University of Glasgow | School of Computing Science

Menu selection for cyclists: comparing voice, touchscreen and button controls

Diego Drago (2384810d)

April 20, 2023

ABSTRACT

As countries aim to build a greener economy, bicycles are emerging as an alternative to cars. With the advent of AR glasses onto the market, many companies have started producing devices targeted to cyclists, which, however, lack informed or consistent user interface design. This paper aims to assess and compare suitable input methods for cycling AR Glasses. We evaluated voice input, touchscreen and button controls in a within-subjects field study (n=21). Participants were asked to complete menu navigation tasks with the input methods while cycling in a straight line. The study showed voice input to impact cycling speed the least and to perform best in subjective measures such as workload, perceived risk, and visual and manual demand. Touchscreen, on the other hand, performed best in terms of task accuracy and completion time. Our work provided insights into the design of interaction for cyclists by assessing the impact of three input methods on cycling performance, task performance, and several subjective measures.

1. INTRODUCTION

As Augmented Reality (AR) headsets become ubiquitous, and as technological improvements make AR glasses closer and closer to an everyday reality, we begin to wonder about their possible applications. Interest and research are growing on the vast and varied benefits AR glasses can offer to casual and professional cyclists. For example, novel safety features are paving the way for the use of AR to increase cyclist safety in traffic [42]. Museums are embracing the potentials of AR glasses to educate and inform visitors, and so is the case for walking or cycling-based open-air museums [26]. Thanks to smart glasses, navigational information can be provided in new and safer ways [9, 13, 8]. Finally, several studies have highlighted how Extended Reality (XR) can be used to improve athletic performance [29, 19, 39]. The USA Road and Track Cycling Teams, for instance, used smart glasses to train for the Rio 2016 Olympics and the Tokyo 2020 Olympics [32]. The teams won a bronze medal and two silver ones. However, since the technology is so new, an industry standard has yet to be set. There are several smart glasses targeted to cyclists available on the market currently, and they lack a common set of interaction modalities:

- Garmin Vision[1] is a display that can be attached to the leg of a pair of glasses. It provides a simple touchpad on the side of the device [2].
- The Solos[2] glasses provide all-purpose up and down mechanical buttons on the frames. These can change the page displayed and control media. Furthermore, this device also offers voice controls [2].
- The Everysight Raptor [43] provide voice commands, and a touch pad on the side arm of the glasses, where users can interact by tapping and swiping. Furthermore, the company sells a remote controller[3] for the glasses. This controller connects via Bluetooth to the glasses and can be attached to the bike handlebar, controlling them via mechanical buttons. Finally, the Raptor also supports the use of the smartphone app as a remote control [2].
- The ActiveLook technology[4] provides a mid-air gesture-based interaction modality, where users can lift one of their hands up to their face and swipe mid-air to change the page. Moreover, they allow users to change information interactively on the smartphone app [5]. A number of commercially available glasses use this technology, like the Cosmo Vision[5] glasses and the ENGO 1[6], which therefore function similarly [31].

Interacting via voice commands makes no significant difference on driving performance [12] and most new cars and phones provide them. However, such an input method is susceptible to a plethora of problems, from privacy issues [30], to security concerns [38, 28] to environmental noise [20]. As such, AR Glasses need to offer alternative methods for situations when speech controls cannot be used. Furthermore, several products provide mid-air gestures-based interaction, though in all of these cases users have to lift their hands from the handlebar, which is widely regarded as unsafe [39, 16]. Frames touchpads and buttons, where users need to lift their hands to the glasses, run into the same design issue. So far, no study assessed the above mentioned input methods, let alone compared them against one other. This paper assessed suitable input methods for AR Glasses designed for cyclists in terms of cycling performance, task performance and several subjective metrics. We first conducted an online survey to identify which input methods casual and professional cyclists used and investigated their perceived safety (Sec. 3). We then conducted a focus group to gain a deeper understanding of the survey's results as well as to identify the

[1] `https://www.garmin.com/en-GB/p/530536`

[2] `https://store.simplifaster.com/product/solos/solos-smart-glasses/solos-smart-glasses/`

[3] `https://pronetkrakow.com.pl/en/x-reality/29-everysight-raptor-controller.html`

[4] `https://www.activelook.net/pages/the-tech`

[5] `https://cosmoconnected.com/products/cosmo-vision`

[6] `https://engoeyewear.com/products/engo-1`

most desirable input methods for cycling AR glasses (Sec. 4). Finally, we carried out a within-subjects comparative field-study (n=21) where we analysed and compared the viability of voice input, button-based input, and touchscreen input for AR Glasses (Sec. 5). We report our findings for each step of this study and discuss where each input method performs best compared to the others.

2. RELATED WORK

Research on the use of AR glasses for cyclists has broadly focused on the output of novel features that the device could offer. Poppinga et al. [33] developed a navigational system that gives vibrotactile feedback to indicate directions, Von Sawitzky et al. [42] developed and evaluated safety features to reduce car crashes, Litvak and Kuflik [26] developed a museum guide app on the Everysight Raptor glasses.

2.1 Input methods for cyclists

No paper studied the input methods for cycling AR glasses. However, mid-cycling input has started being investigated, with a few studies comparing different methods. Hochleitner et al. [16] investigated the paradigms of mid-cycling interactions. In an enactment study, the authors explored possible interaction movements, and identified in the group discussion five factors to bear in mind while designing input methods for cyclists. These were: distraction, balance, (mis-)communication with other road users, terrain and weather conditions, and the Midas Touch problem - where a system struggles to differentiate between intentional gestures and ordinary movements [21]. Furthermore, participants were found to prefer keeping their hands on the handlebar. From these results the authors identified two viable options: wrist-flicking gestures and handlebar controls, which they compared in an outdoor study to the common one-handed smartphone touchscreen interaction. Participants interacted with a smartphone game, and results showed buttons on the handlebar to perform best in terms of task load index and game score. The paper focused on movement-based interaction, and did not take voice controls into consideration.

Dancu et al. [9] compared the mid-air gesture conventionally used to signal turns to a button-based controller mounted on the handlebar. The paper investigated the projection of navigational information and turn signals onto the road. In a first experiment the authors compared displaying a map on a head-up display to projecting it onto the road. The former scored better in terms of subjective workload and perceived safety. In a second experiment, the authors developed a gesture detection system that would project a turn signal onto the road whenever the user would physically signal with their arm. They compared this reality augmentation system with an off-the-shelf signal turning system, where users can interact with a handlebar mounted controller to turn on turn-signalling arrow-shaped LEDs mounted on the back of the bike. Results showed the gesture to perform significantly better than the buttons in terms of workload. Usability was not significantly different, and, in a follow-up survey, 62% of the participants considered the buttons to be safer.

Woźniak et al. [44] designed, implemented and evaluated two remote controllers attached to the handlebars. Allowing users to provide input without having to let go of the handles, the controls are either based on buttons or rotating grips. The authors first performed a 152 participants online survey to gather requirements. Most participants (61%) reported using their smartphone while cycling, and three actions in particular were rated as the most used: controlling music, answering calls and activating the voice assistant. The authors conducted a further experiment comparing the two controllers against the phone touchscreen. The former ones were mounted on the left hand-side of the handlebar, so as to not interfere with the derailleur controls, whereas the smartphone was held and controlled with the right hand. The traditional interaction was found to perform significantly better than the button-based controller in terms of task completion time. Both novel input methods scored better than the one-handed interaction in terms of subjective workload. The rotation-based controller scored better than both others in terms of bicycle tilt. The button-based controller was found to be significantly more usable than the smartphone touchscreen too. We believe future work should investigate a similar comparison but ignoring the need to make space for the derailleur controls. Installing the novel controllers on the left side, although a completely understandable design choice, might impact the task completion time results most and affects the rest of the results as well. With the introduction of electronic gear-shifting systems, we might see an overhaul of the controls on the handlebar. As such, it might be worth evaluating controllers on the right side.

Kosch et al. [23] evaluated and compared three selection modalities for cyclists: gaze dwell time, hand gestures and a multimodal approach that combines the user's gaze and a physical button (MAGIC Pointing). The authors experiment focused on target selection while cycling. In a bike simulator, users were asked to wear the HoloLens 2. While cycling, notifications would pop up in semi-randomised position containing the instruction for which virtual button to press. The authors measured task-completion time, error rate and task load. They found MAGIC Pointing to result in a smaller completion time than both other inputs and gestures to have a smaller completion time than the dwelling method. Furthermore, the authors found gestures to result in higher error rates than dwell time and there to be no significant difference in subjective workload.

