work for VUI. It is difficult for users to remember all options in the menu, due to the limitation of short-term memory. Therefore, the VUI navigation system is not the same as the GUI. The VUI has an intelligent agent that can search the data from the current pool of databases. Currently, many home-device assistants or voice assistant (VA) such as Amazon's Echo and Google Home have an AI technology due to advanced natural language processing and machine learning capabilities. Such advancements of natural language processing allow users to speak to and receive in-context replies in the same way to users' interactions with other human counterparts. Nevertheless, many VUI agents do not understand users' context like the situation that users face and users' preferences. The limitation of VUI needs to be reviewed in the aspect of human-centered design such as the personality of agents, multimodal platform, language, user experience, usability and method approach of each research study.

2. DESIGN OF VOICE USER INTERFACE

2.1 Designing Voice Assistant Personalities

There are two reasons why developers try to create personalities of voice assistants: 1) Increasing the interpersonal level of communication with users. 2) Making voice assistants behave like humans. For the first reason, voice assistants must know users' preferences. Choi et, al [3] created an experiment that the VA knows the user's pattern of usage. For example, VA recorded the previous car rental date and favorite music opened in the car the user played. VA will greet users with past data and make users familiar with it. For the second reason, several VAs have their personalities and mood. Microsoft creates Cortana as an avatar of VA showing their mood through 18 different emotions. Google has a plan to include phrases like "Pretty please" when interacting with kids [4].

Figure 1: **Voice assistant answers the question from the participants, Berdasco et, al [5]**

Yuan et, al. [6] experimented with the VA for children by using personal information. For example, the VA added the name to the greeting conversation. They found that older children were concerned about their privacy, since the setting up VA knew their specific information such as their age and name. On the other hand, younger preferred an interface with naming personalization. They concluded that children preferred personified interface to non-personified interface and naming personalization does not influence children's effectiveness with speech interfaces in a statistically significant way. In the same direction, car drivers are more familiar in the unfamiliar space (i.e., rental car) when using the context-specific voice agent system [3]. Another interesting study is the combination of VA's role. Zhao et, al [7] experiment with two VAs' roles. One is a secretary and another one is the housekeeper role.

The experiment carried out three levels. Two independent VAs were used for personal and work-related issues separately without any synchronization of information between them. Familiar VA which is two VAs was used for personal and work-related separately. Merged VA is the only one VA that could deal with many issues at home or work simultaneously. The result shows that merged VA performs best and the participants also preferred it over the others based on the collected data and mental workload, performance, satisfaction, and perceived usefulness. Subjects saw the future potential to assist with caregiving tasks and human connections among community engagement.

3. DESIGNING VUI FOR SPECIAL PEOPLE

Three types of research on special people are elderly, children, and disabled people. Those special people required VA to act differently.

3.1 VA for elderly

The research suggests elder adult's needs, benefits and barriers with voice-based technology interaction. The main purpose of study is to assist the impaired elderly with the problem of eye and wellness activities. Kowaski et, al. [8] suggested that the VA interface combined with IOT are well adjusted to the needs of elderly. For example, elderly can use their voice to control the lighting, curtain and cleaning robot for their smart home. Moreover, the VA is used as a nursing center to play music for the people over there. The VA also was used for mental impairments for the outdoor environment. Trajkova and Martin-Hammond [9] studies on the reasons why elderly abandoned Amazon Echo. 63% described the limited use of Echo. They preferred other devices such as mobile phones. 18% used their Echo daily. 13% had abandoned the VA and discontinued use because they found no value in its features. 5% were lagging adopters that had interest in Echo. There are some concerns on users' privacy. In shared space, the issue of privacy concerns of other household members. Kowaski et, al. [8] suggested that the VA should give elderly hints when they did something wrong. It should have seamless incorporation into everyday life such as controlling the home automation when their hands are dirty. The VA must be controlled and assured of security, for example, asking VA to turn off the stove safety.

3.2 VA for children

Another group of study is children. Yuan et, al. [6] studied children's information query practices. The speech recognition with children is notoriously difficult described as disfluent and ungrammatical. The experiment is on how children reformulate questions for VA. They found that children struggled and required hints to complete the task. Another research is on how children program the game by using the VUI. The result shows that voice interaction allows children to become more immersed in the game. They can learn programming easily, enjoyably and confidently.

