Designing a Taxonomy for Evaluating Virtual 3D Interfaces and Experiences

TMAC, Toronto, April 12-14th

By Roxanne HENRY

A thesis exhibition presented to OCAD University in partial fulfillment of the requirements for the degree of Master of Arts in Digital Futures

Toronto, Ontario, Canada, April 2019

Declaration of Authorship

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including final revisions, as accepted by my examiners.

I authorize OCAD University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

I understand that my thesis may be made electronically available to the public. I further authorize OCAD University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

OCAD UNIVERSITY

Abstract

Digital Futures

Master of Arts

2019

Designing a Taxonomy for Evaluating Virtual 3D Interfaces and Experiences

by Roxanne HENRY

As virtual reality climbs out of the Trough of Disillusionment, now is the time to be developing design methods and tools. In the effort to make VR as immersive as possible, lots of new hardware technologies have been developed, and design methods are being explored to make better use of them. However, the effect of wearing all this hardware hasn't yet been explored in detail. Is more hardware required for experiences to be more immersive? Does wearing so much hardware ever get overwhelming? This thesis attempts to discover the relationships between virtuality, interaction and hardware encumberment, and offers a new comparative tool which can help designers make more informed decisions about the use of hardware in their experiences.

Acknowledgements

Alexis and Emma, thank you for your advice and guidance.

Alexander, thank you for your help.

A special thank you to the Barford family for your support.

Contents

List of Figures

Chapter 1

Introduction

Virtual reality. It sounds a bit like a contradiction and it feels a little like it too. Unreal unreality. Milgram and Kishino (1994) define virtual reality (VR) as a "[...] completely synthetic world" in which "the participant-observer is totally immersed in, and able to interact with [...]" it. It isn't so simple, however. The paper goes on to explain that sometimes, fully virtual environments or even immersion itself are not always necessary to describe virtual experiences and that there exists an entire spectrum of possibilities where augmented reality (AR) and other forms of mixed reality (MR) reside. AR, being one of the specifically defined classes is described as being an experience that is mostly reality, but with virtual elements superimposed. MR refers to the entire subclass of AR and VR experiences, where reality and virtuality are combined in some way. Finally, full reality and full virtuality exist at opposite ends of the scale (Figure 1.1). This thesis will focus primarily on the more virtual end of the scale for the sake of scope, but some AR experiences are discussed in the effort to contrast and compare.

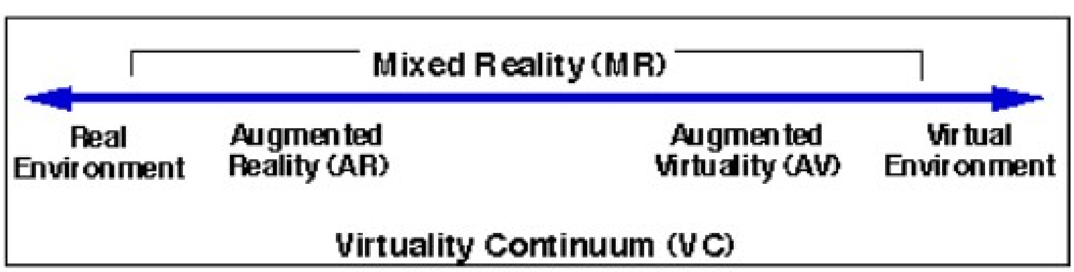


FIGURE 1.1: "Simplified representation of a virtuality continuum" (Milgram and Kishino, 1994)

Virtual reality is rapidly becoming more available to the everyday consumer, though adoption is somewhat stalled, having not "escaped the early adopter, game-oriented segment" (Anderson, 2019). From personal use through gaming, geocaching, and telepresence, to professional use like CAD modelling, training, and remote control of robotics, the technology is becoming more widely accessible and increasingly common (Billinghurst et al., 2015). But according to SuperData, investment in VR software has dropped over fifty percent in the last year, despite the eight percent rise in VR headset shipments (Anderson, 2019)). Meanwhile, a survey inquiring about consumer behaviour patterns has detected that although active VR consumers are very satisfied with their experiences, those who are not consumers are simply uninterested, and researchers have cited that part of this reason is that they have not yet experienced the immersiveness VR can offer. Another cited reason for slow adoption found in this survey was "adoption barriers [...] like price and technological invasiveness" (Boland, 2017).

As for existing software, although some truly wonderful experiences exist, there is a lack of consistency in quality. Often, experiences leave much to be desired, either causing physical discomfort like nausea and dizziness thanks to hardware limitations—latency issues being one of the primary culprits (Jerald, 2015)—or leaving the user confused and lost, through poor mapping of affordances and interface design. The aforementioned survey found that users want to see "more and better" content for VR before adopting (Boland, 2017). These hiccups and design flaws are all part of the natural process of learning to design for new technologies, and it is the consequence of a lack of standardization in the field of interaction design for 3D virtual environments (Jerald, 2015; Malaika, 2015).

In the case of interaction design for flat screen interfaces, certain standards of design have been adopted on the general agreement that they are functionally efficient strategies or commonly agreed upon definitions, called platform conventions or taxonomies respectively (Nielsen, 1994a), and help to keep consistency between different software in order to make it easier for new users to learn how to use it. These standards are typically developed through years of trial-and-error, and building off of what users adapt quickly or not (Reimer, 2005). Despite VR's long and rich history, we have only recently achieved high-resolution head-mounted displays, mass-produced at low enough costs for consumers to begin adopting them (Rubin, 2014). Prior to this, the problem of technical feasibility largely stood in the way of significant progress in the field of interaction design for VR, therefore we are only now in a position to start iterating on design standards for VR. Taxonomies, however, have been under development since the first notions of VR were conceptualized, such as those put forth by Kjeldskov (2001), Milgram and Kishino (1994), Robinett (1992), and Zeltzer (1992) to name but a few.

A lot of the work that has been done on standardization of design practices in VR is scattered between several different industrial players and researchers, all aiming for the same goal, but approaching it with vastly different techniques. Valve, Meta, Oculus, Leap Motion, to name but a few, are, or were, all looking for some way to find better, more intuitive interactions, design methods and best practices which will help users have more pleasant experiences in VR. Valve, for example, a game company developing for both Oculus Rift and HTC Vive HMDs, have presented their research at several conferences, sharing their findings on taxonomizing new techniques for interaction and ways of reducing user discomfort with regards to user teleportation (Malaika, 2015; Abrash, 2014). Oculus, on its front, provide developers like Valve with their own research and findings by compiling developer kits, packed full of information about best practices when it comes to developing for VR, highlighting the "unique physicality" of VR and the best ways to manage this (*Introduction to Best Practices*). Leap Motion, a motion-tracking and AR hardware company, have been busy at work trying to find ways that shortcuts can be adapted for VR, since keyboard shortcuts—like ctrl+c for copying on a computer—aren't available in most VR experiences: they have developed gesture-based shortcuts, being mindful of user fatigue and leveraging pinch motions (Fox and Schubert, 2018).

As for academic researchers, there have been several attempts to describe VR and its surrounding theory in order to make it easier to talk about the new aspects of design which VR brings with it. We have the virtuality scale as briefly introduced above, contributed by Milgram and Kishino (1994). Kjeldskov (2001) contributed interaction categories, describing concrete ways in which users can interact with virtual environments. Additionally, the interaction categories were compared for use with different hardware configurations, indicating which combinations worked more efficiently.

The overall goal these researchers are aiming for, of course, is user satisfaction with VR, and ultimately, user adoption of VR. Consumers have been let down by the hype for VR before, when back in the 90s, and throughout the early 2000s several VR experiences were hyped and flopped, due to the hardware causing discomfort, and lack of adoption by software developers, despite the excitement and promise it showed (Rubin, 2014; Snow, 2007). Commercially available hardware is necessary to experience VR at home, but hardware on its own is nothing but a glorified paperweight without good software to go with it. But software companies are hesitant to invest their energy into VR without the promise of a return on this investment (Anderson, 2019). This is a natural part of the Gartner Hype Cycle, and VR is finally emerging for the "trough of disillusionment" of the 90s. This is apparent in the way VR companies are no longer showing as many flashy demos or generating quite as much buzz as it has in years past, and why consumer adoption of VR has been slow (Anderson, 2019; Boland, 2017; Sinclair, 2018).

As a result, hardware companies are invested in making sure software companies can have an easy time developing for their hardware. The trend among several major HMD companies—Oculus, PlayStation, HTC—is to offer two lightweight motion tracked controllers, something which is allowing software designers to make "cleaner design tradeoffs" (Malaika, 2015), than if there was more variety in the hardware users would have to use to interface with the software. Software companies want their games or experiences to be well received in order to get that return on investment, and to do this, must work closely with the hardware which their software depends on, whether that's just the HMD and their proprietary controller, or custom controllers/interaction tools and additional haptic hardware, etc. Consumers, on the other hand, just want VR that works, to serve one of many desires: some want entertainment, others want simulations and training, while others still want to experience social connections with distant people (Boland, 2017).

The next phase of the Gartner Hype Cycle, "climbing the slope of enlightenment", is where "methodology and tools are added to ease the development process" (Linden and Fenn, 2003), and there is an abundance of research being done to establish these. The wonderful thing about so many different groups putting effort into standardization is that there is a lot of research happening simultaneously and a lot of lessons being learned and knowledge shared: the field is advancing at a rapid rate. Valve is one of the leading industry researchers who have helped to create not only hardware technology through its contributions to the HTC Vive, but who also foster research in software, as well as being host to one of the leading VR experience marketplaces (Lang, 2018). Owlchemy Labs, a small indie game studio, have developed the concept of Tomato Presence, which has been picked up by several developers, including the teams at Valve as being a core design practice for implementing presence (Malaika, 2015). Simply put, this is the concept of transferring the user's intention, or presence, from their own avatar to that of the object they are manipulating, in this case, by replacing the user's hand models to that of the object they have picked up (Figure 1.2) (Bye, 2016; Owlchemy Labs, 2019).

(A) User presence/intention is the hand model (B) User presence/intention signaled to change (C) User presence/intention transferred to tomato

FIGURE 1.2: Example of presence transferring in Owlchemy Labs' Job Simulator (Owlchemy Labs, 2019)

Also considered by the Valve team as an important contribution to the taxonomy of VR is the extension of Fitts' law—a popular model for calculating the difficulty of pointing tasks frequently used in flat-screen interaction design to determine the size of buttons—adapted for 3D environments (Malaika, 2015). Using a spherical coordinate system, instead of the standard Euclidian coordinate system typically used to calculate it, it has become possible to extend the model from use in 2D space to use in 3D space Cha and Myung (2013), which in turn allows designers to use the model to guide the design process by making interacting with interface elements more accessible and less error-prone.

Leap Motion, a motion tracking hardware/software company, for their part, have discovered the value of gesture-based shortcuts to activate different tools (Fox and Schubert, 2018). They have developed a "palm-up pinch to open a four-way rail" system, which allows the user to select different tools within varying contexts using "comfortable" gestures, and without the user having to look at their hands to use it, freeing up some of the mental bandwidth usually spent on navigating interfaces.

However, the unfortunate thing about so many different groups putting effort into the endeavor is that there is such a large amount of variety in design practices being established that it becomes easy to get lost in it all, and best practices are having a hard time surfacing.

1.1 Research Summary

1.1.1 Problem Statement

Designing good user experience for VR is still somewhat of a novelty, at least, in so much as there are still growing pains. In the current landscape, there is a distinct lack of clearly established best practices or comparative models for VR to help designers and developers make informed decisions about experience design.

1.1.2 Importance of VR in the Industry

VR has a lot of potential to be as much a useful tool as a useful toy. From its conception, VR has been used in a lot of practical scenarios, including force feedback for molecular docking, military training simulations, NASA training simulations, telepresence, CAD modelling and more (Mazuryk and Gervautz, 1996; Jerald, 2015). Games were developed for VR in the 90s, but it ended up being such a catastrophic commercial failure, that the 2000s were dubbed "the VR winter" (Jerald, 2015). Today, VR is making a comeback, with commercial hardware finally reaching a point where its quality doesn't make the price inaccessible.

In the coming years, VR is predicted to be adopted in education, training, simulation, modelling, etc., but consumers themselves aren't yet one hundred percent convinced of its ability to deliver (Jerald, 2015). The hardware is ready to go, with HMDs which promise reduced lag and better graphics, and precise motion controllers, but the experiences themselves are extremely hit or miss. In order for VR to take off, consumers have to want to adopt it, but consumers won't be interested in VR systems if they are generally unsatisfied by their experiences.

1.1.3 Inconsistency in the Face of Novelty

In flat-screen interface design, there exist several tools to help evaluate the quality and user-friendliness of designs: Nielsen's heuristic evaluations for general system-to-user interactions, AChecker for accessibility checking, W3C for internet standardization, etc. Many of these are what are called open standards, which Jerald (2015) defines as standards agreed and voted upon by a non-profit organization which accepts memberships from anyone interested in the proceedings. These sorts of tools are developed through decades of lessons learned from trial-and-error.

Due to the history of VR, which has been dominated by the effort to make the hardware feasible, we are only now entering a period in which it is possible to start learning these software design lessons (Sinclair, 2018). To complicate matters further, there are so many options to choose from with regards to hardware—HMDs, input controls, motion trackers, etc.—which can lead to several different branches of best practices to accommodate their requirements. For example, input controls alone can vary from keyboard-mouse, traditional gamepads, joysticks, haptic-enabled exoskeletons, hand-held touch-sensitive controllers, task-specific props, and much more, each offering its own set of advantages and disadvantages. Deciding on which to choose for any given experience is not always trivial.

Additionally, there is a disparity in the process industrial game developers adhere to in order to design UX/UI, versus what indie developers undergo. Industrial game companies tend to play it safer, and are doing intense research into VR, but are reluctant and slow to commit to making games, as some companies have expressed interest in developing for VR several years back, but have yet to produce results (Ashcraft, 2016; Patrick, 2016). Indie developers, however, are leading the race with experimental experiences, not quite steady on their feet (Sinclair, 2018). There are few captivating experiences, released by indie devs who stumble upon new and exciting VR-only affordances (Bye, 2016), and which leverage the unique embodiment mechanics. But there is also an overwhelming amount of lackluster experiences (Bye, 2018). In this climate, some experiences, such as Owlchemy Labs' Job Simulator, are breathtakingly immersive, and players come away excited, surprised by how much time has gone by, and ready to engage in more (Bye, 2016). Other experiences leave players disappointed at best, or confused, nauseous or hurt at worst, unlikely to want to try VR again.

1.1.4 Hypothesis

Developers need evaluation tools and standards in order to be able to validate their design decisions. With the help of these evaluation tools, developers and the industry at large will be better equipped to produce higher quality experiences, leading to higher general consumer satisfaction with VR hardware. With higher consumer satisfaction, opinion of VR will become more favourable, and eventually, this will lead to an improvement in the adoption rates of VR as a suitable technology.

1.1.5 Research Questions

1. What criteria can be used to evaluate interaction design for MR?
2. What are the relationships between the criteria?
   (a) How can we highlight these relationships?
   (b) How do these relationships affect the quality of the experience?
3. What are the steps to create a conceptual tool to help describe, compare, analyze and visualize MR experiences and to improve design?
   (a) How can this tool be used to help make informed design decisions?
4. How can existing design standards validate the conceptual tool?

1.1.6 Goals and Objectives

My goal with this thesis, ultimately, is to provide a conceptual tool from which developers can objectively describe and compare experiences and derive trends in the relationship between the provided dimensions, in order to allow for more informed design choices, which will eventually improve the quality of VR experiences. To do this, I will conduct extensive research on the history and current status of both the field of flat-screen interaction design, as well as VR interaction design. Next, I will synthesize and analyze the information in order to bridge some gaps in the taxonomy of VR, with regards to methods of describing and comparing experiences, and I will create a new conceptual space within which to place experiences. I will select data within specific categories and rate them using the scales of this new conceptual space. Following this, I will use statistical analysis to model the relationships and trends which can be derived from the plotted experiences in the conceptual space. Finally, I will compare the model to existing design standards and taxonomies in order to demonstrate its value within context using an evaluation table. With this new model, it will be possible for designers to make informed design decisions by analyzing experiences within different contexts.

1.1.7 Methodology

This thesis employs an iterative process in order to create and validate a new conceptual space. The five steps are as follows:

1. Literature review
2. Creation of the conceptual space
3. Populating the conceptual space with data
4. Model data through statistical analysis
5. Discussion of final results

Where the "Creation of the conceptual space" is an iterative loop, which persists throughout until the data is modeled. The conceptual space is constructed on a foundation of literature and refined through attempting to map experiences into the space. These experiences are selected based on different contexts and genres, in order to showcase the conceptual space's flexibility. Then, once the conceptual space is finalized and data entered, statistical analysis is performed on the data in order to highlight trends and identify relationships. Finally, the conceptual space is placed back into the field's context by its comparison to existing and similar frameworks, and the results discussed.

1.1.8 Contributions

The major contribution of this thesis will be the conceptual space which can be used to make more informed design decisions. The following are additional contributions, and further information about the tool:

1. a literature review, detailing the history of flat-screen UX/UI design, and its appropriate parallels of the history of VR UX/UI design, as well as a survey of current VR best practices research;
2. a new conceptual space within which it is possible to place experiences in order to visualize the relationships between certain influential factors VR experience design, as well as compare them to other experiences;
3. a model derived from statistical analysis which elaborates on the relationship of these factors and explains what they mean within the context of design;
4. an evaluation and examples of applications for the model, as well as ideas about future research.

1.2 Chapter overview

In the next chapter, I will be exploring the history of VR, flat-screen interaction design, as well as the current climate of the industry. In chapter three, I will be taking a deeper dive into the methodology used and the decisions which have been made in order to accomplish the work required to create the recommendations. In the fourth chapter, I will be diving deeper into the creation of the conceptual space, going over the chosen axes and variations in detail. In the fifth chapter, the model will be explored and explained, the relationships between key factors, and their implication on design being the primary focus. The sixth chapter will go over an extended discussion of the thesis. Finally, in the last chapter, I will discuss the results of the thesis, reflect on the process, its limitations, and elaborate on what future work lies ahead.

Chapter 2

Literature Review

Virtual reality (VR) as we imagine it today, is in the middle of its refinement phase, coming out of its hardware development phase, parallel with the Gartner Hype Cycle's phenomenon of "climbing the slope of enlightenment" (Linden and Fenn, 2003). The Sensorama, a multisensory experience which is largely considered to be one of first immersive, multimodal experiences, was patented by Heilig in 1962, and was one of the first examples of stereoscopic displays, though it wasn't really considered real VR. Sutherland, in 1968, produced a head-mounted display (HMD) with image occlusion called the Sword of Damocles, creating essentially the first fully immersive VR experience. Users could turn their head, and the system would calculate what lines should be visible to the user for each of the user's eyes, making it the very first stereoscopic VR HMD. Both of these examples were large and unwieldy: important proofs of concepts, but not feasible for common applications; as they were stationary, heavy, room scale and didn't allow for free movement. The software in the Sword of Damocles was nothing more than wireframe walls that adjusted to the user's head position. At this point in history, the so-called "technology trigger" (Linden and Fenn, 2003), there were no standards for design since there was no need: the important thing to investigate at the time was “does it work”? With modern hardware being cheaper, lighter, and more accessible to the public, we've finally reached a point where the question “but how can we make it good?” has become imperative.

2.1 What is VR?

In 1965, Ivan Sutherland described a system called the Ultimate Display, which would include a visual monitor, audio display, it would produce taste and smells, and could even provide force feedback through a kinetic display—known today as haptic feedback. According to him, this system of displays could produce the perfect virtual world (Sutherland, 1968). VR, as we know it today comes pretty close. Its current dictionary definition is as such:

> "an artificial environment which is experienced through sensory stimuli (such as sights and sounds) provided by a computer and in which one's actions partially determine what happens in the environment" (Jerald, 2015)

and:

> "the technology used to create or access a virtual reality" (*Virtual Reality*)

If VR software generates the artificial environment, then the most important piece of hardware is the HMD, screen, projection or whatever is used to physically present this environment to the user. Billinghurst suggests that the following are essential characteristics for compelling and immersive experiences: "3D stereoscopic display, a wide field of view (over 80-degrees) and low latency in head tracking" (Billinghurst, 2018). Together, these are used to project or display a virtual environment, and in a lot of cases, interact with it. Recently, the most popular way to experience VR is using a motion-tracking enabled HMD, accompanied by an often hand-held controller which can sometimes include force-feedback.

VR technology, however, hasn't always been as sophisticated as the system which Sutherland described; some would argue that such things as spoken word, kaleidoscopes, and illustrations could be considered "analog VR" Jerald (2015). For the sake of simplicity, we will consider VR to be

1. computer-generated sounds and graphical images presented stereoscopically, which are...

2. projected in real-time to a user such that they are...
3. given the illusion of existing in a completely synthetic environment which aims to be as fully immersive an experience as possible(Mazuryk and Gervautz, 1996; Milgram and Kishino, 1994)

The level of interaction can vary from very little interaction to fully interactive. With this definition, VR is a movie that a user watches through an HMD, seated in a virtual theater, but VR is not a 3D modeled game the user plays on a flat-screen display screen.

2.2 Standards in Flat-Screen Interaction Design

In the field of interaction design, standards are an important tool for designers. New users often rely on past experience and knowledge in order to learn interfaces and ways to interact with media. This was made evident in the evolution of the Desktop style GUI, where once users learned the metaphor, they became used to it, and seeking to leverage user prior knowledge and familiarity, operating system developers continued to use the same metaphor in future iterations (Reimer, 2005). Beyond this, efforts to standardize interaction design for flat screened interfaces have spawned organizations like the W3C for web design which have helped developers looking for answers to questions like "what makes a design intuitive" and "how do I make my designs accessible", and helps to promote long-term growth of the web (W3C, 2018).

A lot of the sources and references on the W3C refer back to something called "usability heuristics" and one of the most popular heuristic evaluations for flat-screen interaction design was proposed by Nielsen (1994b): a list of 10 usability heuristics that aim to ensure ease of use. It was created out of the need to standardize the evaluation methods used by interface designers after they found that most evaluators only caught up to half of usability problems on their own. After asking the evaluators to describe the criteria they were looking for, they were able to aggregate their criteria and simplify the list, thus ensuring that all of the usability problems could be covered. It has since been used as an evaluation tool by interface designers and can be considered a standard of usability for flat-screen interfaces. To be able to standardize something, one must first standardize the language used to talk about. Taxonomies, the classification of concepts and ideas, are an important part of being able to form more complex ideas around concepts and they allow us to move forward in with standardization (Khan, 2017).

2.3 Early VR, Frameworks, and Taxonomies

2.3.1 Hardware and Software

In 1996, VR's primary applications were for training, modelling, telepresence, cooperative working and entertainment (Mazuryk and Gervautz, 1996). In the early 2000s, VR started being used for rehabilitation (Darbois, Guillaud, and Pinsault, 2018). These were during the "peak of inflated expectations" and the subsequent slide "into the trough of disillusionment" of VR's hype cycle, respectively.

Not much has changed about VR applications since, but what has changed is the accessibility to the technology and the quality of the hardware. Though commercial-at-home hardware was available in the 90s, it was very underwhelming and simplistic, and it simply couldn't achieve the level of immersion promised by the hype (Jerald, 2015; Snow, 2007). In fact, full virtual immersion like we imagine wouldn't be possible until about 2010 when the Oculus Rift, with its revolutionary 90-degree field of view (FOV), was designed (Rubin, 2014), thus triggering the climb up the "slope of enlightenment" of the hype cycle (Linden and Fenn, 2003). Other industry giants also designed their own HMDs in the following years, including Sony's Morpheus and HTC's Vive (Thier, 2014; McCormick, 2015). Indeed, early examples of VR took on many forms besides HMDs (which at the time could not provide FOV beyond 80-degrees), including Cave systems, "Fishtank" models, boom-mounted displays and holobenches (Kjeldskov, 2001). With so many different types of hardware available, with so many different levels of FOV, which in turn affects the immersion level, which in turn can affect the type of experiences which are possible for each of the hardware types, it wasn't easy to come up with any kind of knowledge on what makes a good experience. It wasn't for lack of trying, but there was a lot of ground to cover, and since VR was a nearly prohibitively expensive research topic, it was difficult for more people to experiment with the possibilities until more consumer-priced devices became available (Jerald, 2015).

2.3.2 Relevant Frameworks and Taxonomies

One of the most important frameworks and taxonomies in the field for VR is the Virtuality Continuum put forth by Milgram and Kishino in their 1994 paper "A Taxonomy of Mixed Reality Visual Displays", as mentioned in the introduction. This continuum helps to place the different levels of reality and virtuality on a tangible scale, allowing people to talk about mixed reality in a concise and universally understood way and, furthermore, distinguishes the requirements of the different levels of reality/virtuality. The design challenges for AR, for instance, include having a reality upon which the designer has no control to consider before interface designs can be made, whereas, with VR, designers have full control over environment and interface. Having an agreed-upon language with which to talk about the subject, and an agreed-upon definition about what delineates the levels is a good first step to standardization, and this particular framework has been used for discussions about environment since, to discuss the technological requirements of the field.

The second most important framework for this thesis was put forth by Kjeldskov (Kjeldskov, 2001), who introduced a taxonomy on categories of interaction, also introduced in the previous chapter. These were Orienting, Moving, and Acting, and in relation to display types, can help to categorize several experiences. Kjeldskov emphasizes that the relationship between interaction type and display type can help a designer to choose the right display type for their desired interaction type to maximize the advantages each display type offers. Kjeldskov also highlights the importance of categorizing display types as partially immersive, or fully immersive, which relates to Milgram's continuum. Certain display types lend themselves better to certain levels of immersion, which in turn are more or less acceptable on different ends of the continuum.

Another important framework which this thesis will lean on, but which came from a different field, is Noorian and Ulieru's comparative framework (Noorian and Ulieru, 2010). It is slightly similar to the way Nielsen (1994b) conducted the heuristic analysis in that it gathers distinguishing features for each of the proposed candidates and compiles a complete list of possible features which are pertinent. However, this framework also places weight on the features and compares their relevant importance.

2.4 Current Industry and Academic Research

Many industry experts and independent researchers are investigating what makes a VR experience successful—what gets users to enjoy VR, to buy their products. Although most recent VR technology is head-mounted and somewhat portable, there is still a lot of variety in the ways the technology allow for interaction. Some VR systems make use of handheld controllers like the Oculus Rift and the HTC Vive. Others make use of hand tracking like the Meta2, the Magic Leap. The HTC Vive and other HMDs even have a room-scale option, which, combined with the controllers, allow the user to move in real space to interact with the virtual environment. Meanwhile, there has been a lot of research done on the kinds of physical controllers which can give appealing and informative haptic responses to users, like the CLAW and its multi-functional force-feedback system and the work by Provencher in emulating force feedback using tactile feedback.

Most importantly, each are looking for ways to make better VR experiences through experimentation and standardization of their design practices. Abrash released a presentation for Steam Dev Days which outlined the ways VR can be used to enable presence and has advocated for presence to be the primary focus in designing new interactions

(Abrash, 2014). Jerald has published several books detailing perception and the importance of understanding human perception in the process of designing for VR, as well as methods of interaction (Jerald, 2015; Jerald, 2016). Malaika of Valve presented a tentative taxonomy for describing interaction types in VR during a conference (Malaika, 2015). Owlchemy coined the concept of "Tomato Presence" in relation to user intent, in which the user's hand does not always have to be visible for the user to feel immersed as long as the object which represents their intent behaves exactly as they expect it to (Bye, 2016). Meanwhile, Meta, created a whole new operating system with different interaction rules to complement the AR HMD he announced (Meron, 2017). Engineers at Leap Motion have made progress in designing a shortcut and gesture-based interface, leveraging the user's own hands to provide haptic feedback, and have created a custom evaluation setting in order to test and improve it (Fox and Schubert, 2018).

All of this research, however, is being done independently. We are currently in a stage similar to Nielsen's when the need to create a standardized heuristic evaluation for interaction design arose: several experts are dealing with the new technology on their own, coming up with their own frameworks and design practices. However, there could be a benefit in consolidating these practices such that there is a universal and standard framework that anyone could use in order to evaluate their own work.

2.5 Summary

Standards have long helped developers to make good decisions about user interface and experience design. From research about human motor skills to heuristic evaluations, interface designers of flat-screen interfaces have strived to understand what makes an experience or interface a pleasant one. As a result, several standards have been established which offer suggestions for typical design questions.

VR has had several phases of development, following the Gartner Hype Cycle fairly closely. Since the hype peaked and lasted throughout the entire 1900s, and a commercial flop of VR in the late 90s stalled its value in market, plunging VR into the trough of disillusionment, we've only finally entered the next phase almost twenty years later, as interest and hype in VR has been renewed by recent breakthroughs in the quality of the technology.

For this thesis, background research was one of the major tasks, as the results from this step would inform every other step following. It also presented a big challenge in that a lot of the kind of information I was looking for was not available in detail to the public. I was looking for whatever I could find on modern design practices for VR, or for taxonomies that described experiences thoroughly. Unfortunately, it is a common trend in the industry to hoard secrets about intellectual property which can give one company an edge over another as far as proprietary or copyrighted design and best practices go. Most of the information I had access to was through interviews with tech hype and news blogs, or through tech conferences, where only very polished examples would be discussed, and very few of the failing test cases would be mentioned, if at all. Additionally, the discussions about the presented material were often only surface level, barely scratching into the reasoning behind the choices they've made. However, the information I did have access to did shed some light on the current state of design practice in VR: in progress, still trying to find their feet, but working through the major kinks.

As for the historical analysis of VR, I found it was important to understand the way flat-screen interaction design gained its own standards, and what kind of influence these might have had on the developing design standards for VR. Many of the developers working in VR have come from a background of development for flat-screened devices like cellphones or computer monitors, but not all design rules that were useful there can be applied in VR. This prompted questions about which standards and evaluation tools were still applicable in VR, with or without modification, and which ones weren't.

Furthermore, designing for VR has brought about the need for a completely new taxonomy, which introduced new problems to the field of interaction design which flat-screen screens never had Kjeldskov, 2001; Malaika, 2015; Milgram and Kishino, 1994; Robinett, 1992; Zeltzer, 1992. These taxonomies were the key to answering my research questions, and have thus become a major focus of my research.

Chapter 3

Methodology

The methodology for this thesis can be presented as a series of five steps: literature review, which we've already seen, creation of the conceptual space (VAE cube), populating the space with data, model VAE data through analysis, and then a discussion of results (Figure 3.1). The creation of the conceptual space is an iterative one, where each axis or dimension added required the cube's re-evaluation. The literature was already covered in the previous chapter (Chapter 2).

3.1 Creation of conceptual space

The work from this point on becomes a cycle, which continues until a general model can be inferred from the more specific cases. In this step, I have gathered the research and divided it into similar topics: frameworks or taxonomies vs heuristic analysis vs historical context. Each group served a different section of the methodology. For example, frameworks and taxonomies fueled the development of the multidimensional taxonomy by shaping the dimensions. The heuristic analyses were used to validate and evaluate the model. Historical context helped to inform both the need for the tool as well as demonstrate the areas the tool needed to address.

Once the context was determined, it was a matter of establishing the axes and creating quantifiable scales, grounded in the work of previous taxonomies. Works by Milgram and Kishino (1994), Kjeldskov (2001) and Zeltzer (1992) were used to help found the virtuality and activity axes, whereas the encumberment axis had to be created from scratch. As with the rest of this step, this too was an iterative process where corrections to the axes' scales were made as errors occurred.

Finally, a fourth dimension, *genre*, was added to the space in order to be used as an example to showcase the conceptual space's flexibility.

This part of the methodology is explained in detail in Chapter 4.

3.2 Populating space with data

With the space created, it became possible to place existing experiences into it in order to describe and compare them. Using the scales created in the previous step to consistently rate experiences, a selection of games were sorted into a spreadsheet and then mapped into the conceptual spaces. Additionally, the experiences were divided into different genres in order to create a fourth dimension to be used for comparative purposes later on.

Selecting the experiences was a challenge in setting scope. In order to properly showcase the cube's flexibility, it was necessary to concentrate on finding a variety of experiences. To help with this, three contexts were chosen to narrow down the selection of available experiences, and then experiences of a variety of genres were selected out of the most prominent examples of these.

Additionally, the AIP cube by Zeltzer (1992) is also populated with a subset of the experiences. The spread of data is then compared to that of the VAE cube's spread.

A detailed discussion about this step is available in Chapter 4.

3.3 Analysis VAE

Once the conceptual space has been established and data input, analysis of trends and relationships can begin. In this step, statistical mathematical methods were applied to the dataset produced for the cube in order to determine if the experiences within the cube have relationships between one another based on the fourth dimension of *genre*, and whether the value of one axis could be used to predict the value of another. For the purpose of this thesis, only single-variable linear regression was used for this, but it might be useful to investigate multivariate linear regression.

Additionally, the data from the AIP cube is subjected to point-distance calculation and the results compared to those obtained for the VAE in order to discuss the possible relations between the objective and subjective factors.

The results from these methods were then analyzed and discussed as an example of how information in the cube can be used to make informed design decisions.

These methods are described in detail in Chapter 5.

3.4 Summary

This thesis featured a highly iterative process revolving around the creation of a conceptual space and the literature review which supported it. The core of the project is the conceptual space, and every step fed into its creation and validation, including the literature review, and the use of the cube for analysis. The cube was then filled with data, and its flexibility showcased by giving examples of statistical analysis which might lead to informed design decisions. The same was done to the AIP cube, a similar, but more subjective conceptual space, in an attempt to find relationships between the objective factors and the subjective ones. This was done using statistical analysis.

