

YALE-NUS COLLEGE

CAPSTONE REPORT

Building an Interface for Haptic Spatial Navigation

Author:

Hrishi OLICKEL

Supervisor:

Dr. Simon PERRAULT

*A thesis submitted in fulfillment of the requirements
for the degree of B.Sc (Hons)*

March 27, 2018

Declaration of Authorship

I, Hrishi OLICKEL, hereby declare that this Project, the Capstone Report and associated work listed herein, is the end result of my own work and that due acknowledgement has been given in the bibliography and references to ALL sources printed, electronic, or personal, in accordance with the academic regulations of Yale-NUS College.

I acknowledge that the Thesis is subject to the policies relating to Yale-NUS College Intellectual Property (Yale-NUS HR 039).

Signed:

Date:

YALE-NUS COLLEGE

Abstract

B.Sc (Hons)

Building an Interface for Haptic Spatial Navigation

by Hrishi OLICKEL

Sensory substitution is a field of Human Computer Interaction that has long been explored as a possible avenue to replace and improve human sensory perception. This paper explores novel frameworks for designing and evaluating a sensory substitution system based on haptic feedback, in a way that can build a standard extensible base for future research. A new device for haptic spatial navigation, along with a novel encoding protocol is also developed and evaluated. Results show that training is a significant component of such systems, and indicates possible features that can improve the experience of the user. The result is two experiments, in two and three dimensions, that deal with spatial navigation. In these, participants are able to navigate their environments while having replaced one of their sensors with the device.

Acknowledgements

I am very thankful to my capstone supervisor Prof. Simon Perrault for the invaluable guidance in shaping what was a weekend project into a capstone thesis, and for helping me through the many, many obstacles along the way. This project would not have the form it does today without his constant support and valued feedback.

I would also like to thank Aaron Ong and Parag Bhatnagar for help in developing the kernel that was haptic sensory substitution into a real project, and for the apt criticism and many midnights spent poring over the viability of the prototypes.

I would like to thank my capstone examiner, Prof. David Smith, whose contributions both in writing and in person to this report have made it (by multiple accounts) far more understandable yet succinct.

I am extremely grateful for the MCS batch of 2018, whose support and camaraderie in the tough times of the year helped complete this project, and keep me sane as I write the final sentences of it.

I would also like to thank all the participants that made time for the experiment even as it was uncompensated, and stayed seated in a chair in my room, in the blind for much time as I fumbled about trying to reset the experiment.

Finally, I'd like to thank Hebe Hilhorst for her sharp quick-witted criticism and warm support from the very inception of this project. I could not have done it without you.

Contents

Declaration of Authorship	i
Abstract	ii
Acknowledgements	iii
1 Introduction	1
1.1 Motivation	1
1.2 Claims	3
1.3 Related Work	4
2 Development	8
2.1 Goals	8
2.2 Components	10
3 Experiment	12
3.1 Phase 1: Sightless Racing	12
3.2 Phase 2: Virtual Reality Fitts Test	17
4 Results	21
4.1 Experimental Apparatus	21
4.2 Experimental Design	22
4.3 Phase 1	23
4.4 Phase 2	24
4.5 Discussion	28
4.5.1 Training	29

4.5.2	Active vs. Passive information flow	30
4.5.3	Demographics	30
4.5.4	Evaluation	31
4.6	Further Research	32
4.7	Conclusion	32
A	Supporting material	34
Bibliography		35

List of Figures

1.1	Mechanical apparatus for sensory substitution	2
1.2	BRAINPORT implementation	4
1.3	GIST Interface	5
1.4	Structure of TACIT.	6
1.5	Servo arrangement	6
3.1	P1 Device	13
3.2	Servo configuration	14
3.3	FM switching visualization	15
3.4	Implementation of the World Interface.	15
3.5	P1 experiment	16
3.6	Cartesian Vectors in VR	19
3.7	Virtual Reality Setup	20
4.1	Experiment Setup	22
4.2	Phase 1 Error	23
4.3	Phase 2 Time to Target	25
4.4	Phase 2 Timing Graph	26
4.5	Phase 2 Target Error	27
4.6	Phase 2 Individual Timing	28

List of Tables

4.1	Phase 1 Timing ANOVA	26
4.2	Phase 1 Target Error ANOVA	28

Chapter 1

Introduction

We as humans perceive the world through a sensory envelope made up of the signals relayed to our brains through the primary senses - touch, sight, taste, hearing and so on. This sensory envelope (called the *umwelt*) is in many cases limited compared to the current state-of-the-art in technology. We are unable to perceive the full breadth of the electromagnetic spectrum, hear anything beyond a narrow range of frequencies, or feel the digital signals that make up the internet and the collective digital consciousness of the human race. This is without considering those of us that have been deprived of the standard human *umwelt* through congenital defects and injury.

In this paper, we will consider the possibility of using recent advances in technology to expand the sensory envelope through sensory substitution.

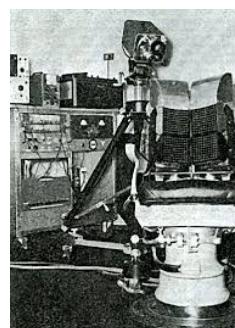
1.1 Motivation

Before working on this project, my primary background consisted of working on Artificial Intelligence networks (Hawkins and Blakeslee, 2007) that focused on adapting the plastic nature of the brain to create easily generalizable learning systems. Having seen such systems work in practice, the possibility of performing such reconfiguration *in vivo* was quite exciting. The possibility of experiencing extrasensory perception as well as tackling cheap

and generalizable solutions for sensory replacement served as the primary motivation.

In 2003, a paper (Bach-Y-Rita, 2003) by a neuroscientist named Paul Bach-y-Rita claimed to have developed a device by which the blind could regain their sight. A series of electrodes attached to the tongue conveyed a low resolution image captured by a camera sitting atop the subject's forehead. Through this rudimentary but ground-breaking procedure, subjects both congenitally blind and temporarily blindfolded could learn to distinguish objects and traffic and catch moving objects with remarkable accuracy with only hours of training time.

Unfortunately, his work did not see much subsequent study, perhaps in part due to an incomplete understanding of how neuroplasticity works (Doidge, 2008).



(A) The image displayed by the mechanical chair. (B) Brobdingnagian mechatronic device with moving attachments.

FIGURE 1.1

Moving further back into history, a Nature paper (White et al., 1969) by the same author 34 years before reported a mechatronic device which consisted of a chair that had a series of vibrating devices attached to the back, and cranks by which a camera could be moved by subjects. 400 lbs in total, the bulky device provided a similar experience, albeit within the limitations of the computing technology of its time.

Despite the experiments revealing that subjects were able to perform complicated visual processing tasks solely based on input from the chair (with only minimal training), the methods proved unusable in the industry due to a multiplicity of factors (Phillips and Zhao, 1993). The devices placed the user in significant discomfort, were complicated to use and relatively low-resolution, all in addition to being prohibitively expensive and limited by the technology of their time.

Since then, a number of devices (*BrainPort V100 Vision Aid*) have been released that attempt to make use of this method, but none have found mainstream commercial success¹.

Given the improvements in technology related to processing power, portability and hardware quality, renewed attempts are clearly needed.

1.2 Claims

This paper presents the following original contributions:

1. A hardware device for haptic sensory substitution along with designs for the construction, as well as a short evaluation of existing hardware and software stacks.
2. Multiple implementations for sensory substitution using haptic feedback, using continuous feedback processes as well as delayed-feedback spatial navigation tasks, each of which include:
 - (a) a front-end for providing visual input to the user during the training phase, as well as useful readouts to the researcher,
 - (b) a transmission protocol, which maps information from the task at hand (spatial coordinates, velocity information, etc) to time-based

¹(perception - Why is sensory substitution not that successful - Cognitive Sciences Stack Exchange)

sensor actuation signals (20° on servo 1, 35° on servo 2, etc) in real-time.

3. Multiple metrics for evaluating the performance of a sensory substitution system, as well as sample tasks that can be used to standardise and compare performance across the board for future research.

