Accessibility within mobile computing for the visually impaired: Comparative study between Voice-to-text vs GUI on Grocery List App

Hwan Kim, Ferose Hameed, Chin Yee Lai, Steven Haddad
Dept. of Electrical Engineering and Computer Science
York University
Toronto, Ontario, Canada M3J 1P3
tk003@my.yorku.ca, feroseh@my.yorku.ca, chinyee8@my.yorku.ca, haddadst@my.yorku.ca
https://github.com/stevenhaddadcs/Grocery-List-App

ABSTRACT
This research addresses the accessibility challenges faced by individuals with visual impairments in mobile computing. A comparative study was performed pitting Voice-to-text input against soft keyboard input in our Grocery List app. Ten participants, composed of visually impaired and non-visually impaired persons, engaged in timed trials, CPU performance analysis, as well as a usability questionnaire. Our findings reveal that according to mean times for all participants Voice-to-text had a 26% faster task completion time to Soft keyboard, highlighting its efficiency. This number increases to 28% when considering only visually impaired participants. However, preferences varied among participants, with 70% of non-visually impaired participants favoring the Soft keyboard, while 90% of visually impaired participants preferred Voice-to-Text. The study also explored CPU performance, revealing a marginal advantage for the Soft keyboard. These findings contribute insights into the nuanced interplay between usability, performance, and user satisfaction in mobile computing accessibility for individuals with visual impairments.

Keywords
Mobile Computing, Voice-to-Text, Visually Impaired, Mobile User Interface, Usability, Android

INTRODUCTION

As more of the world moves online, the increase in need for mobile technology is rapidly growing. In today's society, mobile computing is almost a necessity in everyday life whether it's for work, socially, or personal use. The usability of mobile devices has grown from simple voice communications with physical buttons to fully connected touchscreen devices able to send messages or make video calls, high quality cameras, an effective interface, and so much more.

Usability can be measured by three attributes according to the International Organization for Standardization (ISO), which are; *effectiveness*; *efficiency*; *satisfaction* [2, p. 1-2]. Most mobile interfaces of today display these three attributes very well. With professional UI/UX developers and designers leading research on creating usable UI, most applications and interfaces are outstanding with positive remarks from consumers. Highly developed mobile computing also supports relatively solid performance with the latest mobile phones lasting almost a full day worth of battery.

Despite the advancements in mobile computing, one challenge is the support systems for individuals with impairments such as Visual Impairments. Visually impaired individuals have difficulty interacting with the mobile interfaces, so accommodations must be put in place. Voice-to-Text is a technology that has been present in mobile computing and it provides an alternative input method for users when they can't access the interface tactilely. Through speech recognition, devices are able to listen and help users with their tasks. The development of this technology is important because it is a dependent feature for the visually impaired, but also people who use their mobile devices on the move. With small screens and the shakiness from the person's movement, the majority of the user's focus must be on the device and the task they are trying to accomplish [2, p. 1]. This diverts attention from where they are going and their surroundings. Especially for drivers on the road this can be highly dangerous. To prevent such accidents and for general comfort, there are AI assistants implemented into most mobile devices.

Figures 1 and 2 showcase the AI digital assistants for Apple (Siri) and Samsung (Bixby). The digital assistants provide help accessing applications, searching on browsers, etc. They provide an alternative to quickly access data on your mobile device without having to look at or touch the interface.

This paper aims to explore more into voice-to-text technology and how it can assist users in their lives. The results of a comparative study between voice-to-text and using a keyboard through a regular mobile touchscreen keyboard was conducted. The difference between the performance and usability results of the differing input methods were then examined.

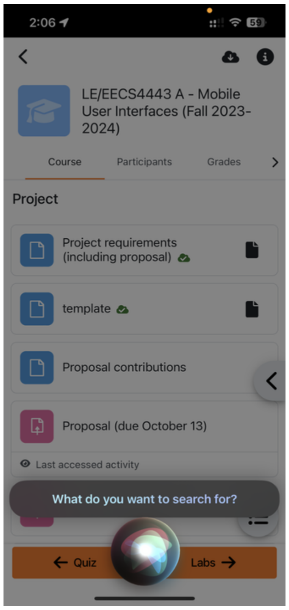

Figure 1. Siri, digital assistant for Apple.

