University of Glasgow | School of Computing Science

Honours Individual Project Dissertation

An eyes-free, multimodal interface for Auditory Headsets

Diego Drago

March 12, 2022

Abstract

Despite the growing interest in novel wearable devices, headsets and earphones are ever-improving technologies that have yet to be introduced to any computing power. Our project aims to leverage their capabilities and design an eyes-free interface for auditory headsets. We developed a multimodal menu interface based on head-orientation tracking and gestural input. We used Unity to implement the audio interface for the Bose Frames and evaluated it in a user study to investigate the performance, usability and workload of three possible layouts. Results show the pie layout to be viable and the UI overall to be promising.

Education Use Consent

Contents

1 Introduction

> Interaction design is about shaping digital things for people's use
>
> *Interaction Design - brief intro*
> *Löwgren*

In this quote, the use of the verb 'shaping' helps us understand the role of the designer. It is not that of an engineer who 'builds', but a role that involves using already present resources to create something new. In this project, we endeavour to make use of already present technologies to develop a novel way to interact with them.

In this chapter, we will introduce the motivations and aims for our project.

1.1 Motivation

With the commercial success of Smart Watches and the growing attention towards novel mobile devices, for the past decade wearable computing has been an increasingly important area of interest for the Mobile Human-Computer Interaction research community.

Although several companies are leveraging the recent improvements in hardware to produce mobile technologies with growing computing power, few have had as widespread a success as Smart Watches, which over the years have undergone significant changes and began providing similar functionalities to smartphones. From computerised motorcycle helmet systems (Dillet (2016), Bohn (2019)) to Smart Rings (Oura (2022)), this decade is experiencing an evolutionary arms race to the development of wearable computing devices.

Furthermore, products like Google Glass launched in 2012 (Goldman (2012)), Amazon Echo Frames announced in 2019 (Smith (2019)) and the Oculus Rift in 2015 (Lang (2019)) brought augmented and virtual reality alike closer and closer. The latter, proving to be a market success story, can now be found in many households as a relatively common gaming console.

Hardware is consistently improving even where computing power has not yet been introduced. Headsets are a good example of this, often providing an arsenal of sensors the use of which is then restrained to sensing gestures or providing non-functional features. Bose Frames, developed by Bose (2022a), are a transparent over-the-ear headset, with integrated directional speakers, two hardware sensors and two virtual sensors that allow users to interact with the device using hand gestures like nodding or double-tapping. This wireless headset takes the shape of a pair of sunglasses that connects to the user's device via Bluetooth, a feature that is made twice as useful as it allows the user's device to not only stream audio but receive both gesture and sensor input alike.

In this project, we aimed to leverage the promising but dormant capabilities of such an off-the-shelf device and explore a possible way to transform them into a proper wearable computer. In particular, we explore how can menus work on an audio-only interface.

An audio-only, i.e. eyes-free, interface offers several advantages. A glaring benefit would provide an alternative interaction method for demographics that are not able to use vision as their input sensing ability, such as visually impaired people, but specific contexts may also arise the need for non-graphic UIs, such as professions that require its users to keep their eyes on the task at hand. Astronauts, Deep sea divers, Sports Referees and manual workers of various kinds might find it useful to be able to access the capabilities of a smartphone without having to rely on their eyes to give and retrieve information.

Conversational interfaces are a tried and tested system that allows audio-only input and audio. Also called dialogue systems, they are another piece of technology that has become ubiquitous in recent years. Usually in the form of voice assistants, they can be found mainly in laptops, smartphones, smartwatches and smart home appliances and it certainly seems to be integrated into every product beginning with "smart-". Advancements in Natural Language Understanding made such a technology incredibly powerful in the span of a decade, as can be seen in the ability of GPT-3 to convincingly pass the Turing test, the new lengths Google went to in order to develop a natural-sounding, if not slightly concerning, voice assistant (Jeff Grubb's Game Mess (2018)). However, precisely these advancements make conversational interfaces a controversial technology for users who highly regard their privacy and other ethical concerns. In an article by Vincent (2018), we notice the stark difference between the public reception of Google Duplex compared to how it was received by the HCI community. Furthermore, in the infamous article by Bender et al. (2021), the authors provide a plethora of reasons to be sceptical of conversational interfaces and their advancements. Privacy concerns are an important weakness in the ethics and the social acceptability of dialogue systems, as both Bender et al. (2021) and Smith (2019) outline, the latter with a particular focus on the Amazon Echo Frames.

The development of a novel eyes-free and audio-only UI should therefore be considered an attempt to fill a niche in the market. Furthermore, it aims to provide users with affordable wearable computing that makes use of already present technology and with a privacy-conscious alternative to the ethically concerning input methods of conversational interfaces.

1.2 Aims

This project aimed to use the capabilities of the Bose Frames to create an "audio augmented reality" where sounds from the environment could be augmented with feedback from a linked phone in order to acoustically display a menu-like interface. We explored how to develop an interface that could allow user interaction with a menu using the input and output modalities of the glasses.

We knew that, by using Bose Frames, the interface output would consist of the sonification (i.e. the "the technique of rendering sound in response to data and interactions", Hermann et al. (2011)) of the common UI element/menu item. We decided to leave exploring the possible input methods for a future study, as they could consist of the various interaction modalities that the Frames could offer.

In particular, we set ourselves the following two sub-objectives:

Firstly, we aimed to investigate the viability of such an interface in a mobile context. We endeavoured to understand and report the effort needed to operate it, how usable it could be, and how well would the UI perform, with special attention to investigating these properties in a mobile environment, such as while walking outdoors. We thus explored whether head tracking was an effective input method in a busy, mobile context and whether users could consistently and precisely interact as they intended with the menu.

Secondly, we investigated what the optimal layout for a menu in such an interface should be. We aimed to find the strengths and weaknesses of each layout in terms of usability and workload.

2 Background

In this chapter, we described similar studies carried out by several authors or similar interfaces proposed and their designs. We will report on their approach to the problem and how they implemented similar or relevant features. Furthermore, we will examine several studies that influenced our decisions over the course of the project.

2.1 The hardware: re-purposing a device

Smartphones and smartwatches, two of the most common examples of mobile computing in everyday life, offer similar interaction methods. A visual display paired with mechanical buttons or with a touchpad or panel, the latter usually combined with the display into a single touchscreen. Though this ubiquitous UI is often expanded with conversational interfaces and other contextually-used interfaces, smartphones have been under the focus of research due to the difficulty in implementing novel input and output techniques (Henze and Rukzio (2015)).

Furthermore, a surge in technological capabilities has sparked the rise of mobile devices with novel interaction methods such as Apple 3d touch Apple (2022) and the Google Squeeze feature Google (2022). Earphones and headsets in particular have been ubiquitous for nearly 60 years (Mark McGill (2020)) and are now sporting hardware that allows them to be used as much more than just earphones. Examples of these technological advancements are several Bose products, like the Bose NC700 Headphones Bose (2022b), which are transitioning from being simple output technologies for more complex devices (such as smartphones) to wearable interfaces which allow a range of functionalities. The aforementioned product, for example, allows users to accept or reject calls, interact with a linked phone's conversational interface and manage volume and music, all without having to interact with the smartphone itself.

Another product with similar functionalities and properties are the Bose Frames, which offer Open Ear Audio technology. In a study by McGill et al. (2020), the authors use the Frames to compare acoustic transparency with acoustic opacity and to investigate the effect of the former on personal audio experiences. The experiment was carried out both indoors and in a mobile context and focused on investigating perceived personal safety, attitudes towards acoustic transparency, and possible changes in usage likelihood, habits, and patterns. The authors showed how acoustic transparency not only appeared safer in mobile contexts but could offer a viable alternative way to consume a range of audio-based content. Among other reflections, the authors also consider how the development of an acoustic UI will require care to avoid overloading users with unsolicited information.

2.2 The interface

As the device chosen allows for both hands-on and hands-free input methods, we aimed to find the best way to leverage the capabilities of the hardware to create the simplest, most intuitive control modalities suitable for a mobile context. In this section, we report on the research done about the possible input modalities for the Bose Frames. Furthermore, we summarize the implementation of several eyes-free interfaces and how authors

Hands-on input methods Chaudhuri et al. (2019) introduces an eye-free calculator for touch-screens that target-free haptic gestures to solve arithmetic tasks. The gestures consisted of taps and directional swipes from one to three fingers, and the system state was communicated by minimal audio feedback. A within-subjects study with 8 participants compared traditional smartphone calculators with the implemented prototype. Participants were able to enter characters 40.5% faster than with a traditional calculator and made 52.2% fewer erroneous calculations. The experiment confirmed the utility of metaphor-based gestures in making novel input methods as intuitive as possible. Ronkainen et al. (2007) carried out a WWW-based survey about the social acceptability of short video clips showing different input methods in public spaces. The survey revealed the preferred interaction methods of a set of 41 participants to be, in order: tapping a smartphone through the pocket, tapping on the side of the device, shaking the phone. performed a usability study on tap input as an eyes-free interaction method. The authors followed the survey with a usability study about the tapping and double-tapping input methods, where 15 participants performed a set of tasks. Though the study did not investigate the intuitiveness of tapping, it showed several issues with the input method: participants interpreted the verb in several ways, ranging in terms of strength, position and number of taps. All these factors presented a challenge while detecting the input. Latency and lack of feedback also confused participants. The poor performance of single tapping led the authors to perform a new, similar study where only double-tapping was considered, which showed double-tapping as a useful and usable eyes-free input method to control a device, even when it's worn on the user.

Menu interface structure and implementation Another article that presented a hands-on interface reported about Foogue, an eyes-free audio interface for smartphones developed by Dicke et al. (2010). The authors proposed a UI for menu navigation, item selection, and window management in order to address the shortcomings of smartphone GUIs. A haptic gestural language was built to allow for a limited number of gesture elements. The authors implemented the menu in a 120° arc in front of the user, where spatialised sounds represented folders and menu items. The user could browse the row of objects by pointing the phone at the target item like a torch, the name of which will then be read to the user by an audio prompt.

Other articles described similarly structured audio interfaces, like the ones implemented by Sodnik et al. (2008) in a study that compared auditory versus visual interfaces while driving. The menus implemented ordered the items in a circle around the user's head, such that users scrolled through the options with a remote to navigate the menu. The item-specific prompts were either played simultaneously, leaving the user's "Cocktail party effect" to indicate the currently selected item, or singularly: only the currently selected item prompt was played. The authors measured mean task completion time, usability was analysed via QUIS questionnaire, and workload via NASA-TLX. Results saw acoustic menus being considered less demanding and more usable than visual ones, though slower.

Alternative, hands-free input methods to interact with audio menu techniques had been researched in the past twenty years with the rise of various fast, powerful, yet affordable technologies, ranging from accelerometers to 3D cameras. 2.5D menus encircling the user saw significant adoption in research, and a previous study by Brewster and Marentakis (2004) used it to compare various input methods in a different mobile context: standing still. The authors compared different input modalities by developing and evaluating a gesture-controlled audio-only user interface for mobile devices. The study arranges the interface menu items around the user. A 3D spatial audio system leverages the "Cocktail party" effect to output multiple feedback streams to the user while still allowing item recognition and localization. The authors investigated three different input methods to browse, either relying on head-direction, hand-direction or using a tablet to orient the sound sources. Similarly, the authors investigated three input methods to select an item: head nodding, a hand gesture resembling a mouse click, and a mechanical button on a tablet stylus. The input conditions were a combination of the above: browsing and selecting with the head, with the hand or with the tablet. The experiment carried out in the study tasked the participants to browse the interface and select the target sound. Deviation angle from target and effective selection angle was measured. The study found that the tablet ensured higher accuracy levels while browsing and selecting. However, participants considered it easier and more comfortable to use either head or hand-based input methods. As such, the paper assessed head-tracking as an effective mobile and eyes-free browsing method and hand gestures as a viable selection method, in each respective combination.