In addition, code for displaying results in real-time, modules for managing servo overload, network latency and other factors was also written by the author.

1.3 Related Work

Some of the work that has been previously undertaken involves aiding the visually impaired in navigation and object recognition. These were helpful in identifying primary methods of communication with the brain through haptic feedback. The knowledge was used to construct early versions of a framework² for providing signals to the brain, in a way that could be quickly iterated upon to find novel solutions.

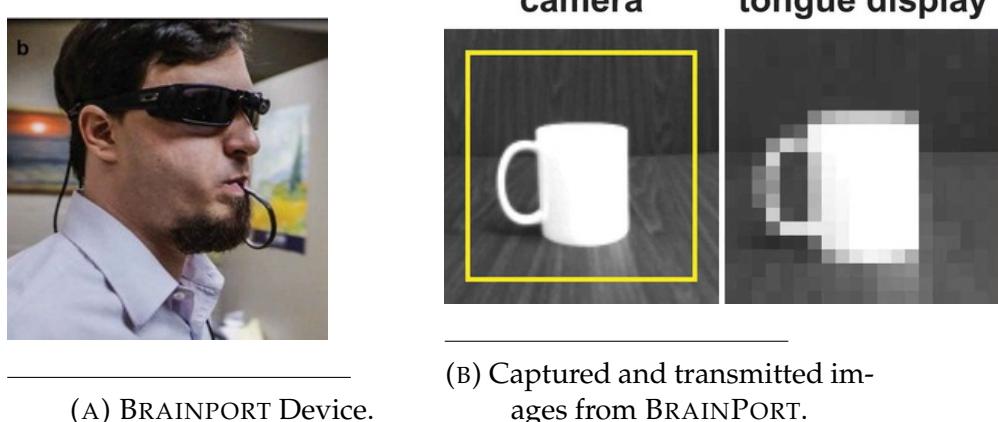


FIGURE 1.2

²In the course of this paper, the terms *framework* and *system* are used to describe the complete end-to-end sensory substitution module.

BRAINPORT (Upson, 2007) implements an industrial version of the aforementioned paper (Bach-Y-Rita, 2003). Figure 1.2a shows the modifications made since the original, resulting in a smaller device that can fit on a pair of eyeglasses and requires an additional battery pack not in view.

Despite having received considerable funding and attention in the media, clinical trials of the device (Nau et al., 2015) encountered a number of problems. Studies report the device as having "poor spatial acuity" (Stronks et al., 2016), as well as severe limitations in temporal resolution. The primary cause of these problems is attributed to the "limited bandwidth of perceived information" that the device is capable of.



FIGURE 1.3: Demonstration of the GIST audio-spatial interface

Since 2007, the advances in processing capability and computer-vision have led to projects that make use of real-time image processing. GIST (Khambadkar and Folmer, 2013) is a device that provides additional information to a user about what he is pointing at. Figure 1.3 demonstrates this device in operation. While initially quite promising, the device cannot be expected to work well in realistic scenarios. Due to the use of bulky external depth sensors that require non-mobile sources of power, the device is not truly portable. In addition, the authors report that scanning any object in the non-dominant side of a user causes the device to recalibrate (due to a consequent small turn of the torso). In addition, the device only reports two slow-updating channels of information. The two channels are the distance to an object and the color (presumably limited by the resolution of descriptive color words in English). One other project (Akhter et al., 2011) attempts to



FIGURE 1.4: Structure of TACIT.



FIGURE 1.5: Internal circuitry demonstrating the arrangement of servo motors.

use smartphones to create and process stereo information for depth sensing, but the project has not reached the testing stage.

The project that comes closest to solving some of these problems is the Tacit Project (*Meet The Tacit Project. It's Sonar For The Blind.*), which attempts to use stereo ultrasound sensors to convert distance information to haptic "squeezes" on the user's arm. Despite not providing any field tests or experiment results, my interest in the project remains because it aligns with the goals we have been forming from review: It is cheap, portable, relatively high bandwidth and places importance on temporal resolution and continuous signaling. Unfortunately, the lack of any experiments demotes us to reading comments and online opinions from users who *do not* possess the device. In addition, Tacit is limited in temporal resolution, as it implements delays of up to 500ms between actuations, effectively being limited to a near 1 Hz frequency. From the projects we've considered, a clear set of goals can be formulated from which to build a framework, and a number of metrics that can be used to evaluate such a system.

A well-known paper (M. Fitts, 1992) in Human Computer Interaction from 1954 outlines Fitts Law, a predictive model that can be used to calculate the time taken for a one-dimensional movement-based task from the distance moved and the size of the target. While one-dimensional movement cannot be easily extended to spatial navigation, there is merit in observing the time taken from an information theory standpoint. Below, we will also attempt

to develop a set of metrics that can be used beyond the scope of this experiment to determine the theoretical maximum bandwidth of sensory substitution systems, if one exists.

Chapter 2

Development

From the projects we've considered, a clear set of goals can be formulated from which to build a framework, and a number of task-based metrics that can be used to evaluate such a framework. The distinction between these is that the goals for the framework concern its construction, designed to maximize the utility of the system to an end-user, while the metrics are intended to be used in measuring performance with comparison in mind. This will be explored at the time of the experiment.

2.1 Goals

For the framework, our goals are as follows:

- **Generalizable Transmission Protocol:** One of the primary challenges of sensory substitution lies in mapping data from the external world to the sensory realm which conveys information to the user. This is also a function of the human sensory infrastructure. For example, hearing operates by mapping specific wavelengths of sound waves to hair strands in the ear that convert the amplitude into intermittent electrical signals. Arguably one of the most important parts of the system, the aim of isolating the transmission protocol is to develop, compare and decide from a number of encodings for information streams encountered in the real world. An emphasis is placed on the protocol not

being application specific in a way that hinders its application to similar tasks. Additional emphasis is placed on minimizing the calibration and testing required before deployment.

- **Continuous Signaling:** Another conspicuously lacking element in past work is the absence of continuous feedback. Systems we have looked at so far continuously snapshot the environmental signals meant to be conveyed, reduce them over a time window to a lower resolution counterpart and then transmit this to the user. In the best case scenario, the user is presented with intermittent morsels of information in stark contrast to the fluid nature of the human *umwelt*. Studies (Kristjánsson et al., 2016) regarding the effectiveness of sensory substitution devices indicate that "spatiotemporal continuity" is significant in maintaining immersion and shortening the training and acclimatization period. Based on this, our framework prioritizes the transmission of continuous signals. Measuring and increasing the frequency (in addition to the bandwidth metrics stated above) of transmitted information is also of value.
- **Effective Bandwidth:** While measurement of effective bandwidth is only possible in a task-based environment (after implementation), it is a priority in design to ensure that no part of the system causes a bottleneck when conveying this information. This will involve making sure that network latency (when involved), as well as the hardware limitations of the device, are understood and accounted for.
- **Spatial Acuity:** The project will primarily focus on spatial navigation and other related objectives for the experiments, as implementing them on a realtime system will also satisfy the temporal acuity goal. The objective is to convey information regarding full 3-dimensional¹ space

¹It is worth noting that the related works we have considered resort to slicing 3-dimensional space to present a 2-dimensional version to the user. It is an interesting question if this approaches a fundamental limit.

and object recognition with the intent of performing tasks.

- **Portability:** The rise of personal mobile devices presents an opportunity that had been closed to most previous studies. Attempting to make the implementation less expensive and more mobile and self-sufficient will be a priority. This means reducing the number of separate modules, reliance on experimentally determined conditions and power consumption just to name a few attributes.

The purpose of these goals is to serve as reasonable guidelines for evaluating the results derived, as well as a possible scaffold on which future extensions can be developed.