3.3 VA for disable

Lauretti et, al. [10] worked on assistive robots by using the VUI with Electromyography (EMG). The main advantage of VUI is the voice recognition instead of other interfaces (such as joysticks) that can be used easily by severe disabilities in upper limbs. The problem that they found is the delay between the vocal command and the robot motion. When the user pronounced the word, the robot arm moved according to the direction. The obtained results pointed out that performance and user acceptance were higher with respect to voice control.

4. MULTIMODAL PLATFORM

Multimodality is a combination of VUI and others. It can be separated into input and output modality. For example, the input could be a combination of graphic user interface or tactile. While output modality is the GUI or physical objects such as robotic arm, interactive device and others.

4.1 VUI and GUI as input

VUI can be used with Graphic User Interface (GUI) as an input. Several applications for the blind use voice to interact and navigate with the system such as mobile banking for the blind. Corbett and Weber [11] developed the GUI that used the icon attached with the number. They designed M-VUI to allow the user with a disability to use voice to navigate the mobile application without touch gestures. The data shows that the user may not be able to utilize the available function. This has negative results for the user. They suggested that more corrective command and context may be more effective to perform the action. In order to help the user, the specific context to lower the amount of the steps should also be considered. Another interesting study, Rong et, al. [12] developed virtual assistance to help as a timely reminder. The research input used the information to imprecise temporal expressions (ITEs) manifestation as a schedule with estimation of booking time. It created a short task and a memo to the virtual assistant's Corkboard. After that, they picked times or dates of a calendar. The study result was significant as the user felt it was easy to use. The input data was very simple. From this research, the user expects the assistant to recognize user activities and be able to prepare for periodic date and time.

4.2 VUI and Tactile as input

There is a question on the driver's performance and response when operating with in-car VUI. The problem is that the driver cannot focus on a long conversation while driving. Due to short-term memory, the driver cannot recognize the previous command menu as a long conversation. In some cases, this can cause an operation error, as the user must wait for the VUI to finish before deciding and they already forget what the key task is. As a result, Jung et, al. [13] created a multiple model Voice+Tactile system to study an experiment on driving situations with VUI.

Pinpad is a tactile input device that can navigate as an ordered menu of GUI. The result shows that Voice+Tactile system stage can shorten the operation time as the user can skip the task without completing the whole sentence. For the gaze performance, the user frequently looks at the pinpad while they are unfamiliar with the device. Although the research did not say whether the task can cause an accident as the user has to take an eye off the road. Subjects suggested that the location of the pin pad was difficult to reach so they have to look at the pin pad to confirm whether they place the finger into the right area. Overall, the users agreed that the Voice+Tactile system is better to use than a voice-only operation.

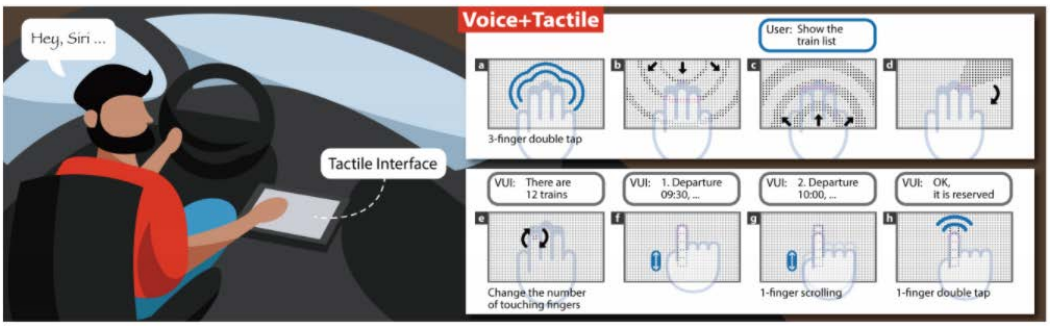

Figure 2: **VUI and object as input Jung et, al [13].**

4.3 VUI and GUI as output

Some mobile applications can use their screens as an output in conjunction with users' voices. The advantage of using GUI as output is that users can visually observe the status and errors all at once on the screen. For example, users do not need to go through the step of booking movie tickets online on many pages. They can use the intelligent voice agent to shorten the navigation. Nevertheless, some parts of the interface requires the screen. For example, it is easy to book the seat location rather than to use the voice command.

