# OPTMIZING VIRTUAL REALITY MULTI-CHARACTER EXPERIENCES USING AFFECTIVE RATINGS

by

**Angshuman Mazumdar**

**A Thesis**

*Submitted to the Faculty of Purdue University*

*In Partial Fulfillment of the Requirements for the degree of*

**Master of Science**

Department of Computer Graphics Technology

West Lafayette, Indiana

May 2021

**THE PURDUE UNIVERSITY GRADUATE SCHOOL**
**STATEMENT OF COMMITTEE APPROVAL**

**Dr. Christos Mousas, Chair**
Department of Computer Graphics Technology

**Dr. Nicoletta Adamo**
Department of Computer Graphics Technology

**Dr. Jorge Dorribo Camba**
Department of Computer Graphics Technology

**Approved by:**
Dr. Nicoletta Adamo

# ACKNOWLEDGMENTS

I would like to thank my graduate advisor, Dr. Christos Mousas, for guiding me through this academic journey, and providing me with valuable knowledge, guidance, and support during my tenure as a Masters Student. I would also like to thank my thesis committee (Dr. Mousas, Dr. Camba, and Prof. Adamo) for providing valuable input during the development of this project.

I would also like to thank my Professors in the Department of Computer Graphics Technology and Purdue University, for helping me grow, both personally and academically, and guide me on the path of becoming a better person and researcher.

I am also hugely indebted to Department of Computer Graphics Technology for providing me with an opportunity for a Teaching Assistantship, which allowed me to explore and add a new dimension to my personal growth. Finally, a big thank you to Purdue University for providing me this platform to do what I love and helping me stay on this path to becoming a better human being.

**Boiler Up!**

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# ABSTRACT

The thesis deals with the study of how virtual multi-character scenarios (primarily crowds) can be synthesized with specific behaviors, so as to induce a negative affect in the user. Virtual crowds are inclined towards being a passive world building factor, rather than a gameplay affecting factor. The study focuses on one main research question: “Is it possible to synthesize a multi-character experience that induce a certain amount of negative affect to participants?” Through the study, the emphasis lies on being able to drive emotions in an effective way, when creating multi-character scenes that need to give off a specific mood or emotion and provide an insight into how the behavior of the collective is able to affect a user’s mindset. The pipeline’s development involved developing a dataset of behaviors to be assigned to the virtual characters. Next an annotation phase assigned the affective scores to the virtual behaviors (34 in total), which (along with several design parameters) were then considered for the total cost of a scenario with a multi-character setup. Using a Markov chain Monte Carlo technique known as Simulated Annealing, the scenes were optimized towards target values of negative affect (namely low, medium, and high target affects). Finally, through the implementation of a user study, the algorithm was validated on synthesizing these targeted affect-driven multi-character virtual reality experiences. The results indicated that the three synthesized experiences (low, medium, and high negative affects) were perceived as expected by participants. Thus, the study concluded by stating that affect-driven multi-character virtual reality experiences can be automatically synthesized in such a way that impacts a user’s affect levels in the way that is expected.

# CHAPTER 1. INTRODUCTION, PROBLEM AND PURPOSE

## 1.1 Introduction

Crowds form an important aspect of our daily lives. In the pre-COVID-19 pandemic period (before the year 2020), it was an overlooked aspect of our daily lives, that humans used to keep in mind, in the passive. Human beings are social animals, and it is this social relationship that makes interaction with each other a key aspect of our lives (McAndrew, 1993). Articles have been written about places where there is no general human presence (but would have been at some point of time) people always associate having negative feelings about the place (McFadden, 2019). Studies often talk about "agent detection mechanisms"—which simply put is an evolutionary process in humans that protects us from harmful elements that may be present at a location (may be a predator or enemy) (Wurm et al., 2018; Maij et al., 2019). This is crucial to understand as many studies show that in many scenarios that affect our personal space (McAndrew, 1993) and our surroundings (Fisher et al., 1992), such as one being amidst a crowd of people (whose emotional quotient and behavior starts of undetermined), where there may be an elevated presence of risk factors according to the human psyche. Thus, when one talks about doing a crowd-based study, or a crowd immersion (in any field, be it virtual reality or other), it is often important to understand what are the important aspects that allow people to be present in a crowd and also the psychological components that work in tandem in the background, to decide what factors go into a person when they are understanding the dynamics as well as the emotional quotient of a crowd, whilst being a present in it.

It is a small portion of this aspect of the human mind that the research attempts to explore into. One of the major end goals of the research, is to understand whether conditions familiar to real-world crowds can be replicated in a virtual environment, so as to invoke an emotional response in a user, inside virtual reality (VR) scenario. The virtual reality-based applications mostly deal with single user point-of-view (POV) during the simulation. These initial simulations such as “Beat Saber,” “Robo Recall,” etc. can be taken up as examples where the focus of the game is entities keep approaching towards the user in the form of waves—thus even though there are a lot of entities present (which is essentially what a crowd simulation would be at its core), they do not display crowd-like behavior, thus not making them a good representation of how users would react to virtual crowds. With the advancement of technology in the field of virtual reality, collaborative scenarios started to emerge with applications such as “Rec Room,” “Star Trek: Bridge Crew,” “World Viz,” etc. However, a majority of these simulations focused on the multiple users being displayed as “floating avatars” and heavily relied on the aspect of each individual avatar being a representation of the actual user, so that the concept of “embodiment” and “presence” (Krogmeier et al., 2018; Porras et al., 2019) can be addressed in the simulation—thus making it ineffective to be considered as a true “crowd” scenario. Thus, there does seem to be a lack in virtual reality scenarios that try to incorporate lifelike crowds, and also there exists a significant gap in research as to the composition of these crowds as well as the response that users have to them.

## 1.2 Problem

This presents a problem when heading into the future of virtual reality-based applications. When presented with the floating avatars, it was found that in small numbers it is effective, but when considered in large numbers, there exists certain problems related to immersion and presence inside the virtual environment, as well as also affecting the way the users move (Koilias et al., 2020). Koilias et al. (2020) found that there was a significant difference when users were put in a virtual crowd, thus moderately associating users' movement with the simulated characters, thus signifying that when put in an actual crowd scenario, there does seem to be some effect on the users' experience.

Another significant research gap exists in the fact that most of the real-world based crowd simulations inside virtual reality have some sort of emphasis on disaster response or crowd dispersion situations (Xu et al., 2019). Thus, a regular crowd study, that does not focus on the user being under duress is a topic that is yet to be explored. Crowd emotion (Carretero et al., 2014; Durupinar et al., 2015) as well as crowd etiquette (Lee et al., 2013), and its role in designing virtual lifelike crowds, is a relatively unexplored area.

This creates a significant gap when further studying emotional quotient of a crowd as a whole, and how emotions of a user can be swayed through the use of a crowded environment. The entities in the crowd and their behavior may be used to give off a specific affect over the users' perception of the crowd, as a whole. Thus, the problem this study focuses on is to address the gap that exists when considering user's avatar relationships with other entities, inside a simulated virtual environment. It attempts to shed light on the parameters and affects that form the basis for driving emotions inside a multi-character virtual environment, and how the fundamental concepts of "immersion" and "presence" can be bolstered when building virtual reality environment.

## 1.3 Purpose

This project explores synthesizing a specific affect in the minds of users—in this context the research focuses on the negative affect—through the use of a computer synthesized and optimized crowd scenario, where the there is a specified target affect that needs to be achieved through the different behaviors displayed by the agents in the crowd. Due to the nature of a majority of virtual reality simulations being single user focused, it is imperative that one studies the emotional responses that users display towards the negative affect, when put in a life-like virtual crowd, and also the factors needed to drive the affect to varying intensities.

For these reasons, this study proposes that users be exposed to an optimized crowd that displays different behavior that has a certain amount of effect towards a target affect (in the study, the chosen affect can be classified to be negative in nature) and try to elicit an emotion off of the user. The research’s primary focus relies on a synthesized crowd being forced to interact with the presence of the user in the environment, in a certain way, through various behaviors that relates to the immersion aspect of a virtual reality simulation. These behavioral cues range from a combination of non-verbal cues, body language, as well as motion of their movement. There is significant research that points towards the fact that when in a crowd, people usually focus on those cues, to activate their agent detection mechanisms, that makes humans vigilant in crowded scenarios (Colombi et al., 2015; Kapadia et al., 2015). Through this, there is an attempt to understand whether there can be an automatic generation of a multi-character environment in virtual reality, that can induce a certain positive or negative affect to the user.

The purpose of this study is to expose participants to a life-like crowd with certain behavioral parameters, and optimize it for displaying negative affect, so as to elicit an emotional response from the users.

The study aims to answer the following research questions:

- **RQ1:** "Is it possible to synthesize multi-character experiences that can induce a certain amount of negative affect, based on a provided level of target negative affect, to participants?"

The deliverable expected from this project is a simulation that takes the user through a virtual crowd that displays various behavioral traits.

### 1.4 Significance of Problem and Purpose

The significance of the problem can be seen in multi-character based virtual reality scenarios. It is essential to understand how multiple characters being in the same virtual environment brings about a change in the user's perception of the environment, the gameplay, the scene, etc. (Huang & Wong, 2018). Understanding emotional response towards crowds is vital, especially in virtual reality scenarios, due to the concepts such as presence and immersion being fundamental concepts of developing virtual reality scenarios. These concepts deal with the belief that the user is in the environment, and definitely play a role in delivering a better experience (Evans & Rzeszewski, 2020). Without the understanding of how multiple characters interact with the user, there will exist this void of information about the different elements of simulation design, which may degrade the quality of future simulations (and in turn future research) by introducing unwanted variables during the simulation.

The significance of this study is that it will be able to provide an insight into impacting future multi-character scenario synthesis inside virtual environments, and also while building up on previous crowd-based studies that deal with crowd composition (Nelson et al., 2019) and crowd density (Koilias et al., 2020)—now there is another dimension that gets added to it, in the form of the emotional response displayed by the user to crowd comprised of characters with annotated behaviors. Another significance of the simulation is that it can further be expanded upon by adding more parameters, as well changing the experience mode of the user, thus also providing as a framework platform for future expanded multi-character setup (crowd focused) studies, synthesis of multi-character narratives, as well as affect based interaction scenarios among a virtual crowd. Since many of the current crowd-based studies are mainly focused on disaster mitigation and hazard scenarios (Xu et al., 2019), it could prove to be a valuable resource when synthesizing certain behavior specific crowds within those management scenarios.

