CS 449 Human-Computer Interaction Term Paper

-

Sabancı University, Faculty of Engineering and Natural Sciences, Department of Computer Science

Project Title:

Evaluating Smartphone Interaction in Middle-Aged Adults/Elderly: A Comparative Study of Siri Voice Commands and Manual Task Execution

Group Members:

Ceren Dinç [cerendinc@sabanciuniv.edu, 28220]

Şimal Yücel [simal@sabanciuniv.edu, 29420]

Zeynep Güçlü [zeynepguclu@sabanciuniv.edu, 27988]

Table of Contents

Abstract

This study embarks on an exploratory journey to explore the complexities of user interactions with smartphone technology, specifically comparing the effectiveness and efficiency of Siri voice commands versus manual task execution among a distinct demographic. Aimed at understanding how mid-aged adults, a group often less emphasized in technology adaptability studies, navigate these two modes of interaction. The research focuses on participants aged 40-60, comprising six males and six females. The research is designed to be engaging and insightful, shedding light on the practicality of voice command technology in everyday smartphone use for this age group. Participants were engaged in a series of five common smartphone tasks: making a phone call, sending a text message, performing a web search via Safari, setting up an alarm, and opening an application. These tasks, which offer a thorough overview of interaction methods, were chosen due to their regularity in everyday smartphone usage. Preliminary findings from the study indicate that while Siri voice commands proved effective for most tasks, particularly excelling in ease of use and user satisfaction, there was a notable lag in performance when it came to opening applications and error rates were also measured. This contrast in efficiency between the two interaction methods provides insightful implications for the design of smartphone interfaces and voice command systems, particularly for the mid-aged demographic. The study's outcomes not only highlight the strengths and limitations of current voice command technology but also pave the way for future enhancements tailored to diverse user groups.

I. BACKGROUND OF THE STUDY

a. Related Literature

The landscape of smartphone user interaction on middle-aged adults has been extensively explored in recent literature, particularly focusing on the comparison between voice-activated assistants and manual operations.

In the study "Hands free - care free: elderly people taking advantage of speech-only interaction" by Linda Wulf et al.(2014), the focus is on the practicality and acceptance of speech-only interaction via Siri among elderly users. This research is particularly relevant to this project, as it demonstrates the high potential and positive user reception of voice-command technology in a specific age group. By connecting an iPhone with Siri to a Bluetooth headset, the study bypassed the traditional GUI and enabled participants to interact with their phones entirely through voice commands. The findings revealed that elderly users greatly appreciated the simplicity and hands-free nature of this interaction, indicating a strong preference for voice commands.

Another study called "Voice and Touch Interaction: A User Experience Comparison of Elderly People in Smartphones" by Carla Tubin et al.(2021), the focus is on how voice interaction impacts the user experience (UX) of elderly people with smartphones. This study is highly relevant to this project, as it explores the efficacy of voice interfaces in enhancing UX for middle-aged adults, a demographic that often faces challenges with traditional touch-based interactions due to age-related physical and cognitive limitations. Through an experiment involving 20 elderly participants, the study found that voice interaction significantly reduced reliance on vision, and provided practicality, speed, and ease, especially regarding motor issues. However, it also identified barriers such as command formulation, speech rate, and learning new technologies.

Lastly, the study called "Evaluating the Usability of the Siri Virtual Assistant on Mobile Devices with Emphasis on Brazilian Elderly Users" by Thiago Silva Chiaradia et al.(2019),

extends beyond age as a factor in usability, highlighting the influence of prior knowledge, education level, and personal motivation on the effectiveness of Siri among elderly Brazilian users. This research, which involved elderly participants completing tasks using Siri and providing feedback through questionnaires, aligns closely with this project. It suggests that while age-related considerations are crucial, a comprehensive understanding of user interaction with Siri also requires examining users' technological familiarity and personal motivations.

