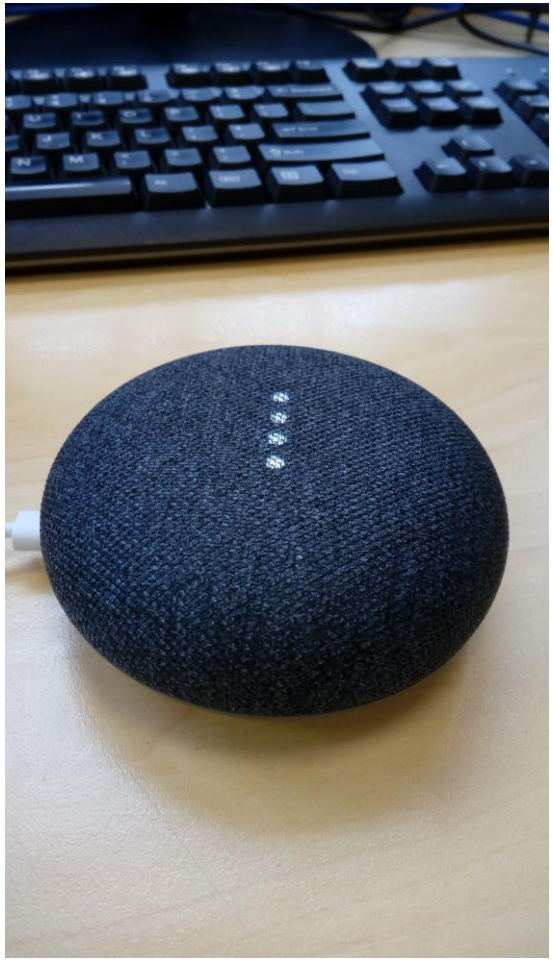

Figure 1: The Google Home Smart Speaker (Mini).

Investigating the Role of User's English Language Proficiency in Using a Voice User Interface: A Case of Google Home Smart Speaker

Aung Pyae
Swinburne Business School
Swinburne University of Technology,
Melbourne, Australia
apyae@swin.edu.au

Paul Scifleet
Swinburne Business School
Swinburne University of Technology,
Melbourne, Australia
pscifleet@swin.edu.au

ABSTRACT

In recent years commercial Voice User Interfaces (VUIs) such as the Google Home Smart Speaker, Amazon's Echo, and Apple's Siri have drawn attention from different user groups; however, these existing commercial VUIs support limited language options for users including native English speakers and non-native English speakers. Also, the existing literature about usability differences between these two distinct groups is limited. Thus, in this study, we conducted a usability study of the Google Home Smart Speaker with 20 participants including native English and non-native English speakers to understand their differences in using the Google Home Smart Speaker. The findings show that compared with their counterparts, the native English speakers had better and more positive user experiences in interacting with the device. It also shows that users' English language proficiency plays an important role in interacting with VUIs. The findings from this study can create insights for VUI designers and developers for implementing multiple language options and better voice recognition algorithms in VUIs for different user groups across the world.

CHI'19 Extended Abstracts, May 4–9, 2019, Glasgow, Scotland, UK

ACM ISBN 978-1-4503-5971-9/19/05.
https://doi.org/10.1145/3290607.3313038

KEYWORDS

Voice user interfaces; Google Home Smart Speaker; usability; user experiences; usefulness

Table 1: Objectives of the Study

Study Objectives
• To investigate differences between native English speakers (NES) and non-native English speakers (NNES) in using GHSS in terms of usability, ease of learning, self-efficacy and usefulness,
• To investigate whether user's English language proficiency correlates to user's experiences in using GHSS in terms of usability, ease of learning, self-efficacy, and usefulness.

Table 2: Pre-defined Tasks to GHSS

Voice Commands to GHSS
To ask the device about the weather today.
To ask the device about the current traffic conditions of a particular location.
To ask the device about the latest world news.
To ask the device to translate from one particular language to another (e.g. English to Mandarin)
To create a personal calendar through a voice command.
To ask the device about a particular search query.
To set up and test an alarm clock.
To use a calculator through a voice command.
To use a dictionary through a voice command.
To ask the device about calendar schedule.
To delete the calendar event.
To play radio on the device.
To play a particular piece of music on the device.
To ask about a famous person/celebrity.
To ask about a particular geographical location (e.g. country/city/town)