2.2 Components

In the interest of modularization, the framework can be subdivided into the following general components, each with as close as possible to a distinct set of constraints, technical stacks and parameters for optimization:

1. **Transmission Interface:** The (primarily) hardware component that transmits (via haptic feedback in our case) previously encoded information to the user. The optimizations here are towards making the system as responsive as possible, minimizing latency, as well as deciding the best components and control circuitry to improve the experience. As we will see later, even a choice of fabric for the instrument case or the type of soldering involved can have a major impact.
2. **Signal Encoder:** Despite the hardware-sounding name, this is primarily software-based, and handles the encoding of incoming sensor streams (with minor prior correction) to outgoing activation signals. From experiments, the location of this module is important - placing it as close

to downstream² as possible leads to lower latency, however there is a delicate balance to be struck. Devices tend to have more processing power as you move upstream, promising significant speedups that can offset the loss from distance.

3. **Application:** This is the module that handles most tasks needed to ensure that the entire system is operational, including a networking module, display management, device enumeration and asynchronous processing, logging, etc.
4. **World Interface:** The World Interface handles all aspects of the task being performed by the user. This may range from simulating a Virtual Reality environment to computer-vision based mapping and tracking of spaces for an Augmented Reality task. Task completion is also measured and reported from this module.

²Terminology: Upstream is closest to the front-end while downstream is closest to the user.

Chapter 3

Experiment

Now we can move to the tasks, as well as the metrics associated with them. The tasks, metrics and results are divided into Phases 1 and 2. Phase 1 explores continuous-feedback based navigation in two dimensions. Phase 2 then builds on the work in Phase 1 to investigate delayed-feedback based spatial navigation in three dimensions using a Fitts' test. The metrics are explained here, but the results are grouped in the results section afterwards.

3.1 Phase 1: Sightless Racing

In Phase 1, the virtual space of a car and race track were mapped to a haptic device. The primary aim of the experiments were to see if users with some training could intuitively drive the car around the track, dodging obstacles and following the road.

With this in mind, the tech stack used is below, in keeping with the module separation outlined in Section 2.2:

- 1. Transmission Interface:** The Transmission interface consists of servos, rated at $1.195 \text{ kg}\cdot\text{cm}^1$ in torque, arranged on a flexible plastic film wrapped around the user's arm. The servos are outfitted with 3 cm arms that rotate down to push lightly into the user's skin, and also connected to a Raspberry Pi Model B, running a python server which communicates

¹Datasheet information for the servo is provided in experiment design.

with the front-end. Figure 3.1 shows the device as it is fitted to the user's arm.

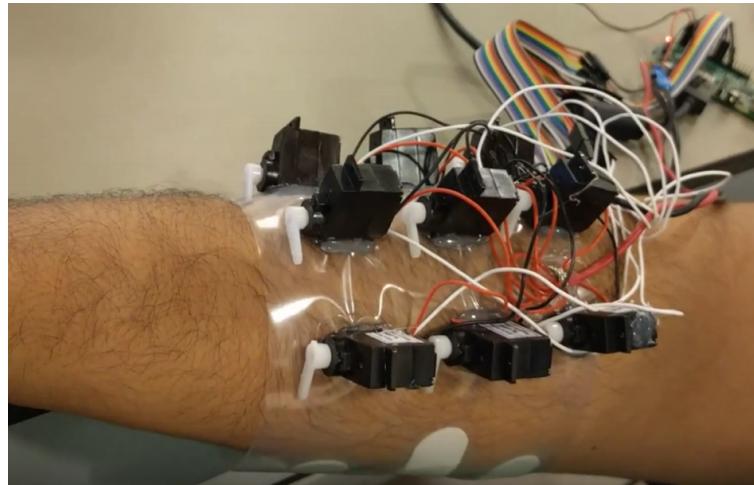


FIGURE 3.1: Phase 1 Transmission interface.

The basic material used for construction was determined after much trial and error. Optimizing for maximum sensitivity, the first configuration enabled the arms of the servos to directly make contact with the skin. While this did achieve the intended purpose, users were overstimulated and described sensations of "sensory overload" and "uncomfortable stretching and pulling". Fabric was the second option, but was overtly effective in dampening sensitivity and preventing users from distinguishing the servos, as well as adding noise due to the texture rubbing against the skin. In addition, fabric proved to quite unhygienic as it was permeable and receptive to sweat and grime, which made it unsuitable for continued use (especially with multiple participants). The ideal option proved to be the ESD-safe plastic bag that the servos arrived in - comfortable without being sharp around creases and stretch-proof, it had the added advantage of not carrying odors.

2. **Signal Encoder:** Design for the Signal Encoder begins with the configuration of servos in the Transmission Interface, which makes use of a

3x3 grid of servos. These are placed in a square on the user's arm (Figure 3.2). Preliminary (with one or two participants) pilot testing helped eliminate configurations such as a strip around the arm, crowded circle placement, and lengthwise across the arm due to lack of sensitivity and general discomfort. The final wide configuration also provided the advantage of being adaptable to other locations on the body, and easier to map in software.

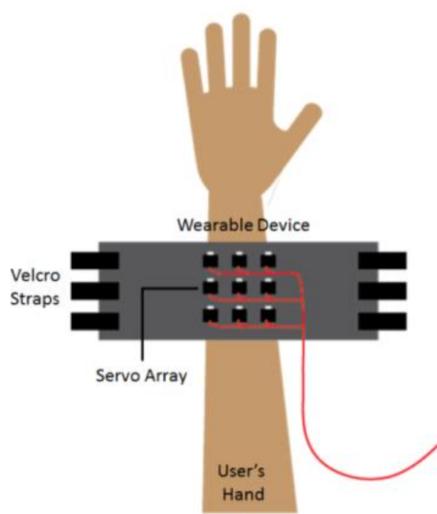


FIGURE 3.2: Servo configuration.

The configuration contributed to the activation patterns in the Signal Encoder. Pilot testing revealed that the four corners were the most identifiable, and that the servos in between contributed to immersion. The information received was consolidated into four analog time streams, which was subsequently converted to discrete values of 16-bit resolution, which was then frequency modulated (*Frequency modulation 2017*) and amplified to produce the activation signals for the corner servos. Figure 3.3 shows how the activation signals are combined. For the servos in between the corners, a discrete signal was generated for them by alternating between their neighbors using a base frequency and likewise coded in frequency with a fixed amplitude. This configuration

allows for four continuous vector streams which can be conveyed to the user.

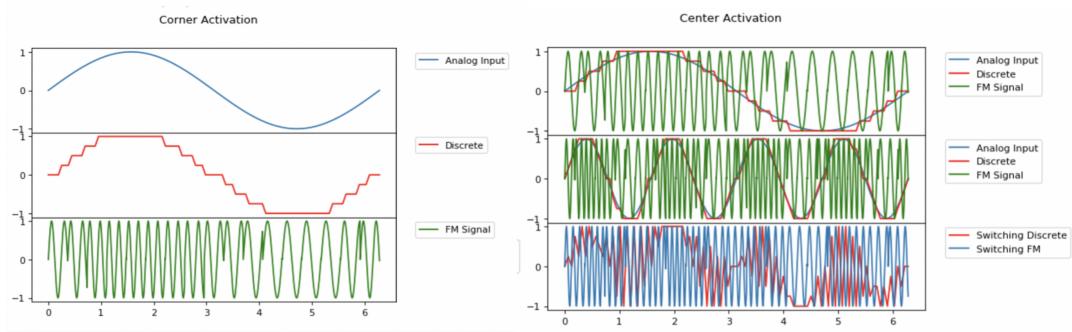


FIGURE 3.3: Visualization of switching FM implemented in activation.

This particular signal pattern proved most effective in shortening training times and providing immersion and the strongest sensory substitution as reported by the early participants.

3. Application: For Phase 1, the application was developed to optimize for fastest turnaround time in code sprints, using python and Javascript to make the process easier. A webserver written in Javascript listens asynchronously for signal broadcasts from the World Interface, and the upstream packet transmission is managed by a python webserver working in conjunction with the World Interface application.



FIGURE 3.4: Implementation of the World Interface.