Head-orientation-tracking We wanted to consider head-tracking as a possible input method, and to understand whether and how to implement head movements as such, we made use of the work done (long before efforts to implement body movement appeared in Mobile HCI) by Pozzo et al. (1990) who studied head kinematics during several locomotor tasks. The authors analysed head angular displacement with a focus on head translation on the vertical axis and rotation in the sagittal plane. Results show that during the various motor conditions, horizontal displacement can reach up to 20°, with a maximum displacement of 10.5°, 11°, 17.5°, and 18° for the walking in place, free walking, running and hopping locomotor conditions respectively.

As technology advanced, research built on previous medical papers started evaluating head-orientation-tracking as an input method, both on stationary computing and mobile devices. LoPresti et al. (2000), for example, investigated how neck range of motion impacted the viability of computer head controls for people with or without disabilities. The authors determined the relationship between neck range of motion and performance. The study analysed Flexion/extension (turning the head up or down), axial rotation (turning the head left or right), and lateral bending (tilting the head left or right). For participants without a disability, the mean range of motion were, respectively, 118°, 137°, and 87°. For participants with disability, 59°, 85° and 36°. Finally, the authors found that vertical cursor movements are faster than horizontal or diagonal movements. Hansen et al. (2018) performed a Fitts' law evaluation of gaze-based and head-based interaction in head-mounted displays. The authors found the two input methods to be equally fast (head pointing at 2.47 bits/s and gaze pointing at 2.13 bits/s) but both slower than using the mouse (3.24 bits/s). They also found gaze to require a larger effective target width (at about 3° of the visual angle) than for head and mouse, both at 2.5°. However, participants scored gaze-based interaction as less physically demanding than head-based one.

Progress in head-orientation-based interfaces grew as the capabilities of affordable sensors were improved, and software that could leverage the powerful hardware that was built-in common smartphones became readily available. Roig-Maimó et al. (2015) developed a head-tracking interface for smartphones by using the phone's camera. The study evaluated its performance at position-select tasks with sitting participants. The screen of the device was divided into 15 regions and evaluated head-tracking as a cursor-like input for mobile devices by testing parameters like gain (the cursor speed) and target width. They found head-tracking to be an accurate input method for targets of 88 pixels and above, and with a gain of 1. However, there is little research about the viability of head-tracking as an input method in a mobile context, i.e. while the user is walking. In a further study, Roig-Maimó et al. (2016) builds on their previously developed interface, this time evaluating it on a tablet. Additionally, in this paper, the authors analyzed the effects of motor impairments on the performance of the interface by involving four participants with multiple sclerosis and twelve able-bodied participants. The research was divided into two studies. In the first one, the participants uncovered the pieces of a puzzle by moving one's head, thus moving the cursor over the tile. The smallest tile size (44 pixels) proved difficult for two motor-impaired participants, though all other participants completed every task. In the second study, the participants selected the tiles in the same manner as in the first study, but with an added 1000 ms dwell-time criterion. The study overall found that tiles should be 76 pixels or above to ensure accessibility and should be placed toward the centre of the screen. Despite the growing number of studies on head-orientation-tracking, there is little research available on the input method in non-stationary mobile contexts, or as an input method for audio-only interfaces.

Finally, a similar and related input method, head-tilting, has indeed been studied in a mobile context by Crossan et al. (2009). The paper explored head-tilting as a hands-free interaction method for mobile interfaces. The authors ran a Fitts' Law style evaluation and tasked participants to select targets by moving a cursor via tilting their heads left or right. This motion was estimated using an accelerometer attached to a hat. Two different methods of moving the cursor were explored. In the first method, the position of the cursor is proportional to the tilt angle. In the second method, the cursor speed is proportional to the tilt angle and the direction reflective of the side the head is tilted. The users performed the task in both static and mobile contexts, where they were asked to follow a figure-eight path. The authors found that participants were significantly slower and less accurate at hitting targets while walking due to signal noise from the walking motion. In particular, while standing still, the position cursor control allowed for the best performance, reliably allowing users to hit 30 pixels targets with a mean accuracy of nearly 90%. In a mobile context, the velocity cursor control on the other hand achieved better performance, with a mean accuracy of 74% compared to the 61% from the position control. Overall, this paper demonstrated that head-tilt can successfully be used as a pointing-based input method in a mobile setting.

3 Requirements

In this chapter, we described the requirements of the product development stage and the project overall. Although the aims of this project were outlined in Chapter 1.2, in this chapter the aims of the project are translated into more specific and in-depth requirements. These were determined following the MoSCoW method, an Agile methodology in which different requirements are labelled and ordered based on their relative importance for the completion of the product. The letters stand for:

- **M**ust **H**ave
- **S**hould **H**ave
- **C**ould **H**ave
- **W**ould like to **H**ave

3.1 Main requirements

To develop a functional User Interface and be able to not only evaluate it but carry out our research aims, we discerned the requirements that concerned the product from those that concerned the evaluation. The former can be further divided into functional and non-functional requirements.

3.1.1 Functional Requirements

The functional requirements of a project determine what structural feature a software should have to manipulate input data and yield output data. This means they describe what the software should do.

The Must-Have requirements followed directly from the project specification and the early discussions about the topic of the project. Furthermore, from these requirements follow the main aims of the evaluation of the User Interface.

1. MH_1: *Users must have the ability to interact with the interface via two separate input methods.* From this requirement follow the latter two.
   Note: It is not a Must Have priority which particular input method we use, as long as the user can perform the two actions.
2. MH_2: *Users must be able to browse the items of a menu-like interface.*
3. MH_3: *Users must be able to select menu items from a menu-like interface.*
4. MH_4: *Users must be able to gather information about the system status while browsing and selecting items.* As the only output method of out device is audio, users must understand what menu items are available and what item are they selecting solely via sound-based feedback.
5. SH_1: *Users should receive feedback about the system state for browsing non-selectable items or for browsing interruption.*

3.1.2 Non-Functional Requirements

The non-functional requirements of a project determine the required performance of the software. In the case of the User Interface, these mainly focused on usability.

1. $\boldsymbol{MH_5}$: *Users must be able to use the interface while in a mobile environment.* While such a context can mean many different situations, we reduced and generalised such aim to two main contexts: walking and standing still.
2. $\boldsymbol{SH_2}$: *Users should be able to reach and select all menu items in the user interface in one browsing action.* This is not a must have as, in the case of many menu items, a scrolling-like feature would enable browsing of hidden menu items (i.e. non-readily selectable).
3. $\boldsymbol{CH_1}$: *Users could be able to easily and continuously reach the user interface in a mobile environment.*
4. $\boldsymbol{CH_2}$: *Users must receive audio output that augments environmental sounds.*
5. $\boldsymbol{WH_2}$: *Users would receive feedback about the status state in a timely manner.* This requirement refers to addressing Bluetooth and software dependent latency.

3.2 Evaluation requirements

Special requirements were formulated ad hoc to be able to evaluate the interface:

1. $\boldsymbol{SH_3}$: *The system should be able to log input data onto a CSV file*
2. $\boldsymbol{WH_3}$: *The system should be able to log output data onto a CSV file*

4 Design

Our design process followed the familiar Product design concept of a *design funnel*. It is characterised in an initial burst in the number of possible designs, which iteratively expands with the introduction of new ideas and shrinks with the selection of a few of those, each time. This process usually terminates with a few remaining designs.

In our case, however, the reduction of the design space to a focal point cannot be achieved, as our project makes up the middle stage, or even the start, of a design funnel. It follows that the "final design" we developed consists of three different designs and, more importantly, it includes the outlook to design, prototype and evaluate many more.

In this chapter we described the design process and the final design used for the interface. In the section 4.1 below, we explain the design development choices we made and their motivations. In section 4.2, we describe in detail the final design for the interface.

4.1 Design funnel

We started with the design process knowing that we wanted to create an interface for an auditory headset, which allowed us to reduce the number of possible interaction methods available to those commonly found on devices like Bose Frames. Furthermore, one of the main aims of the project was to make an audio UI that could offer an alternative to a dialogue system. We thus ruled out speech as a possible input method. Our pool of possible input methods was thus left to three systems: head-orientation-based, gesture-based and a mix-system.

Additionally, we wanted potential users to interact with the interface in a mobile context, i.e. while walking or standing in different locations. We, therefore, decided to use an egocentric design, such that the interface would follow the user's position in space.

However, the relatively few options in input methods did not reflect the multitude of possible menu types. In light of this, we decided to make it an aim of the project to find an optimal menu layout for the audio interface. As per requirements MH_2 and MH_3 we wanted users to be able to browse and select menu items (or buttons) out of a set. The menu envisaged consisted of several items (just like apps on a smartphone home screen) that could be arranged in various layouts. We settled on four main ones to evaluate: doughnut, pie, horizontal, and vertical.

During the first design iteration, we performed a simple Wizard-of-Oz evaluation. We tested the layouts both while sitting and while walking. In particular, we tested the doughnut layout, which consisted of menu items completely surrounding the user. We noticed that a user would have to turn around by 180 degrees to select some buttons, which would prove difficult while standing, let alone while walking. We concluded that the doughnut design would prove too cumbersome a system to use in a mobile context. We thus narrowed down our layouts to three - Horizontal, Pie and Vertical - which we will explain in the following section.

During a second design iteration, we looked at the input methods. We opted for a multimodal system, where input would involve both head-tracking and gestures. We did not consider the other two systems, gesture-based and head-orientation-based, unfit for the interface. We discarded them to make full use of the input methods that Bose Frames could offer. In future work, however, we aim to investigate and compare the above and various other interaction modalities.

Similarly, in this project, we decided to postpone investigating the optimal menu item number for future research. We opted for four buttons as it would allow for the easiest implementation. We aim to study how usable a UI sporting three or five buttons can be and whether there may be situational uses for such layouts.

4.2 Final design

The design we settled on consisted of an egocentric interface, with a spatial, transparent audio output system and a multi-modal input system. The latter was structured as follows: head-orientation allows users to browse through menu items, and a touch gesture allows users to select an item.

4.2.1 Structure

We imagined the interface in development as a flat panel or plane, that sat in front of the user. On the panel, just like a screen, should lie the menu items and UI objects that a user can interact with.

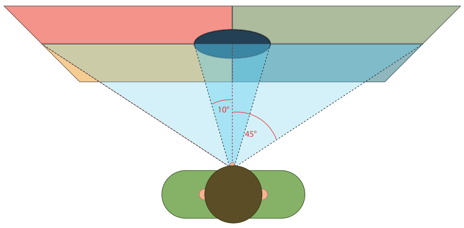
In their study, LoPresti et al. (2000) describe the neck ranges of motion of participants with and without disabilities. In particular, they found that users with disabilities would have on average a range of almost 60° and 85° when turning their heads vertically and horizontally, respectively. Placing buttons farther than at a 60° displacement from the base head orientation would thus eliminate a demographic from the potential user-base. This posed a limit as to how big and spread out we could make our interface. However, we knew that the button sizes would depend on the size of the UI. An interface too small and a user would not be able to control their head direction well enough to interact successfully while walking: when the user's head orientation is less consistent.

To account for this, we decided that the interface height should allow the furthest menu items to be within 30° degrees up and down of the user's natural head orientation. On average, users with disabilities were found to have a greater neck range of motion when turning their heads left and right. Despite this, we decided to apply this same sizing to the width of the interface, such that sizing wouldn't prove to be a confounding factor when evaluating the different layouts. This meant that overall, in terms of the user's field of view, the panel spanned 90° horizontally and vertically (i.e. width and height, see 4.1).

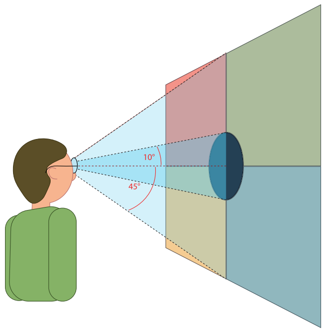
As mentioned in the section 4.1 above, we designed three possible menu layouts for the interface:

- **Horizontal** - This menu layout was build along a single dimension parallel to the floor. In the soundscape that would be reproduced by the audio output, this layout consisted in a row of buttons in front of the user. See 4.2a.
- **Pie** - This layout displayed the menu items around the user's central point of view. It functionally arranges the buttons in the shape of a circle placed in front of the user, each item sitting in one section. See 4.2b.
- **Vertical** - Similar to the horizontal layout, but differing in the orientation of the line the menu items are organised in. In this layout, the buttons are arranged in a vertical column in front of the user. See 4.2c.

