μGeT: Multimodal eyes-free text selection technique combining touch interaction and microgestures

Gauthier Robert Jean Faisandaz, Alix Goguey, Christophe Jouffrais, Laurence Nigay

▶ **To cite this version:**

Gauthier Robert Jean Faisandaz, Alix Goguey, Christophe Jouffrais, Laurence Nigay. μGeT: Multimodal eyes-free text selection technique combining touch interaction and microgestures. 25th ACM International Conference on Multimodal Interaction Paris (ICMI 2023), ACM Special Interest Group on Computer-Human Interaction, Oct 2023, Paris, France. pp.594-603, 10.1145/3577190.3614131 . hal-04353214

HAL Id: hal-04353214
https://hal.science/hal-04353214

Submitted on 19 Dec 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

µGeT: Multimodal eyes-free text selection technique combining touch interaction and microgestures

Gauthier Faisandaz
Alix Goguey
gauthier.faisandaz@gmail.com
alix.goguey@univ-grenoble-alpes.fr
Univ. Grenoble Alpes, CNRS,
Grenoble INP, LIG
Grenoble, France

Christophe Jouffrais
christophe.jouffrais@cnrs.fr
CNRS, IPAL
Singapore

Laurence Nigay
laurence.nigay@univ-grenoble-alpes.fr
Univ. Grenoble Alpes, CNRS,
Grenoble INP, LIG
Grenoble, France

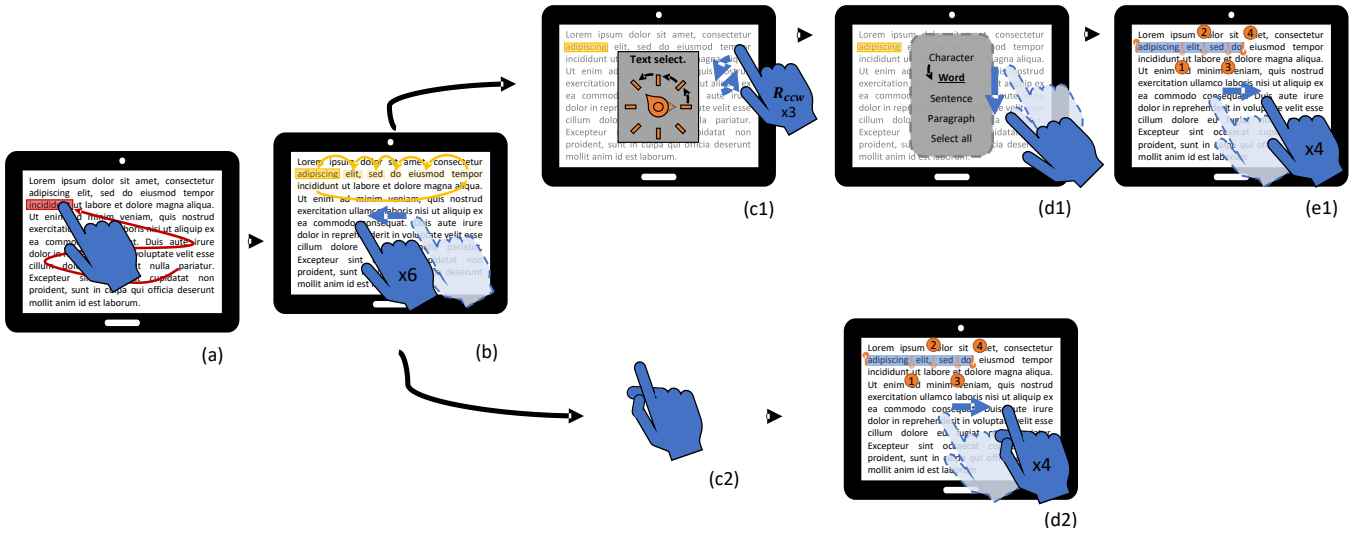
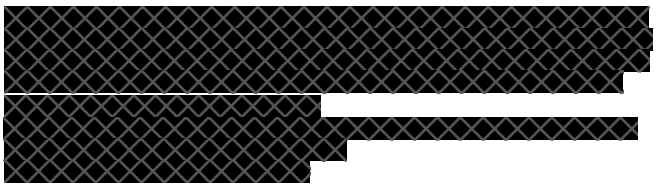


Figure 1: Scenario of text selection, from left to right. The first two steps are common to the two techniques: (a) direct pointing over a word, (b) fine tuning the highlight. Upper line steps for the VO technique: (c1) setting the rotor menu on "Text selection", (d1) setting the sub-menu to "Word", (e1) moving the selection handle to the desired position. Lower line steps for the µGeT technique: (c2) touching the middle finger with the thumb, (d2) moving the ending handle to the desired position.

ABSTRACT

We present µGeT, a novel multimodal eyes-free text selection technique. µGeT combines touch interaction with microgestures. µGeT is especially suited for People with Visual Impairments (PVI) by expanding the input bandwidth of touchscreen devices, thus shortening the interaction paths for routine tasks. To do so, µGeT extends touch interaction (left/right and up/down flicks) using two simple microgestures: thumb touching either the index or the middle finger. For text selection, the multimodal technique allows us to directly modify the positioning of the two selection handles and the granularity of text selection. Two user studies, one with 9 PVI and one with 8 blindfolded sighted people, compared µGeT with a baseline common technique (VoiceOver like on iPhone). Despite a large variability in performance, the two user studies showed that µGeT is globally faster and yields fewer errors than VoiceOver. A detailed analysis of the interaction trajectories highlights the different strategies adopted by the participants. Beyond text selection, this research shows the potential of combining touch interaction and microgestures for improving the accessibility of touchscreen devices for PVI.



CCS CONCEPTS

• **Human-centered computing** → **Interaction techniques**; **Accessibility technologies**.

KEYWORDS

Microgesture, touch, accessibility, visual impairment.

ACM Reference Format:
Gauthier Faisandaz, Alix Goguey, Christophe Jouffrais, and Laurence Nigay. 2023. µGeT: Multimodal eyes-free text selection technique combining touch interaction and microgestures. In *INTERNATIONAL CONFERENCE ON MULTIMODAL INTERACTION (ICMI '23), October 9–13, 2023, Paris, France.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3577190.3614131

1 INTRODUCTION

At least 253 million people live with moderate to severe visual impairment worldwide [2], to whom we will refer in this article as PVI (i.e., People with Visual Impairments). PVI rely more and more on touchscreen devices for everyday tasks [1, 19, 27] such as reading, social interaction, outdoor navigation, object recognition, etc. Improving accessibility of touchscreen devices for PVI is thus very important [27, 36].

Current touchscreen devices rely on accessibility tools, (Android TalkBack and iOS VoiceOver[1]), commonly named as "screen readers". Previous work showed that they are still perfectible [3, 11]. Screen readers "read" the screen's content (2D) as a list of items (1D auditory feedback). Although PVI can access 2D digital content and interact with it, each item must be checked one at a time – which increases the time and number of interactions required to perform a particular action. As current accessibility tools are tedious to use with the touch modality alone, we explore how thumb-to-finger (TTF) microgestures (μG) can be combined with touch inputs in order to simplify the interaction for PVI in routine tasks. TTF microgestures are small, single hand movements made with the thumb on other fingers of the same hand and Faisandaz et al. [14] showed that PVI can perform TTF μG in combination with touch inputs.

In this paper, we focus on the task of text selection, after having consulted fourteen PVI We present μGeT, a technique that enriches touch interaction using quasi-modes triggered by TTF μG (i.e., thumb touching either the index or the middle finger) in a task of eyes-free text selection. The two input modalities, i.e., touch and TTF μG, are combined in a complementary way and the temporal aspect of the combination is parallel, as defined by Serrano et al. [37]. For text selection, the resulting multimodal technique allows users to directly modify the location of the two selection handles (using TTF μG) and the granularity of text selection (using left/right/up/down touch swipes). We conducted two user studies, one with 9 PVI and one with 8 blindfolded sighted people to compare μGeT with a VoiceOver-like (VO) technique as a baseline. Despite a large variability in performance, the two user studies showed that μGeT is globally faster for PVI and similar to VO in terms of text selection time. It also yields fewer errors than VO for both population. A detailed analysis of the interaction trajectories highlights the different strategies adopted by the participants. This study shows how TTF μG can effectively be used by PVI to augment touch inputs, shorten interaction and ultimately simplify routine tasks such as text selection that usually require navigating menus to change parameters.