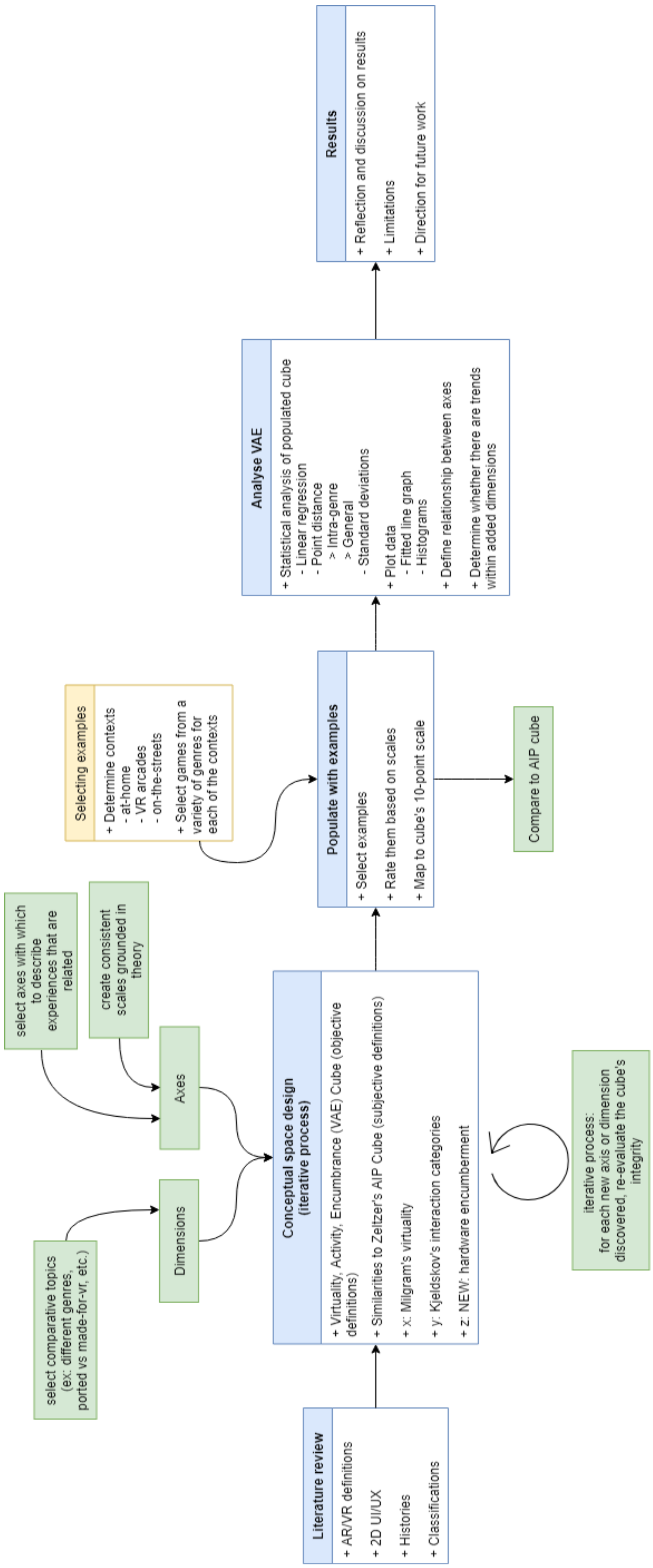


FIGURE 3.1: Chart showing the steps of the methodology.

Chapter 4

Creation of the conceptual space

The bones of this thesis are the conceptual space's components, or dimensions, founded by available literature. Taxonomies introduced by the likes of Milgram and Kishino (1994), Kjeldskov (2001), Zeltzer (1992), are essential in order to describe important concepts about and surrounding VR, and are combined and re-contextualized in this thesis. For example, Milgram and Kishino introduced the concept of virtuality as a scale, allowing us to talk about the many forms virtuality can take. Kjeldskov, on the other hand, has suggested the use of the interaction categories—orienting, moving, acting—to discuss the possible ways users can interact with the environments. He also highlights that the tools, or hardware, the users make use of to interface with the virtual environment can lend themselves better or worse to each of the categories. Zeltzer introduced the AIP cube, a space within which experiences can be described and compared along the axes bearing on user autonomy, interaction, and presence.

The combination of these taxonomies forms the VAE cube, a conceptual space within which games can be described and compared, very similarly to Zeltzer's AIP cube, but where Zeltzer focuses on the more subjective qualities of VR experiences, this cube tries a more objective approach.

4.1 Selecting and Relating Taxonomies

4.1.1 Virtuality

The first important taxonomy this thesis makes use of is the virtuality continuum (figure 4.1) proposed by Milgram and Kishino (1994). This continuum forms one of the axes and is what distinguishes experiences from monitor-based interaction design to artificial reality interaction design, and fully virtual interaction design.

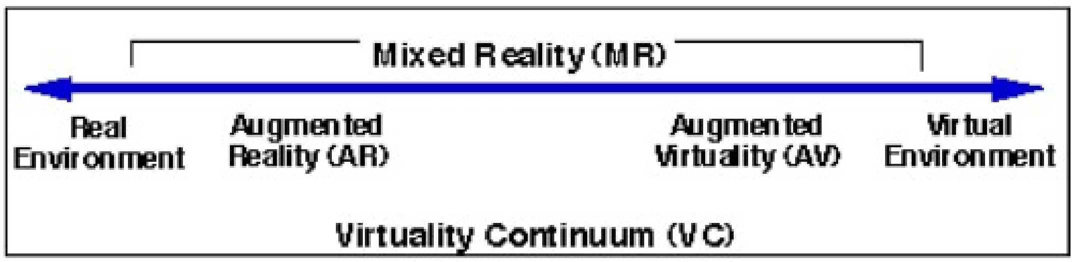


FIGURE 4.1: "Simplified representation of a virtuality continuum" (Milgram and Kishino, 1994)

The continuum allows environments to be placed among the scale into a specific "class" between reality and virtuality. Doing so allows the environment to be defined in terms of its graphical requirements, and thus, its hardware requirements. This can also inform the tools needed for the user to interface with the environment, as real environments will require different tools than fully virtual environments (Milgram and Kishino, 1994).

Milgram's virtuality continuum is presented as a series of six MR "classes", between full reality and full virtuality. The first class describes monitor-based window-on-the-world video displays, and are generally non-immersive. Preferably, these monitors are stereoscopic. The second class is similar to the first, but the monitors become HMDs. The third class still use HMDs, but they have "see-through capabilities", allowing the device to project virtual elements onto an otherwise real-world environment. The fourth class generates the environment virtually, but it is modeled from the immediate the real world, and can still project virtual elements onto this model. The fifth class generates an entirely new environment, whether immersive or not and projects "reality" onto it.

Finally, the sixth class describes partially immersive environments, with fully generated graphics, where reality can "play a role" or "interfere with [...] the generated scene [...]" (Milgram and Kishino, 1994).

This system lends itself well to conversion into a quantified scale (see Figure 4.2), which allows me to accurately rate each experience and provide consistent results for experience virtuality.

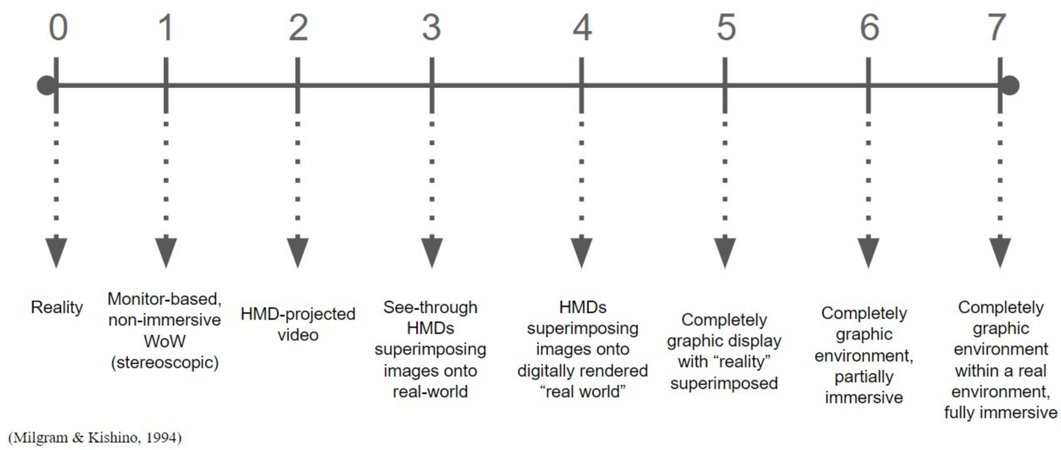


FIGURE 4.2: Quantification of Milgram’s virtuality continuum into a scale.

It is debated whether or not VR or AR will dominate the market (Timson, 2018). While some argue that people’s existing "personal relationship" to their smartphones makes adoption of screen-based AR easier (Schoenfelder, 2018), others still claim that the lack of available hands-free AR hardware might have the opposite effect, and that tech giants like Facebook and Google taking an invested interest in VR might advance it more, though it is possible each will have their niche (Charara and Sumra, 2017). For this reason, though this thesis focuses on games and entertainment experiences primarily, these will cover a broad range of the scale, pulling from both AR and VR libraries.

4.1.2 Activity

Related to the concept of using different hardware in different levels of virtuality, Kjeldskov (2001) discusses hardware requirements on an interaction level, though he focuses primarily on the relationship between the display type and the interaction techniques, determining how well they work in combination. Kjeldskov's interaction techniques serve his interaction categories, as possible ways of manipulating or observing virtual environments, using some kind of interaction device. For this, the device itself plays a "central role" but "does not determine the interaction technique used with it" (Kjeldskov, 2001). The interaction categories are defined like so:

- **Orienting**: Any action a user takes to orient themself in a virtual environment. This includes looking around, whether naturally by moving one's own head, or by rotating the world around the user by using an interaction device.
- **Moving**: The way the user is able to move around in the virtual environment. In some cases, this is a mapping of the user's real movements into virtual space, whilst in others, it requires the world to move around the user, simulating user movement. Often, there is a combination of both. (For the sake of this thesis, teleportation does not count as movement, but would count as an action. Movement implies that the user can somewhat freely navigates the world themself, but teleportation has a set and limited destination, which must be triggered by user action)
- **Acting**: This describes the user's ability to manipulate—select, move, rotate or transform—objects within the virtual environment, as well as the more meta factor of controlling the system itself, usually with some kind of device which maps users' movements, such as a joystick, or controller, or motion tracking.

This is a little bit more challenging to turn into a quantifiable scale (Figure 4.3). As Kjeldskov (2001) suggests, interaction itself is difficult to describe in the context of VR due to its ambiguity as a term. Kjeldskov's definitions of the interaction categories make it clear what types of interaction are useful to discuss within the context of VR, but they talk a little bit less about how much more or less effort the different types require from the user, or how they can be combined. To help mitigate this, Kjeldskov's interaction categories are related to the definition of interaction put forth by Zeltzer (1992): "[...] interaction means the degree of access to model parameters at runtime [...]. The range is from 0 for 'batch' processing in which no interaction at runtime is possible, to 1 for comprehensive, realtime access to all model parameters." The "comprehensive, realtime access to all model parameters" is still a little bit fuzzy, but with this description, we can start quantifying orienting, moving and acting as different levels of "having access to runtime parameters".

So for this scale to work, maximum active participation from the user where the user has "comprehensive realtime access to all model parameters" should equal to 1. Orienting doesn't do much as far as controlling runtime parameters go, and is fairly simple for the user to execute: simply turn one's head, or rotate the world around oneself in order to get an idea of where you're situated. This will have a value of '0.1'. Moving, according to Kjeldskov's description, does not control the system itself in a way that would suggest controlling the system parameters. It can require physical effort if the user's movements are mapped one-to-one, so this will be considered '0.3' points, but moving requires less user physical effort if they are using a joystick, a similar amount to orienting. This will be given the value of '0.2'. Acting can mean a lot of different things. It can range from users being able to interact with objects on a surface level—that is to say, pick it up and rotate it—or it can mean reacting to the environment, around oneself, and changing the environment. For these, we shall use the 4 types of acting that Kjeldskov suggests: selecting ('0.3'), moving ('0.4'), rotating ('0.5'), and transforming ('0.6') objects or the environment. Transforming can also apply to modifying system parameters. Like this, we can add up the different categories as appropriate, totalling to a maximum '1' when maximum user engagement is required, and the bare minimum of engagement is an experience where the user can merely observe the environment around them at '0.1'.

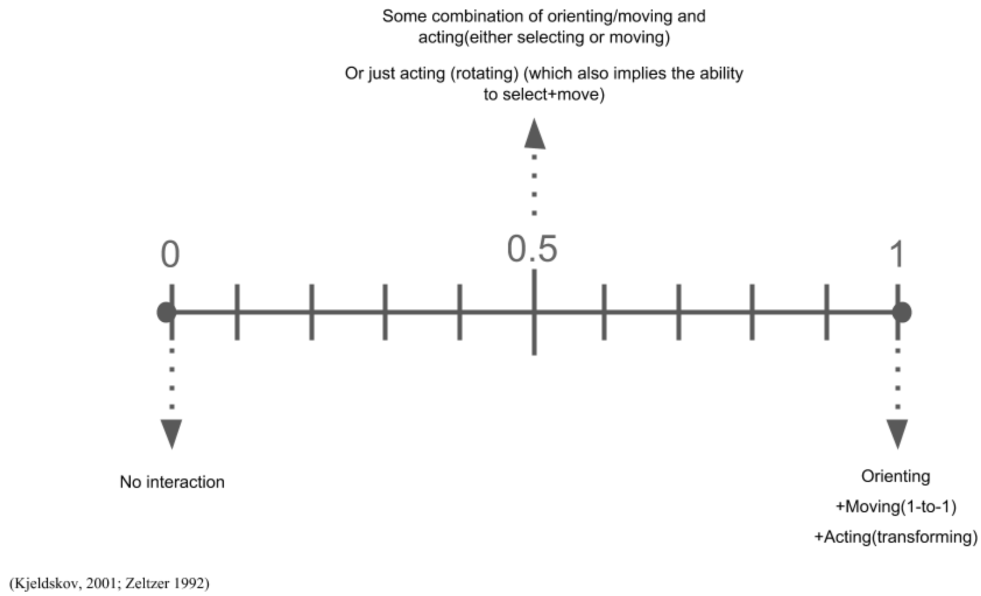


FIGURE 4.3: Quantification of Kjeldskov and Zeltzer's interaction, which describes a user's 'level of activity'

4.1.3 Hardware Encumberment

In both Milgram and Kishino (1994) and Kjeldskov (2001), hardware makes an understated but important appearance. Milgram discusses the dimension of reproductive fidelity with regards to video reproduction technology, outlining at a scale which spans simple wireframes to high definition 3D real-time animation, and the technology required to display it. Additionally, the extent of presence metaphor dimension compares different types of displays and how these have different intentions for user immersion, from simple, monoscopic, flat monitors to stereoscopic HMD displays capable of real-time imaging.

To my knowledge, there hasn't been an extensive amount of research done on how the sheer *amount* of hardware one might sometimes have to use in order to interact with the virtual world while participating in a VR experience can affect user immersion, except for the very telling fact that hardware companies are constantly striving to make lighter and smaller models.

Encumberment in this context will use the following definition: any hardware that displays, computes or is used in interaction between the user and the virtual space. Except for the computer and display—which, in some very rare cases, even this is optional—there aren't, and shouldn't be, any rules about how much more hardware is necessary for experiencing VR. Effectively, there is no fixed possible maximum for how much hardware a user might have to wear for any given experience, so unlike the other scales, this one must be clamped with a maximum value of ten in order to create a finite scale. Additionally, different types of hardware vary a lot in size and weight and have different qualities which affect how much they encumber users. For example, a display on its own can either be just a portable screen or a head-mounted display, and haptic technology can be as simple as vibrating controllers, or a full exoskeleton. This makes adding different types on their own insufficient. Saying an experience has display, computing and interaction hardware vs another which has display, computing, interaction, and haptics does not adequately describe what kind of hardware there is. To tackle this, the hardware had to be divided into functional categories and then given points for the level of encumbrance each type contributed. The points are associated with examples in this situation, as the membership to each group is rather fuzzy:

1. **Computation/Rendering** hardware includes any hardware that computes user interaction and/or renders the virtual environment. Most of the time, these are desktop computers, tethered in some way to the display hardware, but sometimes, they are integrated directly into the display hardware, or worn as a backpack.
   (a) 0: wireless/untethered computer
   (b) 1: tethered computer
   (c) 2: built into display hardware
   (d) 3: backpack computer
2. **Display** hardware is responsible for projecting the virtual environment to the user. This can mean a graphical projection in a cave-like setting, or a smart phone's screen, or an HMD.

(a) 0: projected display (cave-like projection)
(b) 1: small portable screen
(c) 2: light HMD model (e.g. Google Cardboard)
(d) 3: full HMD model

3. **Interface** hardware is comprised of all hardware which the user directly manipulates in order to interact with their environment. Controllers, for example, are interface hardware.
   (a) 0: hands-free/motion control
   (b) 1: light one-handed controller
   (c) 2: medium two-handed controller
   (d) 3: exoskeleton

4. **Motion Tracking** consists of hardware which is used to calculate the user's position in real time-space. These include technology built into other hardware categories, like the display or interfacing hardware, as well as stand-alone motion tracking hardware.
   (a) 0: external/passive motion tracking (camera)
   (b) 1: lightweight puck
   (c) 2: mechanical motion tracker (omnidirectional treadmill)
   (d) 3: exoskeleton

5. **Haptics** are a very large category on their own, and are sometimes built into other hardware but are more often exoskeletons of varying types, including modular and full-body configurations.
   (a) 0: Hypothetical direct neural haptic feedback system
   (b) 1: lightweight localized exoskeleton (ex: Tactsuit pieces)

(c) 2: full-bodied exoskeleton

It is clear that several of the categories exhibit overlap. For instance, an HMD can have some motion tracking, provide haptic feedback, as well as perform its core function of projecting the virtual to the user. For this reason, the order of the hardware categories is based on the importance of the hardware to a VR experience. VR requires some sort of device to compute the virtual, and some kind of display technology to project it to the user. Thus, these have priority over slightly more optional categories like motion tracking, which can be replaced with joystick controls, or even haptics, which certainly enhance the experience, but are not always required. Therefore, if, say the display hardware has built-in computational/rendering power, and motion tracking it will not count as display, computer, and motion tracker (e.g. handheld smartphones); it will simply count as the computer. Similarly, if the interface hardware also acts as a motion tracker, it will not count twice. However, if there are additional motion tracking hardware on the person, these will count on their own.

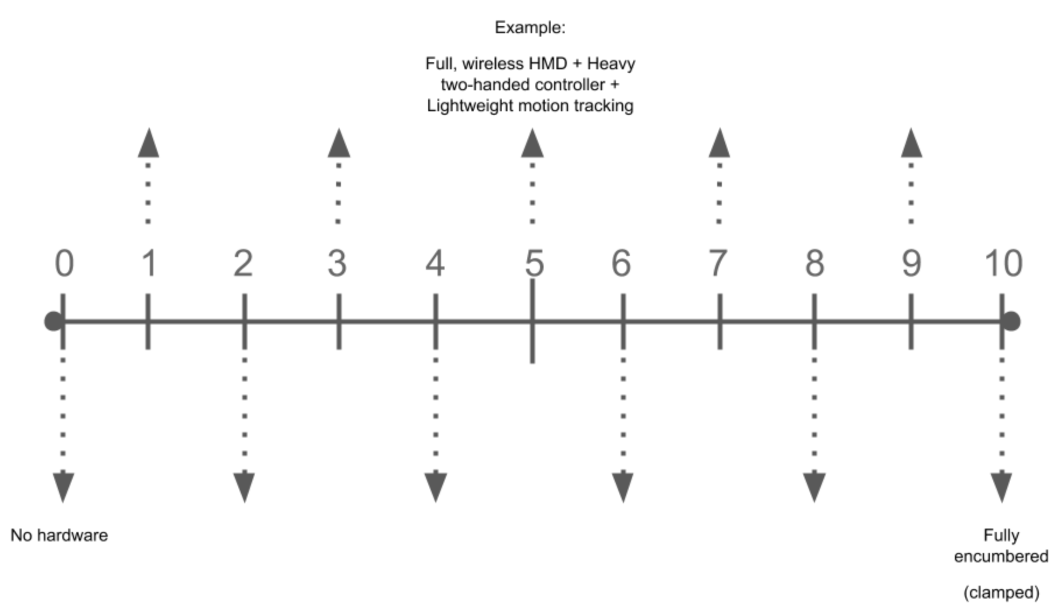


FIGURE 4.4: Level of encumberment scale.

4.2 VAE Cube

With the axes having quantifiable, reliable scales, it is now possible to create the virtual space within which data can be inserted. This space takes the shape of a cube, putting encumberment over virtuality and activity, called the VAE cube. This cube is one of the primary contributions of this thesis.

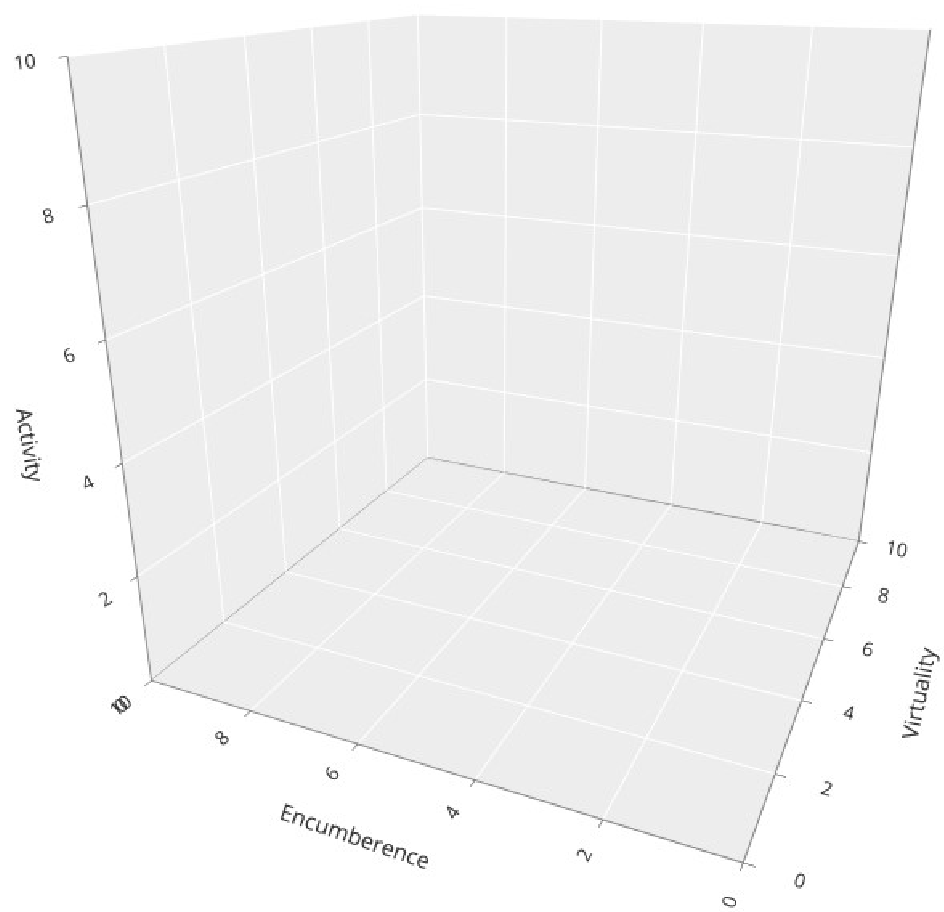


FIGURE 4.5: Virtuality-Activity-Encumberment(VAE) Cube.

This cube can be used to describe, analyze and compare VR experiences. All but two of the extremities of the cube describe experiences which are impossible, either due to contradictions or simply because we do not yet have the technology to accomplish the described experiences yet. The two possible extremities describe default reality and perfect virtuality. However, the spaces in between are all viable, even those close to these extremities.

At the [0,0,0] end of the space, there exists a hypothetical scenario of a user existing in an entirely real environment, unable to interact with it whatsoever. There is no hardware to interface with this real environment. This experience is the first of the impossible scenarios.

Moving along the x axis towards [1, 0, 0], we eventually reach a fully virtual environment. In this hypothetical scenario, the user is existing in the virtual space as they were when they were sitting at [0,0,0], but this environment is entirely synthetic. They cannot interact at all, and there is no hardware on their person which renders the world around her. We do not currently have the technology to achieve this.

Next, we slide up the z axis. Here at [1,0,1] the world is entirely synthetic and the user can engage fully. There is no encumberment here: the world is being generated, and user actions and their effects on the environment are calculated without any hardware on the user's person. Some would qualify this as perfect virtuality: a fully interactable and immersive virtual world (Zeltzer, 1992).

Moving up the y axis for here brings us to the [1,1,1] coordinate. Here we have full virtuality, full interactability, and full encumbrance. In theory, this combination is not necessarily possible: Milgram's virtuality continuum states that any "real physical objects in the user's environment [which] play a role in (or interfere with) the computer generated scene" (Milgram and Kishino, 1994) are not considered fully virtual. The hardware with which the user must interact in order to interface with the virtual world, then, definitely plays a role in the computer-generated scene: it plays the user's role. Thus, the state described by [1,1,1] is merely a theoretical concept instead of an achievable one.

Going back down the x axis, we arrive at 0,1,1, which describes experiences which are entirely real, fully interactive and fully encumbered. This one falls outside the realm of VR games, of course, and the fact that we are removing VR from the equation means the encumberment by VR equipment is suddenly invalid as well: if there is no VR, then the equipment is no longer considered VR equipment, and thus, is simply a person interacting in the real world, wearing a lot of tech.

Speaking of encumberment, we now move down the z axis to visit `[0,1,0]`. This extremity is impossible for the same reasons as the previous one: it describes an experience which is entirely real, but where a user might be wearing a lot of equipment which is no longer considered VR equipment.

Now, we move up along the x axis and meet again with the impossible scenario of a fully virtual experience also having full encumberment at `[1,1,0]`, but this time, there is no possible interaction.

The last extremity, `[0,0,1]` is also the second and last possible experience. It describes the real world: an entirely real environment with no encumberment from interfacing technology, full interaction available to the user.

4.2.1 Comparative dimensions

One of the purposes of the cube is to help visualize comparisons between experiences. This is done by placing new experiences into the cube along with a new dimension. This new dimension needs to tie experiences together into comparable groups, which can be represented in a number of ways. For example, one can insert the genre dimension, to compare where games in different genres sit and determine which, if any, descriptions suit them. Alternatively, one can compare the placement of made-for-VR games, vs that of ported games. One can even compare different studios, different intended audiences, different contexts (experiences meant to educate vs experiences meant to entertain), or whatever other comparisons which can answer questions about trends.

These additional dimensions can even be layered, and a useful dimension to consider adding as a fifth layer while comparing experiences is the quality of the game. Although objectively deciding on the quality of VR experiences isn't a trivial task, it can sometimes be useful to mark down which games you think are successful, and which you don't think are, and then compare these to see if there is a trend between them if you wish to follow this trend. Maybe successful puzzle games tend to be higher on the virtuality and activity scales but lower on the encumberment scale.

Comparative information can then be used to make informed decisions about design choices. If most VR puzzle games don't require a whole lot of encumberment, then users might have this expectation of them. Thus, if you plan on making a puzzle game which uses a lot of hardware to interface with the virtual environment, then you can at least know that this would be considered unusual, and either leverage that fact, or simply choose to ignore it. It's up to you.

This thesis primarily focuses on the comparison between genres, since it guarantees a conveniently large pool of experiences to choose from: every game has a genre.

4.3 Populating the VAE cube

The VAE cube was filled with data in order to run analyses on trends and relationships between axes, as described in chapter 5, as well as to visualize the cube in use and highlight its flexibility with regards to variety. Originally, there was going to be an investigation to compare where "successful" games sat in the cube in relation to less successful examples, but due to time constraints and the challenges associated with finding reliable ratings for experiences, this concept was moved to future works.

4.3.1 Context-based selection

Games selected to add to the sample dataset were chosen based on context, in order to showcase the variety of experiences that can be represented in and described by the cube. It is important to highlight that the cube can work for virtual experiences which are not just VR games and entertainment, but for practical and professional purposes too. Having the different contexts assures that the selection of available experiences can cover the ranges more broadly, even if, for the purpose of staying within the thesis' scope, only entertainment is examined. The three contexts within which games were chosen are the following:

- **at-home experiences**: games a user can install and play easily in most home settings
- **VR arcade experiences**: games available exclusively at VR arcades, where the spaces and equipment tend to be more specialized
- **on-the-street experiences**: games which are meant to be played in a variety of locations

The at-home experiences are meant to cover the range of games which typically make use of very similar hardware setups. For example, a standard hardware setup consists of an HMD and a lightweight controller. Experiences within this context often tend to be entertainment based and come in a variety of genres.

VR arcade experiences tend to be a bit more specialized. Some VR arcades offer experiences which are available in the context of at-home experiences. These are not observed in this situation, but it might be a useful comparison to determine how at-home experience games compare to their twins in VR arcades. For the sake of this thesis, only VR arcade exclusive games are observed. These tend to leverage different hardware setups, often bringing in haptic feedback devices and custom motion tracking. Additionally, social interaction is often featured and some even have the real environments staged to match the virtual one.

Finally, on-the-street experiences cover AR examples, which are more likely to have minimalist hardware setups. This difference in hardware setup, as well as the fact that these experiences are often played in public spaces, mean the requirements for activity might be different or might relate to the other two axes differently.

These three categories together cover a broad range of use cases, even within just the category of games, and will help the showcase the VAE cube's flexibility.

Once these contexts were established, it was a matter of finding games which were prominent in each category, selecting a few examples of different genres for each, and logging them into a datasheet. The datasheet used for this thesis (Figure 4.6) provided options to log the target platform chosen (in the event that more than one target platform exists) and to identify ported games vs made-for-VR games. The datasheet can be modified to include any number of additional axes you might find useful.

ID	Title	Platform	Compatible Hardware	xVirtuality	yEncumberance	zAction	Genre	Port?
0	Beat Saber	PC	HTC Vive	8.6	3.5	10	rhythm	☐
1	Super Hot	PC	HTC Vive	8.6	3.5	10	action	☑
2	Skyrim	Console	Playstation VR	8.6	4	9	fantasy	☑
3	Moss	Console	Playstation VR	8.6	3.5	7	fantasy	☑
4	Somnai	PC	HTC Vive	4.2	3.5	4	location-based	☐
5	Pokemon Go	Phones	Phones	1.4	2	3	location-based	☐
6	Scary Girl	PC	Oculus Rift	7.1	10	10	sci-fi	☐

FIGURE 4.6: Headers used for logging experiences

The Steam library provided an excellent supply of choices for at-home games with a variety of genres and is somewhat hardware agnostic—it doesn't cover the whole range of possible VR headsets but covers many PC-only options. The original intention was to use Steam's rating system to select games which had higher than sixty votes, and which varied in quality. However, it was difficult to find many games with poor ratings that reached the minimum cutoff point, and this idea was scrapped. However, Steam helpfully provides a list of compatible hardware, and basic information about the level of activity required for each experience, making it easier to identify these during the logging process. In some cases, several different hardware setups were listed for the same experience: the most encumbered option was always selected, in order to maintain a consistent selection decision. To cover more ground on at-home experiences, games exclusive to the PlayStation VR system were also selected. These offered somewhat different hardware setups which helped the diversify the dataset.

The search for VR arcade games was a little bit tougher, and involved a lot of generic queries for existing VR arcades and their offered experiences, as well as hunting down review and gameplay footage to assess their virtuality, encumbrance and activity scores.

Again, variety in genre was primarily sought after, which was a challenge in a shooter-dominant field. Some experiences stood out with their uniqueness, including an immersive theater example and a flying simulator. Just as the Steam library made it difficult to find reliable ratings for experiences through its dependency on user action, VR arcades, which rely on professional and user reviews to rate their experiences were equally unreliable in their ratings. Standout professional reviews were always overwhelmingly positive despite individual user reviews making valid complaints. Additionally, the reviews were very hyped in nature and lacked critical analysis of the experience design itself.

On-the-street experiences were unfortunately slightly lost to time constraints, and only one prominent example was included in the dataset.

This data selection process definitely introduced bias. There were several problems with the way the data was selected:

- **Visibility**: due to time constraints, it was impossible to comb through the entire catalogue of available VR experiences.
- **Context**: the primary interest for this exercise was to showcase the VAE cube's flexibility, and games were chosen that could easily be contrasted or compared to one another in varying ways
- **Fetishism of technology and hype**: it was impossible to bypass the way that VR technology is still trapped in the hype cycle's effect of fetishising the technology. This was one of the major contributors to making it difficult to find reliable ratings of VR experiences, as a lot of the reviews were focused on the novelty of VR technology instead of focusing on design aspects of the experience.

Ideally, the dataset would have been double its size for more statistically viable analysis, but due to time constraints, it was restricted to its current size. However, the data does serve to show how the cube works and how data is represented from within it. The full list of experiences is available in Figure 4.7, and the populated cube in Figure 4.8. Genre-specific representations are available in Appendix C.