Less recently, De Waard et al. [11] conducted three studies to investigate the prevalence and effects of mobile phone use on cycling behavior. In a first study, the authors found 2.2% of cyclists in Groningen to be talking on their phone and 0.6% to be entering text. In the second study, they examined the prevalence of phone use in cycling accidents. Finally, in the third and main study the authors investigated the effects of mobile phone use on cycling behaviour. They recorded cycling performance of 24 participants while either talking on the phone while performing simple or complex cognitive tasks, texting, listening to music or just cycling. We should note the talking tasks involved holding the phone to the ear with one hand, and phones with physical keypads were used for texting. The authors collected effort ratings through the Rating Scale Mental Effort (RSME) survey, and risk ratings through the same survey, with slight modifications. The study showed that text messaging resulted in the slowest cycling speed, most frequent and pronounced swerving, poorest object recognition, and highest risk ratings. Talking while performing difficult calculation tasks led to the lowest object recognition scores and highest effort ratings.

Subsequently, De Waard investigated the impact that operating a smartphone touch screen has on cycling [10]. Their study mirrored the one carried out in 2010, with the main objective to compare cycling while texting on a touch screen phone with texting on a physical keypad. Results showed that speed and effort ratings were not affected. Furthermore, object identification and self-reported risk were worse while operating a touchscreen phone than a physical keypad one.

Other research on mid-cycle interaction focuses on developing novel interaction techniques without comparing them to currently used methods. Both Caon et al. [7], and Tan et al. [40] conducted user elicitation studies on hand microgestures as input for cyclists, which allow users to keep their hands on the handlebar.

To the best of our knowledge no published study has evaluated voice commands in a cycling context. In a requirement study, Sörös et al. [39] investigated what features would possible consumers want from using AR glasses in a cycling context. According to their findings, a performance measurements display was the most requested feature, with all participants mentioning it. Other ones mentioned by most of the participants were: performance comparison (with other cyclists or past personal performances), navigational info, calls and other telecommunication features, and simple interaction modalities to avoid taking any or much of the biker's attention. From the requirements collected, the authors developed a voice-controlled Google Glass app. However, they did not evaluate their prototype.

Finally, recent observational studies reported on the uses and intricacies of the interaction methods that cyclists use. In a set of in-the-wild observations (n=414), Al-Taie et al. [3] found that cyclists used various techniques to communicate with other road users. For example, interactions such as arm gestures are used to negotiate right-of-way when lane merging.

Porcheron et al. [34] conducted an ethnography to examine how cyclists interact with mobile devices, providing valuable insights into the different stages involved in the interaction process. These steps include selecting the moments during which to use technology and grip transitions cyclists undergo before and after every interaction.

2.2 Input methods for in-vehicle infotainment systems

Research on in-car input methods can also be useful for our aim. Both drivers and cyclists have to split their attention between the road and the interface when they want to interact, and indeed many of the same solutions have been proposed for cars. Angelini et al. [4] compared speech commands, a touch-board simulating touchscreen, and on-steering-wheel gesture as input methods for in-vehicle infotainment systems. They based their evaluation on interaction performance (task completion time and interaction efficiency), driving performance (driving time, violations and accidents), usability, workload, and emotional response. During the between-subjects experiment, the drivers were verbally instructed which input to give. Touch performed the best in terms of task completion time and speech the best in interaction efficiency. Driving performance, usability, workload and emotive response were not significantly affected by input modality.

Depending on the context, some situations can put additional demands on the driver's visual, auditory, or manual resources. Roider et al. [35] compared different input modalities for in-vehicle interaction. The authors analyzed the specific impacts that situational demands have on gaze, gesture and speech input in terms of driving performance, interaction efficiency, perceived suitability and cognitive workload. During the experiment a head-up display would instruct the user on what task to perform while driving. Each participant would drive four times, with three runs using different input methods and a fourth serving as a baseline. It was found that impairments were greatest when the situational demand addressed the same sensory channel as the used input modality.

3. ONLINE SURVEY

Previous research has found that the main reasons users interact with mobile devices while cycling are controlling music, answering calls and activating the voice assistant [44]. Other studies assessed mobile phone and mobile device usage by cyclists in the countries of the authors [1, 14, 17], which did not always provide unanimous results [45]. However, to the best of our knowledge, no study has investigated what input methods cyclists currently use. We could not exclude the possibility that a significant number of cyclists already use an input method they considered viable, safe and sufficient, independently of the mobile device.

Therefore, to gain insight into current interaction methods and user preferences, we conducted a brief survey. We investigated participants' cycling habits and their simultaneous use of mobile devices. Specifically, we gathered information on the frequency and purpose of cycling, as well as the types of input methods participants use to interact with their mobile devices while cycling. Finally, we inquired about participants' perceptions of the safety of these interaction methods and asked them to explain their responses.

3.1 Participants

We recruited a total of 81 participants from a variety of sources, including social media cycling group posts and university students and staff. We excluded participants who answered "Never" to the question "How often do you cycle?" and ended up with a final sample size of 61 participants (36 female, 22 male, 3 other) aged between 18 and 58 years (M: 28.36, SD: 8.41). Of these, 36% reported cycling 2-3 times a week, 19% less than once a week, 19% 4-6 times a week, 14% every day, and 9% once a week. Furthermore, 59% declared to cycle for recreation, 56% for exercise or sport, and 72% as a means of transportation. Participation in the study was voluntary.

3.2 Results

Most of the participants (75.41%) responded they used at least a mobile device while cycling. The input methods these participants were found to use included: touchscreen interaction (76.09%), mechanical buttons (34.78%), voice commands (4.32%), and AR Glasses touchpad (2.17%). Of the participants that used touchscreen input methods while cycling, 60% perceived it as an unsafe input method, 28.57% perceived it as a safe input method, and 11.43% were unsure. Of the participants that used mechanical button input methods while cycling, 31.25% perceived it as an unsafe input method, 56.25% perceived it as a safe input method, and 13% were unsure. During the study, the participants frequently cited reasons for the lack of safety when using AR glasses while cycling. Eleven participants reported that they needed to lift their hands off the handlebar to interact with their device, while 13 participants noted that the input method(s) they had selected was not eyes-free.

4. FOCUS GROUPS

In view of the results of the online survey, we conducted a focus group study to gain a deeper understanding of the cyclists' opinions. We sought to find out their views on the input methods mentioned in the survey, their views on those provided by smart glasses for cyclists, as well as their views on novel input methods they might come up with. Through this study, we aimed to identify a set of input modalities for cycling AR glasses considered safest and most viable by our participants.

4.1 Methodology & participants

We followed Lazar [25] and Krueger's [24] design for single-category semi-structured mini-focus groups. We decided to perform mini-focus groups since they provide easier logistics and more in-depth discussions. We recruited participants via social media posts and conducted a small screening survey to assess their eligibility based on the following criteria:

1. Having a background in engineering or computing science, defined as having at least a bachelor's degree in the field or five or more years of work experience in these areas.
2. Being an active cyclist, defined as cycling for six or more hours per week or participating in competitive cycling races.

Participants with an engineering or computing science background are more likely to have a good understanding of the technical aspects involved in creating such a product, while active cyclists are more likely to have an understanding of the specific needs and challenges faced by cyclists on the road. We believed that by recruiting participants with these qualifications, the study was more likely to produce relevant and useful results. We recruited 8 total participants (2 female, 6 male, 0 other), aged from 22 to 29 (M: 24.25, SD: 2.11). Each group consisted of 4 participants (1 female in each) and lasted an hour and a half. After the second focus group, we observed that the discussion appeared to have reached saturation and therefore concluded that it was appropriate to terminate the experiment. Participation in the study was voluntary.

We recorded and transcribed audio from each focus group, and then coded the anonymised data using thematic analysis. We chose the following discussion points: everyday input methods mentioned in the survey, input methods integrated into commercially available cycling AR glasses, possible novel input methods, and participant's preferences about these methods. Throughout the discussion, a bicycle was available to allow participants to visually enact and explain their points.

4.2 Results

4.2.1 Input methods emerged from the online survey

First, participants discussed the input methods that emerged from the survey. A unanimous response in favour of voice commands highlighted its favourable perception as the safest and most versatile input method in comparison to all others. The main reasons for it receiving a favourable response laid in the eyes-free and hands-free nature of the interaction.

> *I think voice is not that bad either, [...] and then with the voice you can do more things, so you can say: "Play this song". [...] For deeper input, you would use voice as the safest one.*

The only drawback mentioned was the susceptibility to ambient noise such as traffic or wind noise, as well as a slight concern over privacy and social acceptability:

> *I've tried calling on phones right, so yeah it's safe but it doesn't really work since it's very windy or noisy. I can imagine it's not very reliable when it's windy.*

We found participants to consider touchscreen smartphone interactions, smartwatch interactions and headphone interactions to be unsafe. The most cited reasons included the loss of balance caused by interacting with one hand. Smartwatches had the additional detriment of requiring users to bring one's hand to the opposite wrist. Smartphones were considered the least safe, due to the need to pull them out of pockets.

> *Just because a lot of cycling is balance, and that's a big part of the phone. If I get my head down when I'm looking at my phone, my balance gets quite off.*

Bike computers' touchscreens and buttons were, on the other hand, another interaction method that was seen as safe. Although they require users to lift one of their hands, participants mentioned that a quick input given in the proximity of the handlebar was completely acceptable. Furthermore participants mentioned that touchscreen and buttons input methods both benefit from how familiar they feel to most users.