2 RELATED WORK

Our work builds on previous research on enhancing touch interaction and on μGesture interaction to address a rarely considered problem: increasing the bandwidth of touch screen interaction for accessibility purposes.

2.1 Enhancing the touch modality

Approaches to enhancing touch interaction involve adding one or more modalities as an additional dimension to a touch input to trigger different commands [22]. Also called "touch overloading", approaches on current touchscreen devices commonly use time (e.g, dwell, hold), repetition (e.g., double tap) and/or multiple contacts (e.g., pinch) as additional dimensions to augment the touch modality. Some approaches leverage the screen capabilities to capture shear forces [21], various levels of pressure on the surface [17, 20, 33], unimanual [18, 29] or bimanual [5] multitouch to trigger different commands. Other approaches are based on the identification of the part of the finger (e.g., knuckle, tip, nail) touching the surface, each triggering different commands [22, 31]. Rhythmic patterns were also studied [15]. These approaches either involve complex interactions to perform and learn, or rely on vision to be used seamlessly. Needless to say, drawing-based input requires spatial accuracy [26] and pressure-based input also needs visual feedback such as gauges to be controlled precisely [20, 39]. Hence, these approaches lack accessibility for PVI.

For PVI, we nevertheless note that Kane et al. [24] explored multitouch and bimanuality to design techniques helping users to navigate the screen faster and understand spatial layout. The "Edge projection" technique, specifically, projects the items on screen onto the x- and y-axes on the left and lower edge. Users can then find the required item along the edges, then drag both fingers to the interior of the screen to find the actual location of the item. This technique adds an extra step in the interaction for each input and is not usable in mobile situations. Indeed, this technique requires both hands to be available at all times and is intended for fairly large touchscreens placed on a flat surface. Li et al. [30] used the device's gyroscope to leverage kinesthetic and proprioceptive abilities of users to recall distinct spatial locations in front of them. However, this technique requires mid-air movements that are not discrete for public use and can induce fatigue ("gorilla-arm effect" [23]).

Some approaches specifically tailored for PVI rely on tangible overlays to provide tactile cues [12]. They can be "read" via tactile perception and thus convey spatial information more easily. Examples include interactive raised-line maps [7], tangible widgets (e.g., slider, menu) to be placed on top of touchscreen devices [25] and tactile overlays with horizontal lines to aid reading long texts such as books [13]. However these approaches can be expensive and cumbersome [25, 35]. As each application has a different layout, these approaches would require the use of many custom overlays. Besides, once produced, these tangible add-ons are static. They prevent dynamical content updates (e.g., zooming on a map).

Finally, for PVI, voice can be used to enhance touch interaction. Indeed, voice is frequently used by PVI, notably for textual inputs or punctual discrete commands such as opening an application. Previous research has shown that touch input combined with voice can improve interaction on touchscreen. For instance, Zhao et al. [45] studied the complementary usage of input touch and voice to edit text and correct errors and have showed that the multimodal technique greatly improves the text editing and correcting performance compared to the touch-only modality. But the complementary use of touch and voice relies on a visual context for touch, which makes them unusable by PVI. In addition PVI can be reluctant to talk to their device in public spaces [34], or feel they lack control [4, 10, 44]. That is why, we focus on a new modality with promising characteristics to enrich touch interaction while remaining accessible for PVI: microgestures (μGestures, μG).

[1]Talkback for Android and VoiceOver for iOS

2.2 µGestures for PVI

Wolf et al. describe µGestures (µG) as small motions executed with the hands and fingers [41, 42]. Chan et al. define single-hand µG as finger gestures performed by one hand on itself [9]. µG can be performed eyes-free [14, 41] and are cognitively undemanding [14, 32, 38, 42], making them particularly suited for PVI [14].

Chan et al. did an elicitation study of 1632 µG, in which the thumb was used 88% of the time: these in particular are what we call thumb-to-finger (TTF) µG (i.e., the thumb touching another finger). TTF µG define a promising modality to complement touch and increase expressivity. For instance, a touchscreen device allowing only for simple and double taps, combined with only two TTF µG (e.g., the thumb touching the index or the middle finger), could provide 6 types of tap (i.e., simple tap and double tap without the thumb touching any finger, same with the thumb touching the index, and same with the thumb touching the middle finger). Thus, several studies used TTF µG to increase touch input expressivity [5, 40, 43] and motivate our work on combining TTF µG and touch for PVI.

But for PVI, work on TTF µG is scarce. Boldu et al. developed a head-mounted camera to assist PVI in grocery shopping [6]. The interaction was triggered with a TTF µG performed on a ring mounted on the index. They argued that the TTF µG was the "optimal input gesture" in their context, because it is hand-free and can be performed with minimal effort alongside other tasks. Faisandaz et al. [14] conducted a study to evaluate and compare the usability and comfort of 33 TTF µG to be used in conjunction with touch modality in an eyes-free situation. They found that the absence of vision, combined with the constrained position of the hand, hampers accuracy when performing TTF µG. Consequently, they put forward a set of 8 TTF µG that can be used while the index is touching a surface. They further demonstrated 3 TTF µG in a map-based application to 7 PVI. All participants found this multimodal technique usable and comfortable and were really interested in the interaction possibilities TTF µG could bring. That is why, we study both quantitatively and qualitatively an interaction technique combining 2 TTF µG from this set of 8 TTF µG [14] and touch interaction in a "daily-life" task, without vision.

3 TECHNIQUES

We chose the "daily-life" task of text selection after conducting a 1h long focus group with 14 PVI recruited in a special education center. After introducing TTF µG and context, we made 3 groups and discussed their phone usage (tasks, app, issues) (10 min). We asked them how they would use TTF µG with their phone (10 min), and had them discuss these ideas altogether (10 min). We finished with an open discussion. We also conducted four 1h-long interviews: 3 about general phone usage (with a VO expert and 2 novice), 1 about text manipulation (with a VO expert). Participants of the focus group and interviews were aged from 24 to 58. These participants found that TTF µG could be useful for games and software applications to switch options and parameters. Implicitly, they suggested that the multimodal approach could be promising for simultaneous tasks. They also mentioned how the current accessibility tools and their menus force them into linear interactions, which they found long and tedious, to the point that most of them refuse to use common features of their phone, such as text edition, copying and pasting. For this reason, we chose to focus on a text selection technique without linear menu navigation. Text selection consists of setting the boundaries of a selection field, delimited by two markers (i.e., selection handles). Before describing the designed technique, µGeT, that combines TTF µG and touch, we first recall how text selection is performed with current accessibility tools. In our study, we considered the iOS VoiceOver accessibility tool as the baseline technique.

3.1 Baseline technique: VoiceOver-VO

We use the default configuration of the iOS VoiceOver (VO) accessibility tool, which can be customized in the iOS accessibility settings. VO uses discrete touch inputs to navigate a radial menu with contextual commands (called a rotor, Figure 1-c1), via quick clockwise (CW) and counterclockwise (CCW) rotating gestures with two fingers. There are 8 items in the rotor. Each item of the rotor is a persistent mode. In the following, we refer to the 1st item as being the "North" one, and count up clockwise. To select a portion of text, users must first highlight an entity (e.g., a character, a word, a sentence, in grey highlight IPSUM in the following example) either through a drag on the screen, highlighting the entity being hovered, or by performing a 1-finger swipe left/right (also called flick), moving the highlight over to the previous (or next) entity. Placing the rotor in the 3rd (Character), 4th (Word) or 5th (Sentence) position changes the entity type being highlighted. By default, words are highlighted.

The anchor (i.e., leftmost selection marker, in blue ↓) is placed first, its position being set at the beginning of the highlighted entity (i.e., left of its first character). The anchor cannot be moved otherwise. To select text, users then move the selection handle (i.e., rightmost selection marker, in blue ↑) initially placed at the anchor position. To move the selection handle, the rotor must be placed on the 8th ("Text selection") position. In this mode, the handle is moved leftwards or rightwards using 1-finger swipe left/right. The granularity at which the handle is moved can be changed using 1-finger swipe up/down, circling through 5 sub-modes (Figure 1-d1): "character" (CbC), "word" (WbW), "sentence" (SbS), "paragraph" or "select all". The granularity is initially set to "word". Each change (rotor, sub-modes, updated selection) triggers an audio feedback that reads the menu item or the last entity that was highlighted/selected. Everything comprised between the anchor and the selection handle is considered selected (in yellow highlight IPSUM). To help the users, we added a double tap gesture, which triggers an audio feedback that reads the current selection if any, or simply stops the current audio feedback.