b. Purpose of the Study

The primary objective of this project is to comprehensively analyze how individuals aged 40-60 interact with smartphones, focusing particularly on the comparison between using Siri voice commands and manual task execution. This study aims to identify which of these interaction methods is more effective, user-friendly, and preferred by the elderly demographic. Key areas of focus include assessing the ease of use, task efficiency, error rate, and overall user satisfaction. Additionally, the study seeks to understand the role of age-related factors, such as cognitive and motor skills, in determining the usability and preference for these interaction modes.

c. Significance of the Study

This research holds significant importance in the context of an aging global population and the increasing pervasiveness of technology in daily life. As the elderly population grows, understanding their interaction with current technologies, especially smartphones, becomes crucial. This study's findings will have implications for designing more age-inclusive technologies, improving the digital experience for older adults. Insights gained can guide smartphone manufacturers and app developers in creating more user-friendly interfaces that cater to the needs of older users. Moreover, this research contributes to the broader field of human-computer interaction (HCI) by providing empirical data on the preferences and capabilities of an often underrepresented demographic in technology design. The outcomes of this study could also inform policy and educational programs aimed at increasing digital literacy among the elderly, ultimately enhancing their quality of life and independence.

II. METHODOLOGY

This study aims to discover the differences of user experience between tasks executed by voice commands, specifically with Siri which is a digital assistant belonging to Apple Inc., and manual commands. For this purpose following research questions are investigated:

1. Among these two interaction modes, Siri digital assistant and manual execution, which one performs more efficiently in terms of time and ease of use?
2. In terms of effectiveness, which type of interaction is better in performing common smartphone tasks among middle-aged adults?
3. What could be the effect of age-related factors, such as cognitive skills, while performing the smartphone tasks through Siri and usual methods?
4. What is the task completion rate of middle-aged adults using voice commands as compared to manual task execution?
5. What is the error rate with using Siri commands while performing daily smartphone tasks?

This study is conducted to find answers for these questions. In order to achieve this aim, twelve middle-aged adults were chosen for this research. To ensure that results are not affected by some factors, such as gender, education level and general capability of using smartphones, there are equal numbers of participants from both genders (six men and six women). Additionally, all of them have a university degree as well as use their smartphones at a moderate level in their daily lives. The participants voluntarily participated in this study and initial information about what they should do is given to them.

Participants were asked to perform predetermined smartphone tasks sequentially using voice commands and then manually. The duration for each task was timed and recorded separately. For a robust analysis, participants performed the tasks in the same order, first using voice commands as they learned the tasks on the spot. While determining the tasks, HCI design principles and usability measures were considered. Accordingly, as the study aimed to examine the impact of voice command assistants on the daily actions of middle-aged people, the tasks

were chosen from actions that are familiar and frequently performed by this group. In addition, ensuring that task descriptions were clear and understandable was another point taken into account. The results of the study are intended to guide the future development of increasingly popular voice command assistants, with the goal of making people's lives easier. Generally, these five tasks are tasks that users can perform effortlessly and quickly in their daily lives. Furthermore, in any difficult situation, such as a physical disability or while driving a car, voice commands are expected to be highly functional. The detailed description of the tasks performed in the practical part is as follows:

To ensure that the progress of the study was not affected by phone-related factors (such as the time taken to unlock the phone), all participants performed their manual tasks starting from an already unlocked phone and from the first page of the phone.

***Task 1*:** Participants had to make a call to a person listed in their phone contacts, either using a voice command or by finding the contact manually in the phonebook. (To avoid the influence of alphabetical order, calls were made using the recent calls section)
Siri command: "Hey Siri! Call X."

Task 2: Participants had to send a message to a person in their phone contacts, either through the voice assistant or without using the voice assistant. To ensure the task duration was not influenced by the length of the message, all participants said and typed the same text.
Siri command: "Hey Siri! Send a message to X. 'Hello. How are you?'

Task 3: Participants were required to reach the 'e-government' website using the Safari search engine, either through voice assistant or by typing into the website's search engine themselves. To ensure that typing duration was only influenced by participant skills, they were asked to access the same website.
Siri command: "Hey Siri! Search for e-government on the web."

Task 4: Participants had to set an alarm for a predetermined time (17:00), either using voice commands or manual actions. To avoid influencing the results, it was ensured that

there were no pre-existing alarms for the required time on the participants' phones.
Siri command: "Hey Siri! Set an alarm for 5 PM."