Another research is the educational program for children to learn a program. Jung et, al. [14] designed the conversation game code "Turtle talk" which is a web-based system. It helps children to construct the model with voice order. They learn sequencing and iteration processes through the game. With voice recognition, children can complete all the stages. The result is very interesting as children are more focused on the task and action.

Figure 3: **Turtle talk describes VUI+ GUI Jung et, al. [13].**

4.4 VUI and Physical objects as output.

Lauretti et, al. [10] experimented on the interface performance of VUI with a combination of arm-hand robotic systems (M-IMU/EMG). In this research, the study stimulates under these circumstances. 1. Voice control 2. M-IMU/EMG is combined to control the motion of the robot arm. The data contained an interesting finding of robot motion controlled by voice. This output shows a success rate is 88%, which is higher than the voice control. It is discovered that the performance of multiple models is more accurate and can minimize the error. They found the time delay when using only voice control. It is determined from the combination of experiments. Jung et, al. [13] results confirmed that multiple models are more useful than one model.

Figure 4: **Wizard of Oz model shows how VUI+Physical output Yuan et, al. [6]**

Park & Lim [15] investigates on co-ownership of AI speakers. The study is to redesign the identity of AI speakers and support family activities. From the result, the user is giving a high expectation on AI technology as a helping hand on their daily basis around housework co-hosting support with smart home technology. Another role is a family orientated that connects all family members. Due to tasks given to the AI, the family members are expecting the AI speaker as part of the family because it has human-like voice. As a result, AI speakers should be smart, learning and predictive of family activities. They suggest that it can take a burden from the family when needed.

Meanwhile as a co-ownership device, the privacy of each family member data is concerned. Thus, some personal input and context should only be revealed in individual personal spaces.

5. LANGUAGE

Struggling to interface with voice recognition for non-native English speakers might be the reason for not using the Voice assistant system. Pyae & Scifleet [16] examines how the native and non-native English speakers use VUI. The test is considered on how the device can comprehend the user's English proficiency level. The result shows the non-native struggled on the first attempt of each test. At the end of the test, non-native cannot achieve at one go on the test. Therefore, native English speakers find the VA very useful and help them with easy tasks during the test. In this study, there was no result in the distinction between user accents. The finding shows that English language proficiency is needed for improving the use of VUI.

6. USER EXPERIENCE AND USABILITY

When designers try to design a speech that fits to users, there are no guidelines of VUI before. Murad [17] claims that there are also non-empirical arguments in recent speech-related Human Computer Interaction literature that GUI heuristics can be used for VUI. Since the established set of VUI-specific heuristic is still lacking, GUI evaluation methods have been used for evaluating VUI instead. While the designing for voice is unique and different from designing for graphical interfaces, some important aspects and factors may be abandoned. The following review of existing works are about the factors that influence the VUI that can design for users.

6.1 Speech techniques

The speech is the most problematic when communicating with VA. Several techniques were introduced by researchers. Rong et, al. [12] recommends hyperarticulation. It is a tactic that users of VUI often deploy to resolve a problem by speaking more slowly, clearly, and loudly. When users want to edit the speech in VUI, it is difficult for them to do. Speech is slow for presenting information, is transient and therefore difficult to review or edit and interfere significantly with other cognitive tasks.

6.2 Human factors

Researchers found that children struggled with reformulating questions. Many of them require hints to complete the task [10]. This problem might relate to the working memory especially the children who have less working memory than adults. Shneiderman & Aylett [18] have reported that remembering items requires a cognitive load on audio-only output systems. In terms of human error prevention, Jung et, al. [14] claim that dividing the input into individual words helps on error recognition. Yuan et, al. [6] claims that providing feedback on what was heard could help to clarify what users said and specify formulations too difficult for the stem to understand. With the context of use, VUI can understand users' situation, what they do and where they are now are as important as situational awareness for the machine.