- Highlight - LOREM IPSUM ET DOLOR SIC AMET
- Selection - LOREM ↓↑IPSUM ET DOLOR SIC AMET
- {Input}
- Selection - LOREM ↓ IPSUM ↑ ET DOLOR SIC AMET

3.2 Designed multimodal technique: µGeT

µGeT is a text selection technique that combines µG and touch to augment the input bandwidth. The design rational is to use the index and middle fingers as metaphors of text selection handles. When contacting and keeping their thumb pressed on the index or the middle finger, users respectively "grab" the selection leftmost

or rightmost handle (respectively called First and Last handles, in blue in the previous example ↓↑). These "grabbing" gestures (called μTAP hereafter) trigger quasi-modes (i.e., a mode only active if the trigger action is maintained) in which one can edit the position of the respective handle. Once either of the handles is grabbed, users can move them with three levels of granularity: 1) character level (CbC), through short horizontal swipes which trajectory length is less than 150px; 2) word level (WbW), through long horizontal swipes which trajectory length is more than 150 px; 3) sentence level (SbS), through vertical swipes. Similar to VO, users perform the initial placement of the handles by highlighting an entity through a drag on the screen, placing both handles at the beginning of the word being hovered. Prior to any selection (i.e., no handle was yet grabbed and moved), swipes can be used to adjust the initial placement of the handles. If a selection has been started, drag gestures and swipe gestures on screen with no grabbed handles, have no effect. To reset the selection and start a new one, users must perform a tap on the pinky fingernail. Similar to VO, each change (quasi-mode, updated selection) triggers an audio feedback and a double tap gesture (to listen the current selection if any or to stop the current audio feedback) has been implemented.

The accompanying video figure shows examples of text selection using VO and μGeT.

3.3 Theoretical comparison of μGeT and VO

We hypothesize that navigating menus using VO degrades usability and user experience, as it lengthens completion time, complicates the interaction path, and more generally interrupts the task at hand. We specifically designed μGeT to avoid interruption due to menu navigation by having all selection tools readily available. However, it is not easy to find representative PVI as they are often difficult to reach through standard means of participant recruitment. Thus we first validated our approach from a theoretical perspective: we modeled the tasks with the Storyboard Empirical Modeling tool (StEM) [16], an extension of the Finger-Level Model (FLM) [28] itself based on the Keystroke-Level Model (KLM) [8]. StEM allows us to predict completion times, which we use to compare optimal interaction trajectories using both techniques (μGeT and VO). Figure 2 shows an optimal interaction trajectory for each task. We consider optimal the interaction trajectory using the least number of inputs for a given starting point. There are several optima: an optimum is calculated for a given starting point. In figure 2, for μGeT, all handles are placed at the beginning of the first word containing part of the target, and for VO, all handles are placed at the beginning of the target itself (including inside a word since VO allows a precise initial placement of the handles). We use the following set of StEM operators with times provided by the tool for each modeled task: 1) T (tapping), pressing an on-screen target without knowledge of the starting finger position; 2) R (rotation), rotating an on-screen object with two fingers; 3) F (flicking or swiping), a ballistic linear movement in one of the cardinal directions (up, down, left, right). We also call them swipes in our paper. However, in order to compare the two techniques, we made three assumptions. First, StEM only applies to touch interaction, thus we added a fourth operator (μT) which represents a thumb tap on either the index or middle finger. We have assumed that it is similar to a regular tap (T) in terms

Task	Target & handles starting point ↓↑	μGeT	VO	Delta (μGeT – VO)	VO menu action ratio	μGeT time (s)	VO time (s)	Delta time (s)
T1) Mid-word	μGeT: ↓↑Forty VO: F↓↑orty	$\mu T_{mid}\, F^{\rightarrow}_{long}\, F^{\leftarrow}_{short}$ $\mu T_{ind}\, F^{\rightarrow}_{short}$	(starting from rotor's 3rd position) $R_{ccw}\, R_{ccw}\, R_{ccw}\, F^{\downarrow}\, F^{\rightarrow}\, F^{\rightarrow}\, F^{\rightarrow}$	8 - 14 - 6 actions	4/14 29%	4.2	6.2	-2.0
T2a) 2-words-and-half	↓↑Two Three Four	μT_{mid} $F^{\rightarrow}_{long}\, F^{\rightarrow}_{long}\, F^{\rightarrow}_{long}\, F^{\leftarrow}_{short}$	(starting from rotor's 4th position) $R_{ccw}\, R_{ccw}\, R_{ccw}\, R_{ccw}$ $F^{\rightarrow}\, F^{\rightarrow}\, F^{\rightarrow}\, F^{\downarrow}\, F^{\leftarrow}$	9 - 18 - 9 actions	5/18 28%	4.9	8.0	-3.1
T2b) 2-words-and-half	μGeT: ↓↑Three Four Five VO: Th↓↑ree Four Five	$\mu T_{mid}\, F^{\rightarrow}_{long} F^{\rightarrow}_{long} F^{\rightarrow}_{long}$ $\mu T_{ind}\, F^{\rightarrow}_{short}\, F^{\leftarrow}_{short}$	(starting from rotor's 3rd position) $R_{ccw}\, R_{ccw}\, R_{ccw}\, F^{\rightarrow}\, F^{\rightarrow}\, F^{\rightarrow}$	12 - 12 +0 actions	3/12 25%	6.4	5.3	+1.1
T3) Four-words	↓↑Nine Ten Eleven Twelve	$\mu T_{mid}\, F^{\rightarrow}_{long}\, F^{\leftarrow}_{long}$ $F^{\rightarrow}_{long}\, F^{\leftarrow}_{long}$	(starting from rotor's 4th position) $R_{ccw}\, R_{ccw}\, R_{ccw}\, R_{ccw}$ $F^{\rightarrow}\, F^{\leftarrow}\, F^{\rightarrow}\, F^{\leftarrow}$	9 - 16 -9 actions	4/16 25%	5.1	7.1	-2.0
T4) One sentence	↓↑One Two [...] Thirty.	$\mu T_{mid}\, F^{\downarrow}$	(starting from rotor's 4th position) $R_{ccw} R_{ccw} R_{ccw} R_{ccw}\, F^{\downarrow}\, F^{\rightarrow}$	3 - 12 -9 actions	5/12 42%	1.6	5.3	-3.8
T5a) 2-half-sentences	↓↑three Twenty four [...] Thirty. Thirty one Thirty two	$\mu T_{mid}\, F^{\downarrow}$ $F^{\rightarrow}_{long}\, F^{\rightarrow}_{long}\, F^{\rightarrow}_{long}\, F^{\rightarrow}_{long}$	(starting from rotor's 4th position) $R_{ccw} R_{ccw} R_{ccw} R_{ccw}$ $F^{\downarrow}\, F^{\downarrow}\, F^{\rightarrow}\, F^{\downarrow}\, F^{\rightarrow}\, F^{\rightarrow}\, F^{\rightarrow}\, F^{\rightarrow}$	11 - 24 -13 actions	7/24 29%	6.0	10.6	-4.7
T5b) 2-half-sentences	↓↑eight Twenty nine Thirty. Thirty one [...] Thirty seven	$\mu T_{mid}\, F^{\downarrow}\, (F^{\rightarrow}_{long} \times 13)$	(starting from rotor's 4th position) $R_{ccw} R_{ccw} R_{ccw} R_{ccw}$ $F^{\downarrow}\, F^{\rightarrow}\, F^{\downarrow}\, (F^{\rightarrow} \times 13)$	29 - 40 -11 actions	6/40 15%	15.9	17.7	-1.8

Figure 2: StEM modelling of optimal interaction trajectories using μGeT and VO for each task, and corresponding predicted time. Inputs use for navigating menus using VO are in orange.

of time. We modeled μT in sequence, whereas they could be performed in parallel. Second, the gesture to navigate the rotor menu using VO is different from the rotation gesture (R) described in the StEM. The rotation gesture in StEM is a precise input intended to control a parameter (e.g., the rotation angle of an object) whereas in our situation, it is a "ballistic rotation" that does not need to be finely controlled or targeted. We therefore considered this gesture as equivalent to a flick (F) gesture in terms of time. Third, μGeT involves two types of horizontal swipes: short and long. A linear movement with a length of less than 150 pixels is considered as a short swipe. As neither StEM nor FLM can discern between the two, we counted the time of long flicks as twice that of regular flicks.