Task 5: Participants had to open an application on their phones, either using a Siri command or without. Since the location of the application could affect task completion time, each participant was asked to access an application located on the third screen of their phone, starting from the first screen.
Siri command: "Hey Siri! Open Instagram."

Each task was performed considering the specified restrictions, and completion times were recorded. After the tests were conducted, the gathered data were compared by taking the average completion times of each test across twelve participants. Here, the task durations form the dependent variable. The data transferred to an Excel sheet for evaluation. Furthermore, to collect qualitative data, some open ended questions were asked to participants to evaluate their overall satisfaction level from Siri. Those questions are provided below:

1. Do you think that Siri is easy to use?
2. Will you use Siri in your daily activities?
3. In which situations do you prefer Siri over manual execution?
4. Does Siri accurately understand and execute your commands?

ID	Participants	Task1-Calling	Task1	Task2-Message	Task2	Task3- web search	task3	Task4- setting alarm	Task4	Task5- opening app	Task5
		Siri	Manual	Siri	Manual	Siri	Manual	Siri	Manual	Siri	Manual
1	Esen Güçlü	6,23	7,49	14,28	19,87	9,47	10,8	7,67	11,46	5,72	4,34
2	Gökhan Güçlü	5,57	7,79	13,46	14,2	8,8	11,93	7,13	15,47	4,12	4,46
3	Nihal Dinç	5,34	8	9,67	18,46	4,32	8,63	4,84	9,32	3,35	1,47
4	Hasan Onur Dinç	3,47	6,6	10,14	17,42	6,35	9,52	4,06	5,44	4,46	2,38
5	Aslı Şen	4,46	7,11	11,23	15,42	7	13,63	5,32	8,39	5,39	7,21
6	Umut Şen	3,29	7,24	12,23	17,57	7,36	11,84	4,47	8,78	4,27	4,32
7	Esra Özkan	3,45	6,67	8,63	20,45	8,45	12,36	4,34	9,56	3,87	4,76
8	Kazım Özkan	3,26	8,45	9,22	16,54	7,58	9,47	4,68	8,37	4,49	3,28
9	Gülenay Akarca	4,67	5,89	13,4	17,32	5,39	11,64	5,6	10,75	3,73	3,48
10	Mustafa Akarca	4,78	8	10,53	16,7	8,9	9,22	4,89	9,23	5,47	5,87
11	Gökhan Salcan	5,29	4,54	9,77	18,65	6,48	10,73	6,52	11,19	4,36	3,12
12	Aslım Salcan	5,14	7,34	13,64	19,79	7,69	10,62	4,87	10,71	4,61	3,71
	AVERAGE	4,58	7,09	11,35	17,70	7,32	10,87	5,37	9,89	4,49	4,03

Table 2.1 - Task Completion Times of Each Participant in Both Method

Since users encountered some difficulties with the voice command system, the results in the table show the times in which users succeeded in using the voice command.

III. RESULTS AND DATA ANALYSIS

The study was designed to evaluate the efficiency, effectiveness, and user satisfaction of performing common smartphone tasks using Siri (voice commands) versus manual execution among middle-aged adults. In total there were 12 participants from both genders equally as well as in the same education level. All participants performed 5 different tasks and 10 executions in total.

a. Quantitative Test Results

The quantitative data collected includes task completion time and error rates while performing daily smartphone tasks using Siri voice assistant versus manual execution.

During the test phase, for each task completion times were recorded and then average completion times were calculated for both methods. To understand and compare efficiency of voice command and manual execution bar charts (Figure 3.1 and Figure 3.2) were created for both of them. According to the provided bar charts, Siri has a lower average time in 4 out of 5 tasks compared to manual execution, which indicates better efficiency of digital assistants. Only the task of opening an app was performed manually faster by 0.46 seconds, a negligible difference. However, it is noteworthy that Siri is faster in the messaging task with a significant difference of 6.35 seconds. Since the predetermined message here is frequently used in daily speech, it was little affected by the pronunciation factor. However, the results may vary depending on the complexity and length of the message to be transmitted. Still, Siri has delivered a performance that cannot be ignored.