## 1.5 Definitions and Terms

1. Affect: It can be defined as "Emotion or desire, especially as influencing behavior or action." in the context of psychology. It generally deals with the experience of feeling emotions. ("Affect", 2020)
2. Proxemics: It can be defined as "The branch of knowledge that deals with the amount of space that people feel it necessary to set between themselves and others." ("Proxemics", 2020)
3. VR: Virtual Reality
4. POV: Point of View
5. T: Denotes Mean Affect Target (used in Target Affect Cost Term $C_T$)
6. V: Denotes Target Variance (used in Target Affect Variance Cost Term $C_V$)
7. User avoidance: (In the project's context) When a character in a crowd walks towards the user avatar, fairly along the same path. Motion can be bi-directional.
8. Crossing avoidance: (In the project's context) When a character in a crowd walks across the user's virtual avatar's path, crossing their path in the field of view of the user. Motion can be bi-directional.

## 1.6 Delimitations

There are several limitations that have been put in place so as to make the study feasible to implement as well as record observations:

1. The study only takes the concept of negative affect being produced in the minds of user. All the simulations would be geared towards recording their level of negative affect perceived from the optimized crowd.
2. When undertaking a multi-character virtual reality scenario, users usually have the freedom to move about and interact with other characters, however this study has limited the freedom the user is allowed, so as to build a more focused experience around the user, mainly for experimental purposes. The user does not have to move or do anything interactive (for the scope of this study) and is merely an observer in the experience.
3. Facial emotions have been kept neutral for all the characters in the crowd, as adding those in would mean that there would have been extra variables that needed to be considered, thus increasing the complexity of the simulation.
4. The inclusion of sounds and other audio has been knowingly omitted, to reduce the effect of unaccounted variables on the rating of the behaviors.
5. The developed dataset involves only some of the basic behaviors based on previous knowledge (discussed in Section 2) thus, not all real-world behaviors can be accounted for in the virtual crowd.

## 1.7 Limitations

There are also some limitations that the study has to consider:

1. The study does not account for all the variables involved when being present inside a crowd. These can range from the users' own feelings about presence in a crowd.
2. The study also makes use of a questionnaire for assessment and hence may be opinionated based on the background of the participant.
3. The artificial nature of the experiment definitely does not mimic real life. Thus, there exists certain limitations that creates a low ecological validity, making it difficult to apply the complete findings in a real-world setting.

## 1.8 Assumptions

The study assumes that the participants have had some sort of virtual reality experience and have had an experience of crowd-based scenarios in real life. Also, a Likert scale-based data measure assumes that the strength or intensity of an emotion is linear (from strongly disagree to strongly agree) and assumes that attitudes can be measured.

# CHAPTER 2. REVIEW OF LITERATURE

## 2.1 Search Methodology

The problem at its current stage is to understand the gap that exists between traditional methods of crowd simulation and procedurally generated crowd simulation technique and define some parameters for generating a legible virtual crowd. There are two major components that have been identified for addressing that gap:

1. Macroscopic factors: The key components for developing a virtual crowd, and the parameters that affect the synthesis and behavior of a virtual crowd.
2. Microscopic factors: The composition; how and what the behaviors of each of the characters should be, and what behaviors produce the most significant affect in the user.

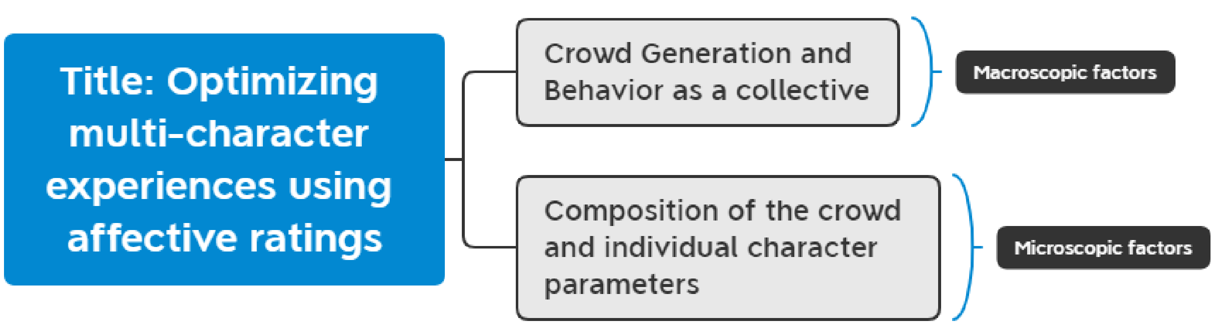



Figure 1. Major components of a virtually generated crowd.

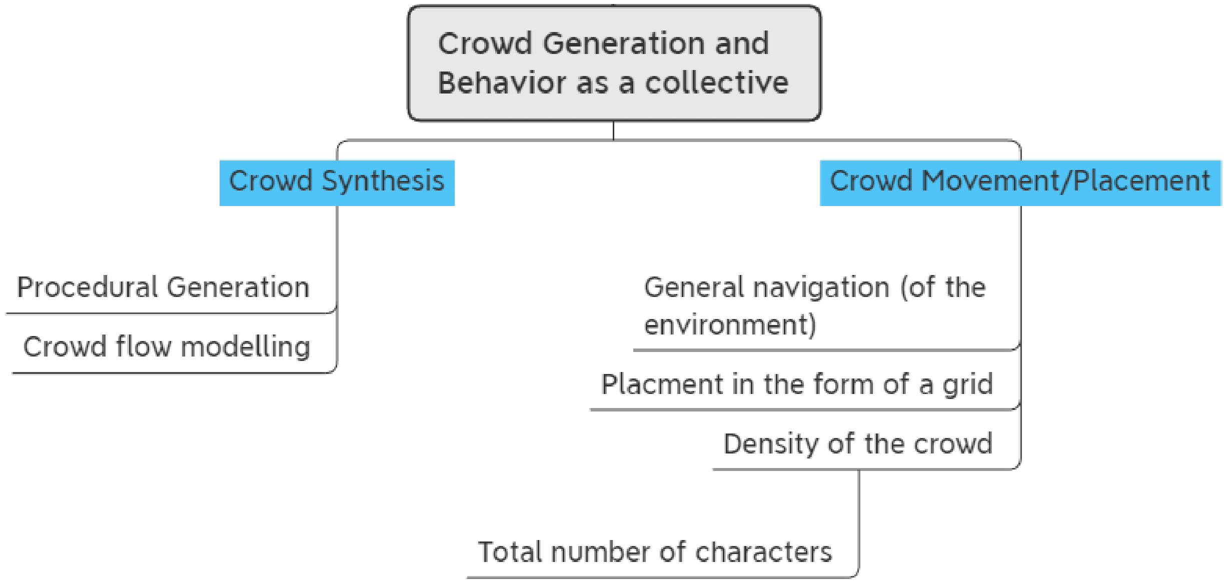



Figure 2. Macroscopic factors' composition.

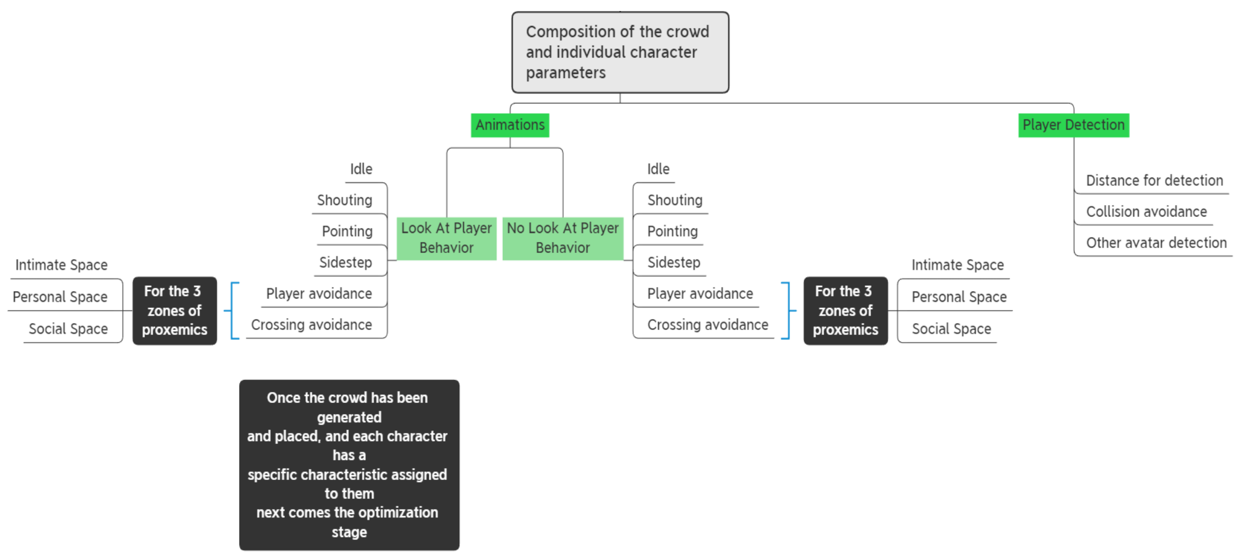



Figure 3. Microscopic factors' composition.

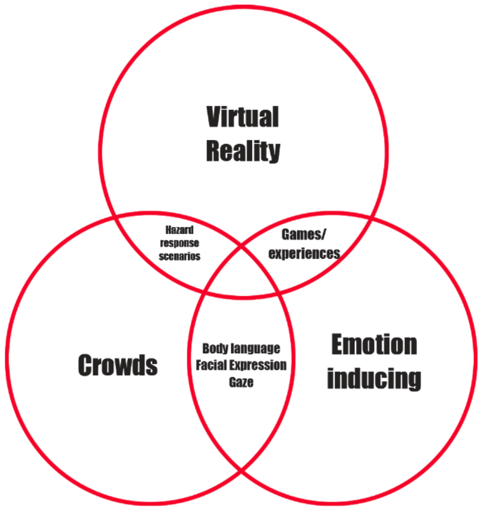



Figure 4. Venn Diagram showing the focus of the study.