We decided the size of the menu items would be directly correlated to the size of the interface. A layout with no intervals between buttons prevents users from miss-clicking those blank spaces. Furthermore, summarising Fitts' law, we knew that selectable UI elements should be as big in size as possible. This is true for head-orientation-tracking, as shown by Hansen et al. (2018).



(a) Width of the interface and the dead-zone in terms of the user's field of view.



(b) Height of the interface and the dead-zone in terms of the user's field of view.

Figure 4.1: *An infographic showing the sizing of the interface. As it is an audio interface, these measures only relate to what the user can select and hover over via the input methods, but the users would not be able to see or perceive the dimensions of the UI panel if not through the feedback sounds. In both (a) and (b) we show that the interface spans 90° of the user's field of view horizontally and vertically, respectively. Furthermore, we show the size of the dead-zone in the centre of the UI panel, which makes up 20° of the user's field of view in either dimension.*

Therefore we decided that the shape and size of the buttons should expand liberally wherever there is space on the interface. In the linear layouts (i.e. the *Horizontal* and *Vertical* ones), the menu items are placed along a single dimension and are centred at the same height or angle, respectively, as the user's head direction, as shown in Fig. 4.2a and 4.2c. However, the height or width of the buttons, in each respective layout, spans the whole interface.

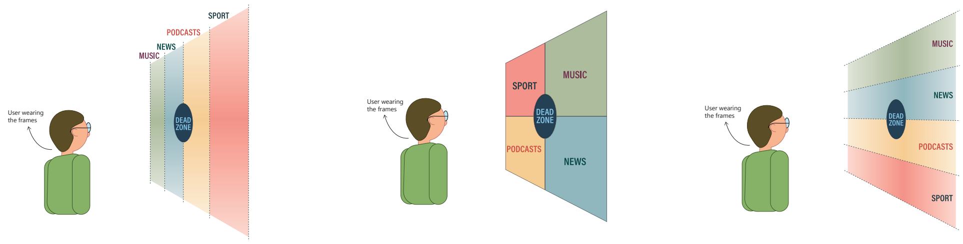
Finally, we decided to add a dead-zone at the centre of the UI panel. The dead-zone is a section of the UI where there is no audio feedback when a user hovers over with their head. From a pilot user study, we found that walking and looking straight was rendered frustrating by the amount of feedback a user would receive as soon as their head direction would turn slightly. Because of this, we included a round dead-zone of 20° in diameter - in terms of the user's field of view - as shown in 4.1. We chose the size of the dead-zone based on recommendations from the participants of the pilot study but, in future studies, it would be interesting to investigate an optimal size.

4.2.2 IO

A user can interact with the interface in two ways: by means of their head orientation and by means of a hand gesture. The direction of the user's head determines whether the user is facing the interface, and what UI object, if any, the user is facing towards. A hand gesture, a double tap, allows the user to select an item.

As the device only offers one output method, audio is the only way users are able to infer the system state. In his famous article, Nielsen (1994) describes ten heuristics that broadly define how a usable interface should behave. In Harley and Nielsen (2018), the authors expand on Nielsen's first heuristic: *visibility of system status*. Appropriate feedback allows users to be in control of their actions and of the system. The authors provide an example: in a GUI, a user knows the position of their cursor by means of a graphical and real-time representation of their mouse pointer. Another example consists in the necessity of feedback after selecting an item. Furthermore, the authors note that "even when users cannot see the effect of an action because the system does not have a screen [...], a minimum feedback that the command was [received] is essential."

We achieve this by means of two types of feedback sounds: verbal and non-verbal. These aim to



(a) Horizontal menu layout. *(b) Pie menu layout.* *(c) Vertical menu layout.*

Figure 4.2: *An infographic showing the three menu layout designs that decided to consider for the project. In (a) the user can turn their head left or right to reach the various buttons, which are arranged in a row. In (b) the user can face any corner of their view to reach the various buttons, which are arranged in a square. In (c) the user can turn their head up and down to reach the various buttons, which are arranged in a column.*

tackle the following system stata: item currently browsed by the user, item selected by the user, whether the user is facing an interactable UI object.

Verbal sounds are used to inform the user about what menu item they are currently browsing. As a user moves and starts to face a new menu item, a verbal feedback is given, naming the relevant button. Verbal feedback informs users of selectable items, but not given when a user hovers over the dead-zone or turns outside of the interface.

Different non-verbal sounds are used to inform the user about each of the three system state changes:

- The user is now facing a different item. This state change triggers both the verbal feedback described in the paragraph above, and an additional sound. This sound is short and non-obtrusive. It functions as a quick and simple feedback that complements the longer and slower verbal one.
- The user has selected an item. The corresponding feedback sound only informs users about selections (double-tapping while facing a selectable item), but not presses (double-tapping while facing the dead-zone or outside of the UI).
- The user has entered a non-selectable area, the dead-zone, or is facing outside the scope of the interface. The feedback consists in an error sound.

5 Implementation

In this chapter, we described the hardware and the software used to implement our product. Insights into the implementation stage aim to explain the decisions behind the use of certain resources and the development overall. Finally, we described the features implemented to be able to conduct the evaluation.

5.1 Hardware and Resources

The main component of this project is the Bose Frames. This device consists of the frame of a pair of glasses in which several pieces of hardware are embedded. Firstly, the Frames include a set of speakers that use Bose Open Ear Audio™ technology, which makes Bose Frames a transparent auditory headset. A set of sensors - an accelerometer and a gyroscope plus virtual capacitive touch sensors and rotation sensors - is easily accessible via the Bose AR SDK. Finally, Bose Frames have built-in Bluetooth technology which allows them to connect to any Bluetooth enabled device via the developer kit tools for searching, connecting and pulling sensor data. We used the Bose Frames from the Alto/Rondo series, as shown in Fig. 5.1.

After a first attempt at implementing the prototype as a Web app using Bose AR Web SDK, privately developed by Qattan (2022), we instead decided to make it using the Unity gaming engine/C# developed by Technologies (2022). This decision was motivated by the wider support, a larger amount of resources available online, and more expansive documentation. The Bose AR Unity SDK package Corporation (2019) provides support for unity-android app development that enables connection to the Frames Bluetooth and pulling and reading of sensor data.

The spatial audio was rendered via Resonance Audio, a library developed by Gorzel et al. (2019) with Google which supports Unity and allows the delivery of high fidelity and realistic 3D soundscapes.

The interface developed on Unity was then built onto an Android platform and ran on a Google Pixel 4. Running the application on a smartphone allowed us to evaluate the UI in a mobile context.

5.2 The interface

We developed our interface over two main classes, one which spawned the menu items and the other which managed user input and corresponding output.

The former was activated at the start of the app activation but could be called as needed from the interface to change the menu layout used. Indeed, we included on the smartphone app a drop-down menu that would allow us to change the layout during run time

The latter class managed the input data pulled from the Frames sensor data. As mentioned above, Bose Unity SDK provides a library that offers tools to read gesture sensor and gyroscope data. We implemented raycasting to keep track of the user's head orientation in relation to the UI panel. As the user faced an item, and the ray cast by the rotation matcher hits the menu items,

Figure 5.1: In our project we used the Bose Frames shown above. The speakers which make it an auditory headset are embedded in the legs of the frames, as are the sensors, battery and Bluetooth hardware. The lenses of the frames were removed for convenience.

a hovering-specific function is called to play feedback audio and report log data, as explained in the section below. The selection of menu items worked similarly and depended on gesture sensor data.

In the first stages of development, we considered the nodding gesture as the item selection input method. However, after testing the early prototypes, we recognised that nodding would interfere with head orientation tracking as browsing depended on it. The horizontal layout was particularly affected, as both browsing and selecting were then performed by turning one's head up and down, which caused a high degree of false-positive among selection calls. Instead, double-tapping proved to be a more stable and consistent input method during the walking tests, though we noted that it was not perfect either. Tapping resulted unreliable when a clear idea of the location of the gesture sensor on the leg of the Frames.

To convey feedback to the user, multiple audio files were stored and managed via an audio manager class and respective object. We recorded voice clips reading the menu items' names and used them for browsing feedback, and we used a package of royalty-free menu UI sounds for the non-verbal feedback.

At this stage in the implementation, the UI would spawn in an arbitrary position compared to the user's head orientation. We developed a toggle that would only show on the android app to recalibrate the UI panel to the current direction of the Frames.

5.3 Experiment-related features

To conduct the evaluation, we also developed features that would allow us to gather data from participant use through the app.

In the evaluation, we would ask participants to select items based on a set of instructions. Since participants would perform the tests in a mobile environment, e.g. while walking, we decided to automatize how we would deliver the instructions. . Every time a user selected a menu item, audio containing the instruction would play. Start and end-of-the-experiment audio feedback was also triggered by menu item selection. Implementing such a feature constituted a trade-off:

- On one hand, users would receive the same timely feedback about the next instruction, without having to rely on the researcher's or their own reflexes to reveal the next task. Implementing an automatic dispensation of the instructions avoided a possible confounding factor.
- On the other hand, receiving an instruction while browsing could cause its audio to overlap with the browsing feedback voice prompts. In the eventuality that both the task instruction and the feedback named different menu items at close time frames, participants could struggle to identify the current target menu item. This could potentially lead to a confounding factor that would link task success rate with the layout most susceptible to this flaw. We attempted to prevent this known flaw while briefing participants, as mentioned in Section 6.4.1

Furthermore, for every test we would log analytic data such as:

- **Event logs** - I.e. all data about user-induced system state changes, i.e. browsing, selecting, and end/start of the experiment. We logged the following event-info:
  - Event type
  - Time
  - The item the user is facing, if any
  - Whether the item is the current instruction target item
- **Head direction** - I.e. the user's head orientation per frame. The orientation is encoded in terms of the 3D coordinates the ray cast from the user's head collides with the UI panel.

6 Experiment design

In this chapter, we described the methodology applied to evaluate our product. We described the main objectives of the experiment, the user study design, and the experiment procedure. Finally, we offer our hypotheses for the results of the evaluation, and a pilot study conducted before the main user study.

6.1 Aims

As outlined in Chapter 1, the project aimed to investigate the implementation of menu-like interfaces for audio-only eyes-free mobile devices. As such, we evaluated the usability and workload of the User Interface proposed, we started investigating what the optimal menu layout for such an interface might look like and we aimed to perform an observational study on how users would interact with the novel interface.

To investigate the optimal menu layout, we decided to compare the three layouts developed on three factors: task completion time, accuracy and impact on walking speed. These measures add on to the scores for usability and workload by providing objective measures. The latter of the three, impact on walking speed, is an objective measure that is directly correlated with workload and offers an alternative to the subjective and self-reported NASA-TLX scores.

6.2 Conditions

One of the independent variables of the experiment derived directly from the aims defined above. Since we aimed to investigate the optimal menu layout among the ones developed, our first IV was "**Menu Layout**" and the levels were:

1. **Horizontal** - The menu layout consisting of the items and the dead-zone being organised along a single dimension, parallel to the ground, such that a user turned their head left or right to browse the interface.
2. **Pie** - the menu layout consisting of a square arrangement for the interface. The items are positioned in each corner of the square, all around the dead zone.
3. **Vertical** - The menu layout consisting of the items and the dead-zone being organised along a single dimension, orthogonal to the ground, such that a user turned their head up or down to browse the interface.

As our experiment aimed to evaluate a mobile interface, we decided that a necessary variable to evaluate over would be **User Posture**. The levels were:

1. **Standing** - Participants should interact with the interface while standing still in a mobile context.
2. **Walking** - Participants should interact with the interface while walking.