4. World Interface: The primary purpose of Phase 1 was to find the best possible signal encoding for real-time time-variant data. The users were presented with a familiar gaming interface resembling OutRun (*Out Run* 2017) (Figure 3.4, which made the initial introduction easier. The World Interface application responded to keyboard controls, and produced the following values as a continuous multi-dimensional vector:

- the current position of the car relative to the center of the road,
- the current position of the car relative to the road on current trajectory after t+4.0 seconds, and
- the positional information of obstacles.

These input vectors are then provided to the signal encoder.

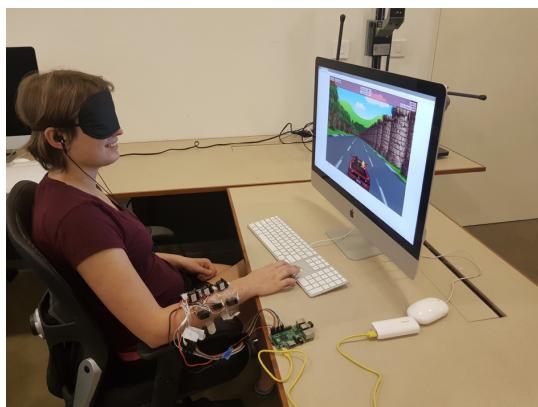


FIGURE 3.5: Phase 1 training setup.

The experimental setup involved a training stage and a testing stage. The training stage involved asking the participants to play the game while wearing the device for a fixed period of time (set to 10 minutes for standardisation) without being blindfolded.

Following this, in the testing stage participants were blindfolded and played the game solely using input from the haptic device.

Due to time and resource limitations, a formal experiment for Phase 1 was not performed. The device was tested at a hackathon on subjects of ages

18–30, and the general responses have confirmed it as an effective method of sensory substitution. With ten minutes of training (and no instruction as to the function of the device), all participants were able to play the game without crashing on a procedurally generated randomized track. In addition, when encountering an obstacle or a turn in the road, all participants were able to reflexively² judge the correct rate of turn and direction without hesitation³.

3.2 Phase 2: Virtual Reality Fitts Test

Phase 2 has the participants perform a Fitts test (a pointing task that measures accuracy in a standardised way) in a Virtual Reality environment, with and without the aid of sight. Shown in Figure 4.1, the participants follow the training-testing structure of Phase 1, where they point and click at various targets in the virtual world.

Iteratively improving on the results and design of Phase 1, Phase 2 is a more complex task in a noisier environment. For one, Phase 1 operated on continuous-feedback. The participants were continuously made aware of success and failure in relation to the task, while Phase 2 provides delayed-feedback, where the completion of a task has a much more time-dilated response. In addition, moving from two dimensions (two vectors) of critical information to three adds the potential for noise and sensory overload. With this in mind, let's consider the modular design where it deviates from Phase 1:

²Interesting feature of note that differentiated other encodings was the report from users that they were 'remembering' the meaning of a signal as opposed to 'feeling' the result of a signal. My intuition is that the latter is a sign of the brain using non-conscious processing to understand the new input. Participants performed better across all categories when they reported that they were 'feeling' instead of 'remembering' (the terms borrowed from participants' recounting the experience).

³Younger participants were able to perform similarly as older participants with significantly less training time. This correlates with literature (Kolb and Gibb, 2011) in that age is negatively correlated with neuroplasticity.

1. **Transmission Interface:** The transmission interface remained identical to the previous phase.
2. **Signal Encoder:** As to be described in the World Interface section, Phase 2 encodes 3D information, in this case the position of user-directed controller relative to a designated target object. The first consideration in the design of the Signal Encoder involved the mapping of spatial orientation in regards to the controller. A number of options underwent preliminary pilot testing, which included a cartesian set of vectors in 3D space, polar coordinates encoding the angle and distance to a target, and other more convoluted sets of representations. The testing revealed the most effective mapping to be a set of cartesian vectors adjusted to the user's orientation, such that the controllers' left, right, up and down were the same as the participants'. Performance was also improved when these vectors were inverted to provide adjustment information rather than absolute information (eg. 'go left to find the target' as opposed to 'you are right of the target' in terms of an instruction). For additional information, a new binary vector was added (an alert variable) that switched on for half a second when the pointer encountered the correct target to measure the effects.

Figure 3.6 shows a representation of the magnitude and origin positioning of the vectors conveyed through the haptic interface.

3. **Application:** The application, due to restraints imposed by the introduction of VR, was developed in C# on the server side. The same software was used on the client side as Phase 1, with some modifications.
4. **World Interface:** As Phase 2 is built on top of Virtual Reality, the World Interface proved to be one of the most complex modules in terms of design and implementation. There are a number of offerings on the market for Virtual Reality, but spatial localisation is not well implemented.

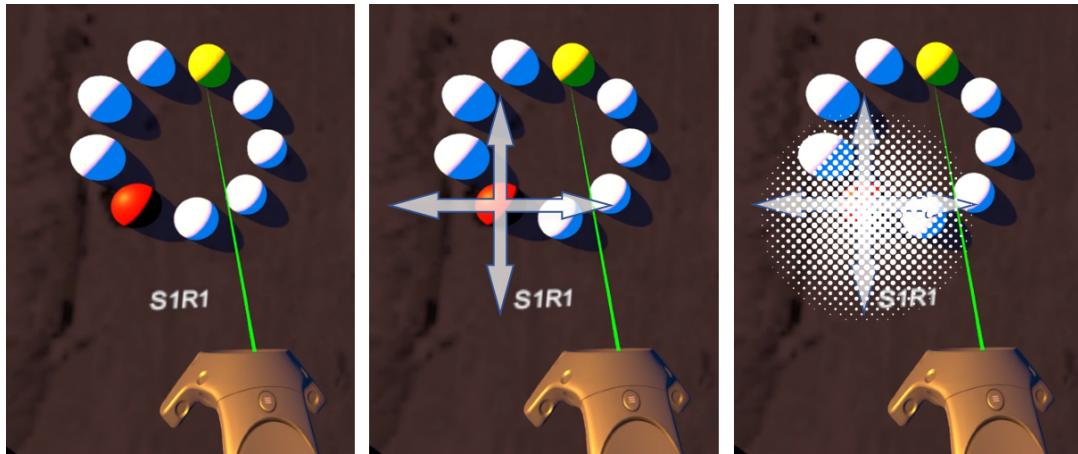


FIGURE 3.6: Overlay representing the cartesian vectors' magnitude and origin.

Tracking an object with certainty across a 3d space is an open problem in computer vision, and acceptable solutions with low latency have only recently become available on the consumer market. The options considered included implementing SLAM (*Simultaneous localization and mapping* 2017) a resource intensive algorithm to map indoor spaces using a single camera, and the two competing Virtual Reality products on the market that offer this functionality, the HTC Vive and the Oculus Rift. SLAM was abandoned as it lacked a robust implementation and would require coding work equivalent in time to a separate capstone to implement. Some open implementations (Yan, 2017) (aivijay, 2017) for SLAM exist, but none that can function within the constraints of a mobile device. In the end, the HTC Vive was chosen for the following reasons: One, the Vive offered easily extensible tracking algorithms in the form of HTC's Lighthouse technology⁴ (Niehorster, Li, and Lappe, 2017), which promised to aid future research. Two, HTC Software was closely compatible with Steam VR and Unity, which enabled faster development and turnaround times.

⁴An interesting video capturing the operation of the Vive's lighthouses is at <https://www.youtube.com/watch?v=5yuUZV7cgd8>

The World Interface (Figure 3.7) contains three primary parts. The first is a virtual Fitts Test board, which contains targets that need to be selected in a randomized order. The second is a controller that exists both virtually and as a physical object held by the participant. The third is a virtual pointer extruded from the controller at the same angle. The participants can click the controller via an attached button, and are asked to do so in order to select the correct target, indicated either through color (for sight) or instruction from the haptic device. Red represents the next target, yellow for any selected target, and green when the right target has been selected.