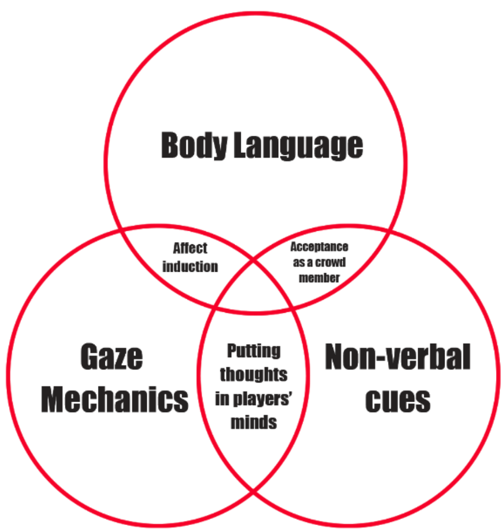



Figure 5. Venn diagram showing the crowd behavior that incudes affect.

**Search strategy:** Most of the terms that will be used revolve around three major concepts: crowds, virtual reality, and affect synthesis and response. Figure 4 and Figure 5 show the relationship between three major concepts. The strategy used was a reverse pyramidal structure, where the common base for the project was identified and then the search words, and search terms for each concept was narrowed down. This is how the searches refined the project topic:

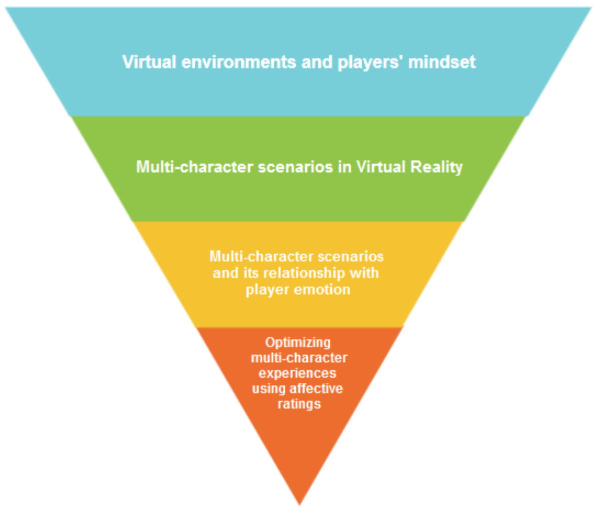



Figure 6. Refining the search strategy.

The library databases that have been identified to have important information are:

1. IEEE Xplore (For the technical implementation of crowd synthesis)
2. ACM Digital Library (For the psychological, and technical aspects of crowd building and user immersion)
3. SAGE Research Methods (For animation techniques and life-like behavior display)
4. GDC Vault (A mix of crowd perception by players, multi-character simulations, and visual demonstrations)

5. Also for searching dissertations ProQuest Dissertations & Thesis Global (PQDT) was used

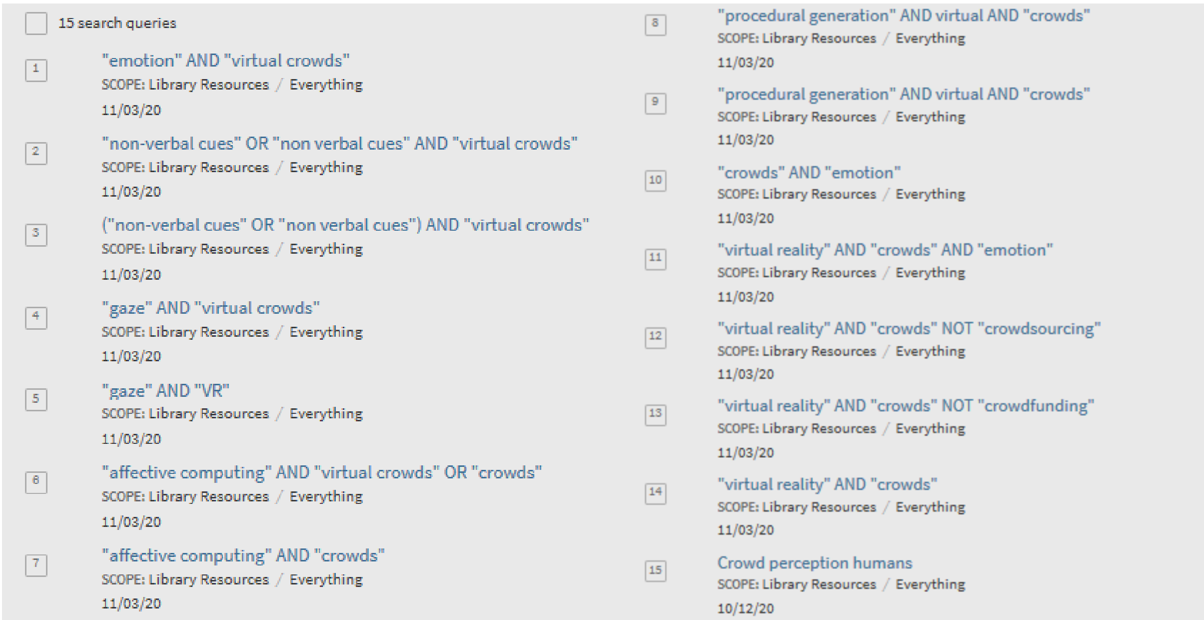



Figure 7. Search History in Purdue Libraries Site.

Table 1. Overview of search results.

| Search Term Number | Total Results | Total Usable |
|---|---|---|
| 1 | 193 | 2 |
| 2 | 15803 | 0 |
| 3 | 0 | 0 |
| 4 | 88 | 2 |
| 5 | 15139 | 4 |
| 6 | 233375 | 4 |
| 7 | 1204 | 3 |
| 8 | 63 | 3 |
| 9 | 61 | 3 |
| 10 | 43135 | 6 |
| 11 | 756 | 4 |
| 12 | 2599 | 2 |
| 13 | 2592 | 2 |
| 14 | 2800 | 4 |
| 15 | 55397 | 3 |

The search methodology has been broken down into two segments: searches related to the simulation and technical aspects of simulation of multi-character scenarios, and searches related to the psychological elements that goes into understanding a person when in a crowd scenario (either real life of virtual). The searches primarily consisted of using databases such as IEEE Xplore, ACM Digital Library, SAGE Research Methods, and GDC Vault. Some notable journals and conferences include Computer Animation and Virtual Worlds (Wiley), Animation Journal (SAGE), ACM CHI and IEEE VR conferences. Some of the tools used for searching for publications are the Purdue Library website, APA PsychNet website, and Google Scholar. For searching for Dissertations, ProQuest Dissertations & Thesis Global (PQDT) was used as well as the direct-to-document links provided by the Purdue Libraries search engine. Keywords used for many of the searches include (but are not limited to) “virtual reality” AND “crowds” NOT “crowdfunding,” “virtual reality” AND “crowd” AND “emotion,” “virtual crowd” AND “emotions,” “procedural generation” AND “virtual crowds,” “group behavior” AND “virtual crowd” OR “crowds,” “affective computing,” or “non-verbal cues” AND “virtual crowds.”

## 2.2 Real World Crowd Dynamics

A lot of the basis of real-world crowd interaction is based on the concept of “Interpersonal Communication”. These individual relationships, withing the population, encapsulates a lot of non-verbal communication cues, eye-gaze, proxemics, etc. (Bailenson et al., 2003). For example, Hall et al. (1968) talks about the importance of space and how it is deeply ingrained in our understanding of patterns in the physical world (along with the cultural background). The “Five-Factor Model of Personality” or more commonly referred to as the OCEAN model (McRae & Costa Jr., 1996) provides an insight into how a person can evaluate another person in a collective using the traits of “openness, conscientiousness, extroversion, agreeableness and neuroticism” (Durupinar et al., 2008). These formulated the basic hypothesis of whether the same rules of perception would apply to certain virtual populations that were generated to induce a certain affect on a user.

## 2.3 Crowd Simulation

Simulating crowds has always been a highly lucrative concept in the field of interactive media and virtual simulations. Historically, crowd simulations were mainly associated with evacuation scenarios and disaster management protocol testing, as illustrated by Xu et al. (2019), Xu et al. (2014), Stamatopoulou et al. (2012), Helbing et al. (2000), etc.

Previous findings shed light on crowd movements in multi-hazard scenarios (Xu et al., 2019; Moussaïd et al., 2016). The study they conducted tried to invoke panic-based emotions in participants through the use of multi-character scenarios and study the response showed by the users. The study was primarily targeted towards generating local avoidance only for simulating panic emotion dynamics of a user in a crowd. Crowd avoidance and dispersive movement seems to be one of the more highlighted features for these types of studies, as also illustrated by Lin et al. (2020) whose findings mainly dealt with the fact that there does not exist any sort of impact of cultural elements of a user, when considering whether or not to follow a crowd. Thus this was one of the key elements in formulation of this study, where the idea was to understand the extent of how much the affect generated by a virtual crowd holds influence over the mindset of a user, and to what extent it can be manipulated.

The first step to understanding how a crowd's behavior can induce an affect in a user was to understand how to simulate a crowd, and the factors that need to be kept in mind when undertaking the synthesis of these multi-character scenarios. Xu et al., 2014 gave a comprehensive idea about simulation of crowds. The study talked about surveying state-of-the-art crowd simulation techniques, their applications, and recent advances and divided the applications into two types: evacuation and training simulations and film productions and video games. They stated that in training simulations visualization is not the primary focus, hence it usually ends up being simple and 2-dimensional. However, in relation to video games, high quality rendering techniques and high-quality animation asset and techniques, for believability, need to be applied. This is where the focus of this study lies—synthesizing a life like crowd that induces a certain affect in the user, and for that an important concept to understand how modelling of natural crowd movement works. The defined crowd simulation techniques are divided into two types: Macroscopic and

Microscopic. Macroscopic deals with crowd systems' simulation as a whole entity and follows the characteristics of flow (making the overall crowd movement look realistic) as well as the interaction of the crowd with the surrounding environment as a collective. In contrast, microscopic deals with the individual behaviors of the elements in a crowd, and their interactions. This study will deal with a hybrid crowd where the microscopic model (i.e. behavior of individual characters in the crowd) will drive the overall macroscopic characteristics (the overall affect of the crowd on the user). Previous studies (Xu et al., 2014; Pelechano et al., 2008) also stated the fundamentals when considering building a crowd simulation—navigation, parallelizing simulations, panic phenomenon, and evacuation system. However, even though these are the fundamental pillars for crowd simulation, it is evident that they are designed, keeping in mind a form of evacuation-based application. Thus, there definitely exists a need for more common generalizable strategies, that can be implemented to crowd synthesis and simulation.

