FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Experiment in Human-Computer Interaction - Evaluation of a Shamanic Interface for Interaction with Cultural Gestures in Virtual Environments

Pedro Pais de Sousa da Costa Carvalho

FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

Mestrado Integrado em Engenharia Informática e Computação

Supervisor: António Coelho

Co-Supervisor: Leonel Morgado

January/2019

Experiment in Human-Computer Interaction - Evaluation of a Shamanic Interface for Interaction with Cultural Gestures in Virtual Environments

Pedro Pais de Sousa da Costa Carvalho

Mestrado Integrado em Engenharia Informática e Computação

January/2019

Abstract

Gesture controlled applications provide a way to streamline interaction between human and computer that deviates from traditional interfaces. While benefits are present, the lack of familiarity of users imposed with novel implementations of so-called Natural User Interfaces may however provide hurdles to their adoption, particularly on systems of higher complexity or for those who present physical impairments.

A Shamanic Interface is a proposal for a semantic bridge between Gesture Recognition and the interpretation and execution of application commands. The intent is to support a customized interaction experience based on gestures already existing within the user's culture. These Cultural 'Emblems' are actions uniquely meaningful within an anthropological context for communicating concepts, and may provide significant benefits against a less culturally charged approach, be it resorting on mimicry of a command, or even imposing of a foreign or disjointed metaphor upon the user.

Being a fairly recent proposal, the Shamanic Interface concept requires further validation to assess the viability of its approach in ensuring easier adoption, higher degree of immersion or the facilitation of more advanced interaction. This work focused on three proclaimed aspects of the postulated approach: Learning Rate and Capacity; Retention and Memorization; and Satisfaction and Immersion, to test the impact of using a Shamanic Interface along these dimensions. This was done by developing a virtual environment where interaction opportunities arise, followed by user questionnaires procedures will be conducted in controlled test environments. Data obtained from these found conclusive evidence to support the former two assertions and suggests methods by which to investigate the last.

Keywords: Shamanic Interface, Gestural Commands, Natural Interaction, Human-Computer Interaction

Acknowledgements

It’s with great esteem that would thank my professors and other people at the labs who helped me with guidance and counceling, and who have provided me with material and lab space. To all the amazing volunteers who took moments of their busy lives and schedules to allow the continued materialization of my work. To my colleagues and friends with whom I’ve shared half a year of chance and empathy. To friends and family in their continued efforts and support during the process of this project. And to Ana, for all the surprises.

Pedro Pais de Sousa da Costa Carvalho

"Why on earth did you make those wings?
You don't seriously think they could let your ferret fly, right?"

Randall Munroe

Contents

List of Figures

LIST OF FIGURES

List of Tables

Chapter 1

Introduction

1.1 Context

Body Motion and Gesture recognition has seen an interesting development with modern technology. We already now see it employed with many forms of sensors and applications. The most common of which includes the touchscreens found on smart devices, which became one of the basic innovations that would shape up the future of interaction between end-user and machine, as well as user expectations for current tech.

Any kind of movement performed by a human can be classified as a gesture, but not all gestures can be considered 'natural'. Natural, in this context, is assumed to be a set of behaviours that come effortlessly and intuitively to the users, in a way they may not even consciously acknowledge. This is an idea that was taken in account by Steve Mann when he introduced the Natural User Interface concept [1]. A NUI is an interface built with organic experience as its primary goal, with which the user should find a higher degree of freedom to explore it without the limitations of the technology surprising the user's anticipations or therefore damaging the ergonomics of the interaction. Thus, it focuses on human factors, the environment and senses a person relies on. Ideally, the interface itself should be effectively invisible to the user, even as they learn to perform more complex interactions with the system. The NUI designation is also later presented as an evolution of interaction paradigms as a whole, following that of the Graphical User Interface[2].

Surrounding the turn of the 2010's decade, a lot of research and development was done into NUIs, particularly in the field personal computing and entertainment. Since the release of the Wii Remote and of the Xbox's Kinect, the gaming industry had an arms race for new true-to-life interactions methods[3], meanwhile, on mobile, accelerometers, gyroscopes, proximity sensors and compasses have become the norm and implicitly expected to be a part of any model's feature set. Besides vision, touch or accelerometer based sensors, examples of applications commonly referred to as NUI may achieve its operation through use of voice recognition, facial expression, gaze direction and biometrics including heart rate or electromyographic sensing. Many devices for each type of input have been developed, even where similar gestures are detected, as an example, the Myo Armband and Leap Motion Controller both register motion of the hands and fingers despite differ-

ent approaches.
Thus, through interest in NUIs, it can be said that for Gesture Interfaces, many have been attempting to create potential standards of interaction that provides larger diversity and scope of use. While NUIs are not predicted to become a predominant form for all future interaction, it's clear they're here to stay and will carve out a mainstay niche. And the same way the GUI has not replaced the Command Line Interface, but rather lowered the barriers of entry for broader use cases and audiences to more complex degrees, the NUI will also not be replacing the GUI, but rather looks to become a facilitator for the scenarios where they do make usage and learning easier.

1.2 Motivation

However, despite given the undeniable relevance of NUIs and their ongoing research towards modern technology, it can't also be undeniably stated that their design philosophy has been sufficiently explored. An opinion that was summarised by Don Norman [4][5] and has since been repeatedly cited, in regards to what he and others felt was the place NUIs currently held during the onset of its surge. He claims that Natural Interfaces are useful, but that they may currently be a misnomer. The discussion further equates their development to the early developments of the GUI, where a lot of actions would be explained through use of metaphors. For the GUI, the popular metaphor that still survives today in vernacular, was that of a work desktop with papers and filing folders strewn about, and a hand to drag and work on them. However, the metaphor was merely a learning aid, and it doesn't directly resemble the actions intended when handling a GUI. For the Natural Interface which purports to better leverage the usage of metaphors, this was not necessarily encountered.
The more remarkable example of a metaphor successfully working as a NUI, but then failing the user expectations was found with a bowling game for the Nintendo Wii system. The Wii Controller is a gestural form of input with buttons on either side allowing users to mimic the motion of grabbing and swinging a bowling ball. Users are supposed to apply pressure on the buttons, and then perform a swinging motion releasing the buttons at its end, analogous to that of using a bowling ball in a real-world environment. However, when players got invested and immersed in the game, occasionally it would be verified that they would also release the controller itself at the end of the swing, throwing the controlled most likely in the direction of the game display. This would take users out of the experience, and they'd then make their plays with more careful and inhibited, yet far less natural, impetus.
The reasons given for the labelled failure of Natural Interfaces is due to them not conforming to the rules, or heuristics, of interaction design that apply beyond the scope of any particular technology[6][7]. Specifically, existing NUIs have issues with the visibility of signifiers and thus also with discoverability of new commands. With the freedom, reliability, feedback and, as seen above, error prevention, leading to users to perform commands they're not even conscious of, and being unaware of how to quickly correct the program after issuing an erroneous change[5][8]. This is the lead up to a lot of scenarios where users either must be taught how to perform certain

commands and are told to perform mimicry of analogue movements, or conversely, of a command that represents a non-kinesthetic concept. This proves to be confusing, and then regularly users find themselves complaining, after extended periods of use, about options that they had no way of knowing existed, of commands not making particular intuitive sense for the application, or even ones that don't seem to work in new contexts with no discernible visible cue to explain the difference. A great deal of concern is given to the need of standards, and exploring the right approach that would actually feel seamless in the hands of users, and also to the difficulty that the behavioural distinctiveness of users presents to either goal.

One the tenants in HCI as a science for its usability concerns, we are told that users should not have to 'radically change to fit in', but rather that 'systems should be designed to match their requirements'[9]. So, one of the solutions presented tried to face the issue regarding ambiguity of input and user spontaneity, was through adaptability. If a technology wishes to allow users to interact with it as they are used to interact with the real world in everyday life, this technology must be malleable enough to each user, and the framing by which it should handle their involuntary suppositions would be their Culture. Culture is rich in gestures and expressions that hold special meaning. Even for concepts that have no physical equivalent and thus can't be simply produced through mimicry of the concept, the depth is such that non-verbal communication is possible. Culture aware systems could provide an answer on not just what is a valid definition of Natural for one user, but separate answers for every group of every ethnicity and upbringing.

As such, there's a need to produce research on this potential of leveraging user culture. Should a methodology prove itself to be feasible, new standards of interaction may be built for it among niches fulfilled by NUI based systems, providing users with more inclusive and immersive experiences, and opening new fields of research. The Shamanic Interface is one among ideas for introducing cultural awareness into systems, focusing on the separation of concerns between gestures classifications and virtual instructions.

1.3 Goals

The main goal for this dissertation is to explore the viability of Shamanic Interfaces as a concept and its application in interaction with virtual environments. Prior work[10], was already accomplished in building a research tool for performing field studies in a controlled environment, as well as verify its playability among a group of users from differing backgrounds. However, the empirical tests still need to be fulfilled, and the tools can still be improved. Some additional concerns were set as future work.

Here, we seek to find empirical insights to the following encompassing hypotheses:

- **Focusing on user culture contributes to the Learning Rate and Capacity of commands**
  One of the very first aspects touched upon the Shamanic Interface proposal was a perceptible learning curve negatively impacting the user experience of Natural Interfaces featuring

commands that scoped beyond of simple mimicry, or complex breadth of instructions and information. The approach to cultural richness is expectable to improve this situation and allow users more freedom thanks to culture having pre-loaded non-kinesthetic abstractions that may be used by designers, as well as permitting the application of semiotics through previously unexplored non-textual signifiers.

- **Focusing on user culture contributes to the Retention and Memorization of commands and concepts**
  Moreover, these benefits and observations are anticipated to produce continuous benefit over the non-cultural uses of natural interfaces, as the interaction would have a longer lasting effect on the user's affectivity, and thus, benefit them in both recollection of content, and recall based on context.

- **Focusing on user culture contributes to the Satisfaction and Immersion of the experience**
  Another concern sat with the receptiveness of users towards a potentially new form of interaction, and how well accepted the naturalism of human gesture when communicating with a machine would be. It is posited that one of the advantages will be allowing users to focus better on the task they wish to perform, rather than on the interface, and thus abstract themselves into the experience with higher degree of sense of presence.

1.4 Thesis Structure

For the remainder of this document each chapter will focus on the background of each of three separate topics, their state of the art and findings that may be relevant.
Chapter 2 will first focus on Culture in HCI and gestures from a cultural perspective, starting with an explanation of the Shamanic Interface proposal, its name. It will shortly reiterate some background knowledge required for the implementaton about Gesture Detection that was relevant to the research tools built for the prior dissertation.
Chapter 3 will cover the planning and development of the required systems and of the user trials. The progression through the chapter will flow roughly after that of the development's work, starting with the reconstruction of the shamanic interface, going through the design elements present in the game, and ending with an overview of the trials, and their goals on each of its phases.
Chapter 4 covers the results and analyisis obtained from the trials. The results will be covered following a similar order to the phases of the user trials, running from pre-test to post-test. Once that's done, the conclusions will summarize how the results fit into the goals posed above 1.3 by this work, and then concludes with some closing words and suggestions for future work.

Chapter 2

Background

2.1 Problem Statement

It's typical for researchers to have a forward-looking aptitude towards their work, and to come up with ideal realities that they wish or find inevitable that will be found fleshed out. This is how in 1991, Mark Weiser described the concept of Ubiquitous Computing[11], a concept that is today taken for granted in the dawn of IoT computing, taking only as a basis the observable trend of miniaturization of electronic components and wireless communication. As such, it's not hard to imagine that the path ahead for Human-Computer Interaction has been broadly discussed, as well as its concerns and criticisms. Gestural Detection technology may appear like a modern innovation of the last decade, but it was sought after for several before, and it took progressive iteration, evolution, to reach this state.

However not all evolutions in a field are continuous, there are points in a macroscopic view of technology where discrete advancements have to be made. This describes the paradigm shifts encountered in the major approaches to HCI. Just like he Graphical User Interface (GUI) came to outperform the Command Line Interface (CLI) in several use cases, now the once predicted Natural User Interface (NUI) shows potentials to outperform the GUI in select applications.

However, just how the initial GUI's were rudimentary, and faced design constraints that were ironed out with the growth of the supporting hardware and with the emergence of technological literacy, so does now the NUI have its own problems. As was further described in the introductory chapter 1.2, the NUI as a concept for Gestural Interfaces is not realized, still has room for cultivation, and the Shamanic Interface proposal intends to contribute through an holistic vision of human communication and computer interaction as a whole.

2.2 Shamanic Interface

The Shamanic Interface takes its inspiration from the science fiction book Freedom, by Daniel Suarez, from which the name is directly lifted. In the fictional plot, the Shamanic Interface is a mechanism by which the characters apply commands to a complex augmented reality system, through use of somatic gestures.

"It's called the shamanic interface because it was designed to be comprehensible to all people on earth, regardless of technological level or cultural background"

The argument suggested by the book was the universality of beliefs in immaterial concepts, and how they may be accessed and communicated through ceremonial and traditional gestures. The futuristic technology described leveraged these rituals, gesticulations, and made it its own form of input, connecting its user experience to acts that almost appear as if magical, where the virtual environments reacts and materializes in accordance to motion in a way that gives a logical and natural impression to anyone. However, while gesticulations and emblematic motion are indeed common to human behaviour, the notion that a single set of gestures that could be understood by all people of the world exists is unfounded and too fantastic for real practical implementation, as seen later in this chapter, namely from the contradictions found within the meaning of same gestures across cultures. For this reason, when Morgado[12] contemplated the solution taking the fantastical interface as a basis, the name "anti-shamanic" was also pondered upon as his proposal subverts and deviates from this conception.
The valuable aspect of focus was the engineering of a system that adapts to the immanent semantics as a form of command for virtual environments. To achieve this, the proposal of the Shamanic Interface required an addendum as such: To decouple the concerns of gestural identification and parameterization, and those of command classification and execution, potentially through distinct software layers. Thus, enabling customized mapping of different collections of cultural gestures onto commands. The ensuing result is that commands in the application layer of a system are independent from the actual motion of the user, and it can conversely make a choice of mappings that best fit the needs and requirements of the user.
In other words, a real Shamanic Interface would be an adaptable system with a customized experience, that allows users to establish links between their learned communal meanings and the application's commands. This is how it is expected for it to tackle the current limitations found in NUIs, where performing those commands usually adopts a mimicry approach. It is expectable that this way, the SI will allow the NUIs to overcome the difficulties of exploration and learning attached to the method. Other problems that this may directly favour NUIs with are the issues of accessibility for users with physical impairments and handicaps who are incapable of performing the current required gestures, such as wheelchair-bound users, and allowing interaction to appear visibly more natural for external observers, thus making usage of this systems more societally acceptable.
Prior work was performed on Shamanic Interfaces[10], including a research paper where a research tool was developed for testing and expanding on the concept of Shamanic Interfaces. The developed application was in a working condition, capable of identifying cultural gestures, however it specifies performing the actual tests as a requirement among future work. In this work we look to give continuity to the development and use of that tool.

2.3 Culture in HCI

Culture is a notoriously difficult term to define. In a more general sense, culture is interpreted in the aspects of a society's artistic productions, their traditions, the system of beliefs held by the populous and their modes of expression. Usually driven by symbols which have values attached, which themselves are regarded thanks to attitude or faith of the community in those values. However the interest in culture is restrained to, from one view in HCI as described by Hofstede[13], the *"collective programming of the mind which distinguishes the members of one human group from another"*. The distinctions between cultures, leads to different approaches to communication, learning, admissibility and work ethic, which in turn, signifies differences in how people will interact with other people, and with technology.
Cultural diversity is a challenge for HCI. The widespread adoption of the GUI is one aspect of what is considered the apparent process of globalization. Businesses outgrow their national boundaries, form a global presence and ship products even to developing countries. Technology itself promotes international growth thanks to the existence of the world wide web, social media, new forms of communication, advertisement and even vectors for outsourced manpower. As such, it's not unusual to find products and services that break the cultural boundaries. This multi-culturalism often proves to be an obstacle towards the usability of the system[14] leading to rejection or slow adoption within certain communities without additional differentiating work applied to it. In recent review works on the subject of culture and HCI, the distinction between visible and invisible attributes or indicators was made, with visible attributes relating to interface design and localization issue, while invisible attributes pertain to Hofstede's theory of cultural dimensions[13][15]. One example of dimension is the Power Distance, which levels how receptive users are to inequalities among other members of a society, a result of individualism[16]. This would be relevant in the case of online gaming, where western audiences find the concept of paying for power unacceptable, but asian players state they find no issue with the practice.
And as such, culture is a highly researched topic in HCI, as no singular approach will exist for an interface because of culture. And this is the obstacle of culturally-aware systems. It is expectable of them to adapt to the users' needs in all stages of HCI design, which means, as opposed to creating different builds for each context, the systems must include cultural models as well as generate adaptation rules. From the visible indicators, these include the presentation of information, the language, the dialog design and the interaction design itself[17][16].
Here, the shamanic interface may be pertinent to the latter, as it models and integrates socio-cultural information relevant to user background.