The tasks where Siri showed the most significant difference were searching for a page on the web and setting an alarm which is almost two times faster with voice commands. Generally, these actions require users to navigate through multiple tabs and distinguish between letters and numbers, which may explain the efficiency difference.

Based on these results, it was concluded that Siri's performance was better than manual operation.

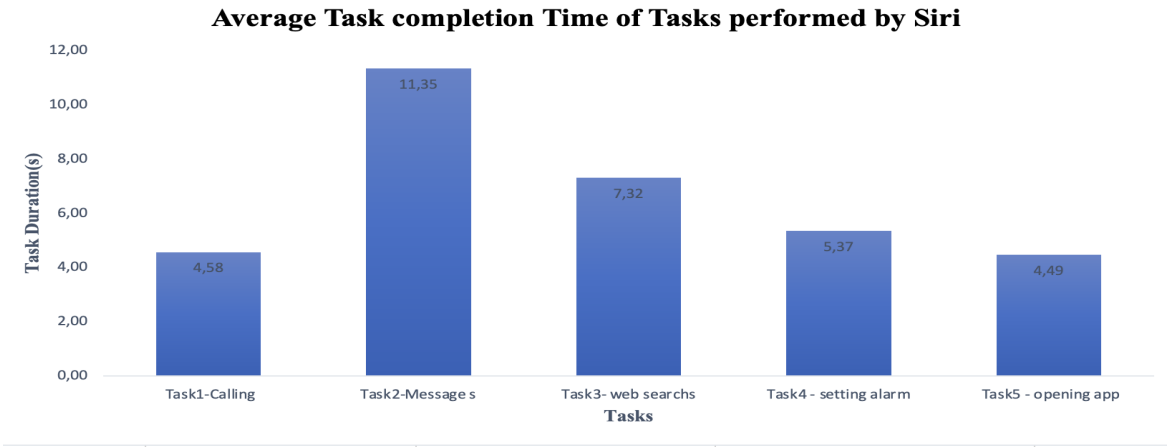

Figure 3.1 - Average Completion Time of Tasks performed by Siri

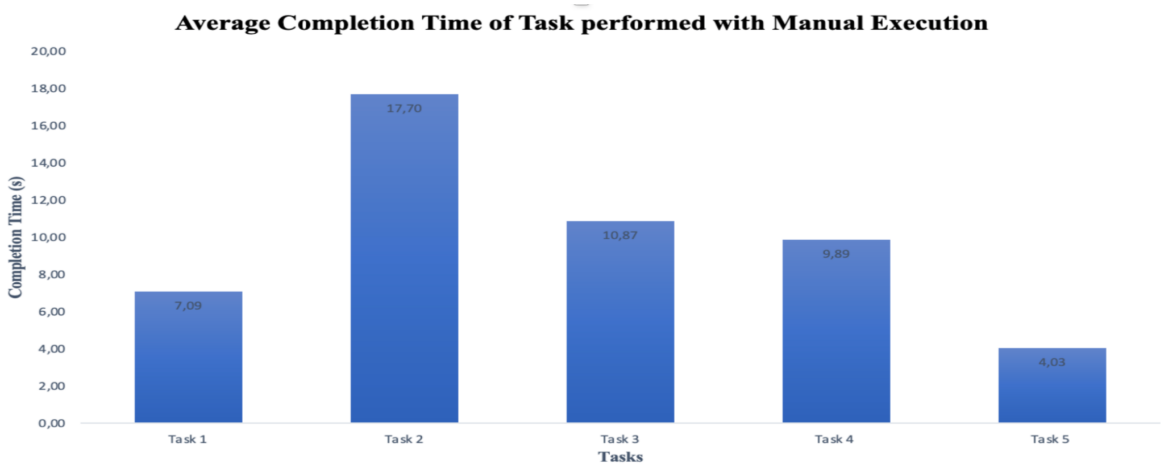

Figure 3.2 - Average Completion Time for Tasks performed by Manual Execution

Furthermore, another critical numerical result is the error percentage. This data indicates the ratio of attempts users needed to correctly complete the voice command tasks. The error rate is a measure of Siri's reliability. According to the bar graph (Figure 3.3), the task with the highest error rate of 75% using Siri was messaging, while setting an alarm was performed with considerable success. Additionally, users made errors 58% of the time when performing tasks like web searches and opening an app. Although this rate is quite high, examining the three tasks with the highest error rates reveals they are