ID	Title	xVirtuality	yEncumberance	zAction	Genre	Port?
1	Super Hot	8.6	3.5	10	action	☑
21	Eve Valkyrie	8.6	3.5	10	action	☑
16	Gorn	8.6	3.5	10	action	☐
27	Arizona Sunshine	7.1	8.5	10	action	☐
3	Moss	8.6	3.5	7	fantasy	☐
2	Skyrim	8.6	3.5	9	fantasy	☑
12	Netflix VR	8.6	3	6	General Entertainment	☑
13	Youtube VR	8.6	3.5	6	General Entertainment	☑
15	Samsung: Snowboard VR	8.6	5.5	1	General Entertainment	☐
23	Birdly	8.6	6.5	3	General Entertainment	☐
5	Pokemon Go	1.4	2	3	location-based	☐
4	Somnai	4.2	3.5	4	location-based	☐
18	Tetris Effect	8.6	3.5	5	puzzle	☑
9	Keep Talking and nobody explodes	7.1	3.5	7	puzzle	☐
25	Time Travel VR	8.6	3.5	7	puzzle	☐
17	Wipeout Omega	8.6	3.5	9	racing	☑
22	Sprint Vector	8.6	3.5	10	racing	☐
24	Pro Race VR	8.6	6.5	8	racing	☑
0	Beat Saber	8.6	3.5	10	rhythm	☐
14	Electronauts	8.6	3.5	10	rhythm	☐
20	Thumper	8.6	3.5	6	rhythm	☑
7	Surgeon Simulator in Zero G	8.6	3.5	10	sci-fi	☐
8	Lone Echo	8.6	3.5	10	sci-fi	☐
10	Employee Recycling Center	8.6	3.5	9	sci-fi	☐
28	Ghostbusters Dimension	7.1	7.5	10	sci-fi	☐
6	Scary Girl	7.1	10	10	sci-fi	☐
26	Amber Sky 2088	7.1	10	10	sci-fi	☐
29	Dreamscape Immersive	7.1	10	10	sci-fi	☐
11	Job Simulator	8.6	3.5	7	slice-of-life	☐
19	Werewolves Within	8.6	3.5	7	social/board game	☐

FIGURE 4.7: Full spreadsheet of experiences used and their respective values.

4.4 Classification and challenges

Classification of the games is fairly straightforward using the scales, however, there is a little bit of mapping required. The virtuality scale only has seven possible points, but the cube is normalized as a `10x10x10` space for consistency. Thus, a mapping function must be used to translate the seven-point scale into a ten point scale.

The games Sprint Vector and Wipeout Omega will be used as examples for inserting an experience into the cube. These are both experiences which qualify for the racing genre, but they differ on a lot of qualities.

The first, Sprint Vector, is a made-for-VR experience by Servios, an establish VR game development company. It is a first-person racing game, where players must sprint against one another to win.

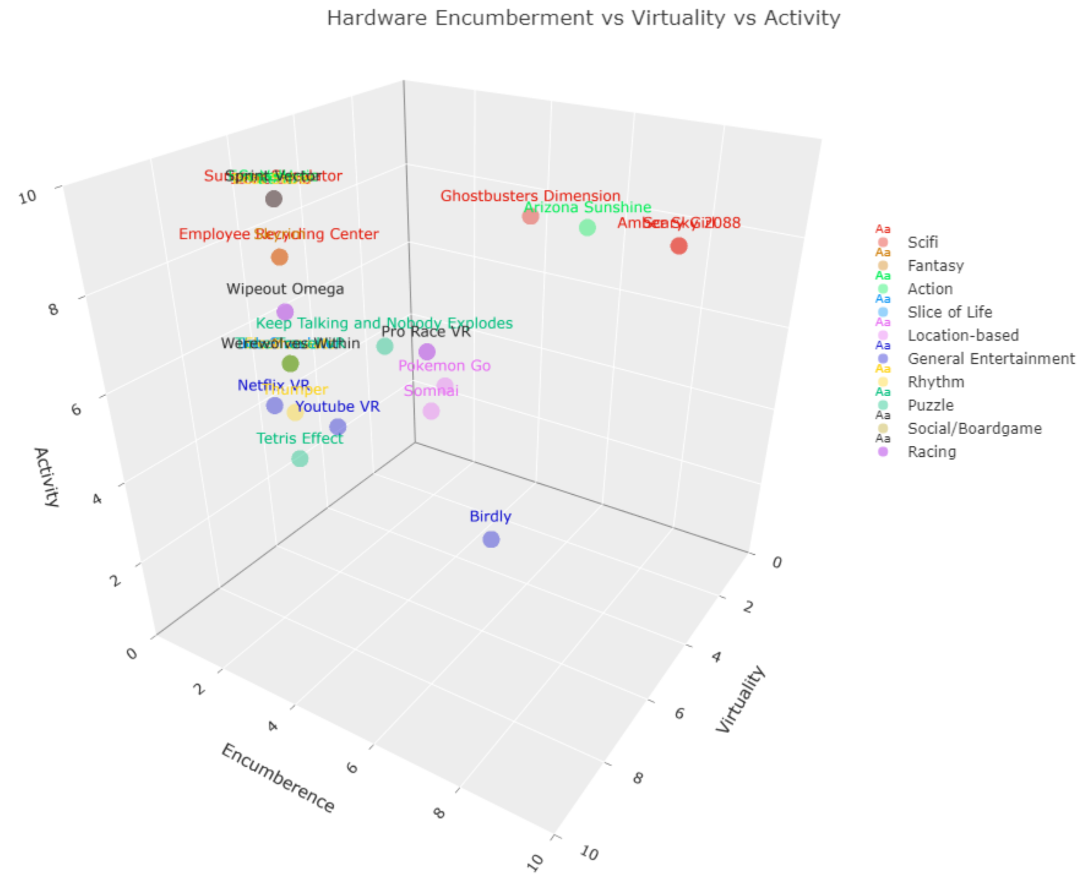


FIGURE 4.8: VAE cube populated with data. The cluster of points on the left represent a majority of the at-home-experiences and can be described with mild encumbrance, high virtuality and a variety of activity levels. The top right experiences show mostly VR arcade experiences, which tend to be more encumbering, but less virtual.

First, the action scale. In this example, locomotion is simulated by users pumping their arms, as though they were running. Turning the body turns the in-game avatar. Additionally, players can fly, climb, shoot objects in the environment and user powerups. Therefore, the player must orient themself `(0.1)`, exert physical energy to move `(0.3)`, and are able to transform objects in the environment `(0.6)`, for a total of `1`–mapped to `10`—action points (Figure 4.9).

Next, the encumberment scale. Sprint Vector is available on many platforms, but the encumberment amounts to the same for each. The player must wear a wired HMD `(2.5)` and use two lightweight controllers `(1)` totally up to `3.5`. There is no additional

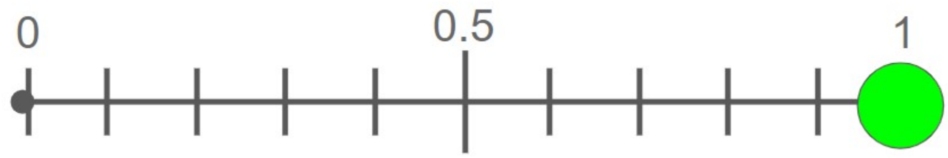


FIGURE 4.9: Sprint Vector's action score

motion tracking, on-person computation, or additional haptics modules on the default setup (Figure 4.10).

FIGURE 4.10: Sprint Vector's encumberment score

Finally, the virtuality of the game. It can best be described as "Completely graphic environment, partially immersive" due to the fact that the user must interface with the environment using hardware which exists in the real environment, netting it a score of `6`, which is mapped to `8.6` for our ten-point scale (Figure 4.11).

WipEout Omega, for its part, is typically a flat-screen racing game by Sony, available exclusively on PlayStation. It's a traditional racing game, in that it can either be played third-person or first-person from within the car's cockpit.

FIGURE 4.11: Sprint Vector's virtuality score

For the action scale, WipEout Omega lets users orient themselves `0.1` and uses a joystick for locomotion `0.2`. Acting is a little trickier to place. The player can definitely move and rotate their vehicle, but this can easily be a moving point instead if played in first-person cockpit mode. However, if played in third-person, the player is now acting on the vehicle, and it can count towards an acting score. Moreover, players can destroy other vehicles, which counts as a form of transforming the environment. Thus, it gets `0.6` for acting. It's total tally is just under Sprint Vector's at `0.9` (mapped to `9` on the ten-point scale) (Figure 4.12).

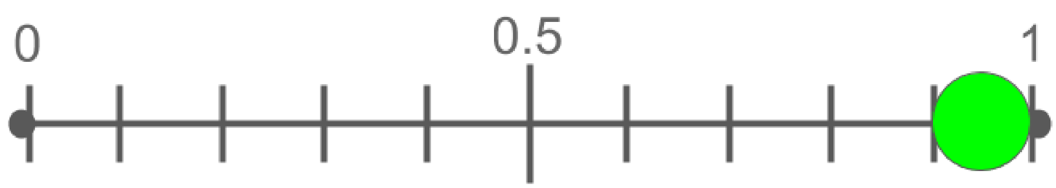


FIGURE 4.12: WipEout Omega's activity score

Encumbrance for WipEout Omega is the same as Sprint Vector's: wired HMD `(2.5)` and lightweight controller `(1)` for a total of `3.5` (Figure 4.13).

Likewise, their virtuality is the same: partially immersive virtual environment where real environment physical objects contribute or interfere with the experience (Figure 4.14).

Despite their similar scores, WipEout Omega and Sprint Vector do differ in play style.

FIGURE 4.13: WipEout Omega's encumberment score

FIGURE 4.14: WipEout Omega's virtuality score

WipEout Omega is a seated game, matching the in-game style of racing from within a vehicle, which is contrasted by Sprint Vector's room scale, nearly one-to-one virtual-to-real sprinting style. Additionally, players embody visible humanoid avatars, which differs from WipEout's disembodied cockpit inhabitant, or even the third-person view. It can be argued that Sprint Vector, being made-for-VR, got the opportunity to leverage VR-specific mechanics, such as the physicality of the controls. But going by their placement in the VAE cube alone (Figure 4.15), one would be hard-pressed to pinpoint where they differ if they knew nothing about the games. It is important to always keep that in sight while analyzing game placement within the cube.

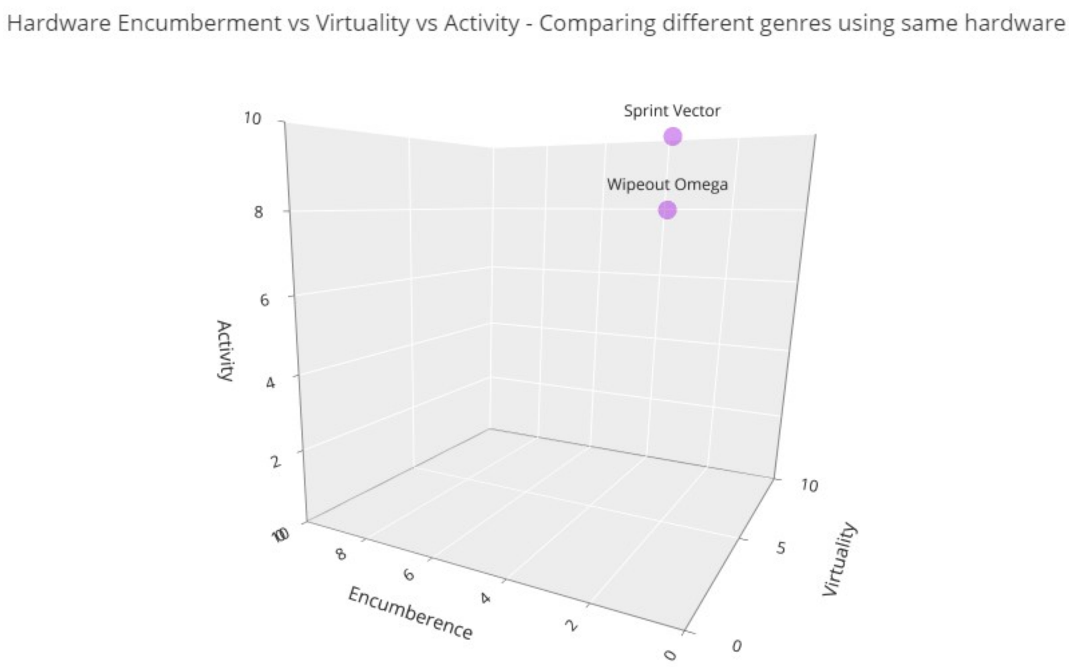


FIGURE 4.15: WipEout Omega and Sprint Vector within the VAE cube.

4.4.1 AIP Cube and subjective qualities

The VAE cube does not attempt to define presence. However, presence is too important a concept in VR to ignore it completely.

Most often with VR games, high presence and immersion are one of the most sought-after goals, but these qualities are very difficult to quantify in measurable ways (Zeltzer, 1992). Robinett (1992) qualifies presence using Zeltzer's definition but specifies that it is measurable by determining which sensory channels are employed. Jerald (2015), however, qualifies it as "an internal psychological state", for which describing it is "just as controversial [as defining consciousness or the feeling of love]". Zeltzer (1992) uses presence as a measurable axis in the AIP cube (Figure 4.16 defining it as "... a rough, lumped measure of the number and fidelity of available sensory input and output channels" (Zeltzer, 1992) and argues that together with the other axes of the cube, one can describe different levels of user engagement, where when each axis reaches its maximum potential, we can achieve ideal VR.

The other axes of the AIP cube are Interaction, whose definition was used to define the activity axis in the VAE cube, and Autonomy, which he defines as the ability for virtual actors to react to the user's input. Even though it is "[...]not clear how to quantify rigorously these components[...]" (Zeltzer, 1992), the AIP cube can still be used as a useful foil to the VAE cube by mapping experiences placed in both of the conceptual tools.

Placing games into the AIP cube would be a subjective exercise: one would have to use their best judgment to determine where to place the experiences, whether it resembles "digital Shakespeare" or whether it approaches ideal virtual reality (Zeltzer, 1992). Then, the same experience would be placed into the VAE cube, and the two can be compared. A large enough dataset in each would highlight trends, if any, between the combination of encumberment/virtuality/activity and autonomy/presence/interaction, if one cube's axis influences an axis on the other cube.

A limited subsection of the VAE dataset was used to populate the AIP cube (Figure 4.17). The omitted experiences were ones I could not justify judging for the subjective axes, since there was either no recorded footage of actual gameplay available, or I have not tried them myself. These were judged based on Zeltzer's definitions for autonomy, interaction, and presence, which were meant to be open-ended enough to include far more types of virtual software such as animation suites and films than games.

For autonomy, the general guideline used was "level of autonomy of virtual actors: if there are virtual actors, how much of their action is in response to the user's input and how much is scripted?". A loose ratio between the number of candid interactions vs scripted goals affected the score for this axis.

As for interaction, Zeltzer's definition included having access to runtime parameters of the simulation. This, surprisingly, excluded parameters such as controlling fast-forward/rewind functions in his paper, so such features were not considered when judging some of the general entertainment experiences. The user's ability to manipulate the environment through interaction and their own freedom to act within it were considered when judging this axis. For example, an experience which lets you open all of the drawers and manipulate all of the tools it presents will have a higher score than an experience which restricts users to only interacting with plot-relevant elements. This, interestingly, varies slightly from the way interaction was measured for the VAE cube's activity axis, as the ability to fast-forward/rewind counted as manipulation and affected the score.

Finally, for judging presence, Zeltzer's definition stated that this should be measured by lumping together the total range of sensations the system can reproduce for the user, as well by considering the mapping between the user's actions and their representation in the virtual world. For example, if a user can freely move 1-to-1 in a virtual environment and their physical actions are faithfully replicated in the virtual environment, this should score higher than if a free-moving user was interfacing with the system more abstractly, say by using a joystick. This measure was used to judge presence.

The resulting mapping into physical space yields the spread seen in Figure 4.18.

With this, it will be possible to compare statistical analyses between the objective VAE cube and the subjected AIP cube, thus identifying if certain VAE trends lead to higher presence. For example, returning to our racing games, Sprint Vector and WipEout Omega, it is interesting to note that they do not feature in the same clusters within the AIP cube, as they do within the VAE cube, despite their very similar scores here too. The only hugely varying statistic between the two is the autonomy WipEout features: this game has an AI mode, which lets players race against semi-autonomous AI racers, but Sprint Vector does not feature such a play mode. Due to this, WipEout Omega inches closer to the center of the AIP cube, whereas Sprint Vector shares its space with other such games which do not feature extensive autonomous virtual actors for the player to interact with.

4.5 Summary

The VAE cube was created on the foundation of taxonomies by Kjeldskov (2001) and Milgram and Kishino (1994), with influence from Zeltzer (1992). Quantitative scales were created for each of the axes, and the conceptual space created. Data was selected to populate the VAE cube based on three contexts: at-home experiences, VR arcade experiences, and on-the-street experiences. This was done to ensure that a wide variety of styles of use were accounted for. Then, prominent experiences from each of these categories were selected and sorted into genre categories. The data was then mapped into both the VAE and AIP cubes, in order to study the data clusters and compare subject analysis to objective analysis. In-depth, statistical analysis will be explained in detail in the next chapter, where the cubes' trends will be compared, and the VAE's axes will be investigated to identify relationships.

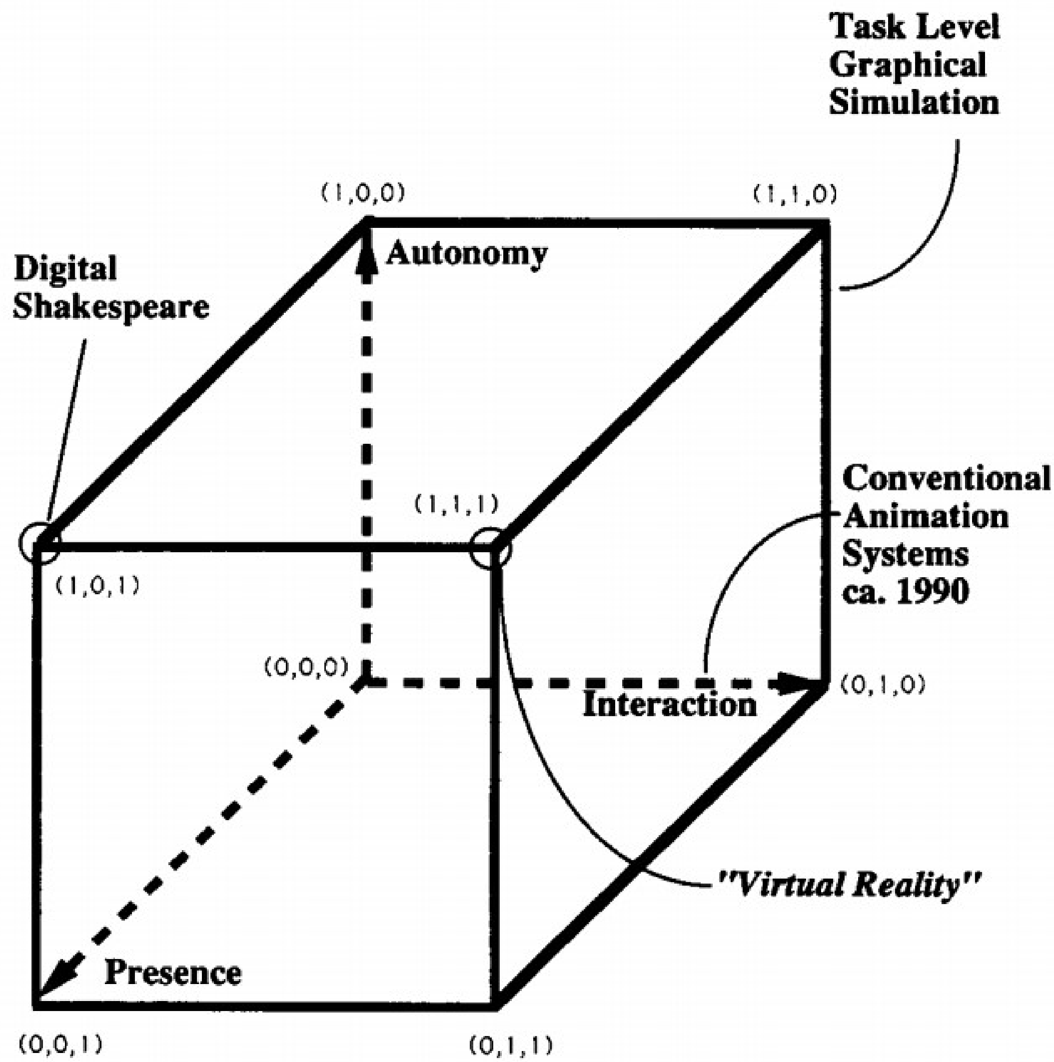


FIGURE 4.16: Zeltzer (1992) AIP cube showing Presence, Autonomy and Interaction axes.

"Autonomy, Interaction, and Presence" by David Zeltzer, in Presence: Virtual and Augmented Reality, Volume 1, No. 1, Winter 1992, ©1992 MIT.

ID	Title	xAutonomy	yInteraction	zPresence	Genre	Port?
1	Super Hot	6	6	6	action	☑
2	Eve Valkyrie	5	6	6	action	☑
3	Gorn	6	7	6	action	☐
4	Skyrim	5	7	6	fantasy	☑
5	Moss	2	5	6	fantasy	☐
6	Netflix VR	0	0	5	General Entertainment	☑
7	Youtube VR	0	0	6	General Entertainment	☑
8	Birdly	0	1	10	General Entertainment	☐
9	Pokemon Go	0	1	1	location-based	☐
10	Tetris Effect	0	6	5	puzzle	☑
11	Keep Talking and nobody explodes	0	8	6	puzzle	☐
12	Wipeout Omega	5	6	6	racing	☑
13	Sprint Vector	0	5	6	racing	☐
14	Pro Race VR	0	8	8	racing	☐
15	Beat Saber	0	5	6	rhythm	☐
16	Electronauts	0	8	6	rhythm	☐
17	Thumper	0	4	5	rhythm	☐
18	Scary Girl	1	6	8	sci-fi	☐
19	Surgeon Simulator in Zero G	0	6	6	sci-fi	☐
20	Lone Echo	1	6	6	sci-fi	☐
21	Job Simulator	5	8	6	slice-of-life	☐
22	Werewolves Within	0	1	6	social/board game	☐

FIGURE 4.17: Dataset used to populate the AIP cube.

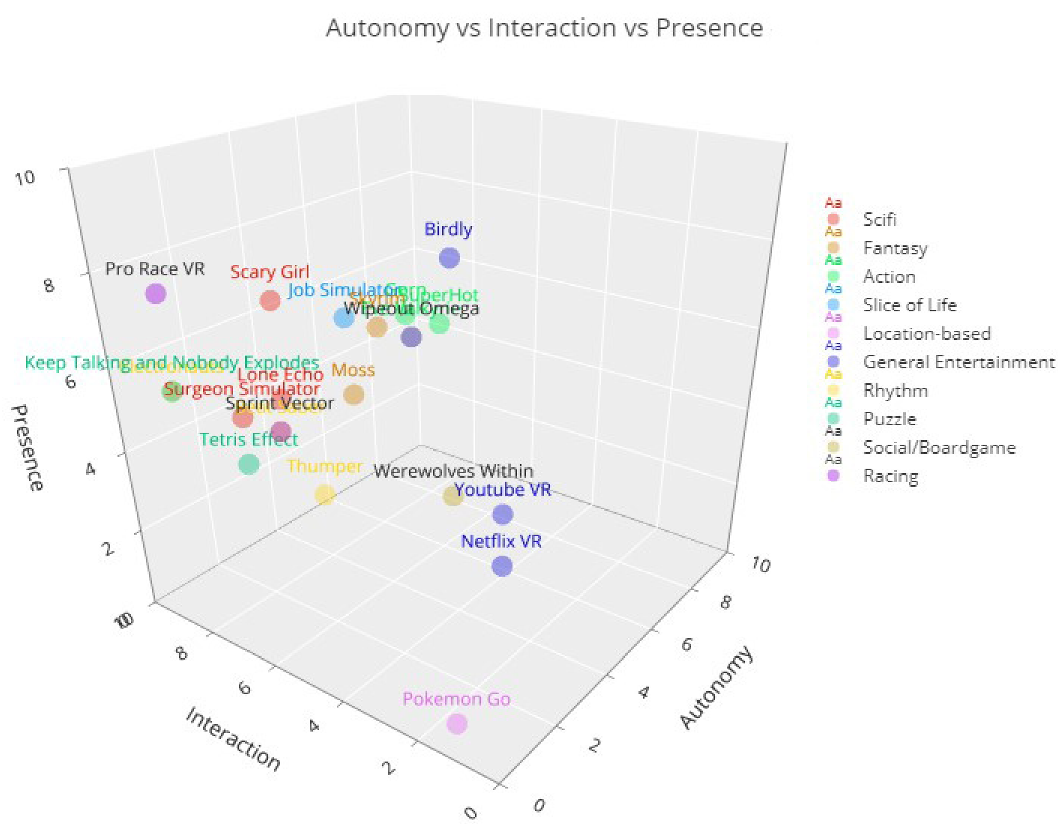
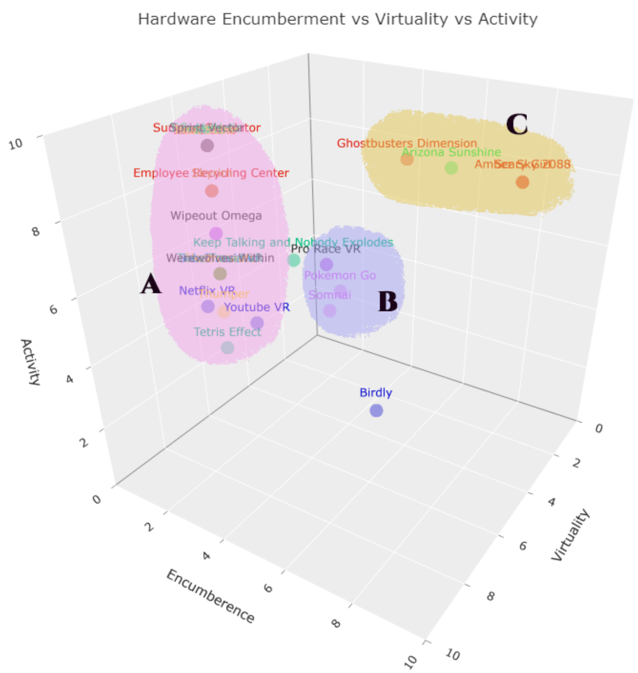
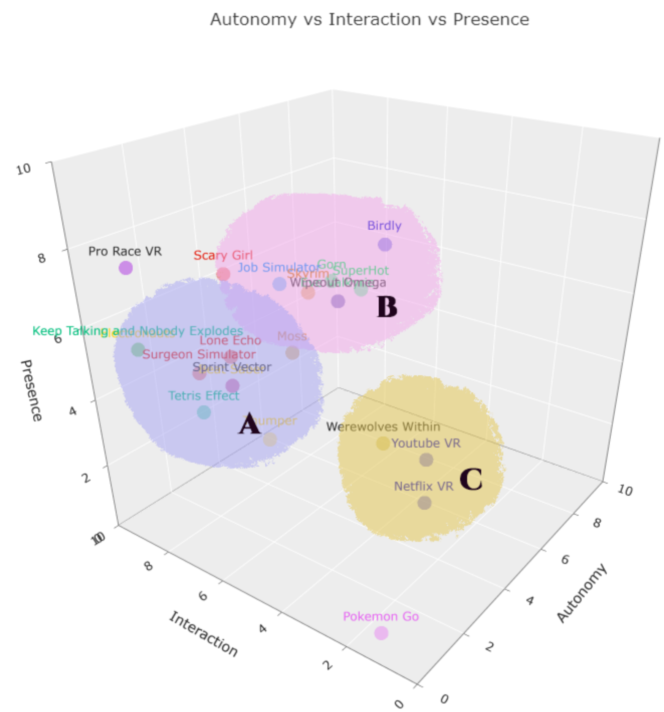


FIGURE 4.18: Populated AIP cube. The spread of data is much less concentrated to a single column like the cluster in the VAE cube. Instead, it spreads out into about three distinct clusters.



(A) Populated VAE cube. Three cluster groups are highlighted: a) represents low-encumbrance, high-virtuality and a variety of activity levels. It is by far the biggest. b) central experiences with mid-to-low-levels for all of the factors. These are almost entirely made of AR experiences. c) mid-virtuality, high encumbrance and high activity. These represent a large portion of VR arcade experiences.



(B) Populated AIP cube. Three cluster groups are highlighted: a) represents mid-to-high interaction, mid-presence, and low-autonomy. This cluster features the puzzle genre, some racing experiences, and many of the scifi experiences. b) central experiences with mid-to-low-levels of autonomy, but mid-to-high for all of the factors. Action, slice of life and fantasy exist in this one, along with some racing. c) this last cluster features all the lower scoring experiences, and includes many simulations.

FIGURE 4.19: VAE and AIP cubes with their respective clusters highlighted.

Chapter 5

Analyzing VAE Relationships

This chapter covers some example analyses which can yield helpful results when the dataset is representative of the population. In this thesis, time constraints limited the dataset to a small number, and to a subset of experiences which were most visible, making it an unreliable dataset. However, it will be used as an example to show tentative results regarding the following approaches:

- linear regression to determine the coefficient of determination between the axes
- point distance calculation between points in the space to determine:
  - the distance between all individual points
  - the average distance between points within the same group (e.g. genre)
- histograms to determine distance frequency
- standard deviation

The primary purpose of this analysis was to determine the relationship between the axes, if any. Its subjective sister, the AIP cube by Zeltzer (1992), has independent axes, but the VAE cube was designed with the assumption that hardware encumbrance was affected both by virtuality and by activity: that the more virtual an experience, the more encumbered the user would have to be—until the ideal direct neural connection of true VR is achieved in which case, hardware encumbrance is irrelevant—or that the more activity was required, the more hardware would be needed to interact with the virtual world. This assumption was based on Milgram and Kjeldskov's own discussions about how different hardware is better suited to different levels of virtuality/interaction.

Genre was selected as the comparative variable for these tests due to genre being a consistent quality, but other comparative variables such as "made-for-VR games vs ported games" or even "games in arcade setting vs games in-home setting", as mentioned in the previous chapter, could yield interesting, if not useful, results.

5.1 Linear Regression

The first approach used for analyzing the dataset was linear regression, typically used for modelling relationships between dependent and independent variables, thus determining whether the dependent variable is predictable based on the value of the independent variable. In the first two cases (Figures 5.1 and 5.3), encumbrance was selected as the dependent variable in order to determine if virtuality and activity might affect the level of encumbrance. The last case (Figure 5.4) graphs the relationship between activity as the dependant and virtuality as the independent variable, to determine if virtuality has an influence on the level of activity for experiences.

Each of the combinations of axes were put through Python's Scipy `linregress()` function to produce the `slope`, `intercept` and `r value`. Then, the result was plotted and the coefficient of determination calculated. For example, here is the code used to calculate the linear regression on encumbrance and virtuality. For the full script used to calculate these statistics, see Appendix A.

```
#obtain the slope, intercept and r value
EVslope, EVintercept,
    EVr_value, EVp_value,
    EVstd_err = stats.linregress(xVirtuality, yEncumbrence)
```

```
#plot original virtuality and encumbrance data
plt.plot(xVirtuality, yEncumbrence, 'o', label='original␣data')

#plot fitted line using calculated intercept and slope
plt.plot(xVirtuality,
    EVintercept + EVslope*xVirtuality, 'r', label='fitted␣line')

#print coefficient of determination using calculated r value
print("coefficient␣of␣determination:␣%f" % EVr_value**2)
```

Each of the regressions were calculated both on the full dataset, as well so on a few intra-genre example, to determine if there was a remarkable difference in the way the axes relate for different genres.

5.1.1 Assumptions

The current state of the cube uses fairly discrete scales. In the case of activity and encumberment, values are determined based on a point system, as explained in the previous chapter, thus, not all points on the scale are accessible. Additionally, it is extremely difficult to quantify activity in a way that lends itself well to linear analysis: what metric can be used to determine that walking as an activity is `X` times more difficult or involved than using a joystick, and where are all the jumps where this needs to be determined? This speaks to the lack of linearity of the scale as it stands, and applies as well to encumbrance and virtuality. Encumbrance, for the same reason, and virtuality, though it lacks the point system, is still not explicitly described as a linear function.

The scales need to be adjusted in order to ensure that, at the very least, the discrete values can be described as exponential, and the dataset enlarged to compensate for the discrete scales. However, the cube is still not a fully linear space and linear regression is used. This thesis only explores a first attempt at analyzing correlation and selects the most common form of regression for simplicity sake. Other forms of regression, however, should be investigated, though this thesis assumes that, should the scales be converted to linearity, linear regression would be a valid analysis for determining relationships between the axes as well.

5.1.2 Full dataset

Encumbrance as a function of virtuality

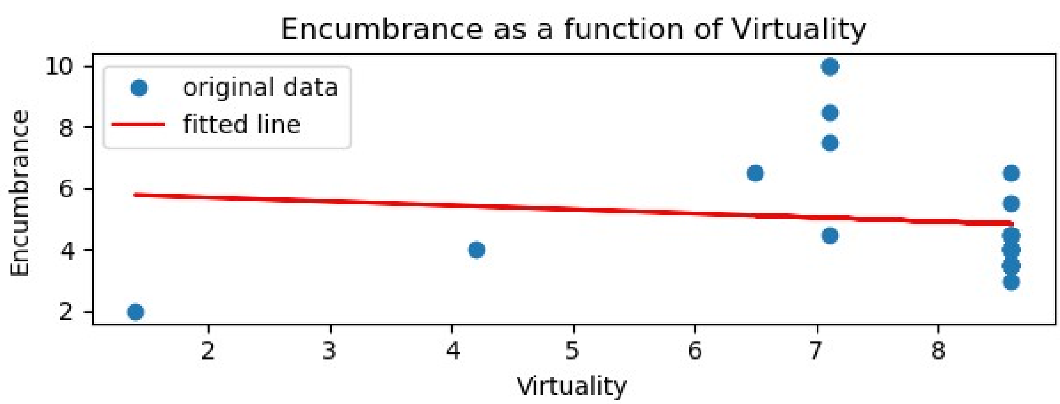


FIGURE 5.1: Graph showing the relationship between encumbrance and virtuality

The resulting coefficient of determination (CoD) for encumberment as a function of virtuality proved to be statistically insignificant, at a mere `0.013439`. Contrary to what was previously assumed, there doesn't appear to be any relationship between encumbrance and virtuality where virtuality might influence the encumbrance value. However, interestingly, this was the stat that changed the most as more games were added to the dataset, particularly when VR arcade games—within this context, only including games which are exclusively available at VR arcades—were added. Prior to the inclusion of these games in the dataset, which account for one-third of the total number of games, the CoD between encumberment and virtuality was as high as `0.391630`, which is still a mere 30%, but significantly higher than the current CoD. A closer inspection of the chart shows that the noticeably more encumbered games are strictly VR arcade games and that they also tend to be slightly less virtual than the at-home VR games (Figure 5.2).