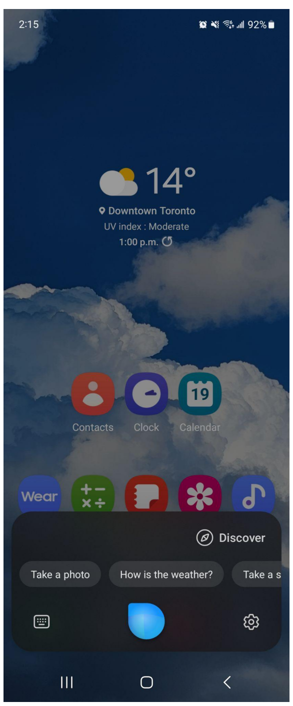

Figure 2. Bixby, digital assistant for Samsung.

Related Work

A study by Yadav and Chakraborty [7] examined the use of Google voice search to support informal learning in four to ten year old children. This research accessed 90 children to understand their use of Google voice search and its potential for informal learning. The study found that children often preferred voice search over textural query-based search to avoid spelling out words and typing, and for convenience. Findings concluded that, with parental guidance, Google voice search is able to support informal education in children aged 4 and older. This underscores the importance of voice technology in facilitating children's learning experiences online.

Cha et al. [1] proposed the design and implementation of a voice-based navigation for visually impaired persons. The paper discusses the urgent need for improved mobility support for blind persons as our society continues to expand. It highlights the challenges faced by blind people in independently navigating their surroundings and presents a solution in the form of a navigation system for blind people using Android-based smartphones that relies on Text-to-Speech (TTS) technology to provide precise location information and utilizes Google Maps API for map data.

Varghese Jacob and MacKenzie [6] conducted a study that explores the utilization of vibration and audio feedback as alternative methods for providing location guidance for the visually impaired when using touch screen devices. The primary objective is to develop an interface that enables users to interact with the touch screen through non-visual cues, facilitating corrective actions. The participants were tasked with locating a predefined set of buttons based on the feedback they received, and data was collected to assess user performance in terms of speed and efficiency. The results indicated vibration feedback required more time to reach the end goal compared to audio feedback, but more participants expressed their preference for vibration feedback as a practical issue in day-to-day life.

The effects of different modes of responding to messages on driving performance was studied by Kurtz et al. [3]. Voice-based input was the focus in this study as it seemingly offers fewer distractions, both visually and manually, at the human-machine interface and it can be perceived as less intrusive, as drivers simply record their spoken message by pressing a button instead of engaging in the process of typing the message out. The participants were asked to complete the Lane Change Task (LCT) in a driving simulator, both in baseline and dual-task conditions. In the dual-task scenarios, participants were required to respond to text messages on a smartphone either by using voice commands or by typing text messages, with varying degrees of task complexity. In-depth analysis of the dual-task runs revealed a noticeable advantage in using voice-based methods for responding to text messages, although this advantage diminished as task complexity increased. The study concluded that voice-based responses generally had a positive impact on driving performance.

Liu et al. [4] conducted a study on the desire of visually impaired individuals to be treated as ordinary consumers significantly influencing their approach to online shopping. The main goal is to broaden the knowledge within the HCI community concerning the needs, behaviors, and psychological aspects of online shopping as experienced by individuals with visual impairments. The visually impaired invest considerable effort in identifying and learning about products that align with their specific visual preferences. The participants exhibit a remarkable attention to the visual aspects of products, even though they themselves cannot perceive them visually. The study elaborates on the effort to appear ordinary as evident in the online shopping practices of visually impaired individuals, challenges faced, their shopping practices with social assistance, and their methods of learning about visual aspects, and provides design recommendations to facilitate and enhance these practices.

The oldest relevant discussion discovered originated in the "Comparison of Elderly and Younger Users on Keyboard and Voice Input Computer-Based Composition Tasks" by Orgozalek and Van Praag [5]. The experiment revealed a distinct indication that there was no improvement in composition tasks when utilizing voice input. However, statistically, participants exhibited a strong preference for voice input over the traditional keyboard method. The study emphasized the absence of a miraculous device capable of universally improving writing skills.