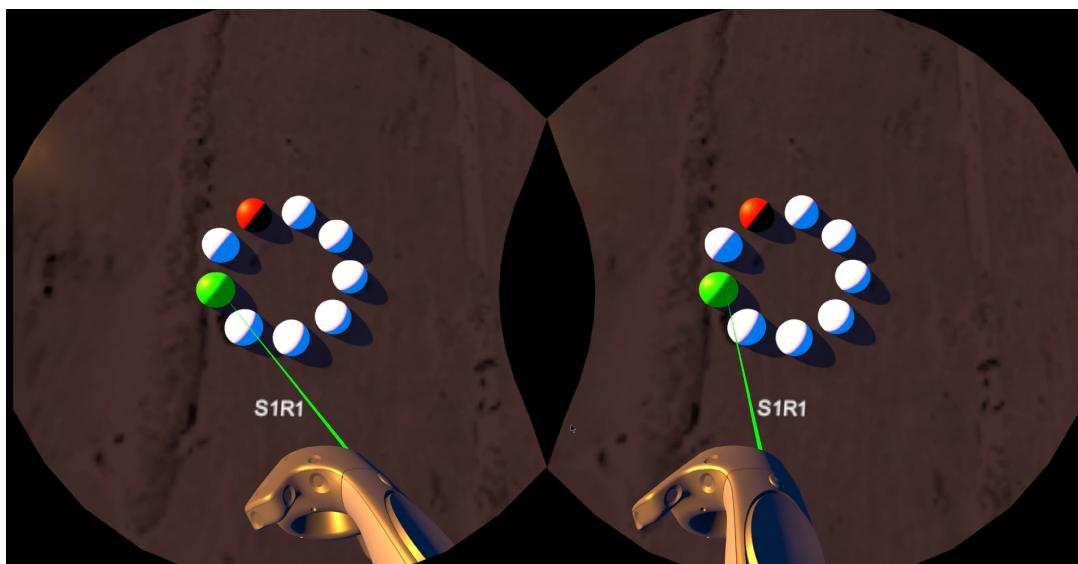


FIGURE 3.7: Virtual Reality Setup for Fitts Task.

Complete with the design, we can now move to the results.

Chapter 4

Results

4.1 Experimental Apparatus

The haptic device mentioned in detail above was made of a Raspberry Pi Model B¹, which was used to drive the nine servos² as part of the haptic device at 3.3V. The Rasberry Pi was run in headless mode, connected via SSH to a BASH terminal for logging. The front-end computer was a custom built desktop³ connected to a Sharp PN-K321 4K (3840x2160) 32" monitor with a Logi Master MX mouse and a Logi CRAFT keyboard. All trials were performed in an isolated room at Yale-NUS, away from distractions.

Figure 4.1 shows the experiment configuration.

For Phase 2, the HTC Vive Virtual Reality headset was used in stock configuration. In addition, the multi-core CPU on the Raspberry Pi was used in Phase 2 for multi-threaded operation, which permitted a web server to receive signals, and another thread on the same server to actuate the servos. Raspbian is not a realtime OS, so some timing variation is expected. However, for the timings in question it should not be noticeable.

¹1.2 Ghz Quad-core ARMv8 CPU, 1 GB RAM, 40 GPIO Pins (with nine pins operating in software PWM mode). The Raspberry Pi was running the Wheezy distribution of Raspbian OS from a 16 GB micro SD card.

²Rated for torque: 1.195 kg-cm at 4.8V, speed: 0.15 sec/ \circ , single-top ball bearing nylon gear servo. Part No. A0090

³Intel Core i7-6700K CPU @ 4.00 Ghz, 16 GB RAM, 64-bit Windows 10 Professional with an NVIDIA GeForce GTC 980 Ti

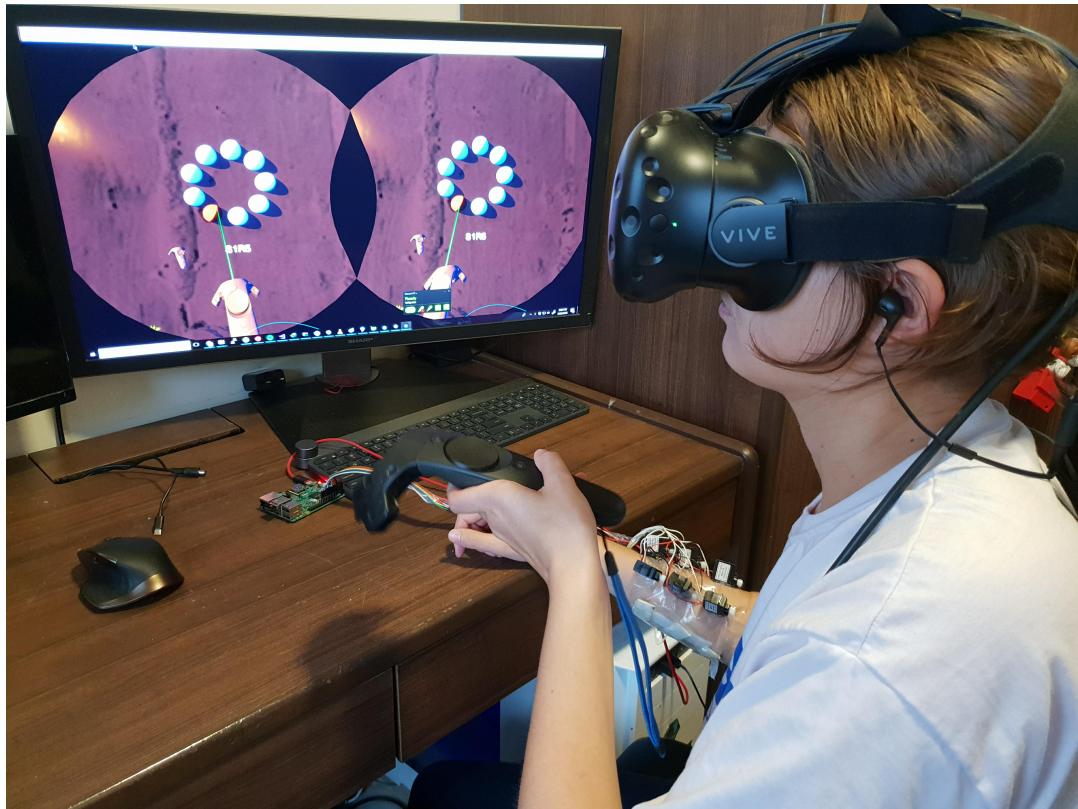


FIGURE 4.1: Experiment Setup showing participant.

4.2 Experimental Design

Pilot testing for Phase 1 was performed at a hackathon, with 7 participants (2 female, 5 male) after which the data was aggregated and anonymized. Each participant was given five minutes of training and five minutes of testing (blind) time, and the participants were between 18 to 25 years of age.

8 participants (2 female, 6 male, 1 left-handed, 7 right-handed) were selected for testing Phase 2 of the prototype. All participants were between 18 to 25 years of age. The experiment lasted 45-minutes on average. Informed consent was obtained from all participants beforehand.

4.3 Phase 1

Phase 1 results were collected primarily in a single dimension: error. The error was measured as the l2norm distance of the car from the centre of the road and timestamped. This data was then interspersed for purposes of anonymity, and is presented in Figure 4.2.



FIGURE 4.2: Phase 1 Error Graph.

From the graph, we can see that the error when playing blind is still within the same order of magnitude as the error with sight. Participants were able to correctly steer the vehicle and complete multiple laps of the track without visual aid. Correcting for the procedurally generated track which has easier and tougher sections, we can see that graphing blind error over non-blind shows a training effect where the error decreases over time⁴.

⁴Note that this is a result of preliminary testing and needs to be confirmed through further experiments.

4.4 Phase 2

In Phase 2, the Fitts' Test is used because it is currently a standard in measuring human-computer interfaces. Fitts' Law is useful in our case, but it is only peripherally relevant to the scope of this paper. Measuring the information-theoretic maximum of a sensory substitution model is something that deserves an exploration in and of itself, and is therefore left to future research. The Fitts' Test however, provides data that can be used by other researchers to compare our data and to compile a growing dataset. It therefore stands to reason that our primary measurements be the same as those in a Fitts' Test. The primary measurements that are common in this modality (MacKenzie and Buxton, 1992) (Chun et al., 2004) are the time taken to reach a target, and the distance from the center of the target when a successful click is made (also called error).