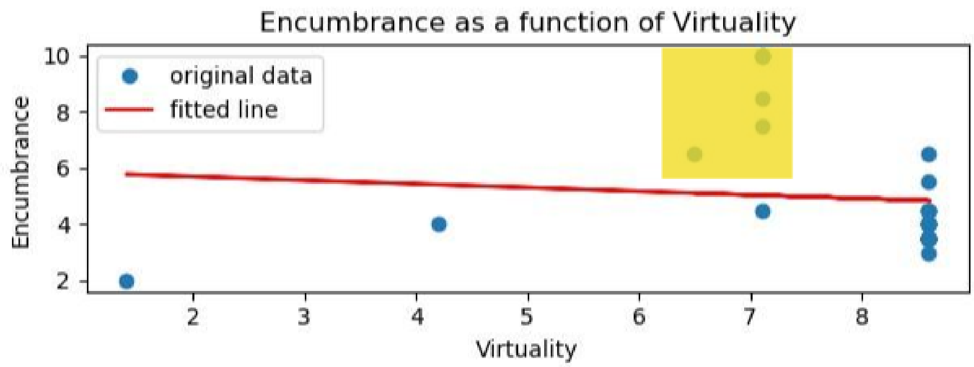


FIGURE 5.2: The highlighted area shows noticeably more encumbered experiences.

This could be caused by a combination of factors including:

- VR arcades are more likely to have the full equipment setups with backpack computers and haptic exoskeletons than at-home game setups
- VR arcades are more likely to have "face-to-face" social interactions between players
- VR arcades are more likely to have physical props with which players can interact

the first of which would increase their encumbrance score, but the last two of which would bring their virtuality score down somewhat.

Encumbrance as a function of activity

Although encumbrance and activity have a slightly higher CoD at `0.082910`, less than 10% is still not significant enough to determine that activity has a strong influence on encumbrance. This was expected: level of activity will depend on the context (genre, purpose, target audience, etc.) of the experience, and the amount of hardware required to perform tasks in all of these contexts can vary greatly. For example, only a joystick

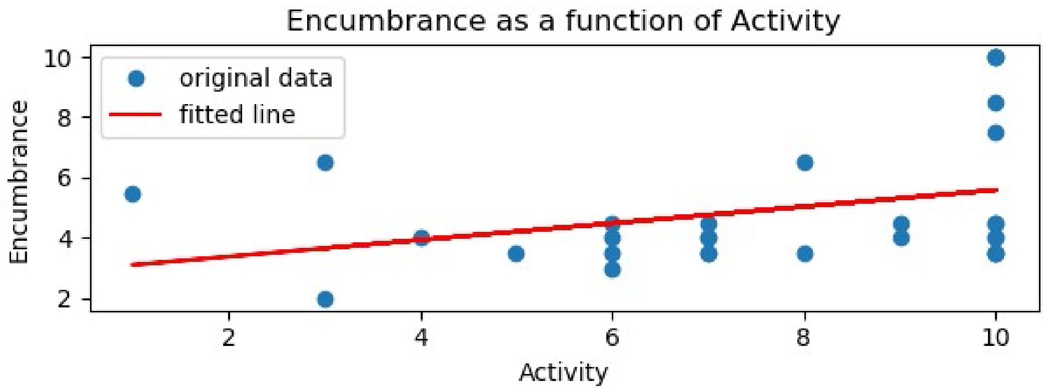


FIGURE 5.3: Graph showing the relationship between encumbrance and activity

is needed to be able to perform all of the interaction categories of the scale, but likewise, a full-body exoskeleton can be used to provide haptic feedback for a roller coaster experience which only supports orientation as an interaction.

Activity as a function of virtuality

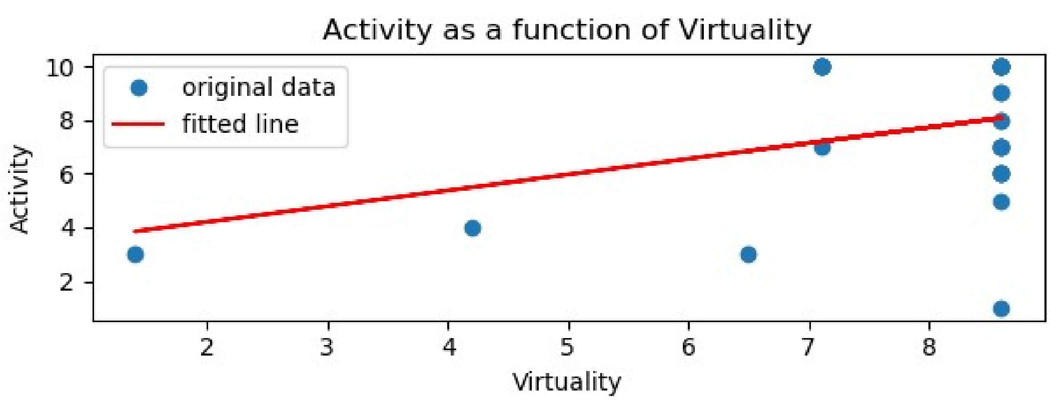


FIGURE 5.4: Activity as a function of virtuality

Similarly to the previous function, activity as a function of virtuality shows a low CoD at `0.130110`. However, this is the highest CoD between all the compared axes at 13%. In the selected samples, the lowest virtuality and activity combo is represented by Pokémon Go, a location-based game which lets players catch creatures using an optional AR portal, but it is the only AR game of its type in the dataset. The next lowest is Somnai, an interactive theater experience, where the player is guided in a real environment over which virtual images are projected. These two experiences alone are the sole occupants of the lower half of the virtuality scale. There is a strong chance the CoD would change drastically if more AR experiences had been included.

5.1.3 Intra-genre

An intra-genre analysis is a bit more difficult with such bare-bones datasets, but these are principally being presented as examples of possible use cases for the VAE cube. Figures 5.5a and 5.5b show the fitted line graphs for the *scifi* and *general entertainment* plots respectively.

Encumbrance as a function of virtuality

For both of these examples, the CoD was significantly higher when mapping encumbrance as a function of virtuality on a per-genre basis than when calculating for the dataset as a whole. The CoD for both *scifi* and *general entertainment* were nearly perfect at `0.92` and `0.98`, respectively. This would indicate that it is fairly safe to say that the higher the virtuality is for these experiences, the lower the encumbrance will be. With the *scifi* dataset, however, it is important to note that each of the high-virtuality/low-encumbrance experiences are from the at-home category, whereas the experiences on the lower end of the virtuality scale are VR-arcade experiences, and the effects of this distinction have already been discussed.

Similarly, the case of *general entertainment* experiences, where a decidedly broad range of possibilities exist, can be further broken down into types. The two low-encumberment experiences at the high end of the virtuality scale are home-theatre types: Netflix VR and Youtube VR. This type of experience requires very little hardware to successfully interface with the virtual world, and indeed, both only make use of an HMD and a lightweight remote. However, the highly encumbered Birdly experience at the low end of the virtuality scale features an entire exoskeleton to allow the user to flap their arms to fly and receive haptic feedback.

The relationship described by this CoD appears to be one which determines an experience's subcategory within a genre.

Encumbrance as a function of activity

For *scifi*, there is a higher CoD than the one calculated for the entire dataset, but it is still fairly low at `0.2`. The reason for this is the same as mentioned above: various levels of activity can be accomplished with varying combinations of hardware. As if to exemplify this, all of the experiences in *scifi* range very highly on the activity scale, with most sitting at `10` and one sitting at `9`, but the hardware encumberment ranges from mild at `3.5` to very high at `10`—keeping in mind that the encumberment scale is clamped down, so many of these experiences have even higher encumberment levels that can be represented by the graph. That being said, the relationship here, again, seems to represent a split in subcategory: highly encumbered games belong to the arcade category, whereas the less encumbered ones belong in the at-home category.

General entertainment, again, features a nearly perfect CoD at `0.98`. The outlier, in this case, is the same: Birdly, which features higher encumberment for lower activity.

Activity as a function of virtuality

Finally, the CoD of activity as a function of virtuality for *scifi* is once again higher than the CoD of the general calculation, but not terribly strong at `0.22`, though the range in question is extremely small. Almost all experiences have a high virtuality and a high activity level, varying only significantly in virtuality. This variation, however, once again reflects the subcategory distinction.

And again, *general entertainement* shows a high, in fact perfect, CoD at `1`, mirroring the Birdly/home-theatre distinction.

5.2 Point distance

Point distance analysis was performed in order to determine if games within the same genre shared more similarities amongst themselves than to other points in the space. The assumption was that there would be, but that there would be variety in how close average distance is for different genres. Next,

The first step was to calculate the intra-genre distances, or the distance between points within the same genre. This was done using Python's Scipy spatial distance function. Each genre's average was stored into a new variable for later use, and plotted into a histogram to visualize the distribution of the resulting values. The following example shows the code used to calculate the distance matrix for the *action* genre. The results are presented in Figure 5.6.

```
#Action point distance
distanceMatrixAc     = distance.pdist(VAEAction)
#Average point distance
averageAc            = np.mean(distanceMatrixAc)
#Plot histogram
plt.subplot(521)
plt.hist(distanceMatrixAc, bins='auto')
plt.title("Action games distance")
plt.ylabel("Frequency")
plt.xlabel("Bins")
```

Next, all the genres' averages were stored in the `genreMatrix` variable to be plotted into its own histogram. Since the *slice of life* and *social* genres only had one example per, it was impossible to calculate point distance, thus, they had to be commented out. In the future when there is more data, these will be re-included in the calculations. This histogram is presented in Figure 5.7a.

```
#Collecting the average distance for all the genres
genreMatrix = np.array([averageAc, averageFa,
    averageGe, averageLo, averagePu,
    averageRa, averageRh, averageSc]) #, averageSl, averageSo])

plt.subplot(529)
plt.hist(genreMatrix, bins='auto')
plt.title("Intra-genre␣Average␣Distance")

plt.ylabel("Frequency")
plt.xlabel("Bins")
```

Finally, the distance between all points was calculated and plotted into its own histogram (Figure 5.7b) to compare with the average intra-genre point distance.

```
#Calculating point distance for all experiences
distanceVAE     = distance.pdist(VAEMatrix)

plt.subplot(5,2,10)
plt.hist(distanceVAE, bins='auto')
plt.title("All␣Points␣Distances")

plt.ylabel("Frequency")
plt.xlabel("Bins")
```

It is clear by comparing the *Intra-genre average distance* and *All points distances* histograms that there is a lot more variety between point distance outside of genre relation. The standard deviation of these datasets supports this, with intra-genre distance showing only `0.6` units of deviation and general point distance showing nearly `3`, at `2.7`.

However, Figure 5.7b shows that the majority of the distances between all points exists within the same range as the average intra-genre distances shown in Figure 5.7a, that is to say within 2 and 4 units. The entire range of distances is about triple this, presenting values between 0 and 12 [1]. Compared to the max spread of values, the 2 unit range of genres is comparatively small enough to be significant.

From this, it is reasonable to conclude that although there is more commonality between experiences within the same genres—this even despite the different play styles and context explored in Chapter 4—several of these genres exist very nearby one-another, enough that it is possible the low distance between genres is simply a result of most points being clustered. The outlying experiences, which are contributing to the larger range of the *All points distances* graph, are the two AR experiences which situate themselves apart from the main cluster (Figure 5.8), and the Birdly experience.

Still, the cluster is large enough that the much closer intra-genre points could very well indicate a trend.

5.3 AIP Analysis

Analysis for the AIP cube should be taken lightly, due to the subjective nature of the scales. However, comparing the results with the VAE analysis could yield some interesting information about intra-genre tendencies with regards to the ever elusive presence factor. In this case, linear regression was not calculated for the AIP cube as the axes were designed to be independent of one another. For the full code used to calculate this, refer to Appendix B.

[1] It is important to note, however, that *All points distances* calculation does include two extra experiences, which cannot feature in the *Intra-genre average distance* calculation.

5.3.1 Point distance

Point distance calculation was performed on the AIP cube in order to obtain information about intra-genre similarities compared to overall similarities between experiences, the same as for the VAE cube. Likewise, the results were plotted into histograms (Figure 5.9 for the breakdown of intra-genre distances, Figure 5.10a for intra-genre average distance, and Figure 5.10b for the distance between all points) and the standard deviation for both intra-genre distance and all points distance. Since an experience was omitted for the *location-based* genre, leaving only a single one behind, this genre's graph was also omitted from the point-distance calculation.

The first noticeable difference is the wider spread of distance values which the AIP intra-distance graph introduces. Specifically, as discussed in Figure 4.19, the AIP cube features a generally tighter spread, despite its three distinct clusters. These clusters, also discussed in the aforementioned figure, contain fairly clean separations of genre, with scifi, puzzle, rhythm and most of racing in mid-presence, high-interaction, low-autonomy, action and fantasy in high-presence, high-interaction, mid-autonomy, with the rest scoring low in all three axes. Additionally, autonomy experienced the most radical jumps: experiences which feature AI had scores in autonomy at all, whereas the rest of the experiences did not or had very little to show, and this might have contributed to the wider range of values. If within a genre, most experiences didn't feature AI, but a single one did, this experience would skew the distance out further.

Opposite to this, there was a smaller, and slightly more even spread of values in the all points distance calculation for the AIP cube than for the VAE cube. Again, this was caused by the more even spread of data in general. Here, the jumps caused by autonomy don't register as clearly. In the VAE cube's graph, it's clear that most experiences fall within the "fairly close" range, as discussed above, but in the AIP graph, nearly half of results are in the mid-range distance, indicating that most points are equally far from one another. This is also perhaps caused by the fact that the dataset features exclusively entertainment-based experiences, which are generally meant to be high-presence and interactive. There are some exceptions, with the home-theatre style experiences having less interaction, and the AR experience not really requiring high interaction or presence for its context.

It would seem, based on this analysis, that though genre seems to have some influence on the placement of experiences, autonomy being such a powerful factor implies that the real dividing factor is whether these experiences feature AI or not. Additionally, the even spread points to very similar scores for all experiences, despite their variety. It is possible that, like the VAE cube's limitations in describing activity, the AIP cube's definitions of presence and interaction don't reflect the reality of the situation. Further experimentation would be required to confirm this.

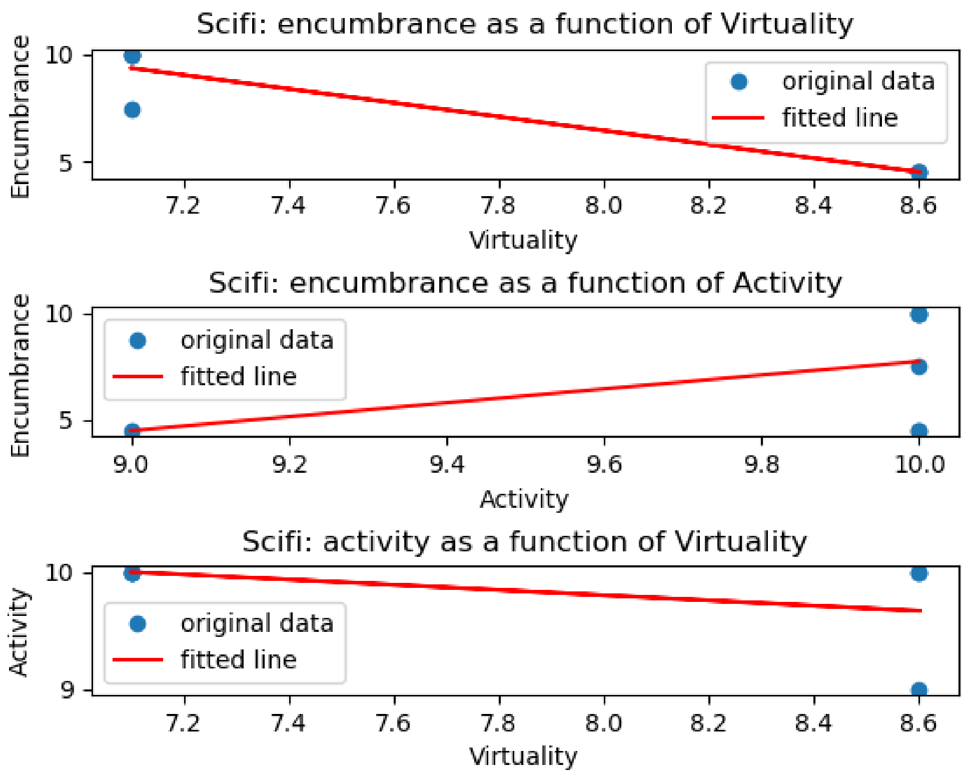
5.4 Summary

Given the limited dataset and tentative results it has yielded in this analysis, a few conclusions can be made about the use of the cube.

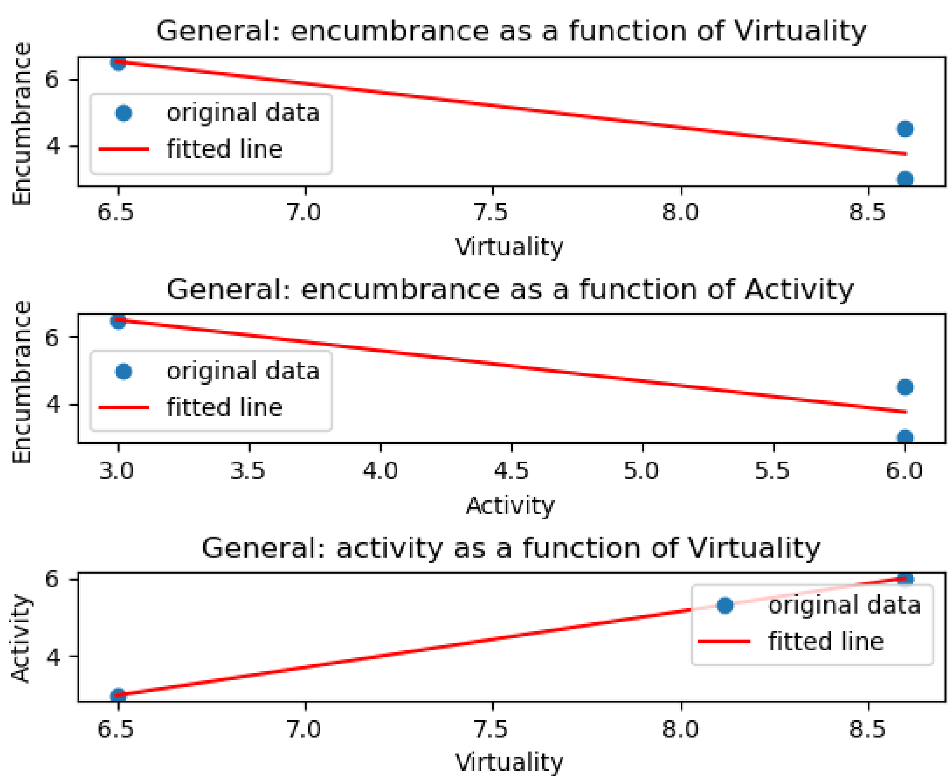
1. Data inserted into the cube can be used to calculate relationship information
    (a) There is less of a relationship between the axes than originally expected.
2. Dimensions added to the cube can be used to compare different parameters and determine if there are relationships between these new axes and the existing ones, or if there are trends among the groups
    (a) Intra-genre experiences share enough similarities between each other in comparison to their non-group neighbours to be considered significant
3. These kinds of results can be used to
    (a) describe experiences by their relationship to others
    (b) stay informed about trends in comparison to others

   (c) make informed design decisions (e.g. being aware that most experiences within a genre share x and y commonalities and deciding to subvert them or play within them)

4. A larger dataset would be needed to obtain more conclusive data about trends in VR at large.



(A) Scifi linear regressions. This dataset contains seven experiences.



(B) General entertainment linear regressions. This dataset contains three experiences.

FIGURE 5.5: Linear regression preformed on example genres.

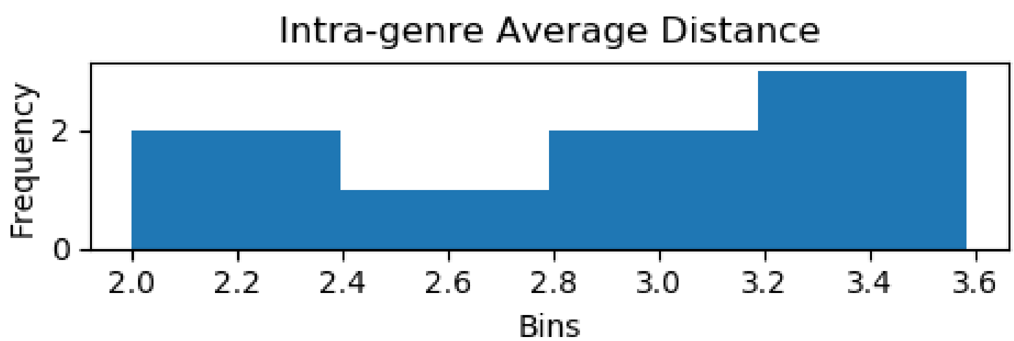
(A) Left: action games; Right: general entertainment

(B) Left: Fantasy games; Right: location-based games

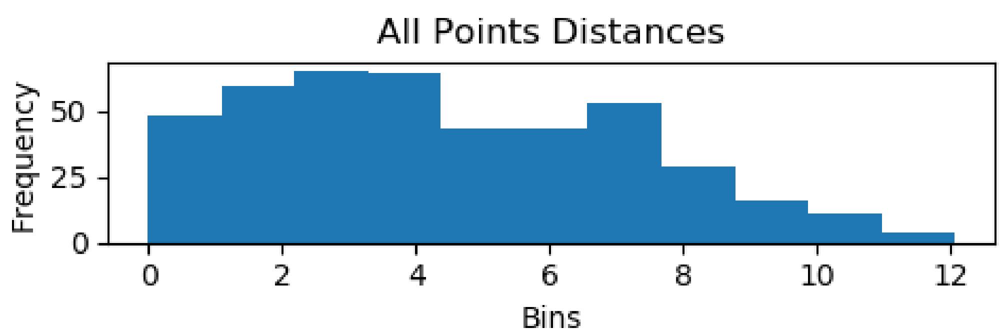
(C) Left: puzzle games; Right: racing games

(D) Left: rhythm games; Right: scifi games

FIGURE 5.6: Histograms for each of the genre's average point distances.



(A) Histogram representing the distribution of average point distance within genre-groups.



(B) Histogram representing the distribution of distances between all points

FIGURE 5.7: Histograms comparing the average distance between points within the same genre-group, and the distance between all points in the VAE cube.

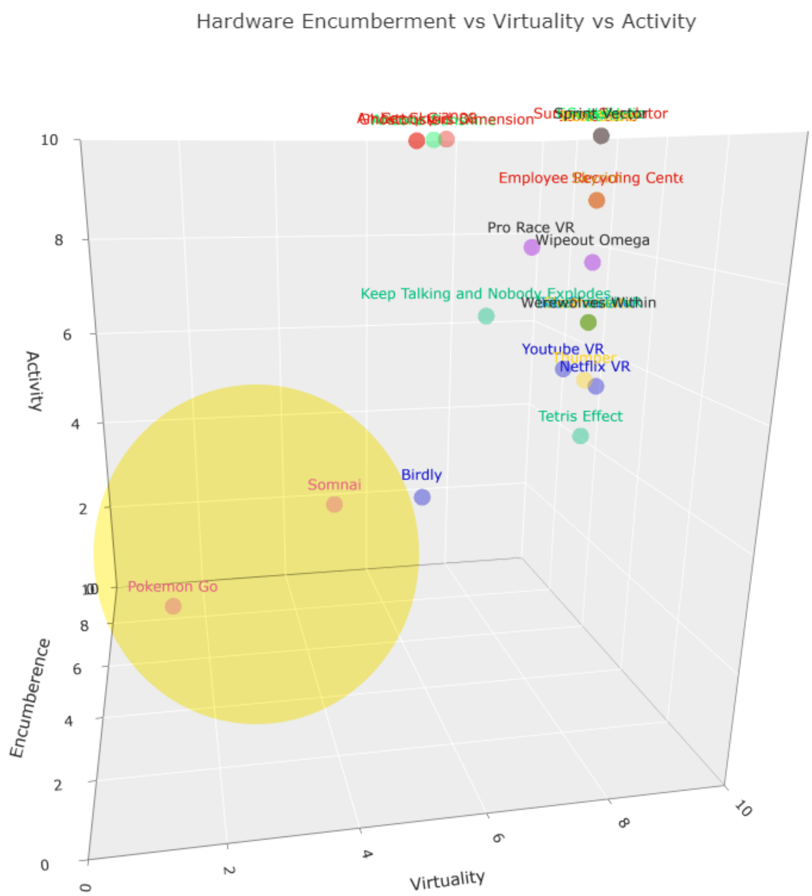
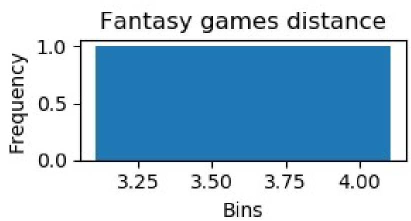


FIGURE 5.8: AR experiences far apart from general cluster of experiences. These two experiences contribute to the larger range of the *All points distances* graph.

(A) Left: action games; Right: general entertainment

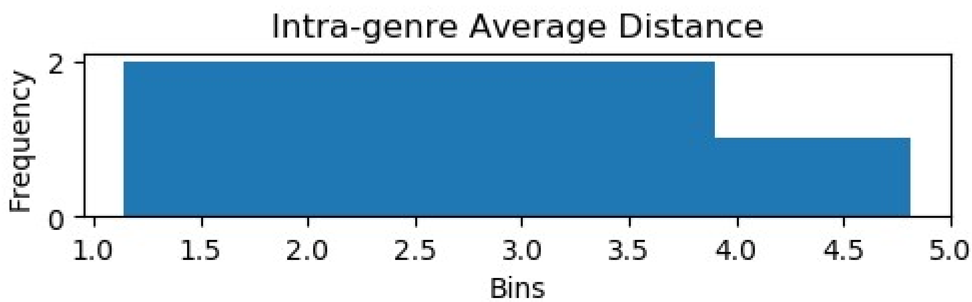


(B) Left: Fantasy games

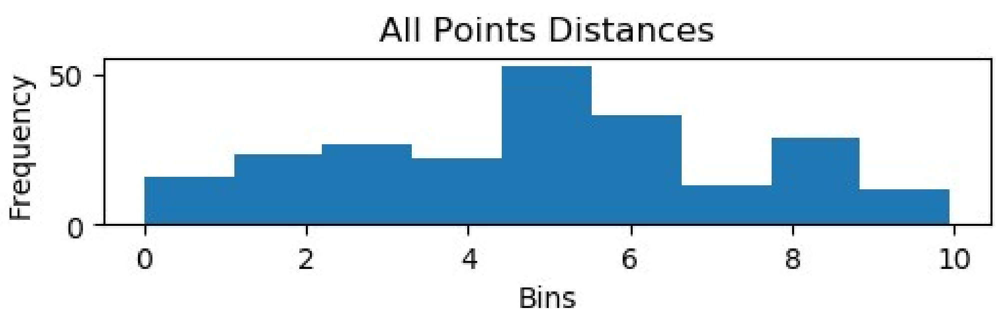
(C) Left: puzzle games; Right: racing games

(D) Left: rhythm games; Right: scifi games

FIGURE 5.9: Histograms for each of the genre's average point distances while mapped into the AIP cube.



(A) Histogram representing the distribution of average point distance within genre-groups mapped in the AIP cube.



(B) Histogram representing the distribution of distances between all points mapped into the AIP cube.

FIGURE 5.10: Histograms comparing the average distance between points within the same genre-group, and the distance between all points in the AIP cube.

Chapter 6

Discussion

In this chapter, various challenges encountered and observations made are discussed in detail. Design practices and taxonomies are discussed within the context of the need for standards, where the cube's motivations are contextualized, and the frameworks used to build it are reiterated. Next, the effects which the Gartner Hype Cycle and the fetishization of technology had on this thesis's research are examined, and an anecdote featuring possible consequences of not carefully navigating these factors is shared. Then, the issues discussed in Chapter 4 about the selection of data are reiterated and the consequences explored. Following this, the data analysis seen in Chapter 5 is criticized to highlight its limitations, and the possible workarounds and consequences these have. Finally, the VAE cube is re-contextualized within the literature and validated through a process of comparison with related works.

6.1 The need for standards

6.1.1 Design practice

Standards in design practices have existed for the design of flat screen interfaces as we know them for decades. For example, the desktop metaphor became a longstanding tool for designing operating systems and the software for them (Reimer, 2005). These standards have helped both designers and users; designers refer to them as a base for making informed decisions about their new interfaces, and users can expect to learn new interfaces based on their existing knowledge of them. Ultimately, this means users will be more likely to like the software, rather than become frustrated with it and abandon it.

But VR has only recently entered its refinement phase. Is standardization of VR design practices premature at this stage? Malaika (2015) says it isn't. Having so many different platforms available as target hardware makes VR design even more challenging than the novel interaction possibilities already do, but having, for example, congruent interaction methods—the dual hand-held motion-tracked controllers which many of these platforms offer—makes it easier to make "cleaner design trade-offs [...] sooner in development". Additionally, Malaika (2015) reiterates that standards are cropping up organically too, in the same way that the flat screen desktop interface has evolved through the propagation of design ideas when they are proven to work (Reimer, 2005).

In fact, such organically created standards are already being established: Owlchemy Labs have coined the term "Tomato Presence", an interaction design method to help mitigate the uncanny feeling of seeing virtual hands holding objects, accomplished by simply replacing the hands with the object they are meant to be holding (Fox and Schubert, 2018; Malaika, 2015). More explicitly, Oculus has released a document outlining some best practices for user experience, locomotion, user input, positional tracking, and more (*Introduction to Best Practices*).

Whether intentionally created or not, standards for interaction design in VR are cropping up. There are even effort being made to amalgamate, test, and confirm these design practices, in order to make it easier to spread the knowledge (Jerald, 2015; Malaika, 2015).

6.1.2 Taxonomies

Less flashy than visible design results, but perhaps as important to design standards, are taxonomies: the standardization of the way we can describe experiences in order to make it easier to talk about them. Many taxonomies exist for VR, some about interaction itself (Kjeldskov, 2001; Zeltzer, 1992), others about core concepts of VR (Milgram and Kishino, 1994; Robinett, 1992). Many of these taxonomies address hardware as an important factor of VR and discuss it in relation to the concepts of virtuality, interaction, immersion, etc. However, the sheer amount of hardware a user can end up wearing for the sake of having a higher degree of interaction, virtuality, what-have-you was rarely every addressed in these discussions. This prompted the thesis question.

6.2 Hype cycle and fetishism of technology

The effects of the Gartner Hype Cycle have contributed both to the slow adoption rate of VR by consumers as well as to the fetishism of technology, which presented challenges to this thesis.

The ’90s with the many failed commercial VR systems (Snow, 2007; Jerald, 2015), and the subsequent "VR winter" (Jerald, 2015) marked the phases of inflated expectations and accompanying fall in consumer interest. Now that consumer interest has returned thanks to the new wave of commercially available hardware, the hype is starting to rise again, as the price of hardware stabilizes, and quality increases. However, this means that average consumers of all kinds are gaining access to VR for the first time. The aforementioned excellent experiences which are available make a profound impact on those who try them, often being so remarkable, that describing the experiences proves to be difficult (Boland, 2017; Jerald, 2016). Consequently, excitement over the novelty often ends up overpowering useful critique (Jerald, 2016). This is something I have observed not only during my attempts at finding valuable ratings for VR experiences—which often laud the amazing new breed of immersion but either entirely fail, or simply gloss over discussing relevant design questions—but also from working within the industry, as well.

This makes it difficult to determine with any certainty which factors contribute to a user’s enjoyment of the experience, and which might hinder it. A personal anecdote:

while working on a new VR experience who's target demographic was children, adult developers were focused on developing a free-roam setup we had, and this elated our adult playtesters. How wonderful to be able to roam an entire warehouse untethered! But what no one questioned was that to make this free-roam experience possible, users had to wear an almost `6kg` backpack computer, on top of the `2kg` haptic feedback vest, not to mention the large two-handed gun prop, and the VR HMD itself. Not many test users complained about this, so it's safe to assume this intense encumberment didn't affect the experience, but the excitement about the new technology distracted from the question at hand: would this be too much weight for a child? This game will be released to their intended audience soon, and only time will tell if this oversight will affect consumer opinion of VR, but it is a clear example of how fetishism of the technology itself can cause one to overlook avoidable, underlying design problems.

It is perhaps possible for the VAE cube to make these kinds of underlying design problems more clear by visualizing the descriptions of experiences along not only the established dimensions but also through comparison by adding additional dimensions.

6.3 Selecting data

Data selection for this thesis was a challenge for a variety of reasons. As previously discussed, it was originally the intention to determine if the VAE cube showed any common trends among experiences which are considered "good" vs those that are considered "bad". During the selection process, however, it quickly became clear that there was not a consistent method for determining if experiences were good or not that could be applied to the wide range of contexts.

At-home experiences which featured on the Steam Store, a popular database for games, could be given a five-star rating, but it was difficult to find games with enough ratings to qualify which weren't the most prominent examples. Games which already enjoy success receive more attention, more downloads, and thus, more ratings. But games which are not to people's liking often fade into obscurity. It's impossible to play all of the games in the library—there are over two thousand results for a "VR Only" search—in order to determine myself if they are bad or not, and even then this would be an extremely subjective rating. Beyond this, games which belonged to the VR arcade and on-the-street contexts lacked a central database at all, making it even more difficult to obtain any kind of objective measure on their quality. The best option which existed was to scour the internet for reviews, but those were biased at best, partly for the fetishism described above, and partly for the issue of paid reviews.