Previous studies (Xu et al., 2014; Paris et al., 2007) also provides deep insight into developing simulations for pedestrian crowds. The authors state the following as major points to consider when developing pedestrian based crowds (which is the focus of both studies): simulations based on vector fields, diversification of motion styles, and perceiving motion transitions. One of these concepts, the diversification of motion, is essential when synthesizing human crowds. The human eye tends to perceive small differences in motions pretty well (Snowden & Freeman, 2004;Vater et al., 2020), and thus it is highly important to take into account the different motion (i.e. the behavior) that each character displays. This is why while calculating the cost term of the overall scene, there is the inclusion of cost terms for variation in behavior for each character as well as adjacent characters. Their research also provides insight that distant viewpoints helped mask unrealistic motion of characters, as well as increasing density helped in hiding motion transition. Thus, the inclusion of a density cost term as well. In terms of density, the study conducted by Dickinson et al. (2018) stated that high density virtual reality scenarios were perceived as uncomfortable by the participants. However, this study is not enough to substantiate major claims, as the researchers claim that the negative perception could have been the cause of agent proximity and "behavioral artifacts of the simulation model."

Most of the crowd simulation is generally thought of as a simulation of a swarm. That is something that this study passively aims to shed light on. Albi et al. (2019) in their research presented an analysis on modelling the dynamics of human crowds (along with vehicular traffic and other swarms) mathematically, with the properties of a swarm. While treating human crowds as a swarm can have its benefits, it also lacks building up on the factors of the individual characters that affect the realism of the virtual crowd, as well as understanding how the emotional spectrum (in general) can be measured.

Simulation of crowds thus needs to be seen apart from swarm simulation, and more as a collective of individuals. It may seem to be quite complex creating a unique personality for each character in a multi-character simulation, however, for the benefit of expanding upon the current research of crowd synthesis, it is essential for this character development feature to be considered. Thus, this is where "procedural generation" comes in.

As evident from studies such as Albi et al. (2019) and Colombi et al. (2016) crowds have been generally formulated as a mathematical model and the simulations are built around them. Also, when generating crowds, the elements tend to already have their characteristics defined and/or pre-determined (Lin et al., 2020 and Xu et al., 2019). This can result in situations where the simulation demands for the crowd's affect to be of one type towards the user, however due to the individual characteristics displayed, the emotions that the user experiences may not always be close to the targeted intensity. That is the reason it is important to understand that in virtual crowd-based scenarios, one of the major components to be considered is the affect that can induce a certain type of emotion, and procedurally generated characters and crowds, whose parameters can be adjusted based on the required affect intensity, can help achieve the same. Refer to Section 2.5 for more details on optimization methods being used for simulations.

## 2.4 Interaction with Virtual Crowds

After the development of a crowd in a virtual scenario, the next question to ask is how to enable users to interact with the crowd. The term "interaction" is being used loosely here. Interaction can refer to anything from the user engaging actively with the characters (such as pushing them away, talking to them, etc.) or a passive interaction where the presence of the crowd affects the experience of the user (such as producing a specific emotion in the user's mind). This study focuses on the more passive aspect of interaction.

Other research (Mossberg et al., 2020; Kyriakou et al., 2016; Durupinar et al., 2016) has examined the various characteristics of a virtual simulated crowd and how it can affect a user's experience in a VR environment. Kyriakou et al. (2016) set up three scenes in their experiment, each with varying levels of crowd interaction. In their results, they found that when the realism factor of the crowd increased i.e. the level of interactivity increased, there was a higher sense of realism and higher level of presence in the users. The main factor was the collision avoidance (as evident from the changes in the first and second scenes) according to the researchers, that heightened the realism of the scenes, with other effects only adding more to the realism factor. The theory of collision avoidance is also backed up by the findings of Sohre et al. (2017) in which they found that the users reported significant changes in reactiveness (increased), intimidation (decreased), and human-like-characteristics (increased) when there was collision avoidance behavior i.e. when the characters moved out of the way when the user was coming it increased the realism factor of the simulation. This provided good insight into a couple of parameters that are perceived important for realistic crowd-based scenarios and showed how introducing just one parameter could affect the perception levels of the user in such a massive way, and also one of the major reasons for this study incorporating a type of collision avoidance in the characteristics of the crowd.

The studies conducted by Kyriakou et al. (2016) and Pettré et al. (2009) did shed new light on crowd simulations however an aspect that they missed out on was the emotional composition of the crowd. Most of the changes in the scenes in the aforementioned study involved changing various physical parameters that could directly be observed by the user or where the user could be put directly in a situation where they had to interact with a character. It did leave a significant gap when it came to using non-verbal cues, body language, as well as an overall affect inducing crowd which could bring about an affect response in the mind of the user, without the user being too involved in the scene.

Volonte et al. (2020) built upon the foundations of crowd simulation and introduced emotional characteristics into the elements of the crowd, which directly affected the user in the simulation. The user was put into a virtual market scene where they had to complete unique tasks which involved interacting with the crowd, who displayed various verbal and nonverbal characteristics. Their results found partial support with their hypothesis. They hypothesized that the "Users will report a similar emotional affection to the condition they experience due to emotional contagion effect." Their results showed that users were not very intensely and emotionally affected by the emotions displayed by the virtual humans. Their research also showed that in the positive emotional spectrum of the simulation the users did experience less negative emotion, thus the point of their partial support of the hypothesis.

Another hypothesis the researchers made, was that the users would interact more frequently in a more positive variant of the crowd, rather than a negative one. They concluded that the results, based on interaction times and number of interactions did support their hypothesis, and thus a positive environment does bring out more engagement in users. Finally, in their third hypothesis, they concluded that in the more positive variant of the crowd, the users observed the body language as well as the facial expressions more, than in the negative variant of the crowd (in the negative variant, the users did not focus a lot on the faces of the characters because of the negative facial expressions).

New research in the field of crowd simulation (Volonte et al., 2020) has also provided significant background on the perception of emotion in crowd-based scenarios, that built up on previous works such as Kyriakou et al. (2016) and Pettré et al. (2009). However, an aspect that seems to be missing is inducing a targeted emotion in the mind of the user, without actively involving the user, and letting the crowd dictate the mindset of the user. Another shortcoming of the scenario could be the fact that when the users are fully immersed in doing a task actively, it may not allow them to fully focus on the non-verbal cues that the characters in the crowd displayed. This is what this study will try to contribute towards i.e. inducing a targeted affect (negative) in the minds of the user, without including an active interaction system. This will enable users to be more vigilant about the characteristics of the crowd, rather than having to focus on multiple things at a time.

Another aspect is the affect gaze produced on participants in a study conducted by Mousas et al. (2019), where they investigated the movement of users in a virtual environment and they attempted to avoid a virtual character, that was a representation of the users' self-avatar (basically the same gender as that of the avatar selected by the user). Their results concluded that the length, duration of the task, and deviation from the straight line's trajectory greatly increased when the self-avatar represented the participants. They also found that when the character that the user had to avoid had the look-at condition on (i.e. their gaze was focused on the user), the users showed significant deviation from the straight path, as well as the gap between them and the character when adjacent to each other. This is one of the reasons for including gaze (i.e. look-at and no look-at) as one of the characteristics in the behavioral dataset, that the characters of the crowd will display.

Interaction with virtual crowds also involves several elements that can either be classified into character interaction or the crowd interaction as a whole. Studies such as Carretero et al. (2019) and Novelli et al., (2013) have dived into the realm of exploring crowd-based scenarios and evaluating the effect on users by the crowd, as a whole rather than focusing on individual elements of the crowd. It is important to understand why crowds need to be evaluated as an entity, rather than a collection when it comes to eliciting a response from the user. Carretero et al. (2019) carried out their study by dividing a crowd into individuals and small groups. This ensures that when the simulation occurs, the users are exposed to a more natural feeling crowd, who may display a collective emotion, thus making the impact of the displayed characteristic much more intense. Their results stated that emotion perception of these groups in the background was perceived significantly however it had no influence on the judgement of the foreground characters. They concluded by stating that the valence of characters in the background are also essential when conveying emotions in groups. Thus, what one can understand is that when considering the crowd as a whole, there needs to be special attention paid to all the characters in the crowd (as a whole) rather than just focusing on the foreground characters, to add onto the life-like nature of the crowd.

Finally, another factor that needs to be considered when talking about interaction in virtual spaces is the concept of proxemics and personal space. Personal space invasion can directly tie into the emotional response of the user, due to pre—conceived ideas of safety and threat response (Iachini et al., 2014). Previous studies have shown that users and users alike display a negative reaction to when their personal space is violated in VR scenarios (Wilcox et al., 2003). In their study, they concluded that there was an increase in avoidance behavior when virtual agents approached the users and invaded their personal space. Iachini et al. (2014) also provided insight into reach and comfort distance inside immersive virtual environments through their research. They studied two conditions for each reachability and comfort distance where either the virtual agent approached them or vice versa. In their results, they discovered that in the passive condition (where the virtual agent moved towards the user), the participants had a larger comfortability distance, which meant that they did not want the unknown virtual agent to come very close to them as they did not feel comfortable with it. In the active condition (where the user moved towards the virtual agents), they displayed much lower distance of comfort, thus implying that the user was more comfortable being in a denser crowd with a greater number of characters.

This ties in to one of the parameters that is considered for the annotation terms, where there will occur crowd agents invading the three zones of proxemics of the user (intimate, personal, and social spaces). Thus, this forms the basis of this study having a mix of "active condition" and "passive condition" (Iachini et al., 2014) based behavior in the dataset that will be used to create the characters. There are various aspects that the aforementioned paper does not touch upon—it does not consider any action that the virtual agents are doing, nor does it consider multiple agents and how those behaviors affect the user in the scenario.