> *P1: The cycling computer sticks out of the middle of the handlebar, so while you're cycling you can view the road while also looking at the device itself.*

> *P2: I guess it's also a balancing issue, like* [when] *you move your hand over for the smartwatch. It feels much more in control with the cycle computer.*

4.2.2 Input methods integrated into commercially available cycling AR glasses

Hand gestures as implemented by currently available products were ill-considered because of the need to lift one's hand from the handlebar. Additionally, cyclists can communicate with other road users via arms gestures, as shown by Al-Taie [3], and participants expressed concerns about accidentally sending wrong signals. However, some users suggested using micro-gestures or finger gestures as an alternative way to provide gestural input. This novel iteration of the input method was highly considered and often mentioned, mainly envisaged as having similar benefits to handlebar-embedded buttons.

Frames temple touch-pads were not highly considered, mainly due to the inherent need to lift a hand from the handlebar and the lack of accuracy they had personally experienced with such an input method.

4.2.3 Novel input methods

Furthermore, we asked the participants to come up with as many input methods suited for the cycling context as they could. Though ours was not a user elicitation study, results closely paralleled those found in previous studies [16], and could be mainly categorised by the zone of interaction. Users discussed smart gloves, elbow gestures, head gestures, lip reading input, blinking-based input, and foot gestures. Buttons embedded in the handlebar were the most mentioned and the most credited. Their benefits include immunity to the Midas-touch problem (explained in Sec.2.1), being largely eyes-free and only requiring minimal finger movement. A drawback was the possible lack of space on the handlebar.

> *If there [are] any obstacles it can be dangerous if you're not holding your handlebar with both hands. I think buttons on the handlebar would work better, as there is enough space for them, so you can easily reach them with your fingers.*

However, all other inputs were considered to face either the Midas-touch problem, social acceptability issues or weather-related problems in the case of glove-based input.

4.3 Discussion

The focus group aimed to identify safe and viable interaction modalities for AR glasses. Participants discussed problems at the core of mid-cycle interaction design, such as distraction, balance, misinterpretation of gestures by other road users, and varying terrain and weather conditions [16]. Participants identified five input modalities: voice input, handlebar-buttons, bike-computer touchscreen and buttons, and micro-hand gestures. These methods were considered safe and viable, as they were deemed to have minimal impact on the aforementioned factors.

5. COMPARATIVE STUDY

Based on the insights gained from the focus group, we identified a set of input methods that were deemed most trusted by participants. To further evaluate the effectiveness and feasibility of these methods, we conducted a user study that involved a field experiment. Through this study, we aimed to quantitatively compare the identified input methods and gain more generalisable and externally valid results. Therefore, we developed the following research questions:

RQ1 *How do the selected input methods impact task performance?*

RQ2 *How do the selected input methods impact cycling performance?*

RQ3 *How do the selected input methods perform in terms of perceived safety, privacy-preservation, social acceptability, usability and task effort?*

To simplify the design, we combined two input methods into one as they both provided mechanical button-based interaction, with only a significant difference in the location of the buttons. As a result, we chose to eliminate the bike computer button input method. With this decision made, we proceeded to test the remaining ones. Among them, the gesture-based input involved using micro-hand gestures while holding the handlebar to interact with the AR interface. However, we found this method to be impractical due to the limitations of the technology used to detect hand movements. We integrated Ultraleap Gemini hand-tracking software[7] both with a Leap Motion hand gesture detector[8] and an Ultraleap Stereo IR 170[9] one. In particular, we encountered challenges with reliably detecting the user's hands while holding a bicycle's handlebar. As a result of these limitations, we excluded mid-air micro-gestures from further evaluation and focused on the remaining input methods.

5.1 Apparatus

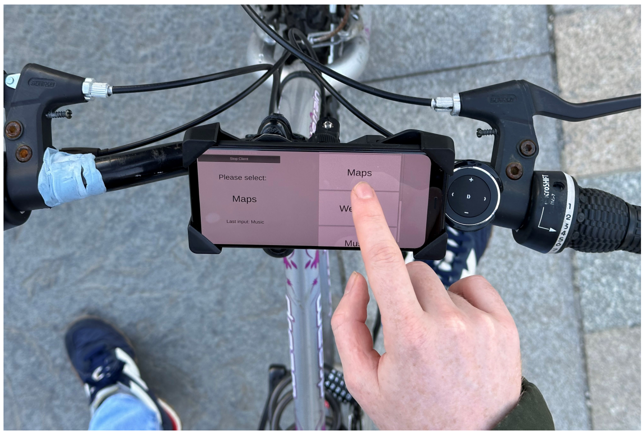

Figure 1: Apparatus set up for right-handed participants. The screen is showing the touchscreen menu for right-handed users. The smartphone is mounted on the right side of the handlebar, with the 'input' area as close as possible to the handlebar grip without covering the controller input device. This has been placed as close to the right hand as the off-the-shelf controller mount allowed, i.e. just left of the shifter. In the image, the user is completing the touchscreen condition tasks.

We carried out the experiment using a unisex hybrid bike. As the study involved comparing three input methods, our set-up involved a laptop and three peripherals: a smartphone, a headset and a controller. Additionally, we used a Livlov speed sensor[10] to measure cycling performance. We connected this sensor to the Wahoo Fitness app, which enabled us to gather detailed speed data at a granularity of one second intervals.

As AR glasses would have served mainly to dispense task instructions, we used a Google Pixel 4[11] smartphone instead to provide that information. We mounted the phone on the same side of the handlebar as the dominant hand.

---

[7] `https://www.ultraleap.com/tracking/gemini-hand-tracking-platform/`

[8] `https://www.ultraleap.com/product/leap-motion-controller/`

[9] `https://www.ultraleap.com/product/stereo-ir-170/`

[10] `https://www.amazon.co.uk/LIVLOV-Computer-Smart-Phone-Bluetooth-Wireless/dp/B08XNHDN6F?ref_=ast_sto_dp`

[11] `https://www.amazon.co.uk/Google-Pixel-64GB-Just-Black/dp/B07ZJKMXP9`

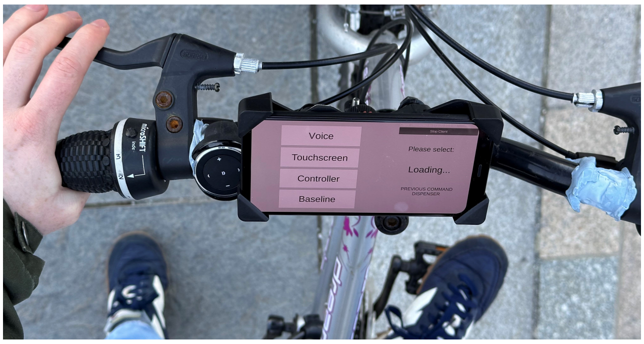

Figure 2: Apparatus for left-handed participants. The smartphone (in left-handed layout mode) is mounted on the left side of the handlebar. The screen is showing the main menu for left-handed users. Similarly to the apparatus layout for right-handed participants, the controller device has been placed just right of the left shifter.

The software setup consisted of a twin app developed on Unity version 2020.3.21f1. The first app ran on a Dell XPS laptop and featured WiFi communication with the smartphone, which hosted the second app, via the Unity plugin Mirror[12]. All peripherals were connected to the laptop and the phone was used to pick up touchscreen input and provide task instructions. Given that the experiment involved menu-item selection tasks, we divided the smartphone screen into two sections: the 'input' area and the 'instruction dispenser' area. The 'instruction dispenser' area simply provided the instruction for the next task, i.e. which menu item to select. Our 'input' area consisted of a main menu (see Figure 3) from which participants could access four further sub-menus. Since our study revolved around comparing three input methods, three of the sub-menus allowed the selection of the items via either voice, touchscreen or controller. A fourth sub-menu served to gather cycling performance data. The controller and touchscreen menus were designed to be identical. They feature six large buttons, with a large font (see Figure 4). Our choice of this menu structure was motivated by its versatility and suitability for various use cases, ranging from operating media players and navigation systems to searching contacts and retrieving cycling performance data. We implemented a period after each selection that deactivated button interactability. This prevented possible miss-clicks caused by the bumpiness of the road, thus reducing the likelihood of noisy data.

To overcome the logistical challenge of not having access to handlebar-embedded controls, we chose to implement such input as a controller mounted on the handlebar. This modification allowed us to simulate the experience of using embedded controls and ensured that the button-based input method was still available for testing. In addition, we conducted a thorough review of available off-the-shelf options for media buttons and identified the Satechi media button[13] as the most suitable for our study. This was based on a combination of factors, including its popularity with online reviewers in comparison to other similar products, as well as its design aligning with current trends of sleek and waterproof interfaces. For example, the Everysight Raptor Controller (Sec. 1) features a similar size, shape and button layout. By choosing this particular input device, we aimed to increase the external validity and generalisability of our results, while also ensuring that the technology used was both practical and relevant to current market trends. It featured five mechanical buttons laid out in a cross (See Figure 2). However, we only used the vertical three buttons to interact with the controller-based interface. We mounted the controller on the side of the user's dominant hand, like the smartphone, as shown in Figure 1. This positioning required users to grip the handlebar closer to the centre, which has been suggested as a viable option for providing input to digital devices while cycling by Porcheron et al. [34].