Scrapping this intention, then, became necessary, and the goal of analyzing the data in the VAE cube was shifted from being able to show trends prominent within good examples vs bad examples to focusing more on the comparative analyses of other added dimensions.

Additionally, a critical underestimation of the time it would take to select examples and map them into the cube lead to having a much smaller dataset than intended. At the time of planning, it was assumed that experiences would, at the very least, list their hardware requirements and that basic research or testing would be required to determine the level of activity—this was the case for experiences found in the Steam Store. However, it was challenging to determine the hardware requirements and level of activity for experiences found for the other two contexts, which again, didn't have their own central databases, and preferred to showcase in-game scenes rather than players. It took almost twice the time required to find all the data needed for mapping when collecting data for VR arcade and on-the-street experiences than it took for at-home experiences available on Steam or on PSVR.

This means that to save time, data selected was aimed more at showcasing the cube's flexibility by showing a variety of experiences, than for accurate representation of all available experiences.

6.4 Analyzing data

Due to the aforementioned smaller-than-expected dataset, the goal of the statistical analysis had to be shifted from an in-depth survey of existing experiences and their apparent trends, to simply showing examples of possible statistical methods which can be applied to the VAE cube which was given *genre* as a fourth dimension. The chosen statistical methods might possibly change depending on the applied fourth dimension—or however many dimensions are added—but the ones used in this thesis are general enough and speak to the question of finding trends and relationships, which were the original intention for the cube's use.

The analysis itself highlighted a few concerns:

- The first was that all but two of the cube's extremities are unreachable. The rest are either impossible due to a lack of technology to make the described scenario possible, or due to contradictions within the definitions of the meeting axes. The effect this has on the statistical analysis of the cube has not yet been determined.
- There were many experiences which appeared similar in the dataset, based on the mapping to the axis scales, but which are, when observed in reality, very different. A prime example of this was the comparison of the two racing games, WipEout Omega and Sprint Vector. One is a seated experience, which was ported from flat-screen to VR, and the other is a made-for-VR room-scale game, which leverages the uniquely physical mechanics of VR. These two scored so similarly, there was only one point of difference between them on the activity axis. This was reflected in the intra-genre point distance analysis, and it is as of yet unknown if the similarity is warranted due to their relationship through genre, or if there is more work to be done in the refinement of the scales, in order to let them distinguish these kinds of experiences more accurately.
- Attempting to objectively analyze experiences might be unfair, or misleading. The interpretation of the results will always be subjective, but great care must also be

taken to assure that the analysis itself is done without bias. Using the previous example, it is unclear if comparing ported games with made-for-VR games will yield useful results, and might instead lead the designer to make wrong assumptions about cause and effect. Even though ported VR games should follow at least the bare minimum of VR design standards which assure a user's physical comfort, they will still often lack the ability to leverage VR's unique physicality on a game-mechanic level. This is where being aware of context is important, and subjectivity must be injected into the analysis, which otherwise tries to be as objective as possible.

The last of the examples brings back the AIP cube developed by Zeltzer (1992). The AIP cube focuses on highlighting the subjective qualities of experiences, such as presence and immersion. Although it lacks a clear way "to quantify rigorously these components" (Zeltzer, 1992), it can be a helpful tool if used in combination with the VAE cube. There is no avoiding the need for the subjective lens while talking about designing experiences for VR, and perhaps trying to objectively analyze experiences introduces too many ways of losing sight of that.

Another advantage of using the AIP cube along with the VAE cube is that the AIP cube has room for discussion about the context of the experience, which is absent from the VAE cube current model. Whereas the VAE cube considers sensory hardware and its effect on users in its description of experiences, it does not account for whether or not it should even be there. Zeltzer mentions that it is necessary to "identify carefully the sensory cues that must be provided for a human to accomplish the task" (Zeltzer, 1992), and whether the implementation of haptic hardware in any given experience is a result of the aforementioned hype and fetishisation around technology, or whether it was carefully considered for its ability to convey important sensory cues is not observable in the VAE cube. Only the fact that there *is* an excess of hardware for the user is represented in the cube, not the context which might justify it. Thus, using the two cubes together might give a more complete description of the experience, which can help interpret the analysis results better.

6.5 Related works

Throughout this thesis, related works were discussed in order to contextualize the VAE cube. Some of these works are tools or taxonomies which can be used to help make better design decisions by providing a framework within which designers can place their own experience in order to situate it within a certain context. These works were compared to the VAE cube, using criteria that highlight the cube's advantages and features, in order place it firmly within the context of the field.

The first related work is a list of usability heuristics for the design and evaluation of interfaces put forth by Nielsen (1994b). This list is a set of ten recommendations which can guide designers by providing a sort of checklist of software features and functions called factors that make interfaces more usable[1]. These factors are task agnostic: they can be applied to any number of new software products, and some even to physical hardware. The criteria for any system to be "good" about any of the recommendations—visibility of system status, for example—is subjective and depends entirely on the type of tasks the software is designed to accomplish. The qualification for this specific recommendation uses ambiguous language such as "through *appropriate* feedback" and "within *reasonable* time" (emphasis mine), which in turns allows for a wide range of possible solutions which can fit the description. Finally, the list was designed to be, itself, as easy to use as possible, allowing designers to clearly see where their systems are failing, and ultimately make a decision about how to fix it.

Next was the virtuality continuum presented by Milgram and Kishino, 1994, which was converted into an axis for the VAE cube. This one serves as a taxonomy to define the different levels of virtuality, and the appropriate hardware to use for each level. It does not, however, attempt to judge experiences or interactions, or even the level of immersion which each level might provide, simply stating, instead, the hardware requirements for each, and how much reality and virtuality bleed into the experience.

[1]The cited paper only lists seven of them, but more have been added and are available in an article online (Nielsen and Molich, 1994)

Milgram expressly mentions that the continuum can be used to "meaningfully compare [data]", one of the key functions of the VAE cube itself.

Following this is the AIP cube by Zeltzer, 1992, a very similar conceptual space to the VAE cube. It provides three axes for the quantitative evaluation of experiences, which can then be placed into space. This tool is to be used, again, to evaluate through comparison, and describe experiences. It is applicable to a wide variety of software uses, from animation software and such tools to entertainment and focuses primarily on the subjective aspects of virtual experiences: user presence, actor autonomy, and user interaction. Placing experiences into the space can show trends over time, and despite that the axes are made to be independent of one another in a broad sense, comparative analysis can be done within different contexts in order to better situate a new experience.

Next, Kjeldskov's interaction categories were evaluated. These, like the virtuality continuum, were adapted to become an axis of the VAE cube. These interaction categories, which describe specific types of interactions that pose unique challenges in VR, are defined and then appropriate hardware is mapped to them based on the techniques one can use to apply them. This makes the tool easy to refer to, as designers can simply select their desired outcome and in doing so, will obtain a recommendation. Since the right choices are dependent on the system's context, these mappings are subjective. Systems can be compared within this taxonomy by mapping the experience's specific context to several options in order to decide which trade-offs are better. This tool as well can be applied to a very broad range of experiences, the interactions categories are mutually exclusive and independent, meaning that if the new system does not require *orientation* as an interaction, they do not need to consider it.

Very similarly, the interaction patterns introduced by Jerald (2015) are easy to use and can be broadly applied. In fact, Jerald's system shares all the same features as Kjeldskov's but is a little more granular about its category distinctions. Additionally, the interaction patterns are not mapped to appropriate hardware but are instead mapped to interaction techniques, focusing more on the user. Like all previously mentioned works, this one can be used comparatively. Finally, Valve's high-level and informal taxonomy of VR input was evaluated (Malaika, 2015). It is a breakdown of input types for interaction in VR, and then a mapping to appropriate channels, or use cases. This taxonomy focuses on the affordance of the virtual environment and how these can affect interactions. It is fairly easy to apply, though a lot of subjective consideration is needed. Additionally, its application not limited to games, despite Valve's status as a games-centric hardware and software company.

A full comparative table is available in Figure 6.1.

Tool or Taxonomy	Objective	Subjective	Quantitative	Qualitative	Easy to apply	Accounts for wide range of examples	Allows for comparison	Allows points to be presented over time
10 Usability Heuristics for User Interface Design (Nielsen, 1994)		✔	✔		✔	✔	✔	N/A
Virtuality Continuum (Milgram and Kishino, 1994)	✔		✔			✔	✔	✔
AIP Cube (Zeltzer, 1992)		✔	✔			✔	✔	✔
Interaction Categories (Kjeldskov, 2001)		✔		✔	✔	✔	✔	N/A
Interaction Patterns (Jerald, 2015)		✔		✔	✔	✔	✔	N/A
Taxonomy of VR Input (Malaika, 2015)		✔		✔	✔	✔	✔	N/A
VAE Cube	✔		✔			✔	✔	✔
VAE+AIP Cubes	✔	✔	✔			✔	✔	✔

FIGURE 6.1: Table showing the comparison between different tools and taxonomies which aim to help make design decisions.

The following criteria for comparison were chosen for the indicated reasons:

- **Objective/Subjective**: Few of the studied taxonomies presented an evaluation of objective factors of VR experience design. This thesis partly set out to explore the possibility of making such an objective tool. Though subjective evaluations are excellent at capturing information about the human factor of design, they are not reliable or repeatable. Objective evaluations, for their part, are more reliable, but often tend to lose the human part of human-centered design. It is common to use a combination of both types, as each can answer different questions (Dünser and Billinghurst, 2011).

- **Quantitative/Qualitative**: Both qualitative and quantitative evaluations of experiences are important to have a well-rounded idea of the experience's quality. For example, half the related works present quantitative data, which is easy to visualize and react to, whereas the other half provide qualitative analyses of the experience, which relate to the user experience itself. Like the above criterion, it is a common form of evaluation for interaction design (Dünser and Billinghurst, 2011).

- **Easy to apply**: The next few are features which cropped up often enough in the related works and that could be considered useful based on their near universality. Ease of use was an explicit objective of some of these works, and with good reason (Jerald, 2015; Kjeldskov, 2001; Nielsen, 1994b). A tool is only useful if it can do its job.

- **Accounts for a wide range of examples**: Tools which can be applied to more than a few specific cases, especially in design, are extremely useful, as there are many, many different types of experiences and contexts within which these experiences can exist.

- **Allows for comparisons**: Comparative analysis can yield a lot of information in the negative space in between experience, and creates context. This, in turn, allows designers to make more informed design decisions. By determining what another experience looks like within the context of a certain evaluation, one can then place their own and detect some missing pieces to reach that same position. For example, an experience which a developer finds has excellent usability can be put through Nielsen's heuristics evaluation. Then, they can run their own experience through it and by comparing their scores, the developer can identify what factors they are lacking to match the desired score.

- **Allows points to be presented over time** This one is applicable only to the works which provide quantitative results, but it is an important feature to have. Similarly to the comparison feature, as it is in a way, a comparison, it allows a designer to place their experience within a broader context and make more informed design decisions. This one specifically can be used to detect trends, and help make decisions about next steps or new innovations.

The first comparative criteria are whether the tool or taxonomy uses objective or subjective criteria. One of the goals with the VAE cube was to create an objective evaluation tool which could help hint at desirable qualitative aspects. Contrasting this is Zeltzer's AIP cube, which measures subjective qualities in order to describe and compare experiences in a quantitative space (Zeltzer, 1992).

In a similar vein, the way the criteria are measured was compared as either quantitative or qualitative. The VAE cube is measured using quantitative scales, and, despite the subjective nature of the values, so do the AIP cube and usability heuristics (Nielsen, 1994b; Zeltzer, 1992).

The next criterion is the tool's flexibility. It was very important for the VAE cube to apply to the broad range of possible experiences, since VR has so many applications, as mentioned in Chapter 1. This thesis focused on games as a subset, but it should be as useful for other types of experiences as well, like training, simulations, CAD modelling, and more. All of the related works also feature this criterion, a clear sign that such flexibility in tools and taxonomies is useful.

Similarly, each tool in the related works allows for comparison of experiences. This is a helpful feature for tools meant to guide design decisions: a designer who can compare their own experience to that of one they think is good—for whatever metric is important to quantify good in that specific scenario—can identify what they are missing in order to achieve the quality they want. Alternatively, one can perform statistical analysis on comparative qualities for the objective tools, which can help to clarify relationships or trends.

Finally, the tools which can be converted into conceptual spaces are all capable of showing points over time. This can be a helpful feature for visualizing trends over time, and to help predict what might work in the future. Although Kjeldskov's Interaction Categories were converted into a scale to be used as an axis in the VAE cube, it had to be combined with Zeltzer's AIP *Interaction* axis in order to become quantifiable, and cannot on their own map data.

The VAE cube on its own, as highlighted by the table, does not cover all of the features available on the list. However, when used with the AIP cube, it can be applied to a broader range of uses, all but quantitative measure. This allows it to cover the gap of relating to the subjective qualities which are so important to designing for VR.

6.6 Summary

The VAE cube ended up suffering a little bit from the very phenomenon it was trying to help mitigate: inconsistency in the face of novelty. Even though it is the right time to be standardizing and taxonomizing, data selection for analysis of such newly created tools presents a challenge, due to the effects hype and technology fetishism. Much work remains in the refinement of the tool, even if the concept appears to be viable, as its comparison to related works highlights.

Chapter 7

Conclusion

This thesis revolved around the development and validation of the VAE cube, a conceptual tool which can help describe, compare, analyze, and visualize virtual experiences within a specific context. Using this cube and its flexible analysis options by inserting additional dimensions, designers can make informed design decisions which can lead to higher quality virtual experiences.

This tool was directly based off the works of Milgram and Kishino (1994), Kjeldskov (2001), and Zeltzer (1992), and inspired by the desire to fill in the perceived gap in the way designers and researchers talked about hardware's effect on immersion.

7.1 Standards and the VAE cube

Though VR has existed for a long time, it has not long been the feasible consumer phenomenon it is now. Mapping the current state of VR to the Gartner Hype Cycle, we are currently climbing up the "slope of enlightenment", which is described as the phase where "[f]ocused experimentation and real-world experience by an increasingly diverse range of enterprises lead to a better understanding of the technology's applicability, risks and benefits" (Linden and Fenn, 2003). Certainly, the phase's promised new "methodologies and tools [aimed at easing] the development process" are emerging. From Owlchemy Labs' Tomato Presence, to Valve's high level taxonomies for interaction, Oculus' design practice manual, and even Leap Motion's research into gesture-based shortcuts, interaction design standards for VR are sprouting everywhere (Fox and Schubert, 2018; Malaika, 2015; *Introduction to Best Practices*; Reimer, 2005).

Through researching these establishing taxonomies and design standards, I have identified a perceived lack in the discussion of how hardware by itself can effect immersion. Though some taxonomies discuss hardware, it is only in relation to how it can be used in their respective contexts, and it does not address the question of how hardware encumberment itself can affect experiences. For example, Milgram and Kishino (1994) discusses the kinds of display hardware which are compatible with each type of virtuality, and Kjeldskov (2001) discusses which interaction types are best suited for use with which display types. In Zeltzer (1992), the AIP cube hardly even address the role hardware plays in the rendering, interfacing and displaying virtual environments, outside of a brief mention of input modalities and the power a graphics engine needs to render certain types of scenes.

Thus, the VAE cube was created to describe experiences using hardware encumberment as a foil to Milgram's virtuality and Kjeldskov's interaction categories. As a conceptual tool, it offers the dual role of tentative taxonomy and a space within which analysis of experiences can be done in order to make more informed design decisions.

7.2 Assumptions

A few assumptions were made in order to have a direction to aim for over the course of this thesis. However, not all of them came to fruition. Below, these assumptions are discussed, and hypotheses about what went wrong are presented.

7.2.1 Relationships

Based on the tight relationship hardware has with the concepts in question for both Milgram and Kishino (1994) and the virtuality continuum, and Kjeldskov (2001) and the interaction categories, it was assumed that the VAE cube would also reflect this dynamic, that there would be a strong linear relationship between virtuality and encumberment, and perhaps a less strong but just as important relationship between activity and encumberment. This relationship would be dependent: the more virtual or active an experience, the more hardware would be required to interface with the virtual environment. Additionally, it was believed that there would exist a sort of "sweet spot" for these relationships to exist within where experiences which were considered more immersive would exist. Due to limitations described further down, the last assumption could not be tested. Finally, it was assumed that it would not only be possible to add dimensions to the cube to create more nuanced queries, but that if that added dimension were *genre*, then it would be possible to highlight trends within genres.

With the dataset represented in Chapter 4, tentative linear regression between axes shows that there was not nearly as much of a linear relation between hardware encumberment and the other two axes as originally thought. In fact, virtuality and activity had a more significant relationship than these did. It is impossible to draw certain conclusions with how insignificant the dataset is, but, as discussed in Chapter 6, it is possible that the custom-made encumberment axis is in need of refinement.

The same dataset, put through point-distance analysis this time, showed that there was, however, a demonstrable trend showing that same-genre experiences shared more similarities with one another than to other experiences.

7.2.2 Quality of experiences

As discussed in Chapter 6, the quality of experiences is a lot more nuanced to evaluate than previously thought. The VAE cube attempted to highlight the subjective qualities which are used to describe experiences using objective parameters, but the cube has not yet been found to accomplish this. The "sweet spot" for higher quality experiences, which was assumed to exist in the VAE cube, couldn't be verified within the scope and time line of this thesis, but it is, I believe, an important aspect of evaluating VR experiences, and should be researched further. This is possibly achievable by using the VAE cube and the AIP cube (Zeltzer, 1992) together, as described in Chapter 4.

Despite this, designers can still use the VAE cube to analyze trends between comparative dimensions, which can help them make more informed design decisions. For example, if it is clear based on point-distance analysis that most horror games tend to exist within a certain range of coordinates compared to other genres, then the designer can choose to intentionally subvert this trend, or to work within it. In this case, users will be used to a certain hardware setup and level of activity/virtuality in relation to horror games, and a new experiences which subverts that might meet resistance, unless the subversion was performed with a specific intention.

7.3 Limitations

The thesis was rather ambitious, as most are, in its original scope. Additionally, some unexpected challenges presented themselves, which complicated adjusted scope. The following is a list of the limitations which I have experienced in trying to complete this thesis:

- As mentioned in the discussion, fetishism of technology and hype made it difficult to obtain an objective perspective of experiences. This challenged the original intention to use experience quality as the original comparative axis along with genre, as it was assumed that the slope of enlightenment phase of the Gartner Hype Cycle would have meant that the hype would have died down more. Instead, it was very difficult to find discussions about experience quality which didn't feed into either hype or technology fetishism.

- Small sample size for the dataset made the analysis part of this methodology significantly less reliable than originally intended. It was assumed that there would be a relationship between axes, but it is difficult to say if the lack of one found during the analysis was caused by this small dataset or not. It is almost certain that the intra-genre linear regression would present different results if it was redone with a larger dataset, and it is possible the same it true about the general linear regression.
- The cube itself ended up having its own limitations: it was not able to highlight subjective qualities, nor give any clear indication about how to fix any perceived differences in datapoints, as it was assumed it would. Instead, it requires the designer to extract meaning themself, requires a large dataset to be effective for certain analyses, and needs a separate subjective analysis to accompany it.
- AIP cube's definition of Presence is possibly a little bit limiting and old fashioned, as discussed in Chapter 5. It is possible that finding a more nuanced subjective foil is necessary, or to update Zeltzer's definitions to suit newer ways of describing presence and interaction.
- One of the original goals that had to be dropped due to time constraints was to validate the axes using human participants. If this had been possible, this thesis would have been able to combine subjective user opinion about what constituted as too much encumberement, and about the effort required for tasks, which would have helped to refine the scales more. As the scales exist presently, it is possible for seated experiences with a high level of environment manipulation actions to be only a single coordinate point off from room-scale experiences with the same level of manipulation. This feels like a strangely small representation of difference when these experiences are placed in the cube, compared to how different these experiences feel to play in reality.
- The VAE cube's extremities are inaccessible, for various reason. This could potentially affect statistical analysis.

- Time constraints limited the number of statistical analyses which could be performed to a few examples, but other interesting analyses would have included a temporal analysis of VR experiences throughout VR's lifetime, as well the use of other comparative dimensions such as *ported games* vs *made-for-VR*, or *physically social VR experiences* vs *virtually social VR experiences*.

7.4 Lessons learned and speculations

The cube, although it is a sufficient proof of concept, has a long way to go before it can become a reliable tool. One of the major problems it has, where objectivity and statistical analysis is concerned, is the non-linearity of the space. The axes, being discrete, prevent the reliable application of certain useful statistical analyses. What makes them explicitly discrete is the use of the point sum system. Thus, a new system, or perhaps even a new function, would have to be devised in order to appropriately quantify these axes so that they might be continuous and linear.

This opens up a new question for, at the very least, the activity scale: what is the best way to calculate activity, if there even is a best way, which takes into consideration both the relationship activity has with hardware, as well as the active or passive way this technology is used to take this action. Should the physicality of the action be taken into consideration at all, given that it might be impeded or affected by encumbrance, and if so, how does one measure this as objectively as possible? Answering such questions will be important for making the cube more robust, consistent and reliable.

Another problem with the discreteness of the scales is that it offers only very specific situations which the experiences must fit into in order to be placed reliably into the scale. Should VR technology evolve in an unforeseen direction, the scales might no longer be sufficient for describing experiences, and would need to be changed in order to do so. Ideally, the updated, linear and continuous scales would be able to handle such evolution by allowing more generalizations to be done. Should the cube work in this way, it might then be possible to predict future evolving trends.

With continuous scales, designers and developers would be able to engage in in-depth statistical analysis of a chosen dataset. For example, a designer might want to tap into new territory, perhaps, and would use the cube to identify where fewer or no experiences exist yet, and would be given the combination of level of activity/encumberment/virtuality they would need to achieve to reach that spot. Alternatively, one could map experiences over time and compare the landscapes, determine if there have been shifts in trends throughout the years, and be the first to push through at the forefront of the trend.

7.5 Direction for future work

A few directions for future work presented themselves throughout the course of the thesis:

1. The time should be taken to subjectively rate VR experiences in order to have a database of experiences from which comparative analysis can be done.
2. A more intuitive model of the cube should be developed, as it would help designers make clearer informed decisions, and it could be done at a fraction of the time.
3. It would be a good exercise to update the AIP cube's definitions to better match more recent definitions of presence and interaction.
4. The scales need to be refined to allow for more room to distinguish between seated experiences from room-scale experiences. This can be done by engaging users to determine a finer set of requirements for what kinds of hardware constitute as encumbering, and which kinds of interactions are more physically taxing.
5. It is important to investigate the effect which the inaccessibility of so many of the cube's extremities has on the statistical analysis of experiences. If experiences can never hope to reach a certain set of coordinates, this should be accounted for in the analysis, and interpretation of results on the spread/clustering of points.

6. A statistical analysis should be done using a larger and more general dataset. This will give a better understanding of the relationship between axes, and should preferably be done after the scales have been refined.

7. Run additional analysis on the cube using other comparative dimensions in order to have a broader understanding of the axes relationships, and reveal more potential use cases for the cube.

7.6 Summary

The following research questions were answered during the course of this thesis:

1. What criteria can be used to evaluate interaction design for MR?
2. What are the relationships between the criteria?
   (a) How can we highlight these relationships?
   (b) How do these relationships affect the quality of the experience?
3. What are the steps to create a conceptual tool to help describe, compare, analyze and visualize MR experiences and to improve design?
   (a) How can this tool be used to help make informed design decisions?
4. How can existing design standards validate the conceptual tool?

A conceptual tool, presenting three criteria for evaluation—virtuality, activity and encumbrance—was created to help designers describe, compare, and analyze experiences. These criteria were picked based on the apparent lack of discussion on the topic of hardware encumberment in relation to virtuality and activity, and each taxonomy or concept was converted into a scalar dimension. These dimensions together form the VAE cube.

Data was inserted into the cube, in order to demonstrate its use cases. It was assumed the cube could highlight the relationship between encumberment as a dependant axis of both virtuality and activity, but no significant relationship could be found using linear regression. Additionally, the cube was not found to have the ability to passively show that the criteria had any effect on experience quality. Instead, it showed an uncertain but significant relationship between the criteria and the added comparative dimension, *genre*, which opens up possibility that the cube has more use as a comparative tool, than as a standalone conceptual space.

At this point, it is not believed that the objective analysis of experiences alone is enough to make any firm decisions based on data presented in the cube. Running statistical analysis, like point-distance, can help developers make better design decisions with regards to trends in the comparative dimension by providing a visual representation of the trends, but an injection of subjectivity is necessary to apply the results in a meaningful way. It is believed this can be done by mapping the VAE cube's data into the AIP cube, and comparing the mapped results' coordinates.

7.7 Closing thoughts

The goal of this thesis was to provide a conceptual tool which developers could use to describe and compare experiences. From there, they could leverage the relationship between axes and results from comparative statistical analysis to make more informed design decisions, in the hopes of improving their VR experiences. In an effort to do this, I have reviewed relevant literature, both in the fields of flat-screen interaction design, and VR interaction design, from which I have lifted useful taxonomies and tools. From these, I was able to identify a gap in the discussion: hardware encumberment and its affect on experience quality. Using these same tools, I was able to create axes for a new conceptual space, called the VAE cube, within which experiences could be placed. Following this, I have selected a set of experiences from within different contexts, and have rated the experiences so that they could be mapped into the space. This data was then analysed using statistical methods. I found no significant relationship between axes in the general dataset, but genre-based analysis showed that there was some potential that genre brought about some of the relationships. Additionally, I used point-distance calculation to show that intra-genre experiences shared more similarities to each other than to other data points. I have also preformed a subjective analysis, by mapping using the AIP cube (Zeltzer, 1992), and I did not find that any one section of the VAE cube indicated that experiences would be more likely to feature higher presence. However, I have determined that both the VAE cube and the AIP cube's descriptions did not always seem to reflect reality and that refinement of both of their axes would be required before making any reliable claims. Finally, I have re-contextualized the VAE cube within the field of interaction design tools by comparing it to related works.

This tool alone cannot answer or solve all of the design questions and problems. It can help designers make decisions about what kind of hardware encumbrance is common in certain comparative contexts, and when combined with a subjective measure, might be able to point at which combinations work most effectively to achieve presence. However, the tool still needs to be stress tested with larger and more diverse datasets, and the axes refined to allow more nuance. Additionally, the cube was designed to be used with comparative dimensions, and only one such dimension was explored in this thesis. It would be very interesting to see what kinds of results different analyses with different comparative, even temporal dimensions yield.

Bibliography

Abrash, Michael (2014). *What VR Could, Should, and Almost Certainly Will Be Within the Next Two Years*. URL: http://media.steampowered.com/apps/abrashblog/Abrash%20Dev%20Days%202014.pdf.

Anderson, Mae (2019). *Remember virtual reality? Its buzz has faded at CES 2019*. Pop Science News. URL: https://phys.org/news/2019-01-virtual-reality-ces.html.

Ashcraft, Brian (2016). *Report: Nintendo Is Researching VR*. URL: https://kotaku.com/report-nintendo-is-researching-vr-1782797521.

Billinghurst, Mark (2018). *COMP Lecture1 - Introduction to Virtual Reality*. Lecture. URL: https://www.slideshare.net/marknb00/lecture1-introduction-to-virtual-reality.

Billinghurst, Mark et al. (2015). "A survey of augmented reality". In: *Foundations and Trends R in Human–Computer Interaction* 8.2-3, pp. 73–272. URL: https://is.muni.cz/el/1433/podzim2015/PA198/um/59482554/A_Survey_of_Augmented_Reality.pdf.

Boland, Mike (2017). *What Do Consumers Want To Do in VR? (report)*. Analytical reviews of AR and VR. URL: https://arinsider.co/2017/10/10/what-do-consumers-want-to-do-in-vr-report/.

Bye, Kent (2016). *'Job Simulator' and the Magic of Hand Presence*. URL: https://www.roadtovr.com/job-simulator-magic-hand-presence/.

– (2018). *A Candid Assessment of the Industry From a Leading Indie VR Developer*. URL: https://www.roadtovr.com/candid-assessment-of-the-vr-industry-from-a-leading-indie-developer/.

Cha, Yeonjoo and Rohae Myung (2013). “Extended Fitts’ law for 3D pointing tasks using 3D target arrangements”. In: *International Journal of Industrial Ergonomics* 43.4, pp. 350–355.

Charara, Sophie and Husain Sumra (2017). *Virtual reality v augmented reality: Which is the future?* URL: https://www.wareable.com/vr/virtual-reality-vs-augmented-reality-which-is-the-future.

Darbois, Nelly, Albin Guillaud, and Nicolas Pinsault (2018). “Do Robotics and Virtual Reality Add Real Progress to Mirror Therapy Rehabilitation? A Scoping Review”. In: *Rehabilitation research and practice* 2018. URL: https://www.hindawi.com/journals/rerp/2018/6412318/.

Dünser, Andreas and Mark Billinghurst (2011). “Evaluating augmented reality systems”. In: *Handbook of augmented reality*. Springer, pp. 289–307. URL: https://www.researchgate.net/profile/Andreas_Duenser/publication/226920353_Evaluating_Augmented_Reality_Systems/links/09e4151242ea63d87e000000/Evaluating-Augmented-Reality-Systems.pdf.

Fox, Barrett and Martin Schubert (2018). *Exclusive: Validating an Experimental VR/AR Shortcut Interface with Flaming Arrows & Paper Planes*. URL: www.roadtovr.com/validating-experimental-vr-ar-shortcut-interface-leap-motion/.

Jerald, Jason (2015). *The VR book: Human-centered design for virtual reality*. Morgan & Claypool. URL: https://sci-hub.tw/10.0000/dl.acm.org/2792790.

– (2016). *The Best Way to Interact Within VR*. URL: https://uploadvr.com/the-best-way-to-interact-within-vr/.

Khan, Sarah (2017). *An Introduction to Taxonomies*. uxbooth. URL: http://www.uxbooth.com/articles/introduction-to-taxonomies/.

Kjeldskov, Jesper (2001). “Interaction: Full and partial immersive virtual reality displays”. In: *Proceedings of IRIS24, University of Bergen, Bergen*, pp. 587–600. URL: http://vbn.aau.dk/files/218295496/Kjeldskov_C1.pdf.

Lang, Ben (2018). *Valve's New Site Affirms Role of VR in Company's Future, Teases "Top secret" Games*. URL: https://www.roadtovr.com/valves-new-site-affirms-role-of-vr-in-companys-future-teases-top-secret-games/.

Linden, A. and J. Fenn (2003). *Understanding Gartner's Hype Cycles*. Strategic Analysis Report R-20-1971. Gartner, p. 12. URL: https://www.bus.umich.edu/KresgePublic/Journals/Gartner/research/115200/115274/115274.pdf.

Malaika, Yasser (2015). "Interaction Design in VR: The Rules Have Changed (Again)". In: *Proceedings of Game Developers Conference Europe. Retrieved from http://www. gdcvault.com/play/1022810/Interaction-Design-in-VR-The*.

Mazuryk, Tomasz and Michael Gervautz (1996). "Virtual reality-history, applications, technology and future". In: URL: https://www.cg.tuwien.ac.at/research/publications/1996/mazuryk-1996-VRH/TR-186-2-96-06Paper.pdf.

McCormick, Rich (2015). "Valve showing off new virtual reality hardware and updated Steam controller next week". In: *The Verge*. URL: https://www.theverge.com/2015/2/23/8094817/valve-virtual-reality-hardware-gdc-2015.

Meron, Gribetz (2017). *Meron Gribetz (Meta): The Future of Productivity Is Spatial: Introducing Meta's AR Workspace*. URL: https://www.slideshare.net/AugmentedWorldExpo/meron-gribetz-meta-the-future-of-productivity-is-spatial-introducing-metas-ar-workspace (visited on 07/24/2018).

Milgram, Paul and Fumio Kishino (1994). "A taxonomy of mixed reality visual displays". In: *IEICE TRANSACTIONS on Information and Systems* 77.12, pp. 1321–1329. URL: https://www.researchgate.net/publication/231514051_A_Taxonomy_of_Mixed_Reality_Visual_Displays.

Nielsen, Jakob (1994a). *10 Usability Heuristics for User Interface Design*. World Leaders in Research-Based User Experience. URL: https://www.nngroup.com/articles/ten-usability-heuristics/.

– (1994b). "Enhancing the explanatory power of usability heuristics". In: *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. ACM, pp. 152–158. URL: http://cs.ashoka.edu.in/cs102/papers/Nielsen94-Enhancing-Heuristics.pdf.

Nielsen, Jakob and Rolf Molich (1994). "10 Usability Heuristics for User Interface Design". In: *Nielsen Norman Group*. URL: https://www.nngroup.com/articles/ten-usability-heuristics/.

Noorian, Zeinab and Mihaela Ulieru (2010). "The state of the art in trust and reputation systems: a framework for comparison". In: *Journal of theoretical and applied electronic commerce research* 5.2, pp. 97–117. URL: https://scielo.conicyt.cl/scielo.php?pid=S0718-18762010000200007&script=sci_arttext&tlng=pt.

Oculus. *Introduction to Best Practices*. URL: https://developer.oculus.com/design/latest/concepts/book-bp/.

Owlchemy Labs (2019). *Tomato Presence!* URL: https://owlchemylabs.com/tomatopresence/.

Patrick, Kyle (2016). *GT Sport VR Functionality Limited to Dedicated "VR Tour Mode"*. URL: https://www.gtplanet.net/gt-sport-vr-functionality-limited-to-dedicated-vr-tour-mode/.

Reimer, Jeremy (2005). "A History of the GUI". In: *Ars Technica* 5. URL: http://www.cdpa.co.uk/UoP/Found/Downloads/reading6.pdf.

Robinett, Warren (1992). "Synthetic experience: A proposed taxonomy". In: *Presence: Teleoperators & Virtual Environments* 1.2, pp. 229–247.