6.3 Modeling users

As VUI became more popular in the past few years and have been embedded in many commercial products available to use in daily life, users still encounter problems and obstacles using VUI. The nature of VUI was intangible and not yet fully developed technology of Natural Language Processing, making it having low usability and learnability. However, some researchers have already explored those unsolved challenges. Myers et, al. [19] mentioned that the users behavior patterns when interacting with unfamiliar VUI, user modeling techniques, can be classified in two main categories: model-based approaches (based on behavioral and psychological theory) and bottom-up (based on data analytics). His paper classified users into three groups by using the bottom up approach with an unfamiliar VUI. Three groups of users are novice, intermediate and expert based on user behavior patterns. If machines can model users, they can understand the users and context of use. Not so much research has been done in this area, though.

7. METHOD APPROACH

Thirty research papers were classified based on the method they used in research into 13 categories (see appendix). Seven researchers studied the user acceptance of VUI [7, 9, 21-25]. This shows that the VUI technology needs to be improved in terms of trust, privacy and usability. Many researchers used the interview, questionnaire and survey methods to conduct the research [21-23, 26], since the VUI related to the speech that should be conducted with the qualitative method such as content analysis.

Many studies use empirical methods with the task assignment. They are most used to study the interaction between subjects and intelligent agents [3, 5, 8, 13, 15, 24, 27]. Another technique that most researchers used is the Wizard of Oz method [3, 5]. It is a process that allows subjects to interact with VUI without knowing that the responses are being generated by a human rather than a computer by having someone behind-the-scenes who is typing message with text-to-speech software. These interesting two research papers [15, 23] used the NASA-TLX to measure the drivers' workload while using the in-vehicle voice assistants. Some researchers utilized the existing data set such as the history of using Alexa and Google home [14, 26, 27], for example, the data set with the machine learning techniques of 100 expressions of elderly and caregiver users on Twitter. Some researchers have used the statistical technique to analyze the data by coding the data, then continued to analyze such data by using clustering analysis, factor analysis or correlation [17, 19, 20, 23, 28]. There are some studies that tried to improve and apply the voice system such as improving the noise reduction [29], using VUI for programming [14] and combining VUI and EMG for controlling the robot [11].

8. CONCLUSION

The development of VUI can be done in both systems and users' sites. The system should be improved in order to the efficiency and effectiveness of use. From the review, several research is in the beginning stage since the VUI starts to combine an intelligent agent who can understand users more than ever. There are several aspects of using the VUI according to the age group. Elderly users have a problem of product acceptance and children show the difficulty how they reformulate query. On the other hand, the disabled people have a lot of potential to use VUI for future use to help them control their home devices. Another problem of VUI is the language issue. Non-native people were struggled more than native speakers. This problem could be alleviated by the speech recognition for any specific language and accents. Lastly, several implications of voice-only command were eliminated by the combination of multimodal platform. This has increased the decisiveness of VUI. The speech technique could be used for communication between humans and machines but the future development of how to use speech that is clearer and shorter would be an interesting study.

REFERENCES

[1] IBM Shoebox (2020). IBM Archives, Access on Oct 14, 2020. https://www.ibm.com/ibm/history/exhibits/specialprod1/specialprod1_7.html

[2] Bell System Technical Journal (1940), p. 509.

[3] Choi S, Kim Y, Han E, Son J, Cho J. (2018) Designing conversational voice user interface for improving intimacy of shared invehicle. In: Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers [Internet]. New York, NY, USA: Association for Computing Machinery; 2018. p. 33–37. (UbiComp '18). Available from: https://doi.org/10.1145/3267305.3267638

[4] Dasgupta R. (2018) Voice User Interface Design: Moving from GUI to Mixed Modal Interaction. 2018.

[5] Berdasco, A,,López, G Diaz, I Quesada, L Guerrero, L.A. User Experience Comparison ofIntelligent Personal Assistants: Alexa, Google Assistant, Siri and Cortana. Proceedings 2019, *31*, 51.