The voice menu, on the other hand, displayed no buttons, instead relying on the Unity Windows Speech Package to detect keywords. This package offers phrase recognition capabilities on Windows platforms using the operating system language. Since we are based in Glasgow, United Kingdom, we set our OS language to "*English (UK)*". To capture the speech commands, we provided participants with a Conambo Bluetooth Headset V5.1[14], selected for its ability to filter out a significant amount of background noise (as claimed by the manufacturer), and deliver clear voice audio. Despite this technology, we found during our pilot studies that the speech recogniser could fail several times in a row in detecting keywords. We found this to be dependent on both environmental noise and the participant's accent. We wanted to avoid this limitation in the technology to excessively impact the performance of the voice input. We implemented the voice sub-menu such that, if the system failed to recognise a keyword for a period of 20 seconds, the system would log an incorrect item selection and provide the next instruction. Furthermore, we used the headset to provide users with audio feedback at each item selection.

We implemented a left-handed version of the software which switches the position of the areas, as shown in Figure 4. For the right-handed version, the instruction dispenser area sits to the left of the screen and the input area sits at the right of the screen - the latter on an area of roughly the size of a Garmin Edge[15] display: 3.5" (88.9 mm) diagonal. For the left-handed version, this layout was reversed.

5.2 Experiment Design

We conducted a within-subjects study where each participant completed three conditions. Our independent variable was the input method used, and each condition allowed the participants to interact with the apparatus with one of the following: voice commands, touchscreen or button-based controller. We adopted a Latin square to avoid order bias. We measured the effects of the independent variable through a set of eight dependent variables, which encompassed both objective and subjective aspects of the participants' experience. The objective measures consisted of percentage preferred cycling speed (PPCS), which reflects the percentage change in the participant's cycling speed, as well as Task Completion Time (TCT) and task accuracy. In contrast, the subjective measures included perceived risk, privacy, so-

[12]https://mirror-networking.com/
[13]https://satechi.net/products/satechi-bluetooth-button-series?variant=34058399049
[14]https://www.amazon.co.uk/Wireless-Earbuds-Bluetooth-Mini-Headphones/dp/B088CVZ5DL
[15]https://www.garmin.com/en-GB/p/731136

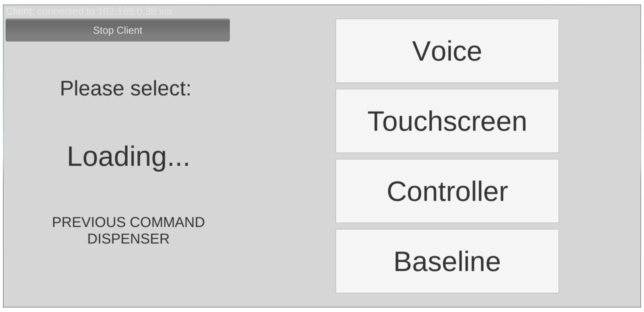

Figure 3: Screenshot of the smartphone app showing the main menu for right handed participants. On the right is the 'input' area with the buttons that direct users to the three condition sub-menus and the 'Baseline' sub-menu, used to gather participants' preferred cycling speed. On the left, we see the 'instruction dispenser' area, which for this menu serves no purpose.

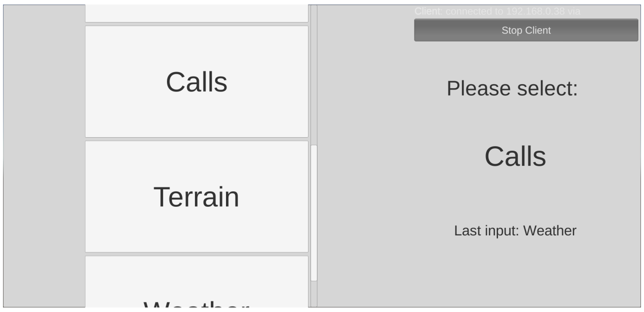

Figure 4: Screenshot of the left handed touchscreen menu. The 'input' area is to the left of the screen. Participants in the midst of the touchscreen condition would see this screen on the smartphone mounted on the handlebar, and select the given instruction. In this case, the instruction asks them to select the "Calls" item.

cial acceptability, usability, and workload

5.2.1 Track

As we decided to carry out a field experiment, we selected an outdoor route that would serve us in different aspects:

- The track should ensure passerby safety.
- The track should ensure participant safety.
- The track should offer realistic conditions for the experiment.

In view of such requirements, we chose a tract of the Kelvin Way: a pedestrian, relatively uncrowded, long, wide and straight road. The track featured dashed lines and a gentle decline. The former provided a visual guide that aided participants in cycling in a straight line. The latter led us to design the study such that participants would always cycle downhill, starting from the top of the track. We chose this route to minimise the effect of fatigue bias. The track is approximately 430 metres long and features a drop in altitude of roughly 4 metres from start to end. We provided every participant with the same starting and end position for every condition. If participants happened to arrive at the end position while still in a condition, they were instructed to turn around and proceed to complete tasks on the way back to the starting position.

5.2.2 Procedure

After reading an introduction script to the participants, we let them sign a consent form. In the script, we asked participants to cycle at their preferred pace and to complete each task as fast as possible. However, we remarked that their safety and the environment took priority and that, if needed, they could take their time completing a task or stray from the dashed line to overtake the pedestrians that might be in the way. Finally, we asked them to return to the dashed lines once having overtook potential pedestrians. Afterwards, we asked participants to fill out a demographic questionnaire and then let them put on a bike helmet, the headset, and adjust the bicycle seat to their height of preference. We allowed a static familiarisation period of 2 minutes per input method menu. We then asked participants to cycle in a straight line for a minute. The "baseline" button (see Figure 3) on the main menu gave participants instruction on when to stop to come back. This period allowed us to gather their preferred cycling speed.

The user study consisted of three runs down the track while cycling and while interacting with the interface. Depending on the condition, users would therefore use either the phone's touchscreen, the controller or voice commands to interact. We provided each participant with 24 item-selection tasks per condition. We did so via the 'instruction dispenser' area of the smartphone which provides the instruction as text. The order of the items to select was assigned randomly per condition. After each condition, we asked the participants to complete a survey comprising the following surveys:

- Rating Scale Mental Effort (RSME) [46] to measure perceived risk. The RSME is a uni-dimensional scale which ranges from 0 to 150. A rating of 12 denotes *"almost no effort"*, 58 is marked as *"rather much effort"*, and 112 as *"extreme effort"*. We adopted a modified version of the RSME used by De Waard et al. [10], which swaps the word *"effort"* for *"risk"*. We will refer to this scale as RSMR from now on.
- Perceived Privacy Risk (PPR) [27] to measure perceived privacy risk. This survey consists of 4 questions answered on a 7-point scale, from *"strongly disagree"* to *"strongly agree"*.
- International Positive and Negative Affect Schedule Short-Form (I-PANAS-SF) questionnaire to measure positive and negative social-acceptability-related affects [18]. The survey asks participants to rate 10 statements on a 5-point scale (*"not at all"* to *"extremely"*). Each statement inquires about one of 10 affective states. The five positive ones are: active, determined, attentive, inspired, and alert. The five negative states are: afraid, nervous, upset, hostile, and ashamed.
- System Usability Scale (SUS) questionnaire to measure usability [6].
- Raw NASA-Task Load Index (NASA-TLX) questionnaire to measure workload [15].
- To assess the impact on different attentional resources, we introduced three questions based on the aspects identified by Roider et al. [35]. Each question focused on one of three resources: visual demand, manual demand, and auditory demand. We modelled them after

the NASA-TLX sub-scales, and are measured on a 21 point scale. The questions were the following:

- *How visually demanding was the task?*
- *How manually demanding was the task?*
- *How auditorily demanding was the task?*

At the end of the experiment, we carried out a short semi-structured interview with each participant asking the advantages and disadvantages of each input method.

5.2.3 Hypotheses

Based on related literature, our online survey results and our focus group analysis, we formulated the following hypotheses:

H1 *Participants will cycle the fastest while using the voice input method.*

H2 *Participants will select items most accurately while using the controller input method.*

H3 *Participants will take the longest amount of time to select items while using the voice input method.*

H4 *Participants will rate the touchscreen input method the worst in terms of perceived risk.*

H5 *Participants will rate the voice input method the worst in terms of privacy preservation.*

H6 *Participants will rate the voice input method the worst in terms of social acceptability.*

H7 *Participants will rate the controller input method the best in terms of usability.*

H8 *Participants will rate the touchscreen input method the worst in terms of workload.*

5.3 Results

We analysed our data using a one-way ANOVA with post-hoc Tukey tests via the python packages SciPy [41] and statsmodels [37], respectively. For the latter test, we used an alpha value of 0.05. The study was approved by our University ethics committee.