## 2.5 Affective Computing Theoretical Framework

The theoretical framework this paper is based on is the topic of Affective Computing. Affective computing can be defined as an interdisciplinary field that focusses on techniques that relate to studying, understanding, and simulating human affect (Picard, 1997; Tao & Tan, 2005). *It is usually responsible for taking a deeper dive into the world of psychology and cognitive science, generally using the methodologies and techniques from the field of computer science, to process these affects.*

The study deals with the emotion classification aspect of affective computing. The emotions used for the basis of computing feature an expanded list of emotions (both positive and negative), which are proposed to convey a majority of the emotional spectrum of humans. The emotions can be listed as: “Anger, Disgust, Fear, Happiness, Sadness, Surprise” (these first six comprise the original emotions Ekman proposed), “Amusement, Contempt, Contentment, Embarrassment, Excitement, Guilt, Pride in achievement, Relief, Satisfaction, Sensory pleasure, Shame.” (Ekman, 1999).

Body gestures are also a major component of affective computing. Prior research has shown that body gestures can be effectively used to understand the psychology of a person. Carretero et. al (2014) reported findings in their study that a change in the body language of background characters did have a significant impact in the perception of emotion by the user. Another study by Carratero et al. (2014) also reported that users were able to report the perceived behaviors by the characters in a crowd as the same as the mood the researchers were aiming for while doing the study. Thus, applying it in reverse, where body language gets perceived by the user, is another factor that can help synthesized crowds control the affect levels in users. Depending on situations, the body gestures can be simple or complex. They can also be in varying intensities, which may help in creating different intensities of the affect in the simulation.

However, due to new developments being developed in the field of Human Computer Interaction, the model proposed by Picard, was contrasted with a newer model (Boehner et al., 2007) which was classified as the “interactional” model (compared to Picard’s “information” model). The new model’s idea was to aid users understanding their own emotions when placed in a simulation, rather than an algorithm or a computer just identifying the affect and emotion, and in some cases replicating human emotions, which the information model had suggested. The interactional model was much more emotion focused rather than algorithm focused, and it brought in the concept of subjective evaluation of affective experiences in humans. This introduced the concept of understanding emotions backed by the social and cultural experiences of a person, which shapes their viewpoint of the outside world and their individual role in a collective.

Thus, this forms the framework for the study. It focusses on taking, almost a reverse approach to what Picard’s research stated. The idea here is to build affect driven experiences in the realm of virtual reality and use the experiences and feelings that the users report on, to build a better framework for future simulations, as well as build a foundation for crowd-based studies that are driven by a crowd that displays a certain behavioral characteristic.

## 2.6 Synthesis and Optimization of Virtual Reality Experiences

Now that it has been established what the parameters are for the composition of a crowd, as well as what dictates affect induction in a user, it is vital to look into the process of synthesizing the crowd. Crowd synthesis can be a very unique aspect to world building. As stated in Section 2.1, most techniques that have been used until now follow swarm creation logic and mathematical models, that treat these crowds as just a lot of entities. There are more to virtual crowds than just a group of characters, rather they should be viewed as individual elements with certain characteristics that form a collective, and establish a relationship between them, the other characters, and the users.

In one study, the researchers exposed volunteers to immersive and semi-immersive VR setups, which were based in an "open-space mall" (Kyriakou et al., 2016) with a crowd and the users had a specific task where they had to follow a character, and they had three levels: (1) a low interacting crowd where they ignore the user; (2) a medium interacting crowd which avoids colliding with the user; (3) a highly interactive crowd where the characters do basic social interaction actions with the user, and avoid colliding with them. This methodology is pretty sound as it covered the three possible levels of any activity (high, medium, and low) and due to that it was able to generate a defining relationship between crowd engagement levels and the user's experience. It also helps understand between the minor nuances between two individual levels, and which of the parameters seems to affect the most, out of all the parameters that come with a crowd (such as collision avoidance). Using multiple runs for data collection and multiple scenarios has also proven to be effective from the previous studies because having repeated runs introduces less variance in the collected data (Kyriakou et al., 2016). This also points this study in the direction of having a dataset of behaviors and answers the question to why there should be a variation in crowd composition. The dataset of character behaviors ensures that there is ample amount of variation in the synthesized crowd, and also helps to drive the affect level to different intensities.

Once the crowd has been synthesized, there arises the question of the user being able to identify, understand, and validate the emotions being conveyed by the crowd. Novelli et al. (2013)

focused in their study on a more survey-based approach to understand the nuances to how people in a crowded situation classify it as inducing a positive effect. Understanding one end of the spectrum of emotion provides a baseline for what to work towards when trying to induce the other end of the emotional response spectrum (negative affect for this study). Their study ended with the conclusion that more people showed a positive reaction toward the crowded environment when they positioned themselves in a more central location of the crowd, and that once self-identifying themselves as a part of the crowd seemed to elicit a more positive reception toward the crowd.

Thus, based off of on that, to elicit a negative affect from the crowd, the study aims to imbibe the negative feelings described by Ekman (1999) by trying to make the user not feel as a part of the crowd (in a straightforward way). Since the physical location also matters, the simulation will focus on the user moving linearly and not having a stationary position (so that the users do not feel that they are the central point around which the crowd is being synthesized).

One of the major aspects of the synthesis of this study's crowd is procedurally generating levels that can adjust according to a final targeted cost term for a parameter. Xie et al. (2019) talked about procedurally generating levels designed for providing a target exercise intensity cost, and then developing a level around it. Their study incorporated various cost terms associated with the generation of a level, and that provided with a high degree of controllability to achieve the desired level of affect intensity for each scenario. In other crowd studies such as Zhang et al. (2019), procedural generation can also be seen but this time their study used it to add, modify, or remove certain poses for their character, based on the users' gaming experience or emotion, during the gameplay. Their results showed a generally positive review by the participants and proved that their methodology used for procedural generation was able to dynamically adjust parameters to develop user experience driven levels. Both the above-mentioned studies show that for a level to be highly adaptive and target driven, there needs to exist a problem statement that determines an overall cost function for the entire experience. This term comprises of various individual costs that are associated with the components that make up the level (as discussed in the previous sections).

There is, however, very low evidence of the previously discussed algorithm (Li et al., 2020; Zhang et al., 2019; Xie et al., 2019) being used to optimize virtual populations and virtual crowds. These studies have their optimization implementation on the level or environment itself, rather than the entities or the behaviors in the virtual environment, or on elements that have a direct impact in inducing a specific affect in the user. What this may imply towards, is the optimization implementation is trying to contribute passively to the user's experience inside the virtual environment. Most of them do not focus on actively trying to achieve affect manipulation. Even though there have been multiple implementations, in various domains, of the proposed optimization technique, very rarely does it deal with a virtual collective's behavior.

This study considered previously published work on human interaction with virtual characters as well as prior work on optimizing virtual reality experiences and combined such knowledge to explore whether is possible to automatically synthesize virtual reality multi-character experience that elicit certain emotional responses of participants.

# CHAPTER 3. RESEARCH METHODS AND PROCEDURES

## 3.1 Introduction

To recap, crowds form a vital part in the everyday lives of human beings. These relationships between multiple people can be established in almost all walks of life. However, the translation of effective multi-character setups, into a simulated world has not always been the most efficient. Most of the virtual reality simulations are seemingly geared towards collaborative experiences when it comes to multi-character setups. These vastly differ from general crowd-based experiences, due to the change in dynamics—these being more collaborative and mostly interactive, compared to multi-character crowd scenarios, which have higher levels of individual presence, and lower interactivity levels.

*Problem*

The problem this study focuses on is to address the gap that exists in establishing meaningful and affect driven experiences when it comes to the field of multi-character-based scenarios in virtual reality, that help bolster the concepts of user immersion and presence, as well as the design decisions they influence when synthesizing an optimized affect driven experience.

*Purpose*

The purpose of this study is to expose users to life-like multi-character scenarios (a virtual crowd with certain behavioral parameters), and optimize it for displaying negative affect, so as to elicit an emotional response from the users and validate whether affect driven multi-character scenarios can be automatically synthesized for a targeted user response.

*Significance*

The significance of the study is that it will provide ample insight into the relationships that get established when a user is exposed to a multi-character-based scenario, and how the emotions of one may drive the response from the other (in the scope of this study, only one way relationship is being explored, where the manipulation of the crowd behavior is expected to have an impact on the user's own emotions). If the dynamics between a user and a virtual crowd is established, this could lead to methods in the future, where certain emotions can be triggered just by the behavior of the crowd, thus giving a whole new dimension to narratives inside interactive media (such as video games, virtual simulations, virtual experiences etc.).

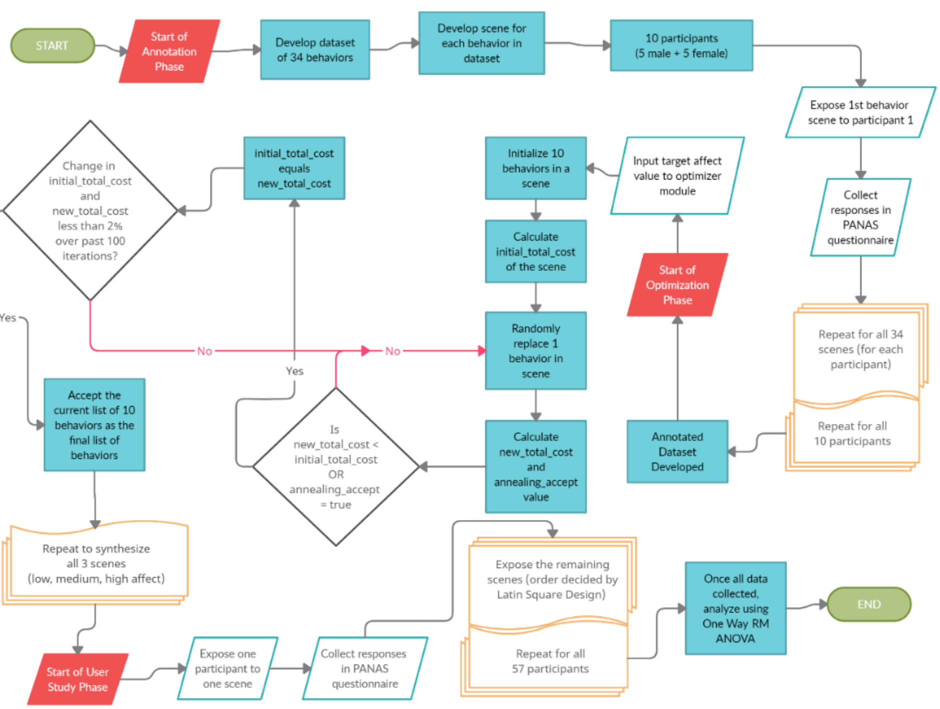



Figure 8. Experimental Design flowchart.