[6] Yuan Y, Thompson S, Watson K, Chase A, Senthilkumar A, Bernheim Brush A J. (2019) Speech interface reformulations and voice assistant personification preferences of children and parents. Int J Child-Computer Interact [Internet]. 2019; 21:77–88. Available from: http://www.sciencedirect.com/science/article/pii/S2212868918300898

[7] Zhao J, Patrick Rau P L. (2020) Merging and synchronizing corporate and personal voice agents: Comparison of voice agents acting as a secretary and a housekeeper. Comput Human Behav [Internet]. 2020; 108:106334. Available from: http://www.sciencedirect.com/science/article/pii/S0747563220300881

[8] Kowalski J, Jaskulska A, Skorupska K, Abramczuk K, Biele C, Kopeć W. (2019) Older Adults and Voice Interaction: A Pilot Study with Google Home. In: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems [Internet]. New York, NY, USA: Association for Computing Machinery; 2019. p. 1–6. (CHI EA '19). Available from: https://doi.org/10.1145/3290607.3312973

[9] Trajkova M, Martin-Hammond A. (2020) "Alexa is a Toy": Exploring Older Adults' Reasons for Using, Limiting, and Abandoning Echo. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems [Internet]. New York, NY, USA: Association for Computing Machinery; 2020. p. 1–13. (CHI '20). Available from: https://doi.org/10.1145/3313831.3376760

[10] Lauretti C, Cordella F, di Luzio FS, Saccucci S, Davalli A, Sacchetti R. (2017) Comparative performance analysis of M-IMU/EMG and voice user interfaces for assistive robots. In: 2017 International Conference on Rehabilitation Robotics (ICORR). 2017. p. 1001–6.

[11] Corbett E, Weber A. (2016) What Can I Say? Addressing User Experience Challenges of a Mobile Voice User Interface for Accessibility. In: Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services [Internet]. New York, NY, USA: Association for Computing Machinery; 2016. p. 72–82. (MobileHCI '16). Available from: https://doi.org/10.1145/2935334.2935386

[12] Rong X, Fourney A, Brewer R N, Morris MR, Bennett P N. (2017) Managing Uncertainty in Time Expressions for Virtual Assistants. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems [Internet]. New York, NY, USA: Association for Computing Machinery; 2017. p. 568–579. (CHI '17). Available from: https://doi.org/10.1145/3025453.3025674

[13] Jung. J., Lee. S., Hong. J., Youn. E., Lee. G. (2020) Voice+ Tactile: Augmenting In-vehicle Voice User Interface with Tactile Touchpad Interaction. CHI 2020 Paper. Available from: http://dx.doi.org/10.1145/3313831.3376863

[14] Jung H, So S, Oh C,Kim H, Kim J. (2019) TurtleTalk: An Educational Programming Game for Children with User Interface. CHI 2019 Extended Abstracts.

[15] Park. S., Lim. K. Y. (2020) Investigating User Expectations on the Roles of Family -shared AI speakers. CHI 2020 PAPER. Available from: http://dx.doi.org/10.1145/3313831.3376450

[16] Pyae. A., Sciflleet. P. (2018) Investigating Difference between Native English and Non-native English Speaker in Interacting with a Voice User Interface: A case of Google Home. Available from: https://doi.org/10.1145/3292147.3292236

[17] Murad C. Tools to Support Voice User Interface Design. In: Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services [Internet]. New York, NY, USA: Association for Computing Machinery; 2019. (MobileHCI '19). Available from: https://doi.org/10.1145/3338286.3344424

[18] Shneiderman, B. (2000) The Limits of Speech Recognition. *Commun. ACM* 43, 9 (Sept. 2000), 63–65. Available from: http://dx.doi.org/10.1145/348941.348990

[19] Myers CM, Grethlein D, Furqan A, Ontañón S, Zhu J. (2019) Modeling Behavior Patterns with an Unfamiliar Voice User Interface. In: Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization [Internet]. New York, NY, USA: Association for Computing Machinery; 2019. p. 196–200. (UMAP '19). Available from: https://doi.org/10.1145/3320435.3320475

[20] Kim D, Lee H. Effects of user experience on user resistance to change to the voice user interface of an in‐vehicle infotainment system: Implications for platform and standards competition. Int J Inf Manage [Internet]. 2016;36(4):653–67. Available from: http://www.sciencedirect.com/science/article/pii/S0268401216302298