1 INTRODUCTION

According to Cho [1], conversation has become a key interface in human-computer interaction (HCI) as virtual assistants (VAs) are integrated into smart phones and smart home devices. In recent years, commercial Voice User Interface (VUIs) products with intelligent VAs have been available in the consumer market, including the Google Home with Google Assistant, Amazon' Echo with Alexa, and Apple's Siri. Described simply, a VUI is "*a user interface that uses speech input through a speech recognizer and speech output through speech synthesis or prerecorded audio*" [2, p1], that allows users to interact with a smart system through voice-based commands and audio-based responses [3]. Pearl [4] describes how VUIs bring some important advantages to users including hands-free and natural interaction. According to Myers et al. [5], commercial VUIs have been growing in popularity and have been integrated into users' daily lives in terms of smart-home assistants, information seeking, entertainment, and personal information management.

Since 2016, most releases of the Google Home Smart Speaker (GHSS) (see Figure 1) have been in countries where English is the official language, including the United States, Canada, the UK and Australia. Furthermore, existing commercial VUIs, including GHSS, provide only limited language selection options for users [6]. For instance, in the Google Assistant application, users can choose from only six languages: English, French, German, Italian, Japanese, and Spanish for primary interaction [7]. Amazon's Alexa and other commercial VUIs are similarly limited in the language options they provide for users. As a result, users who use English as a second language or non-native English speakers are required mainly to use English to interact with GHSS unless their native or first language options are available. In the United States, one in Five U.S. Residents speaks a language other than English at home [8] and this trend is similar in other English language countries, like the UK and Australia, where GHSS is officially available with a sizable proportion of the population speaking English only as a second language (e.g. foreign students and immigrants). Thus, non-native English speakers must be viewed as an important part of the consumer market for GHSS.

While the number of studies investigating the usability and usefulness of GHSS and consumer VUIs is increasing, for example, Pyae and Joelsson's [7] online survey of the usability and usefulness of GHSS reported that GHSS is an easy and user-friendly device for users, the researchers have also stressed the importance of further detailed inquiry into its usability and usefulness. Similarly, Wallace and Morris [9] conducted a study to investigate the usability, needs and user preferences for smart speakers by military veterans with mild traumatic brain injury (mTBI) and post-traumatic stress disorder (PTSD). Although their findings show the potential of GHSS to support health and wellbeing, the researchers also indicate the importance of further research into the usability challenges of smart speakers for this population. Building from this, the main objectives of this study are stated in Table 1.

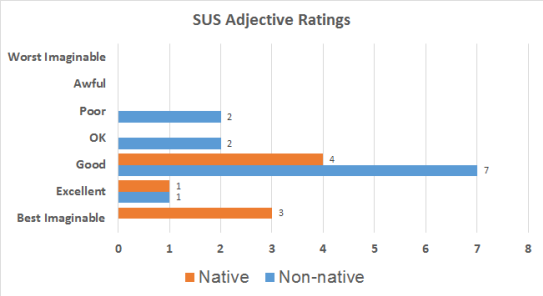

Figure 2: SUS adjective ratings

Table 3: Ease of Learning in GHSS

Ease of learning items	NNES		NES	
	M	***SD***	***M***	***SD***
I learned to use it quickly.	4.5	0.5	4.5	0.5
I easily remembered how to use it.	4.3	0.7	4.6	0.5
It easy to learn how to use it.	4.4	0.5	4.6	0.5
I quickly became skillful with it.	3.5	0.9	4.5	0.5

Table 4: Participants' Self-efficacy in GHSS

Self-efficacy items	NNES		NES	
	M	***SD***	***M***	***SD***
I feel confident giving a command to this device.	3.7	0.7	4.4	0.6
I feel confident receiving voice-based results from the device	3.6	1.0	4.1	0.8

Table 5: Participants' attitudes towards GHSS

Attitude items	NNES		NES	
	M	***SD***	***M***	***SD***
Using it is a good idea.	3.7	0.7	4.3	0.7
Using it is a pleasant experience.	3.6	0.8	4.3	0.7
Using it is beneficial to me.	3.3	0.9	4	0.9