## 3.2 Research Type

The type of research that the study aims to perform is experimental. The experiment can be labelled as a "Lab Experiment" (Falk & Heckman, 2009). The conditions under which the experiment will be done are controlled, and that is why it can be categorized under that label. This enables the experiment to be highly replicable, due to standardized procedures, and precise control of certain circumstances (since this study deals with the psyche of users to some extent) (Camerer et al., 2016).

The research aims to establish this, by first conducting a data annotation process, where the crucial elements (relating to affect) that comprise of crowd behavior will be analyzed. Initially, a set of variables have been selected as the base parameters that affect a person's psyche amidst a crowd—number of people making direct eye contact and gaze follow (Mousas et al., 2019; Sun et al., 2017; Volonte et al., 2020), prevalence of verbal and non-verbal cues (Carretero et al., 2014), obstacles (Kyriakou et al., 2016), and entity movement (Volonte et al., 2020; Xu et al., 2019). From the initial study, the plan is to annotate the initial dataset that is comprised of various in-crowd character behaviors to see which one of the selected variables seem to have the most impact on the psyche of the user inside the crowd. After the completion of the initial annotation phase, the second phase is to use the annotated character behavior dataset to develop a scenario and optimize it according to the developer's preferences and induce the target emotion in the user (Amiri & Sekhavat, 2019). This is done by providing the simulation with a target negative affect value and based on that creating a composition of characters in the crowd with multiple different behaviors. The final purpose and aim of the annotation stage combined with the optimization stage, is to expose users to multiple levels with the target negative affect already provided (either low, medium, or high) and use the PANAS scale (Watson et al., 1988) to collect the user response for data analysis. This final stage is done in order to validate whether it is feasible to automatically synthesize a crowd-based scenario, considering that a target affect value has been provided (in this case various intensities of negative affect), and based on the generated crowd whether the targeted reaction is displayed by the user.

## 3.3 Population and Sample

The population for the experiment comprised of all the Graduate and Undergraduate students of Purdue University. An *a priori* power analysis was conducted using the G-Power (Faul et al., 2009). With a low-to-medium affect size $f = .30$, and the non-sphericity correction to be $\varepsilon = .70$, G-Power recommended 40 participants.

In the initial data annotation phase, the sample of participants consisted of 10 Graduate students as participants and comprised of 5 males (age: $M = 24.40, SD = 2.96$) and 5 females (age: $M = 25.00$, $SD = 1.41$), all of them belonging to the Computer Graphics Technology Department of Purdue University. This will ensure that for the data annotation phase, the values for the affect response get established by people who are relatively comfortable with virtual reality experiences—so as to eliminate the factor of any negative effect of being in virtual reality becoming an unwanted variable.

The final study comprised of all undergraduate and graduate students of Purdue University. In total 57 students (42 male, 14 females, and 1 other) within the age range of 18 to 29 years old ($M = 19.44$, $SD = 2.28$) formed the pool of participants. All participants were recruited through emails, posters, and word of mouth. The participants gave informed consent, and no monetary compensation was provided for their time. This study was approved by Purdue's Institutional Review Board.

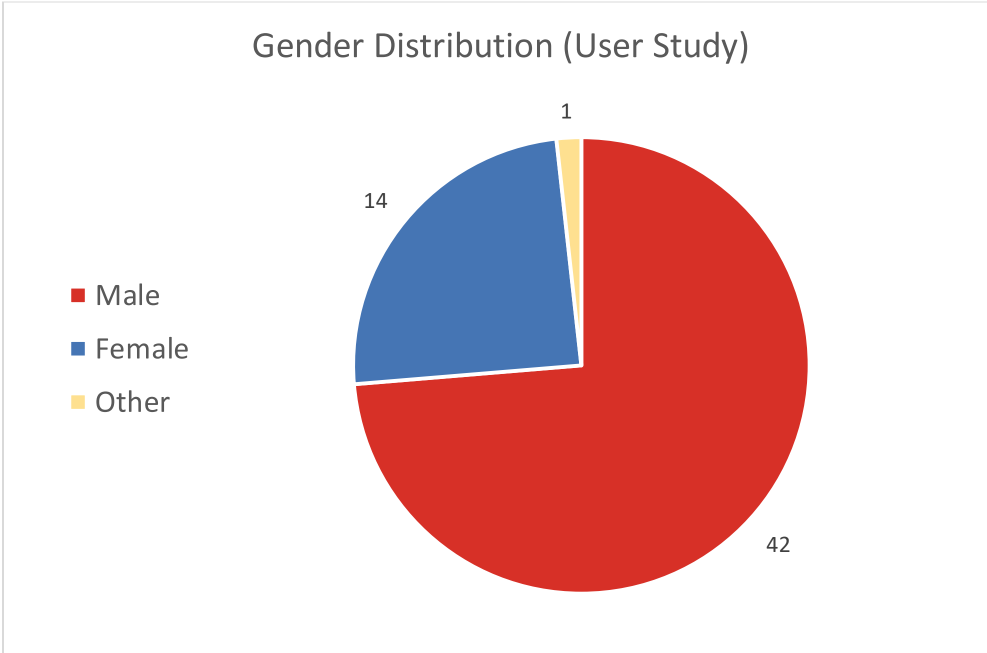



Figure 9. Pie chart displaying the breakdown of the participant's genders.

## 3.4 Instrumentation

For the annotation and user study phase, an Asus Republic of Gamers Scar III laptop was used, with an NVIDIA GeForce RTX 2070 graphics card, and 16 GB of RAM. The virtual reality component was implemented using an Oculus Quest head mount display, which was connected to the laptop using the Oculus Link cable throughout the duration of the phases. The application was developed using the Unity game engine and ran at 60 FPS in the Oculus Quest.

For the recording of affect intensity and response that the user feels during the experience, the PANAS Scale will be used (Watson et al., 1988). It is a scale that can be used as a measure of emotion, using the different words that are listed on it that are measured in intensity using the Likert scale (Likert, 1932; Joshi et al., 2015). Originally, the Likert scale was developed to only use 5 points of measurement, however for this study having a 100-point scale will serve much better as it will allow the users' responses to be measured on a more precise level and will be more appropriate for optimization purposes. On the 100-point scale, the lowest value was 1 which denoted “Not at all negative” for the scene the participant just experienced, and the highest value was 100 which denoted “Extremely negative”.

The PANAS scale has been known to be very reliable for studies (Cronbach alpha coefficient was 0.86—0.90 for positive affect and 0.84—0.87 for negative affect) (Magyar-Moe, 2009; Taber, 2017). For negative affect specifically (since this study mainly focusses on the emotions based on the negative affect), there was significant convergent validity between measures of stress, aversive events, general dysfunction, distress, and discriminant validity with measures of social activity (Watson et. al., 1988). Research works such as Dickinson et al. (2018) and Zibrek et al. (2018) state that the PANAS scale is an effective measure of gauging the affect of users inside a virtual environment, even though very few virtual crowd-based studies have made use of it. Since in fields outside of VR simulations have successfully used in conjunction to provide reliable participant feedback, combined with the high reliability of these tools, it can be considered to be a good approach when it comes to collecting participant response data.

## 3.5 Key Variables

Keeping the PANAS Scale in mind, some of the key variables being recorded were the negative emotions that could be displayed by the participant and determined the questions in the questionnaire. They were geared towards asking the user about emotions dealing with negative affect, experienced during the simulation—the point-scale values were the dependent variables. The four categories of negative emotions (the independent variables), that formulated the four questions were:

1. Upset
2. Distressed

3. Alert
4. Nervous

The original Likert scale featured five different intensities of arousal based on a 5-point Likert Scale—"Very Slightly or not at all", "A little", "Moderately", "Quite a bit", and "Extremely" (Watson et al., 1988). For the 100-point scale, the range will be 1 (which can be denoted as "Very slightly or not at all") to 100 (which can be denoted as "Extremely") as stated previously. All the intermediate values will suggest a mixed response with a bias towards the representations on either end of the scale. These variables and the intensities will first help establish the extent to which these emotions are experienced by the user in the simulation. Scaling up to a 100-point Likert scale that denotes the same levels of arousal, is important for the optimization algorithm, since the value will not need to be normalized, saving up on calculation times. This is done, so as to have a more precise calculation of the affect's impact that each synthesized crowd will have on the user.

Finally, for each experience, certain costs were identified and mathematically formulated, that would determine the total cost of the experience—which would be the main component of optimization calculation. The subsequent sections discuss each cost term in detail, but as an overview three cost terms contribute to the total cost of an experience:

1. Affect Cost ($C_A$ )—The driving value of the simulation. Using this, the affect intensity of the experience is set (either low, medium, or high intensity of the virtual crowd).
2. Affect Variance Cost ($C_V$ )—The variance needed in the displayed behavior of the of the characters in the crowd, as a whole, in-turn introducing a variation in the affect displayed by the crowd.

3. Duplicate Behavior Cost ($C_D$)—This makes it so that no two behaviors are repeated in the experience.

## 3.6 Simulation Design

The experiment comprised of three different stages: the initial data annotation stage, the optimization and synthesis stage, and the final user-response collection stage. There were a total of 30 characters (3D models) that formed the pool from which the characters for generating the crowd were selected from. The design characteristics of these characters followed a "common-person" look i.e. there existed no inhuman or non-human characteristic, appearance-wise.

The data annotation stage's main role was to develop a dataset with annotated values for the behaviors identified as key components of elements of a human crowd. This dataset is constructed by using motion sequences (walk, point, yell, etc.) as well as scripted behaviors (Look At participant's position in the virtual space). To add onto that, the study considered interpersonal space's Proxemic Model, as a trigger to prompt the behaviors to start being played. The Proxemic Model includes three zones that are identified for every human (based on real-world parameters)—(1) Intimate Space, (2) Personal Space, and (3) Social Space. There are a total of thirty four (34) behavior conditions that the dataset contained (No Look At Idle behavior will generally not be dependent on Proxemic Model). Table 2 outlines the hierarchy of the scenes that were included in the dataset.