Our initial intention was to evaluate users across three levels of user posture: standing still, walking and sitting. However, the pilot study highlighted how time-consuming performing experiments for just the walking conditions were. Indeed, these could take up to eight minutes per task, which meant that for nine conditions an experiment with one participant could take up to more than an hour. As such, we decided to put aside the "Sitting" level for another future study. This allowed us to have a total of six conditions to perform our experiment over, rather than nine, reducing our estimated experiment duration to fifty minutes.

In summary, our conditions were:

1. $C_{S,H}$ - Participants stand still while testing the horizontal layout.
2. $C_{S,P}$ - Participants stand still while testing the pie layout.
3. $C_{S,V}$ - Participants stand still while testing the vertical layout.
4. $C_{W,H}$ - Participants walk while testing the horizontal layout.
5. $C_{W,P}$ - Participants walk while testing the pie layout.
6. $C_{W,V}$ - Participants walk while testing the vertical layout.

6.3 Design

The dependent variables measured followed directly from the aims of the experiment:

- **Accuracy rate** - The percentage of the double-taps that correctly selected the target menu item.

  Note: *Double-taps* differ from *selections* as they do not have to be performed while facing a selectable menu item, but instead can be done facing anywhere, on the UI panel or not. This means that double-taps that occured while the user was facing non-target menu items, the dead-zone, or outside of the bounds of the interface were all counted as misses.
- **Completion time** - The amount of time needed to perform all the tasks.
- **Decrease in preferred walking speed** - The speed loss experienced while performing the experiment in the walking conditions.

On top of these objective measures, we kept track of usability and workload qualitative scores. We asked participants to fill out SUS and NASA-TLX questionnaires after each condition. These rounded out performance data with participant input about the interface. Finally, a semi-structured interview at the end of the experiment gathered exploratory, qualitative data from the participants. In particular, it offered in-depth insight into participants' user experience and actions.

The experiment was within-subjects, as that allowed to find fewer participants to gather the same amount of data and to avoid confounding variables such as age or gender. However, this led to the issue of order effects. In particular, the experiment is susceptible to both learning effects and participant fatigue. To remedy this, we counterbalanced the condition order. The starting condition would be different for every participant, for each block of six participants.

Another factor that could introduce confounding factors is the experiment setting. Laboratories offer more controlled conditions to carry out experiments when compared to studies carried out 'in the wild.' However, such controlled environments are not as realistic. Brewster (2002) concluded that users could enter a lower amount of data in mobile contexts, reducing performance. We thus decided to test our mobile interface outdoors.

Similar to the studies of Zhao et al. (2007) and Brewster and Marentakis (2004), who used real-world menu items for their experiments, we also used menu items themes that users would find familiar and easily recallable. In particular, we used the categories: "Music," "News," "Podcasts," and "Sports." Furthermore, like Sodnik et al. (2008), we chose to arrange them in the same order for each condition.

Order consistency offered a trade-off between realism and the chance to encounter confounding factors. Randomised instructions could potentially lead to data that did not encompass all menu items equally. We thus randomised instructions such that participants would select each menu item two times. Each condition would therefore comprise eight tasks.

6.4 Methodology

The study comprised four stages:

1. Briefing,
2. Walking test and familiarization,
3. Main tasks and questionnaires
4. Final questionnaire and interview.

6.4.1 Briefing

Participants were instructed on the experiment aims, its approximate duration, and the need for participant involvement. We summarised the experiment in general terms and then described the User Interface. The instructions on how to interact with the UI included structure, input methods, output methods, and known flaws. We disclosed the latter to prevent common pitfalls: difficulty in double-tapping, audio-feedback overlap while browsing, and feedback delay. Furthermore, we informed participants about a known issue of the experiment, where the instruction audios can overlap with hover-feedback voice prompts, causing possible uncertainty when receiving an instruction. As the implementation of an automatic instruction dispenser offered a trade-off, we speculated that disclosing the known issue to participants would mitigate the possible confounding factor. These issues are discussed more in detail in Section 7.5.5.

Participants were also informed of the ethical considerations of the experiment. We made sure participants knew their data was anonymous, that they could withdraw without consequences, and that they could ask questions at any time.

6.4.2 Walking test and familiarization

Before the experiment had started, we asked participants to perform a walking test. However, they were not informed that the test aimed to measure their natural walking speed. Studies like Brito et al. (2000) show that informing the participants of the study measures primes participants, inducing a change in their performance.

They were asked to walk in a straight line for 14 meters. A two-meter mark and a twelve-meter mark delimited a ten-meter central section. We measured the time taken by each participant to travel the ten meters. We thus derived the natural walking speed, which functioned as a control measure.

We then allowed participants to familiarize themselves with the interface. For a maximum of five minutes, participants browsed and selected menu items from the menu layout that corresponded with their first condition.

6.4.3 Main task

Once familiarised with the interface, we carried out the main tests. All standing still conditions were structured as such: Participants would standstill in a pre-defined spot, looking towards the end of the road. They were then asked to select any item and follow the instructions given. They were greeted and informed that the experiment had begun. Then, a voice instructed the participants to press one of the four menu items. Once they selected an item - not necessarily the correct one - or even the dead zone, the voice gave the participant a new item to press. Once a user had completed eight instructions, the interface informed them that the experiment had

Figure 6.1: The researcher following a participant while they perform a walking condition set of tasks. Walking behind the participants during the evaluation stages allowed us at once to: keep the Frames Bluetooth-connected to the UI android app, measure the distance the participant walked in the duration of the experiment, and monitor on the phone the participant's interactions and behaviour. Furthermore, the researcher walked behind the participants to avoid social priming bias.

ended. Walking conditions would be structured similarly. Participants positioned themselves at the starting line. They were then asked to begin walking as they double-tapped to start the experiment. Once they had completed the eight tasks, they were asked to stop walking. We would then measure the distance walked. As the Bose frames were paired via Bluetooth to the Google Pixel phone, the author closely followed the participants during the tasks. Walking behind the participants, rather than next to them, was necessary to avoid walking-speed bias.

Participants completed this set of eight tasks for each condition, every time followed by the SUS and NASA-TLX questionnaires. Participants were allowed to take breaks between trials.

6.4.4 Final questionnaire and interview

Finally, we asked participants to fill out a demographic questionnaire and whether they would answer some questions while being recorded. In the demographic questionnaire, we gathered participant data such as age, gender and professional status.

The interview was semi-structured to allow participants to bring new discussion points to the conversation. The interview framework consisted of the following open-ended questions:

- What did you think of the hardware?
- What did you think of the interface and its layouts?
- What did you think of the input methods?
- What did you think of the output methods?
- What would you change?
- In what contexts, if any, would you see yourself, or others, using this interface?

6.5 Hypotheses

We formulated the following hypotheses:

1. ***Hypothesis*** $_1$: Participants will be able to select menu items significantly more accurately while standing still than while walking, due to the lack of head-orientation noise that stems from mobile contexts like walking. This entails that we foresee $C_{S,H}$, $C_{S,P}$ and $C_{S,V}$ to yield significantly higher accuracy scores.
2. ***Hypothesis*** $_2$: Participants will be able to select menu items significantly faster while standing still than while walking, due to the lower workload needed to stand still compared to walking. This entails that we foresee $C_{S,H}$, $C_{S,P}$ and $C_{S,V}$ to yield significantly lower duration scores.
3. ***Hypothesis*** $_3$: Participants will be able to rate the workload of the interface significantly higher while standing still than while walking, due to the obstacles present in a "in the wild" setting and higher amount of multitasking needed to walk compared to standing still. This entails that we foresee $C_{S,H}$, $C_{S,P}$ and $C_{S,V}$ to yield significantly higher workload scores.
4. ***Hypothesis*** $_4$: Participants will grade the *Pie* layout significantly better compared to the other two in terms of usability. This is due to the ability to browse the menu items in the *Pie* layout to be faced directly from the dead zone, without having to go through other items on the way there. This entails that we foresee $C_{W,P}$ and $C_{S,P}$ to yield better scores from the SUS questionnaire.

6.6 Pilot Study

We conducted an initial pilot study on two participants (one male, one female). The purpose of the pilot study was to start testing the interface on real participants and to test the experiment design.

As we decided the experiment would be performed in a mixed in-the-wild, we tested various locations on the University Campus to understand what conditions were optimal to conduct such a test. Two characteristics noted by the participants were that they believed a quiet and spacious environment to be important. Among the various locations tested, we decided that the initial stretch of Kelvinway, at the start of the university campus, was the best candidate. Kelvinway, Glasgow, is a straight, wide, well-maintained pedestrian street that crosses a park. This offers three benefits: it is quiet, it offers an easily navigable path without curves or bumps and it is usually not very crowded. Participants here did not have to worry about cars or bikes, and pedestrians offered a safe obstacle, appropriate for a mobile context.

Furthermore, the participants gave useful feedback about the interface itself. I had not yet implemented a dead zone at the centre of the menu. This meant that, while walking straight, participants would constantly hear the hovering feedback while looking straight ahead. This was annoying and overstimulating and was forcing users to walk straight while keeping their head at an angle, such that they would activate the feedback as little as possible. As such, as explained in chapter 4 we decided to add a dead zone to the centre of the interface. This dead-zone would provide no verbal audio feedback when hovered over and would allow users to look straight without accidentally interacting with the system.

The participants also noticed a lack of feedback when they would turn their heads outside of the bounds of the interface. This meant that when looking at the extremities of the UI, they would not know if they were facing a menu item or were out of bounds. We, therefore, added non-verbal feedback to denote such an action. Furthermore, we chose to add this feedback to when users would face the dead zone too, and as such the sound used would signify, in general, that a user was not facing a clickable menu item.

7 Results

In this chapter, we present the results of the evaluation we conducted. Furthermore, we discuss their significance, whether they attest to the validity of our hypotheses (Section 6.5), and how they relate to the wider literature.

7.1 Participants

Nineteen participants took part in the study, 10 male, 7 female, 1 genderfluid and 1 nonbinary/agender. All were recruited from the university community and volunteers. As our experiment was within-subjects, any multiple of the number of conditions, six, was acceptable. However, the data from one of the subjects was not usable, hence the extra one. The average age of the participants was 22 (mean: 22.5, standard deviation: 1.4), most belonging to the 18-24 age range and only one to the 25-34 one. Most of the subjects were university students, with only 1 participant being employed full-time. Of the participants, 4 work in engineering or computing fields.

All the participants had normal or corrected-to-normal hearing. Participants could use prescription glasses in addition to the HMD if they so needed, 2 did, both male.

7.2 Quantitative measures

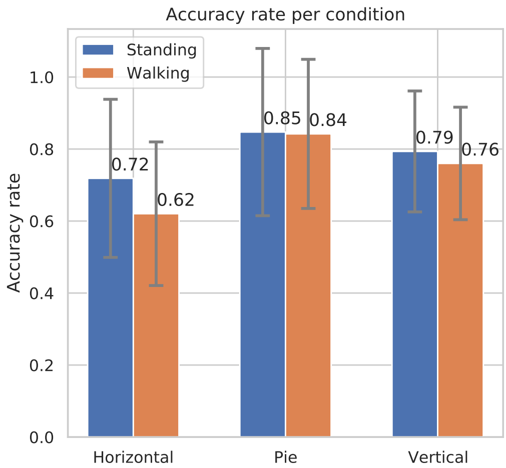
To obtain these statistical results we used a Two-Factor with Replication ANOVA test from the Analysis ToolPak extension pack for Excel. ANOVA was performed on *Posture*Layout* to assess the impact of these independent variables on accuracy, task duration, usability, workload and walking speed.