The Fitts' Test was performed in four trials, each comprising of five sessions. Each session represented one completion of the Fitts Test board, with all targets being selected. Of the five sessions, information from the first was discarded and marked as training. Among the four trials, two independent variables were chosen and varied between the trials. These variables were whether the participant had sight (blind vs non-blind), and whether an alert tone (as described earlier) was given (alert vs non-alert).

In addition, the time taken to reach an individual target is also affected by the distance between the two targets. This was corrected by dividing the total time taken for each target by the distance between the previous target and the intended target.

Figure 4.3 shows how the average time to target, corrected for distance, varies across the two parameters. Changing the blind component affects the average by more than two orders of magnitude, which is very different from the results we observed in Phase 1. This is presumably due to the added

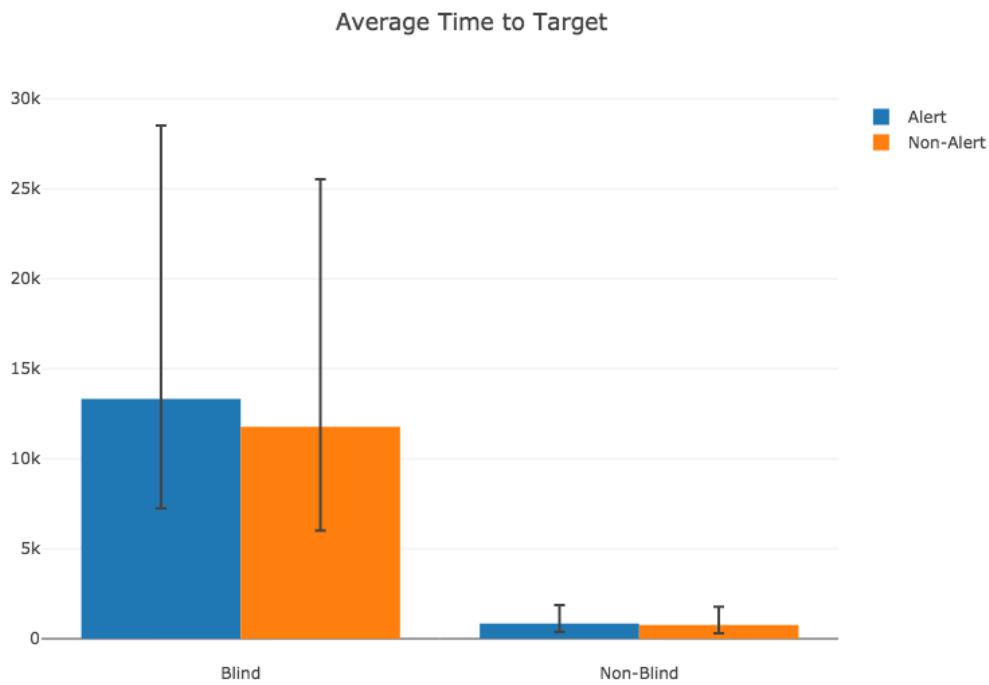


FIGURE 4.3: Average time to target (adjusted for distance).

noise and variation involved in the system, along with the delayed-feedback mechanism at play.

We can also see that there is a lot more variation in the blind trials, and this was observed in practice by the researcher as the participants often vacillated heavily in deciding which direction they needed to move in. What is interesting is that providing an alert notification increases the amount of time taken to hit a target. While counter-intuitive, in practice it was observed (and reported by participants) that the alert notification worked counter to the participant training him/herself to learn the other signals. What was even more surprising was that this remained true even when the participants could see the targets in front of them. The alert tone proved to be a distraction that lengthened training times and possibly prevented the participants from 'feeling' the signals being conveyed.

ANOVA confirms this intuition as we can see in Table 4.1. There is a

Effect	Sum Squares	dF	F	p
Alert	1.391602e+09	1.0	17.028212	3.900152e-05
Blind	2.760588e+10	1.0	337.79681	9.512239e-68
Alert:Blind	1.402356e+09	1.0	17.159801	3.642130e-05
Residual	1.143309e+11	1399.0	-	-

TABLE 4.1: ANOVA results for timing.

strong effect on blind vs non-blind on the time taken to hit a target ($p < 9.6 \times 10^{-68}$). There is also a significant effect from the alert variable ($p 3.9 \times 10^{-5}$).

There is significant interaction between the two variables, but this could be due to a number of reasons. From conversations with participants, the alert variable is also the most affected by latency. Participants reported latency problems with the alert variable far more than they did with the other vectors, and this is understandable since the alert variable is binary while the others are continuous. Non-blind trials tend to have a higher velocity of movement for the pointer than blind, since the participant is immediately sure of which direction to move in and how quickly. My educated guess is that the increased velocity accentuates problems with latency, which then affect the results differently. A participant while blindfolded can be quite sure that he has passed over the target when the alert parameter activates, but he or she is far more likely to be confused if they aren't blindfolded.

Figure 4.4 further illustrates this uncertainty, as we can see that the results vary widely between trials, rounds, sessions and users.

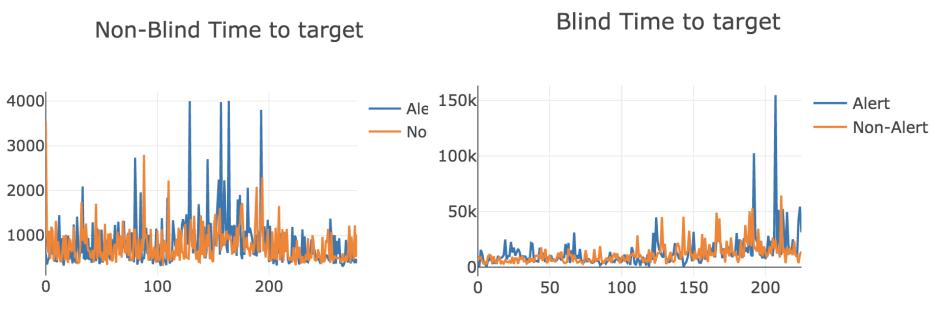


FIGURE 4.4: Time to target per round.

Moving on to the second measurement, the distance from the center of each target to the successful click (which we can call target error) is illustrated across the parameters in Figure 4.5

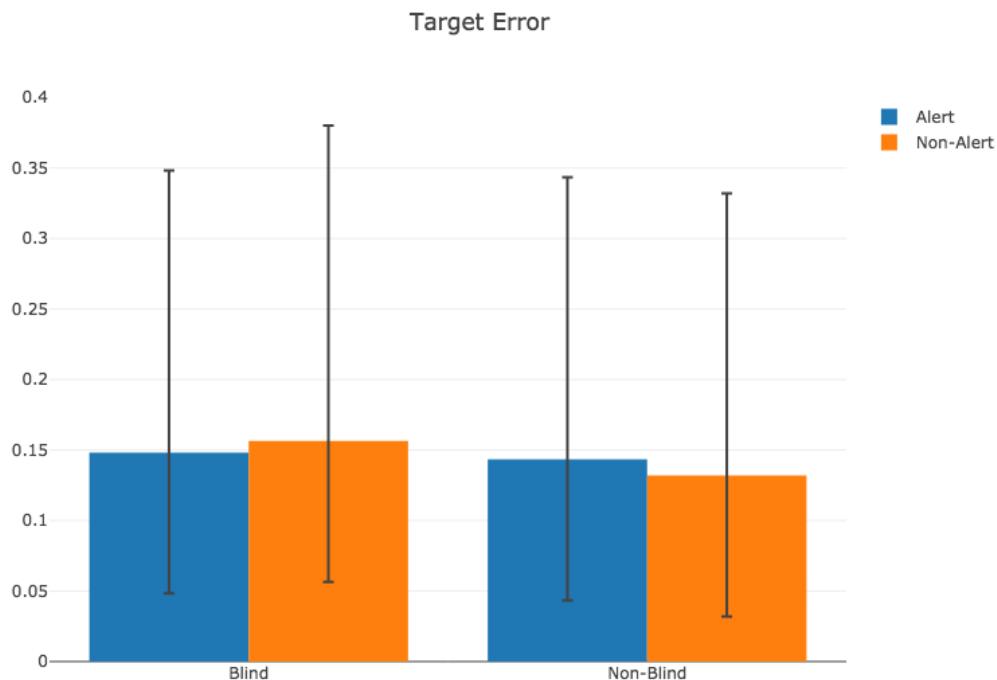


FIGURE 4.5: Average Target Error.