METHOD

This user study used both qualitative and quantitative measurements, as experience, understanding, and time were measured. The results were then analyzed with a focus on both usability and performance.

Participants

We recruited a total of 10 participants for our experiment, comprising 5 visually impaired individuals and 5 individuals without visual impairments. This was done to assess the usability of the app for the target audience as well as for general users.

All the participants had either normal vision, impaired vision, or low vision. Specifically, five participants who typically wore glasses were asked to remove them. None of the participants reported known visual or vestibular disorders, such as color or night blindness, dyschromatopsia, or a displacement of balance.

Participants were selected from local community centers, universities, and organizations catering to individuals with visual impairments. Demographically, participants were all aged between 20 to 60 years, with an equal distribution of males and females. Relevant prior experience with mobile technology and voice-to-text technology was noted.

The participation in this study was entirely voluntary, and participants were not provided with any incentives or compensation for their involvement. Individuals chose to take part in the study based on their own willingness and interest in contributing to the research.

Hypothesis Statement

Usability Hypothesis:

H_0: The time taken to create a grocery list using voice-to-text in our app will be equal to the time taken using traditional text input methods.($\alpha = 0.05$)

H_a: The time taken to create a grocery list using voice-to-text in our app will not be equal to the time taken using traditional text input methods.

Performance Hypothesis:

H_0: The CPU performance using voice-to-text in our app will be equal to the CPU performance using traditional text input methods.($\alpha = 0.05$)

H_a: The CPU performance using voice-to-text in our app will not be equal to the CPU performance using traditional text input methods.

Statistical Tests:

Considering the nature of the hypotheses, paired-sample t-tests were employed to compare the usability and performance between the two apps for both user groups. This test was deemed appropriate since the same participants used both apps, allowing for paired comparisons.

Apparatus

In our study, the apparatus chosen for all user performance trials was the Samsung Galaxy A51. This specific device was selected to maintain consistency across trials and minimize confounding variables that could arise from using different hardware.

Figure 3. Main page of Grocery List App

Procedure

The participants were categorized into two balanced groups: five individuals with visual impairments and five without. To control for order effects, we employed a counterbalancing strategy, wherein the groups performed their tasks in reverse sequence to one another. Before commencing the trials, each participant was allotted time to become acquainted with the application to ensure familiarity with the interface.

To evaluate the voice-to-text and keyboard input methods, participants were instructed to conduct a series of tasks. For the voice input trial, they were asked to use voice commands to add items such as baby formula, avocados, and apples to the grocery list, then modify the list by deleting items yogurt and chips and adjusting the quantity of items brocolli and apples to 3, and lastly clearing the entire list using the 3 dots menu clear option.

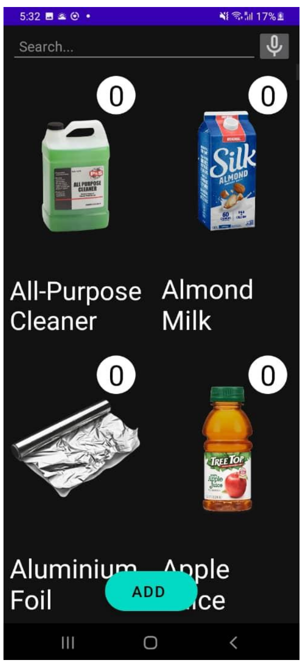

Figure 4. Voice-to-text and keyboard search method to add items.

For the keyboard input trial, the same items were to be searched and manipulated to produce the same end result except using keyboard input.

This methodical approach allowed us to directly compare the effectiveness and efficiency of voice-to-text versus keyboard inputs in a controlled environment, as depicted in the accompanying figures that illustrate the application's main page and the item adjustment interface.

Figure 5. Increment or decrement quantity of item

Design

This study employed a 2 x 2 mixed design. The two factors that were evaluated were the type of input method (voice-to-text vs. keyboard) and the type of user (visually impaired vs. not visually impaired). Each participant used both methods, enabling a comparison of their performance and preferences. Usability metrics, such as task completion time—the time each participant took to complete a given task in both apps—and performance metrics, such as CPU performance, were measured. Subjective feedback was gathered through a usability questionnaire.