7.2.1 Accuracy

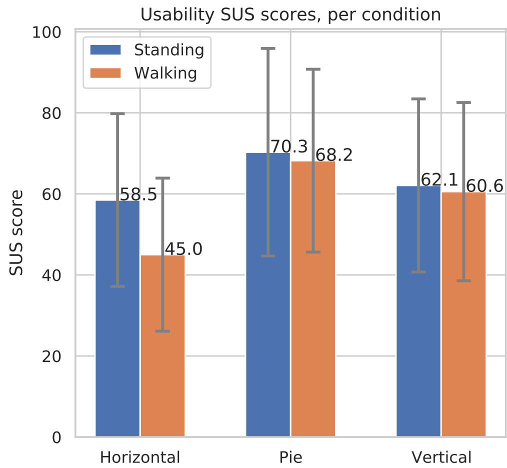
We defined accuracy as the number of double-taps that resulted in a correct selection, i.e. a selection of the target menu item requested by the task instruction. It is the ratio of the number of correct selections to the number of presses.

Among the layouts, the *Pie* layout performed the best, with a mean accuracy rate of 0.844 or 84.4%, as shown in Fig. 7.1a. The *Horizontal* layout performed the worst at 0.669 or 66.9%, more than 10 percentiles behind the *Vertical* layout and more than 15 behind *Pie*. Overall the layouts produced significantly different accuracy rates ($p < 0.005$). We found *Horizontal* to be significantly less accurate than both *Pie* ($p < 0.001$) and *Vertical* ($p < 0.05$), whereas results were not significantly different between *Pie* and *Vertical*. A complete summary of the accuracy rates can be found in Table 7.1.

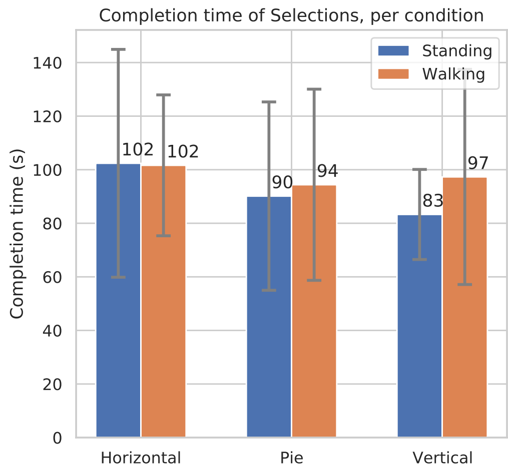
There was no significant difference in accuracy between the *Walking* conditions and the *Standing* ones. Therefore ***Hypothesis***$_1$ was not supported. This is surprising, as, in other user studies, such as the one conducted by Crossan et al. (2009) about head-tilting, walking conditions are reported to be less accurate.



(a) Accuracy



(b) Usability SUS scores



(c) workload NASA-TLX scores

Figure 7.1: *The above graphs show the performance of the interface in the six conditions. Plot (a) shows how accuracy differed among layouts and whilst standing or walking. The Horizontal layout was significantly less accurate than both the Pie and Vertical ones. Plot (b) shows how participants SUS ratings differed among layouts and whilst standing or walking. The Horizontal layout was rated significantly worse than the Pie one, whereas no significant difference in accuracy was found between Vertical and any of the other two. Plot (c) shows how participants TLX ratings differed among layouts and whilst standing or walking. Both layout and posture were significantly impacted, with walking and Horizontal performing the worst.*

Table 7.1: Summary of the mean and standard deviation of the ***accuracy rates*** *for each of the six conditions. The table shows how the Pie layout performed better than the other two, with little difference between the two Posture levels.*

	Horizontal		**Pie**		**Vertical**		Total	
	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Standing	0.7186	0.2196	0.8471	0.2323	0.7934	0.1680	0.7864	0.2113
Walking	0.6203	0.1995	0.8421	0.2071	0.7600	0.1563	0.7408	0.2071
Total	0.6694	0.2127	0.8446	0.2169	0.7767	0.1608		

Table 7.2: (Summary of the mean and standard deviation of ***task completion times*** *for the six conditions. Though no significant difference was found between conditions, we can notice horizontal to perform slightly worse.*

	Horizontal		**Pie**		**Vertical**		Total	
	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Standing	102.39	42.53	90.14	35.16	83.29	16.81	91.94	33.63
Walking	101.63	26.30	94.38	35.68	97.30	40.16	97.77	34.01
Total	102.01	34.85	92.26	34.98	90.29	31.16		

Most conditions reflected the same differences found among the layouts. The horizontal layout, however, improved of almost 10 percentiles from the 0.62 accuracy rate of $C_{W,H}$ to the 0.72 one of $C_{S,H}$. No interaction was found ($p > 0.05$).

7.2.2 Completion time

Overall, no correlation was found between either the *Layout* Independent Variable, the *Posture* one, or an interaction of the two ($p > 0.05$ for all three analyses). Although, as we can see on Table 7.2, results seem to indicate that the *Horizontal* layout and the *Walking* condition seem to cause a slower completion time, the results are not statistically significant ($p > 0.05$). Therefore, $\boldsymbol{Hypothesis_2}$ was not supported.

7.2.3 Usability

Overall, the various layouts received significantly different results ($p < 0.005$), with the *Pie* layout performing the best, yielding a mean SUS score of 69.2, the only one above average. The *Horizontal* layout performed overall the worst, with a mean SUS score of 51.7, almost 10 points behind *Vertical* at 61.3 and almost 20 behind *Pie*. Amid the layouts, we found a significant difference between the Horizontal and Pie layouts ($p < 0.005$) but not between Horizontal-Vertical or Pie-Vertical, both $p > 0.05$. A complete table with the mean SUS scores and their standard deviation can be found in Table 7.3. $\boldsymbol{Hypothesis_4}$ is therefore supported by the results.

No significant difference was found in usability between using the interface while standing still compared to using it while walking.

Condition $C_{S,P}$, in which users stood still while interacting with the *Pie* layout interface, performed the best among all others, with an above-average mean SUS score of 70.3. It was followed by condition $C_{W,P}$, which trailed behind by 2 units at 68.2. The worst scoring condition was $C_{W,H}$, with a SUS score of 45, more than 23 points below average and 25 below the best performing condition. No interaction was found ($p > 0.05$).

Table 7.3: Summary of the mean and standard deviation of ***SUS scores*** *for the six conditions. In green are the SUS scores of the conditions that scored above 68 points, i.e. above average. We can see how only the Pie layout achieved above average scores, both in the standing and in the walking Posture levels.*

	Horizontal		**Pie**		**Vertical**		Total	
	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Standing	58.47	21.31	70.28	25.60	62.08	21.35	63.61	22.96
Walking	45.00	18.88	68.19	22.55	60.55	21.98	57.92	22.97
Total	51.74	20.99	69.24	23.80	61.31	21.37		

Table 7.4: Summary of the mean and standard deviation of ***NASA-TLX scores*** *for the six conditions. Both the Pie and the Vertical layouts achieved better scores than the Horizontal one. Standing conditions also achieved a significantly better score than the Walking ones.*

	Horizontal		**Pie**		**Vertical**		Total	
	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Standing	49.02	19.74	33.87	20.86	43.39	17.24	42.09	19.99
Walking	62.39	13.48	48.04	17.46	49.26	16.96	53.23	17.07
Total	55.70	17.99	40.95	20.27	46.32	17.11		

7.2.4 Workload

Overall, *layout* had a significant impact on NASA-TLX scores ($p < 0.005$). The *Horizontal* layout caused the highest workload, with an average score of 55.70, more than 5 points than the *Vertical* layout (at 46.32, $p < 0.05$) and almost 15 more than the *Pie* one (at 40.95, $p < 0.005$), both differences being significant. No significant difference was found between the *Pie* and *Vertical* layouts ($p > 0.05$).

As expected, NASA-TLX scores differed significantly between *Standing* and *Walking* conditions ($p < 0.005$), with the former one scoring on average 42.09, 10 points less than the latter ones.

Overall, Condition $C_{S,P}$ received the best scores, averaging at 33.87, almost 10 points lower than the second-best condition, $C_{S,V}$, at 43.39, and almost half the mean score of Condition $C_{W,H}$ at 62.39, the condition with the worst result. A full rundown of the scores can be found in Table 7.4.

7.2.5 Walking speed

Walking speed was found to slow down significantly while using the interface ($p < 0.005$), with a mean decrease of 22.7% ($St.Dev. = 16.5$). However, Despite the predictable and significant difference that *Layout* entailed in the NASA-TLX scores, no significant impact on walking speed was found among any of the layouts ($p > 0.05$). This is highly surprising, and we discussed the topic in more detail in Section 7.5.3.

7.3 Qualitative measures

We performed an open coding analysis of the transcribed semi-structured interviews at the end of each experiment, as described by Corbin and Strauss (1990). Participants provided comments, ideas and critiques which we grouped by the topics described in Section 6.4.4 and be the topics brought up most frequently by participants. Although the number of themes identified was too high to report them all, we present the most frequent and significant ones.

Preferred layout The *Pie* layout obtained the overwhelming majority of votes when asked about the preferred layout of those experimented upon, except for P7, P14 and P16. P1 described how the above layout is particularly more appropriate for walking contexts, as it permits users to turn their heads less to reach all the menu items than the other two layouts, thus allowing them to more easily look ahead while walking.

Both participant P8, P11, and P14 mentioned they preferred different layouts in the different *Posture* conditions: *Horizontal* while standing and *Pie* while walking. Both P14 and P17 mentioned that they considered being able to customize the layout to the user's preference an important feature.

Scrolling-induced audio overlap P1, P3, P4, P5, P6, P8, P12, P13 all mentioned that, in the linear layouts, the act of "scrolling" needed to select the furthest menu items was clunky, i.e. having to browse over non-target items on the way to face the target one. Since browsing is followed by audio feedback about the menu items that the user faced, the act of scrolling creates an overlap in different pieces of verbal feedback.

Audio Output Both participants P3, P13, P15 and P17 felt that the verbal audio was triggered too often in the current design of the interface. P13 suspected the length of the verbal feedback sounds was the reason for the overlap and thus frustration. P17 believes that in everyday contexts, users would become familiar with the interface after little time, and would be able to de-activate verbal audio feedback if desired, especially in mobile contexts where: "*There would have to be a way to turn off the voice prompts, because, when you're looking around you don't want to hear voice prompts thrown at you everywhere.*

Double-tapping Input P2 spoke favourably of the input method, saying that they *"felt kind of cool"*. On the other hand, many other participants (P3, P4, P5, P6, P8, P9, P11, P14, P15, P16) found it unresponsive, inconsistent and frustrating, and lamented how *"...it was a bit clunky to get the tap to register"* (P3). Participants P11 and P13 in particular noticed they were struggling to double-tap due to the Frames not being fitted to their head size. We found this comment to be particularly important and we discussed it further in Section 7.5.2. However, some participants (P3, P5, P14) that criticised the method also mentioned how they liked the concept, and the execution was something that needed to be ironed out.

Safety concerns P4, P6, P8, P10, P12 found themselves to be concerned about the safety of using the interface in a busier environment, such as a street, or while biking or while driving. P6 stated that *"I think your vision also gets impaired because you are focusing so much on your hearing aspect that you can't concentrate on your surroundings as much, which for instance if you're crossing a road might be dangerous"*.

P2, however, commented that after sufficient familiarization with the system, they would imagine such an eyes-free interface to allow users to be more aware of their environments, thus being more secure.

Potential contexts of use P4, P7, P9, P11 stated that they would only use such an interface in a stationary environment, like standing still or sitting. P10 mentioned they would imagine this interface to be useful for *"Some worker in a manufacturing line, when they are standing and have their hands full"*. P7, P8, P10, P17, P18 mentioned they would find the product useful in an everyday context. P5, P13 and P16, on the other hand, stated that they would not consider the interface very useful on a day-to-day basis, but that in more particular contexts they might.

Dead-zone Participants P1, P6, P9 and P16 thought the dead zone was too small and found that a dead zone of such a size was not big enough to account for the head bobbing that happens while walking, causing frustration.

Social Acceptability P10 and P11 commented that they felt they were being stared at because of the Frames and their interaction methods. In particular, the act of turning their head toward non-visible menu items was received as socially unacceptable. P18 in particular identified this problem for the *Horizontal* layout. P13 stated that they did not feel like they looked silly while using them, and felt comfortable in a social setting. They also stated, however, that, for the *Horizontal* and the *Vertical* layouts they *"...felt more uncomfortable in a social sense looking left or right more than [they] did, looking up and down."* P17 commented that this interface felt more socially acceptable than voice interfaces.