From the graph, it doesn't seem that the presence of the alert variable has a strong effect on the target error. This is surprising as many participants had reported that they felt "less accurate" if the alert was given, since they would click much sooner and closer to the edges of a target. What is also interesting is that the small effect the alert variable has on the cumulative average is affected by whether the participants are blindfolded. This is very likely within the margin of error, as ANOVA (Figure 4.2) fails to report any significance in the effect of the alert variable. It does however, place the effect of blindfolding as significant ($p < 1.0 \times 10^{-4}$, as expected) along with the interaction of the two variables. The true cause of this interaction is unknown, and as in the prior case, further study is needed.

Effect	Sum Squares	dF	F	p
Alert	0.000336	1.0	0.058371	0.809125
Blind	0.086352	1.0	14.995669	0.000113
Alert:Blind	0.038955	1.0	6.764772	0.009396
Residual	7.975467	1385.0	-	-

TABLE 4.2: ANOVA results for target error.

To further investigating the myriad interactions here, the results are shown by individual. Figure 4.6 shows a much clearer picture of how each individual performs against the variables. Here we can see that the communal indicators are clearly misaligned with individual outcomes when the alert variable is introduced. In fact, the choice of individual clearly seems significant in determining how the alert variable affects the results.

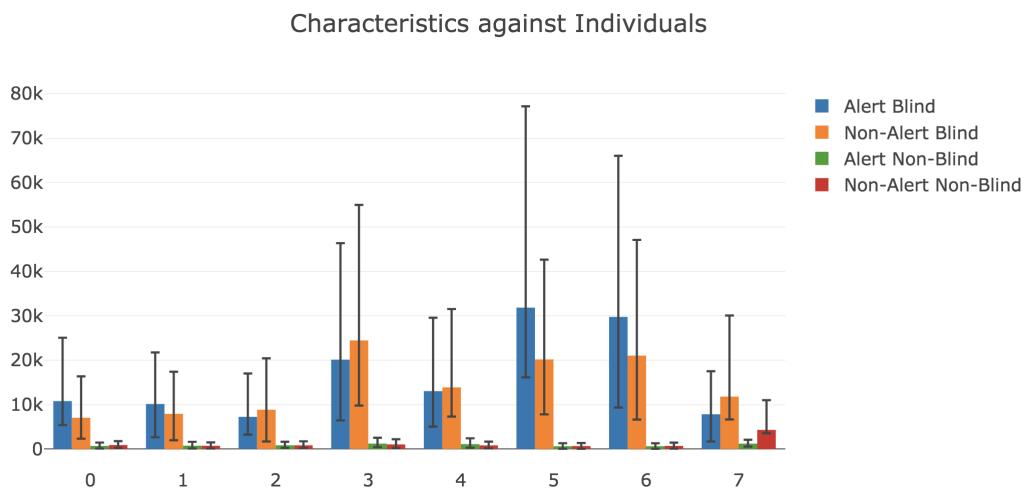


FIGURE 4.6: Timing and variables grouped against individuals.

4.5 Discussion

While a number of insights can be gleaned from the results gathered here, the strongest one is that increasing the number of dimensions merits a scaling-up of the sample size almost as much, perhaps even more so than the number of independent variables being modified. The uncertainty involved in human

learning makes evaluating the effectiveness of more complex systems a challenging task. In reference to the related work that we have covered, it is clear that a mathematical model for determining sample size *relative* to the complexity of the transmission protocol needs to be developed. However, the nature of the tasks used in the experiments here, which modify feedback latency, number of vectors, overall bandwidth and the measurements therein, present a good framework against which to evaluate future work.

That being said, a few insights found in conducting the experiments that explain the behavior of the system can be laid out.

4.5.1 Training

In a sensory substitution module that aims to use neuroplasticity as a way of reconfiguring senses, much importance is placed on the participants being able to intuitively grasp the information being provided. In both the experiments conducted, I was surprised by the effect training time had on the performance of an individual, and in unexpected ways. Person 0 (from Figure 4.6) was aquainted with the experiment for the first time, and had to leave within a minute of starting due to scheduling conflicts. When they returned the next day, their performance had remarkably improved despite there being no conscious recollection of trying to improve their score. Likewise, participants that spent the first of the randomized sessions without access to the alert variable tended to perform much better than those that didn't. As in Phase 1, there seems to be a direct correlation between training time and performance, and while this seems obvious from the premise none of the related work that was consulted seemed to make use of or measure this parameter.

In addition, a lack of training time can lead to haphazard results that defy analysis. The first minute or two of most participants' time (well beyond the session that is often allocated for 'getting used to' an experiment) was spent

devising strategies to improve performance. For example, one of the participants preferred to circle the pointer in ever-tightening concentric circles, preferring to feel for the shift in vector direction across the axes as opposed to identifying and following them. However, given long enough all participants began to develop and intuitive sense for the information and saw marked increase in performance.

In my opinion, there appeared to be a battle of the conscious desire to remember and strategise a good solution, against the intuitive learning that took place given enough time. However, this merits further discussion and study.

4.5.2 Active vs. Passive information flow

Something that seemed to influence the participants was the active nature of the feedback being given. Comparing across both experiments, Phase 1 utilized a passive information approach whereby following the right path resulted in no response from the system, where Phase 2 did the inverse. From conducting the experiments, participants reported to have a stronger preference for a system that only provided corrective input as opposed to a surplus of information. This was an unexpected discovery that I believe can aid in the development of sensory substitution systems, where a passive approach is often overlooked.

4.5.3 Demographics

Finally, such experiments could benefit from tighter control over the demographics. While this increases the effort on the researcher's part, variation across the participants in age, build, gender and even mental state seemed

to have a strong effect on the responses garnered therein. From prior experience in Human Computer Interfaces, this was much more pronounced here than in interfaces that required only a conscious level of learning.

4.5.4 Evaluation

From the measurements made so far, we can compile a useful metric for evaluating the performance of such a system. The salient points of such a metric would include:

1. Tests that can be replicated across various platforms and interfaces, such as the Fitts' design and driving simulation presented here. Both tests can be extended to the real world, or subsequently made more complex to evaluate maximum possible bandwidth of a system in virtual space.
2. Identification of the properties of such a system, which include the type of feedback loop being created, the encoding being used and its ability to be used as a generic transport for multiple forms of data, as well as the possible mapping and domain space of output and input information.
3. A testing paradigm that focuses on improving the training and testing performance of the system as opposed to just test performance.
4. Additional methods for measuring loss/failure rate of the system at the high resolution. An example would be the measurement of drift/error in Phase 1, and the adjusted timing metric used in Phase 2.

4.6 Further Research

Due to the limited resources and scope of this paper, there is much to be desired in further work. Understanding of the cognitive cycles of learning and intuition seems to hold the key to building high-bandwidth systems that can replace or enhance entire senses. As a next step, experiments in understanding the relationship between individual makeup, training time and training modality might be in order. Longer training sessions with intermittent testing periods can help measure the change as the brain comes to accept the new sense and begin to process the new reality.

The design framework, experiment structure and measurement protocols described herein were built with extensibility in mind. Future work could involve building on top of the encoding protocols to facilitate better cognitive acceptance of the new sense, as well as using standard tests such as Fitts' across multiple interfaces to compare and contrast possible best practices.