2.4 Cultural Gestures

One important task to prior to the development of the goals was to answer the question of what gestures are, and how does motion relate to meaning. We know that gestures play a role in communication and learning[18][19], but we seek out a definition of gesture that would provide an understanding of the information by them encoded. The field of gestural study covers this as well as other topics[20], such as their kinetics and shape, however, the typology of consequence, is the relationship between gesture and verbalizations. Gestures can be produced in conjunction with speech, in alternation or replacement of speech, or as their own utterances.

Gestures had prior classifications but were categorized, based past work in the field, in a singular set by Adam Kendom[21] along a continuum of closeness to speech or, opposite to that, to metaphorical allusion. Wherein, the following five types of gestures were distinguished and still today referenced[22][23]: **Gesticulations** are spontaneous movements that accompany speech. These consist of 90% of the gestures performed by humans and are most usually started imperceptibly, often accompanying strokes in the semantic pattern of utterances, usually synchronizing or preceding emphasis and pauses. Their prevalence is universal in human nature, and not necessarily applicable as an act of communication, as the interlocutor in a phone conversation will often perform these gestures as an aid in speech production; **Speech-Framed Gestures** are gestures that are a part of the spoken sentences themselves. These replace words directly. They're no co-expressive the same way gesticulations were, instead replacing a grammatical hole in the message; **Pantomimes** are used within a similar fashion, but these do not aid verbal communication. Epitomized with the description of "Dumb-show", it's any exaggerated mimicry that describes a shape, an action or any other physical concept. **Emblems** which are familiar symbology conventionalized within a specific culture, translating directly to a specific significance. The meanings are vastly diverse, ranging from polite, to less-than-so, such as they're addressed informally as 'quotable gestures'. Despite their innate substance, they can naturally be interweaved with one another or with speech itself as a form of gesticulation. Their meaning is powerful and long lived, such as that many emblems have outlasted their own historical roots, the languages that could describe their meanings they co-existed with. **Sign languages** which are fully fledge lexicons, with their own linguistic structures, including that of grammatical patterns and vocabulary. These do not directly correlate to that of a native language and have evolved out of the necessity of coordination with speech, with the practice of signing and speaking frequently proving to be a disruption for practitioners.

Of these, the ones that most interest our work are the emblems. Culturally charged gestures with a basis on the background their users originate from, such as the thumbs up, or the tongue protrusion. While an accomplished shamanic interface wouldn't restrict itself to simply detecting emblems, these have the strongest link between a natural movement and an inferred interpretation, and thus, are the most promising towards the primary contribution of the proposal. Also, on their own, they function as idioms, potentially having multiple denotations, and as thus, a rigid system would not be able to utilize one without facing eventual need of teaching them to the user, or directly clashing with the expectation of users from different contexts.

In the past work [10], an example list of emblems was researched. The following table 2.1 is a summary of information that was sourced from David McNeill's article[24] on gestures, however, the examples were also themselves further sourced to Desmond Morris' findings on an earlier book[25] from which the accompanying figure 2.1 is sourced. With the prospect of having a more complete set of gestural emblems as a basis for the requirements of the thesis' work, a later dictionary-like publication by Morris featuring over three hundred gestures that could be contemplated as meaningful was obtained.

2.5 Gesture Detection Methodologies

Gesture Recognition and Gesture Classification are two distinct aspects of the final application. Classification pertains to the attribution of meaning to a detected Gesture, and is therefore handled by the cultural layer, as each gesture may have different interpretation dependent on the culture chosen. Recognition is the prior step and is involved in how the system identifies meaningful gestures from among all gestures performed by the human users, in real time.

There are a couple of approaches to handle recognition, between mathematical models and soft computing. Before tackling those, first it'd be important to quickly list the types of gestures that can be recognized and how they're processed. These depend on the supporting instruments, which target different portions of the human body, obtain different sensor stimuli: Electric, Optic, Acoustic, Magnetic and Mechanic. These devices include: Gloves, Body Suits, Optical Trackers among others. Furthermore, vision-based techniques are incredibly varied and have several factors differentiating among themselves, by which broad fields of research and business are formed. The basic structure of a Gesture Recognition controller's analysis involves two main tasks: Segmentation and Feature Extraction, features which are forwarded into a recognition module to build models. Segmentation is the extraction of the limbs of interest from background and determination of its location, while feature extraction is the determination of valuable data and cues among the segments. Not all input methods require the latter step, for example, magnetic sensors.

The Recognition module follows up with different approaches: *Hidden Markov Models* are a process governed by an underlying Markov chain with a finite number of states, and a random set of functions, each associated with each state. The transitions between states are based on probabilities. After each of a discrete amount of time, the system will be in one state and will observe a new symbol that feeds into the functions which will either yield a new state for the system, or output a recognized gesture for the followed chain. The chain involved a lot of mathematical modelling, producing a lot of deterministic integration, however, it can be seen as merely a sequence of observations and states, and thus it received the term "Hidden". *Particle Filtering* or alternatively, Sequential Monte Carlo, are approximations to simulation-based methods. Works by representing probabilities of noisy and partial samples, building a predictive model for the likelihood of following states. The benefit of these over grid-based filters such as conventional Markov models, is these do end up modelling uncertainty. *The Finite State Machine* approach, by which a gesture is ordered by a sequence of states that vary in space and time. Each state is a datapoint of trajectory data, and the gesture is divided by each substantial change in trajectory data, sampled in a 2D space. Addition of gestures is achieved by constructing new FSM models and each gesture is matched to all the deterministic FSM's. This does mean that adapting the system to

Table 2.1: 20 examples of Emblems, their meaning and distribution

Sign Name	Cultural Background	Meaning
Fingertip Kiss	Holland, Belgium, Yugoslavia and Turkey	Praise
	Portugal, Sardinia, Malta and Corfu	Salutation
FingerCross	England, Scandinavia, parts of Sicily, and Yugoslavia	Protection
	Corfu and Turkey	Breaking a friendship
The Nose Thumb	Everywhere	Mockery
The Hand Purse	Italy, Sardinia and Sicily	Query
	Portugal, Greece and Turkey	Good
	Belgium and France	Fear
	Holland and Germany	Emphasis
Cheek Screw	Italy, Sicily and Sardinia	Good to eat
	Germany	Crazy
Eyelid Pull	Italy	Watch out, be alert
	Everywhere else	I'm alert
Forearm Jerk	North America	Italian Salute
Flat-Hand Flick	Belgium, France, Italy and Greece	Beat it
Ring		OK
	Tunisia	Threat
	Brazil	Insult
	Germany, North Italy, Northern Sardinia and Malta	Orifice
	Belgium, France and Tunisia	Zero
Vertical Horn	Spain, Portugal and Italy	Cuckold
Horizontal Horn	Malta and Italy	Protection
Fig		Sexual comment
		Insult
Head Toss		Negation
		Beckoning
Chin Flick		Disinterest
		Negation
Cheek Stroke		Thin and ill
Thumb Up		OK
Teeth-Flick		Nothing
		Anger
Ear Touch		Effeminate
		Watch out
Nose Tap		Secrecy
		Insult
Palm-back V-sign		Victory
	Britain	Sexual insult

Figure 2.1: Representation of emblematic gestures

more gestures requires incremental computational power, as well as adding winning criteria between gestures to choose the most likely when multiple are matched. *Soft Computing*, which is a number of techniques which involve computational intelligence and computer learning with high degree of tolerance for imprecisions and uncertainty. The system can be trained even while in use and adapts to users. Methods include fuzzy logic, genetic algorithms, artificial neural networks. These, however, require a large amount of data and iterations to find adequate robustness.

2.6 HCI: Goals and Definitions

Human Computer Interaction is a term that gave rise in the 80's to describe a then emerging field of study. Jenny Preece describes this in her book[26] as rising acknowledgement that the focus of interest of interface designers was no longer just that of the design of the interface itself, but also to the aspects that relate to the interaction between human and computer. She further names the primary three concerns and branches of study of HCI: The design; the evaluation; and the implementation, of interactive computing systems. So rather than being considered a singular discipline, HCI is often thought of as a joint effort from multiple areas of research, as each concern can benefit from contributions arriving from various sources.

The escalation of interest HCI came as a reaction to the equally explosive appearance of new challenges that beset it. Technology advances rapidly, and with it, so did the rate at which new user experience opportunities with their own collective of hindrances to cohesive and proper interaction. Problems such as, the broadness of audiences and environments, users of different ages and outdoor and public settings, such as the range of activities enabled by the hardware, such as the forms of information and physical objects that hardware can transmit.

And thus, HCI sets for itself the purpose of matching the needs and requirements of users in through careful consideration during the design stages. The HCI experts must ponder about the organizational and social factors, about appeal and about efficiency and effectiveness of the interaction in order to assemble a valuable system, with the core belief that users should not emph"radically change to fit in". It set goals, and set principles, which were found to be taken as the artful and professional process by which good design is put to use. In better understanding of what makes applications emph"interactive, instructional and effective", Preece names the goal of HCI as the design of emph "computer systems that are safe, efficient, easy, and enjoyable to use as well as functional"[26].

This obviously involves a lot of discussion. Lessons relevant to HCI cannot simply be reproduced from related established, for example from cognitive psychology [27] and requires basic research on the new domains HCI is tackling, as a whole range of different factors affects the learnability, accessibility and memorability of an interactive system that might have not other fields. And those factors can be messy and hard to track. From the human side of interaction, we have communication theory, cognitive psychology, linguistics, socio-cultural background, meshing with the machine's choice of graphical presentation, operating design, physical input, all of which further coupled with the effect of an uncontrolled and unpredictable, distracting environment, as is normal to the case of personal computing. This leads to a lot contested definitions in HCI, as was the exemplified above on the section about culture 2.3. Despite clear and valuable conceptual developments, the above does contribute to its over-reliance on empirical data and usability as a metric[28][27].

2.7 Evaluation

Again, looking through Rogers and Preece's work[29], one finds four common evaluation paradigms used in HCI research: *"quick-and-dirty" evaluation, usability testing, field studies and predictive evaluation*. Among them, the one that provide the most interest for the desired findings are field studies (see fig 2.2). Field studies are performed with the aim of better understanding how users behave naturally and how a technology impacts them. The goals of a field study include finding opportunities for innovative technology, determining requirements for designs, facilitating introduction of paradigms and evaluating the natural assimilation of the products.
In practice, a field study involves two primary techniques: Observing the user; and asking users their opinions. Testing user performance is also often relevant as an observation, however that is not necessarily a part of the process. Observation should be performed while utilizing support tools and technologies, such as taking notes, recording audio and video, keeping interaction logs of the experiments. The evaluator should not disturb the users during the observation procedure, as a person's presence may be considered obstructive. As for asking users what they think, it should be done separately from evaluation, and can be done through an interview or a questionnaire.
One of the more important tasks for the planning aspect, is to identify what will be the practical issues facing the testing environment and the participants, and plan around them to ensure that every user is faced with the same conditions, and that results obtained pertain to the test's design and not to other factors. These issues may be: The design of each task, choosing users of interest, preparing and recording the test conditions, and choosing how to run the tests. Establishing requirements and objectives early allows an easier decision of the remainder of the important aspects, such as establishing dates, times and places, gaining contacts, and choosing what recording techniques will be employed based on available equipment. For the testing stage, a couple focus point are prevalently relevant, such as preparing and photographing the controlled environment with minimal distractions, obtaining consent of recording the interaction, writing down observations as they occur and user answers as is without personal flair, and attempting to be as unobtrusive as possible. Analysis should always occur later, but any observations that seem peculiar must be noted down as they happen and clarified with the users after the session is completed, to obtain some insights. It's also important to perform a pilot observation session (real or not) to get a feel of what to expect and test out any observation sheets.
Without having a better feel for what's the expected cultural background of voluntaries, or without having a clear schedule for the environment, the questions of who and how cannot be objectively answered. However, the flow of the sessions can be pre-emptively decided. Due to the need of

evaluating the capacity and retention of gestural commands in the system, it makes sense to employ more than a single session per user. As such, the study will be performed in two phases, an initial longer one and a second, shorter one focused more on user performance. Each session will have 5 parts to it: The *Introduction*, wherein the user is welcomed to the controlled environment and recording consent issues are handled; a *Warmup* where users are introduced to the tool and asked about their familiarity with it, and are given some easy instructions that may be required for setup; the *main session*, during which they will perform their tasks with lesser involvement from the overviewing researcher and will have to perform tasks in escalating complexity and difficulty; *Cool Off* period, during which they are interviewed about their experience, their difficulties and some more confusing aspects of the session will be clarified; and finally, a *Closure* in which the recording is stopped and the user is thanked and led out of testing. The second of the two phases will have a much shorter warmup session, and the researcher will have much lesser involvement in the main session, as users should by then be more aware of what is the flow of the program and what tasks they should perform.

Evaluation paradigms	"Quick and dirty"	Usability testing	Field studies	Predictive
Role of users	Natural behavior.	To carry out set tasks.	Natural behavior.	Users generally not involved.
Who controls	Evaluators take minimum control.	Evaluators strongly in control.	Evaluators try to develop relationships with users.	Expert evaluators.
Location	Natural environment or laboratory.	Laboratory.	Natural environment.	Laboratory-oriented but often happens on customer's premises.
When used	Any time you want to get feedback about a design quickly. Techniques from other evaluation paradigms can be used–e.g., experts review software.	With a prototype or product.	Most often used early in design to check that users' needs are being met or to assess problems or design opportunities.	Expert reviews (often done by consultants) with a prototype, but can occur at any time. Models are used to assess specific aspects of a potential design.
Type of data	Usually qualitative, informal descriptions.	Quantitative. Sometimes statistically validated. Users' opinions collected by questionnaire or interview.	Qualitative descriptions often accompanied with sketches, scenarios, quotes, other artifacts.	List of problems from expert reviews. Quantitative figures from model, e.g., how long it takes to perform a task using two designs.
Fed back into design by . . .	Sketches, quotes, descriptive report.	Report of performance measures, errors etc. Findings provide a benchmark for future versions.	Descriptions that include quotes, sketches, anecdotes, and sometimes time logs.	Reviewers provide a list of problems, often with suggested solutions. Times calculated from models are given to designers.
Philosophy	User-centered, highly practical approach.	Applied approach based on experimentation, i.e., usability engineering.	May be objective observation or ethnographic.	Practical heuristics and practitioner expertise underpin expert reviews. Theory underpins models.

Figure 2.2: Characteristics of different evaluation paradigms

Chapter 3

Development

As previously stated, this work is a continuation on the thesis by Tiago Susano Pinto[10]. The goal of their work was the creation of a research tool that utilizes the shamanic interface concept, so empirical research can be performed with it.

The developed software materializes the shamanic interface concept through its cultural layer, which is responsible for generating classifiers. These are through which the application will interchange gesture and command information. The cultural layer achieves this by storing relations between sequences of gestures, culture and meaning, and linking these to gesture models based on chosen culture. The models themselves are built following a Hidden Markov Machine approach, which in turn is available through the Accord.NET Framework for supervised learning. Models were recorded from hundreds of example gestures. As for the playable environment itself, it was built in Unity and it is controlled using a Leap Motion controller exclusively, which detects hand gestures performed by the user.