7.4 Observational research

To understand participant head movements and directions when interacting with our interface, we logged the coordinates on the UI panel users faced per frame.

In the *Horizontal* layout, even though participants could select items at any vertical angle within the 90° (i.e. 45° up and down natural angle) height of the UI panel, most selections happened within the 30° range (i.e. 15° up and down natural angle). Similarly, for the *Vertical* layout, despite ample space users mostly remained within 15° left or right of their natural angle looking straight ahead. In Fig. 7.2a and Fig 7.2b we can see a heatmap of the collective user head direction in relation to the UI panel. It is important to note that head orientations that did not face the interface were not recorded.

In the *Pie* layout, it is interesting to see how participants do indeed seem to be able to, as mentioned in the interviews, avoid unnecessary browsing and going directly from one target menu item, to the dead-zone, to the next target menu item, as we see in Fig. 7.2c. Furthermore, users seemed to avoid the upper-central section of the UI panel more so than the lower-central, centre-left or centre-right sections. We can attribute a possible reason for this: turning their heads up while walking prevented users from clearly seeing where they were walking, whereas facing down still allowed easy enough navigation of the street.

Finally, in all layouts, the dead zone was consistently and prominently used as an anchor point to go to between interactions. As shown in all the graphs of Fig. 7.2, the dead-zone was the area with the highest counts of head orientation coordinates, meaning that users were facing it the most of any other section of the interface.

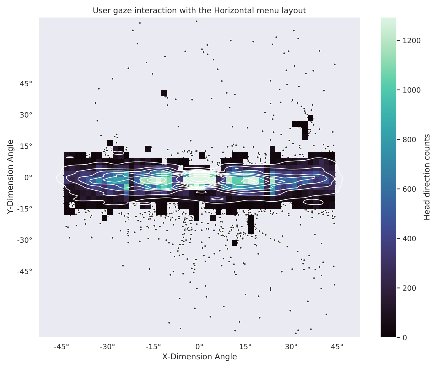
7.5 Discussion

The evaluation showed this interface has potential for further development and adoption as a viable, mobile, eyes-free user interface.

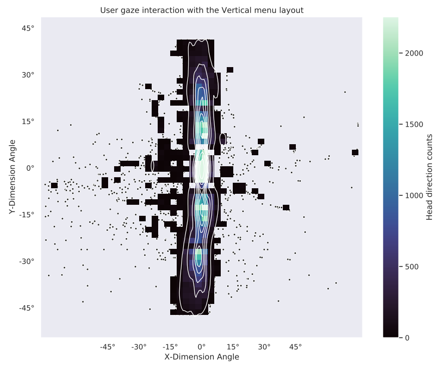
7.5.1 Interface structure

The results produced coherent and useful evidence for the assessment of the three layouts. Three of our quantitative measure studies, two subjective and one objective, as well as our qualitative research agreed on the viability of the *Pie* layout and its superiority over the *Horizontal* one. The *Vertical* layout received mixed results. Due to the results, we recommend avoiding using a horizontal layout for wearable interfaces that depend on head-orientation-tracking input.

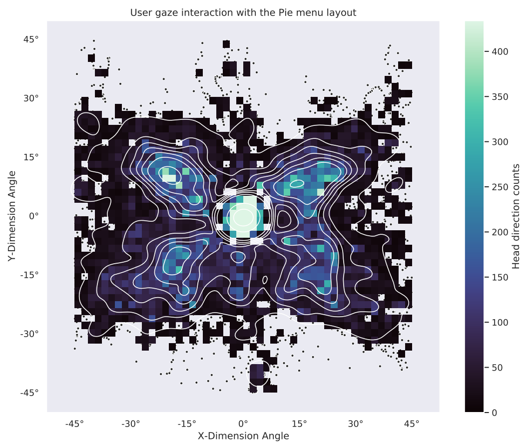
The clearest outcome is the apparent inferiority of the *Horizontal* layout compared to the other two. In fact, it was shown to be significantly worse than the other layouts both in terms of *Accuracy* and *Workload*. Furthermore, it was shown to be significantly less usable than the *Pie* layout according to the subjective SUS scores. A possible reason for its worse results has to do with social acceptability. As mentioned in the qualitative research, the act of turning one's head left or right, while in a public space, can be seen as more socially improper than looking up or down. Another possible reason takes into consideration the head-bobbing that occurs while walking. A study by Pozzo et al. (1990) confirmed that head rotation displacement along the sagittal (i.e. horizontal) plane could reach amplitudes of 20°.



(a) *User head orientation input heatmap using the Horizontal menu layout.*



(b) *User head orientation input heatmap using the Vertical menu layout.*



(c) *User head orientation input heatmap using the Pie menu layout.*

Figure 7.2: *The graphs show the coordinates on the UI panel that users faced the most, i.e. what parts of the interface users would most frequently face. A higher head-orientation count corresponds to a higher directional frequency. The coordinates of the user head orientation in relation to the UI panel are presented in terms of the vertical and horizontal angle displacement between the shown direction and the "base" direction, i.e. when looking straight ahead. The (0°, 0°) corresponds with the user looking forward and at horizon level. Dots represent single head orientation data points. Squares represent areas where more than 1% of the head directions were found. The colour of the squares indicates the count of head orientations. The lines represent a kernel density estimate (KDE) visualisation of the distribution of observations.*

The *Pie* layout obtained good results in terms of *Accuracy*, *Usability* and *Workload*. Its results did not differ significantly enough to obtain a clear supremacy over the *Vertical* layout, though its measures and scores *are* are consistently better. This is reflected in its nearly unanimous popularity among participants as shown by the qualitative research. P13 mentioned how a possible reason for this layout being so usable was its similarity to the same familiar grid-like layout used for smartphone home screens.

Some participants mentioned how they preferred different layouts in different conditions, though this was not reflected in the interaction p-values for any of the measures calculated. However, we should highlight that the set of layouts implemented in our experiment is not exhaustive, and as such different layouts that were not considered may perform better in specific conditions.

Given our observational study about the head-orientation-tracking interactions, we can recognise potential areas of improvement for both the linear layouts. We can indeed assess that users did not tend to use the extra space given to select the ribbon-like buttons of the *Horizontal* and *Vertical* layouts, as explained in Section 4.2.1. Such a space could potentially be used for extra or utility buttons (e.g. "Undo" or "Go back" commands). Furthermore, as shown in Fig. 7.2c, users tended to avoid the upper-central area of the interface more so than other liminal areas. This section could also be a candidate area for utility features.

The Dead-zone was a feature that was added as a consequence of the pilot study, where participants lamented the high amounts of feedback while looking straight ahead. As shown from the observational research, it was the area that was faced the most during the experiments. We attribute this to two possible and non-exclusive reasons:

- The dead-zone being in the centre of the UI panel also corresponded with the users looking straight ahead, which is particularly useful while walking.
- The dead-zone functioned as a section of the user's head-orientation space where they could not receive any feedback. As reflected in the interviews, participants found frustrating the high amount of verbal feedback that browsing entailed, thus returning to the feedback-less centre as a "safe zone".

  Note: However, the latter motivation can not be complete. Users could potentially keep facing a particular section and still not receive any feedback before moving to face the next target item. The common trend of returning to the centre can be then attributed to the head-position corresponding with facing the dead zone being the most comfortable while walking or standing still.

Additionally, four participants mentioned they would prefer the dead zone to be bigger, a particular comment that was consistently linked to the over-stimulation caused by the excessive amount of feedback. Further studies should look at the optimal dead-zone size that could, at the same time, allow users to not have to turn their heads too much to select menu items and still feel like they can move and not involuntarily trigger any audio feedback.

7.5.2 Interface Input and Output

The measure of accuracy revealed an interesting flaw in the interface: users are not able to select items very accurately. With *Pie* yielding the best results with a mean accuracy of 84.4%, most participants would miss-click at least once in five clicks using the *Vertical* layout and almost once every three clicks while using the *Horizontal* one. Despite these results, only *Pie* and *Horizontal* were found to differ significantly. We offer several possible explanations for this:

- Our implementation of the automatic instruction dispenser caused a confounding variable. As commented by eight participants, the linear layouts suffered from the way the hovering feedback prompts were implemented. In these layouts, if participants immediately returned to the dead-zone after selecting one of the two farthest items they would head an overlap of the instruction and hovering-feedback audios.

Caveat: This explanation, however, does not explain why the *Pie* layout was not found to be significantly more accurate than *Vertical* one. Although no participant brought up or criticised this known issue in the interviews, even when prompted, we cannot reject the possibility of such an implementation causing a confounding factor.

- As mentioned in the Section above, the natural head bobbing that occurs while walking described by Pozzo et al. (1990) might have particularly affected the performance of the *Horizontal* layout.
- Two participants commented during the interviews how they found double-tapping particularly difficult and attributed the cause to the size of the Frames used in the experiment. The Frames did not fit them correctly. Furthermore, two other participants wore prescription glasses under the Bose Frames during the experiment. We suspect these aspects might be confounding factors that affected the performance of the *Vertical* layout, which required participants to look down to select two of the four items, thus making the frames looser on the user's head.
- Finally, ten participants considered the double-tapping input to be inconsistent and unresponsive. This could be related to the size of the Frames but could also be a fault of the hardware itself. Bluetooth latency and the positioning of the optimal spot to double-tap both could have affected the accuracy ratios of all three layouts.

Several participants critiqued the implementation of the verbal audio feedback, particularly when discussing browsing in the *Horizontal* and *Vertical* layouts. The overlap in the menu-item descriptions while browsing seemed clunky and frustrating. This point of view seems to be reflected in the SUS questionnaire scores, where the *Pie* layout, which does not suffer from this implementation flaw, gained consistently higher scores. Furthermore, such an overlap might prove particularly frustrating for users on the autism spectrum or with other neurodivergent traits: the abundance of stimuli makes the interface *more difficult* to use, rather than helping. One possible way to allow users to receive feedback about the menu item they are facing without providing too much feedback would involve:

1. activating the non-verbal audio feedback as soon as the state change happens,
2. and to start the verbal one only after the participant had "hovered" over a button and held their head orientation there for a set amount of time.

This would allow scrolling without having pieces of audio feedback overlap, while still offering such an important type of feedback.

7.5.3 Postures

Standing and walking conditions did not differ significantly in many measures. Accuracy, completion time and usability all were unaffected. We can thus assume that our interface is fit for being used in both mobile contexts. Workload, on the other hand, was shown to be significantly more while walking. Furthermore, these results also reflect comments gathered in the qualitative research, where four participants states that they would only use the interface in a stationary context, and not in a mobile one. This is expected, since "walking while performing a secondary task [...] increases cognitive workload" (Hoang et al. (2020)), and does not necessarily imply that our interface could only be usable while standing still. On the contrary, *Walking* conditions yielded above average results in terms of SUS scores and did not significantly perform worse in terms of usability and accuracy, therefore indicating that the product is viable in both stationary and walking contexts.

An unexpected result saw the walking speed results not mirroring the significant difference that *Layout* entailed for NASA-TLX scores. This is surprising because walking speed and workload both indicate cognitive workload. We attribute two possible meanings to this:

- Data was not collected carefully enough. Noise was introduced whilst measuring natural walking speed, the distances walked by participants during their *Walking* conditions, or both.
- All three layouts did not differ in workload enough to produce a significant effect on walking speed.

7.5.4 Other

In this section, we discussed topics that were not the main aims of the evaluation but that were of high interest for the product itself and possible future work about the interface.

Five participants independently brought up the topic of social acceptability, without being prompted. Four commented negatively about the impact of the interface and the frames feeling uncomfortable in a social environment, while one commented positively. Social acceptability holds an important influence on the success rate of novel pieces of technology or interaction methods. Studies like the one by Kelly and Gilbert (2016) aim to develop ways to measure such a factor and to allow developers to evaluate novel devices and interfaces. Future work on our product might look at these measures to gauge how socially acceptable is our interface. Another interesting piece of feedback was brought by P17, who positively commented about the social acceptability of such an interface when compared to voice commands, as it provided a silent input method.