the activities that involve the most speaking and require the highest precision in pronunciation. In other words, the high error rate could stem from the users' language skills. A better comparison was made by looking at the average completion times for these tasks, and logical support was found for the error rate's relevance to user interaction.

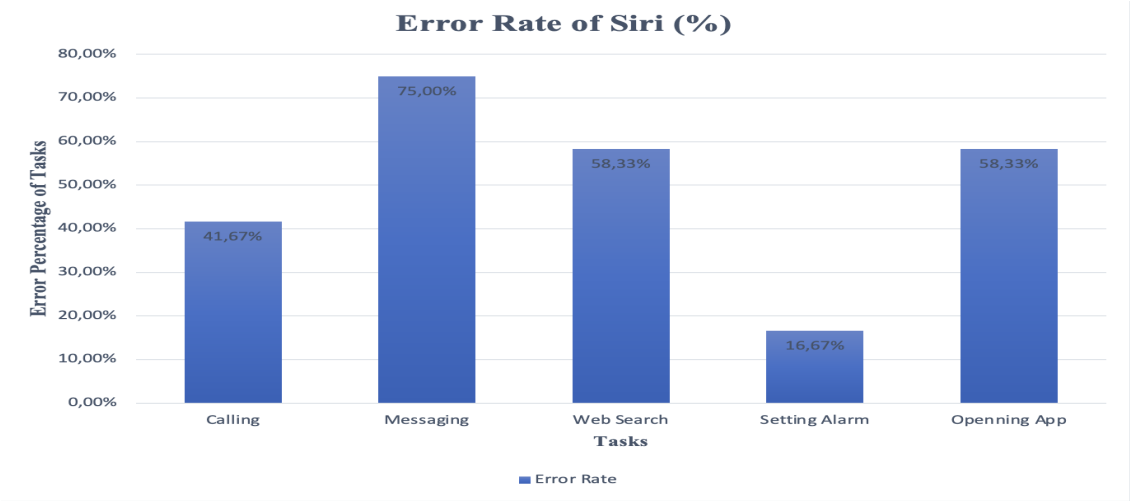

Figure 3.3 - Error Rate Percentage of Siri Voice Assistant

b. Qualitative Test Results

The qualitative data of the study involves the answers of interview questions which were asked to participants after conducting the tests. The answers to the open ended questions were recorded and analyzed. The feedback from participants are evaluated in 4 categories, which are ease of use, daily use preference, situational preference and accuracy of assistant.

1. **Ease of Use:** To begin with, some participants found Siri is easy to use after some trials or after learning how to use it. This suggests a learning curve, which can get better after some experience. Some of them claim that they have struggles with the execution since it is hard to describe the task to the assistant and initiate interaction with it .

2. **Willingness to use Siri in Daily Activities:** The responses vary in terms of using voice commands rather than manual commands. Several participants preferring manual actions or would not use Siri at all. On the other hand, a couple of them indicated that they see the utility of Siri in certain tasks during their daily lives.These results suggest that even the full replacement of voice commands will not happen for them, with an improvement Siri can become preferable for daily smartphone activities.

3. **Situational Preference:** The gathered responses from participants emphasize that Siri can be preferred in scenarios where manual execution is less convenient. Driving, working and when their hands are not free are some scenarios which are exemplified by participants. However, there is still a significant number of users who would not use Siri regardless of the situation.

4. **Accuracy and Execution of Commands:** Responses regarding accuracy of Siri are various. While some of the answers indicate that Siri does not understand and execute commands correctly, a significant portion of participants suggest that it completes the tasks without misunderstanding. The problems encountered by users are calling the wrong person, sending incorrect messages, and searching for wrong websites. These errors can be particularly discouraging for the users. On the other hand, these errors can be related with user-based problems such as pronunciation as mentioned in the methodology section. In that case, better voice recognition or interaction can be developed in the future.