2 METHOD

In this study, we recruited 20 participants from an Australian university through a poster advertisement on campus. The inclusion criteria for participants included, age restriction (18 years and above), self-assessment of English language proficiency (for both native English speakers and participants who use English as a second language), and individual consent. The study took place at a usability lab at the university and comprised three stages: pre-study, in-study, and post-study. In the pre-study, we used a questionnaire to ask participants about their demographics and prior experience in using VUIs. It also included an exercise for the self-assessment of English language proficiency based on the Common European Framework of Reference for Languages (CEFR); participants used a self-assessment grid to report on their own level of English language proficiency [10] for four different language skills: listening, reading, spoken interaction, and spoken production and rated their own proficiency from A1 (1 - *the lowest*) to C2 (6 - *the highest*). The researchers confirmed their agreement with the participants assessment as part of this process. The pre-study took approximately 20 minutes. During the in-study stage, the participants were asked to execute 15 pre-defined tasks that included giving voice commands to GHSS (see Table 2). For instance, the participants gave a voice-command to check the weather of a particular location (e.g. "*What is the weather in Melbourne today?*") with the researchers observing the interaction between the participants and GHSS. The in-study stage took approximately 30 minutes to complete. In post-study stage, we used the System Usability Scale (SUS) [11] and User Experience Scale (UES) [12] questionnaires to understand the participants' experiences in using GHSS. We used 5-point Likert scales for all questionnaires, ranging from "*Strongly Disagree* (1)" to "*Strongly Agree* (5)". We also conducted a short interview session with each participant to gain further insight into the users' experience of interacting with GHSS. The post-study took 20 minutes approximately while the complete study took approximately 1 hour and 30 minutes for each participant. In the analysis below, native English speakers are described using the abbreviation of 'NES' and non-native English speakers as 'NNES'.

3 DATA ANALYSIS AND FINDINGS

Eight (8) native English speakers (40%) and 12 non-native English speakers (60%) participated in the study. There were 17 male participants (85%) and 3 female participants (15%). While most were either students or staff members of the university, participants also included some members of the surrounding community. The average age of the participants is (M=29). All had prior experiences in using VUIs such as Apple's Siri or Google Assistant, however only 6 had used GHSS prior to the study. With regard to the participants' English language proficiency, as might be expected, native English speakers had noticeably high scores: Listening (M=6, SD=0), Reading (M=5.8, SD=0.4), Spoken Interaction (M=6, SD=0), and Spoken Production (M=5.6, SD=0.4). For the non-native English speakers, average scores for each English skill were: Listening (M=4.7, SD=1.2), Reading (M=5.2, SD=0.6), Spoken Interaction (M=4.6, SD=0.9), and Spoken Production (M=4.6, SD=0.9). Regarding the usability of GHSS, we used adjective ratings to translate the participants' System

Table 6: User's Satisfaction towards GHSS

Satisfaction items	NNES		NES	
	M	***SD***	***M***	***SD***
I am satisfied with it.	3.4	0.9	4	0.5
It is fun to use.	4.6	0.5	3.6	1.0
I feel I need to have it.	2.8	1.0	2.9	0.7
It is pleasant to use it.	3.9	0.9	4.3	0.6
It works the way I want to work it to work.	3	1.0	3.5	1.0
I would recommend it to a friend.	3.4	1/1	4.1	0.5

Table 7: Usefulness of GHSS

Usefulness items	NNES		NES	
	M	***SD***	***M***	***SD***
It is useful.	3.6	0.5	4.5	0.5
It would save me time by using it.	3.2	1.0	4.3	0.7
It meets my needs.	3.2	0.7	4.4	0.7
It does everything I would expect it to do.	2.6	1.1	3	1.1
It could make the things I want to accomplish easier to get done.	3.1	0.8	4.1	0.8

Usability Scale (SUS) scores into descriptive terms [11]. For instance, the SUS score 90.1 and above is equivalent to the term "*Best Imaginable*" while the SUS score 12.5 and below is described as "*Worst Imaginable*" [11]. The findings from the SUS analysis show that NES participants had noticeably high SUS scores: 3 at "Best Imaginable" (37.5%), 1 "Excellent" (12.5%), and 4 "Good" (50%) respectively. The NNES group had 1 "Excellent" (8.3%), 7 "Good" (58.3%), 2 "OK" (16%), and 2 "Poor" (16%) respectively. We found that all NES had more positive SUS scores to the usability of GHSS; whereas, some of the NNES had negative scores in SUS. The SUS adjective ratings of both NES and NNES are given in Figure 2. Regarding ease of learning, we found that the NES had a relatively high mean score (M=4.6, SD =0.1) and the NNES also had a high mean score (M=4.2, SD=0.5). The individual score for each item in ease of learning can be seen in Table 3. For participant self-efficacy (see Table 4) we found that the NES had a high confidence in giving commands to GHSS (M=4.4, SD= 0.7) and that they were also confident to receive voice-based results from it (M=4.1, SD=1.0), whereas the NNES had average mean scores in giving commands (M=3.7, SD=0.6) and similar scores in receiving results from GHSS (M=3.6, SD=0.8). Concerning participant attitudes in using GHSS (see Table 5), we found that the NES had more positive attitudes (M=4.2, SD=0.1) compared to the NNES who had just average scores in their attitudes towards using GHSS (M=3.5, SD=0.2). Regarding comfort of use, the findings show that the NES were slightly more comfortable (M=4.3, SD=00.70) than the NNES (M=4.0, SD=0.9). Both NES satisfaction (M=3.7, SD=0.5) and NNES (M=3.5, SD=0.6) had above average scores when it came to satisfaction in using GHSS (see Table 6). Lastly, for the usefulness of GHSS, the NES had a relatively high score (M=4.0, SD=0.6) while the NNES had an average score (M=3.1, SD=0.4) (see Table 7).