Rubin, Peter (2014). "The inside story of oculus rift and how virtual reality became reality". In: *Wired. com*. URL: https://www.wired.com/2014/05/oculus-rift-4/.

Schoenfelder, Vinnie (2018). *AR + VR: Why Augmented Reality on Smartphones is Winning*. URL: https://www.captechconsulting.com/blogs/augmented-reality-versus-virtual-reality.

Sinclair, Brendan (2018). *VR's slow growth ideal for indies - Moss dev*. URL: https://www.gamesindustry.biz/articles/2018-08-20-vrs-slow-growth-ideal-for-indies-moss-dev.

Snow, Blake (2007). "Feature: The 10 Worst-Selling Consoles of All Time". In: URL: https://web.archive.org/web/20070508035815/http://www.gamepro.com/gamepro/domestic/games/features/111823.shtml.

Sutherland, Ivan E (1968). “A head-mounted three dimensional display”. In: *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*. ACM, pp. 757–764. URL: https://www.cise.ufl.edu/research/lok/teaching/ve-s07/papers/sutherland-headmount.pdf.

Thier, Dave (2014). “Sony Announces 'Project Morpheus:' Virtual Reality Headset For PS4”. In: *Forbes*.

Timson, Evelyn (2018). *VR vs AR: Which will become the dominant technology?* URL: https://www.itproportal.com/features/vr-vs-ar-which-will-become-the-dominant-technology/.

Virtual Reality. URL: https://www.merriam-webster.com/dictionary/virtual%20reality.

W3C (2018). *ABOUT W3C*. Community. URL: https://www.w3.org/Consortium/.

Zeltzer, David (1992). “Autonomy, interaction, and presence”. In: *Presence: Teleoperators & Virtual Environments* 1.1, pp. 127–132.

Appendix A

Python script for analysis of VAE cube

This Python script uses the Scipy and Matplotlib libraries in order to leverage their methods to calculate linear regression, point distance, and standard deviation, and then graph the results using classic slope graphing methods, as well as histograms.

```
from scipy import stats
from scipy.spatial import distance
from scipy import ndimage
import matplotlib.pyplot as plt
import numpy as np

#full list of experiences [xVirtuality, yEncumbrance, zActivity]
VAEMatrix = np.array([
    [8.6, 3.5, 10], #BeatSaber
    [8.6, 3.5, 10], #SuperHot
    [8.6, 4, 9],    #SkyrimVR
    [8.6, 3.5, 7],  #Moss
    [4.2, 3.5, 4],  #Somnai*
    [1.4, 2, 3],    #Pokemon Go
```

```
    [7.1, 10, 10],  #Scary Girl*
    [8.6, 3.5, 10], #Surgeon Simulator
    [8.6, 3.5, 10], #Lone Echo
    [7.1, 3.5, 7],  #Keep Talking and Nobody Explodes
    [8.6, 3.5, 9],  #Employee Recycling Center
    [8.6, 3.5, 7],  #Job Simulator
    [8.6, 3, 6],    #Netflix VR
    [8.6, 3.5, 6],  #Youtube VR
    [8.6, 3.5, 10], #Electronauts
    [8.6, 5.5, 1],  #Samsung Snowboard*
    [8.6, 3.5, 10], #Gorn
    [8.6, 3.5, 8],  #Wipeout Omega
    [8.6, 3.5, 5],  #Tetris Effect
    [8.6, 3.5, 7],  #Werewolves Within
    [8.6, 3.5, 6],  #Thumper
    [8.6, 3.5, 10], #Eve Valkyrie
    [8.6, 3.5, 10], #Sprint Vector
    [6.5, 6.5, 3],  #Birdly*
    [8.6, 6.5, 8],  #Pro Race VR*
    [8.6, 3.5, 7],  #Time Travel VR*
    [7.1, 10, 10],  #Amber Sky 2088*
    [7.1, 8.5, 10], #Arizona Sunshine*
    [7.1, 7.5, 10], #Ghostbusters Dimension*
    [7.1, 10, 10]   #Dreamscape Immersive*
    ])

#SCIFI MATRIX
VAEScifi = np.array([
    [8.6, 3.5, 9],  #Employee Recycling Center
```

```
        [7.1, 10, 10],  #Amber Sky 2088
        [7.1, 7.5, 10], #Ghostbusters Dimension
        [7.1, 10, 10],  #Dreamscape Immersive
        [7.1, 10, 10],  #Scary Girl*
        [8.6, 3.5, 10], #Surgeon Simulator
        [8.6, 3.5, 10]  #Lone Echo
    ])

#SLICE OF LIFE MATRIX
VAESlice = np.array([[8.6, 3.5, 7]]) #Job Simulator

VAEAction = np.array([
    [8.6, 3.5, 10],    #SuperHot
    [8.6, 3.5, 10], #Gorn
    [8.6, 3.5, 10],        #Eve Valkyrie
    [7.1, 8.5, 10]         #Arizona Sunshine
        ])

#RHYTHM MATRIX
VAERhythm = np.array([
    [8.6, 3.5, 10],        #Electronauts
    [8.6, 3.5, 10],        #BeatSaber
    [8.6, 3.5, 6]    #Tumper
        ])

#LOCATION-BASED
VAELocation = np.array([
    [1.4, 2, 3],      #Pokemon Go
    [4.2, 3.5, 4]        #Somnai
```

```
        ])


#FANTASY MATRIX
VAEFantasy = np.array([
    [8.6, 3.5, 9],      #SkyrimVR
    [8.6, 3.5, 7]               #Moss
    ])


#PUZZLE MATRIX
VAEPuzzle = np.array([
    [7.1, 3.5, 7],      #Keep Talking and Nobody Explodes
    [8.6, 3.5, 5],    #Tetris Effect
    [8.6, 3.5, 7]       #Time Travel VR
    ])


#GENERAL ENTERTAINMENT MATRIX
VAEGeneral = np.array([
    [8.6, 3, 6],        #Netflix VR
    [8.6, 3.5, 6],      #Youtube VR
    [6.5, 6.5, 3]       #Birdly
    ])

#SOCIAL/BOARDGAME
VAESocial = np.array([[8.6, 3.5, 7]])    #Werewolves Within
```

```
#RACING MATRIX
VAERacing = np.array([
    [8.6, 3.5, 10],    #Sprint Vector
    [8.6, 3.5, 8],  #Wipeout Omega
    [8.6, 6.5, 8]      #Pro Race VR
        ])



#Assign each column of the full list of experiences to their own variable
#In order to access all of their respective axes' scores at once
xVirtuality = VAEMatrix[:,0]
yEncumbrence = VAEMatrix[:,1]
zActivity = VAEMatrix[:,2]

xScifi = VAEScifi[:,0]
yScifi = VAEScifi[:,1]
zScifi = VAEScifi[:,2]

xAction = VAEAction[:,0]
yAction = VAEAction[:,1]
zAction = VAEAction[:,2]

xRhythm = VAERhythm[:,0]
yRhythm = VAERhythm[:,1]
zRhythm = VAERhythm[:,2]

xPuzzle = VAEPuzzle[:,0]
yPuzzle = VAEPuzzle[:,1]
zPuzzle = VAEPuzzle[:,2]
```

```
xGeneral = VAEGeneral[:,0]
yGeneral = VAEGeneral[:,1]
zGeneral = VAEGeneral[:,2]

xRacing = VAERacing[:,0]
yRacing = VAERacing[:,1]
zRacing = VAERacing[:,2]

#Decide which analysis to print
printLinReg      = False
printDistance    = False
printStanDev     = False
printGenreLine   = True


#============LINEAR REGRESSION================

if printLinReg:

        print('') #newline

        #====Encumbrance as a function of Virtuality=========
        EVslope, EVintercept,
            EVr_value, EVp_value,
            EVstd_err = stats.linregress(xVirtuality, yEncumbrence)

        #print coefficient of determination
        print("Encumbrance␣as␣a␣function␣of␣Virtuality")
```

```
print("coefficient␣of␣determination:␣%f" % EVr_value**2)
print('') #newline

#plot linear fit
plt.subplot(311)
plt.plot(xVirtuality, yEncumbrence,
    'o', label='original␣data')
plt.plot(xVirtuality, EVintercept +
    EVslope*xVirtuality, 'r', label='fitted␣line')
plt.title("Encumbrance␣as␣a␣function␣of␣Virtuality")
plt.ylabel("Encumbrance")
plt.xlabel("Virtuality")
plt.legend()

#====Encumbrance as a function of Activity=========
EAslope, EAintercept,
    EAr_value, EAp_value,
    EAstd_err = stats.linregress(zActivity, yEncumbrence)

#print coefficient of determination
print("Encumbrance␣as␣a␣function␣of␣Activity")
print("coefficient␣of␣determination:␣%f" % EAr_value**2)
print('') #newline

#plot linear fit
plt.subplot(312)
plt.plot(zActivity, yEncumbrence,
    'o', label='original␣data')
plt.plot(zActivity, EAintercept +
```

```
    EAslope*zActivity, 'r', label='fitted␣line')
plt.title("Encumbrance␣as␣a␣function␣of␣Activity")
plt.ylabel("Encumbrance")
plt.xlabel("Activity")
plt.legend()



#====Activity as a function of Virtuality=========
AVslope, AVintercept,
    AVr_value, AVp_value,
    AVstd_err = stats.linregress(xVirtuality, zActivity)

#print coefficient of determination
print("Activity␣as␣a␣function␣of␣Virtuality")
print("coefficient␣of␣determination:␣%f" % AVr_value**2)

#plot linear fit
plt.subplot(313)
plt.plot(xVirtuality, zActivity,
    'o', label='original␣data')
plt.plot(xVirtuality, AVintercept +
    AVslope*xVirtuality, 'r', label='fitted␣line')
plt.title("Activity␣as␣a␣function␣of␣Virtuality")
plt.ylabel("Activity")
plt.xlabel("Virtuality")
plt.legend()
```

```
        #print total number of games
        print('')#newline
        print('No of games: ', VAEMatrix.shape[0])
        print('')#newline
        print('——————————————————————————————————————')

elif printGenreLine:

        #====Find and replace the Genre in this section to match
        #====The genre you'd like to calculate a lingres for

        print('') #newline

        #====Encumbrance as a function of Virtuality=========
        EVslope, EVintercept, EVr_value,
            EVp_value, EVstd_err = stats.linregress(xRacing, yRacing)

        #print coefficient of determination
        print("Racing: encumbrance as a function of Virtuality")
        print("coefficient of determination: %f" % EVr_value**2)
        print('') #newline

        #plot linear fit
        plt.subplot(311)
        plt.plot(xRacing, yRacing, 'o', label='original data')
        plt.plot(xRacing, EVintercept +
            EVslope*xRacing, 'r', label='fitted line')
        plt.title("Racing: encumbrance as a function of Virtuality")
```

```
plt.ylabel("Encumbrance")
plt.xlabel("Virtuality")
plt.legend()

#====Encumbrance as a function of Activity=========
EAslope, EAintercept,
    EAr_value, EAp_value,
    EAstd_err = stats.linregress(zRacing, yRacing)

#print coefficient of determination
print("Racing:␣encumbrance␣as␣a␣function␣of␣Activity")
print("coefficient␣of␣determination:␣%f" % EAr_value**2)
print('') #newline

#plot linear fit
plt.subplot(312)
plt.plot(zRacing, yRacing, 'o', label='original␣data')
plt.plot(zRacing, EAintercept +
    EAslope*zRacing, 'r', label='fitted␣line')
plt.title("Racing:␣encumbrance␣as␣a␣function␣of␣Activity")
plt.ylabel("Encumbrance")
plt.xlabel("Activity")
plt.legend()



#====Activity as a function of Virtuality=========
AVslope, AVintercept,
    AVr_value,
```

```
        AVp_value, AVstd_err = stats.linregress(xRacing, zRacing)

    #print coefficient of determination
    print("Racing: activity as a function of Virtuality")
    print("coefficient of determination: %f" % AVr_value**2)

    #plot linear fit
    plt.subplot(313)
    plt.plot(xRacing, zRacing, 'o', label='original data')
    plt.plot(xRacing, AVintercept +
        AVslope*xRacing, 'r', label='fitted line')
    plt.title("Racing: activity as a function of Virtuality")
    plt.ylabel("Activity")
    plt.xlabel("Virtuality")
    plt.legend()


#============POINT DISTANCE===================

elif printDistance:

    #Action point distance
    distanceMatrixAc         = distance.pdist(VAEAction)
    #Average point distance
    averageAc                = np.mean(distanceMatrixAc)
    #Plot histogram
    plt.subplot(521)
    plt.hist(distanceMatrixAc, bins='auto')
    plt.title("Action games distance")
```

```
plt.ylabel("Frequency")
plt.xlabel("Bins")

#General Entertainment point distance
distanceMatrixGe        = distance.pdist(VAEGeneral)
#Average point distance
averageGe                       = np.mean(distanceMatrixGe)
#Plot histogram
plt.subplot(522)
plt.hist(distanceMatrixGe, bins='auto')
plt.title("GenEnt␣games␣distance")
plt.ylabel("Frequency")
plt.xlabel("Bins")

#Fantasy point distance
distanceMatrixFa        = distance.pdist(VAEFantasy)
#Average point distance
averageFa                       = np.mean(distanceMatrixFa)
#Plot histogram
plt.subplot(523)
plt.hist(distanceMatrixFa, bins='auto')
plt.title("Fantasy␣games␣distance")
plt.ylabel("Frequency")
plt.xlabel("Bins")

#Location-based point distance
distanceMatrixLo        = distance.pdist(VAELocation)
#Average point distance
averageLo                       = np.mean(distanceMatrixLo)
```

```
#Plot histogram
plt.subplot(524)
plt.hist(distanceMatrixLo, bins='auto')
plt.title("Location-based games distance")
plt.ylabel("Frequency")
plt.xlabel("Bins")

#Puzzle point distance
distanceMatrixPu        = distance.pdist(VAEPuzzle)
#Average point distance
averagePu                         = np.mean(distanceMatrixPu)
#Plot histogram
plt.subplot(525)
plt.hist(distanceMatrixPu, bins='auto')
plt.title("Puzzle games distance")
plt.ylabel("Frequency")
plt.xlabel("Bins")

#Racing point distance
distanceMatrixRa        = distance.pdist(VAERacing)
#Average point distance
averageRa                         = np.mean(distanceMatrixRa)
#Plot histogram
plt.subplot(526)
plt.hist(distanceMatrixRa, bins='auto')
plt.title("Racing games distance")
plt.ylabel("Frequency")
plt.xlabel("Bins")
```

```
#Rhythm point distance
distanceMatrixRh        = distance.pdist(VAERhythm)
#Average point distance
averageRh               = np.mean(distanceMatrixRh)
#Plot histogram
plt.subplot(527)
plt.hist(distanceMatrixRh, bins='auto')
plt.title("Rhythm games distance")
plt.ylabel("Frequency")
plt.xlabel("Bins")

#Scifi point distance
distanceMatrixSc        = distance.pdist(VAEScifi)
#Average point distance
averageSc               = np.mean(distanceMatrixSc)
#Plot histogram
plt.subplot(528)
plt.hist(distanceMatrixSc, bins='auto')
plt.title("Scifi games distance")
plt.ylabel("Frequency")
plt.xlabel("Bins")

#===The following only have one data point,
#   thus making average and point distance calculation null
#   Uncomment these sections when there is more data===

#distanceMatrixSl       = distance.pdist(VAESlice)
#averageSl              = np.mean(distanceMatrixSl)
#plt.subplot(529)
```

```
#plt.hist(distanceMatrixSl, bins='auto')
#plt.title("Slice of life games distance")
#plt.ylabel("Frequency")
#plt.xlabel("Bins")

#distanceMatrixSo       = distance.pdist(VAESocial)
#averageSo                    = np.mean(distanceMatrixSo)
#plt.subplot(5,2,10)
#plt.hist(distanceMatrixSo, bins='auto')
#plt.title("Social games distance")
#plt.ylabel("Frequency")
#plt.xlabel("Bins")



#Collecting the average distance for all the genres
genreMatrix = np.array([averageAc, averageFa,
    averageGe, averageLo,
    averagePu, averageRa,
    averageRh, averageSc]) #, averageSl, averageSo])
print('Genre matrix: ')
print(genreMatrix)
print('')
plt.subplot(529)
plt.hist(genreMatrix, bins='auto')
plt.title("Intra-genre Average Distance")

plt.ylabel("Frequency")
plt.xlabel("Bins")
```

```
        #Calculating point distance for all experiences
        distanceVAE             = distance.pdist(VAEMatrix)
        #Calculating average distance between points
        averageAll                      = np.mean(distanceVAE)
        print('All points distances average: ')
        print(averageAll)
        print('')

        plt.subplot(5,2,10)
        plt.hist(distanceVAE, bins='auto')
        plt.title("All Points Distances")

        plt.ylabel("Frequency")
        plt.xlabel("Bins")

        #Calculating standard deviation between all distances
        print('Standard Deviation between overall point distance: ')
        PDstandardDeviation = ndimage.standard_deviation(distanceVAE)
        print(PDstandardDeviation)



        #plt.imshow(averageAll, cmap='hot', interpolation='nearest')



#==============STANDARD DEVIATION====================

elif printStanDev:
        print('Standard Deviation: ')
        standardDeviation = ndimage.standard_deviation(VAEMatrix)
```

```
        print(standardDeviation)




if printLinReg or printDistance or printGenreLine:
        plt.subplots_adjust(left=None,
            bottom=None, right=None, top=None,
            wspace=1, hspace=1)
        plt.show()
```

Appendix B

Python script for analysis of AIP cube

This Python script uses the Scipy and Matplotlib libraries in order to leverage their methods to calculate point distance, and standard deviation, and then graph the results histograms.

```
from scipy import stats
from scipy.spatial import distance
from scipy import ndimage
import matplotlib.pyplot as plt
import numpy as np


#full list of experiences [xAutonomy, yInteraction , zPresence]
AIPMatrix = np.array([
    [0, 5, 6], #BeatSaber
    [6, 6, 6], #SuperHot
    [5, 7, 6], #SkyrimVR
    [2, 5, 6], #Moss
    [0, 1, 1], #Pokemon Go
```

```
    [1, 6, 8],  #Scary Girl*
    [0, 6, 6],  #Surgeon Simulator
    [1, 6, 6],  #Lone Echo
    [0, 8, 6],  #Keep Talking and Nobody Explodes
    [5, 8, 6],  #Job Simulator
    [0, 0, 5],  #Netflix VR
    [0, 0, 6],  #Youtube VR
    [0, 8, 6],  #Electronauts
    [6, 7, 6],          #Gorn
    [5, 6, 6],    #Wipeout Omega
    [0, 6, 5],    #Tetris Effect
    [0, 1, 6],    #Werewolves Within
    [0, 4, 5],    #Thumper
    [5, 6, 6],  #Eve Valkyrie
    [0, 5, 6],  #Sprint Vector
    [0, 1, 10], #Birdly*
    [0, 8, 8],  #Pro Race VR*
    ])

#SCIFI MATRIX
AIPScifi = np.array([
    [1, 6, 8],  #Scary Girl*
    [0, 6, 6],  #Surgeon Simulator
    [1, 6, 6]   #Lone Echo
    ])

#SLICE OF LIFE MATRIX
AIPSlice = np.array([[5, 8, 6]])        #Job Simulator
```

```
AIPAction = np.array([
    [6, 6, 6],  #SuperHot
    [6, 7, 6],  #Gorn
    [5, 6, 6]   #Eve Valkyrie
    ])

#RHYTHM MATRIX
AIPRhythm = np.array([
    [0, 8, 6],  #Electronauts
    [0, 5, 6],  #BeatSaber
    [0, 4, 5]   #Thumper
    ])

#LOCATION-BASED
AIPLocation = np.array([[0, 1, 1]])     #Pokemon Go



#FANTASY MATRIX
AIPFantasy = np.array([
    [5, 7, 6],  #SkyrimVR
    [2, 5, 6]   #Moss
    ])



#PUZZLE MATRIX
AIPPuzzle = np.array([
    [0, 8, 6],  #Keep Talking and Nobody Explodes
    [0, 6, 5]   #Tetris Effect
    ])
```

```
#GENERAL ENTERTAINMENT MATRIX
AIPGeneral = np.array([
    [0, 0, 5], #Netflix VR
    [0, 0, 6], #Youtube VR
    [0, 1, 10] #Birdly
    ])

#SOCIAL/BOARDGAME
AIPSocial = np.array([[0, 1, 6]])    #Werewolves Within


#RACING MATRIX
AIPRacing = np.array([
    [0, 5, 6], #Sprint Vector
    [5, 6, 6], #Wipeout Omega
    [0, 8, 8]  #Pro Race VR
        ])


#Assign each column of the full list of experiences to their own variable
#In order to access all of their respective axes' scores at once
xAutonomy = AIPMatrix[:,0]
yInteraction = AIPMatrix[:,1]
zPresence = AIPMatrix[:,2]


#Decide which analysis to print
```

```
printDistance = True
printStanDev = False



#============POINT DISTANCE===================

if printDistance:

        #Action point distance
        distanceMatrixAc      = distance.pdist(AIPAction)
        #Average point distance
        averageAc             = np.mean(distanceMatrixAc)
        #Plot histogram
        plt.subplot(521)
        plt.hist(distanceMatrixAc, bins='auto')
        plt.title("Action␣games␣distance")
        plt.ylabel("Frequency")
        plt.xlabel("Bins")

        #General Entertainment point distance
        distanceMatrixG       = distance.pdist(AIPGeneral)
        #Average point distance
        averageGe             = np.mean(distanceMatrixGe)
        #Plot histogram
        plt.subplot(522)
        plt.hist(distanceMatrixGe, bins='auto')
        plt.title("GenEnt␣games␣distance")
        plt.ylabel("Frequency")
        plt.xlabel("Bins")
```

```
#Fantasy point distance
distanceMatrixF      = distance.pdist(AIPFantasy)
#Average point distance
averageFa            = np.mean(distanceMatrixFa)
#Plot histogram
plt.subplot(523)
plt.hist(distanceMatrixFa, bins='auto')
plt.title("Fantasy games distance")
plt.ylabel("Frequency")
plt.xlabel("Bins")


#Puzzle point distance
distanceMatrixPu     = distance.pdist(AIPPuzzle)
#Average point distance
averagePu            = np.mean(distanceMatrixPu)
#Plot histogram
plt.subplot(525)
plt.hist(distanceMatrixPu, bins='auto')
plt.title("Puzzle games distance")
plt.ylabel("Frequency")
plt.xlabel("Bins")

#Racing point distance
distanceMatrixRa     = distance.pdist(AIPRacing)
#Average point distance
averageRa            = np.mean(distanceMatrixRa)
#Plot histogram
```

```
plt.subplot(526)
plt.hist(distanceMatrixRa, bins='auto')
plt.title("Racing games distance")
plt.ylabel("Frequency")
plt.xlabel("Bins")

#Rhythm point distance
distanceMatrixRh     = distance.pdist(AIPRhythm)
#Average point distance
averageRh            = np.mean(distanceMatrixRh)
#Plot histogram
plt.subplot(527)
plt.hist(distanceMatrixRh, bins='auto')
plt.title("Rhythm games distance")
plt.ylabel("Frequency")
plt.xlabel("Bins")

#Scifi point distance
distanceMatrixSc     = distance.pdist(AIPScifi)
#Average point distance
averageSc            = np.mean(distanceMatrixSc)
#Plot histogram
plt.subplot(528)
plt.hist(distanceMatrixSc, bins='auto')
plt.title("Scifi games distance")

plt.ylabel("Frequency")
plt.xlabel("Bins")
```

```
#===The following only have one data point,
#    thus making average and point distance calculation null
#    Uncomment these sections when there is more data===

#Location-based point distance
#distanceMatrixLo   = distance.pdist(AIPLocation)
#Average point distance
#averageLo          = np.mean(distanceMatrixLo)
#Plot histogram
#plt.subplot(524)
#plt.hist(distanceMatrixLo, bins='auto')
#plt.title("Location-based games distance")
#plt.ylabel("Frequency")
#plt.xlabel("Bins")

#distanceMatrixSl   = distance.pdist(AIPSlice)
#averageSl          = np.mean(distanceMatrixSl)
#plt.subplot(529)
#plt.hist(distanceMatrixSl, bins='auto')
#plt.title("Slice of life games distance")
#plt.ylabel("Frequency")
#plt.xlabel("Bins")

#distanceMatrixSo   = distance.pdist(AIPSocial)
#averageSo          = np.mean(distanceMatrixSo)
#plt.subplot(5,2,10)
#plt.hist(distanceMatrixSo, bins='auto')
#plt.title("Social games distance")
#plt.ylabel("Frequency")
```

```
#plt.xlabel("Bins")



#Collecting the average distance for all the genres
genreMatrix = np.array([
averageAc, averageFa, averageGe,
averagePu, averageRa, averageRh,
averageSc]) #, averageLo, averageSl, averageSo])
print('Genre matrix: ')
print(genreMatrix)
print('')
plt.subplot(529)
plt.hist(genreMatrix, bins='auto')
plt.title("Intra-genre Average Distance")

plt.ylabel("Frequency")
plt.xlabel("Bins")

#Calculating point distance for all experiences
distanceAIP              = distance.pdist(AIPMatrix)
#Calculating average distance between points
averageAll                        = np.mean(distanceAIP)
print('All points distances average: ')
print(averageAll)
print('')

plt.subplot(5,2,10)
plt.hist(distanceAIP, bins='auto')
plt.title("All Points Distances")
```

```
        plt.ylabel("Frequency")
        plt.xlabel("Bins")

        #Calculating standard deviation
        #between all genre distance averages
        print('Standard␣Deviation␣between
␣␣␣␣␣␣␣␣␣␣␣␣point␣average␣distances␣per␣genre:␣')
        PDstandardDeviation = ndimage.standard_deviation(genreMatrix)
        print(PDstandardDeviation)
        print('')

        #Calculating standard deviation between all distances
        print('Standard␣Deviation␣between␣overall␣point␣distance:␣')
        PDstandardDeviation = ndimage.standard_deviation(distanceAIP)
        print(PDstandardDeviation)



        #plt.imshow(averageAll, cmap='hot', interpolation='nearest')



#==============STANDARD DEVIATION====================

elif printStanDev:
        print('Standard␣Deviation:␣')
        standardDeviation = ndimage.standard_deviation(AIPMatrix)
        print(standardDeviation)
```

```
if printDistance:
        plt.subplots_adjust(left=None, bottom=None,
            right=None, top=None,
            wspace=1, hspace=1)
        plt.show()
```

Appendix C

Genre-filtered Representations in the VAE cube

The following are representations of the VAE cube's dataset filtered by genre.

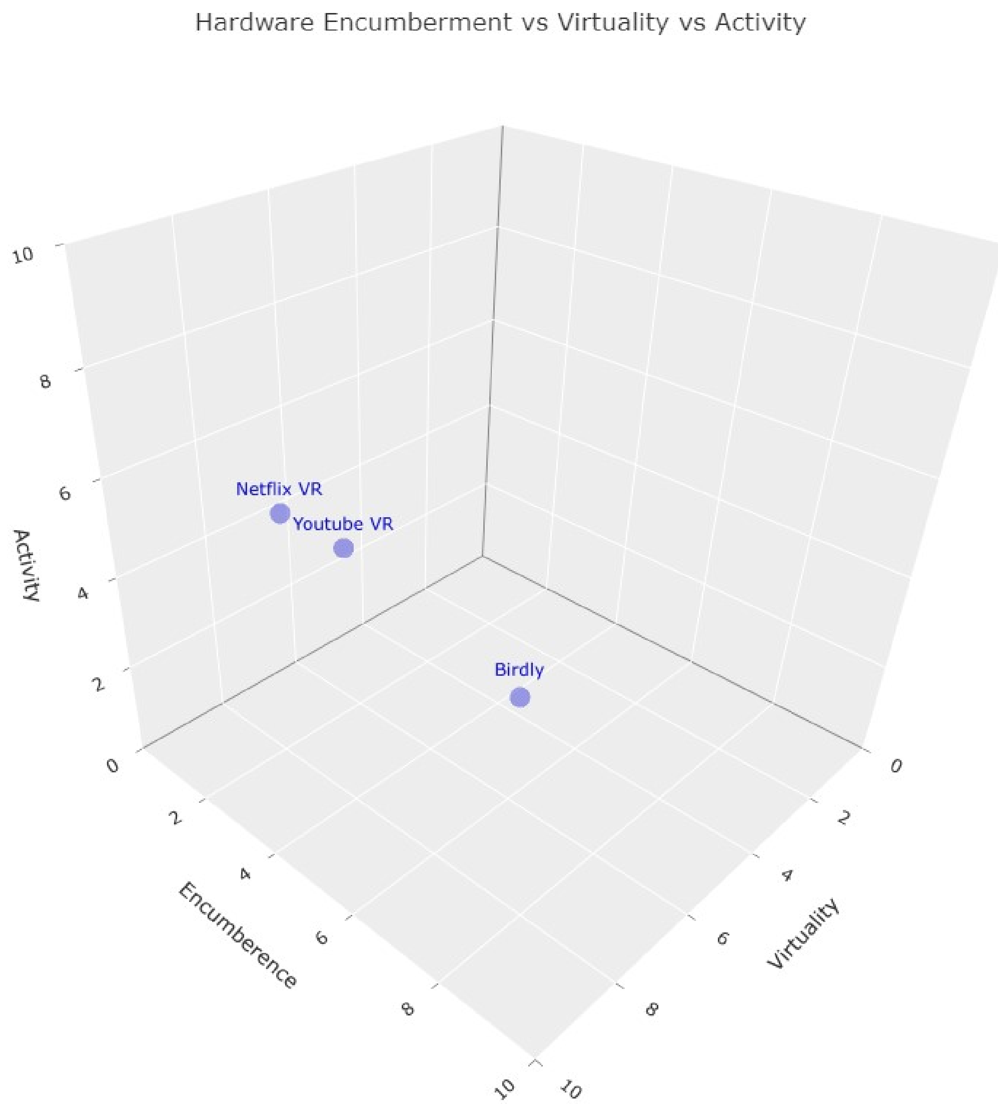


FIGURE C.1: VAE cube filtered to show experiences which belong to the *general entertainment* genre.

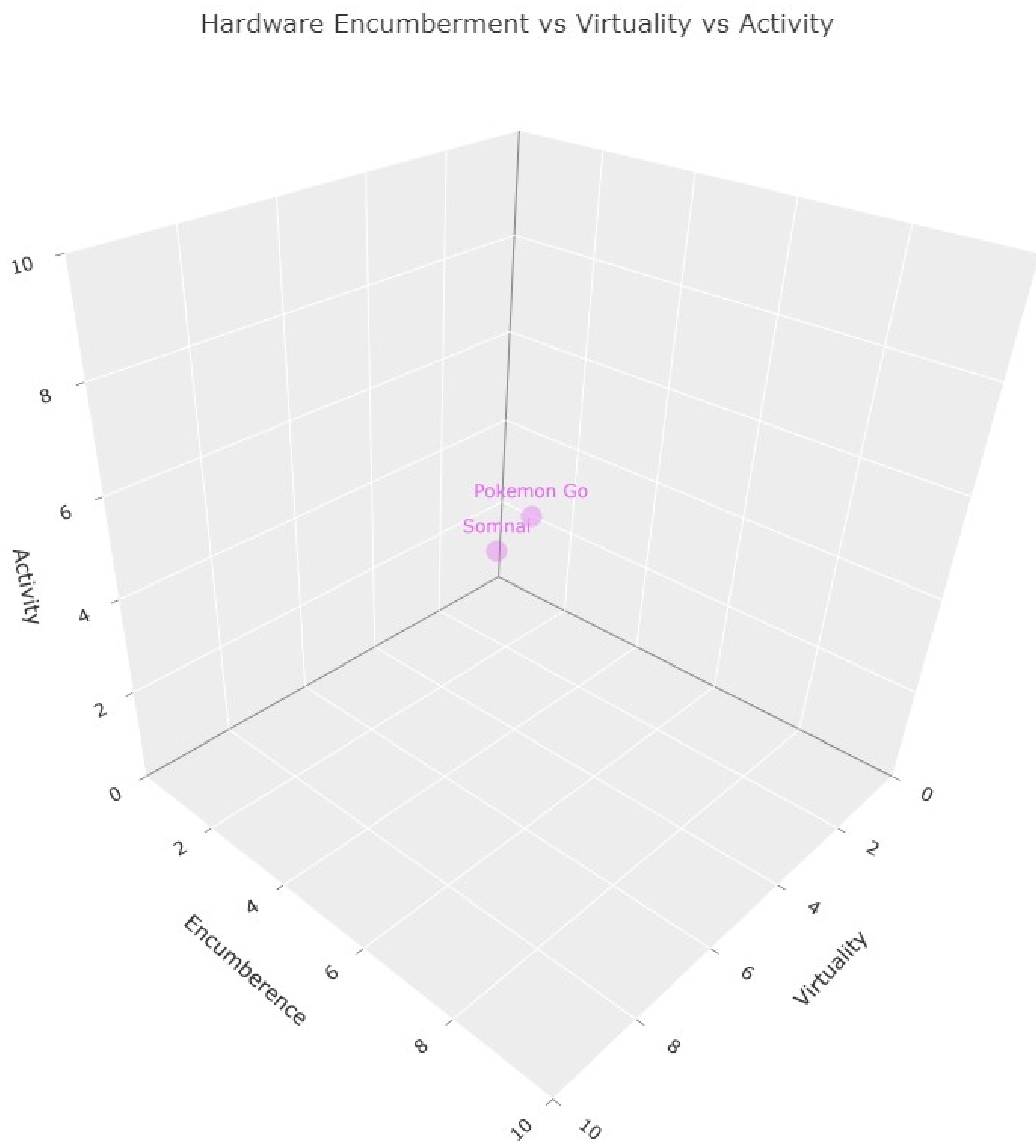


FIGURE C.2: VAE cube filtered to show experiences which belong to the *location-based* genre.