A different set of five participants additionally mentioned, again independently and without being prompted, that they do not perceive such an interface to be safe to use while walking in a street, while biking or driving, as it requires the user to turn their head. It is interesting to notice that although the interface was eyes-free, people do still recognize the danger of using the interface in high-stake mobile contexts.

7.5.5 Limitations

The evaluation overall identified several areas of improvement for the product. The design and implementation of the audio output arose as an important feature deserving revision. As mentioned by McGill et al. (2020), the design of acoustic interfaces requires extensive research and careful design to avoid user over-stimulation. Research to investigate the viability of delayed verbal hovering feedback when compared to immediate feedback would therefore be invaluable.

Furthermore, our decision regarding the relative size of the UI panel in relation to the user's head rotation implied a trade-off between comfortable neck ranges of motion as described by LoPresti et al. (2000) and Fitts' law (Hansen et al. (2018)), and as such, we cannot guarantee it to be viable for all demographics: users with disabilities that affect their neck motion range might benefit from a customisable UI panel size.

7.5.6 Experimental limitations

A known flaw in the implementation of the experiment-related features was the possibility of overlap of audio feedback and the instructions audio. The issue, although tentatively tackled by debriefing participants of the user study, might have introduced a confounding variable in the evaluation of the linear layouts. Further or related research should investigate alternative ways to combine audio feedback with task instructions.

A second issue arose during our user study, where two participants highlighted that the device used for the experiment was not fitted to their heads, causing frustration with the input methods and possibly introducing a confounding variable.

8 Conclusion

Mobile computing is evolving, as the technology that makes it possible becomes more and more powerful. The widespread availability of ever-improving hardware like accelerometers and gyroscopes and the refinement of sensor-data analysing software both contribute to fostering growth in mobile human-computer-interaction and wearable computing overall. With the introduction of head-mounted Displays like Oculus Rift in the market and their growing presence in everyday households, we foresee that the boundaries of their social acceptability and usage will expand to include public and working spaces. Furthermore, conversational interfaces currently form the bulk of eyes-free computing, most commonly appearing in the shape of voice assistants on smartphones, laptops, or smart-home appliances. Providing an alternative to conversational interfaces could prove useful and successful in contexts where voice input is not possible or proper.

In this report, we discussed the motivation and aims of our project, followed by a review of related literature. We explain how our product was designed, implemented and evaluated, and the results of our study. In this final chapter, we provide a summary of the product and the findings of its evaluation, with a recap of the limitations of this project as well as the future work to be carried out in order to further investigate this promising novel interface.

8.1 Summary

In this project, we leveraged the capabilities of the Bose Frames to design and develop a novel, audio-only mobile user interface, focusing in particular on ways to implement a menu. A user study conducted suggests that our UI provides a usable and viable menu technique.

Menu items are placed in 3-dimensional space in front of the user according to three possible layouts. Users can browse through them by orienting their heads in space: facing toward a menu item corresponds to hovering over a button with a mouse pointer. Selection is implemented via a double-tapping gesture on the leg of the Frames while facing the desired item. Users receive feedback about the system state solely by audio, which leverages a set of verbal and non-verbal sounds to indicate browsing and selection actions.

We designed three possible ways to display the items: a horizontal row, a vertical column, and in the shape of a square. In a user study with 18 participants, we evaluated this interface to investigate its usability, workload and performance across different layouts and mobility conditions (*Standing* and *Walking*). We found that users were significantly less accurate using the *Horizontal* layout, selecting correctly only a third of the time. Although the *Pie* layout, with which participants averaged at almost 85% accuracy, produced consistently better results than the other ones, neither layout nor posture entailed a significant difference in any other objective measure like *completion time* and *decrease in preferred walking speed*. Subjective measures were significantly affected by *Layout* for Usability and by *Layout* and *Posture* for Workload. Usability was measured using the *System Usability Scale* (SUS), where the *Pie* layout was the only layout with above-average results, both while *Standing* and *Walking*, achieving a mean score of 69.4, and performing significantly better than the *Horizontal* one. Workload was measured using the *NASA-Task Load Index* assessment tool, where both the *Pie* and *Vertical* layouts (averaging at 40.95 and 46.32 respectively) received significantly better scores than the *Horizontal* one, which only managed 55.7. Qualitative research of the interviews conducted with each participant confirmed *Pie* to be the preferred layout used, provided insight into the social acceptability and safety perception of the interface and highlighted several areas of improvement of the product, some of which investigated via an observational study of the head orientation tracking interaction method.

The user interface developed offers an affordable and feasible route toward a novel, eyes-free piece of wearable computing.

8.2 Future work

Whereas this study focused on the optimal menu layout for the interaction techniques used, the set of layouts trialled was far from being exhaustive, and different layouts should be considered. As brought up by a participant, for example, their preconception of the *Pie* layout, and their conjectured candidate for the optimal layout, consisted of a cross-shaped one. Furthermore, observational research about all three developed layouts highlighted areas in the interaction space that could potentially make space for extra buttons or features. Indeed, our study about eyes-free, head-orientation-based menu interaction aimed to lay the foundations for further research that might instead consider:

- Different layouts
- UI panel shape and size
- Menu item shape and size
- Number of items
- Browsing input method
- Selection input method

Furthermore, our observational research on the use of the layouts highlighted how users believe is the most efficient way to use each layout, but additional studies on the most efficient head-orientation-tracking input patterns could point out ways to improve or alter the current layouts.

As mentioned above, we aim to investigate the optimal input methods in future studies. A completely hands-free interface that made use of head movements as head gestures could result beneficial overall or in particular use cases.

Given the challenge met in developing a usable audio output method to display system state, it might be interesting to investigate novel techniques.

- An auditory cursor that, continuously, rather than incrementally, informed the user of the state of the head-tracking cursor, similar to the continuous visual feedback of a real mouse cursor,
- A delayed hovering audio feedback technique, that would more discreetly release system state information, thus avoiding feedback overlap during scrolling-type events.

Finally, it would be interesting to analyse the social acceptability and safety of the developed interface, comparing them to those of dialogue systems and other mobile devices both in private, public, and work environments.

A Acknowledgements

We thank Professor Stephen Brewster for his invaluable support throughout the year. Furthermore, we want to thank Dr Mark McGill for his guidance with Bose AR Unity SDK, and all the participants that contributed to our user evaluation.

B | Ethics approval

School of Computing Science
University of Glasgow

Ethics checklist form for $3^{rd}/4^{th}/5^{th}$ year, and taught MSc projects

This form is only applicable for projects that use other people ('participants') for the collection of information, typically in getting comments about a system or a system design, getting information about how a system could be used, or evaluating a working system.

If no other people have been involved in the collection of information, then you do not need to complete this form.

If your evaluation does not comply with any one or more of the points below, please contact the Chair of the School of Computing Science Ethics Committee (matthew.chalmers@glasgow.ac.uk) for advice.

If your evaluation does comply with all the points below, please sign this form and submit it with your project.

---

1. Participants were not exposed to any risks greater than those encountered in their normal working life.

   Investigators have a responsibility to protect participants from physical and mental harm during the investigation. The risk of harm must be no greater than in ordinary life. Areas of potential risk that require ethical approval include, but are not limited to, investigations that occur outside usual laboratory areas, or that require participant mobility (e.g. walking, running, use of public transport), unusual or repetitive activity or movement, that use sensory deprivation (e.g. ear plugs or blindfolds), bright or flashing lights, loud or disorienting noises, smell, taste, vibration, or force feedback

2. The experimental materials were paper-based, or comprised software running on standard hardware.

   Participants should not be exposed to any risks associated with the use of non-standard equipment: anything other than pen-and-paper, standard PCs, laptops, iPads, mobile phones and common hand-held devices is considered non-standard.

3. All participants explicitly stated that they agreed to take part, and that their data could be used in the project.

   If the results of the evaluation are likely to be used beyond the term of the project (for example, the software is to be deployed, or the data is to be published), then signed consent is necessary. A separate consent form should be signed by each participant.

   Otherwise, verbal consent is sufficient, and should be explicitly requested in the introductory script.

4. No incentives were offered to the participants.

   The payment of participants must not be used to induce them to risk harm beyond that which they risk without payment in their normal lifestyle.

5. No information about the evaluation or materials was intentionally withheld from the participants.
   Withholding information or misleading participants is unacceptable if participants are likely to object or show unease when debriefed.

6. No participant was under the age of 16.
   Parental consent is required for participants under the age of 16.

7. No participant has an impairment that may limit their understanding or communication.
   Additional consent is required for participants with impairments.

8. Neither I nor my supervisor is in a position of authority or influence over any of the participants.
   A position of authority or influence over any participant must not be allowed to pressurise participants to take part in, or remain in, any experiment.

9. All participants were informed that they could withdraw at any time.
   All participants have the right to withdraw at any time during the investigation. They should be told this in the introductory script.

10. All participants have been informed of my contact details.
    All participants must be able to contact the investigator after the investigation. They should be given the details of both student and module co-ordinator or supervisor as part of the debriefing.

11. The evaluation was discussed with all the participants at the end of the session, and all participants had the opportunity to ask questions.
    The student must provide the participants with sufficient information in the debriefing to enable them to understand the nature of the investigation. In cases where remote participants may withdraw from the experiment early and it is not possible to debrief them, the fact that doing so will result in their not being debriefed should be mentioned in the introductory text.

12. All the data collected from the participants is stored in an anonymous form.
    All participant data (hard-copy and soft-copy) should be stored securely, and in anonymous form.

Project title __An Eyes-Free, Multimodal Interface For Auditory Headsets_

Student's Name _____Diego Drago___________________________

Student Number ______2384810d___________________

Student's Signature ______Diego Drago______________

Supervisor's Signature __S Brewster___________________

Date ____08/04/22_________

C | SUS questionnaire

SUS

The System Usability Scale (SUS) provides a “quick and dirty”, reliable tool for measuring the usability. It consists of a 10 item questionnaire with five response options for respondents; from Strongly agree to Strongly disagree.

*Required

1. I am participant number: *

2. This is task *

Mark only one oval.

- Standing, Horizontal
- Standing, Vertical
- Standing, Pie
- Walking, Horizontal
- Walking, Vertical
- Walking, Pie

3. I think that I would like to use this system frequently *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree						Strongly Agree

4. I found the system unnecessarily complex *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree						Strongly Agree

5. I thought the system was easy to use *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree						Strongly Agree

6. I think that I would need the support of a technical person to be able to use this system *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree						Strongly Agree

7. I found the various functions in this system were well integrated *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree						Strongly Agree

8. I thought there was too much inconsistency in this system *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree						Strongly Agree

9. I would imagine that most people would learn to use this system very quickly *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree						Strongly Agree

10. I found the system very cumbersome to use *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree						Strongly Agree

11. I felt very confident using the system *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree						Strongly Agree

12. I needed to learn a lot of things before I could get going with this system *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree						Strongly Agree

D NASA-TLX questionnaire

Figure 8.6

NASA Task Load Index

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

Name	Task	Date

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?

Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort — How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low — Very High

E Demographic questionnaire

Demographic Questionnaire

*Required

1. I am participant number *

2. I consent to take part to this experiment *

   Mark only one oval.

   - Yes
   - No

3. How old are you? *

4. Which gender do you identify as? *

   Mark only one oval.

   - Male
   - Female
   - Prefer not to say
   - Other:

5. What is your education level? *

Mark only one oval.

- Secondary education (e.g., GCSE)
- High School Diplomas/A levels
- Technical/Community College
- Undergraduate Degree (BS, B.Sc.)
- Graduate Degree (MA, M.Sc. etc.)
- Doctorate Degree (PhD or higher)

6. What is your current professional status? *

Tick all that apply.