To sum up, we used the following operator times: **669ms for T** (since no precise screen location is required, we used the maximum amplitude to derive the operator time); **669ms for μT**; **216ms for F and F_{short}**; **432ms for F_{long}**; and **216ms for R**.

As both VO and μGeT use sequences of ballistic flick and rotation gestures, which start and end their movement in the air, rotations (or flicks) will necessarily be modelled as TR (or TF). In order to simplify the reading of figure 2, we simply use R for TR and F for TF. However, we take both operator times into account.

As theoretical results suggest potential benefits, we proceeded with an experimental study.

4 EXPERIMENT DESIGN

The study investigates, for a task of text selection in eyes-free situation, whether μGeT is faster, requires less inputs, and ranks better compared to the VO baseline technique.

4.1 Participants

9 PVI and 8 sighted volunteers (6 participants self-identified as women: 4 PVI, 2 sighted) aged from 24 to 58 (Mean 33, SD 9.1) participated in this experiment. In the following sections, we refer to these participants as P1 to P9 for the PVI and P10 to P17 for the sighted participants (SP). All the participants were right-handed and blind-folded during the experiment. None of the participants reported any mental or perceptual impairments other than visual.

All of them use a touchscreen device daily; 8 PVI among 9 own an iOS device and have used VoiceOver before. None of the SP had ever used VoiceOver. No participant had experience with TTF µG.

4.2 Apparatus

The experiment was conducted on a 10.1" tablet (Huawei MediaPad T5 Lite, 24*15,5 cm), laid flat on a table in front of them. Input capabilities were provided through both the tablet (touch inputs) and custom-made rings worn on the right hand (µGesture inputs). The tablet ran a full-screen blank web-page which relayed all touch events, via websockets, to a laptop. Rings were made using an elastic fabric to accommodate various finger sizes and embed patched squares of conductive fabric and yarn. Each conductive surface played the role of a capacitive touch sensor connected to an Arduino board. They were 2cm×2cm large to cover the middle phalanx of the index and middle fingers, and the nail of the little finger. The Arduino board relayed press and release contacts on each sensor to the laptop via a wired serial connection. On the laptop, a Python written software ran the experiment, received and processed data from both the rings and the tablet, and displayed a window on the laptop allowing the experimenter to monitor the progress. We used an absolute mapping between the tablet screen and the text area on which tasks were performed. In our experiment, we implemented µGeT as well as a VoiceOver like technique, allowing more control over the experimental logs.

4.3 Experimental procedure

We manipulated participants' hands to show the gestures required in the experiment then we had them go through a tutorial. When they felt ready, they were instructed to select a section of a text as quickly and accurately as possible in a tablet sized view displaying three sentences of pseudo-text over a span of thirteen lines. To only evaluate the selection process and not the memorization of the text, we used numbers from one to sixty two, written in order in full letters (i.e., "One Two Three [...] Sixty Sixty one Sixty two."). Each word is counted separately (i.e., "twenty one" counts as two words). The resulting pseudo-text was divided into three sentences: "One" to "Thirty", "Thirty one" to "Fifty four", "Fifty five" to "Sixty two").

Each participant completed a block of five text selection tasks per technique. The order of the techniques was counter-balanced across participants. Each task involved a different target type. Between the first and second block, target types were the same but at different locations in the text, so no two tasks were identical. These target types were inspired from a previous study on text selection [17], and were meant to cover a variety of situations. Target types were: 1) **4-words**, selecting four consecutive words; 2) **sentence**, selecting one sentence; 3) **2-words-and-half**, selecting two words and the following three characters, or selecting the last three characters of a word and the following two words 4) **2-half-sentences**, selecting the last four words of a sentence then thirteen words, or selecting the last thirteen words of a sentence then the first four words of the next one; 5) **mid-word**, selecting 3 characters in the middle of a word of 5 or 6 characters. The goal behind this panel of tasks was not to represent real-world behaviors, but rather to assess the usability of both techniques in more or less demanding contexts. The target type order was always the same: from "easy" to "hard" (the same order in which the target types are listed above). We purposefully chose to gradually increase the difficulty to ease the learning process of our participants and keep the duration of a session under an hour. The target was read aloud by the experimenter at the beginning of each trial. Participants could ask the experimenter to repeat the target as many times as necessary during the trial. A trial ended when the participant told the experimenter s.he completed the selection. If the target was not selected or was partially selected, the trial was counted as an error. The experimenter started the next trial upon request of the participant. Participants could take a break whenever they wanted in between trials. Once a block has been completed, participants answered a raw Nasa-TLX (perceived workload) and a UMUX-LITE (general usability) questionnaires about the technique used in the block. Once the second block has been completed and the questionnaires answered, participants were asked which technique they preferred and why. The experimenter concluded the experiment by collecting open comments and answering any questions. On average, the experiment lasted one hour.

5 RESULTS

To compare VO and µGeT efficiency, we primarily focus on **selection time** which spans between the end of the text selection-handles placement phase and the end of the trial. Additionally, we also compare **move time** (which spans between the first touch or µGesture input recorded and the beginning of the selection phase), **number of input** actions per trial, **errors**, **cognitive load**, **user experience** and **user preference**.

Since sighted people (SP) completed 2 more tasks than people with visual impairment (PVI), we report their results separately. We report medians and Interquartile Ranges (IQR) for continuous data (i.e., selection time, move time, number of inputs), use Wilcoxon signed-rank tests as they are not normally distributed, and report the rank biserial correlation (r) effect size. For non-continuous data (i.e., cognitive load, user experience), we report means and Standard Deviations (SD). We use Mann-Whitney U tests and report the rank biserial correlation (r) effect size. We use a Two proportion z-tests for binary data (i.e., number of errors). We also studied in detail all the interaction trajectories. We report a summary of strategies we observed as well as potential explanation for the behaviors that stand out. We use Two proportion z-tests for this purpose.

In total, 9 PVI × 3 tasks = 27 trials, and 8 SP × 5 tasks = 40 trials were completed per interaction technique. We removed one trial from the SP results (P11, condition VO, task "sentence") as the participant wrongly thought to have completed the trial, but did not select anything.

5.1 Quantitative analysis

Figure 3 summaries the **selection time** and **move time** analysis.

5.1.1 PVI. For PVI, the median **selection time** is shorter for µGeT (11 seconds, IQR 22.5) than for VO (21.6s, IQR 22.7). A Wilcoxon test show a significant difference between the two techniques for PVI ($p = 0.014$, $r=0.534$). **Selection times** vary greatly between each task type (figure 5). For PVI, a Wilcoxon test shows significant differences in selection time between the two techniques for the "four-words" ($p=0.027$, $r=0.822$) and "sentence" ($p=0.012$, $r=0.911$)

Data	Vision status	IT	Median (s)	IQR	Test	Effect size
Selection time	PVI	µGeT	11.0	22.5	Wilcoxon signed-rank p = 0.014*	r = 0.534
		VO	21.6	22.7		
	Sighted people	µGeT	21.7	22.3	Wilcoxon signed-rank p = 0.548	r = 0.113
		VO	20.5	15.2		
Move time	PVI	µGeT	29.2	34.6	Wilcoxon signed-rank p = 0.162	r = 0.312
		VO	42.0	58.7		
	Sighted people	µGeT	21.0	18.5	Wilcoxon signed-rank p = 0.239	r = 0.265
		VO	31.9	20.0		
# inputs per trial	PVI	µGeT	8	6.5	Wilcoxon signed-rank p = 0.012*	r = 0.555
		VO	14	10		
	Sighted people	µGeT	11	12	Wilcoxon signed-rank p = 0.24	r = 0.265
		VO	12	4.8		
# errors	PVI	µGeT	0 (total)	/	Two proportion z-test p = 0.004*	/
		VO	7 (total)	/		
	Sighted people	µGeT	1 (total)	/	Two proportion z-test p = 0.545	/
		VO	2 (total)	/		

Figure 3: Summary of the results regarding Selection time, Move time, Number of inputs and Number of errors, for both populations (PVI and SP), by interaction technique.