User Performance study

Independent Variables: Keyboard, Voice-To-Text

Group: Visually Impaired, Not Visually Impaired

Dependent Variables: Task Completion Time (seconds)

CPU Performance study

Independent Variables: Keyboard, Voice-To-Text

Dependent Variables: CPU Performance (%)

Usability Feedback

Subjective feedback from participants regarding the ease of use and overall experience with each app, gathered through the usability questionnaire.

Each participant used both apps, allowing for a comparison of their performance and preferences. Considering that each of the 10 participants used both apps, the total number of trials for user performance study was 10 participants x 2 input methods x 3 trials each = 60 trials. The total number of trials for CPU performance study was 2 input methods x 3 trials each = 6 trials.

RESULTS AND DISCUSSION

The group effect on task completion time was not statistically significant ($F_{1,8}$ = 0.085, ns). So, counterbalancing had the desired effect of eliminating order effects. The main results of interest are the usability measures for task completion time and performance measures for CPU performance. Additionally, the subjective impressions of the participants regarding the two methods are of interest.

Task Completion Time

The grand mean of task completion time was 87.6 seconds. The voice-to-text method was 26% faster, with a mean task completion time of 74.57 seconds, compared to the soft keyboard method, which had a mean task completion time of 100.63 seconds. This result indicated that the voice-to-text method outperformed the keyboard method.

The paired t-test result yielded a T-statistic of 66.678 with a corresponding p-value of 0.0002. The calculated confidence interval indicated that the true difference between Voice-to-Text and Keyboard means fell between -27.7381 and -24.3753. The negative values signified that Voice-to-Text had a lower mean than the keyboard. With 95% confidence, the interval excluded zero, suggesting that the difference was not equal to zero. The p-value of 0.0002, significantly smaller than 0.05, indicated a statistically significant difference. Consequently, the null hypothesis, the time taken to create a grocery list using voice-to-text in our app would be equal to the time taken using traditional text input methods, was rejected.

One of the reasons for the significant difference in task completion time was that it took a longer time to spell each word when using the keyboard search method, whereas voice-to-text search method took less time to obtain the same search result. Another contributing factor was that misspelling words in the keyboard search method resulted in no search results, leading to a longer overall time compared to voice-to-text.

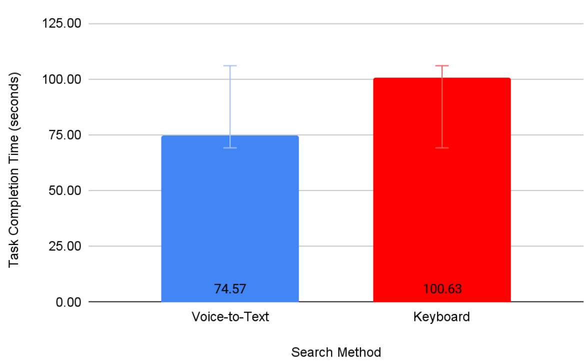

Figure 6. Chart depicting the average time for task completion (sec) for Voice-to-text and Keyboard search methods.

The participants with visual impairment, with an overall mean task completion time of 91.62 seconds, took 9.6% longer on average compared to those without visual impairment, who had a mean task completion time of 83.58 seconds.

However, it is noteworthy that according to the means for those with visual impairment the trials with the voice-to-text input method were 28% faster than the trials using the soft keyboard. With the mean completion time of voice-to-text being 76.76 seconds and the mean completion time of the soft keyboard at 106.47 seconds.

In comparison to non-visually impaired participants, it was observed that the mean completion time of voice-to-text 72.38 seconds and the mean completion time of the soft keyboard at 94.78 seconds. Resulting in just a 24% faster time for voice-to-text in comparison to the soft keyboard input.