So, in summation, the work involved developing three primary independent necessary components: The Shamanic Interface itself, which is implemented as library. A gesture recorder which is used to collect hand gesture data. And finally, a game volunteers are supposed to complete during their trials.

During development of this work, all these components have been changed, either in small amounts for compatibility or feature addition, or as a complete rewrite and redesign. Additionally, completely new work was performed for the purposes of the user trials.

3.1 Methodology

The primary focus and methodology used in this work centres around a user trials in a controlled environment with a piece of immersive interactive content featuring hand movements. The main research goal is answering a number of questions regarding the approach and usage of a Shamanic Interface as a core building block of such kind of interactive content. These questions have not been previously answered due to the lack of prior research, and thus the intent is to formulate conclusive or precursory insight towards the realization of more complex implementations or systems

for further research or prototype use.
Briefly, the proposed increased benefits of enabling culturally enriched meaning through gestures by use of a Shamanic Interface, and the goals of this research, are the following:

- Learning Speed and Capacity.
- Richness and Depth of interaction
- Retention and Memorization of commands and concepts.

It's obvious that a diversity of domains is being looked at, such as psychology, culture and User Experience. Additionally, given that some or all of these benefits are already well-documented perks of Natural User Interfaces, which may cause trouble in evaluating the exact impact that is specific to the SI. Furthermore, a useful indicator of positive effect from one aspect of a user may be correlated with another n such a way that dissociating the two will lose the compounded benefit, which is a large concern in UX, such as visual and audible coherence.
It's important to find a method by which to distinguish the baseline benefit of a novel and immersive experience, and those specific to the application of the Shamanic Interface. Thus, it was reasoned that the user trials would be divided upon two groups. The first group would perform a culturally enriched experience similar to the previous works' game, grounded on completing tasks in an openly explorable virtual space. The second group would have to perform the exact same experience, however, the aspects that made the first culturally enriched have to be removed, without this affecting other variables, such as for example, the choice of gestures would still need to make a semblance of logical conveyance for the intended meaning, just not one that would be the person's choice within the culture.
There're more concerns and even risks involved with making assumptions involved with the trials. One such assumption was with that of the designation of a Cultural Experience, and its broad meaning in literature review. To tackle this, it was necessary to include segments of interviews with the users at different timings of the trial dedicated to the problem. A semi-structured pre-test interview with exploratory intents, and a structured post-test interview handled step-by-step by the researcher following a protocol. Including these surveys, lengthens the total time required to complete a trial, which added to the expectation of an instructional step towards teaching the game to the volunteer, leads to the potential of apathetic exhaustion as a factor.
Additionally, and separate from those issues, one of the required questions relates to the measure of memorability of the experience. This prompts the demand of a follow-up experience and, or, interview that that all volunteers partake in after a given delay of at least two weeks. This double phased trial allows determining the long-term impact given to the users, as well as verifying or inquiring about other behavioural indicators that may prove relevant.
Thusly, changes to the Shamanic Interface and the development of the game were made with the early interest in mind with that a short, repeatable experience with swappable commands would be required. Before coming back and describing the specifics of the trials, this document will first go through the changes that the technical components of the implementation. So many of these changes may incur reasoning that relates to usability or the UX aspects of the application.

3.2 Shamanic Interface

Speaking broadly, the Shamanic Interface acts as a module responsible for bridging the tasks of gesture detection and their meaningful interpretation. Towards that purpose, past work identified a number of requirements that must be met and a suggested architecture that meets them, and then followed them in implementation. One such requirement is it should have a Logic Layer that not just carries out gestures and commands as data structures, but also performs the mapping between the two. These data pairs would then in turn have an added dimension corresponding to each culture types that the system is aware of. This core component that handles the data mapping logic is called the *Cultural Layer*. Booting a Shamanic interface into an application involves creating Cultural Layer and keeping it loaded until it generates the next necessary component.
The input the cultural layer takes to function are a set of gesture models, a list of commands and meanings expected from them, and finally, a selected user culture. With these, it produces the Shamanic Interface's other core component, a *Classifier*. The Classifier's purpose is to take gesture data as input, and output what command best fits the gesture performed. For all intents and purposes, the Classifier may act as an abstraction black box for the two steps of interpreting the data's best fit with one of the internal gesture models, and then the reverse lookup between that model and the selected culture towards a given command, the latter being the output.
With this architecture setup, two major points of communication between the Application's logic and the Shamanic Interface's logic are defined. The first is during the Shamanic Interface's Initialization process, which occurs a single time when it loads its pre-built Gesture Model files and receives a list of commands required and a choice of culture. The other moment is a repeatable event, the Classification process, during which data is fed to the Classifier and it outputs the expected command.

Previous work defined an application that followed this basic architecture. However, this operation is insufficient for the purposes of the present work's design. One major change in the requirements is the need for multiple classifiers during completion of the game, hence, the ability for the Shamanic Interface to react to context switching, such that gestures that previously should be recognized as commands no longer should be recognized in the same manner, if at all.
This is where the concept of *State* is introduced. States are a smaller selection of Commands which will generate a different Classifier. Instead of being discarded from memory, the Cultural Layer is always kept in memory, ready to generate new Classifiers on demand. There's a number of advantages to this approach. Compared to receiving commands but ignoring it at application level, there are benefits to classification accuracy in generating a classifier with only the fewer actions required. This also provides a larger degree of potential adaptability to the interface based on implementation, given that it provides a method by which the application can fall back to alternate classification. With this change, an additional stage of communication is added between the Application and the Shamanic Interface.

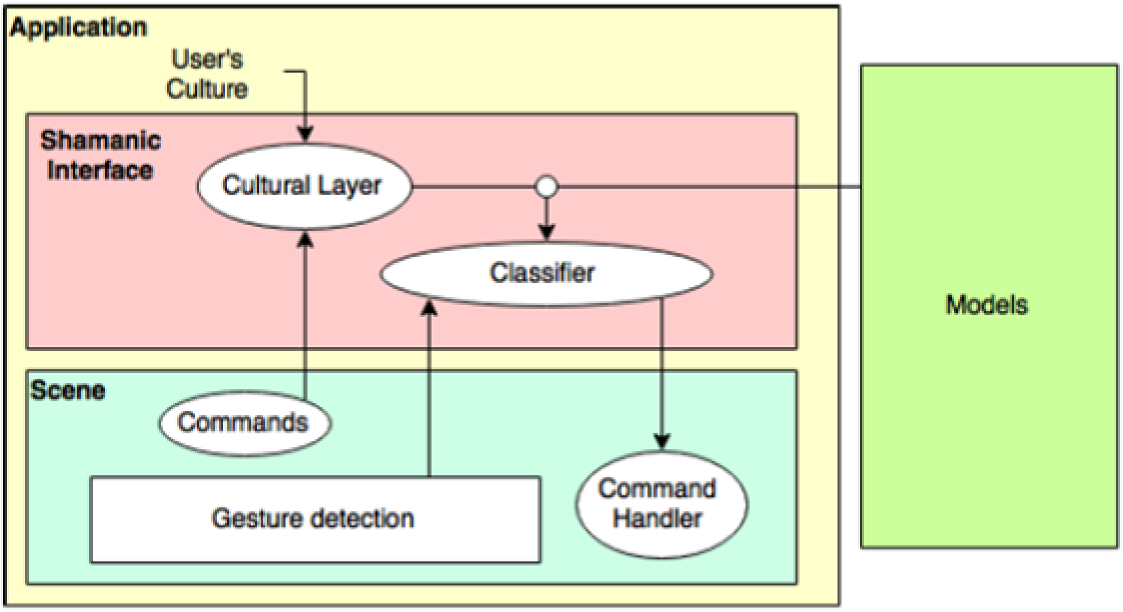

Figure 3.1: Basic Architecture of a Shamanic Interface

3.2.1 Gesture Classification and Recording

The method used for Gesture Classification is specific to each implementation of a Shamanic Interface. Optimally, classification and type of models used would be an entirely decoupled and independent process so the interface can be more adaptable to new solutions. Nevertheless, given the scope of the present work being very small and, while the Shamanic Interface in its state, given the separation of roles between the cultural layer and the classier, could easily be further refactored to make use of alternate models, the module employed makes use of a single gesture detection method and would require further work to generalize its model usage.

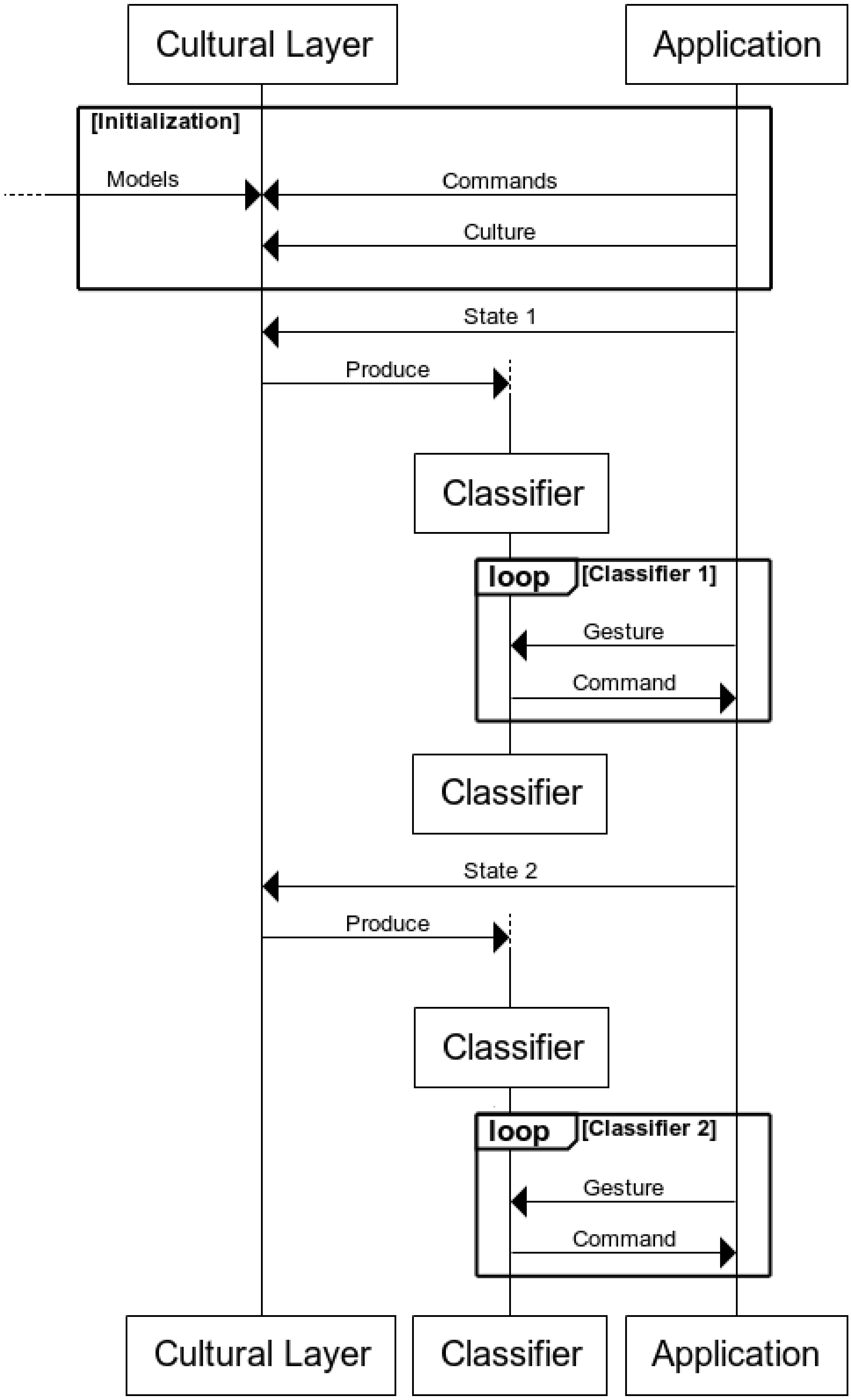

Figure 3.2: Interaction Diagram for generation and usage of multiple Classifiers

As stated previously in the opening of this chapter, the models used to generate Classifiers with are built using an Accord.NET library's Hiddden Markov Learning module. The Classifier is loaded with gesture classification models based on Hidden Markov Model, in simple terms, state machines where each state has probabilistic transitions. The Classifier acts as a filter, outputting the most likely match between an input noisy set of data values (the recorded gestures) and each of the models. The filter yields matches above a threshold probability, and defaults to the neutral gesture, the first one loaded, if none of the models has a good enough fit. To achieve this setup, it was required to create, and therefore train, these gesture classification models to load the Classifiers with. Thus, the models were trained by employing an unsupervised learning algorithm called the Baum-Welch Learning Algorithm, and utilizing a Forward Topology for the Hidden Markov Models. The number of states permitted in the Model's topology is customizeable before training, with more states often resulting in higher likelihood of accurarcy, but requiring longer amount of times to produce the model. Models were trained with static tolerance, but with different levels of Forward states, starting at 6 and going up to 20, with all but the best matching model discarded and the remaining one then tested and incorporated into the shamanic interface. More information on training these models can be found on [30].

As for the recording of Gestures to produce the models with, little work was done to the prior Gesture Recorder application beyond compatibility necessary with the Shamanic Interface's end. Sequences of Gestures are recorded using the same Leap Motion device that will later perform Gesture Detection during the game's application. These Sequences are captured as Frames, structures detailing hand and finger directions and movement during the recording segment. A total of 200 sequences were recorded for each gesture and each hand, to be used as training data for the models (4000 samples spread accross 20 classes).

3.3 Game

Alluded to above, the application developed and to performed during the user trials is a culturally enriched experience involving hand gestural control. Put another way, the application is a Game, which is how it will be referred to henceforward in this document, and evaluation involves observing users perform its completion.
It's a first-person perspective game, controlled exclusively by means of hand gestures detected by the Leap Motion Controller. The game runs on the Unity Engine, making use of Leap's 3.2 SDK to give it the ability to both show the user's hand within the game world as a virtual representation, and also to obtain all the frame data from its detection. This frame data is then redirected it to the Shamanic Interface's Classifier, which handles the interpretation of Gestures into commands. The user goals for the game is to explore the environment and finish a number of tasks, which are distributed in a series of rooms. The game is mostly linear, and the controls are very simple, featuring basic forward and back movement, sideways turning and then a number of commands

executed by a cultural emblem.

3.3.1 Tasks Definition and Design

From the very start, the game was thought of and designed to act as a frontend for the delivery of segregated and quantifiable cultural events. The idea of creating a gameplay loop involving separate *Tasks* predates this work and is a method often employed in user trials involving usability testing. What's essential about the tasks is it permits highlighting observable and surveyable data within measurable parameters. These include: User Error, Annoyances and frustrations, Speed of Completion, Attitudes of satisfaction, Requests for help among others that may emerge during trial.

The tasks are independent from one another, one action can not alter the function of another task, however, each individual task may involve performing at least more than one action given that these follow a natural flow in the view of the user. This independence extends to the task of moving the character. While performing a task, the player is locked in place and may not move until the task is complete, so that the participant isn't confused or distracted with technical aspects of control besides that which they contextually expected to perform.

The trials have two different user group, as informed previously. One group is given a natural and cultural experience, while the second group is given the same experience without the cultural components. The difference is established in the Task Definition. The first group, denominated the Cultural Group, performs tasks by means of actions corresponding to Emblems for their culture that fit the task's context. And a minimum amount of attrition is predicted in negative observable parameters for this group. Meanwhile, the other group, the non-Cultural group, performs the same tasks but the corresponding gestures are evocative of the mimicry or the general context the task is inserted in, but is not a culturally validated first choice.

With the validation of three independent helpers of the same Portuguese culture and age as the eventual participants, two sets of tasks and corresponding command hand-gestures was decided upon as detailed on table 3.1. All cultural group emblems were previously selected from a list of well-known and widely spread emblems to mid-Western European cultures, and this verification made sure they were recognizable to Portuguese people. Furthermore, tasks were separated into 2 groups, one, tasks T1 through T4, involving mandatory goalposts the user must go through during the trial, and the other, tasks O1 through O4, involving optional tasks they may choose to skip to end the trial earlier.