Data	Vision status	IT	Median	IQR	Test	Effect size
Mental demand	PVI	µGeT	3	2	Mann-Whitney U p= 0.653	r=0.116
		VO	3	3		
	Sighted people	µGeT	6	1.5	Mann-Whitney U p= 0.825	r=0.078
		VO	6	1.25		
Physical demand	PVI	µGeT	1	1	Mann-Whitney U p= 0.083	r=0.457
		VO	2	0		
	Sighted people	µGeT	2	2.5	Mann-Whitney U p= 0.870	r=0.062
		VO	2	2.25		
Temporal demand	PVI	µGeT	2	1	Mann-Whitney U p= 0.778	r=0.086
		VO	2	2		
	Sighted people	µGeT	3.5	1.5	Mann-Whitney U p= 0.703	r=0.125
		VO	3.5	1.25		
Performance	PVI	µGeT	5	2	Mann-Whitney U p= 0.745	r=0.099
		VO	5	3		
	Sighted people	µGeT	6	1.25	Mann-Whitney U p= 0.867	r=0.062
		VO	6	0.5		
Effort	PVI	µGeT	2	2	Mann-Whitney U = 0.855	r=0.062
		VO	2	1		
	Sighted	µGeT	5	2	Mann-Whitney U p= 1	r=0.016
		VO	5	1.25		
Frustration	PVI	µGeT	1	2	Mann-Whitney U p= 0.660	r=0.124
		VO	1	1		
	Sighted	µGeT	4	3	Mann-Whitney U p= 0.704	r=0.126
		VO	4	3.25		
Usefulness	PVI	µGeT	7	1	Mann-Whitney U p= 0.164	r=0.370
		VO	6	2		
	Sighted	µGeT	5.5	1	Mann-Whitney U p= 0.234	r=0.344
		VO	5	0.5		
Ease of use	PVI	µGeT	6	2	Mann-Whitney U p= 0.238	r=0.321
		VO	7	1		
	Sighted	µGeT	6	2	Mann-Whitney U p= 0.357	r=0.281
		VO	5.5	1.25		
Preferred technique	PVI	µGeT / VO	5 (56%) / 4 (44%) (N=9)		/	/
	Sighted	µGeT / VO	4 (50%) / 4 (50%) (N=8)		/	/

Figure 4: Summary of the results regarding the two questionnaires (raw Nasa-TLX and UMUX-Lite), for both populations (PVI and SP), by interaction technique.

tasks in favor of µGeT (figure 5). The median **move time** is shorter for µGeT than for VO (29.2s µGeT vs. 42s VO) (21s µGeT vs. 31.9s VO for SP). This observation holds when broken down for each task type (figure 5). However, Wilcoxon tests show no significant difference. The median **number of inputs** is lower for µGeT than for VO (8 µGeT vs. 14 VO) (figure 3). A Wilcoxon test shows a significant difference between both techniques (p=0.012, r=0.555) (figure 3). When broken down by task type, tests show a significant difference for "four-words" (p=0.024, r=0.866) and "sentence" (p=0.028, r=0.844) tasks in favor of µGeT. Regarding the **number of errors**, only 7 out of the 54 PVI trials resulted in text selections that did not match the target, all of them with VO. A Two proportion z-test shows a significant difference (p=0.004) between the two techniques. We measured the perceived **cognitive load** for each technique using a raw Nasa-TLX questionnaire and the **user experience** using a UMUX-Lite questionnaire. Figure 4 shows the mean subjective scores per technique and SD for each of the 6 items of the Nasa-TLX questionnaire (i.e., mental demand, physical demand, temporal demand, overall performance, global effort, frustration level) and the 2 items of the UMUX-Lite questionnaire (i.e., usefulness and usability). For each item of the two questionnaires, Mann-Whitney U tests show no significant difference between both techniques. Finally, **user preference** did not reveal any difference, both techniques were equally liked (figure 4).

5.1.2 Sighted. For SP, the median **selection time** is shorter for VO (20s, IQR 15.1) than for µGeT (21.7s, IQR 22.3) A Wilcoxon test does not show a significant difference between the two techniques for SP (p = 0.548). A Wilcoxon test shows significant differences in **selection time** between the two techniques for the "mid-word" tasks (p=0.008, r= 1), in favor of VO (figure 5). The median **move time** is shorter for µGeT than for VO (21s µGeT vs. 31.9s VO). This observation holds when broken down for each task type (figure 5). However, Wilcoxon tests show no significant difference. The median **number of inputs** is lower for µGeT than for VO (11 µGeT vs. 12 VO) (figure 3). A Wilcoxon test shows no significant difference between both techniques (p=0.582) (figure 3). When broken down by task type, tests show a significant difference for "mid-word" (p=0.022, r=1) task in favor of VO. Regarding the **number of errors**, 3 out of the 79 trials were erroneous: 2 with VO and 1 with µGeT. A two proportion z-test shows no significant difference (p=0.545) between the two techniques. We measured the perceived **cognitive load** for each technique using a raw Nasa-TLX questionnaire and the **user experience** using a UMUX-Lite questionnaire. Figure 4 shows the mean subjective scores per technique and SD for each of the 6 items of the Nasa-TLX questionnaire (i.e., mental demand, physical demand, temporal demand, overall performance, global effort, frustration level) and the 2 items of the UMUX-Lite questionnaire (i.e., usefulness and usability). For each item of the two questionnaires, Mann-Whitney U tests show no significant difference between both techniques. Finally, **user preference** did not reveal any difference, both techniques were equally liked (figure 4).

5.2 Analysis of the interaction trajectories

To better understand the differences in performance between the two techniques, we used Thematic analysis to derive types of behavior (patterns) in all the interaction trajectories that made participants deviate from the optimal trajectories. We recall that an optimal trajectory is using the least number of inputs for a given starting point, as defined in Section 3.3. One author did the parsing and analysis. There are 3 main types: **disorientation**, **inertia** and **mistake**. Figure 6 summarizes the classification of all trials.

Disorientation groups behaviors that delay the selection task because users take extra steps to understand the state of the system. Observed behaviors are:

- Users triggering time consuming audio feedback to read the current text selection.
- Users pausing or hesitating (from 3 to 15 seconds) between inputs.
- Users repeatedly moving a selection handle back and forth to understand the text locally.
- Users trying to cross over selection handles, which by design is not possible for both techniques. Since there is no

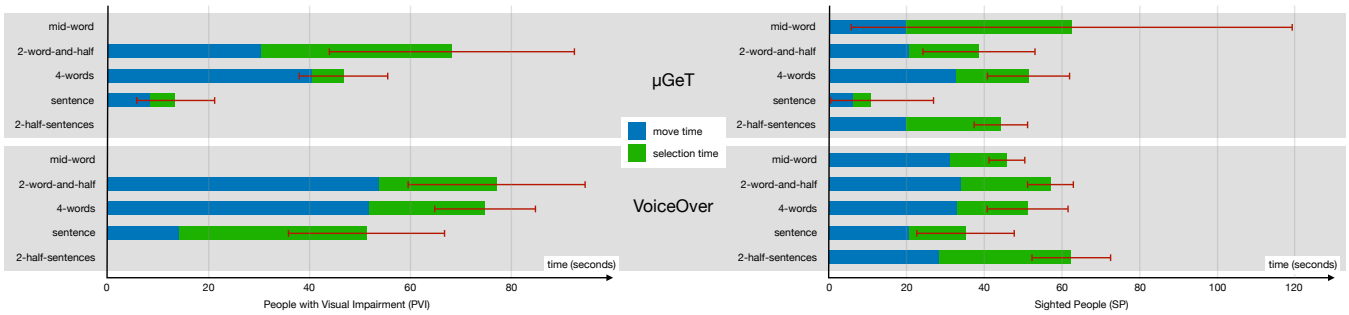


Figure 5: Median move and selection times with IQR, organized by task type and technique (PVI on the left, SP on the right).