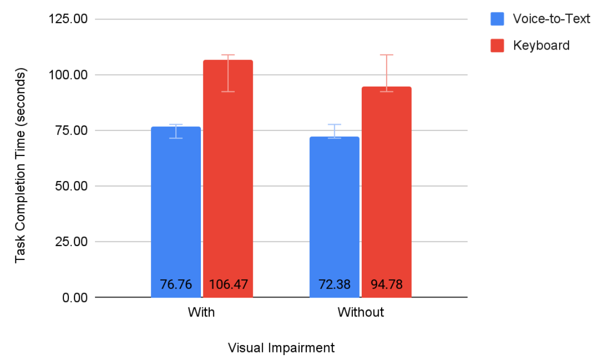

Figure 7. Chart depicting visually impaired participants's average task completion time (sec) for Voice-to-text and Keyboard.

CPU Performance

The grand mean for CPU performance was 8.5%. The keyboard search method consumed 30.0% more CPU performance, with a mean of 10% over three trials. This suggested that the voice-to-text method exhibited better CPU performance compared to the keyboard method, although the observed difference was not statistically significant (T-statistic = 1.9640, p = 0.1885). The confidence interval estimated that the difference between the voice-to-text and keyboard methods was between -9.57 and 3.57. Therefore, with 97% confidence, the difference was equal to zero, and the p-value of 0.1885 is greater than 0.03. Consequently, we lacked sufficient evidence to reject the null hypothesis regarding performance.

The reason for no significant difference in CPU performance can be attributed to the implementation of a straightforward speech-to-text functionality in the Android app using the SpeechRecognizer class. While CPU performance is typically influenced by the complexity of voice recognition software, in this instance, the app was designed with a voice recognizer configured to retrieve a maximum of 5 results. Subsequently, these results were compared with the list of items stored in the app.

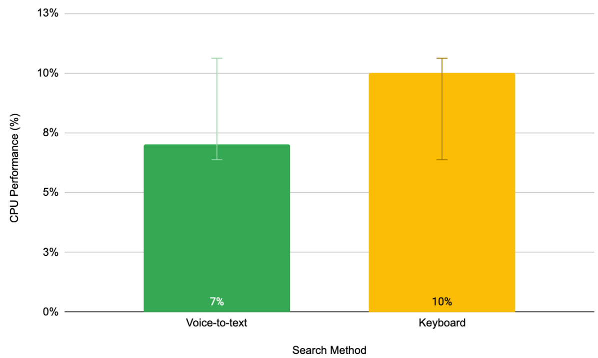

Figure 8. Chart depicting the average CPU utilization of Voice-to-text vs Keyboard for 3 trials each respectively.

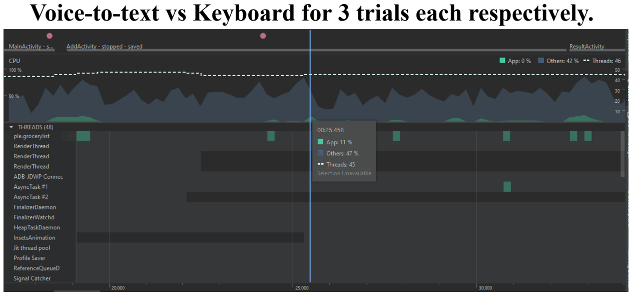

Figure 9. Screenshot of CPU performance when keyboard search method was performed on 1 trial

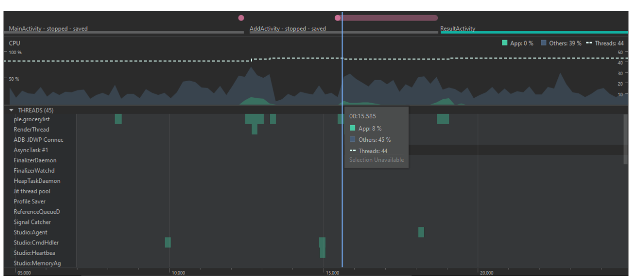

Figure 10. Screenshot of CPU performance when voice-to-text search method was performed on 1 trial

Participant Feedback

We assessed user satisfaction through feedback, and the overall findings regarding subjective preference revealed that participants without visual impairment with a percentage of 70% favored the keyboard method over the voice-to-text method. On the other hand, 90% participants with visual impairment expressed a stronger preference for the voice-to-text method compared to the keyboard method.