5.3.1 Participants

We recruited 21 participants (8 female, 12 male, 1 other), mostly right-handed (2 left-handed), aged from 22 to 29 (M: 23.8, SD: 1.8). Of these, 13 participants reported cycling less often than once a month, 3 participants report cycling 1-3 times a month, one participant reported cycling 2-3 times a week, 5 participants reported cycling 4 or more times a week. Participation was voluntary. Additionally, 19 participants reported cycling for transport, 10 for recreation, and 6 for exercise or sport. Participants reported interacting with mobile devices via the following input methods while cycling: 18 via touchscreen, 14 via mechanical buttons, 1 via voice controls and 2 via touchpad.

5.3.2 Accuracy

A One-way ANOVA found significant differences in **accuracy** (F:8.25, p:<0.001). The pairwise Tukey test showed the touchscreen input method to perform significantly better than both the voice (Meandiff.:0.09, p:<0.01) and controller (Meandiff.:-0.11, p:0.001) input methods as shown in Figure 5 and Table 1. We found no significant difference between the controller and voice input methods in the effect on item selection accuracy (Meandiff.:0.02, p:0.78). Therefore, *H2* was not supported.

	Accuracy (%)		TCT (s)		PPCS (%)	
	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Controller	89.29*	7.04	3.92	1.01	-21.70*	21.27
Touchscreen	98.61†	2.41	3.09*	0.77	-14.33	11.03
Voice	87.30*†	14.93	4.78*	2.14	-6.97*	12.88
ANOVA	F(2,60):8.25, p:<0.001		F(2,60):7.23, p:<0.005		F(2,51):3.96, p:<0.05	

Table 1: Means and Standard Deviations of the objective measures of Accuracy, PPCS and Task Completion Time. Results were analysed via one-way ANOVA and difference between conditions were detected via pairwise Tukey HSD tests. Significantly different pairs are marked by the symbols * and †.

5.3.3 Task Completion Time

One-way ANOVA also found significant differences in **task completion time** (F:7.23, p:<0.005). We used *seconds* as the unit of measurement. The pairwise Tukey test showed the touchscreen input method to perform significantly better than the voice input method (Meandiff.:1.68, p:0.001) as shown in Figure 6 and Table 1. We found no significant difference between the controller and either the touchscreen (Meandiff.:-0.83, p:0.16) or voice (Meandiff.:0.86, p:0.14) in the effect on TCT. Therefore, *H3* was partly supported.

5.3.4 Percentage preferred cycling speed

The Wahoo app failed to register data for three participants, therefore the speed results are based on N = 18. One-way ANOVA found significant differences in **PPCS** (F:3.96, p:<0.05). The pairwise Tukey test showed the voice input method to perform significantly better than the controller (Meandiff.:14.73, p:<0.05) input method as shown in Figure 7 and Table 1. We found no significant difference between the touchscreen and either the controller (Meandiff.:7.37, p:0.34) or voice (Meandiff.:7.36, p:0.35) in the effect the input method had on preferred cycling speed. Therefore, *H1* was partly supported.

5.3.5 Perceived risk

A one-way ANOVA test found significant differences in **perceived risk**, i.e. the **RMSR scores** (F:15.50, p:<0.001). The pairwise Tukey test showed the voice input method to perform significantly better than both the controller (Meandiff.:-33.76, p:<0.001) and touchscreen (Meandiff.:-28.19, p:<0.001) input methods as shown in Figure 9 and Table 2. We found no significant difference between controller and touchscreen in the effect on the perceived risk of the interaction (Meandiff.:-5.57, p:0.67). Therefore, *H4* was partly supported.

5.3.6 Perceived workload

One-way ANOVA also found significant differences in **perceived workload** (F:4.60, p:0.01), corresponding to the total score combined out of the NASA-TLX sub-scales. The pairwise Tukey test showed the voice input method to perform significantly better than both the controller (Meandiff.:-10.52, p:<0.05) and touchscreen (Meandiff.:-12.18, p:<0.05) input methods as shown in Figure 8 and Table 2. There was no significant difference between the controller and touchscreen input methods in terms of the total NASA-TLX score (Meandiff.:1.66, p:0.92). Therefore, *H8* was partly supported.

Three of the **NASA-TLX sub-scores** parallel the total score closely as shown in Table 4. Mental demand, physical demand and effort all were found to feature significant differ-

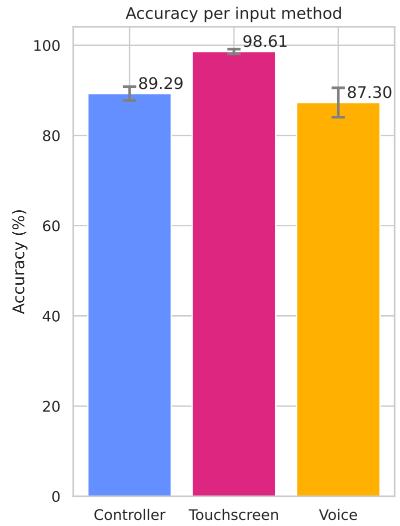

Figure 5: Bar chart for the accuracy ratio per input method. Error bars show standard error.

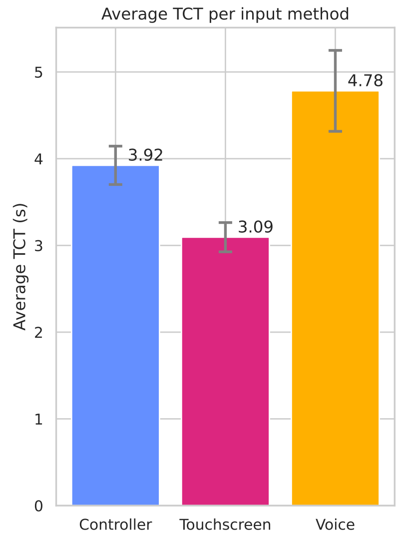

Figure 6: Bar chart for the task completion time per input method. Error bars show standard error.

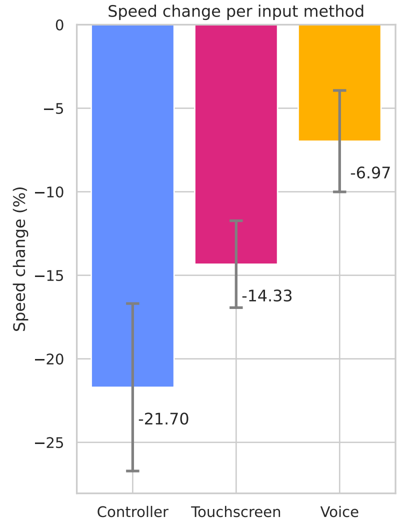

Figure 7: Bar chart for the percentage preferred cycling speed per input method. Error bars show standard error.

	RSMR		PPR		I-PANAS-SF (Pos.)		I-PANAS-SF (Neg.)		SUS		NASA-TLX (Total)	
	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Controller	54.38*	24.10	10.62	6.64	12.24	3.81	9.14	3.44	62.38	14.61	40.24*	13.65
Touchscreen	48.81†	23.88	8.67	5.04	12.43	3.61	8.48	2.56	66.90	18.39	41.90†	18.17
Voice	20.62*†	13.42	12.48	4.98	12.38	4.21	8.86	3.69	73.69	14.33	29.72*†	8.97
ANOVA	F(2,60):15.50,p:<0.001		F(2,60):2.43,p:0.09		F(2,60):0.01,p:0.99		F(2,60):0.22,p:0.80		F(2,60):2.70,p:0.08		F(2,60):4.60,p:0.01	

Table 2: Subjective measures. Means and Standard Deviations of the scores of the RMSR, PPR, I-PANAS-SF, SUS and NASA-TLX surveys. Results were analysed via one-way ANOVA and difference between conditions were detected via pairwise Tukey HSD tests. Significantly different pairs are marked by the symbols * and †.

	Visual Demand		Manual Demand		Auditory Demand	
	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Controller	40.24*	24.52	36.90*	26.71	29.76*	19.81
Touchscreen	45.00†	33.54	35.48†	21.62	24.05†	18.94
Voice	18.81*†	16.95	13.57*†	12.66	35.00*†	18.81
ANOVA	F(2,60):21.99,p:<0.001		F(2,60):28.03,p:<0.001		F(2,60):10.24,p:<0.001	

Table 3: Means and Standard Deviations of the attention resources demand questions scores. Results were analysed via one-way ANOVA and difference between conditions were detected via pairwise Tukey HSD tests. Significantly different pairs are marked by the symbols * and †.

ences by ANOVA. The **mental demand** score (F:6.09, p:<0.005) revealed participants were affected by the voice condition significantly less than both controller (Meandiff.:-21.43, p:<0.05) and touchscreen (Meandiff.:-26.19, p:<0.005) conditions. The **physical demand** score (F:8.03, p:<0.001) revealed participants were affected by the voice condition significantly less than both controller (Meandiff.:-23.33, p:<0.005) and touchscreen (Meandiff.:-21.90, p:<0.005) conditions. The **effort** score (F:4.97, p:0.01) revealed participants were affected by the voice condition significantly less than both controller (Meandiff.:-20.48, p:0.01) and touchscreen (Meandiff.:-17.86, p:<0.05) conditions. However, one-way ANOVA found **no** significant differences in either **temporal demand** (F:1.71, p:0.19), **performance** (F:2.29, p:0.11), or **frustration** (F:0.07, p:0.93).