Finally, within the modular components of the framework lie the possibility of integrating some form of learning algorithm to tailor responses in real-time. Participants responded best when a continuously varying feedback loop was established, and the measurements laid out in the results section can be used to fine-tune the signals in a way that this loop is made stronger. While this is definitely outside the scope of this work, it is what seems to me the natural progression of the science.

4.7 Conclusion

In this paper, new frameworks, metrics and experiments for performing sensory substitution were explored in a way that made use of new advances in technology. The investigations demonstrated the need for more frameworks

that can be generalized towards conveying information as opposed to being successful at a specific task. The experiments found that the effects were strongly related to the dimensionality of input data, as well as the particular encoding being used.

The experiments were successful in developing a complete framework that enabled users to replace—for the purposes of spatial navigation in two and three dimensions—an existing sense in a manner that lowered cognitive load and the possibility of sensory overload. This has applications in Robotic Surgery and information dense environments such as combat fighters, MMORPGS, and data processing.

The paper also put forth a number of metrics that expanded upon existing literature in a way that makes future work easier to compare and build on top of. Of the configurations tried, successful configurations by way of hardware design and software encoding were found that best suited sensory uptake, as well as provide a modular platform whose results pointed to continuous-feedback based haptic systems as a way of replacing and enhancing the *umwelt*.

Appendix A

Supporting material

Interactive versions of the graphs used in this paper can be found at <https://plot.ly/toonistic/>, along with the datasets used.

The code used in this experiment¹ along with detailed datasets and debug logs can be found at <https://github.com/hrishioa/goosebumps>.

A video of the experiment setup for Phase 1 is located on Youtube at <https://youtu.be/5U9C2m8akOc>.

¹Strictly of research quality and to be considered at best an alpha release.

Bibliography

aivijay (2017). *lsd_slam_noros: LSD-SLAM*. original-date: 2015-05-10T01:41:56Z.

URL: https://github.com/aivijay/lsd_slam_noros.

Akhter, S. et al. (2011). "A Smartphone-based Haptic Vision Substitution system for the blind". In: *2011 IEEE 37th Annual Northeast Bioengineering Conference (NEBEC)*, pp. 1–2. DOI: 10.1109/NEBC.2011.5778565.

Bach-Y-Rita, Paul (2003). *Seeing with the Brain: International Journal of Human-Computer Interaction: Vol 15, No 2*. URL: http://www.tandfonline.com/doi/abs/10.1207/S15327590IJHC1502_6?journalCode=hihc20.

BrainPort V100 Vision Aid. URL: <https://www.wicab.com>.

Chun, Kwonsoo et al. (2004). "Evaluating haptics and 3D stereo displays using Fitts' law". In: *Proceedings. Second International Conference on Creating, Connecting and Collaborating through Computing*. Proceedings. Second International Conference on Creating, Connecting and Collaborating through Computing, pp. 53–58. DOI: 10.1109/HAVE.2004.1391881.

Doidge, by Norman (2008). "Brain That Changes Itself: into the abyss". en-GB. In: ISSN: 0307-1235. URL: <http://www.telegraph.co.uk/news/health/3355721/Brain-That-Changes-Itself-into-the-abyss.html>.

Frequency modulation (2017). en. Page Version ID: 808647064. URL: https://en.wikipedia.org/w/index.php?title=Frequency_modulation&oldid=808647064.

Hawkins, Jeff and Sandra Blakeslee (2007). *On Intelligence: How a New Understanding of the Brain Will Lead to the Creation of Truly Intelligent Machines*. en. Google-Books-ID: Qg2dmntfxmQC. Macmillan. ISBN: 978-1-4299-0045-4.

- Hoefer, Steve. *Meet The Tacit Project. It's Sonar For The Blind.* URL: <http://grathio.com/2011/08/meet-the-tacit-project-its-sonar-for-the-blind/>.
- Khambadkar, Vinitha and Eelke Folmer (2013). "GIST: A Gestural Interface for Remote Nonvisual Spatial Perception". In: *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*. UIST '13. New York, NY, USA: ACM, pp. 301–310. ISBN: 978-1-4503-2268-3. DOI: 10.1145/2501988.2502047. URL: <http://doi.acm.org/10.1145/2501988.2502047>.
- Kolb, Bryan and Robbin Gibb (2011). "Brain Plasticity and Behaviour in the Developing Brain". In: *Journal of the Canadian Academy of Child and Adolescent Psychiatry* 20.4, pp. 265–276. ISSN: 1719-8429. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3222570/>.
- Kristjánsson, Árni et al. (2016). "Designing sensory-substitution devices: Principles, pitfalls and potential". In: *Restorative Neurology and Neuroscience* 34.5, pp. 769–787. ISSN: 0922-6028. DOI: 10.3233/RNN-160647. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5044782/>.
- M. Fitts, Paul (1992). "The Information Capacity of the Human Motor System in Controlling the Amplitude of Movement". In: *Journal of experimental psychology. General* 121, pp. 262–9. DOI: 10.1037/h0055392.
- MacKenzie, I. Scott and William Buxton (1992). "Extending Fitts' Law to Two-dimensional Tasks". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '92. New York, NY, USA: ACM, pp. 219–226. ISBN: 978-0-89791-513-7. DOI: 10.1145/142750.142794. URL: <http://doi.acm.org/10.1145/142750.142794> (visited on 03/26/2018).
- Nau, Amy C. et al. (2015). "Acquisition of Visual Perception in Blind Adults Using the BrainPort Artificial Vision Device". In: *The American Journal of Occupational Therapy* 69.1, 6901290010p1–6901290010p8. ISSN: 0272-9490.

- DOI: 10.5014/ajot.2015.011809. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4281706/>.
- Niehorster, Diederick C., Li Li, and Markus Lappe (2017). "The Accuracy and Precision of Position and Orientation Tracking in the HTC Vive Virtual Reality System for Scientific Research". In: *i-Perception* 8.3, p. 2041669517708205. ISSN: 2041-6695. DOI: 10 . 1177 / 2041669517708205. URL: <https://doi.org/10.1177/2041669517708205> (visited on 03/26/2018).
- Out Run* (2017). en. Page Version ID: 809908534. URL: https://en.wikipedia.org/w/index.php?title=Out_Run&oldid=809908534.
- Phillips, Betsy and Hongxin Zhao (1993). "Predictors of Assistive Technology Abandonment". In: *Assistive Technology* 5.1. URL: <https://cs.brynmawr.edu/Courses/cs380/fall2010/readings/Phillips1993.pdf>.
- SE, SE. *perception - Why is sensory substitution not that successful - Cognitive Sciences Stack Exchange*. URL: <https://cogsci.stackexchange.com/questions/9135/why-is-sensory-substitution-not-that-successful>.
- Simultaneous localization and mapping* (2017). en. Page Version ID: 807615182. URL: https://en.wikipedia.org/w/index.php?title=Simultaneous_localization_and_mapping&oldid=807615182.
- Stronks, H. Christiaan et al. (2016). "Visual task performance in the blind with the BrainPort V100 Vision Aid". In: *Expert Review of Medical Devices* 13.10, pp. 919–931. ISSN: 1743-4440. DOI: 10 . 1080 / 17434440 . 2016 . 1237287. URL: <https://doi.org/10.1080/17434440.2016.1237287>.
- Upson, S. (2007). "Tongue Vision". In: *IEEE Spectrum* 44.1, pp. 44–45. ISSN: 0018-9235. DOI: 10 . 1109/MSPEC . 2007 . 273043.
- White, Benjamin et al. (1969). "Vision Substitution by Tactile Image Projection". En. In: *Nature* 221.5184, p. 963. ISSN: 1476-4687. DOI: 10 . 1038 / 221963a0. URL: <https://www.nature.com/articles/221963a0>.

Yan, Qinrui (2017). *Contribute to ORB-SLAM-Android development by creating an account on GitHub.* original-date: 2016-03-30T23:11:59Z. URL: <https://github.com/castoryan/ORB-SLAM-Android>.