One of the participants in favor of the keyboard method cited reason such as

difficulty using voice-to-text in noisy surroundings

Another participant who preferred the keyboard method said the following

challenges in obtaining accurate results with my accent

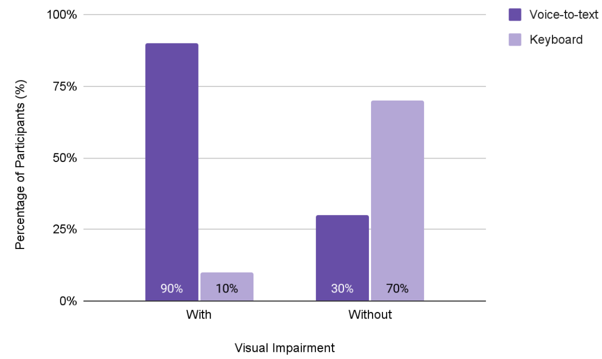

Figure 11. Chart depicting the percentage of participants with and without visual impairment on their preferred search method

In contrast, participants who preferred the voice-to-text search method mentioned that it was

easier to use due to difficulties in spelling certain words

Most participants with visual impairment had similar reason as follow

more accessible for individuals with visual impairment

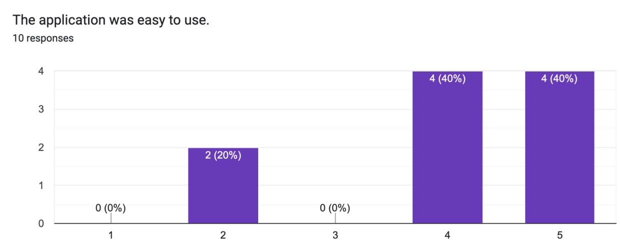

Figure 12. Chart depicting the rating for usability

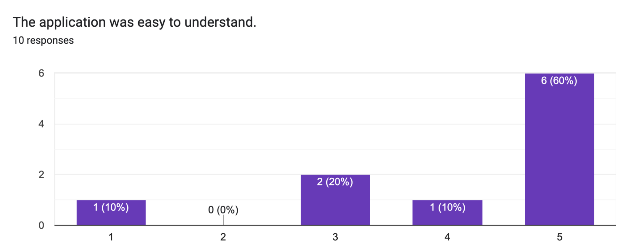

Figure 13. Chart depicting the rating for understanding

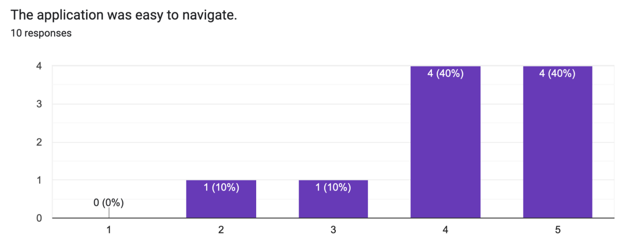

Figure 14. Chart depicting the rating for navigation

Overall, the participants provided favorable and preferential ratings for using the application, reporting an overall positive experience in terms of usability, navigation, and understanding. Participants also provided suggestions on improving the applications used for the user study. A participant recommended enhancing the voice input button by making it larger and placing it in a more accessible area. Another suggestion from a different participant was to incorporate a delete button for easily removing items.

CONCLUSION

In a rapidly evolving world of mobile computing, Accessibility features must be considered for those that need the accommodation. With that in mind, developing a mobile UI must always consider the three attributes to usability, which are *effectiveness*, *efficiency*, and *satisfaction* [2, p. 1-2]. Effectiveness in terms of how well the UI executes its tasks, efficiency as in amount of time and resources that the UI must use to complete its tasks, and finally satisfaction, meaning the overall level of contentment that the users feel when interacting with the UI. That being said, the goal of this user study was to analyze and compare the usability and performance between two common mobile input methods, that being, Voice-to-text vs Soft keyboard. In doing so, a mobile grocery list app was developed with accessibility features such as the variety of input methods to search for a certain item to add to your current list. Voice-to-text or a soft keyboard could be used to search for a specific item.