Within the game, Tasks are performed in isolation from one another, including the movement. There're two reasons for this separation. Firstly, so that a user is not encumbered by the possibility of issuing the wrong order while attempting to solve a different prompt. This could be the case if movement is available at all times, and a task's completion was based on proximity to the task's location. Secondly, so that a user is capable of identifying a prompt in the first place, as opposed to, for example, the possibility of a user attempting to complete a task with a non-interactive part

<table>
<tr><th>Sigle</th><th>Task</th><th>Cultural Group</th><th>Non-Cultural Group</th></tr>
<tr><td rowspan="4">M1</td><td>Move Forwards</td><td colspan="2">Point Forwards (Index)</td></tr>
<tr><td>Move Backwards</td><td colspan="2">Point Backwards (Thumb)</td></tr>
<tr><td>Turn Left</td><td colspan="2">Point Left (Thumb)</td></tr>
<tr><td>Turn Right</td><td colspan="2">Point Right (Thumb)</td></tr>
<tr><td>T1</td><td>Call to Come Close</td><td>Closing a hooked Index</td><td>Wave</td></tr>
<tr><td>T2</td><td>Display Impatience at delay</td><td>Look at opposing wrist</td><td>Open hand forward</td></tr>
<tr><td rowspan="2">T3</td><td>Call for Help</td><td>Wave / Raise Index</td><td>Thumbs Up</td></tr>
<tr><td>Direct Towards Object</td><td>Point At</td><td>Pointing, Finger gun style</td></tr>
<tr><td>T4</td><td>Celebrate Victory</td><td>Raising Fist Pump</td><td>Thumbs Up</td></tr>
<tr><td>O1</td><td>Silence</td><td>Index Over Lips</td><td>Thumbs Down</td></tr>
<tr><td>O2</td><td>Frame Photo of a flower</td><td>Square Corners with indexes and thumbs</td><td>Pinching the imaginary flower</td></tr>
<tr><td>O3</td><td>Pick Telephone</td><td>The “Shaka” hand</td><td>Raise hooked Index upwards</td></tr>
<tr><td>O4</td><td>Shoo Away</td><td>Strike air from inwards to outwards</td><td>Wave</td></tr>
</table>

Table 3.1: List of Gestures performed on each Task by the Cultural and Non-Cultural groups.

of the environment, expecting it to respond.
However, Tasks are not meant to be separated within the *‘physical’* space of the virtual world. If that were to be the case, then there would be no requirement to have movement as an option, but rather, all tasks could be tested individually one at a time in each of their own virtual spaces. It was, however, considered that the experience would prove to be more immersive and memorable to the users, thus closer to a natural experience, by giving the player the agency and freedom of the movement option between tasks. The opposing solution was compared to a test or a checkmark examination.
To ensure that the game offers no confusion in regard to what current task the game is prompting, and where tasks are located, there is a requirement for both visual and interactive feedback from the game. As such, it was necessary that Task transition fit the following criteria:

- Clearly identifiable visual labelling of a task’s location and spatial range of its correspondent state transition.
- Clearly identifiable differentiation between a completed and an incomplete task.
- Clearly identifiable differentiation between a mandatory and optional task.
- Clearly identifiable differentiation between each task type.

The design that fulfils these criteria was that of the Interaction Space, detailed in image 3.3. The interaction space is comprised of two main elements. One being a colour filled floor-level ring that signals to the player the radius the task will take place in and its current state. With a mere glance, the players are able to tell what stage of priority a task has towards their progress and if they should approach it or not in regard to that information, fulfilling 3 of the above requirements.

The fourth requirement is fulfilled by the other element, which is a collection of surrounding environmental and interactive props that produce light, sound and movement.

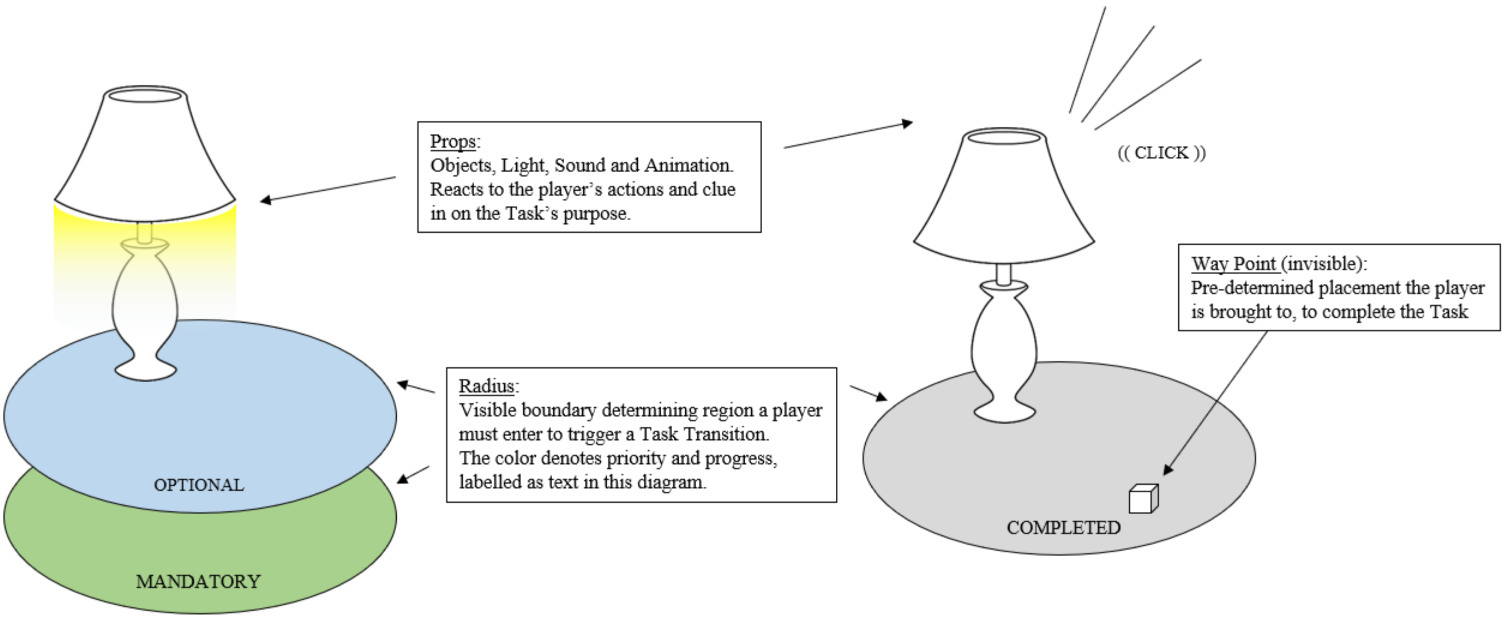

Figure 3.3: Overview of Interaction Space's Design Elements

3.3.2 Player Character Design

Given that the game was supposed to be a natural and immersive experience, no attention was desired to be brought to a character the user controls. Rather than creating one, the intent was to create an outward sense of belonging into the virtual world, as if the user could say "I am the character on-screen". Implementing a person even just through aesthetic design could compromise that plan. This was the reason why the game was built with its first-person perspective, that would eliminate the need for any visible character traits on-screen or any latent thought of piloting something over those of self-control.

However, merely showing nothing has one significant issue related to the novelty of the technology used. The Leap Motion is primarily used as an accessory apparatus to Virtual Reality experiences and games, and VR is itself a still uncommonly used technology, with many folks not having first hand experiences with it. The volunteers were predicted to not be completely aware of correct usage or limitations of the device unless if the game provided such feedback live as it occurs, and there isn't enough time during a user trial to reach the level or proficiency required to innately recognize problems as they arise. A virtual representation of the player's hand isn't just a suggestion brought up by other applications of the Leap Motion device, but a requirement that must be present within the game, as then, the user is capable of discerning not just if they're doing the correct gesture from the representation, but also if the application finds that they're doing the correct gesture as well, or if small corrections will be required.

Here, a decision is made that is at odds with the problem that caused the choice itself. What should the hand look like? There're essentially only two possibilities: Either the hand looks realistically, but inevitably appears different from the user's own real hand and a dissonance is established; or the hand is made to appear like an unambiguously digital approximation of a hand with simpler shapes. There were two at least two hand models provided by the Leap Motion's library of assets that made either choice possible without much trouble. Given the two choices, no benefit is posed by the first option, and a lesser consistent experience among volunteers was predicted, not to mention the potential source for distraction provided by the more visually enticing. There was a third option, giving the user a choice of hand model among a larger number, however, that could also pre-emptively give too much of an importance to the hand's look, which is an undesirable outcome if avoidable.
As such, during the game, the player can see a hand's model built of simple gray and blue smooth shapes denoting fingers and joints, functioning as a proxy to the player's own hand, and reacting to movement by performing the same gestures and staying at a similar depth and distance as the player's own real hand would, mimicking a first person's field of view with a raised hand.

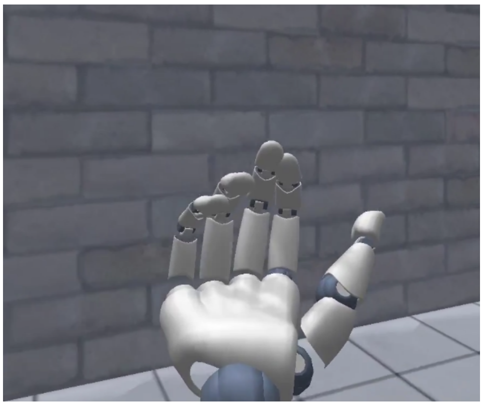

Figure 3.4: Screen Capture of the in-game Hand Model

3.3.3 Enviroment Design

The game occurs in a single virtual environment. Once it begins, there are no loading screens and any pauses happen only due to the game's own interactive make up rather than to needing to load more assets. The only complex background processing required involves the classifier switching and even this should be a seamless and instant process. In other words, the game is small enough to be fully loaded at all times. Yet, it should also have been big enough that there isn't a perception overload to the player with too many tasks and options present at a time. A large multiplicity of cues can confuse or frustrate the user. Given that there are 8 total tasks in the game separated in two groups of four, it was found appropriate to separate these into pairs of two tasks, one mandatory and one optional, and placed within four rooms. Additionally, to streamline the process of finding these tasks, the rooms were laid out in sequence through connecting corridors.

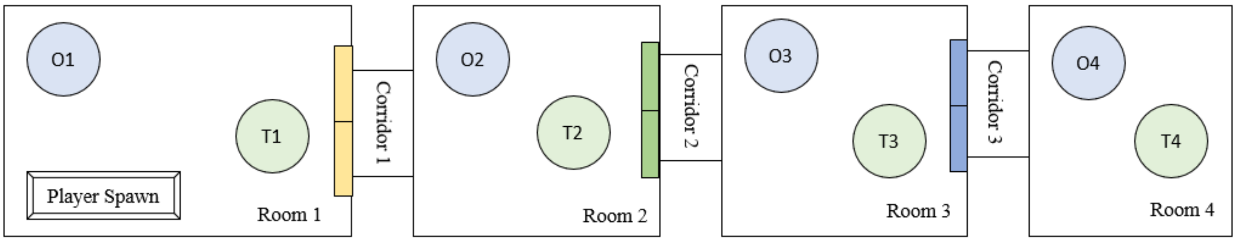

Figure 3.5: Overview of the virtual enviroment's Room Design

One of the things that such a sequential line-up opens to the game is the chance of someone missing out on the presence of a task. This is undesirable, as detracting user choice from the completion of a task gives off less information about the task itself, than someone intentionally avoiding it or performing it and giving a bad grade. Opposed, to it, it does permit to observe one of the subtler factors of immersive experiences, assumingly, the user's level of comfort and confidence with the controls, which are often overstated in self-assessments. Ultimately, the proposition of an intuitive cue for progress trumps over alternatives.

Rooms are therefore separated by gates that do not open until the containing mandatory task is finished. Moreover, the transition between rooms is not just identified by the crossing of a gate, but also visually by the rooms themselves, as the looks of the floor, walls and ceilings changes drastically upon entering a corridor and matches with the following area. Also, to better attach the completion of a mandatory task with its correspondent gate, these gates were colored, and a key was integrated as part of the task's completion.

Other guiding elements are also present within rooms. Green arrows appear near gates, corridor entrances or on empty walls to help direct the user and reinforce that they're heading the correct direction. Non-interactive plants are present on empty corner to fill the empty space and give the impression of a location of no interest to the game's progress. Also, the empty play area with no object of interest is progressively funnelled in size, as the player is expected to make lesser mistakes in their movement as they progress and would find less of a reason to take large exploratory measures with the controls.

This last point touches on the subject of room layout, which brings up the question of task placement. Of the two tasks in each room, the mandatory task is always placed closest to the exit gate and furthest to the entrance. The optional tasks however are progressively brought into the most direct path of the player. In order: O1 is on an indentation to the left of the room, O2 is off to the left but within the regular room boundaries, O3 is to the left of the path, but directly in front of the player from where they would come into the room, and finally, O4 is completely in the way of the player and they must sidestep it if they don't want to compete the task. Additionally, O1 and O2 represent tasks that don't lend themselves to having very much of an immediate urgency in nature, while O3 and O4 represent something that demands hurriedness from the user, and something that is just bothersome and has to be dealt with right away, thus justifying their progressive shift from staying away to getting in the way.

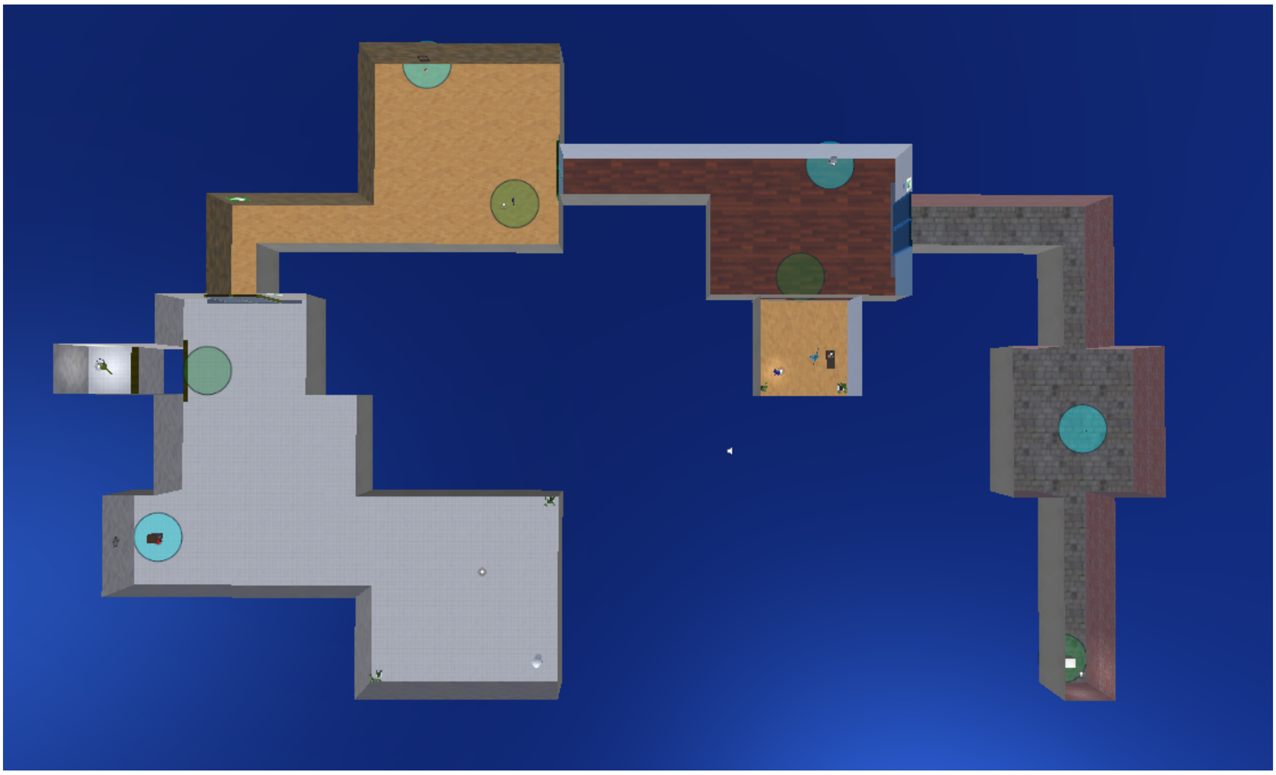

Figure 3.6: Overhead view of the game's virtual enviroment

3.3.4 Game Logic

For its implementation on Unity, the game has 4 core elements. These are the *GameController*, *CharacterController*, *InteractionSpaceController* and *CinematicHandler* scripts. Many other scripts were relevant to the game's function, however, all of those had a small role within restrained contexts, and worked in a master, slave manner to one of these core scripts, such as, for example, the RadomFlight handler script, that made task O4's insects move erratically, does not interact with any script besides the CinematicHandler and does so only on one moment in the game, during the closing transition of the O4 interaction space.
The *GameController* is tasked with initializing the game's Shamanic Interface and defining state definitions that the game will require. It is also the game controller that will dispense a new classifier during task transitions. Thus, this Is an entry point for the Character Controller and the Interaction Space Controller during those times. It is also the GameController that defines the game's culture, and it is here that I found the single difference between the game build presented to the Cultural and Non-Cultural groups. The Non-cultural group plays with an undefined Culture and thus the game uses the SI's default gestures. Meanwhile, the Cultural group defines its culture

as “PT” and those gestures get overridden by the pairs belonging to it. The code snippet corresponding to gesture efinition can be found on appendix A.1. There’s a single Game Controller object in the game space and the other core scripts have a reference and access to it at any given moment.