- Student
- Employed
- Self-Employed
- Unemployed
- Other:

Technical proficiency

In the following questionnaire, we will ask you about your interaction with technical systems. The term 'technical systems' refers to apps and other software applications, as well as entire digital devices (e.g., mobile phone, computer, TV, car navigation).

Please indicate the degree to which you agree/disagree with the following statements.

7. I like to occupy myself in greater detail with technical systems. *

Mark only one oval.

	1	2	3	4	5	6	
Completely disagree							Completely agree

8 Bibliography

Apple. 3D Touch - User Interaction - iOS - Human Interface Guidelines - Apple Developer, 2022. URL https://developer.apple.com/design/human-interface-guidelines/ios/user-interaction/3d-touch/.

E. M. Bender, T. Gebru, A. McMillan-Major, and S. Shmitchell. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? . In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pages 610–623, Virtual Event Canada, Mar. 2021. ACM. ISBN 978-1-4503-8309-7. doi: 10.1145/3442188.3445922. URL https://dl.acm.org/doi/10.1145/3442188.3445922.

D. Bohn. Using Amazon's Echo Loop ring is like whispering a secret to Alexa, Sept. 2019. URL https://www.theverge.com/2019/9/25/20884044/amazon-echo-loop-smart-ring-hands-on-photos-alexa.

Bose. Bluetooth Audio Sunglasses | Bose, 2022a. URL https://www.bose.co.uk/en_gb/products/frames.html.

Bose. Smart Noise Cancelling Headphones 700 | Bose, 2022b. URL https://www.bose.co.uk/en_gb/products/headphones/noise_cancelling_headphones/noise-cancelling-headphones-700.html.

S. Brewster. Overcoming the Lack of Screen Space on Mobile Computers. *Personal and Ubiquitous Computing*, 6(3):188–205, May 2002. ISSN 1617-4909. doi: 10.1007/s007790200019. URL https://doi.org/10.1007/s007790200019.

S. A. Brewster and G. Marentakis. A Study on Gestural Interaction with a 3D Audio Display. In D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, S. Brewster, and M. Dunlop, editors, *Mobile Human-Computer Interaction - MobileHCI 2004*, volume 3160, pages 180–191. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004. ISBN 978-3-540-23086-1 978-3-540-28637-0. doi: 10.1007/978-3-540-28637-0_16. URL http://link.springer.com/10.1007/978-3-540-28637-0_16. Series Title: Lecture Notes in Computer Science.

C. Brito, K. E. Kutzko, and M. Wall. Demonstrating experimenter and participant bias. In *Investigative Ophthalmology & Visual Science*, pages 94–97. PubMed, July 2000. ISBN 978-1-4338-2714-3. doi: 10.1037/0000024-020.

B. Chaudhuri, L. Perlmutter, J. Petelka, P. Garrison, J. Fogarty, J. O. Wobbrock, and R. E. Ladner. GestureCalc: An Eyes-Free Calculator for Touch Screens. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '19, pages 112–123, New York, NY, USA, Oct. 2019. Association for Computing Machinery. ISBN 978-1-4503-6676-2. doi: 10.1145/3308561.3353783. URL https://doi.org/10.1145/3308561.3353783.

J. M. Corbin and A. Strauss. Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative Sociology*, 13(1):3–21, Mar. 1990. ISSN 1573-7837. doi: 10.1007/BF00988593. URL https://doi.org/10.1007/BF00988593.

B. Corporation. Welcome to the Bose AR Unity SDK, Sept. 2019. URL `https://developer.bose.com/bose-ar/unity`.

A. Crossan, M. McGill, S. Brewster, and R. Murray-Smith. Head tilting for interaction in mobile contexts. In *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '09, pages 1–10, New York, NY, USA, Sept. 2009. Association for Computing Machinery. ISBN 978-1-60558-281-8. doi: 10.1145/1613858.1613866. URL `https://doi.org/10.1145/1613858.1613866`.

C. Dicke, K. Wolf, and Y. Tal. Foogue: eyes-free interaction for smartphones. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*, MobileHCI '10, pages 455–458, New York, NY, USA, Sept. 2010. Association for Computing Machinery. ISBN 978-1-60558-835-3. doi: 10.1145/1851600.1851705. URL `https://doi.org/10.1145/1851600.1851705`.

R. Dillet. Fusar Mohawk Turns Any Helmet Into A Powerful Smart Helmet, Jan. 2016. URL `https://social.techcrunch.com/2016/01/07/fusar-mohawk-turns-any-helmet-into-a-powerful-smart-helmet/`.

D. Goldman. Google unveils 'Project Glass' virtual-reality glasses, Apr. 2012. URL `https://money.cnn.com/2012/04/04/technology/google-project-glass/index.htm`.

Google. Use gestures on your Pixel phone - Pixel Phone Help, 2022. URL `https://support.google.com/pixelphone/answer/7443425?hl=en`.

M. Gorzel, A. Allen, I. Kelly, J. Kammerl, A. Gungormusler, H. Yeh, and F. Boland. Efficient Encoding and Decoding of Binaural Sound with Resonance Audio. Audio Engineering Society, Mar. 2019. URL `https://www.aes.org/e-lib/browse.cfm?elib=20446`.

J. P. Hansen, V. Rajanna, I. S. MacKenzie, and P. Bækgaard. A Fitts' law study of click and dwell interaction by gaze, head and mouse with a head-mounted display. In *Proceedings of the Workshop on Communication by Gaze Interaction*, COGAIN '18, pages 1–5, New York, NY, USA, June 2018. Association for Computing Machinery. ISBN 978-1-4503-5790-6. doi: 10.1145/3206343.3206344. URL `https://doi.org/10.1145/3206343.3206344`.

A. Harley and J. Nielsen. Visibility of System Status, June 2018. URL `https://www.nngroup.com/articles/visibility-system-status/`.

N. Henze and E. Rukzio. Mobile Human-Computer Interaction. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '15, pages 2475–2476, New York, NY, USA, Apr. 2015. Association for Computing Machinery. ISBN 978-1-4503-3146-3. doi: 10.1145/2702613.2706690. URL `https://doi.org/10.1145/2702613.2706690`.

T. Hermann, A. Hunt, and J. G. Neuhoff. Chapter 1: Introduction | The Sonification Handbook. In *The Sonification Handbook*, pages 1–6. Logos Publishing House, Berlin, Germany, 2011. URL `https://sonification.de/handbook/chapters/chapter1/`.

I. Hoang, M. Ranchet, R. Derollepot, F. Moreau, and L. Paire-Ficout. Measuring the Cognitive Workload During Dual-Task Walking in Young Adults: A Combination of Neurophysiological and Subjective Measures. *Frontiers in Human Neuroscience*, 14:592532, 2020. ISSN 1662-5161. doi: 10.3389/fnhum.2020.592532.

Jeff Grubb's Game Mess. Google Duplex: A.I. Assistant Calls Local Businesses To Make Appointments, May 2018. URL `https://www.youtube.com/watch?v=D5VN56jQMWM`.

N. Kelly and S. Gilbert. The WEAR Scale: Developing a Measure of the Social Acceptability of a Wearable Device. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '16, pages 2864–2871, New York, NY, USA, May 2016. Association for Computing Machinery. ISBN 978-1-4503-4082-3. doi: 10.1145/2851581.2892331. URL https://doi.org/10.1145/2851581.2892331.

B. Lang. Oculus Rift S Revealed with Inside-out Tracking, Resolution Bump, & New Ergonomics, Mar. 2019. URL https://www.roadtovr.com/oculus-rift-s-specs-release-date-announcement-gdc-2019/. Section: GDC 2019.

E. LoPresti, D. M. Brienza, J. Angelo, L. Gilbertson, and J. Sakai. Neck range of motion and use of computer head controls. In *Proceedings of the fourth international ACM conference on Assistive technologies*, Assets '00, pages 121–128, New York, NY, USA, Nov. 2000. Association for Computing Machinery. ISBN 978-1-58113-313-4. doi: 10.1145/354324.354352. URL https://doi.org/10.1145/354324.354352.

J. Löwgren. *Interaction Design - brief intro*. 2022. URL https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/interaction-design-brief-intro.

Mark McGill. CHI 2020 Video Talk: Acoustic Transparency and the Changing Soundscape of Auditory Mixed Reality, May 2020. URL https://www.youtube.com/watch?v=R8U5HFJXB98.

M. McGill, S. Brewster, D. McGookin, and G. Wilson. Acoustic Transparency and the Changing Soundscape of Auditory Mixed Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–16, Honolulu HI USA, Apr. 2020. ACM. ISBN 978-1-4503-6708-0. doi: 10.1145/3313831.3376702. URL https://dl.acm.org/doi/10.1145/3313831.3376702.

J. Nielsen. 10 Usability Heuristics for User Interface Design, Apr. 1994. URL https://www.nngroup.com/articles/ten-usability-heuristics/.

Oura. Oura Ring: Accurate Health Information Accessible to Everyone, Apr. 2022. URL https://ouraring.com.

T. Pozzo, A. Berthoz, and L. Lefort. Head stabilization during various locomotor tasks in humans. *Experimental Brain Research*, 82(1):97–106, Aug. 1990. ISSN 1432-1106. doi: 10.1007/BF00230842. URL https://doi.org/10.1007/BF00230842.

Z. Qattan. Bose AR Web SDK, Jan. 2022. URL https://github.com/zakaton/Bose-Frames-Web-SDK. original-date: 2019-01-23T20:02:54Z.

M. F. Roig-Maimó, J. Varona Gómez, and C. Manresa-Yee. Face Me! Head-Tracker Interface Evaluation on Mobile Devices. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '15, pages 1573–1578, New York, NY, USA, Apr. 2015. Association for Computing Machinery. ISBN 978-1-4503-3146-3. doi: 10.1145/2702613.2732829. URL https://doi.org/10.1145/2702613.2732829.

M. F. Roig-Maimó, C. Manresa-Yee, J. Varona, and I. S. MacKenzie. Evaluation of a Mobile Head-Tracker Interface for Accessibility. In K. Miesenberger, C. Bühler, and P. Penaz, editors, *Computers Helping People with Special Needs*, Lecture Notes in Computer Science, pages 449–456, Cham, 2016. Springer International Publishing. ISBN 978-3-319-41267-2. doi: 10.1007/978-3-319-41267-2_63.

S. Ronkainen, J. Häkkilä, S. Kaleva, A. Colley, and J. Linjama. Tap input as an embedded interaction method for mobile devices. In *Proceedings of the 1st international conference on Tangible and embedded interaction*, TEI '07, pages 263–270, New York, NY, USA, Feb. 2007. Association for Computing Machinery. ISBN 978-1-59593-619-6. doi: 10.1145/1226969.1227023. URL https://doi.org/10.1145/1226969.1227023.

D. Smith. Amazon Echo Frames: Here's what you didn't know, Oct. 2019. URL https://www.cnet.com/tech/mobile/amazon-echo-frames-heres-what-you-didnt-know-about-amazons-new-smart-glasses/.

J. Sodnik, C. Dicke, S. Tomažič, and M. Billinghurst. A user study of auditory versus visual interfaces for use while driving. *International Journal of Human-Computer Studies*, 66(5):318–332, May 2008. ISSN 1071-5819. doi: 10.1016/j.ijhcs.2007.11.001. URL https://www.sciencedirect.com/science/article/pii/S1071581907001553.

U. Technologies. Unity Real-Time Development Platform | 3D, 2D VR & AR Engine, 2022. URL https://unity.com/.

J. Vincent. Google's AI sounds like a human on the phone — should we be worried?, May 2018. URL https://www.theverge.com/2018/5/9/17334658/google-ai-phone-call-assistant-duplex-ethical-social-implications.

S. Zhao, P. Dragicevic, M. Chignell, R. Balakrishnan, and P. Baudisch. Earpod: eyes-free menu selection using touch input and reactive audio feedback. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, pages 1395–1404, New York, NY, USA, Apr. 2007. Association for Computing Machinery. ISBN 978-1-59593-593-9. doi: 10.1145/1240624.1240836. URL https://doi.org/10.1145/1240624.1240836.