Table 2. Behavior table for the developed dataset.

| | LookAt | | | NoLookAt | | |
|---|---|---|---|---|---|---|
| | **Intimate** | **Personal** | **Social** | **Intimate** | **Personal** | **Social** |
| **Walk Across** | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| **Walk Towards** | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| **Pointing** | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| **Yelling** | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| **Sidestep** | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| **Idle** | ✓ | ✓ | ✓ | | ✓ | |

The primary tool used for developing the simulation is the Unity game engine. For the animations, they were downloaded from Mixamo[1] which has free animation resources. These were then used in the animator controllers of the engine, so as to add functionality to the crowd characters that could show the behavior when the simulation runs.

In each of the annotation scenes, a pre-determined number of characters (which is developer defined) are spawned in to form a default crowd, with one specific behavior, and then users move virtually through the crowd. For the data annotation stage, the scenes that will be comprised of a set of locations (70 for the annotation phase) for the characters in the crowd to spawn in. The number of characters for the annotation scenes will be pre-set to be a mix of 10 random characters from the pool of available models.

There were 34 scenes developed in total, each representing one behavior combination. All the characters had the same behavior for the annotation phase, however, since each participant had to go through all of the 34 scenes, the characters appearance and location was set randomly to eliminate all negative effects (such as carry over bias, and other appearance related bias) in their responses. Thus, in this way, it was ensured that the data collected was related to the behavior and not the appearance.

[1]https://www.mixamo.com/

Figure 10. Example behaviors. Example behaviors that could be assigned to a virtual character and that were used in our project. From left to right: idle, point, walk. Top row has no Look At, and bottom row has Look At functionality.

Figure 11. Annotation Scenes Example. Example scenes that were used for the annotation phase. From left to right: Look At Idle, Look At Point, Look At Yell.

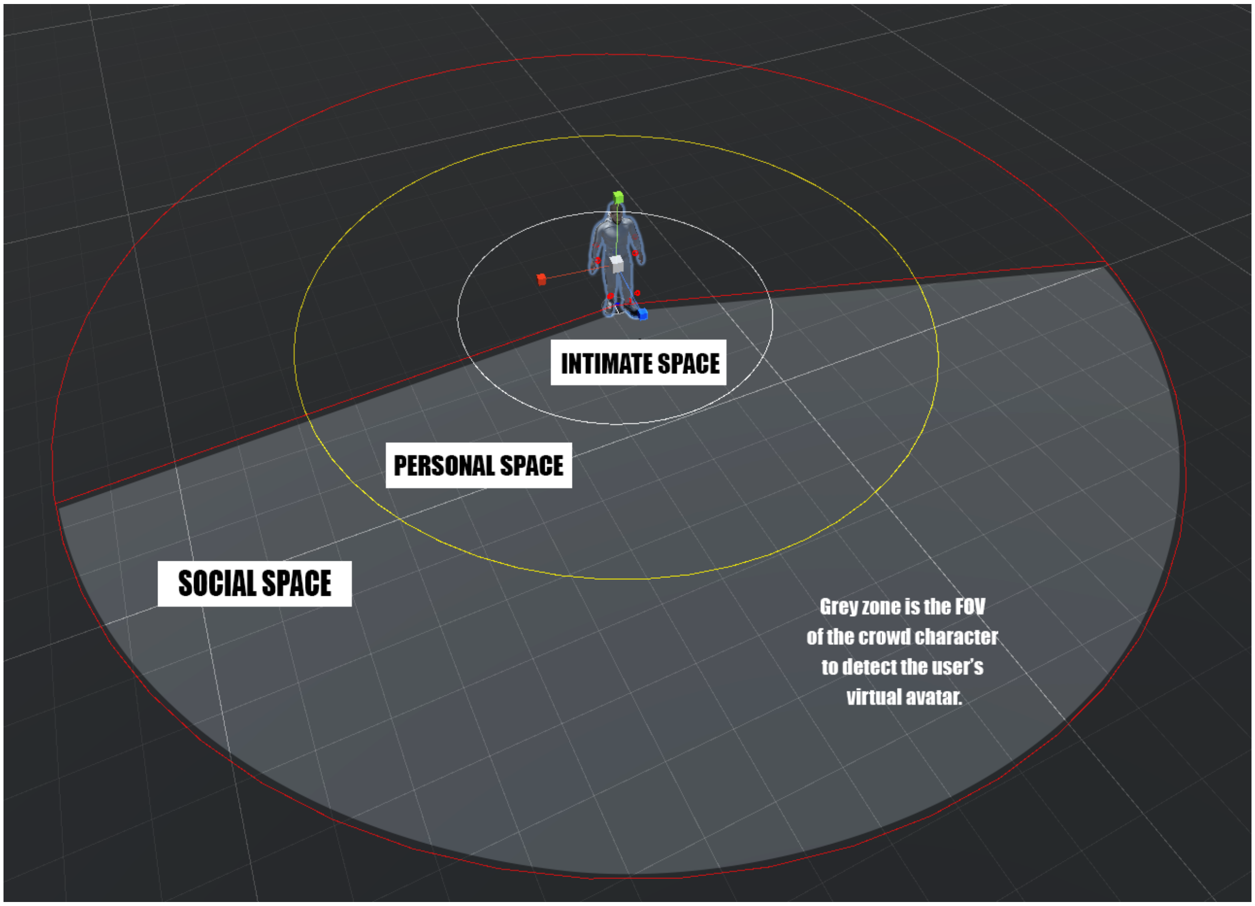



Figure 12. Crowd character with proxemic zones.

When the simulation starts, the crowd is spawned into the randomly assigned locations with the pre-determined behavior for that experience. The screen fade-out of black for the participant, and they are exposed to the virtual scene in a non-abrupt way. The participant's avatar moves through the scene using Unity's Navigation-mesh component, thus not requiring any physical movement or controller input. After moving through the crowd, once the participant avatar reaches the end, the screen fades to black to signify the end of the experience. The participants go through each of the 34 experiences in a random order determined by a Latin Square Design, to eliminate any bias. Figure 13 displays a location of the characters for spawning, as well as the start and end points of the simulation. Figure 14 shows the virtual environment that was developed for each scene. Figure 15 displays an initial synthesized scene. The styling of the environment was designed to be on-par with that of the characters, so as to not be a factor that takes away from the participants understanding the character behaviors. Also, the same stretch of environment was used for all the scenes, both in annotation phase and final user study phase, so as to keep the non-crowd elements as a constant.

At the end of each level, the users rated their negative affect of the simulation, based on the PANAS scale, and specified negative affect parameters (listed above). Each scene took around 2 minutes to complete (along with answering the questionnaire), and hance the total duration of the annotation phase for each participant was around 1 hour and 30 minutes. The questionnaire responses were used to determine and establish a baseline for the values required for the affects, so that a virtual crowd can be automatically synthesized during the optimization and synthesis stage to give off a target negative affect.

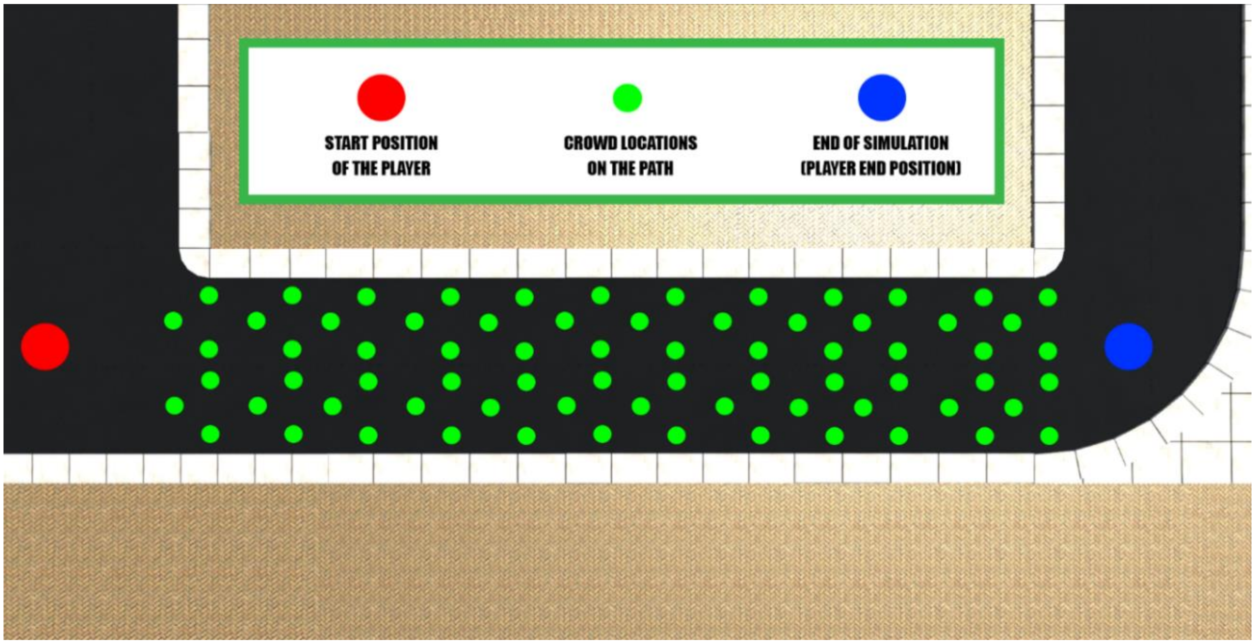



Figure 13. Virtual road that was used in the study with locations of start position and end position of user's avatar, and the locations where the crowd characters were generated.

Figure 14. The virtual environment (in-engine footage).

Figure 15.  A synthesized scene with the virtual characters.