The Character Controller is a long script with definitions related to accepting input from the Leap Motion device (or Keyboard, for debugging purposes) to move the player’s camera around the virtual environment or to complete tasks. It’s here that the game checks if the classifier yields the correct gesture and if it’s also here that the game informs the rest of the system what gesture is currently being performed. It’s also where the game’s current valid options for control are changed, and where collision between a player and an interaction space’s region is handled. Ironically, the Character Controller is not statically aware of any command definition, merely matching what the classifier informs it, with the current Task. Its task completion code being agnostic to the definitions allows it to function if new definitions are set up. In summation, anything related to player input or to changing what a player can do has its entry point here.

However, it’s the Interaction Space Controller that defines what those gestures are for each task, and if they’ve been performed for long enough to accrue to the task’s completion. While there is a single Character Controller there’s a different instance of the Interaction Space Controller attached to the corresponding game objects. Every interaction space object has its own script, with its own different internal settings correspondent to a different Task. Upon collision of the Player and an Interaction Script, the scripts belonging to each attach to each other, and a back process of communication ensues through the 3 Scripts. The Interaction Space controller informs the Character Controller of what Task is performed within the interaction space, and the Character Controller then requests the Game Controller To feed that task’s correspondent State into its Shamanic Interface. Finally, the Game Controller yields the resulting Classifier back to the Character Controller, and the Character Controller will from then on inform the Interaction Space every frame that a gesture was performed. State definition code snippets can be found in A.2.

This happens relatively instantaneously, albeit, there is some leeway to give the Shamanic Interface some time for returning the Classifier, as at the very first moment of the collision between player and Interaction Space, an initial task transition phase, called player repositioning, is executed. During these, the character controller actually removes control from the player for three seconds and forcefully moves them to an appropriate spot where they will perform the task’s gestures.

And, finally, the Cinematic Handler, which synchs and performs visual and audible alterations to the environment during intermediate and final task transition phases. From moving and animating object, to dimming lights and turning off sounds, this provide the player with the feedback that a Task was completed by changing the task’s props in some manner. Only the Interaction Space Controller communicates with it, which means that granting player control after a Task’s completion must be done through it as a middleman between the Cinematic Handler and the Character Control Script.

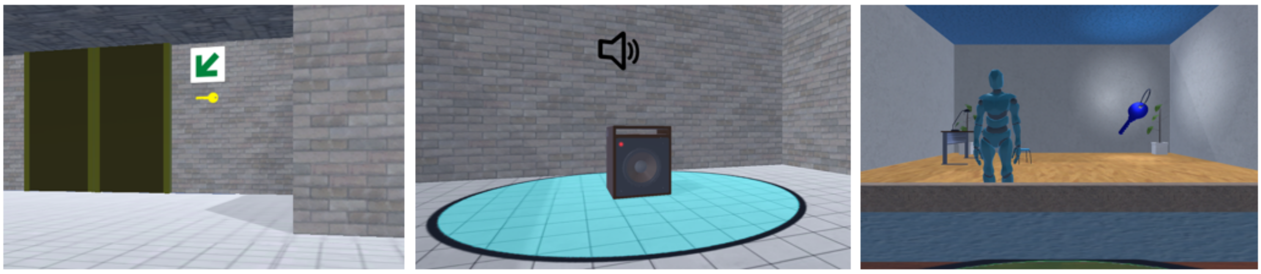

Figure 3.7: Screenshots of the Game Enviroment

3.4 Volunteer Trials

This works' field research is composed of, as determined before, user trials in a controlled environment with the goal of finding and collecting insight in regard to the usage of Cultural Awareness as the primary solution for the issue of user spontaneity and adaptability, and the hypotheses stated within the Goals section1.3.

The planning surrounding the user trials was carried out with the intent of facilitating the forms of techniques employed in data collection useful for evaluation, those being observation data and user opinion data. These are performed concurrently to differing degrees of focus in accordance to different parts of each trial. These parts of a trial roughly follow the directions of the outline provided earlier in the section on HCI Evaluation2.6, however, they will be detailed further in the following sections and an overview provided is provided in image 3.8.

3.4.1 Pre-Test

The pre-test is the initial part of the trial, and it matches up roughly to the Introduction and Warm-Up phases of a field study plan. Users are introduced to the controlled environments, the room and equipment, and are initially asked for their demographic data and some first impressions and some very basic questions, such as, if they have knowledge of the apparatus used in gesture detection presented, the Leap Motion. They are also allowed to test out the Leap Motion's diagnostic visualizer for up to 30 seconds just to get a rough idea of its function.

However, there's two bigger main components to the Pre-test after the initial introduction. A Cultural Survey and a Game Explanation.

The Cultural Survey is done before any information about the game is given to the volunteer. They are presented with a simple black text on white background slide show containing several snippets indicating a different message. Each message represents one of the game's task, and the volunteer is tasked with performing the gestures they feel best describe that action or message. The main objective of this exercise is checking if the user performs a majority of the same moves as the ones the Cultural Group uses during the game. This is a requirement of this work as a counter-measure against certain potentially erroneous assumptions made during development, such as, most directly, the choice of gestures the tasks use. Should the majority of users name the same gestures as the Cultural Group, this helps pre-emptively validate the game as a cultural experience. Should the

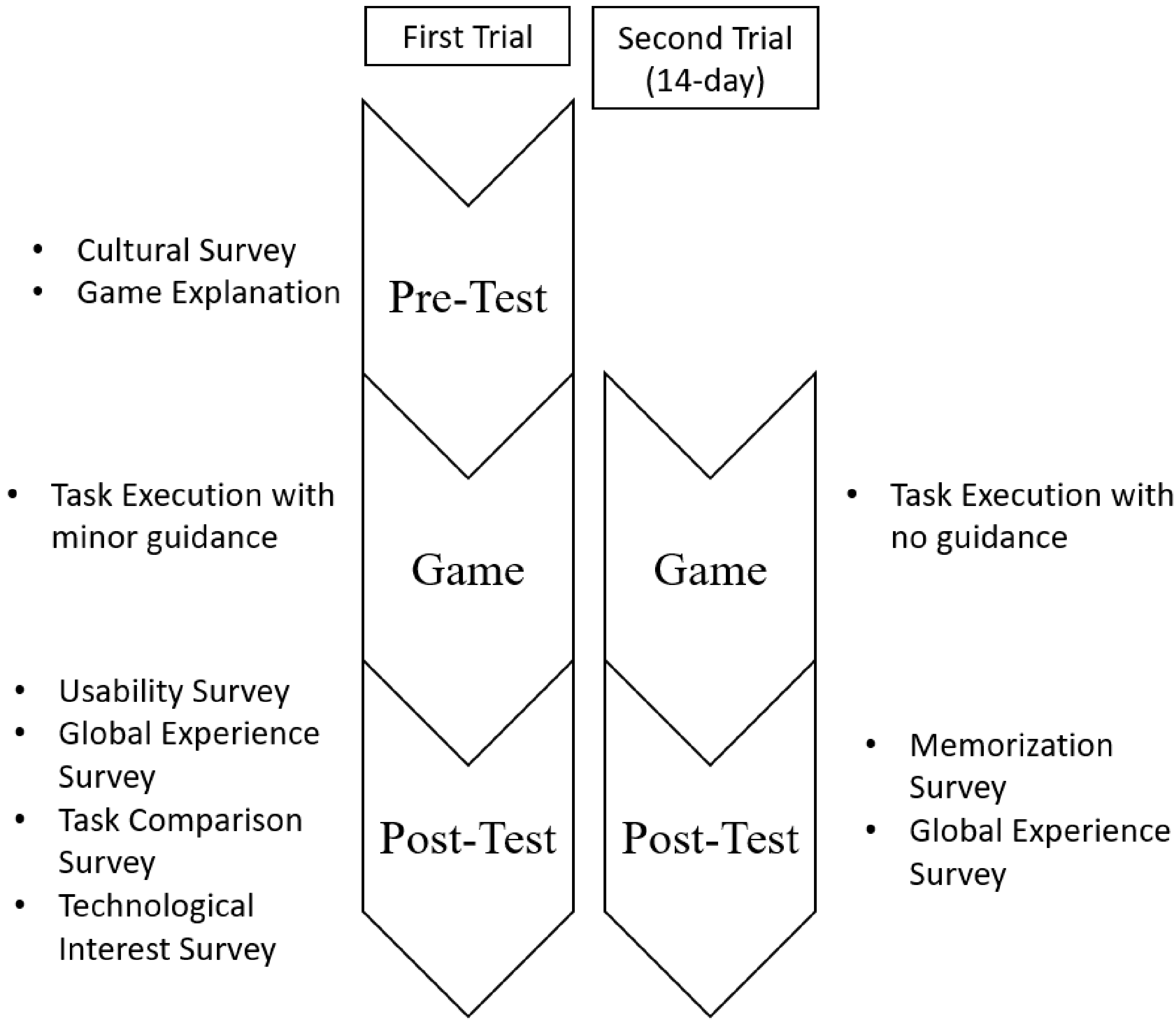

Figure 3.8: Overview of User Trial's Sections.

majority name a specific different gesture, that should, conversely, invalidate that task for result-checking.

The Game Explanation happens right afterwards. Here, the volunteers are shown a depiction of the in-game tasks they'll have to perform, in the same order as the messages they just read through, both in the slide show and with a separate physical document. There are different explanations for the cultural and non-cultural group, however, those difference are never stated. The volunteers are given instructions on how to perform the tasks, as well as the differences between the Interaction Spaces and the method by which the game progresses. Finally, they're asked if they understand the purpose of the gesture used in the task. Their comments and understanding are written down and recorded. Here the non-cultural users are implicitly given an opportunity to name their preference for one of the tasks they performed previously.

3.4.2 Game Observation

Following the explanation of the game, the volunteers are ready to play the game proper. Here, they are asked to perform the game fully. The volunteers are given a few more pointers to aid them on the way, mainly with getting used with movement, but are never outright given the actual gestures required for a task. They are also prompted for help if they appear to have difficulties at certain steps but are only given it if they accept or request it.
Any special observation is recorded, priority given to unexpected displays of frustration or amusement. But some aspects of observing the user are most important for this segment. The Time Elapsed, the Number of User Errors, if Help was Required, and if the Correct Gesture was actually performed. These are likely to provide the most qualitative insight and are discernible enough to make an immersion scoring out of, and as such, the protocol developed for the trial is to document these values for every single task and every single volunteer, with both help from recordings (when permitted) and the game's own logging of each task.

3.4.3 Post-Test

After the game is completed, the volunteer is given a moment to rest and then they're asked to go over a questionnaire and answer it. This is closely related to the Cool-Off period of a field research and is the adequate time to investigate questions in the form of a survey. Volunteers aren't asked to answer on their own, but rather, all of the questions are read out to the users and they're made clearly aware of what each question is asking, slowing down the pace between questions if they're answering too fast, and prodding for certainties in responses when a bout of indecision arises.
There are three parts to the post-test questionnaire. The first part is a usability survey. Since the game was rewritten from scratch, it stands that its usability is still unproven and must be evaluated, however, the real reasoning behind making this investigation is that many aspects of usability have crossover with immersion and satisfaction signifiers. Thus, this is an insightful exercise towards the goals of this work and helps frame the way by which further questioning will be administered.
The questions presented in the first part are adapted from the System Usability Scale. All scores of the scale are present and answers are scaled such as a 0-100 usability score is possible. Some other satisfaction-related questions were dropped here, particularly, that of the least enjoyable or most concerning aspect of the game, to ensure users don't leave out details about their frustrations.
The second part of the questionnaire goes over every Task and asks two questions on a scale for each of the gestures. The two aspects are surveyed are the ease of the gesture, and the natural fit of the gesture for the task. Since the users may find different interpretations for the scale, this one is meant to be taken as a comparative measure between each of the gestures, scoring them in terms of 'below than ideal'. This should provide some final insights into the choice of gestures and their

applicability for the task, but moreover, should provide a focal point of comparison between the Cultural and Non-Cultural groups.
The third part is short, and merely looks into the volunteer's experience with computers on their daily lives and their interest and pre-existing capabilities using other natural systems. This was a concern that could arise if groups of volunteers appeared to have vastly different expectations and performances between one another.

3.4.4 Second Trial

After performing the trial, the users are thanked and allowed to depart. However, their contribution isn't over at this point. Given that one aspect of the proposed benefits to the shamanic interface approach has a large emphasis on memorization of commands, these have to be tested after a delay of time with no exposure to the system. All participants are therefore asked to come back and fulfil a second follow-up trial after a period of 14 days minimum. The goal here is to assess if their memory of the content still matches the actual experience. The second trial has two parts, first the volunteers perform the game again, and then they're given a new post-game questionnaire.
Absolutely nothing is changed with the game builds themselves, as the experience must remain consistent with the previous. Also, the same indicators have primary focus during observation, those being Number of User Error and so on. However, a small difference is with the protocol is that time, the users are not given any of the previous helpful pointers and are not prompted on their need of help if they're seen struggling. The players must request help themselves if they appear to struggle and are allowed to proceed with failure if they don't remember the task gesture but are certain of an erroneous one. This compels users to rely only on their own recollection of the solutions and in taking their time, rather than rely on communication with the researcher. The benefit of taking this approach is that it highlights their confidence on the gestures given without it needing to be taken in survey. To that end, the users are also not corrected on gestures until the second trial is over and they're no longer under observation.
The second post-game questionnaire is different from the first trials. This questionnaire has two parts. The first goes over every task and surveys users on their ease remembering each gesture. This is, again, meant to highlight a difference between each of the two Cultural groups, as the Non-cultural is predicted to mention greater barrier against the Cultural group. Additionally, there may also be a significant disparity between self-confidence and erroneous resolution of gestures in this aspect between the two groups, which may be of interest to record. As for the second part of the questionnaire, instead of focusing mostly on usability like the previous, this time, users are prompted over categories of experiential quality indicators. They're asked about the impact, agreement and value of several aspects of the experience which they may rate on a 5-point scale between Strong Disagreement and Strong Agreement, although the majority just kept to simple a binary dimension between Yes and No or Good and Not Good answers. Listing the categories, their affective expectations, their feelings, their self-consciousness, the repeatability and their willingness to share with and recommend to others, their thoughts on the applicability and usability of

the type of experience among different groups of people, the technological quality of the experience, and sense making and meaningful feedback.

Chapter 4

Results

The user trials were conducted at the Faculty of Engineering from University of Porto within labs and other rooms, in isolation from each other and from the presence of distracting elements such as other people not part of the test. Excluding a user requested to conduct a pilot trial ahead of the remainder, from a much larger pool of signed up people a total of 17 users performed the trials, 16 of which saw the second trial to completion. The research observer for the trials was the author of this document, who sat beside the volunteer at a table with all the equipment and required paper files. User sessions have been recorded with given explicit consent.
Results, observations and analysis of the data will be given in this paragraph. The impact this analysis had on the hyposteses will be laid out in sumation on the Conclusions Section 4.5.