Regarding the correlation between the participants' level of English proficiency and their user experiences, we conducted the correlation analysis using SPSS software. The findings show that there is a positive correlation between the participants' English language proficiency and the ease of learning when interacting with GHSS ($r= 0.60$, $p < 0.01$). We also found that the participants' level of English proficiency is positively correlated to the users' experience of usability of GHSS ($r=0.70$, $p < 0.01$). However, we found that the participants' level of English language proficiency is not correlated to their perception of the usefulness of GHSS. Similarly, we found that there is no correlation between the English language proficiency of participants and their self-efficacy or satisfaction in using GHSS. According to the participants' comments about the usability and their experience in using GHSS, we found that the NES had relatively positive experiences although some participants highlighted the usability issues they encountered during task execution (see Table 8). According to the NNES' comments about the usability of GHSS, we found that there were mixed reactions including both positive and negative views (see Table 8). Overall, the findings from the interview show that NES participants considered GHSS fun to use and useful and, apart from some significant usability challenges, most NNES participants also thought GHSS was fun and

Table 8: Participants' Comments towards the Usability of GHSS

PNo	Comments	Type of User
#1	"*I think the, the experience is very interesting. I think it's, it's good, very good*".	Native English Speaker
#6	"*the system understood clearly when it listened to me but its - its database couldn't adapt to different changes of questions, different types of wordings. Um, other than that it seems like it had its own database to answer certain questions only*".	Native English Speaker
#7	"*well, I wouldn't put user friendly, ur, definitely not with me. Ur, I could see that it was, ur, lots of effort to just make this one activated but it might be useful for some others. Um, at the moment I'm not sure, ur, about this*"	Non-native English Speaker
#19	"*very easy to use, provided useful information. I'm not sure that I would really need to have such a device, but I can see that it would be helpful sometimes".*	Non-native English Speaker

**PNo= Participant's No*

mentioned their intention to use it in the future. Both user groups considered GHSS to be a useful device for the household in terms of entertainment, information seeking, and smart home automation, however some participants (NES and NNES) were concerned about security and about being recorded by the device-service while interacting with it. Sixteen (16) of the 20 participants mentioned that they would not use GHSS in an emergency situation, whereas 4 participants commented that hands-free, voice commands could be useful in an emergency situation. Lastly, 15 out of 20 participants mentioned that they would not trust GHSS for seeking health related information or advice, while 5 participants said that they would use it for seeking general health knowledge and advice. Based on the findings from researcher observation during participant task interactions with GHSS, we noted the following usability challenges for both NES and NNES participants (see Table 9).