Data	Vision status	IT	Count	% of occurrence	Test
Optimal trajectories	PVI	µGeT	12	44%	Two proportion z-test p= 0.048
		VO	5	19%	
	Sighted	µGeT	10	25%	Two proportion z-test p= 0.080
		VO	4	10%	
Disorientation: Supplementary audio feedback	PVI	µGeT	3	11%	Two proportion z-test p=0.329
		VO	1	4%	
	Sighted	µGeT	19	48%	Two proportion z-test p= 0.003
		VO	3	8%	
Disorientation: Pauses (> 3s)	PVI	µGeT	16	59%	Two proportion z-test p= 0.398
		VO	19	70%	
	Sighted	µGeT	34	85%	Two proportion z-test p= 0.020
		VO	24	62%	
Disorientation: Back & forth in text	PVI	µGeT	6	22%	Two proportion z-test p= 0.049
		VO	1	4%	
	Sighted	µGeT	10	25%	Two proportion z-test p= 0.013
		VO	2	5%	
Disorientation: Trying to cross handles over	PVI	µGeT	0	0%	/
		VO	/		
	Sighted	µGeT	4	10%	/
		VO	/		
Inertia: Navigating VO menus in the longest direction	PVI	µGeT	/		/
		VO	14	52%	
	Sighted	µGeT	/		/
		VO	17	44%	
Inertia: Performing a valid sequence of selection inputs in a sub-optimal mode	PVI	µGeT	1	4%	Two proportion z-test p= 0.049
		VO	6	22%	
	Sighted	µGeT	2	5%	Two proportion z-test p= 0.398
		VO	4	10%	
Mistake: Spoiling the current selection	PVI	µGeT	4	15%	Two proportion z-test p= 0.036
		VO	0	0%	
	Sighted	µGeT	8	20%	Two proportion z-test p= 0.003
		VO	0	0%	
Mistake: Confuse interactions	PVI	µGeT	9 (small. gran.) / 2 (menu/sub)	33% / 7%	Two proportion z-test p= 0.121 (small. gra
		VO	4 (smaller granularity only)	15%	
	Sighted	µGeT	15 (smaller granularity only)	38%	Two proportion z-test p= 0.048
		VO	7 (smaller granularity only)	18%	
Mistake: Overshoot selection or item menu	PVI	µGeT	3 (text only)	11%	Two proportion z-test p= 0.276
		VO	6 (menu only)	23%	
	Sighted	µGet	12 (text only)	30%	Two proportion z-test p= 0.001 (text)
		VO	1 (text) / 3 (menu)	3% / 8%	

Figure 6: Summary of the results regarding the interaction trajectories, for both populations (PVI and SP), by interaction technique.

audio feedback for this specific case, users could spend some time trying to actively understand what is happening, before moving on.

Inertia groups behaviors that delay the selection task because users adopt a sub-optimal strategy. Observed behaviors are:

- Users navigating VO menus, which wrap-around, in the "wrong" direction (e.g., going from 1st to 8th rotor item clockwise, instead of counter-clockwise).
- Users performing a valid sequence of selection inputs in a sub-optimal mode (e.g., selecting a sentence word-by-word instead of using the sentence selection mode).

Mistake groups behaviors that largely hinder the selection task and impose extra correction steps to achieve the task. Observed behaviors are:

- Users spoiling their current selection by (un)selecting a large portion of text, forcing to redo finer granularity selections.
- Users confusing interactions (e.g., performing one TTF instead of the other, moving the rotor menu instead of the sub-menu).
- Users overshooting their selection or item menu.

These errors are not mutually exclusive: several of them can be made during a trial or even at the same time.

6 DISCUSSION

In the following, we discuss the results for people with visual impairment (PVI) and sighted people (SP) separately.

PVI - overall performances. Overall, PVI performed better with µGeT than with VO (figure 3). When looking at interaction trajectories, 12 out of 27 (44%) are optimal with µGeT, and only 5 out of 27 (19%) with VO (significant difference, p=0.048). With µGeT, sub-optimal trajectories are mostly due to mistakes (use of a smaller granularity – 9/27, 33%) and disorientation (back and forth - 6/27, 22%). With VO, they are mostly due to mistakes (menu overshot – 6/27, 22%), and inertia (use of sub-optimal levels of granularity – 6/27, 22%). Pauses are very common in both techniques (µGeT: 16/27, 59%; VO: 19/27, 70%). Our results also show that PVI performed the text selections with a sub-optimal granularity 22% of the time with VO (compared to only 4% with µGeT) (p=0.045), which exponentially lengthened the selection time (e.g., P2 took 19.5s to select a sentence word by word). While in proportion, PVI seem to make less sub-optimal actions with VO than with µGeT (12 with VO, 26 with µGeT), they still are faster with µGeT. For PVI, it could suggest that: 1) sub-optimal actions are less costly in µGeT than in VO; and/or 2) sub-optimal behaviors are easier to correct or have less inertia with µGeT than with VO.

SP - overall performances. Overall, SP performed similarly with both techniques (figure 3). When looking at interaction trajectories, 10 out of 40 (25%) are optimal with µGeT, and only 4 out of 39 (10%) with VO (no significant difference, p=0.08). With µGeT, sub-optimal trajectories are mostly due to disorientation (supplementary audio feedback – 19/40, 48%, back and forth – 10/40, 25%), and due to mistakes (use of a smaller granularity – 15/40, 38%, selection overshot – 12/40, 20%). With VO, they are mostly due to mistakes (use of a smaller granularity – 7/39 trials, 18%). No other behaviors happened more than 10% of the time with VO. Pauses are very common in both techniques, but higher with µGeT (µGeT: 34/40, 85%; VO: 24/39, 62%; significant difference p=0.02) While in proportion, SP seem

to make much less sub-optimal actions with VO than with μGeT (17 with VO, 66 with μGeT), they still perform equivalently with both techniques in terms of selection time. Probable explanations exposed for PVI could therefore hold as well for SP.

Task performances. PVI performed better with μGeT in the "4-words" and "sentence" tasks, and similarly in the "2-words-half" tasks, while SP performed similarly with both techniques for all tasks, except the "mid-word" tasks, where they perform better with VO. Except for the "sentence" task, our observations differs from expected theoretical predictions. StEM models only account for motor actions, meaning that it models "expert" users, which by nature is not the case in our experiment. However, we could have expected similar ranking between both techniques. Therefore, differences we see could reveal the difference in cognitive work that participants have to go through to complete their tasks. For instance, tasks "4-words" and "sentence" do not require changing levels of granularity and approach ecological tasks. In both cases, we could hypothesize cognitive work to be low hence matching the motor predictions. For the other tasks, the mental demand could have eroded away the small expected motor differences.

Paradigm differences. If we look at pauses and number of inputs, results reveal a major difference between both techniques. VO yields more inputs but seem to require less thinking time, as opposed to μGeT. We hypothesize that the VO menu acts as a guide and provides an explicit and structured way to interact with the system: menu items can be only be navigated linearly, the choices of what to do next is reduced to 4 options (i.e., previous item, next item, (de)select next entity), and audio feedback explicitly states the new mode. On the contrary, μGeT relies on memory, like keyboard shortcuts: commands are accessible at all times but users need to to know the mapping between actions and commands This paradigm difference may become exacerbated as the complexity of the task increases.

For VO, we detected inertia behavior in menus (i.e., cycling through the "wrong" way): 14/27 trials (52%) for PVI, 17/40 trials (44%) for SP (no significant difference, p=0.52). While it increased the number of inputs it did not substantially increase selection time. They are likely due to a lack of familiarity with the technique, thus resulting in a random choice of direction for both populations.

For PVI, time per input is significantly shorter (Wilcoxon test p=0.020, r=0.867) for the "easiest" tasks ("4-words", 0.95s IQR 0.79) than for the "hardest" tasks ("2-words-half", 2.13s IQR 0.62) with μGeT, which is not the case with VO where all three task types have similar Time per input.

PVI and SP differences. With μGeT, SP asked significantly more supplementary audio feedback than PVI (19/40 trials, 48% for SP compared to 3/27 trials, 11% for PVI, p=0.0016). SP also paused significantly more with μGeT than with VO. SP also overshot more with μGeT (12/40 trials, 30% for μGeT compared to 1/40, 3% for VO, p=0.001). SP had more sub-optimal strategies in μGeT compared to PVI (0.84 sub-optimal actions per trial on average to 0.44 for PVI) but less in VO (0.44 per trial to 0.96 for PVI). Related to the previous point, this could show that SP are simply less familiar with eyes-free interaction and therefore rely more on structured audio feedback such as that provided by VO. In μGet, since such structured feedback is not provided, they have to pause more often and rely on their mental model of the system.