4.1 User Sample Description

As remarked above, the trials involved 17 participants, from whom 16 have concluded the whole two phases of trial and one has failed to attend the second trial. The participants were divided in two groups, the cultural group with 9 participants, and the non-cultural group. Given that participant number 5, the person who did not have the availability to show up to the follow-up trial, belonged to the Cultural Group, this leaves the results with 17 cultural evaluations during the pre-test, and then a balanced 8 evaluations for the Cultural Groups and 8 evaluations for the Non-Cultural Group.
The participants were all portuguese students from University of Porto, belonging to different faculties and courses, although the majority reports studying a field of engineering. As such, the age gap is constrained to the 18-24 gap, and demonstrate at least some domain knowledge in technology, even if none showcases any experience with the equipment. Still, two particular users, number 1 and number 8, stood out from the remaining where Technological and Domain Interest was looked at, to which both have been further questioned and both have answered to currently work in the field of study of Informatics.
The sorting of volunteers between Cultural and Non-Cultural groups was done in the most randomly and balanced approach available, while weighting the possibility of volunteer drop-out.

The first user was tested and observed within the Cultural Group and then each participant was interleaved to the Non-Cultural Group and the Cultural once again in accordance to their scheduling order, which is equivalent to their participant numbering. As such all odd-numbered participants would belong to the Cultural Group and all even ones would belong to the non-Cultural. However, a slight alteration was found necessary, by questioning the users early on if they knew someone else who would also participate in these trials. Two volunteers claimed to know each other, number 4 and number 7, and as such, participant 6 was slated to belong to the Cultural Group ahead of time so that 4 and 7 could both belong to the same group. This such change as to not have them influence each other's memories with erroneous impressions of the other's trials by sharing them between each other. Sharing impressions of the first experience with other people was not forbidden and is actually encouraged as one of the potential immersivity signifiers, however there may be a potential unexpected effect from the sharing of mismatching experiences that was thought best avoided. By giving them both the same experience, they could instead only reinforce or reconsider aspects they have witnessed first-hand and make their own impressions for the second experience survey.
Table 4.1 lists user demographics as well as findings on the user's technological interests.

Volunteer Number	Gender	Main Hand	Age	Tecnological Interest	Group
1	M	R	23	9	Cultural
3	M	R	21	7	
5	M	R	21	8	
6	F	R	22	4	
9	M	R	24	5	
11	M	R	23	6	
13	M	R	21	3	
15	F	R	18	8	
17	F	R	22	7	
2	F	R	18	6	Non-Cultural
4	M	R	23	8	
7	M	R	19	7	
8	M	R	21	10	
10	F	R	24	7	
12	M	L	23	7	
14	M	R	22	6	
16	M	R	23	7	

Table 4.1: Listing of Volunteer Demographics Findings and Sorting within Cultural and Non-Cultural Groups.

4.2 Cultural Experience Validation

The pre-test was a common section that all 17 volunteers performed, still unaware of the User Trial's goals or even general setup beyond the knowledge that cultural gestures would be an essential part of the observation. With no pre-emptive knowledge of the experience and with, at this point, very little suggestive information as to what will be required of them, they were surveyed, using simple language and a simple presentation, on what they found was the correct set of emblematic gestures that would solve or address each of the specific Tasks they would later have to employ in the game. Before beginning, the users were assured that no such thing as a wrong answer could possibly exist as this was a direct survey of their personal opinions, and they were put at ease that none of the data lifted would result into later questioning, as this data would have no relation with the experiment ahead.
Ideally, this would be performed on a different set of users beforehand, which would even allow for a restructuring of the methodology employed on the user trials as a whole, exploring aspects such as natural discovery of commands without guidance. However, given the limited availability of volunteers and the likelihood of dropout in a two-part experiment, it was found hard to do so within the means reachable.
If a Task's chosen gesture is correctly selected and fit for the purposes of a Cultural Experience, then when given the opportunity to perform the Task on any independent context, a person should most of the time opt to perform that same well-fit gesture over another. Admittedly, certain scenarios may change the chosen gesture due to certain contextual cues lending towards different interpretations, and on top of that, not all gestures that abide by this are necessarily emblematic. For example, working with certain tools may represented by an emblematic gesture, by limiting the shape of the tool to a different arrangement may lead the person to perform a less emblematic gesture, and instead perform mimicry of the action. This concern was actually predicated in the design of the game by the optional Task O3, answering a phone, which demanded the choice on in-game representation of the ringing Phone be an object with a handset piece that can be held, as opposed, for example, a more modern looking device. This is why in the pre-test, not just were users asked to perform the gesture they find correct without a particular context in the barebones slideshow, but they were also asked if the game's taught gesture made sense for the context.

4.2.1 Invalidated Tasks

Table 4.2 is the result of the first part of the pre-test, where all of the volunteers were asked about performing the chosen gesture.

Two tasks with concerning response arise in a suspiciously distinct manner. Of the one or two gestures each user made for the slide correspondent to Task O2 and Task T2, only very few actually performed the expected Gesture during the Cultural Validation. This punctuates the possibility of having made the wrong assumption for each of the two during the initial brainstorming or subsequent steps involved in this selection of Gestures. Alternatively, it could also signify that

	O1	T1	O2	T2	O3	T3.1	T3.2	O4	T4
V1	Y	Y	N	N	Y	N	Y	Y	Y
V2	Y	Y	N	Y	Y	Y	Y	Y	Y
V3	Y	Y	Y	N	Y	Y	Y	Y	N
V4	N	Y	N	Y	Y	Y	Y	Y	Y
V5	Y	Y	N	Y	Y	Y	Y	Y	Y
V6	Y	Y	N	N	Y	N	Y	Y	Y
V7	N	Y	N	N	Y	Y	Y	Y	Y
V8	Y	Y	N	N	Y	Y	Y	Y	Y
V9	Y	Y	N	N	Y	N	Y	Y	Y
V10	Y	Y	N	N	Y	N	Y	Y	Y
V11	Y	Y	N	N	Y	Y	Y	Y	Y
V12	Y	Y	N	N	Y	N	Y	Y	Y
V13	Y	Y	N	N	Y	Y	Y	N	Y
V14	Y	Y	N	N	Y	Y	Y	Y	Y
V15	Y	Y	N	N	N	N	Y	Y	Y
V16	Y	Y	N	N	Y	Y	Y	Y	Y
V17	Y	Y	N	N	Y	Y	Y	Y	N

Table 4.2: User-oriented Validation of Task gesture choice by mean of blind survey of natural cultural emblems.

the presentation itself was misleading or not clear enough.
To that end, all the diverging answers provided by the users were recorded, in an attempt to figure out if there's a consistent pattern as to what went wrong in the design of the Task.
Task O2 involves taking a snapshot of a flower and framing it in an initially empty rectangular mount. Already, with care, from this Task description could the problem be perceived, as well as the eventual source of the disparity. The Task describes two actions, the photograph snapping, and framing of the revealed photo. While in design, the action focused upon was the framing action, most users instead focused on the photographic action itself, which involved a mimicry of the snapshot acting. Out of all 16 users that performed a different action, every single one was recorded performing a gesture called the *"Camera Click"*. This gesture was not found as a symbolic emblem while first researching and planning for this work, which was partially the reason that led to the wrong assumption that it wouldn't be the most natural response to the task's proceedings. There was some worry about this Task towards the impact it may have had on the global experience, however this worry was put at ease with the second component of this pre-test, as can be seen on table 4.3, every user in the cultural group stated that the *Framing* emblem was recognizable and fit the game's context. Every user, but one, Volunteer number 11, who left the comment "Knows the gesture, but wouldn't see the connection without an explanation".
Task T2 is a slightly more complicated matter. The task features a large clock and a waiting sign room, and the users are meant to emote their impatience at a long wait. Only three people gave the expected gesture of looking at or tapping the opposite wrist, however, when looking over the other given gestures, no pattern was recognized. A group of three volunteers performed another well fit

gesture of bringing their hand to the side of their head, while two volunteers crossed their arms and tried appearing displeased, a gesture that doesn't involve detectable hand movements. The remaining gestures were singleton examples of emblems corresponding to hurrying other people, threats and then other non-emblematic gestures of disconcerting reasoning. Unlike with the O2 task, were all users performed one, this makes a total of only 7 users performing an adequate emblematic gesture for the given scenario in T2, and 5 users performing non-emblematic gestures. This is an indication that the scenario chosen for task T2 may have not been an apt one for this game.

As such, the data collected for these two tasks is ignored while doing a discrete assessment of the Tasks, since it can't be fully justified that differences found between the groups are due to the cultural differences of the experience. These, however, still take a part of the global assessment of the experience, as table 4.3 does provide verification that the users understand and recognize that the choice of gesture is fit to both Tasks, even if not their first choice.

4.2.2 Preeminent Non-Cultural Misfits

Past the Cultural Validation, volunteers were taught the game's intended gestures and asked to comment on them, particularly by request of mentioning if they had prior knowledge of the emblem required and if they found it fitting to the Task shown. Table 4.3 is the result of this request. As is evident, the Cultural group almost completely shown agreement with the choice of gesture. This is was an expected outcome.

But for the Non-Cultural Group, the intent was never to obtain an acquiescent response, but rather a mixture. The objective of the gesture selection for the Non-Cultural Group was to avoid the natural choice, but still pick something that could make sense given an alternate context of the task. For example, the task O1, silencing an object, is performed by showing a thumbs down gesture of disapproval at the speaker prop. The contrast between what's asked and what gesture is required at play here is that, while the thumbs down is an emblematic gesture for the volunteers and could be used in a different scenario where the objective is communicating the request of lowering of a music's volume, this is not very sensical if the recipient of the gesture isn't a person. Ideally, all tasks would have a balanced mixture of volunteers finding the gesture fitting and unfitting, but some gestures have got an overtly positive or overtly negative response.

The Non-Cultural Group participants seemed to focus most on the negative at this step, which is why the presence of task O2 as a major negative read is preeminent. For the positively approached gestures, a difference between the response and the actual performance in practice, specially on the Second Trial, could provide more insight than just user opinion. But for a preeminent negative task, this means that the gesture chosen for the non-cultural group may be far too detached from the task to make a fair comparison between it and a cultural equivalent. The reasoning is that this unique contrast between O2 and the rest of tasks may make it more memorable or notable than the remainder for the Non-Cultural, which is an undesirable factor for the results obtained.

As such, this compounds that Task O2 should be ignored during data analysis. Users were not informed and were still requested to perform it if they felt like it alongside the remaining Optional Tasks, so as to not alter the experience between first and second trial. No special attention being brought to it or to Task T2.

Cultural Group	O1	T1	O2	T2	O3	T3.1	T3.2	O4	T4
V1	Y	Y	Y	Y	Y	Y	Y	Y	Y
V3	Y	Y	Y	Y	Y	Y	Y	Y	Y
V5	Y	Y	Y	Y	Y	Y	Y	Y	Y
V6	Y	Y	Y	Y	Y	Y	Y	Y	Y
V9	Y	Y	Y	Y	Y	Y	Y	Y	Y
V11	Y	Y	N	Y	Y	Y	Y	Y	Y
V13	Y	Y	Y	Y	Y	Y	Y	Y	Y
V15	Y	Y	Y	Y	Y	Y	Y	Y	Y
V17	Y	Y	Y	Y	Y	Y	Y	Y	Y

Non-Cultural Group	O1	T1	O2	T2	O3	T3.1	T3.2	O4	T4
V2	Y	Y	N	Y	Y	Y	Y	Y	Y
V4	N	Y	Y	Y	N	Y	Y	Y	Y
V7	Y	N	N	Y	Y	Y	N	Y	Y
V8	N	Y	N	Y	Y	Y	Y	Y	Y
V10	Y	Y	N	Y	N	Y	Y	Y	Y
V12	Y	N	N	Y	Y	N	N	Y	Y
V14	Y	Y	N	Y	N	Y	Y	Y	Y
V16	Y	Y	N	Y	Y	Y	Y	N	Y

Table 4.3: Difference in consensus and recognition between gestures used in the Cultural and Non-Cultural Group.

4.3 Task Execution and Observation

The game is played twice by the volunteers, once with some initial direction by the observing researcher, and then again without any aid, unless if requested. During the game, the observer's role is to record the volunteer's performance in completing the task. Tracking their total game time, their completion time per task, the number of errors they commit, when they have requested help, which tasks the players couldn't solve correctly, how often the system has caused issues to the players. However, they're also there to provide support if necessary, and to keep the volunteer sharing about their impressions, as other valuable information may be better accumulated during the process of the play session itself.

4.3.1 Movement and Time as a Measure

One of the elements that the observation was meant to register was the total time taken to complete the game, which would be a function of how fast the players managed to learn and recall the task completion gestures. However, this ended up being moreover, influenced by time taken moving around the virtual environment.

Both times taken to complete tasks and times taken to become accustomed to movement were predicted as important values to take in consideration, however, this latter wasn't an important metric, as movement had the same approach on both the Cultural and Non-Cultural Groups. It was going to be tracked in the interest of showing important observation, it became clear that doing so the intended way was going to be impossible. Several volunteers never found a moment where they said they got used to the controls to the very end of the experience, and even found trouble the second time. Meanwhile, other users were immediately proficient with it. Nevertheless, all of them still claimed that the movement was the least enjoyable part of the game. The main complication stated being the transition between the left movement and the right movement, as transitioning between the two involved doing too wide of a movement, and the camera wouldn't pick up on it right away. But moreover, the most impactful effect of movement over the data was that it made the total game time a noisy metric. Rather being able to take that information and attribute a conclusion to the cultural impact, partial conclusions about technological impact needed to be made as this provides a clear influence. Total times in the first trial were much slower for the Non-Cultural Group Volunteer, but there's a lot of variance to the values, and for the second trial the values were all much too similar between the two groups, without it being possible to support any conclusion in regard to the change. Individual conclusions may be made, as the fastest times observed in the first trial belong to users reportedly with a high technological interest score (V1, V8, V15), and two proven to be outstandingly fast in the second trial again (V8, V15). Also, there is a user (V9) that was outstandingly slow in the second trial, and who struggled with movement to a much larger extent than everyone in both groups, who coincidentally also reported to have a lower technological interest score than the median. In effect, this user may have struggled with retention of this aspect of the game. However, the lowest times on both trials do not have a definite pattern to the technological interest scores, and even the volunteer who showcased most trouble did not self-report the lowest of all scores. It's very possible that a correlation exists for higher user confidence with technology and gestural command retention, but, given that technological impact was not the focus of this study and given that the not enough rigor on the was employed to explore the matter and circumvent the potential issue of self-reporting biases, any further analysis was dropped. Reiterating, the thought given to this section is a justification of why movement timing observations were not entirely helpful in drawing meaningful conclusions.

For clarity, table 4.4 shows the Total Game Time data of every participant including averages and standard deviation on that data.

First Trial

Cultural	Total Time	Non-Cultural	Total Time
V1	322	V2	642
V3	348	V4	550
V5	355	V7	433
V6	405	V8	356
V9	415	V10	470
V11	336	V12	355
V13	446	V14	556
V15	322	V16	361
V17	425		
Average	374.89	Average	465.38
Std. Dev.	47.84	Std. Dev.	108.68

Second Trial

Cultural	Total Time	Non-Cultural	Total Time
V1	357	V2	315
V3	329	V4	361
V6	372	V7	388
V9	530	V8	258
V11	327	V10	321
V13	307	V12	304
V15	267	V14	358
V17	347	V16	314
Average	354.5	Average	327.38
Std. Dev.	77.91	Std. Dev.	40.49

Table 4.4: Total elapsed time between game start and completion of both Cultural and Non-Cultural Groups for both First and Follow-Up Trial's observations

4.3.2 Volunteer Number 8

Volunteer number 8 stands out as an unexpected anomaly to the data. To better highlight why this is the case, the following table 4.5 is collated to the document. It should be noted that, for the purposes of better showcasing the difference between this one and the other volunteers in the Non-Cultural Group, data from the O2 and T2 tasks was not left out.