4 DISCUSSION AND CONCLUSION

In this study, we aimed at investigating the usability, ease of learning, self-efficacy, and usefulness of GHSS between NES and NNES participants. We were also interested to know if a user's level of English language proficiency would correlate to the usability, ease of learning, self-efficacy, and usefulness of GHSS. The findings show that there was a noticeable difference between NES and NNES in using GHSS. We found that the NES had better and more positive user experiences than the NNES in using GHSS. They learned to use the device quickly and it was easier for them to use. Overall, NES participants found the device user-friendly and they saw the potential for its use in a range of different contexts (e.g. information seeking and entertainment, household automation). Furthermore, they felt confident in interacting with the device and more often than not they obtained clear answers from the device. Their positive experiences led to relatively high satisfaction scores in using GHSS. For NNES participants, we found that their user experience of GHSS was 'average' overall. Although they mentioned that the device is relatively easy for them to learning how to use, they encountered significant challenges in using it, especially in using English to structure sentences, choose the 'right' words (vocabularies) and pronounce instructions in a way that GHSS recognised. Despite the challenges in using this device, most NNES participants still felt confident in interacting with GHSS and their satisfaction scored above average. The findings show clearly that a users' level of English language proficiency does have a profound influence on the ease of learning and usability in interacting with GHSS [13]. Despite this, we found that a users' level of English language proficiency did not influence their sense of self-efficacy and usefulness of the system. In conclusion, GHSS is relatively easy and user-friendly especially for native English speakers. Regardless of their user experiences, both native English and non-native English speakers are interested in using it and they both suggested the potential of it to be used in different contexts. The findings also highlight that the ability and proficiency of English language plays a vital role in interacting with GHSS. Based on the findings from this study, we recommend that in designing VUI, voice recognition algorithms should be further improved to be effective for different user groups including those who use or do not use English as a first language. More importantly, designers and developers of VUI should take into account implementing multiple languages for

Table 9: Usability Challenges

Usability Challenges
• First, GHSS cannot easily distinguish multiple commands in one voice interaction and this created difficulties for participants aiming to engage conversationally. • Second, GHSS's response is often too quick for users who may give commands slowly or pause in the middle of a command in order to construct their enquiry and this can lead to incomplete and inaccurate responses from GHSS. • Third, in general information seeking, the answers provided by GHSS may be too short and generic and this can lead to user disappointment as they are often seeking to know more about the matter of their enquiry. • Fourth, in translating from one language to another, most of the time GHSS achieved only a simple translation, it often redirected users to a secondary website for further information and this contributed to user disappointment. • Fifth, users often needed to rephrase or re-structure the English statement in giving commands to GHSS and this was a source of both frustration and disappointment. • Lastly, we found that both NES and NNES indicated that they would prefer to see a combination of both visual and voice results supported by GHSS.

different users in consumer markets. The usability challenges highlighted in this study can be insightful for future implementation of VUIs. The limitation of this study includes relatively small sample size and the gender imbalance among participants, the duration of the study and limited tasks pre-defined for this study. In the future, we aim at conducting a usability study with a larger sample size and wider range of tasks in using GHSS.

REFERENCES

[1] Janghee Cho. 2018. Mental models and home virtual assistants (HVAs). In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. DOI:https://doi.org/10.1145/3170427.3180286

[2] Dirk Schnelle-Walka. 2012. I tell you something. In *Proceedings of the 16th European Conference on Pattern Language of Programs, 10*. DOI: https://doi.org/10.1145/2396716.2396726

[3] Interaction Design Foundation. 2018. Voice User Interfaces. Retrieved January 6, 2019 from https://goo.gl/FQXscC

[4] Cathy Pearl. 2016. Designing voice user interfaces. O'Reilly Media.

[5] Chelsea Myers, Anushay Furqan, Jessica Nebolsky, Karina Caro, and Jichen Zhu. 2018. Patterns for how users overcome obstacles in voice user interfaces. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*.

[6] Google Home Help. 2018. Retrieved January 6, 2019 from https://goo.gl/6q53kk

[7] Aung Pyae and Tapani N. Joelsson. 2018. Investigating the usability and user experiences of voice user interface: A case of Google Home smart speaker. In *Proceedings of the* 20th *International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct MobileHCI' 18*, Barcelona, Spain

[8] Center for Immigration Studies. 2018. Retrieved January 6, 2019 from https://goo.gl/iiMTej

[9] T Wallace and J Morris. 2018. Identifying barriers to usability: Smart speaker testing by military veterans with mild brain inquiry and PTSD. In: Langdon P., Lazar J., Heylighen A., Dong H. (eds) Breaking Down Barriers. CWUAAT 2018. Springer, Cham.

[10] European Language Portfolio (ELP). 2018. Retrieved June 23, 2018, from https://goo.gl/LC9yDu

[11] Aaron Bangor, James Miller, and Philip Kortum. 2009. Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of Usability Studies*, 4(3), 114-123.

[12] Heather L.O'Brien et al. 2018. A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. *International Journal of Human-Computer Studies*, 112, 28-39.

[13] Aung Pyae and Paul Scifleet. 2018. Investigating differences between native English and non-native English speakers in interacting with a voice user interface: A case of Google Home. In *Proceedings of the 30th Australian Conference on Computer-Human Interaction (OzCHI 2018)*, Melbourne, Australia.