7 CONCLUSION

Touch modality remains the primary means of interaction for tablets and smartphones but still presents usability challenges for people with visual impairment (PVI). This paper focuses on the use of microgestures (μG) as a secondary input modality to enhance touchscreen interaction bandwidth in eyes-free situations. We proposed μGeT, a multimodal eyes-free text selection technique, that combines touch interaction with thumb-to-finger μG. We theoretically and experimentally compared μGeT to the iOS VoiceOver (VO) accessibility tool. Our results show that μGeT outperforms VO in terms of selection speed for PVI. Interestingly, selection time is equivalent with both techniques for sighted people. With μGet, sighted people performed more actions, required more audio feedback and paused more than PVI to complete the task. We hypothesize that this is due to the difference in familiarity with audio-based interaction without the visual channel. Moreover, our results show that μGeT outperforms VO in terms of number of inputs of the optimal interaction trajectories. However, participants often deviate from these optimal trajectories. A detailed analysis of the trajectories showed that μGeT is more likely to produce sub-optimal actions that lengthen the interaction trajectory, causing errors and confusion. However, these sub-optimal actions individually have a small impact on the selection time because they can be quickly corrected. On the contrary, VO is less likely to produce sub-optimal actions, but the errors are more difficult to correct, resulting in more failures. We put forward the following explanation: VO, similar to graphical menus, provides a structural guide that reduces the need to remember the commands but that lengthens the interaction. μGeT is similar to keyboard shortcuts: users must know the commands beforehand and remember them to use them efficiently. But all commands are directly accessible at all times, which encourages users to change their course of action. We expect that with more expertise, both populations would get closer to the modeled optimal interaction trajectories. We therefore plan to further investigate this effect of expertise with our technique μGeT. We also plan to test other tasks suggested by our participants, such as common copying and pasting tasks, and tasks in the context of games.

8 ACKNOWLEDGMENTS

This work has been partially supported by the French National Research Agency (ANR) project MIC (ANR-22-CE33-0017) and by the LabEx PERSYVAL-Lab (ANR-11-LABX-0025-01). We also thank our partners: Laboratory Cherchons pour Voir, CNRS, CESDV-IJA and UNADEV Toulouse.

REFERENCES

[1] Carl Halladay Abraham, Bert Boadi-Kusi, Enyam Komla Amewuho Morny, and Prince Agyekum. 2021. Smartphone usage among people living with severe visual impairment and blindness. *Assistive Technology* 0, 0 (2021), 1–8. https://doi.org/10.1080/10400435.2021.1907485

[2] Peter Ackland, Serge Resnikoff, and Rupert Bourne. 2017. World blindness and visual impairment: despite many successes, the problem is growing. *Community eye health* 30, 100 (2017), 71–73.

[3] Nancy Alajarmeh. 2021. The extent of mobile accessibility coverage in WCAG 2.1: sufficiency of success criteria and appropriateness of relevant conformance levels pertaining to accessibility problems encountered by users who are visually impaired. *Universal Access in the Information Society* (2021). https://doi.org/10.1007/s10209-020-00785-w

[4] Grace M. Begany, Ning Sa, and Xiaojun Jenny Yuan. 2016. Factors Affecting User Perception of a Spoken Language vs. Textual Search Interface: A Content Analysis. *Interact. Comput.* 28 (2016), 170–180.

[5] Hemant Bhaskar Surale, Fabrice Matulic, and Daniel Vogel. 2017. Experimental Analysis of Mode Switching Techniques in Touch-based User Interfaces multi-touch; touch input; mode switching. (2017). https://doi.org/10.1145/3025453.3025865

[6] Roger Boldu, Denys J.C. Matthies, Haimo Zhang, and Suranga Nanayakkara. 2020. AiSee: An Assistive Wearable Device to Support Visually Impaired Grocery Shoppers. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 4, Article 119 (dec 2020), 25 pages. https://doi.org/10.1145/3432196

[7] Emeline Brule, Gilles Bailly, Anke Brock, Frédéric Valentin, Grégoire Denis, and Christophe Jouffrais. 2016. MapSense: Multi-sensory interactive maps for children living with visual impairments. In *Conference on Human Factors in Computing Systems - Proceedings*. Association for Computing Machinery, 445–457. https://doi.org/10.1145/2858036.2858375

[8] Stuart K. Card, Thomas P. Moran, and Allen Newell. 1980. The Keystroke-Level Model for User Performance Time with Interactive Systems. *Commun. ACM* 23, 7 (jul 1980), 396–410. https://doi.org/10.1145/358886.358895

[9] Edwin Chan, Teddy Seyed, Wolfgang Stuerzlinger, Xing-Dong Dong Yang, and Frank Maurer. 2016. User Elicitation on Single-Hand Microgestures. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 3403–3414. https://doi.org/10.1145/2858036.2858589

[10] Eric Corbett and Astrid Weber. 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* (Florence, Italy) *(MobileHCI '16)*. Association for Computing Machinery, New York, NY, USA, 72–82. https://doi.org/10.1145/2935334.2935386

[11] Rafael Jeferson Pezzuto Damaceno, Juliana Cristina Braga, and Jesús Pascual Mena-Chalco. 2018. Mobile device accessibility for the visually impaired: problems mapping and recommendations. *Universal Access in the Information Society* 17, 2 (2018), 421–435. https://doi.org/10.1007/s10209-017-0540-1

[12] Julie Ducasse, Anke M. Brock, and Christophe Jouffrais. 2018. Accessible Interactive Maps for Visually Impaired Users. In *Mobility of Visually Impaired People*. Springer International Publishing, 537–584. https://doi.org/10.1007/978-3-319-54446-5_17

[13] Yasmine N. El-Glaly, Francis Quek, Tonya Smith-Jackson, and Gurjot Dhillon. 2013. Touch-Screens Are Not Tangible: Fusing Tangible Interaction with Touch Glass in Readers for the Blind. In *Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction* (Barcelona, Spain) *(TEI '13)*. Association for Computing Machinery, New York, NY, USA, 245–253. https://doi.org/10.1145/2460625.2460665

[14] Gauthier Robert Jean Faisandaz, Alix Goguey, Christophe Jouffrais, and Laurence Nigay. 2022. Keep in Touch: Combining Touch Interaction with Thumb-to-Finger μGestures for People with Visual Impairment. In *Proceedings of the 2022 International Conference on Multimodal Interaction* (Bengaluru, India) *(ICMI '22)*. Association for Computing Machinery, New York, NY, USA, 105–116. https://doi.org/10.1145/3536221.3556589

[15] Euan Freeman, Gareth Griffiths, and Stephen A. Brewster. 2017. Rhythmic microgestures: Discreet interaction on-The-go. *ICMI 2017 - Proceedings of the 19th ACM International Conference on Multimodal Interaction* 2017-Janua, September (nov 2017), 115–119. https://doi.org/10.1145/3136755.3136815

[16] Alix Goguey, Géry Casiez, Andy Cockburn, and Carl Gutwin. 2018. Storyboard-Based Empirical Modeling of Touch Interface Performance. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3174019

[17] Alix Goguey, Sylvain Malacria, and Carl Gutwin. 2018. Improving Discoverability and Expert Performance in Force-Sensitive Text Selection for Touch Devices with Mode Gauges. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (2018). https://doi.org/10.1145/3173574

[18] Alix Goguey, Daniel Vogel, Fanny Chevalier, Thomas Pietrzak, Nicolas Roussel, and Géry Casiez. 2017. Leveraging finger identification to integrate multi-touch command selection and parameter manipulation. *International Journal of Human-Computer Studies* 99 (mar 2017), 21–36. https://doi.org/10.1016/J.IJHCS.2016.11.002

[19] Nora Griffin-Shirley, Devender R Banda, Paul M Ajuwon, Jongpil Cheon, Jaehoon Lee, Hye Ran Park, and Sanpalei Nylla Lyngdoh. 2017. A Survey on the Use of Mobile Applications for People who Are Visually Impaired. *Journal of Visual Impairment & Blindness* 111 (2017), 307–323.