ID	Participants	Do you think that Siri is easy to use?	Will you use Siri in your daily activities?	In which situations you prefer Siri over manual execution?	Does Siri accurately understand and execute your commands?
1	Esen Güçlü	Generally easy,hardest part is waking up the assistant.	I don't think so	Calling when I am at outside	Not in first trial
2	Gökhan Güçlü	I tried couple of times at the beginning, once I learned how to use it was easy.	I prefer manual execution	Calling and messaging while driving	I tried couple of times then I learned its usage
3	Nihal Dinç	I think it was very easy, I learned to use it very quickly.	I can use to call someone	Messaging when my hands are not free	Yes it does
4	Hasan Onur Dinç	If it can understand properly it is easy, otherwise it is not	Definetely	Calling and messaging	Fails in first trial of web searching
5	Aslı Şen	It is not easy in first trial	Maybe for some tasks	When I am a little bit far from my phone	Fails in sending correct message
6	Umut Şen	I use Siri sometimes so it was very easy	I use Siri for messaging and calling in general	driving and far from phone	Yes it does
7	Esra Özkan	It did not understand first command sometimes, so I struggled	I don't prefer to use	I wont prefer	It called wrong person in first trial
8	Kazım Özkan	I never used it before, once I learned, it was easy	Maybe, but not too often	Driving or working	Yes it does
9	Gülenay Akarca	I think it is cumbersome, but easy to use	I won't	I wont prefer	Yes it does
10	Mustafa Akarca	It was very easy	For some tasks it can be useful	driving and far from phone	Yes it does
11	Gökhan Salcan	For the first test, it did not understand me, so it was hard to explain the task	I won't use it	I wont prefer	I couldn't call the person I want at first
12	Aslım Salcan	It was generally easy but I do not prefer to use	I won't use it	I wont prefer	It sent wrong message

Table 3.1 - Post Interview Questions for the Participants

IV. DISCUSSION AND CONCLUSION

a. Interpretation of Results

The study clearly demonstrates that voice-based commands can be executed more rapidly than manual commands. Among the five tasks, messaging had the longest average completion time; however, when compared to the average completion time of manual messaging, the voice assistant's performance was still observed to be better. The variance in messaging task can be attributed to the complexity and variability inherent in text-based communication, which often requires more detailed and context-specific inputs that Siri may not handle as efficiently as manual typing. While Siri excels at tasks with straightforward and predictable commands, it struggles with tasks requiring more nuanced language and personalization. In line with the numerical and verbal results from the middle-aged participant group, while the tasks performed with Siri were quite fast, there were contradictions with the comments they verbally expressed. Despite most participants describing the use of the voice command assistant as easy after their initial interactions, their reluctance to use it in the future has led to contradictory outcomes in the research questions.

b. Technological Familiarity Among Middle-Aged Adults:

Another research question, the effects of age-related user motor and cognitive skills, can be offered as an explanation for this contrast. As mentioned above, there was a noticeable hesitance among middle-aged participants to adopt voice commands for everyday smartphone use. Concepts such as their inclination to give voice commands, limited experience with voice-command systems and the adaptation of new technologies explain the prejudices of people as well as their preference for more traditional, manual methods of interaction towards performing tasks with voice input. At the same time, the error rate research question to which an answer is sought can again be explained by the effects of users' indifference and their familiarity with technology.

c. Comparison with Literature:

The observed trends in our study are consistent with the broader literature, which suggests that while voice-command technologies like Siri can enhance user experience and efficiency, their effectiveness is closely tied to the nature of the task and the user's familiarity with the technology. Studies like those by Linda Wulf et al.(2014) and Carla Tubin et al.(2021) as discussed in the literature review, have highlighted the benefits of voice-command technology for older users, especially in terms of reducing reliance on visual and motor skills. However, our study adds a new dimension by focusing on middle-aged adults' technological proficiency and adaptability, a demographic that has received less attention in this context. Finally, usability heuristics for user interface design should be considered for future studies. Especially, error prevention, recognition rather than recall and match between system and the real-world principles are crucial for this system, since it is important to reduce the memory load of users by adapting the system to their daily language (Nielsen et al., 2009). Moreover, users tend to not use a system if they encounter lots of problems. To avoid errors, users should be guided properly by the assistant and the system should understand their way of talking. Similarly, according to Nielsen, the system should be designed to match with users' language (Nielsen et al., 2009). If the system adopts the pronunciation style of a particular user who is the owner of the phone, then while the error rate will decrease, usage of the voice assistant will increase respectively.

d. Practical Implications:

For UX designers, researchers, and technology developers, our study underscores the importance of designing voice command systems that cater to a broader range of user proficiencies and preferences. It highlights the need for intuitive and user-friendly interfaces that can accommodate users who are not as technologically adept. Furthermore, the study suggests that there is significant potential in developing educational programs or tutorials to help familiarize middle-aged adults with voice-command technology and alleviate apprehension towards its use.

e. Critique and Recommendations:

A notable limitation of our study is its relatively narrow focus on a specific set of tasks and a specific demographic group. Future studies could expand on this by including a more diverse range of tasks, particularly those requiring more complex and varied voice inputs. Inclusion of a more diverse participant pool in terms of age, technological background, and cultural context would also provide a more comprehensive understanding of the usability and acceptance of voice-command technologies like Siri.

f. Suggestions for Further Research:

Further research should delve into understanding the barriers and facilitators to technology adoption among middle-aged adults. This includes investigating how personal factors such as tech-savviness, cognitive abilities, and lifestyle influence the preference for voice-command technologies. Additionally, exploring the role of user interface design in enhancing the intuitiveness and accessibility of voice command systems for this demographic is crucial.

There is also a need to examine the long-term adaptation and satisfaction with voice-command technologies in everyday use. Longitudinal studies could provide insights into how user preferences and competencies evolve over time with continued exposure to these technologies.

Moreover, it would be beneficial to investigate the potential of adaptive voice-command systems that learn and adjust to individual user preferences and speech patterns. Dialect is a factor to consider. For the use of voice commands to become widespread and to increase user experience, the system must be adapted to the pronunciation patterns of the users. One of the biggest problems many users have with these systems is that they feel that the system does not understand them. Therefore, further studies and developments should include the dialect proneness of the digital assistant.

Lastly, given the growing importance of voice-command technologies in various aspects of life, studies should also focus on their impact on the social and psychological aspects of users,

particularly in the context of independence and self-efficacy among middle-aged and older adults. Considering the rapid proliferation of voice command tools today and the conveniences they offer, they can possess much greater competencies in the future. Attention should be drawn to two significant groups that could benefit from even such relatively simple daily phone activities. The first group consists of individuals without literacy skills. This segment represents those who need to adapt to this technology to communicate through voice commands, interact with the world via the internet, and not be excluded from other activities that could simplify their lives without any reading or writing ability. Another group is visually impaired individuals, although today touch screens can provide a braille keyboard, the voice command assistant allows them to perform a variety of phone actions more quickly.

V. REFERENCES

Chiaradia, T. S., Seabra, R. D., & Mattedi, A. P. (2019). Evaluating the usability of the Siri virtual assistant on mobile devices with emphasis on Brazilian elderly users. In *16th International Conference on Information Technology-New Generations (ITNG 2019)* (pp. 437-441). Springer International Publishing.

Tubin, C., Rodriguez, J. P. M., & De Marchi, A. C. B. (2021, October). Voice and touch interaction: a user experience comparison of elderly people in smartphones. In *Proceedings of the XX Brazilian Symposium on Human Factors in Computing Systems* (pp. 1-7).

Wulf, L., Garschall, M., Himmelsbach, J., & Tscheligi, M. (2014, October). Hands free-care free: elderly people taking advantage of speech-only interaction. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational* (pp. 203-206).

Nielsen, J. (2009). Chapter 5. In *Usability engineering* (pp. 115–151). essay, Kaufmann.