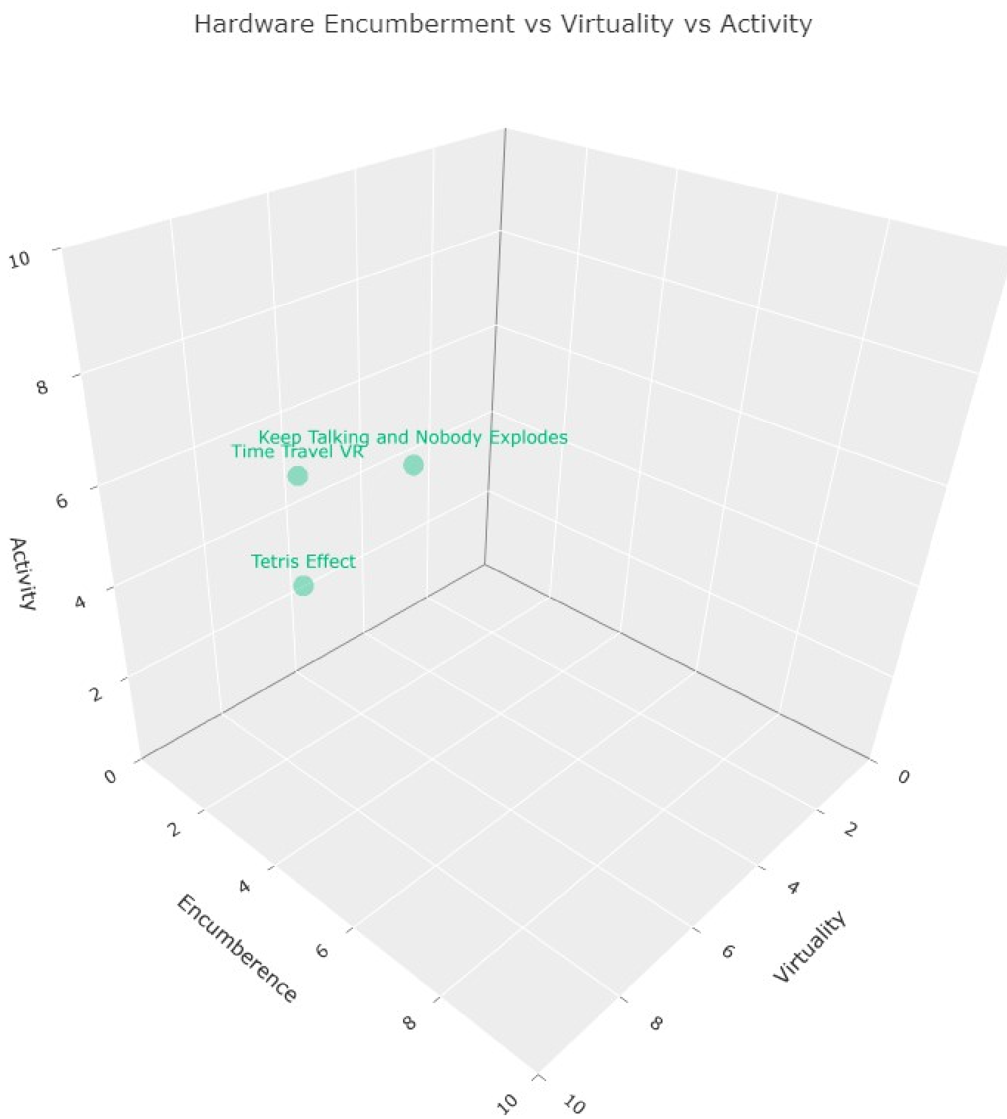


FIGURE C.3: VAE cube filtered to show experiences which belong to the *puzzle* genre.

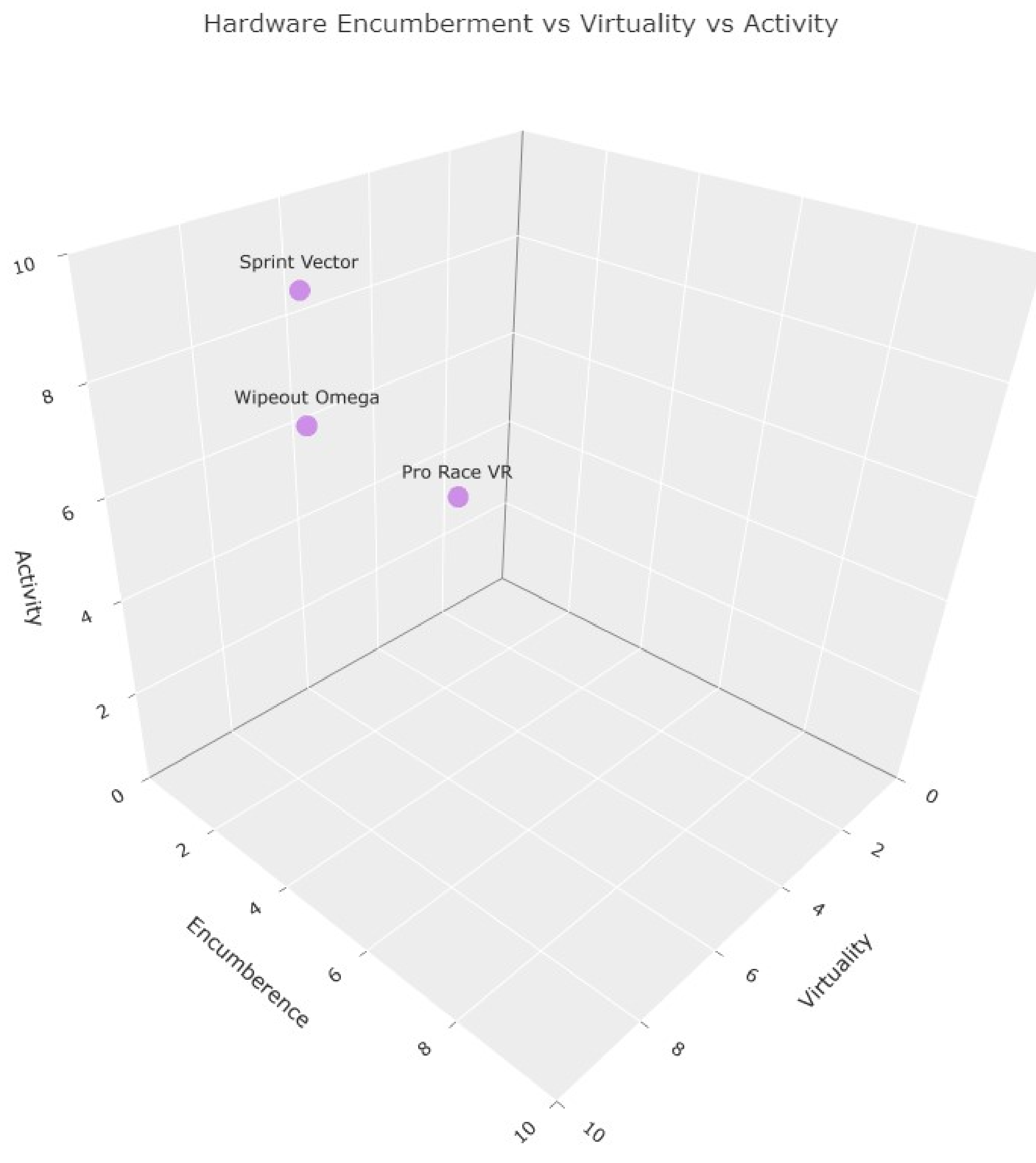


FIGURE C.4: VAE cube filtered to show experiences which belong to the *racing* genre.

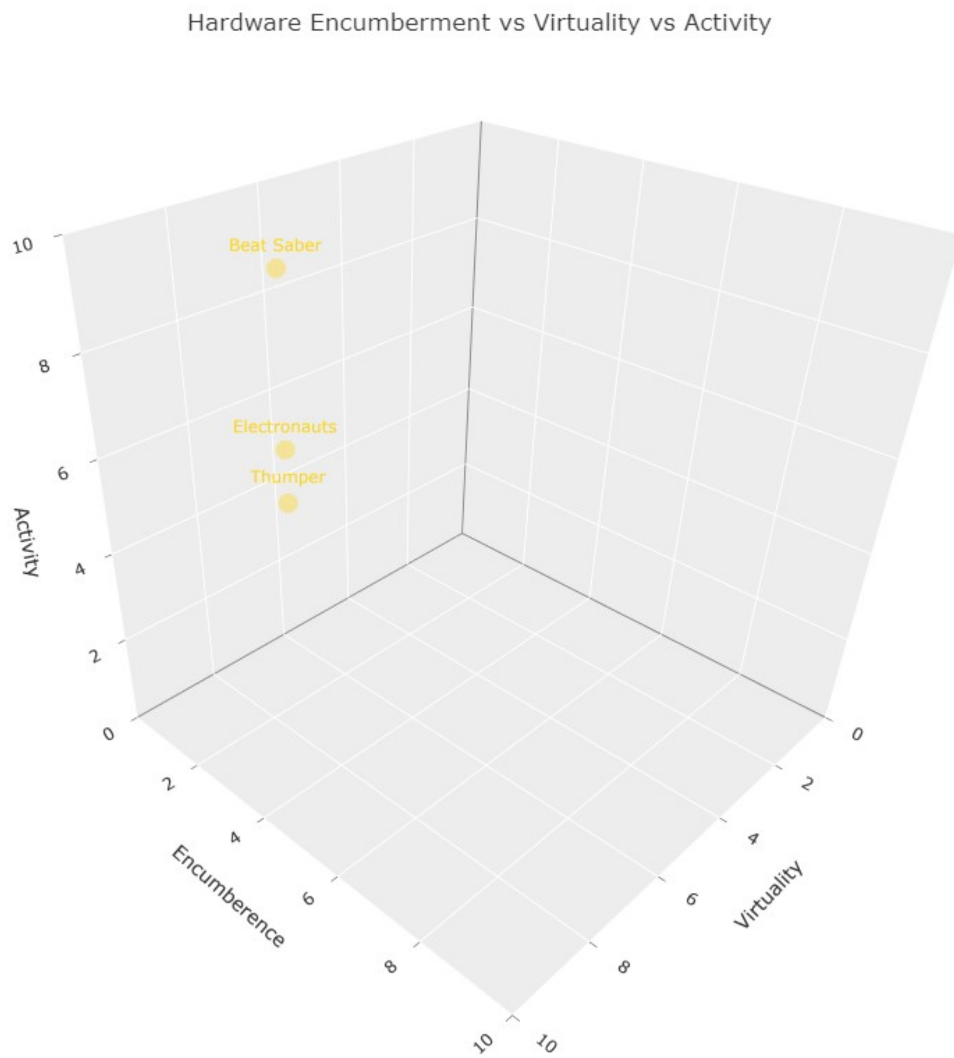


FIGURE C.5: VAE cube filtered to show experiences which belong to the *rhythm* genre.

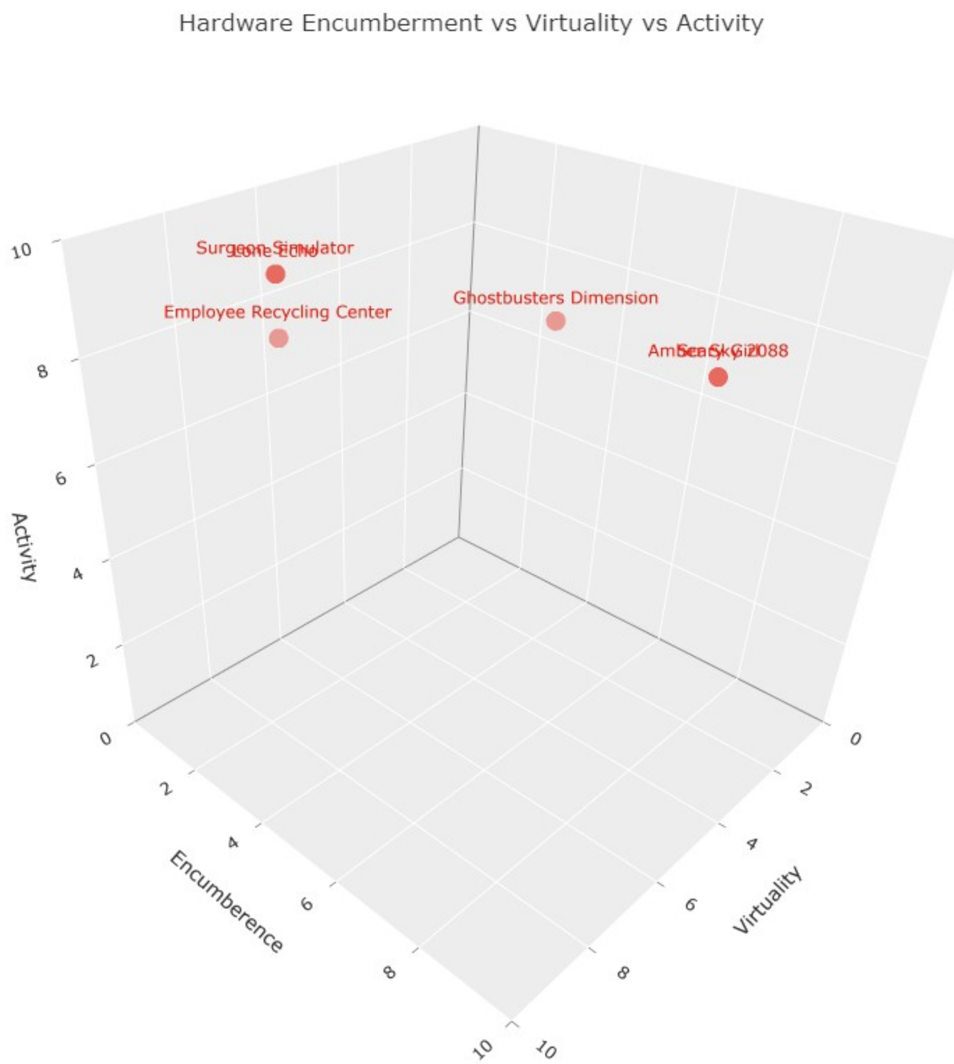


FIGURE C.6: VAE cube filtered to show experiences which belong to the *scifi* genre.

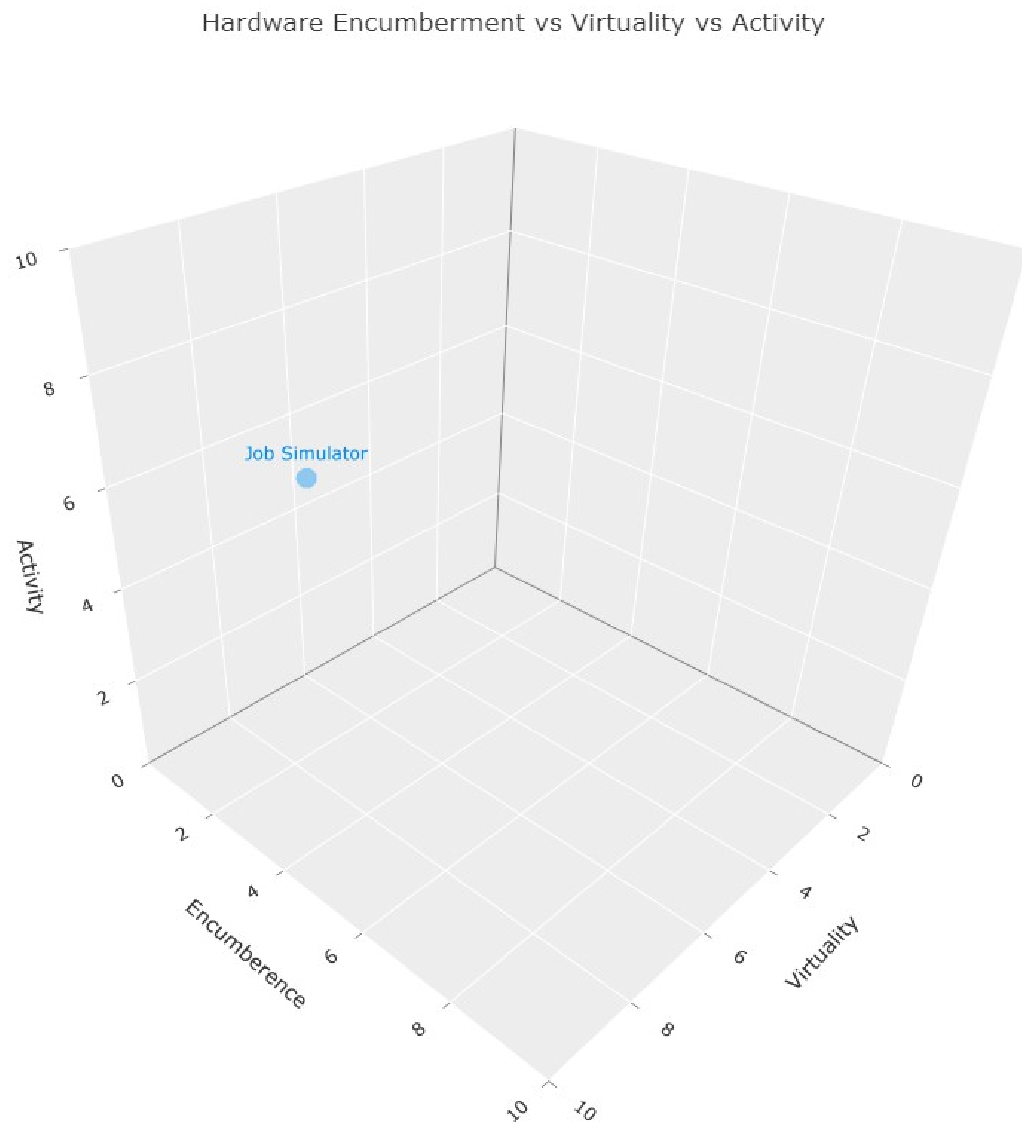


FIGURE C.7: VAE cube filtered to show experiences which belong to the *slice of life* genre.

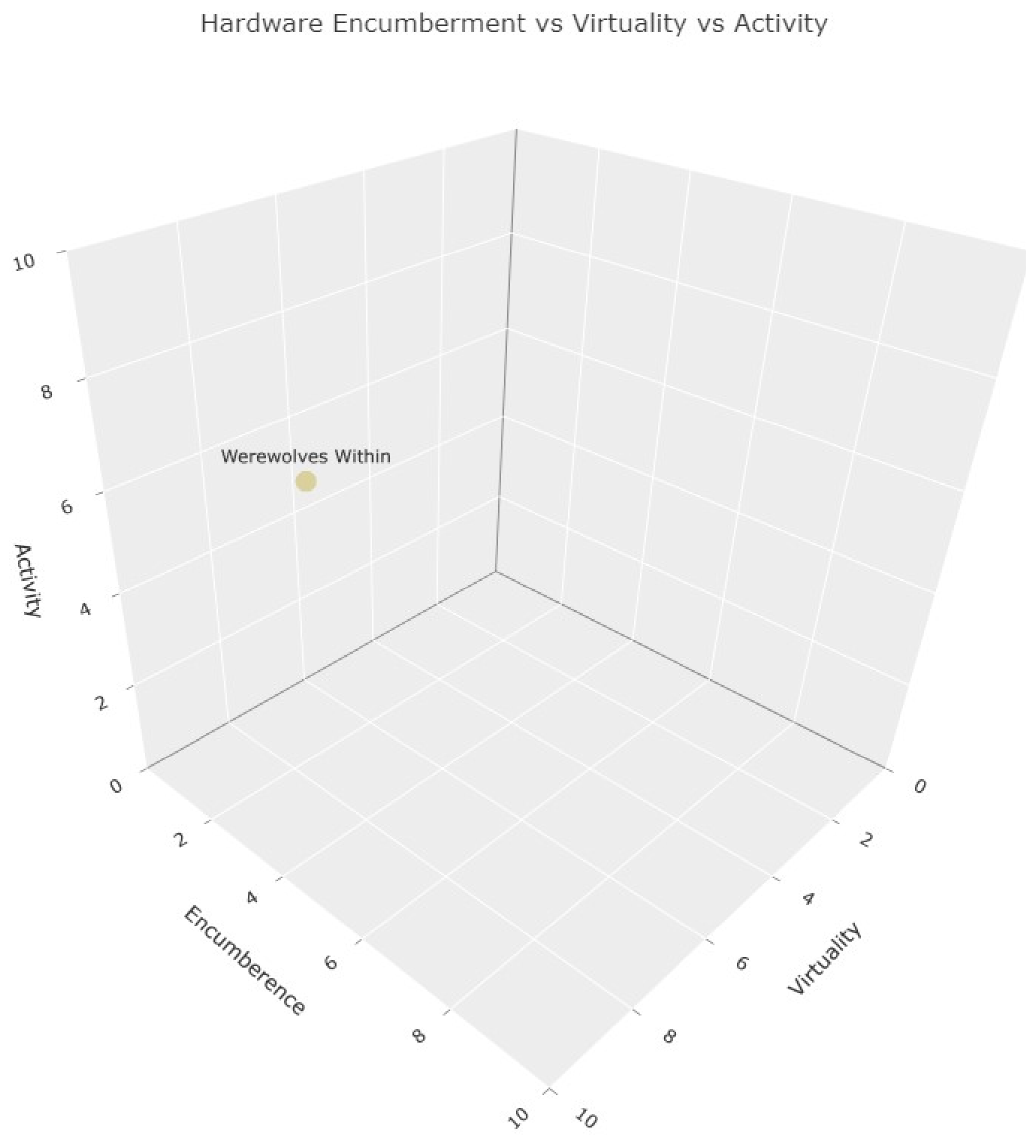


FIGURE C.8: VAE cube filtered to show experiences which belong to the *board-game/social* genre.

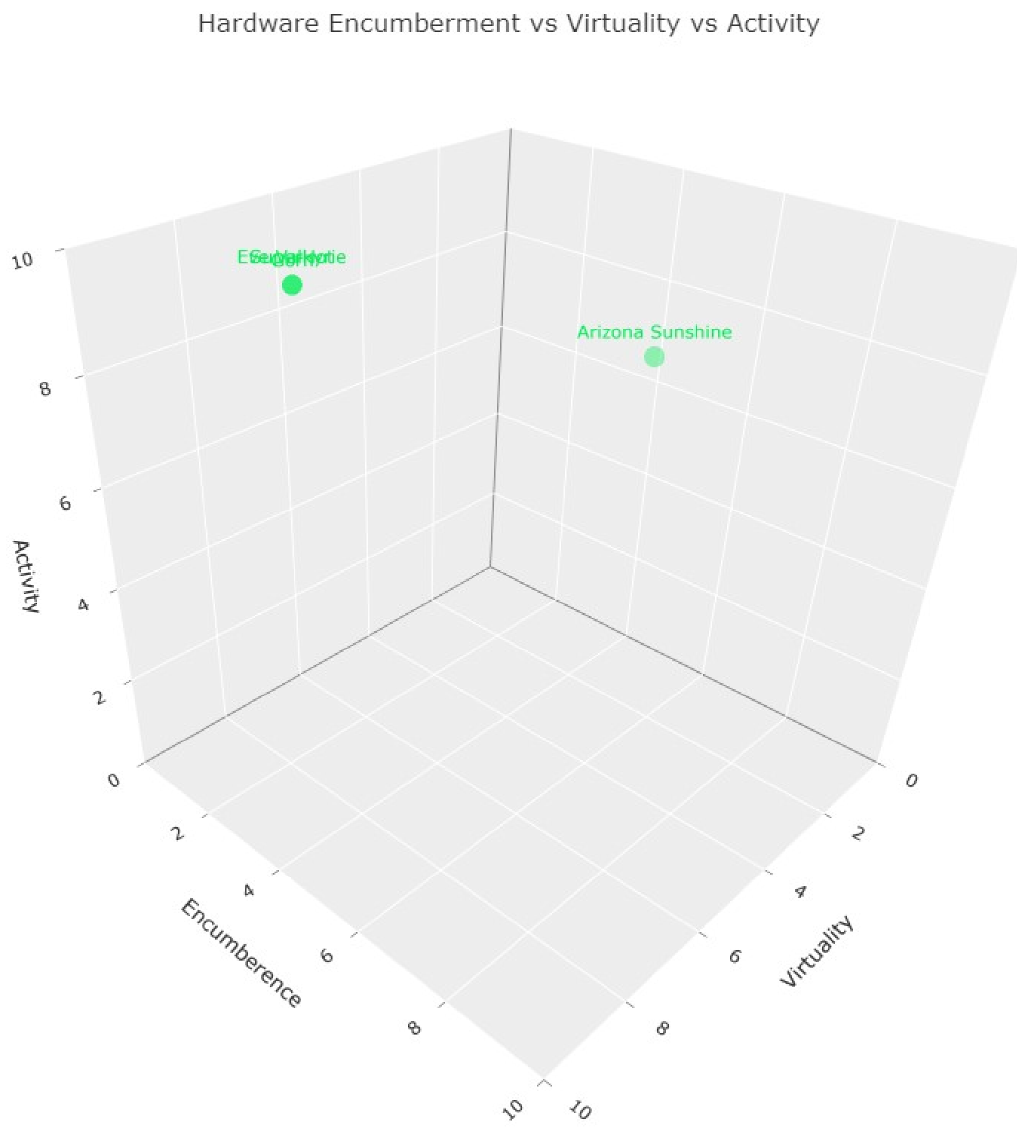


FIGURE C.9: VAE cube filtered to show experiences which belong to the *action* genre.

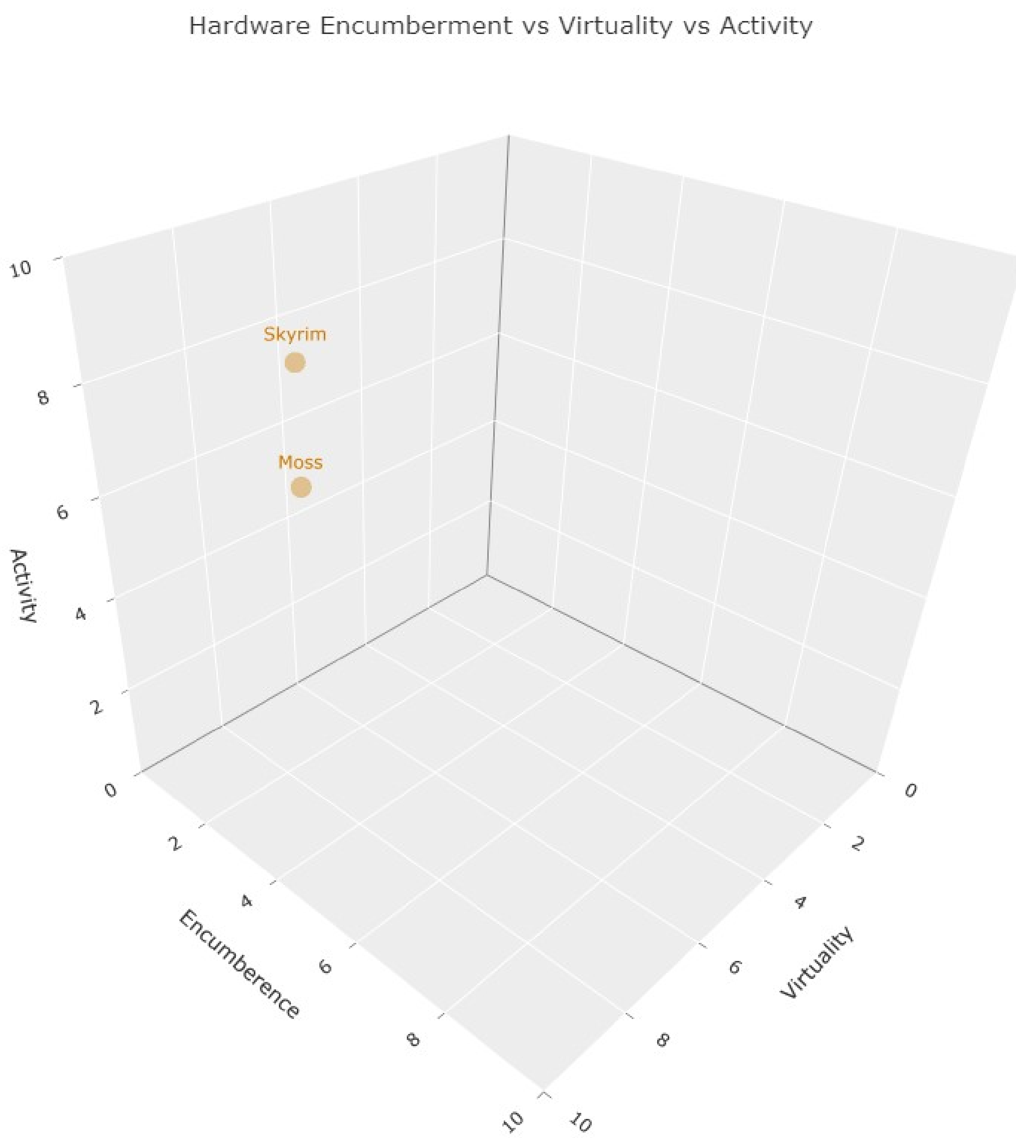


FIGURE C.10: VAE cube filtered to show experiences which belong to the *fantasy* genre.

Appendix D

Genre-filtered Representations in the AIP cube

The following are representations of the AIP cube's dataset filtered by genre.

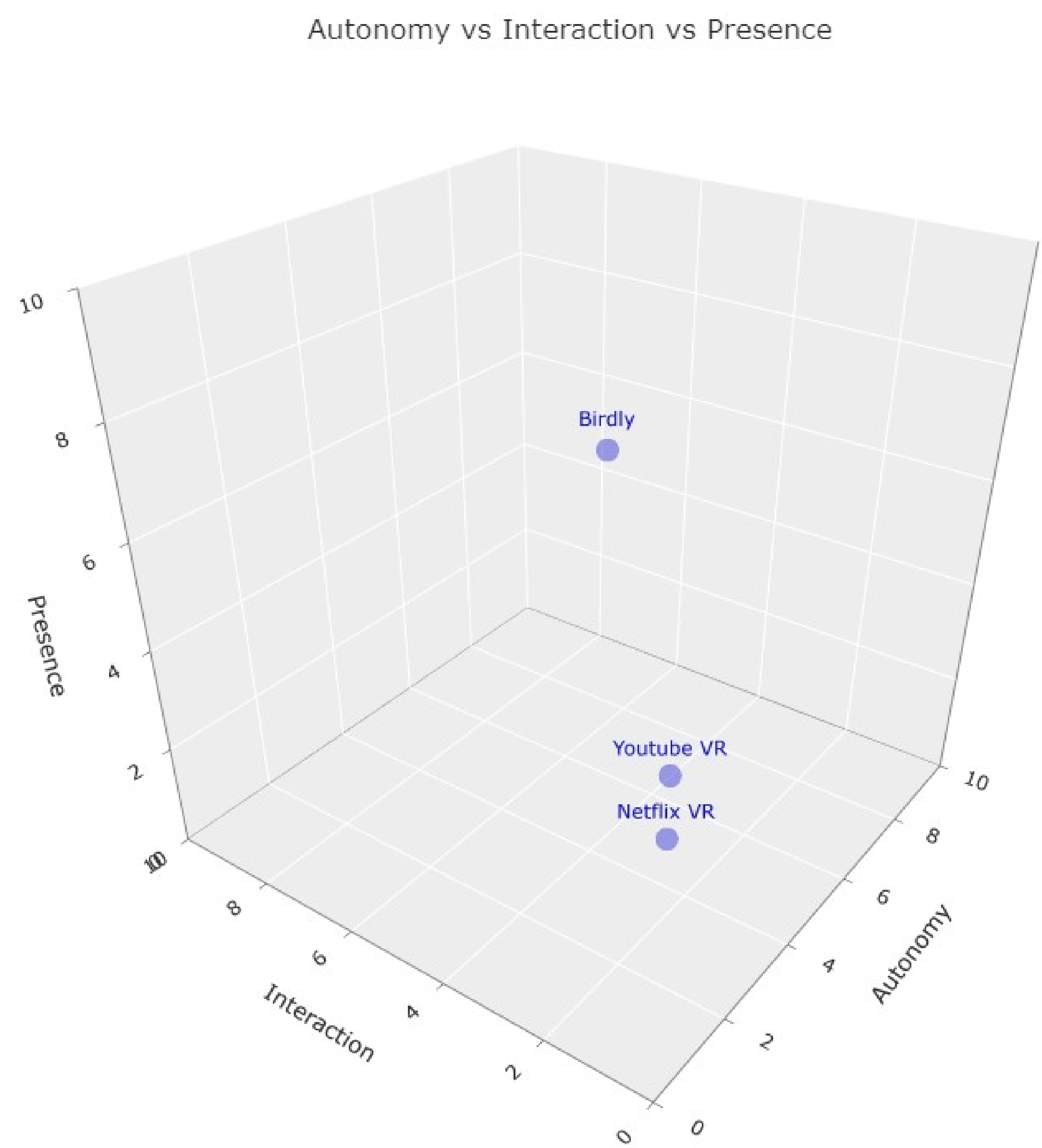


FIGURE D.1: AIP cube filtered to show experiences which belong to the *general entertainment* genre.