[20] LeeLik Hang, LamKit Yung, LiTong, BraudTristan, SuXiang, and HuiPan. 2019. Quadmetric Optimized Thumb-to-Finger Interaction for Force Assisted One-Handed Text Entry on Mobile Headsets. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (sep 2019), 1–27. https://doi.org/10.1145/3351252

[21] Chris Harrison and Scott E Hudson. 2012. Using Shear as a Supplemental Two-Dimensional Input Channel for Rich Touchscreen Interaction. (2012).

[22] Chris Harrison, Julia Schwarz, and Scott E Hudson. 2011. TapSense: Enhancing Finger Interaction on Touch Surfaces. *Proceedings of the 24th annual ACM symposium on User interface software and technology - UIST '11* (2011). https://doi.org/10.1145/2047196

[23] Juan David Hincapié-Ramos, Xiang Guo, Paymahn Moghadasian, and Pourang Irani. 2014. Consumed Endurance: A Metric to Quantify Arm Fatigue of Mid-Air Interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 1063–1072. https://doi.org/10.1145/2556288.2557130

[24] Shaun K. Kane, Meredith Ringel Morris, Annuska Z. Perkins, Daniel Wigdor, Richard E. Ladner, and Jacob O. Wobbrock. 2011. Access Overlays: Improving Non-Visual Access to Large Touch Screens for Blind Users. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology* (Santa Barbara, California, USA) *(UIST '11)*. Association for Computing Machinery, New York, NY, USA, 273–282. https://doi.org/10.1145/2047196.2047232

[25] Shaun K Kane, Meredith Ringel Morris, and Jacob O Wobbrock. 2013. Touchplates: Low-Cost Tactile Overlays for Visually Impaired Touch Screen Users. *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility* (2013). https://doi.org/10.1145/2513383

[26] Shaun K Kane and Jacob O Wobbrock. 2011. Usable Gestures for Blind People: Understanding Preference and Performance. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2011). https://doi.org/10.1145/1978942

[27] Akif Khan and Shah Khusro. 2021. An insight into smartphone-based assistive solutions for visually impaired and blind people: issues, challenges and opportunities. *Universal Access in the Information Society* 20, 2 (2021), 265–298. https://doi.org/10.1007/s10209-020-00733-8

[28] Ahreum Lee, Kiburm Song, Hokyoung Blake Ryu, Jieun Kim, and Gyuhyun Kwon. 2015. Fingerstroke time estimates for touchscreen-based mobile gaming interaction. *Human movement science* 44 (2015), 211–224.

[29] G. Julian Lepinski, Tovi Grossman, and George Fitzmaurice. 2010. The design and evaluation of multitouch marking menus. *Conference on Human Factors in Computing Systems - Proceedings* 4 (2010), 2233–2242. https://doi.org/10.1145/1753326.1753663

[30] Frank Chun Yat Li, David Dearman, and Khai N. Truong. 2009. Virtual Shelves: Interactions with Orientation Aware Devices. In *Proceedings of the 22nd Annual ACM Symposium on User Interface Software and Technology* (Victoria, BC, Canada) *(UIST '09)*. Association for Computing Machinery, New York, NY, USA, 125–128. https://doi.org/10.1145/1622176.1622200

[31] Guanhong Liu, Yizheng Gu, Yiwen Yin, Chun Yu, Yuntao Wang, Haipeng Mi, and Yuanchun Shi. 2020. Keep the Phone in Your Pocket: Enabling Smartphone Operation with an IMU Ring for Visually Impaired People. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 2 (jun 2020). https://doi.org/10.1145/3397308

[32] Denys J. C. Matthies, Bodo Urban, Katrin Wolf, and Albrecht Schmidt. 2019. Reflexive Interaction: Extending the Concept of Peripheral Interaction. In *Proceedings of the 31st Australian Conference on Human-Computer-Interaction* (Fremantle, WA, Australia) *(OZCHI'19)*. Association for Computing Machinery, New York, NY, USA, 266–278. https://doi.org/10.1145/3369457.3369478

[33] Takashi Miyaki and Jun Rekimoto. 2009. Graspzoom: Zooming and scrolling control model for single-handed mobile interaction. *MobileHCI09 - The 11th International Conference on Human-Computer Interaction with Mobile Devices and Services* (2009). https://doi.org/10.1145/1613858.1613872

[34] Aarthi Easwara Moorthy and Kim-Phuong L Vu. 2015. Privacy Concerns for Use of Voice Activated Personal Assistant in the Public Space. *International Journal of Human–Computer Interaction* 31, 4 (2015), 307–335. https://doi.org/10.1080/10447318.2014.986642

[35] Matt Rice, R. Daniel Jacobson, Reginald G. Golledge, and David Jones. 2005. Design Considerations for Haptic and Auditory Map Interfaces. *Cartography and Geographic Information Science* 32, 4 (2005), 381–391. https://doi.org/10.1559/152304005775194656 arXiv:https://doi.org/10.1559/152304005775194656

[36] João Ricardo dos S. Rosa and Natasha Malveira C. Valentim. 2020. Accessibility, Usability and User Experience Design for Visually Impaired People: A Systematic Mapping Study. In *Proceedings of the 19th Brazilian Symposium on Human Factors in Computing Systems* (Diamantina, Brazil) *(IHC '20)*. Association for Computing Machinery, New York, NY, USA, Article 5, 10 pages. https://doi.org/10.1145/3424953.3426626

[37] Marcos Serrano and Laurence Nigay. 2009. Temporal Aspects of CARE-Based Multimodal Fusion: From a Fusion Mechanism to Composition Components and WoZ Components. In *Proceedings of the 2009 International Conference on*

Multimodal Interfaces (Cambridge, Massachusetts, USA) *(ICMI-MLMI '09)*. Association for Computing Machinery, New York, NY, USA, 177–184. https://doi.org/10.1145/1647314.1647346

[38] Adwait Sharma, Joan Sol Roo, and Jürgen Steimle. 2019. Grasping Microgestures: Eliciting Single-Hand Microgestures for Handheld Objects. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19, Chi)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300632

[39] Craig Stewart, Michael Rohs, Sven Kratz, and Georg Essl. 2010. Characteristics of Pressure-Based Input for Mobile Devices. *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10* (2010). https://doi.org/10.1145/1753326

[40] Hsin Ruey Tsai, Te Yen Wu, Min Chieh Hsiu, Jui Chun Hsiao, Da Yuan Huang, Yi Ping Hung, Mike Y. Chen, and Bing Yu Chen. 2017. SegTouch: Enhancing touch input while providing touch gestures on screens using thumb-to-index-finger gestures. In *Conference on Human Factors in Computing Systems - Proceedings*, Vol. Part F1276. ACM, New York, NY, USA, 2164–2171. https://doi.org/10.1145/3027063.3053109

[41] Katrin Wolf. 2016. Microgestures—Enabling Gesture Input with Busy Hands. In *Peripheral Interaction: Challenges and Opportunities for HCI in the Periphery of Attention*, Saskia Bakker, Doris Hausen, and Ted Selker (Eds.). Springer International Publishing, Cham, 95–116. https://doi.org/10.1007/978-3-319-29523-7_5

[42] Katrin Wolf, Anja Naumann, Michael Rohs, and Jörg Müller. 2011. Taxonomy of Microinteractions: Defining Microgestures Based on Ergonomic and Scenario-Dependent Requirements. In *Proceedings of the 13th IFIP TC 13 International Conference on Human-Computer Interaction - Volume Part I (INTERACT'11, Vol. 6946 LNCS)*. Springer-Verlag, Berlin, Heidelberg, 559–575. https://doi.org/10.1007/978-3-642-23774-4_45

[43] Haijun Xia, Tovi Grossman, and George Fitzmaurice. 2015. Nanostylus: Enhancing input on ultra-small displays with a finger-mounted stylus. *UIST 2015 - Proceedings of the 28th Annual ACM Symposium on User Interface Software and Technology* (nov 2015), 447–456. https://doi.org/10.1145/2807442.2807500

[44] Mary Zajicek, Richard Wales, and Andrew Lee. 2004. Speech interaction for older adults. , 122–130 pages. https://doi.org/10.1007/s10209-004-0091-0

[45] Maozheng Zhao, Wenzhe Cui, IV Ramakrishnan, Shumin Zhai, and Xiaojun Bi. 2021. Voice and Touch Based Error-Tolerant Multimodal Text Editing and Correction for Smartphones. In *The 34th Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '21)*. Association for Computing Machinery, New York, NY, USA, 162–178. https://doi.org/10.1145/3472749.3474742