	Total User Errors	Total Forgotten Gestures	Total Help Requests
V2	5	4	0
V4	8	4	1
V7	6	4	0
V8	0	0	0
V10	4	4	0
V12	7	7	1
V14	8	3	2
V16	6	3	1

Table 4.5: Count of focused performance failures commited by Non-Cultural Group members during the Second Trial

Despite being in a group that consistently showcases struggles with the game, Volunteer number 8 has performed no observable mistakes. Additionally, they were also the fastest of all participants in completing the tasks and were one of the fastest users in getting used to the movement options. And in the first trial, they may have committed user errors, but all of them were related to aid requests and ensuring they were not about to perform an error ahead of time. Similarly, all delays in completing any of the Tasks by the user were due to them taking a moment to think and recall on their own, and they have removed their hand from within the Leap Motion's tracking

radius to prevent any accidental reading. In other words, they learned the system's functions and restrictions instantly.
Nonetheless, they claimed to have never experimented with systems entirely controlled by means of hand gestures, and to have a low amount of experience with other similar ones, such as Virtual Reality environments with hand tracking controllers. They did however, state that they had a larger degree of experience with consoles with wide arm gestures such as the Nintendo's Wiimote, but claimed to have never tried the more similar tracking approach used by the Sony Move. Thus, the major characteristic the user had going for them, was their earlier self-reported high interest in the field of Human Interaction, and their efforts towards it given their study of technology at the same university the trials were performed in.
In terms of determining if the user is a statistical outlier, the data changes slightly depending based on whether or not the full sample of data from both trials is used or not. Assuming a Normal Distribution of users, the probability of the user's Standard Score being observed is below 3.44% (-2.124) with data from both trials and below 1.70% (-1.828) taking data from only the second trial. Additionally, the level of technological interest showcased is above 2,794% (1.886). There's enough of a suspicion the user lands two standard deviations above the median in performance, however, choosing a correct cut-off is a well-known matter known to produce false positives in a test group with low amount number of test items or low population [31] [32]. With two oddities among score on a two-tailed cut-off, there's still least a 5% chance that the user was incorrectly leveraged as an outlier, and the rigor self-employed in determining further measures beyond performance errors is of question, as well as the potential misgivings of the sample size or methodology. As such, despite the tabled data crossing the 0.05 alpha level on a two tailed cut-off (+/- 1.96) for the second trial on its own, there's a fear that eliminating the user may be a mistake on the part of the analysis.

4.3.3 Performances, Memory and User Errors

User performance was registered during the game, this was the focal point of the observation protocol. Odd behaviours, mistakes, failure at remembering the correct gesture and help requests were the primary measures employed, targeted primarily at moments during which Tasks were being completed. Table 4.6 provides an overview of a summary of the data collected for both Cultural and Non-Cultural groups during both of the trials, with the dimensions: Average Task Time, Total Number of User Errors, Total Number of Failures (Didn't use the correct gesture), Total Number of Help Requests. The choice towards using totals rather than converted scores here is found apt due to both groups having the same size, 8 participants, and thus any pre-processing of the data would just lead to the same results. It ought to be highlighted that additional time information was already noted on 4.4.

Of the data displayed, time ended up being the least insightful. It was already discussed prior that total game time did not yield many discernible results due to the influence movement had on

	O1				T1				O3			
	Time	Error	Fail	Help	Time	Error	Fail	Help	Time	Error	Fail	Help
C1	8.63	2	0	1	8.56	0	0	2	6.22	0	0	0
C2	7.25	0	0	0	10.88	2	0	0	6.57	2	0	0
NC1	10.17	7	3	2	6.75	1	1	1	5.20	3	3	1
NC2	7.50	6	3	0	8.25	5	2	0	8.50	8	6	2

	T3.1				T3.2				O4			
	Time	Error	Fail	Help	Time	Error	Fail	Help	Time	Error	Fail	Help
C1	6.33	0	0	0	6.89	0	0	0	6.33	0	0	0
C2	6.25	0	0	0	8.88	0	0	0	6.00	0	0	0
NC1	7.50	5	2	2	11.13	6	1	3	7.00	2	1	1
NC2	7.38	7	5	0	10.25	5	3	0	6.17	5	5	0

	T4			
	Time	Error	Fail	Help
C1	7.33	0	0	2
C2	11.75	0	0	0
NC1	8.00	2	1	2
NC2	9.25	1	1	0

Table 4.6: User Performance Data during the First and Second Trials.

the players, without this being one of the differentiated aspects between players. However, discounting all time spent on moving the character, still this doesn't prove to be a much conclusive metric. By further organizing data and preparing a time comparison table between the Cultural and Non-Cultural Group 4.1, while it may appear that the Non-Cultural Group has more trouble in quickly solving each of the tasks, this is neither a consistent observation, nor was it possible to prove with 95% confidence that such was the case. Additionally, it's not possible to use the data as supporting evidence that the more user error performed by the groups for a task, the slower the group was. Using regressions analysis, the value of R-square is a minute 0.5688, which is far from enough to call a good fit between the two.

Naturally, that does seem to point out that fitting Cultural Emblems are a beneficial method of approaching a gestural control scheme, as it seems to lower the negative impact seen both aspects for the Non-Cultural Group. However, there needs to be a tell that makes that clear within the data. To that end, the mistakes committed by this Non-Cultural Group need to be looked at. Were they making a mistake because of the difficulty of the Gestures themselves? Were they confusing the correct gesture from another Task with the one they were trying to Solve? As such, all Non-Cultural Group's erroneous gestures were registered and looked, and with that, the break-down table 4.7 was made, manifesting the most conspicuous of type of error.

Of all the mistakes committed, roughly 75% were Gestural mistakes. These are when the user attempts to perform a gesture that is wrong for the Task and doesn't stop their attempt until after

		O1	T1	O3	T3.1	T3.2	O4	T4
Trial 1	C	8.63	8.56	6.22	6.33	6.89	6.33	7.33
	NC	10.17	6.75	5.20	7.50	11.13	7.00	8.00
Difference		1.54	-1.81	-1.02	1.17	4.24	0.67	0.67
Trial 2	C	7.25	10.88	6.57	6.25	8.88	6.00	11.75
	NC	7.50	8.25	8.50	7.38	10.25	6.17	9.25
Difference		0.25	-2.63	1.93	1.13	1.38	0.17	-2.50

Figure 4.1: Tast Execution time comparison Table between Cultural and Non-Cultural Groups

completing the gesture. The remaining 25% are other types of mistakes which also include errors related to gestures, but that did not involve a full attempt with that given wrong gesture. Now, of these 28 Gestural Mistakes, all except 2 were Emblematic Substitutions. Here, Emblematic Substitutions are mistakes where the volunteer performed an emblem belonging to their culture instead of the requested command, and as would be expected, their choice of Emblems for the task was for the most part, chosen from among gestures the volunteers had performed during their initial Cultural Survey in the pre-test. What this seems to imply is, despite being told clearly not to use those emblems as they would be considered wrong, the participants' memory still made them recall their emblematic gestures over the gesture they actually performed the first time they performed the game. Which is evidence that Emblems are the more natural way performing those tasks. For clarity, the 2 non-emblematic gestures involved Pantomimes, one with the user attempting to say "here" while pointing out proximity to themselves with their hand, and another outright performing Mimicry, grabbing and dragging an invisible object.
The final row refers to External Emblematic Substitutions. Here, the word External refers to the Game's gestural command set, and thus, this breaks downs how many of the Emblematic Substitutions brought gestures that weren't present on the game at all. What's important about this difference, is to show how many of the above were split between merely forgetting the gesture and attempting to solve it in the way felt right, or how many actually recalled performing a certain gesture, and were trying to fit it within a context that made sense. One such example of the latter was the Task O1, as all 4 of the non-external substitutions involved the volunteers performing Task O2's gesture, thus, it's very likely that all 4 of these volunteers were trying to recontextualize a gesture they knew rather than successfully recalling the correct gesture taught. Meanwhile, the other two participants that did use an External gesture, performed what would have been the Cultural Groups' command. More on this will be brought up on section 4.4.2.
This row, however, is a bit open to interpretation, namely in regard to Task T3.1. Also noted in the table, all 5 users performed the exact same mistake of doing a Wave motion to solve this task, which is the method by which two other tasks are solved in the Non-Cultural Group. What's noteworthy about Task T3.1 is it involves calling the attention of a blue humanoid helper that is fully animated in reaction to the participants. And one of the actions this animated helper performs

	O1	T1	O3	T3.1	T3.2	O4	T4
Total Mistakes	6	5	8	7	5	5	1
Gestural Mistakes	6	4	3	5	4	5	1
Emblematic Substitutions	6	3	2	5	4	5	1
External Emblematic Substitutions	2	1	2	0*	4	3	1

* All 5 performed the same gesture: Wave; And didn't realize it was wrong.

Table 4.7: Breakdown of Error Types made by the Non-CulturalGroup in the second trial.

to the player once their attention was caught, is Waving their hand in front of its face. There's a strong reason to believe that the users were actually recalling the response of the helper, rather than misremembering the context in which a waving gesture fits within the game.
Ultimately, nearly every substitution ended up resembling the users Cultural Surveys or even the Cultural Group's gesture set. There are strong reasons to believe that well-fitting Emblematic Gestures had an impact in both the short-term learning attainment rate of the game's gesture set, but also of the long term memorization, against lesser-fitting emblematic or options.

4.4 Survey Data

4.4.1 Usability Report

Given that this game was built from scratch, there's one important aspect that was not previously researched. A proper usability test was never performed on it thoroughly. Some aspects of the game that were verified rougher, such as movement, may have perhaps had a chance for improvement with time given towards taking user feedback and reworking of those trouble spots.
So why is measuring usability important during the trials then? Because ease of use, learnability and repeatability are some of the most important indicators by which satisfaction is evaluated, and the majority of methodologies used in Human-Computer Interaction research to validate assumptions into concrete ideas are variations of usability tests. This present work itself is similar in nature to one method of testing usability, known as Split Testing, whereas two versions of an application are handed to different groups and variations in key metrics are determined. In sum, the measurement of usability is indissociable from any form of research involving user interaction.
To perform this study on the system's usability, we resorted to the System Usability Scale[33], the SUS. The SUS is an industry standard questionnaire consisting of ten elements to which users respond in a scale between strong disagreement and strong agreement. These questions were adapted to the game into the post-game survey and rearranged, however the function of the SUS's scoring is intact. Table 4.8 shows these scores.

Between the two groups it's apparent that the Non-cultural group had a lesser impression related to the game's usability. While this was a unanimous complaint, the lowest SUS scores were also coming from the volunteers who most felt the need to verbalize their discontent with the

Cultural	SUS Score
V1	87.5
V3	80
V5	90
V6	85
V9	90
V11	87.5
V13	85
V15	85
V17	92.5

Non-Cultural	SUS Score
V2	65
V4	77.5
V7	72.5
V8	80
V10	80
V12	87.5
V14	65
V16	75

Table 4.8: System Usability Scale scores determined by the users of the Cultural and Non-Cultural Groups

game's movement. The SUS score ranges from 0 to 100, however these are not percentages spread over a uniform distribution. There's no single guideline by which all SUS scores are interpreted, but what is usually accepted as a good parameter, is that a desirable threshold to surpass is a SUS score of 68, which represents an approximated average of most SUS questionnaires, or alternatively a score of 80, which is associated with high task completion[34], although these values may vary depending on the researcher heading the questionnaire, sample sizes and sample selection. There are more benchmarks and percentile breakdowns, but in broad terms, it's valid to state that the Cultural Group's is set above the Non-Cultural Group by at least a ranking in terms of its usability, and thus, the choice of gestures has provided a benefit to the Cultural Group.

4.4.2 Confidence and Culturally-Driven Misplaced Confidence

After the games were completed, the participants were asked about each of the individual tasks they performed on a number of aspects. Aspects such as how they felt about solving them, if the gesture felt natural, among others. One of which, exclusive to the second trial, was their confidence in remembering the task's gestures, and, conversely, if they had trouble remembering it. A few more questions not related to the tasks directly allowed to further get a feel for their personal evaluation on the matter.

From these, it was possible to rate each user's global confidence in solving the game's tasks on a scale ranging from 0 to 30. However, this is not a metric that's very useful on its own. Beyond the fact that a user's confidence may be related to personal facets and quirk's, we're also not looking into if the game on its own inspires confidence, but rather if the cultural emblems within a fitting context are a major differentiator for the users otherwise experiencing the same game. This means that merely comparing the Cultural and Non-Cultural Group's Confidence Score wouldn't be enough.

It was considered important to go back to the game performance assessment and attempt to categorize the sources of confidence in separate ways. The confidence score would remain a global perspective of the user's evaluation, but a sub-score of it would list out how much of that confidence was misplaced and borne of a user performing the wrong action and still clearing the task. Thus, we may create two scores, a participant's Confidence score, and a Misplaced Confidence score. But this approach is still insufficient, as the cultural component is missing. It's by taking in account the types of Gestural Mistakes observed earlier 4.3.3 that this misplaced confidence becomes relevant.

Ideally for the thesis' hypotheses, for each user that committed a mistake in the game yet claimed confidence that they solved it correctly, they would have done so because they misremembered a well-fitting cultural emblem in place of their actual mistaken gesture. As such, another score is required to be broken down from confidence, that of the Culturally-Driven Misplaced Confidence.

	Total	Misplaced	Cultural
V1	26	0	0
V3	27	0	0
V6	24	0	0
V9	23	0	0
V11	19	0	0
V13	30	0	0
V15	26	0	0
V17	28	0	0

	Total	Misplaced	Cultural
V2	21	9	8
V4	22	12	8
V7	20	10	10
V8	30	0	0
V10	24	12	10
V12	25	20	20
V14	23	8	4
V16	22	3	3

Table 4.9: Confidence Scores showcased by Cultural and Non-Cultural groups during the second trial

Since the Cultural group made no definite gestural mistakes, their misplaced scores can only be zero, and their culturally-driven misplacement, while possible in practice, would require issues with the task design itself, which was precisely the reason why two were eliminated from the evaluation. One way or another, while confidence is an important metric for the global experience assessment, the whole Cultural group is not really insightful for this current approach. However, as for the Non-Cultural Group, a lot of the score is valid. Since nearly the complete majority of the gestural mistakes were Emblematic Substitutions, it's no surprise that a lot of the confidence the users had misplaced was sourced to their cultural expectations. Manifestly, 85% of all of the misplaced user confidence score was due to the Non-Cultural trial intentionally breaking their natural conventions. Given an application with more direct forms of feedback where it comes to user failure, this data favors the gesture set employed by the Cultural group enormously.

4.4.3 Global Experience and Immersiveness Indicators

To properly evaluate the global experience of the game in both game's post-tests, a choice was factors considered relevant would be required. Giving the volunteers more freedom in reporting their relevant determinants was considered and would be valuable for the general evaluation. Such as the PANAS scale, for a tallying of subjective mood. However, the need for a comparative evaluation between test groups dictated that the post-test surveys would have to be more focused at a pre-determined set of elements between the two user groups, that covered more aspects than affectivity.

With other questions in the post-test, the game, the Content component in the UX triad, was already looked at in several manners. A proper UX questionnaire ought to look into the other two elements of interaction: The Person, their experience, their anticipations, their characteristics, their relationships; and the Form, the technology by which the game is delivered, the extra content not part of the game's primary progression. The latter is interesting mostly in aspects that may have affected the former, as people's focus and attention is performed in a generally scattered yet not truly expectable manner. What some find amusing, distracting and correctly perceive, others may fail to because of form factors. That is to say, most of the questions in the post-game were majorily introspective, framing questions from the perspective of what the participants felt and responded to, rather than questioning about the game (E.g. "I have enjoyed myself" instead of "The game was enjoyful").