5.3.7 *Visual, manual and auditory demand*

Additionally, one-way ANOVA also found significant differences in the scores given for the **visual demand** (F:21.99, p:<0.001), the **manual demand** (F:28.03, p:<0.001), and the **auditory demand** (F:10.24, p:0.001) questions as shown in Table 3. In the case of visual demand, the pairwise Tukey test showed the voice input method to be significantly less demanding than both the controller (Meandiff.:-7.43, p:<0.001) and touchscreen (Meandiff.:-8.86, p:<0.001) input methods. We found no significant difference in the effect on visual demand between controller and touchscreen input methods (Meandiff.:1.43, p:0.58). In the case of manual demand, the pairwise Tukey test showed the voice input method to be significantly less demanding than both the controller (Meandiff.:-9.48, p:<0.001) and touchscreen (Meandiff.:-8.86, p:<0.001) input methods. We found no significant difference in the effect on visual demand between controller and touchscreen input methods (Meandiff.:-0.62, p:0.90). In the case of auditory demand, the pairwise Tukey test showed the voice input method to be significantly **more** demanding than both the controller (Meandiff.:4.0, p:<0.001) and touchscreen (Meandiff.:4.14, p:<0.001) input methods. We found no significant difference in the effect on visual demand between controller and voice input methods (Meandiff.:-0.14, p:0.99).

5.3.8 *Privacy, usability and social acceptability*

Finally, one-way ANOVA found **no** significant effects caused

	Mental Demand		Physical Demand		Temporal Demand		Performance		Effort		Frustration	
	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Controller	40.24*	24.52	36.90*	26.71	29.76	19.81	45.24	29.26	48.33*	23.26	35.71	27.94
Touchscreen	45.00†	33.54	35.48†	21.62	24.05	18.94	62.62	26.67	45.71†	25.51	32.86	26.81
Voice	18.81*†	16.95	13.57*†	12.66	35.00	18.81	58.81	27.06	27.86*†	19.59	35.24	25.52
ANOVA	F(2,60):6.09, p:<0.005		F(2,60):8.03, p:<0.001		F(2,60):1.71, p:0.19		F(2,60):2.29, p:0.11		F(2,60):4.97, p:0.01		F(2,60):0.07, p:0.93	

Table 4: Means and Standard Deviations of the NASA-TLX subscales. Results were analysed via one-way ANOVA and difference between conditions detected via pairwise Tukey HSD tests. Significantly different pairs are marked by the symbols * and †.

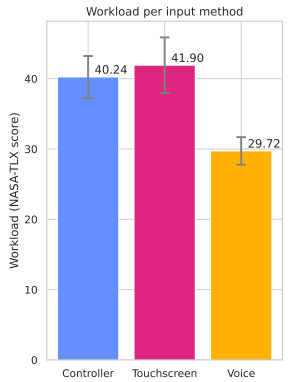

Figure 8: Bar chart for the NASA-TLX Total score per input method. Error bars show standard error.

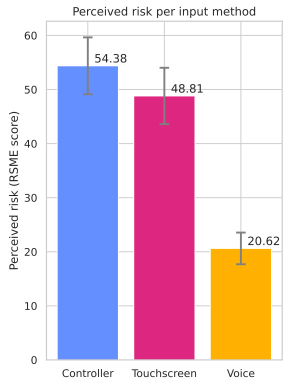

Figure 9: Bar chart for the RSMR Perceived Risk score per input method. Error bars show standard error.

by either **Perceived privacy (PPR) scores** (F:2.43, p:0.10), **usability (SUS) scores** (F:2.70, p:0.08), or **Social acceptability (I-PANAS-SF) scores** (F:0.01, p:0.99 for the positive emotions score and F:0.22, p:0.80 for the negative ones). Therefore, *H5*, *H6*, and *H7* were not supported.

5.4 Exit Interviews

We conducted a short semi-structured exit interview at the end of each user study. We investigated the participant's perception of the benefits and disadvantages of each input method. We used Initial Coding [36] to analyze the interviews. Afterward, we organized the codes based on their relevance and frequency, including those that were mentioned by more than one-third of the participants. We included relevant excerpts to illustrate the identified codes.

5.4.1 Controller

Fourteen participants mentioned reasons as to why they believed the controller method was unsafe. Eleven participants (P2, 4, 5, 6, 13, 14, 15, 16, 17, 18, 19) attributed such perception to the specific design of the off-the-shelf model used in the experiment:

> *P16: You couldn't quite tell which button was which. If the buttons were slightly more 3D and a bit more distinguishable then it would be easier.*

Furthermore, seven participants (P1, 2, 4, 7, 12, 17, 20, 21) mentioned that the controller input felt less safe because it required users to look at the screen while navigating.

5.4.2 Touchscreen

The touchscreen input method attracted similar comments, with fifteen participants (P1, 2, 4, 5, 6, 7, 10, 12, 13, 14, 15, 16, 18, 19, 21) mentioning that touchscreen felt unsafe as it inherently required the user to look at the screen instead of paying attention to the road.

> *P14: I just felt like my eyes were glued to the screen more than they should have been, so it didn't feel as safe because I would keep looking up and down.*

5.4.3 Voice

Eighteen participants mentioned some cons when it came to interacting via the voice input. In particular, fifteen (P1, 2, 3, 4, 5, 6, 7, 8, 10, 12, 14, 15, 17, 19, 21) found the dialogue system to struggle to detect words. Often cited reasons were difficulty in understanding the user's accent and wind noise:

> *P17: It would just regularly not pick up what I was saying, especially if I was going fast, which I gather would be something to do with the wind.*

Eleven participants criticised the input method's poor social acceptability:

> *P1: It was very awkward to be screaming at your phone and just the social acceptability of it being kind of low.*

Finally, twelve participants (P1, 2, 3, 5, 6, 11, 14, 15, 16, 17, 18, 19) praised the voice input for being eyes-free, and therefore feeling more safe.

6. DISCUSSION

6.1 Limitations

It should be noted that our study is subject to the following limitations. Our experiment compared three input methods, each using a different component of the apparatus. For the button-based input, we used an off-the-shelf Bluetooth controller. While this product is commonly used in mobile contexts and we chose it for its similarity to commercially available cycling AR glasses button controls, we found that its design was not optimized specifically for cyclists. As a result, participants struggled distinguishing the buttons, which could have biased the comparison. Future work should instead use button controllers with 3D, easily distinguishable buttons, such that users are able to interact with it without excessive effort. Furthermore, our software used the Window system speech recognition technology used by the Cortana conversational interface. Although widely used, this technology is not without limitations. Like other state-of-the-art voice-to-text software, it struggles to filter out environmental noise completely and has difficulty understanding regional accents. We selected a track that included a gentle descent to reduce fatigue bias. However, the track descent also introduced wind noise that affected the speech detection capabilities of the system. Although wind noise is a known challenge [20] for voice input in outdoor contexts, a planar track without additional environmental noise would have provided a fairer and more generalisable comparison. Additionally, we noted that participants did not need to pedal to maintain their preferred cycling speed while going downhill, which could have potentially affected the measure. Regarding regional accents, it is expected that the voice input technology can gather speech commands from anyone, regardless of their accent but that is not yet the state of the art. As we carried out our experiment in Glasgow, Scotland, with participants from an international group of students, most of the participants had 'regional' accents that were not included in the available set of English varieties the Windows voice recognition model is trained on. Although inevitable, given the location and participant demographics, future studies should consider using a more diverse set of participants or using speech recognition technologies that are trained on a wider range of accents.

6.2 Objective measures

The study examined three objective dependent variables: accuracy, percentage preferred cycling speed, and task completion time. Significant results were found in all three measures.

Regarding *RQ1*, our results that the touchscreen input method had the least impact on task performance, achieving the highest accuracy and shortest TCT. In contrast, the voice input method performed relatively poorly, despite not requiring menu item navigation. We attribute this to the difficulties fifteen participants experienced with the voice input system due to either wind noise or their accent. This resulted in a slower selection process. Additionally, we speculate that the participants' familiarity with the touchscreen input method may have contributed to its better performance.

Our study found that the touchscreen input method was significantly more accurate than both the controller and the voice input. We attribute these results to two factors. Firstly, as mentioned previously, the voice input method struggled to detect input from specific accents and in noisy environments, which may have impacted its accuracy. Secondly, the choice of controller used in our study may have contributed to lower accuracy for the controller input method. The Satechi media button lacks 3D buttons, featuring instead a flat and homogeneous design, which made it difficult for some participants to distinguish between buttons. Eleven participants reported misclicking for this reason and some even reported pressing the selection button situated between the up and down buttons while moving their finger between them.