After the development of the app, a user study was conducted based on the search features of the app. A set number of participants were recruited with half of the population having some sign of visual impairment. The participants were instructed to complete a list of tasks given to them using the app and its two input methods. These trials were timed in order to track the usability of the two input methods. The results of the usability study concluded that Voice-to-text was faster than the soft keyboard which was contrasting to the hypothesis made before the trials took place. Furthermore, participants with a visual impairment overall took slightly longer to complete the tasks compared to those without visual impairments.

As for the performance study, it was conducted separately with the usability study. A mobile device was used to measure the CPU performance using the profiler in the Android Studio software. The CPU consumption was tracked by its peak when utilizing the two input methods in the grocery list app. Although the difference was statistically not too significant, the soft keyboard was still slightly better than the Voice-to-text input method, but not enough evidence was found to reject the null hypothesis.

In conclusion, following the user study, it is noted that when comparing Voice-to-text vs Soft keyboard in terms of usability and performance, Voice-to-text showed an advantage in input speed while the soft keyboard displayed a slight advantage in terms of CPU performance. Therefore there is no clear answer on which input method is superior. In the end, preference or user satisfaction is what decides which input method to use. The opinions of the participant's satisfaction were divided when assessing the answers to the questionnaire that was imposed after the user study. Some individuals posited that the keyboard was easier to use with dissatisfaction with the voice input due to factors such as surrounding noise or the speech recognizer not detecting accents properly. Meanwhile, other users such as some of the visually impaired individuals stated that the Voice-to-text feature was preferred because it is faster for them since the keyboard's characters are small and tedious.

Limitations were evident in the user study, such as time constraints, low participant sample size, difficulty recruiting participants with more severe visual impairments, and the overlooking of ethnic backgrounds, potentially impacting voice recognition accuracy and the study's generalizability. That being so, future research should focus more into gesture based input on top of the voice input to further investigate the comparison using a more realistic research environment and to draw conclusions on how to accommodate better for those that need it. Furthermore, while this research study focused on visually impaired people in terms of quality of sight, future research should branch deeper into different kinds of visual impairment like colour blindness. Ultimately, this comparative study provided a brief glimpse into the current state of accommodation in today's mobile technology and it provided a better understanding of where to further develop and research within the topic of accommodation in mobile computing.

ACKNOWLEDGMENT

We are grateful to the professor Dr. Marios Fokaefs, the York University library, our participants, and research sources for their essential and tremendous contributions to this research paper.

REFERENCES

1. Cha, J.S., Lim, D.K. & Shin, Y.-N. Design and implementation of a voice based navigation for visually impaired persons. 5. 61-68 (2013).

2. Harrison, R., Flood, D. & Duce, D. Usability of mobile applications: *literature review and rationale for a new usability model*. J Interact Sci 1, 1 (2013).

3. Kurtz, M., Oehl, M., & Sutter, C. "Just One Short Voice Message" *–Comparing the Effects of Text- vs. Voice-based Answering to Text Messages via Smartphone on Young Drivers' Driving Performances*. Safety 7(3), 57 (2021).

4. Liu, G., Ding, X., Yu, C., Gao, L., Chi, X., & Shi, Y. "I Bought This for Me to Look More Ordinary": *A Study of Blind People Doing Online Shopping. CHI 2019.. Association for Computing Machinery*, 372, 1-11 (2019).

5. Ogozalek, V. & Van Praag, J. Comparison of elderly and younger users on keyboard and voice input computer-based composition tasks. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '86). Association for Computing Machinery*, 205–211. https://doi.org/10.1145/22627.22372 (1986).

6. Varghese Jacob, S. & MacKenzie, I. S. Comparison of Feedback Modes for the Visually Impaired: Vibration vs. Audio. *Proceedings of 20th International Conference on Human-Computer Interaction - HCII 2018 (LNCS 10907)*, 420-432 (2018).

7. Yadav, S. & Chakraborty, P. Using Google voice search to support informal learning in four to ten year old childre*n*. *Educ Inf Technol* 27, 4347-4363 (2022).