The following were the indicators chosen: Emotional Impact; Internal Expectations; Self-Consciousness; External Expectations and Sharing; Recall and Recognition; Enjoyment and Repeatability; Subjective Sense of Comfort; Technological and Methodological Impact; Symbolic Feedback and Sense Making. Some indicators were also ascertained in other works but didn't feel like they fit well with the current, such as Monetary Value, the willingness to pay money for similar experiences; or Reputation, which did not make sense given the fact the technology was almost entirely unknown to the participants prior. These were asked to the volunteers through questionnaire, and the answers were weighed to a scale ranging from 0 to 50, the higher the better. Table 4.10 shows the scores for each indicator, for each of the Cultural and Non-Cultural Groups.

Cultural Group	score
Emotional Impact	43.75
Internal Expectations	37.50
Self-Consciousness	47.92
External Expectations and Sharing	43.75
Recall and Recognition	43.75
Enjoyment and Repeatability	43.75
Subjective Sense of Comfort	43.75
Technological and Methodological Impact	40.63
Symbolic Feedback and Sense Making	37.81

Total
382.60

Mean
42.51

Std Deviation
3.31

Non-Cultural Group	score
Emotional Impact	50.00
Internal Expectations	28.13
Self-Consciousness	45.83
External Expectations and Sharing	43.75
Recall and Recognition	24.31
Enjoyment and Repeatability	46.88
Subjective Sense of Comfort	47.92
Technological and Methodological Impact	34.38
Symbolic Feedback and Sense Making	21.41

Total
342.59

Mean
38.07

Std Deviation
11.12

Table 4.10: Immersiveness Factors and Scores between the Two Groups

There was a tendency by the users to give consistent responses sticking mostly to full or light agreement. This is why, despite having different number of dimensions (questions), 4 of the parameters in the Cultural Group had the same score in a pattern. The Non-CulturalGroup scores lower than the Cultural overall, and it would appear that three of the factor are very much lower than the Cultural's. But it's not credible enough that this is a significant difference without data analyticals.
Performing a double ailed two-sample t-test, we can attempt to validate an hypothesis by bidding to prove an opposing one. By seeking to credit that the samples have similar characteristics, we could perhaps find suggesting evidence at a given level of significance that the games had different impacts on the players. However, at a significance level of 0.05, we already find no convincing enough evidence (t= 1.149818588, tc= 2.262157163) that the two samples differ significantly. And given that at, at p-value 0.05, there's at least 23% (and typically close to 50%) chance of incorrectly suggesting the alternate hypothesis, this proves to not be a valid vector by which difference between the two groups can be established. In summation, the immersiveness indicators were inconclusive.

4.4.4 Loose Observations

Both cultural and non-cultural groups were requested to name the most memorable part of the game. Out of 4 exceptions, both parties named the same element. Those exceptions include 2

users speaking of their frustration with movement, and 2 speaking of the moment of victory. All remaining 13 users singled out the blue animated humanoid NPC due to the richness of its responsiveness and interaction. Also, as noted earlier, this NPC reacted by performing its own gestures, two of which were done in a similar fashion to the Cultural Group's (Waving and Pointing), which may have influenced the Non-Cultural Group's memory of the experience.
On the second part of the second trial's post-test, after the being questioned about the tasks, the volunteers were asked about substitutions to the gesture set, while also being reminded of what the actual gesture set was. Out of 19 (17 from the Non-Cultural Group) suggestions, only 1 did not involve a cultural gesture, and only an additional 2 didn't involve emblematic gestures. Non-Cultural Group participants were fixated on gestures that ended up belonging to the Cultural gesture set.

4.5 Conclusions

This work reported on a number of findings of two sets of user trials to assess impacts of measures on the quality of several aspects of UX, by means of, primarily, a comparative test. The purpose of these findings was to validate three initial propositions offered before the start of planning and development of the work. Of these three items, two have been found conclusive, while another did not come to be, and should require a different approach with a number of alterations.

- **Focusing on user culture contributes to the learning rate and capacity of commands; and to the Retention and Memorization of commands and concepts**
  On several passages throughout section 4.3 it is referenced that the majority of the Cultural Group performed exceptionally well during both game trials. Very few errors and failures were registered on the first trial, and coming into the second, the group even flourishingly carried on to manifest growth and improvement over the earlier trial. This in contrast to the Non-Cultral Group, which had trouble assimilating and present the corect commands, and in due, saw a worsening development of performance after the two week delay. Aspects of confidence, anticipation and memory were also among the descriptors looked into that illustrated a reduced positive impact on the Non-Cultural Group independent from the performance evaluation. Coming back to said evaluation, it becomes palpable that the Non-Cultural Group didn't just become unsucessful and fell short based on the choice of gesture set given, but moreover, also due to the group seeking out the what was the approapriate gesture for the game's context, which coincided with the Cultural Group's gesture set. This persistence even poisons the Non-Cultural recolection, making them feel certainty and assurance which has been ultimately misplaced, indicating that the Cultural gesture set was not just prefered among the two, but potentially, the most natural choice for the situation.

- **Focusing on user culture contributes to the satisfaction and immersion of the experience**

Two measures were performed primarily towards this during the work. The usability report was insightful, proving that there was a highly significant difference *(p=0.0031)* in the game between the Groups in terms of the system's capabilities to deliver a satisfying entertaining experiece. But one set of measures wasn't enough, and one another more comprehensive invstigation into immersiveness and satisfaction did not produce further corroboration. It's not improbable that this benefit is still factual, and it's possible to suppose from certain observations and moments in the game, particularly involving task T3, that this could be better observed with greater quality of superficial design.

4.6 Closing Thoughts and Future Work

With this work, we have consolidated the idea that cultural emblems are natural and can symbolically benefit users to organically with the initial barrier of entry to using a NUI. By testing a game built around the context of solving gestures, users are prone to resorting to these emblems they know of, and are moreover, it makes it accessible if recall the commands if necessary. But more needs to be tested and done so more rigorously.
Satisfaction and Immersion have not been properly justified, as noted above, but there's more to investigate. The current findings all aim towards benefits involving the initial usage hurdle of an application, it tests out easy to perform tasks, even if these involve multiple steps. While there was a plan to look into separating tasks into easy and hard tasks, this never quite came into fruition. The justification for Optional and Mandatory tasks was such that, besides the very first optional task, these would all involve larger struggle towards resolution. Some ideas for what this difficulty would ensue were written down, such involving gesturing towards objects with motions that would make sense only if aiming to address an intelligent receiver. However, neither was any solution really acceptable, the approach itself didn't seem very lenient towards a good analysis. Still, it would be deemed really useful to make one potential analysis of the upper parameters of complexity. To test out if adaptable Cultural Gestures do provide a benefit to Depth of Interaction. While learning and memory as advantages may be upheld right now, what degree of richness the interaction achieves is not.
Another aspect of the application was unsatisfactory. Player Movement was too contrived by a composite issue of the transition between movement gestures, the Leap Motion having certain faults in perceiving movement with the palm pointed upwards, and with the game itself not smoothing out the rotation speed of turning very gracefully. However, in hindsight, it's very possible that movement in general may have not been a very insightful act to observe in the first place. Additionally, movement is not something that is customarily performed in a natural fashion even in recent immersive systems, such as virtual reality experiences, which relies heavily on either unconventional player teleportation to achieve large distances or using a controller's analog stick for advancing through the world, both approaches which clearly unfit for a fully Natural Interface.

Thus we suggest the following directives towards a development of future system: Improve the

Classifier, bring a new and better approach to the recognition of hand gestures such as machine learning approaches; Create a Virtual Reality environment to test the participants in a more immersive setting, or alternatively, create an Amplified Reality game such that the environment is virtually super-imposed over the real world setting; Focus primarily on the Task setup, remove long duration requirements between each focal point of the observation, such as movement; Make a good definition of task failure and give players a fixed number of attempts or time to complete them, make sure to employ proper feedback within the application on both aspects; Create a difficulty curve, make it so some of the tasks are clearly devised as more difficult than others; Test multiple cultural settings, find valid gesture sets for more than one culture and verify that the benefits permeate all of these over a non-cultural control group.

References

[1] Steve Mann. *Intelligent Image Processing*. John Wiley & Sons, Inc., New York, NY, USA, 2001.

[2] NUIGroup Wiki FAQ, http://wiki.nuigroup.com/Natural_User_Interface, Accessed: 2018-06-09.

[3] Marco Roccetti, Gustavo Marfia, and Angelo Semeraro. Playing into the wild: A gesture-based interface for gaming in public spaces. *Journal of Visual Communication and Image Representation*, 23(3):426 – 440, 2012.

[4] Donald A. Norman. Natural user interfaces are not natural. *interactions*, 17(3):6–10, May 2010.

[5] Donald A. Norman and Jakob Nielsen. Gestural interfaces: A step backward in usability. *interactions*, 17(5):46–49, September 2010.

[6] Ben Shneiderman. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 3rd edition, 1997.

[7] Jakob Nielsen and Rolf Molich. Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '90, pages 249–256, New York, NY, USA, 1990. ACM.

[8] Alessio Malizia and Andrea Bellucci. The artificiality of natural user interfaces. *Communications of the ACM*, 55(3):36–38, 2012.

[9] Jenny Preece, David Benyon, and Open University. *A Guide to Usability: Human Factors in Computing*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1st edition, 1993.

[10] Tiago Susano Pinto. An approach of a research tool based in a shamanic interface. 2015.

[11] Mark Weiser. The computer for the 21 st century. *Scientific american*, 265(3):94–105, 1991.

[12] Leonel Morgado. Cultural awareness and personal customization of gestural commands using a shamanic interface. *Procedia Computer Science*, 27(Dsai 2013):449–459, 2013.

[13] Geert Hofstede. Cultural differences in teaching and learning. *International Journal of intercultural relations*, 10(3):301–320, 1986.

[14] Aaron Marcus and Valentina-Johanna Baumgartner. Mapping user-interface design components vs. culture dimensions in corporate websites. *Visible Language Journal*, 38(1):1–65, 2004.

[15] Alma T Mintu. Cultures and organizations: Software of the mind, 1992.

[16] Abdul Azeez Erumban and Simon B De Jong. Cross-country differences in ict adoption: A consequence of culture? *Journal of world business*, 41(4):302–314, 2006.

[17] Elke Duncker, Javed Anjum Sheikh, and Bob Fields. From global terminology to local terminology: a review on cross-cultural interface design solutions. In *International Conference on Cross-Cultural Design*, pages 197–207. Springer, 2013.

[18] Karen J Pine, Nicola Lufkin, and David Messer. More gestures than answers: children learning about balance. *Developmental psychology*, 40(6):1059, 2004.

[19] Manuela Macedonia and Katharina von Kriegstein. Gestures enhance foreign language learning. *Biolinguistics*, 6(3-4):393–416, 2012.

[20] Adam Kendon. An agenda for gesture studies. *Semiotic review of books*, 7(3):8–12, 1996.

[21] David McNeill. So you do think gestures are nonverbalp reply to feyereisen (1987). 1987.

[22] David Mcneill. Gesture: A psycholinguistic approach. 2014.

[23] S. Mitra and T. Acharya. Gesture recognition: A survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 37(3):311–324, May 2007.

[24] David Mcneill. Hand and mind: What gestures reveal about thought. 27, 06 1994.

[25] Peter; Collett Peter Morris, Desmond; Marsh. *Gestures, their origins and distribution*. Maddison Books, 1979.

[26] Jenny Preece, Yvonne Rogers, Helen Sharp, David Benyon, Simon Holland, and Tom Carey. *Human-Computer Interaction*. Addison-Wesley Longman Ltd., Essex, UK, UK, 1994.

[27] Thomas K. Landauer. Let's get real: A position paper on the role of cognitive psychology in the design of humanly useful and usable systems. In RONALD M. BAECKER, JONATHAN GRUDIN, WILLIAM A.S. BUXTON, and SAUL GREENBERG, editors, *Readings in Human–Computer Interaction*, Interactive Technologies, pages 659 – 665. Morgan Kaufmann, 1995.

[28] Yvonne Rogers. New theoretical approaches for human-computer interaction. *Annual review of information science and technology*, 38(1):87–143, 2004.

[29] Jenny Preece, Yvonne Rogers, and Helen Sharp. *Interaction design: beyond human-computer interaction*. John Wiley & Sons, 2002.

[30] César de Souza. Sequence classifiers in c#: Hidden markov models. https://www.codeproject.com/Articles/541428/Sequence-Classifiers-in-Csharp-Part-I-Hidden-Marko, Accessed: 2018-12-28.

[31] Ronald K Hambleton. On the use of cut-off scores with criterion-referenced tests in instructional settings. *Journal of Educational Measurement*, 15(4):277–290, 1978.

[32] Loring J Ingraham and Christopher B Aiken. An empirical approach to determining criteria for abnormality in test batteries with multiple measures. *Neuropsychology*, 10(1):120, 1996.

[33] John Brooke et al. Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7, 1996.

[34] Jeff Sauro. *A practical guide to the system usability scale: Background, benchmarks & best practices*. Measuring Usability LLC Denver, CO, 2011.

REFERENCES

Appendix A

Code Snippets

A.1 Cultural Definition of Gestures within Game Controller as part of the Cultural Layer's Initialization

```
public static void InitCulturalLayer() {
  //Common Gestures
  culturalLayer.AddDefaultGesture("NOTHING", "OPEN_HAND");
  culturalLayer.AddDefaultGesture("CANCEL", "WAVE");
  culturalLayer.AddDefaultGesture("NO", "THUMBS_DOWN");
  culturalLayer.AddDefaultGesture("YES", "THUMBS_UP");
  culturalLayer.AddDefaultGesture("THUMBS_DOWN", "THUMBS_DOWN");
  culturalLayer.AddDefaultGesture("THUMBS_UP", "THUMBS_UP");
  culturalLayer.AddDefaultGesture("FRONT", "POINT_FRONT");
  culturalLayer.AddDefaultGesture("BACK", "POINT_BACK");
  culturalLayer.AddDefaultGesture("RIGHT", "POINT_RIGHT");
  culturalLayer.AddDefaultGesture("LEFT", "POINT_LEFT");

  //List of default gestures here.
  culturalLayer.AddDefaultGesture("QUIET", "THUMBS_DOWN");
  culturalLayer.AddDefaultGesture("COME_HERE", "WAVE");
  culturalLayer.AddDefaultGesture("PHOTO_FRAME", "OK_RING");
  culturalLayer.AddDefaultGesture("IMPATIENT", "STOP");
  culturalLayer.AddDefaultGesture("PHONE", "LIFT_POINTER_FINGER");
  culturalLayer.AddDefaultGesture("ATTENTION", "THUMBS_UP");
  culturalLayer.AddDefaultGesture("POINT_AT", "POINT_FINGERGUN");
  culturalLayer.AddDefaultGesture("GO_AWAY", "WAVE");
  culturalLayer.AddDefaultGesture("CELEBRATE", "THUMBS_UP");

  //List cultural gestures here
  culturalLayer.AddCultureGesture("QUIET", "PT", "QUIET_NEW");
  culturalLayer.AddCultureGesture("COME_HERE", "PT", "COME_HERE");
  culturalLayer.AddCultureGesture("PHOTO_FRAME", "PT", "PHOTO_FRAME");
  culturalLayer.AddCultureGesture("IMPATIENT", "PT", "IMPATIENT");
```

```
  culturalLayer.AddCultureGesture("PHONE", "PT", "SHAKA_DOWN");
  culturalLayer.AddCultureGesture("ATTENTION", "PT", "ATTENTION");
  culturalLayer.AddCultureGesture("WAVE", "PT", "WAVE");
  culturalLayer.AddCultureGesture("POINT_AT", "PT", "POINT_AT");
  culturalLayer.AddCultureGesture("GO_AWAY", "PT", "SHOO");
  culturalLayer.AddCultureGesture("CELEBRATE", "PT", "FIST_PUMP");
}
```

A.2 State Definitions within the Game Controller to be sent to the Cultural Layer by demand of the Character Controller

```
public static State GoAwayState() {
  State state = new State("Shoo Away State");
  state.AddGesture("NOTHING");
  state.AddGesture("GO_AWAY");
  return state;
}

public static State VictoryState() {
  State state = new State("Victory State");
  state.AddGesture("NOTHING");
  state.AddGesture("CELEBRATE");
  return state;
}

public static State PhotoState() {
  State state = new State("Photo State");
  state.AddGesture("NOTHING");
  state.AddGesture("PHOTO_FRAME");
  return state;
}
```