In regards to *RQ2*, the voice input performed best in terms of percentage preferred cycling speed, therefore impacting cycling performance the least. These findings are consistent with previous research [10, 11], which has shown that cyclists slow down significantly when interacting with their voice and even more when having to text via manual input.

6.3 Subjective measures

In regards to *RQ3*, we see that for most measures (perceived privacy, negative emotions, positive emotions and usability) there is no significant difference between the three inputs.

Regarding the impact on the feeling of privacy, we speculate that the input method did not significantly affect the perceived risk of privacy loss due to two reasons. Firstly, it is unlikely that users would input sensitive information while cycling. Secondly, as *P12* mentioned in the exit interview: "*If you're cycling at a decent enough speed, then people can't hear the whole conversation, so* [privacy] *shouldn't matter*". This suggests that the speed of cycling may make it difficult for others to overhear conversations, reducing concerns about privacy.

Our findings show no input to be significantly more socially acceptable than any other. This is consistent with Angelini et al.'s findings that there is no difference in negative or positive affect between touch and voice input. Surprisingly, despite the voice condition requiring participants to speak loudly while surrounded by others, as mentioned in several exit interviews, participants did not report negative emotions related to social acceptability. However, the touchscreen condition required participants to interact with a smartphone while cycling with one hand, potentially causing feelings of self-consciousness, as expressed by P20: "*I really felt like I shouldn't be looking at a screen as much when I'm cycling in a busy street.*"

We observed notable differences in usability between the three input methods, although they were not statistically significant (p=0.08). We believe that these differences could be attributed to the varying levels of familiarity that participants had with the input methods. As shown in the online survey, a great number (76.09%) of cyclists interacts with mobile devices via touchscreen input while cycling, and a slightly smaller amount (34.78%) via mechanical buttons. Research has shown that a higher level of user experience with an interface is associated with higher usability ratings [22]. Therefore, it is possible that participants' familiarity with touchscreen input led to it being perceived as more usable than the other methods. Future work should investigate and compare the usability of the three input methods recruiting participants with similar levels of experience with all three. Additionally, the voice input method faced difficulties in detecting input, as noted by 15 participants who reported issues with the system's ability to recognize words, possibly due to their accents or wind noise. This could have had a negative impact on the usability ratings reflected in the SUS scores.

Our results show that the input method significantly impacts two measures: perceived risk and workload.

Our study show that participants perceived the voice input method to be significantly less risky than the other two. These findings are consistent with De Waard's research, which demonstrated that cyclists perceive manual input as less safe than speech input.

The scores from the NASA-TLX survey differed significantly between input methods. The total score, which provides an indication of the workload of an interface, was found to be significantly lower for the voice input compared to either of the other two methods. Twelve participants provided positive feedback during the exit interview and praised the voice input for feeling safer and easier to use due to being largely eyes-free. It is worth noting that while only one participant mentioned the hands-free aspect of the voice input, more than half of the participants praised it for not requiring them to take their eyes off the road. Based on these findings, future research should investigate and compare various eyes-free input methods, including touch input located on the frames of AR glasses, to validate and quantify the necessity for hands-free input methods for cyclists. It is worth noting some of the sub-scales of the NASA-TLX survey. The performance sub-scale revealed that participants did not feel a significant difference in performance between the input methods, despite accuracy measures showing a clear difference between touchscreen and the other two methods. One possible explanation is that voice input inherently offers users the certainty about which command was given, while touchscreen input lacks it. Similarly, touchscreen input lacks the mechanical haptic feedback that the controller input offers. As such, although users performed better with touchscreen input in terms of accuracy, they did not perceive it as such. In terms of the effort sub-scale, our results are consistent with the findings of De Waard et al. It is worth noting that while the authors used the RSME scale to measure effort, we used one of the sub-scales of NASA-TLX. Despite the difference in measurement tool, we found the voice condition to require significantly less effort than manual input methods.

Finally, in terms of visual and auditory demand, the voice input outperformed both other methods with participants giving significantly better scores. This result was expected as the voice input is both eyes-free and hands-free, allowing cyclists to focus more on the road and their surroundings, and being less demanding on their balance and dexterity. However, the voice input performed significantly worse in terms of auditory demand. This was expected, as shown in the study by Roider et al., where speech interaction was found to impair auditory attentional resources significantly more than gaze or manual interaction.

6.4 Final thoughts & Future work

While all three input methods tested proved usable, voice input emerged as the safest and least demanding, despite its slower task completion time and lower accuracy. Based on these findings, we recommend integrating voice input as a necessary method in future cycling AR glasses. However, alternative input methods are still needed, as voice input does not perform as efficiently as touchscreen, particularly in noisy environments and with unfamiliar accents due to the current limitations of voice-to-text technology. Touchscreen performed well in overall task performance, likely due to the participant's familiarity with it. Although button controls did not perform significantly poorly, they often underperformed compared to the other two input methods and never outperformed them. One of the main reasons for this result seems to reside in the model of controller used in the experiment, and future work should investigate with different button-based products or solutions. Moreover, it is recommended that future studies explore the potential of micro-hand gestures as an alternative input method. However, significant advancements in the current state-of-the-art hand-tracking technology are necessary before its feasibility can be tested and validated. Finally, it is worth noting that our study did not incorporate cycling AR glasses, opting instead to rely on the existing apparatus to provide task instructions for the experiment. While we don't believe this had a significant impact on the performance of the voice or touchscreen input methods, we suspect the controller input method would perform better in terms of workload, and likely other measures, if participants were able to interact with the interface without having to look down while navigating a menu. In light of this, future studies that compare button controllers to other input methods for cycling AR glasses should consider incorporating them into their apparatus.

7. CONCLUSIONS

Our study aimed to investigate and compare suitable input methods for AR Glasses designed for cyclists. We conducted an online survey to identify which input methods cyclists used. We found touchscreen input and mechanical button-based input to be the most common methods used. We conducted a focus group to investigate the survey's results and to identify the most desirable input methods for AR glasses. We found users to consider voice and controller inputs as the most viable, followed by micro-hand gestures. Touchscreen input was considered acceptable only for bike-computer-like devices. Finally, we analysed and compared the viability of voice input, button-based input, and touchscreen input for AR Glasses and found that the voice input method performs best in terms of percentage preferred cycling speed, perceived risk, and workload. However, touchscreen input performed best in terms of accuracy, and task completion time. Results show voice input to be a viable input method for cycling AR glasses, with touchscreen as an acceptable alternative method. Future work should investigate and compare novel input modalities.

Acknowledgments. I would like to thank Professor Stephen Brewster for his invaluable support throughout the year. Furthermore, I would also like to thank the members of the Multimodal Interaction Group for their guidance. Finally, I would like to thank my family and friends, who supported me through my time at university.

8. REFERENCES

[1] E. Adell, A. Nilsson, and K. Kircher. Cyclists' use of mobile IT in Sweden: usage and self-reported behavioural compensation. Nov. 2014.

[2] Adventure Art (Chris Owen Halper). Cycling HUD comparison - Solos Wearables, Everysight Raptor, Varia Vision, Recon Jet review, July 2018.

[3] A. Al Taie, Y. Abdrabou, S. Macdonald, F. Pollick, and S. Brewster. Keep it Real: Investigating Driver-Cyclist Interaction in Real-World Traffic, Jan. 2023. Conference Name: 2023 CHI Conference on Human Factors in Computing Systems (CHI '23) Meeting Name: 2023 CHI Conference on Human Factors in Computing Systems (CHI '23) Place: Hamburg, Germany.

[4] L. Angelini, J. Baumgartner, F. Carrino, S. Carrino, M. Caon, O. A. Khaled, J. Sauer, D. Lalanne, E. Mugellini, and A. Sonderegger. A comparison of three interaction modalities in the car: gestures, voice and touch. In *Actes de la 28ième conference francophone sur l'Interaction Homme-Machine*, IHM '16, pages 188–196, New York, NY, USA, Oct. 2016. Association for Computing Machinery.

[5] Brad Dean. Engo Hud Ar Glasses | Pain and Cycling, Nov. 2021.

[6] j. Brooke. SUS: A 'Quick and Dirty' Usability Scale.

In *Usability Evaluation In Industry*. CRC Press, 1996. Num Pages: 6.

[7] M. Caon, R. Süsse, B. Grelier, O. A. Khaled, and E. Mugellini. Gesturing on the Handlebar: A User-Elicitation Study for On-Bike Gestural Interaction. In S. Bagnara, R. Tartaglia, S. Albolino, T. Alexander, and Y. Fujita, editors, *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018)*, Advances in Intelligent Systems and Computing, pages 429–439, Cham, 2019. Springer International Publishing.