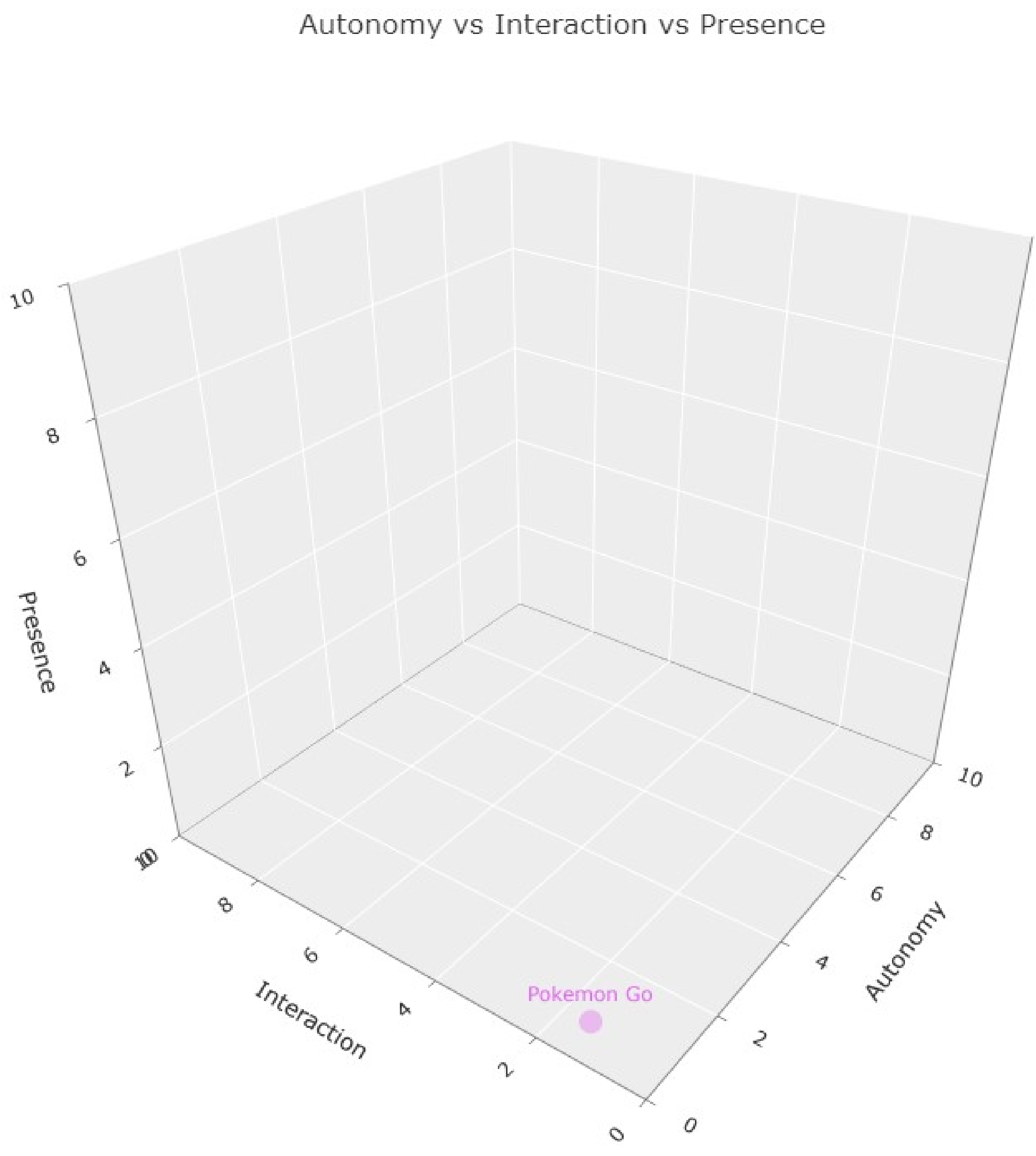


FIGURE D.2: AIP cube filtered to show experiences which belong to the *location-based* genre.

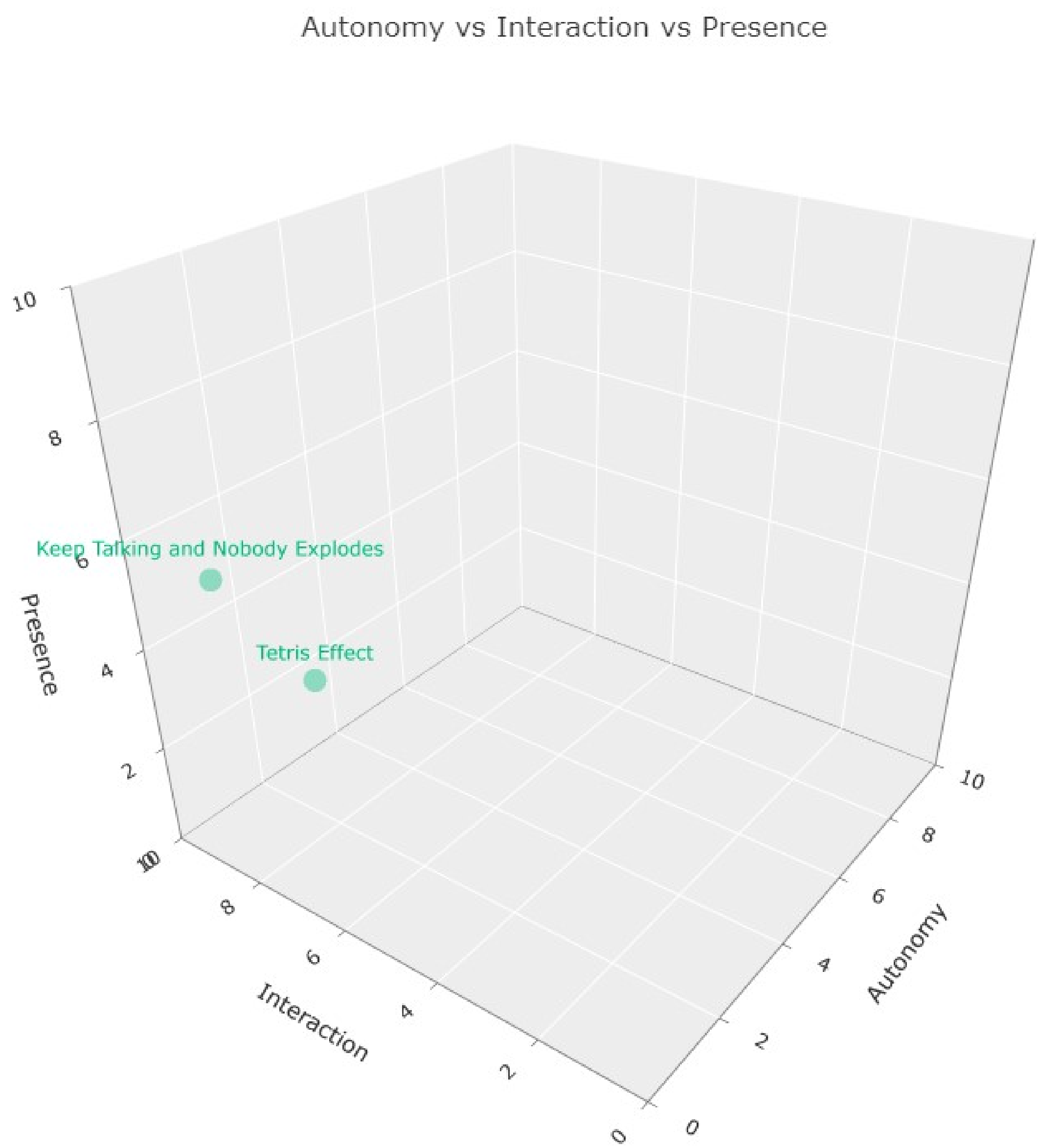


FIGURE D.3: AIP cube filtered to show experiences which belong to the *puzzle* genre.

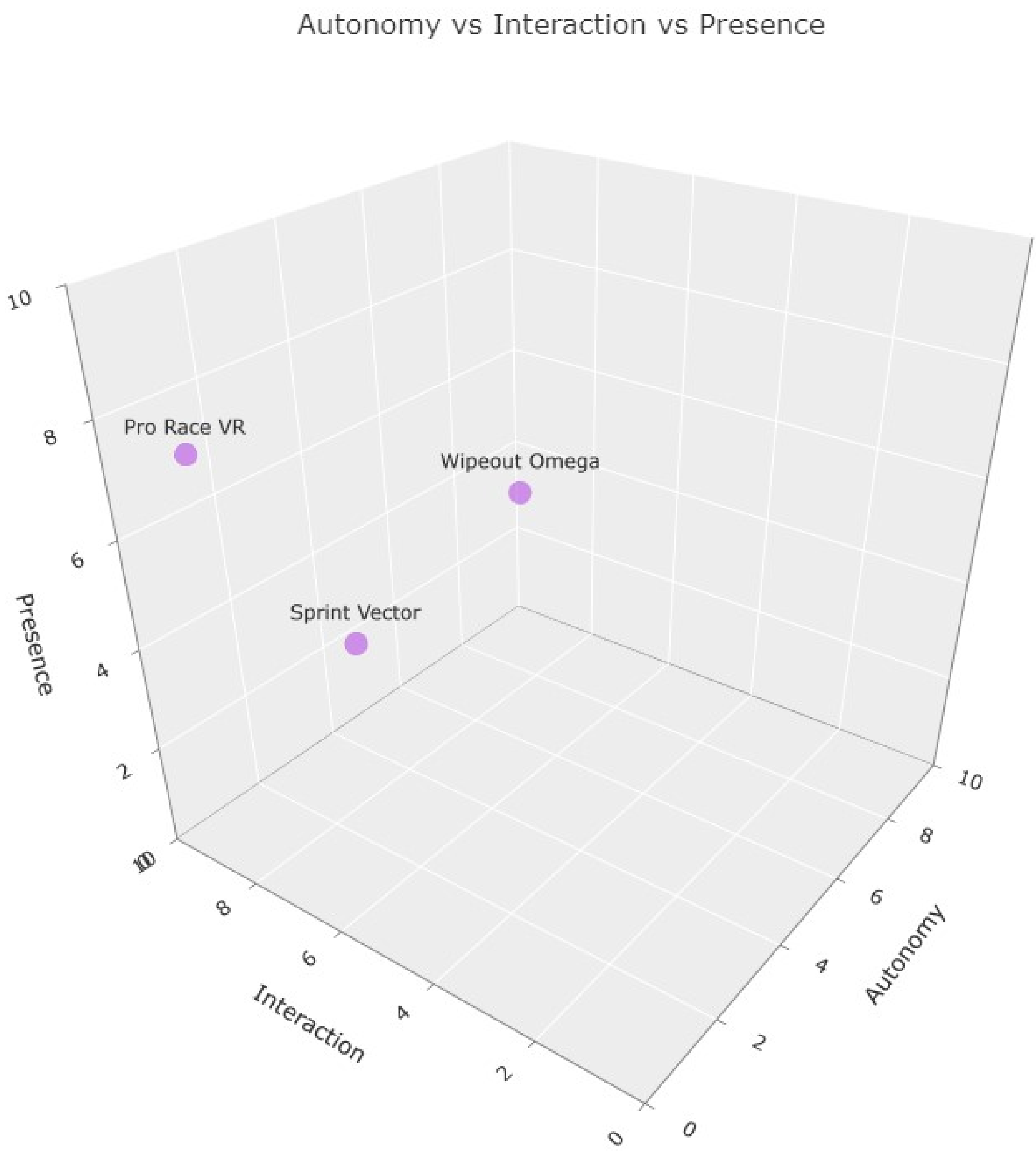


FIGURE D.4: AIP cube filtered to show experiences which belong to the *racing* genre.

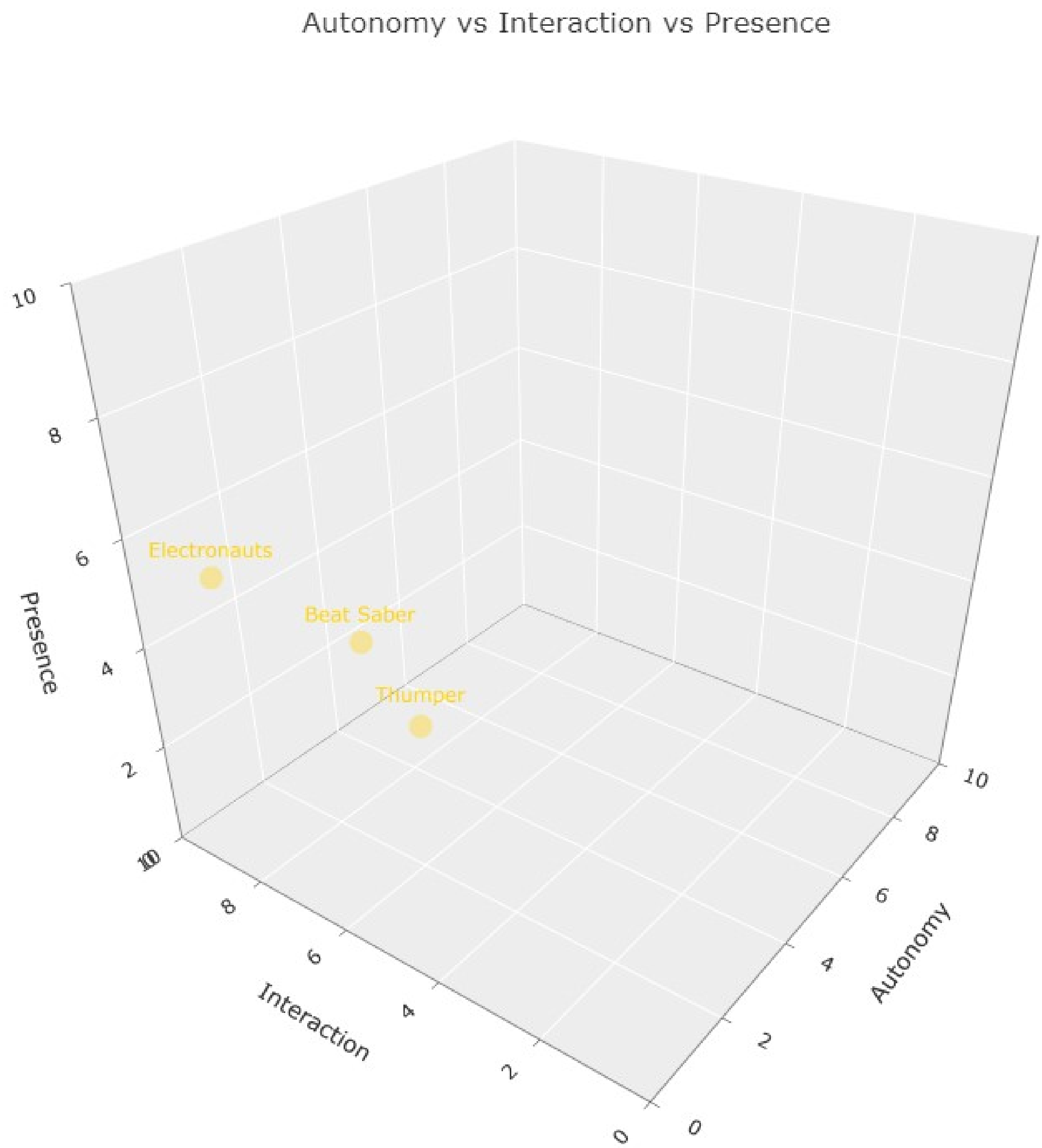


FIGURE D.5: AIP cube filtered to show experiences which belong to the *rhythm* genre.

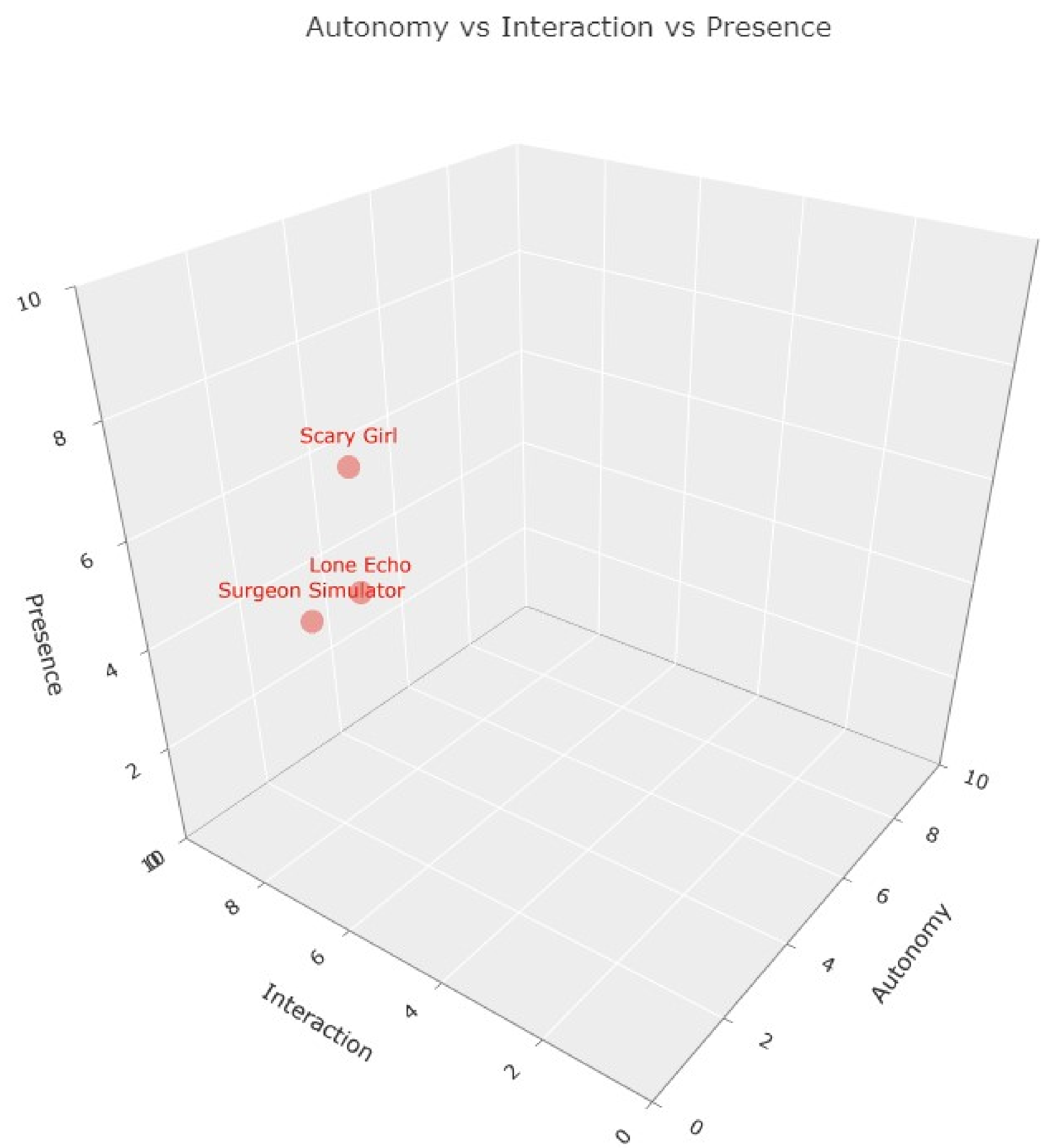


FIGURE D.6: AIP cube filtered to show experiences which belong to the *scifi* genre.

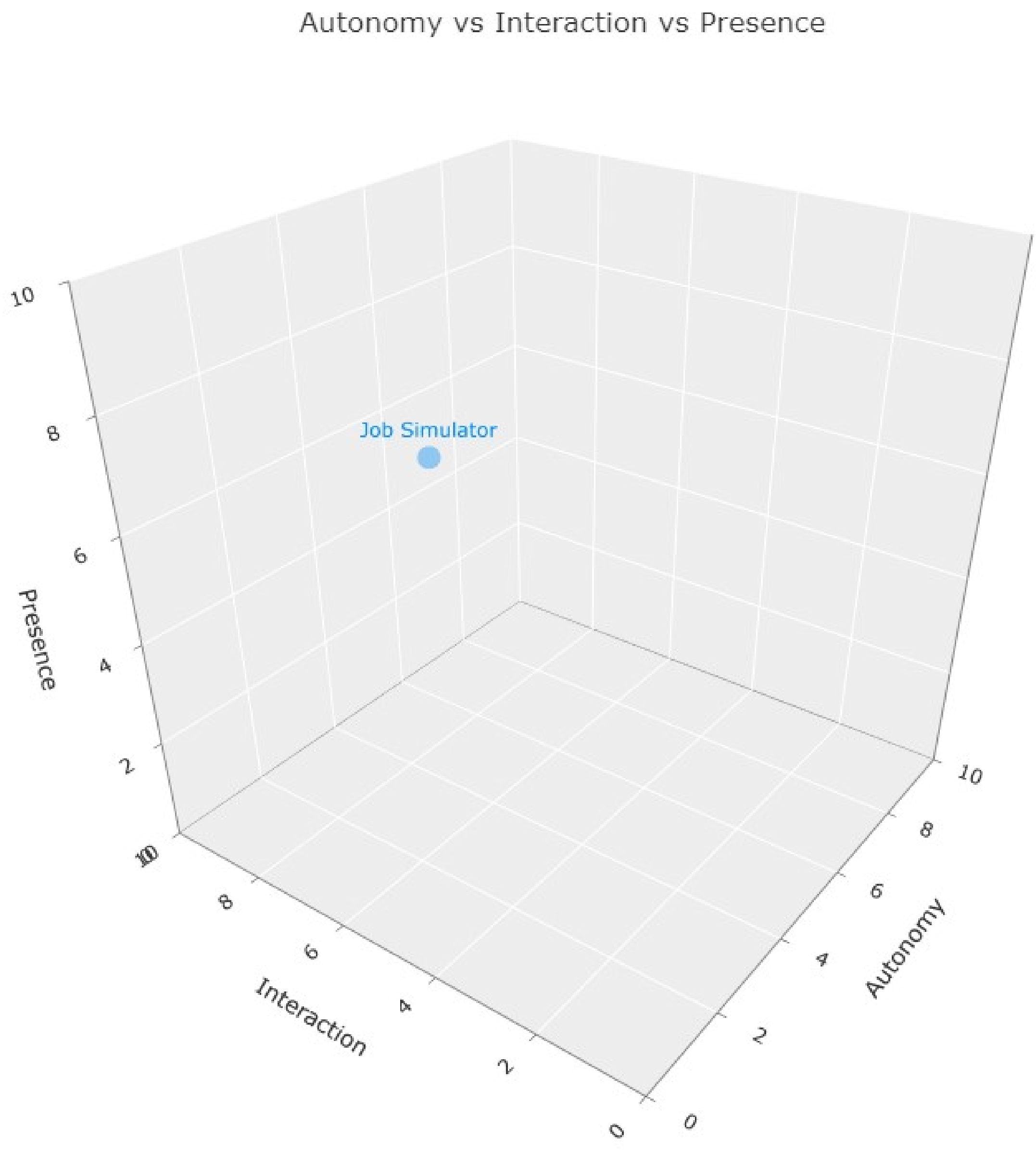


FIGURE D.7: AIP cube filtered to show experiences which belong to the *slice of life* genre.

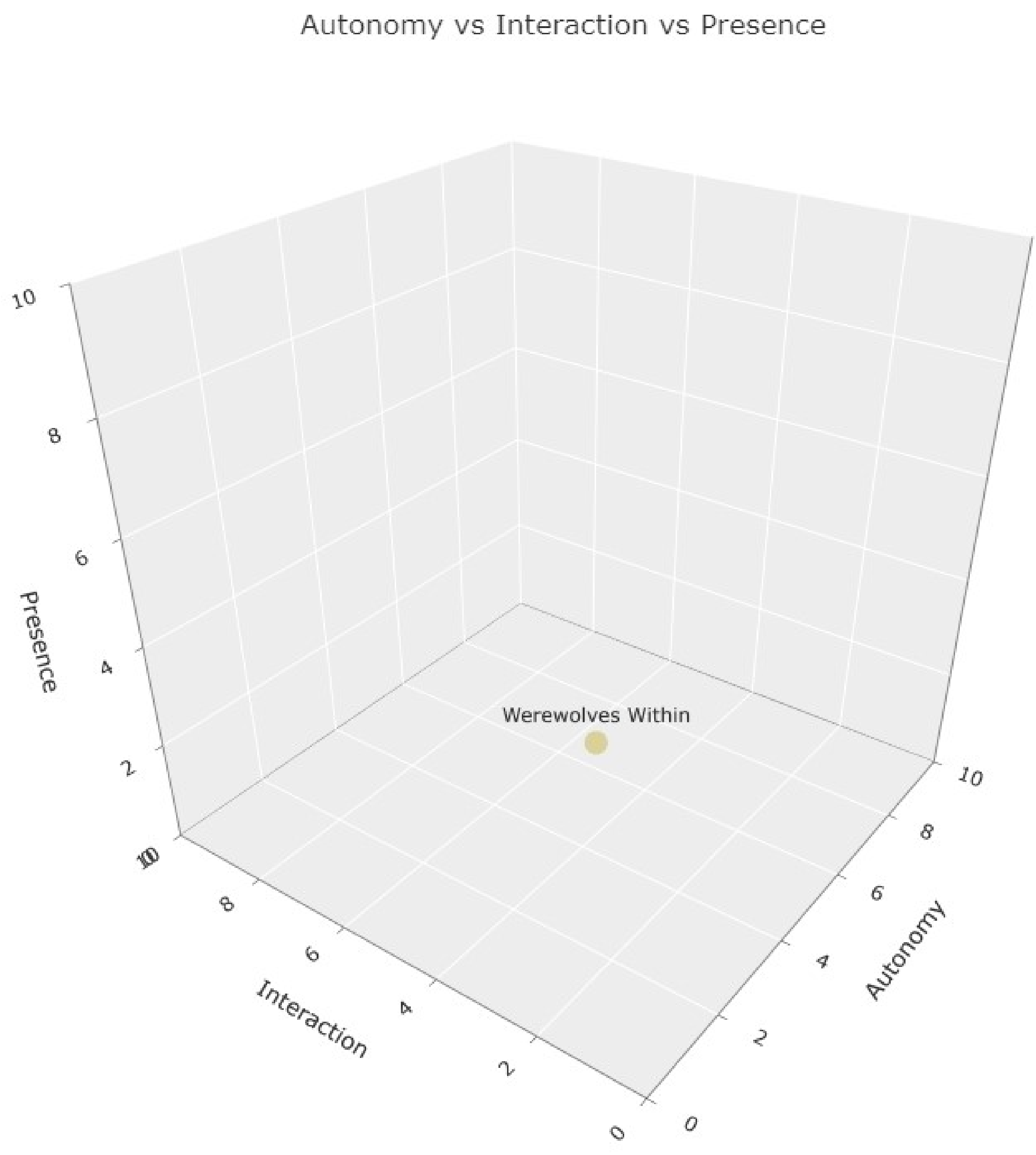


FIGURE D.8: AIP cube filtered to show experiences which belong to the *board-game/social* genre.

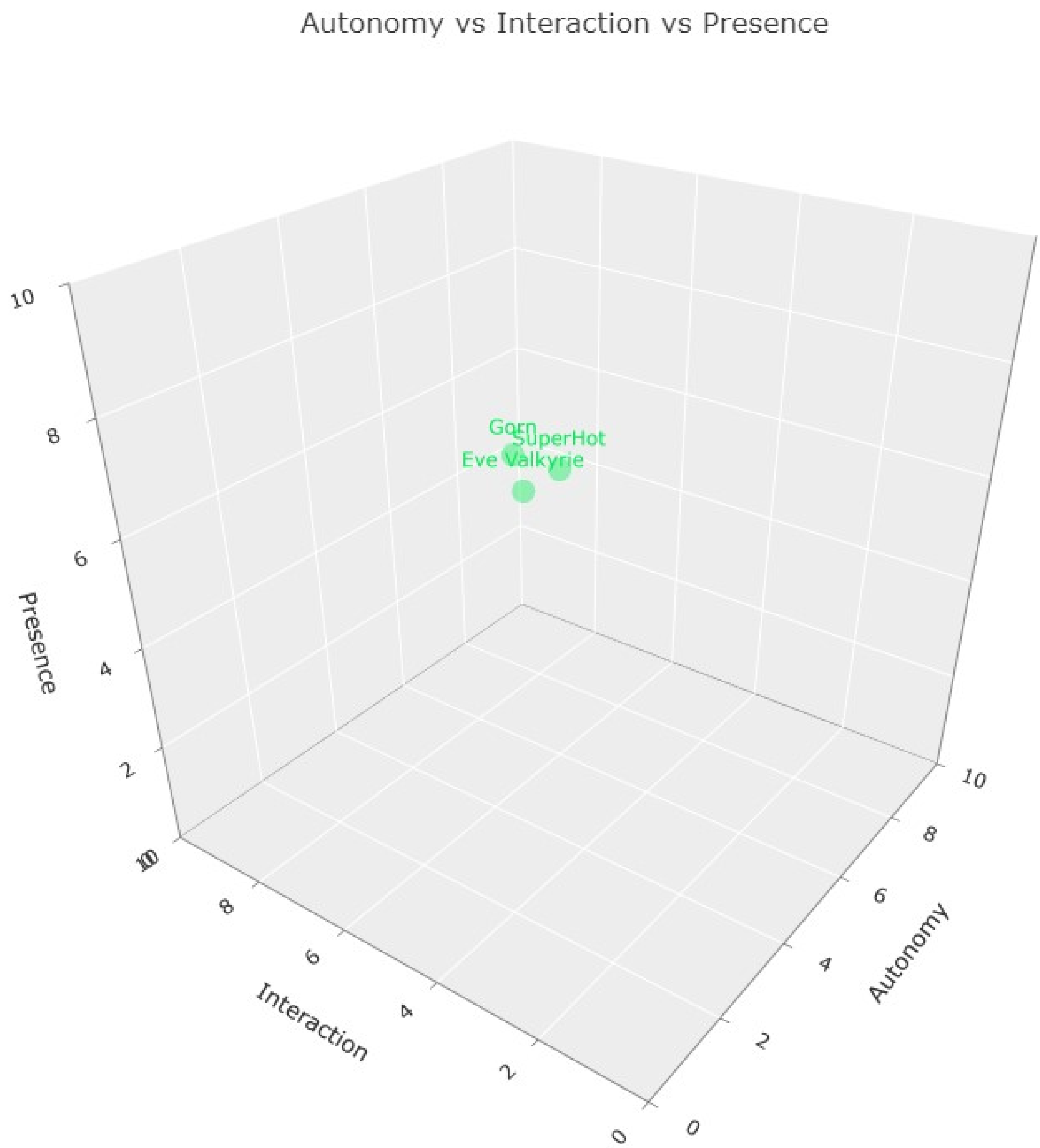


FIGURE D.9: AIP cube filtered to show experiences which belong to the *action* genre.

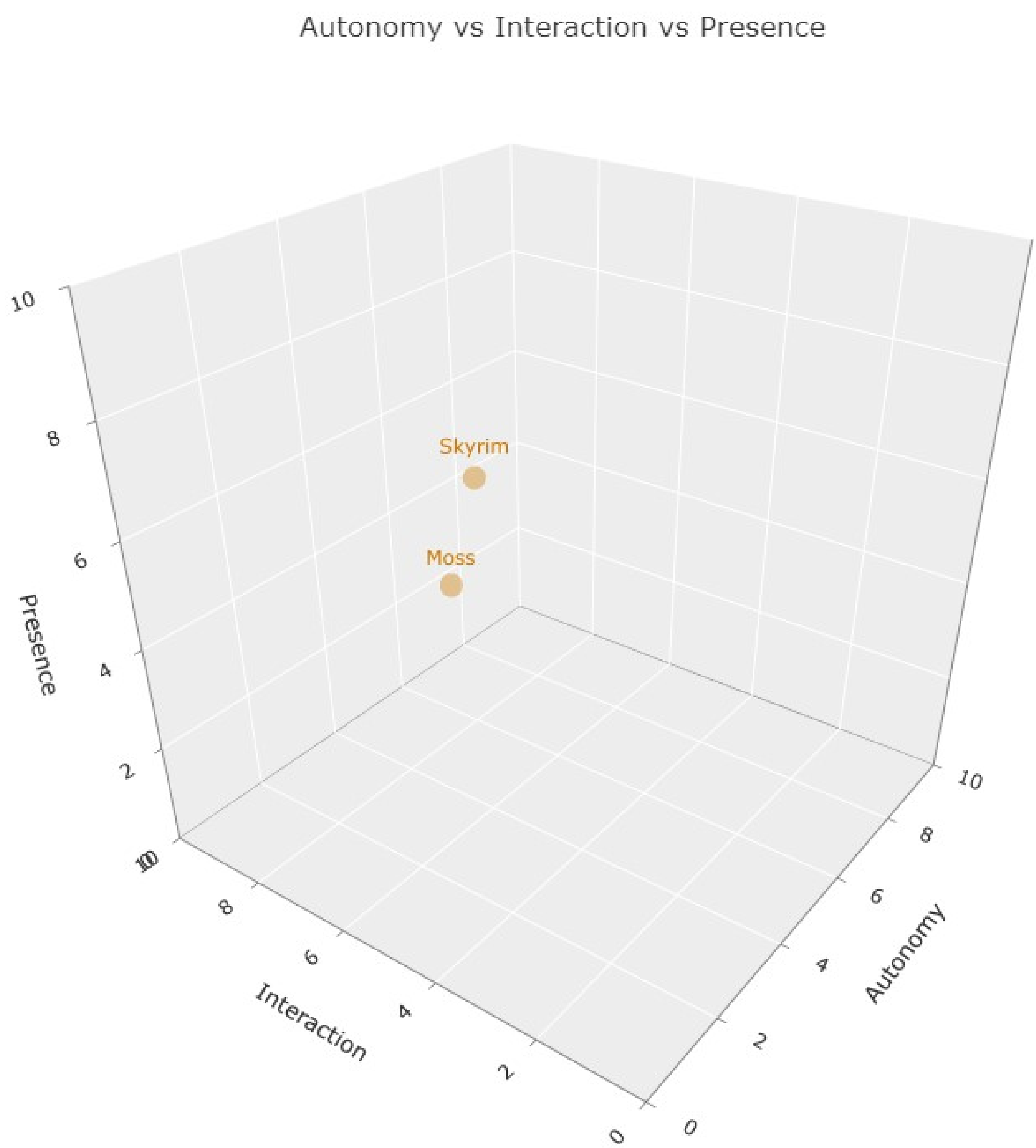


FIGURE D.10: AIP cube filtered to show experiences which belong to the *fantasy* genre.