[21] McLean G, Osei-Frimpong K. Hey Alexa … examine the variables influencing the use of artificial intelligent in-home voice assistants. Comput Human Behav [Internet]. 2019. 99(28–37). Available from: http://www.sciencedirect.com/science/article/pii/S0747563219301840

[22] Braun M, Mainz A, Chadowitz R, Pfleging B, Alt F. (2019) At Your Service: Designing Voice Assistant Personalities to Improve Automotive User Interfaces. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems [Internet]. New York, NY, USA: Association for Computing Machinery; 2019. p. 1–11. (CHI '19). Available from: https://doi.org/10.1145/3290605.3300270

[23] Ziman R, Walsh G. (2018) Factors Affecting Seniors' Perceptions of Voice-Enabled User Interfaces. In: Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems [Internet]. New York, NY, USA: Association for Computing Machinery; 2018. p. 1–6. (CHI EA '18). Available from: https://doi.org/10.1145/3170427.3188575

[24] Kiseleva J, Williams K, Jiang J, Hassan Awadallah A, Crook AC, Zitouni I. (2016) Understanding User Satisfaction with Intelligent Assistants. In: Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval [Internet]. New York, NY, USA: Association for Computing Machinery; 2016. p. 121–130. (CHIIR '16). Available from: https://doi.org/10.1145/2854946.2854961

[25] Thakur N, Han CY. (2018) An Approach to Analyze the Social Acceptance of Virtual Assistants by Elderly People. In: Proceedings of the 8th International Conference on the Internet of Things [Internet]. New York, NY, USA: Association for Computing Machinery; 2018. (IOT '18). Available from: https://doi.org/10.1145/3277593.3277616

[26] Ammari T, Kaye J, Tsai JY, Bentley F. (2019) Music, Search, and IoT: How People (Really) Use Voice Assistants. ACM Trans Comput Interact [Internet]. 2019 Apr;26(3). Available from: https://doi.org/10.1145/3311956

[27] Myers C, Furqan A, Nebolsky J, Caro K, Zhu J. (2018) Patterns for How Users Overcome Obstacles in Voice User Interfaces. In: Proceedings of the 2018 CHI Conference on Human

Factors in Computing Systems [Internet]. New York, NY, USA: Association for Computing Machinery; 2018. p. 1–7. (CHI '18). Available from: https://doi.org/10.1145/3173574.3173580

[28] Nguyen QN, Ta A, Prybutok V. (2018) An Integrated Model of Voice-User Interface Continuance Intention: The Gender Effect. Int J Human–Computer Interact [Internet]. 2019 Sep 14;35(15):1362–77. Available from: https://doi.org/10.1080/10447318.2018.1525023

[29] Xu C, Li Z, Zhang H, Rathore AS, Li H, Song C. (2019) WaveEar: Exploring a MmWave-Based Noise-Resistant Speech Sensing for Voice-User Interface. In: Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services [Internet]. New York, NY, USA: Association for Computing Machinery; 2019. p. 14–26. (MobiSys '19). Available from: https://doi.org/10.1145/3307334.3326073

APPENDIX

Table 1: Method Approach

	NASA-TLX	Wizard of Oz	Questionnaire/ Survey	User Acceptant	Statistics analysis	Empirical method	User satisfaction	Interview	Data set	Expert review	Observation/User study	Eye tracking	System performance
[3]		●				●							
[5]		●				●							
[6]				●									
[7]						●						●	
[8]				●				●					
[9]													●
[10]								●					
[11]								●			●		●
[12]									●				
[13]	●										●		
[14]						●							
[15]						●	●						
[16]										●			
[17]					●								
[18]					●								

	NASA-TLX	Wizard of Oz	Questionnaire/ Survey	User Acceptant	Statistics analysis	Empirical method	User satisfaction	Interview	Data set	Expert review	Observation/User study	Eye tracking	System performance
[19]			●	●									
[20]	●	●	●	●									
[21]					●								
[22]			●	●	●								
[23]								●					
[24]				●		●							
[25]			●					●	●				
[26]				●									
[27]							●						
[28]									●				●
[29]						●							
[30]								●	●		●		●
Sum	2	3	4	7	4	7	2	6	4	1	3	1	4