[8] I. Chaturvedi, F. H. Bijarbooneh, T. Braud, and P. Hui. Peripheral vision: a new killer app for smart glasses. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*, IUI '19, pages 625–636, New York, NY, USA, Mar. 2019. Association for Computing Machinery.

[9] A. Dancu, V. Vechev, A. A. Ünlüer, S. Nilson, O. Nygren, S. Eliasson, J.-E. Barjonet, J. Marshall, and M. Fjeld. Gesture Bike: Examining Projection Surfaces and Turn Signal Systems for Urban Cycling. In *Proceedings of the 2015 International Conference on Interactive Tabletops & Surfaces*, ITS '15, pages 151–159, New York, NY, USA, Nov. 2015. Association for Computing Machinery.

[10] D. De Waard, B. Lewis-Evans, B. Jelijs, O. Tucha, and K. Brookhuis. The effects of operating a touch screen smartphone and other common activities performed while bicycling on cycling behaviour. *Transportation Research Part F: Traffic Psychology and Behaviour*, 22:196–206, Jan. 2014.

[11] D. de Waard, P. Schepers, W. Ormel, and K. Brookhuis. Mobile phone use while cycling: Incidence and effects on behaviour and safety. *Ergonomics*, 53(1):30–42, Jan. 2010. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/00140130903381180.

[12] A. W. Gellatly and T. A. Dingus. Speech Recognition and Automotive Applications: Using Speech to Perform in-Vehicle Tasks. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 42(17):1247–1251, Oct. 1998. Publisher: SAGE Publications Inc.

[13] E. Ginters. Augmented reality use for cycling quality improvement. *Procedia Computer Science*, 149(C):167–176, Jan. 2019.

[14] C. Goldenbeld, M. Houtenbos, E. Ehlers, and D. De Waard. The use and risk of portable electronic devices while cycling among different age groups. *Journal of Safety Research*, 43(1):1–8, Feb. 2012.

[15] S. G. Hart and L. E. Staveland. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In P. A. Hancock and N. Meshkati, editors, *Advances in Psychology*, volume 52 of *Human Mental Workload*, pages 139–183. North-Holland, Jan. 1988.

[16] W. Hochleitner, D. Sellitsch, D. Rammer, A. Aschauer, E. Mattheiss, G. Regal, and M. Tscheligi. No need to stop: exploring smartphone interaction paradigms while cycling. In *Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia*, pages 177–187, Stuttgart Germany, Nov. 2017. ACM.

[17] M. Ichikawa and S. Nakahara. Japanese High School Students' Usage of Mobile Phones While Cycling. *Traffic Injury Prevention*, 9(1):42–47, Feb. 2008. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/15389580701718389.

[18] J. Karim, R. Weisz, and S. U. Rehman. International positive and negative affect schedule short-form (I-PANAS-SF): Testing for factorial invariance across cultures. *Procedia - Social and Behavioral Sciences*, 15:2016–2022, Jan. 2011.

[19] G. Kim and F. Biocca. Immersion in Virtual Reality Can Increase Exercise Motivation and Physical Performance. In J. Y. Chen and G. Fragomeni, editors, *Virtual, Augmented and Mixed Reality: Applications in Health, Cultural Heritage, and Industry*, Lecture Notes in Computer Science, pages 94–102, Cham, 2018. Springer International Publishing.

[20] T. Kitagawa and K. Kondo. Evaluation of a Wind Noise Reduction Method Using DNN for Bicycle Audio Augmented Reality Systems. In *2018 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW)*, pages 1–2, May 2018. ISSN: 2575-8284.

[21] P. Kortum. *HCI Beyond the GUI: Design for Haptic, Speech, Olfactory, and Other Nontraditional Interfaces*. Elsevier, June 2008. Google-Books-ID: 9uzEukFA4ZQC.

[22] P. Kortum and M. Johnson. The Relationship Between Levels of User Experience with a Product and Perceived System Usability. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 57(1):197–201, Sept. 2013. Publisher: SAGE Publications Inc.

[23] T. Kosch, A. Matviienko, F. Müller, J. Bersch, C. Katins, D. Schön, and M. Mühlhäuser. NotiBike: Assessing Target Selection Techniques for Cyclist Notifications in Augmented Reality. *Proceedings of the ACM on Human-Computer Interaction*, 6(MHCI):197:1–197:24, Sept. 2022.

[24] R. A. Krueger and M. A. Casey. *Focus groups: a practical guide for applied research*. SAGE, Los Angeles London New Delhi Singapore Washington DC, 5th edition edition, 2015.

[25] J. Lazar, J. H. Feng, and H. Hochheiser. Chapter_8 - Interviews and focus groups. In J. Lazar, J. H. Feng, and H. Hochheiser, editors, *Research Methods in Human Computer Interaction (Second Edition)*, pages 187–228. Morgan Kaufmann, Boston, Jan. 2017.

[26] E. Litvak and T. Kuflik. Enhancing cultural heritage outdoor experience with augmented-reality smart glasses. *Personal and Ubiquitous Computing*, 24(6):873–886, Dec. 2020.

[27] N. K. Malhotra, S. S. Kim, and J. Agarwal. Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model. *Information Systems Research*, 15(4):336–355, Dec. 2004. Publisher: INFORMS.

[28] Mariella Moon. Lasers can silently issue 'voice commands' to your smart speakers, Nov. 2019.

[29] N. Mascret, G. Montagne, A. Devrièse-Sence, A. Vu, and R. Kulpa. Acceptance by athletes of a virtual

reality head-mounted display intended to enhance sport performance. *Psychology of Sport and Exercise*, 61:102201, July 2022.

[30] S. Mennicken and E. M. Huang. Hacking the Natural Habitat: An In-the-Wild Study of Smart Homes, Their Development, and the People Who Live in Them. In J. Kay, P. Lukowicz, H. Tokuda, P. Olivier, and A. Krüger, editors, *Pervasive Computing*, Lecture Notes in Computer Science, pages 143–160, Berlin, Heidelberg, 2012. Springer.

[31] J.-M. Morgan. Roundup: Three new super lightweight ActiveLook-powered smart sunglasses for cyclists, Nov. 2021.

[32] Olympics. The Cycling Smart-Glasses with Realtime Data | The Tech Race, July 2017.

[33] B. Poppinga, M. Pielot, and S. Boll. Tacticycle: a tactile display for supporting tourists on a bicycle trip. In *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '09, pages 1–4, New York, NY, USA, Sept. 2009. Association for Computing Machinery.

[34] M. Porcheron, L. Clark, S. A. Nicholson, and M. Jones. Cyclists' Use of Technology While on Their Bike. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, pages 1–15, New York, NY, USA, Apr. 2023. Association for Computing Machinery.

[35] F. Roider, S. Rümelin, B. Pfleging, and T. Gross. The Effects of Situational Demands on Gaze, Speech and Gesture Input in the Vehicle. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 94–102, Oldenburg Germany, Sept. 2017. ACM.

[36] J. Saldaña. *The coding manual for qualitative researchers*. SAGE, Los Angeles, 2nd ed edition, 2013. OCLC: ocn796279115.

[37] S. Seabold and J. Perktold. Statsmodels: Econometric and Statistical Modeling with Python. pages 92–96, Austin, Texas, 2010.

[38] J. Shang and J. Wu. Secure Voice Input on Augmented Reality Headsets. *IEEE Transactions on Mobile Computing*, PP:1–1, Aug. 2020.

[39] G. Sörös, F. Daiber, and T. Weller. Cyclo: a personal bike coach through the glass. In *SIGGRAPH Asia 2013 Symposium on Mobile Graphics and Interactive Applications*, SA '13, pages 1–4, New York, NY, USA, Nov. 2013. Association for Computing Machinery.

[40] Y. Tan, S. H. Yoon, and K. Ramani. BikeGesture: User Elicitation and Performance of Micro Hand Gesture as Input for Cycling. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '17, pages 2147–2154, New York, NY, USA, May 2017. Association for Computing Machinery.

[41] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, and P. van Mulbregt. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3):261–272, Mar. 2020. Number: 3 Publisher: Nature Publishing Group.

[42] T. von Sawitzky, P. Wintersberger, A. Löcken, A.-K. Frison, and A. Riener. Augmentation Concepts with HUDs for Cyclists to Improve Road Safety in Shared Spaces. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI EA '20, pages 1–9, New York, NY, USA, Apr. 2020. Association for Computing Machinery.

[43] R. Wilmott. Everysight Raptor Smartglasses review, Dec. 2019.

[44] P. W. Woźniak, L. Dekker, F. Kiss, E. Velner, A. Kuijt, and S. F. Donker. Brotate and Tribike: Designing Smartphone Control for Cycling. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 1–12, Oldenburg Germany, Oct. 2020. ACM.

[45] K. L. Young, A. N. Stephens, S. O'Hern, and S. Koppel. Australian cyclists' engagement in secondary tasks. *Journal of Transport & Health*, 16:100793, Mar. 2020.

[46] F. R. H. Zijlstra. Efficiency in work behaviour: A design approach for modern tools. 1993.