The second stage is short but an important one. After annotating all values for the negative affect in the dataset, it is then fed into an algorithm which tries to adjust the total cost of the scenario based on a target negative affect provided by the developer. This is achieved by the algorithm looking at the affect values of each of the behaviors in the dataset and finding a good combination of behaviors that gives the closest to the target value. The final outcome of this stage was to synthesize a list of 10 characters for a scene with low target affect, 10 characters for a scene with medium target affect, and 10 characters for a scene with high target affect crowds. Section 3.7.4 discusses in-depth about the proposed optimization algorithm. Figure 16 shows a setup for the Optimization component. In the Optimization script, the “Accepted Char Behaviors” is the final list of behaviors that the optimization algorithm outputs after calculating and comparing the total costs. The “Total Characters” value drives the lengths of the “Accepted Char Behaviors” and other internal working list lengths, thus making it highly modular. The other two important terms are “Cost_Total_Proposed” and “Cost_Total_Current”, which are the driving variables for the optimizer. Once the optimizer has found the specific end condition (mentioned in Section 3.7.4), the “Optimize Done” Boolean becomes true, indicating the user that the list of characters has been generated. It should be noted that all other variables have been exposed for aesthetic purposes—usually these would be hidden and non-editable.

The final stage involves a user study to validate the effectiveness of the algorithm. The visual layout of the scene remains the same i.e., it is the same environment, and the same path for the participant’s virtual avatar, however there are some differences. First, the number of locations where the characters can spawn has been reduced to 10 instead of the initial 70, for each of the three scenes. The number of characters remain the same at 10, however a pre-set selection of 10 characters are selected to stay the same throughout the scenes. This ensures that the participants can only focus on the behaviors of the entities in the crowd, rather than their appearance. The locations of where the entities spawned was also kept constant, so as to remove any error that may creep in the response, due to changed location of the characters. All the three scenes will consist of the characters in the crowd having different behavior, and the participants will get exposed to all the three scenes in a random order, determined by a 3 variable Latin Square Design methodology.

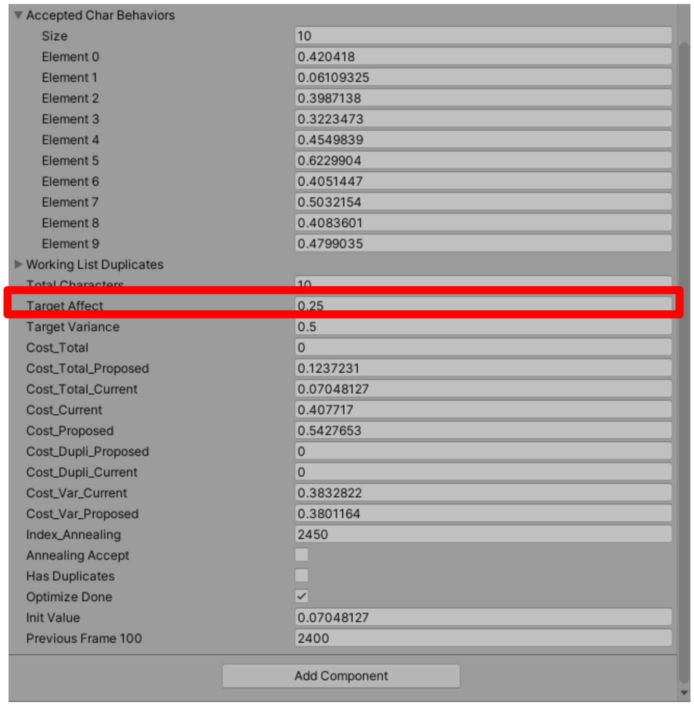



Figure 16. Optimization script example in Unity (target affect is set to 0.25—not a condition for the study).

## 3.7 Problem Formulation and Optimization

One of the primary goals of the simulation in this study is to automatically synthesize an affective experience. The experience, denoted as $E$ consists of a set defined number of virtual characters $c_i$ assembled in a pre-set sequence and is defined as:

$$E = [c_1, c_2, c_3, \dots, c_n]$$

For each experience, three cost terms were developed (also known as design decisions) that will drive the total cost equation of each experience. These design decisions are: the affect cost ($C_A$), the affect variance cost ($C_V$), and the duplicate behavior cost ($C_D$). The final total cost function can be then expressed as:

$$C_{Total}(E) = w_A C_A + w_V C_V + w_D C_D$$

where $w_A$, $w_V$, and $w_D$ are weights associated with the cost terms and their values determine the priority of each cost term to influence the total cost function and drive the experience's final affect levels. It can also be noted that besides these three proposed terms, additional terms may be implemented based on the design decisions of the developers. The cost terms are defined in detail in the following sections.

### *3.7.1 Affect Cost*

The affect cost term encodes the mean target affect that should be exhibited by the multi-character experience. This value is defined as the value that should be achieved by averaging the annotated affect values of all the characters' behaviors inside the experience. This quantity is customizable by the developer and can be weighted so as to prioritize its importance in calculating the total cost of the experience. The affect cost is defined as:

$$C_A(E) = \frac{1}{|E|}\sum_{c_i} A(c_i) - \sigma_A$$

where $|E|$ is the total number of characters, $\sigma_A$ is the target mean affect value of the synthesized experience (e.g., when $\sigma_A = 0.25$, it means that the average negative affect of all characters should be close to $0.25$), and $A(c_i)$ is the negative affect value for the $c_i$ virtual character of the synthesized scene

### *3.7.2 Variance Cost*

The variance cost is used to constrain the variance of the affect values (and in turn the behavior) that should be included in the synthesized scene—hence addressing the topic of variance of behavior across the characters (either the scene ending up with high or low variance). This cost for each experience $E$ can be defined as:

$$C_V(E) = \frac{1}{|E|} \sum_{c_i} (A(c_i) - \bar{A})^2 - \sigma_V$$

where $\sigma_V$ denotes the target affect variance and $\bar{A}$ denotes the mean negative affect of the characters in the scene. It should be noted that when a higher value is assigned to $\sigma_V$, characters with higher affective variations will be included in the scene.

*3.7.3 Duplicate Behavior Cost*

The final cost term dictates whether or not the same behavior will be applied to multiple characters in the synthesized experience. This cost is defined as:

$$C_V(E) = \frac{1}{\frac{|E|!}{(2!\,(|E|-2)!)}} \sum_{c_i, c_j} \Gamma\,(c_i, c_j)$$

where $\frac{|E|!}{(2!(|E|-2)!)}$ returns the total number of combinations between $c_i$ and $c_j$ that are a pair of characters currently in the experience, and Γ returns 1 if the characters have the same behavioral characteristic or 0 if they are different behaviors, based on the evaluation criteria as follows:

$$\Gamma\left(c_i, c_j\right) = \begin{cases} 1 & if\ \mathcal{B}(c_i) == B(c_j) \\ 0 & otherwise \end{cases}$$

where $\mathcal{B}(c_i)$ and $\mathcal{B}\left(c_j\right)$ represent the behavior of characters $c_i$ and $c_j$ respectively.

*3.7.4 Optimization*

The main component for optimization for this study, was addressed using a Markov Chain Monte Carlo Method (Brooks, 1998) called “simulated annealing” (Kirkpatrick et al., 1983) with a Metropolis-Hastings state-searching step (Chib & Greenberg, 1995). The process of optimization begins by initializing an initial set of characters with different behaviors assigned to them, and the system calculates the total cost $C_{Total}(E)$ of that set of characters. Then in the next iteration, the system proposes a new configuration $E'$ and then the algorithm computes the proposed total cost $C_{Total}(E')$. This new configuration is achieved by randomly selecting one of the characters $c_i$ in the $E$ configuration, and then assigning a random behavior from the dataset to the character. The acceptance criteria of a new configuration is achieved when the cost of the proposed configuration, $C_{Total}(E')$, is lower than the current total cost of the current configuration, $C_{Total}(E)$. Simulated annealing works on a factor called “temperature”, which is used to control the greediness of the algorithm in finding the most optimal solution. The temperature gets set as $t = 1.00$ at the beginning of the optimization and is reduced by 0.10 every 200 iterations. As the temperature parameter decreases, the optimizer tries to find the most optimal solution and over time only the optimal solutions start getting accepted. The optimization is completed when the difference in the total cost of the current and proposed configurations is less than 2% over the past 500 iterations. Following presents the optimization pseudocode:

```
1. do while(annealing_Index <= annealing_Iteration)
2.  for every 100th iteration, threshold_value = threshold value - 25
3. generate random_acceptance_criteria
4. if (random_acceptance_criteria < threshold value)
5.  accept annealing (accept bad or unoptimized values)
6. else (reject annealing)
7. Calculate_Total_Cost function executed (calculates 3 cost terms)
8.  total_cost = target_affect_cost + target_variance_cost + duplicate_behavior_cost
9. if ((total_proposed_cost < total_current_cost) OR (annealing_accepted = true))
10.set current_behaviors_list as accepted_behaviors_list, and check whether OptimizeDone
   is true;
11. if OptimizeDone is true, then stop optimizing
12. Replace one behavior for current_behaviors_list and calculate total cost
13. if ((total_cost – initial_cost)/ total_current_cost) < 0.02) and its true for atleast
   100 frames
14. set OptimizeDone = True
```

The other values, unless specified otherwise, were set to be: $\sigma_V = .50$, $w_A = 1.00$, $w_V = .10$, and $w_D = .10$. Since the total characters in each scene needed to be 10, the algorithm considers the affect values of all 10 characters at a time. The weights of the cost terms were implemented to allow the developer to either prioritize certain costs, or give equal priority to all cost terms, to create scenarios with different levels of affect, variance, etc. Figure 17 displays the progression of the developed scenes for low, medium, and high target affect.

## 3.8 Summary

Thus, for the purpose of this study, the aim is towards developing an optimized scenario that is geared towards the negative affect that can be induced in one's mind, when present in a crowded environment. A couple of reasons why the study has been structured in this way: (1) it is seen that humans tend to display a more negative reaction while initially being a part of a crowd, if they do not self-identify themselves as a part of the crowd (i.e. immersion) (Novelli et al., 2013); and (2) because by focusing on one targeted emotion, it will provide a better understanding of one end of the comfort spectrum, and based on that, one can move forward in the direction of creating a more positive scenario in the future (Volonte et al., 2020). A reason for using the proposed optimization method is that it was mostly used to procedurally develop and optimize game levels but not a lot has been explored into optimizing it for targeted behavior-based scenarios. Another significance can be provided from the fact that in many scenarios in video games as well as virtual reality simulations, the developers intend to induce a certain amount of positive or negative affect to the user. Thus, understanding whether if it is possible and feasible to automatically synthesize these polarizing affect inducing crowds, it will enable developers to deliver these experiences more efficiently in interactive media.

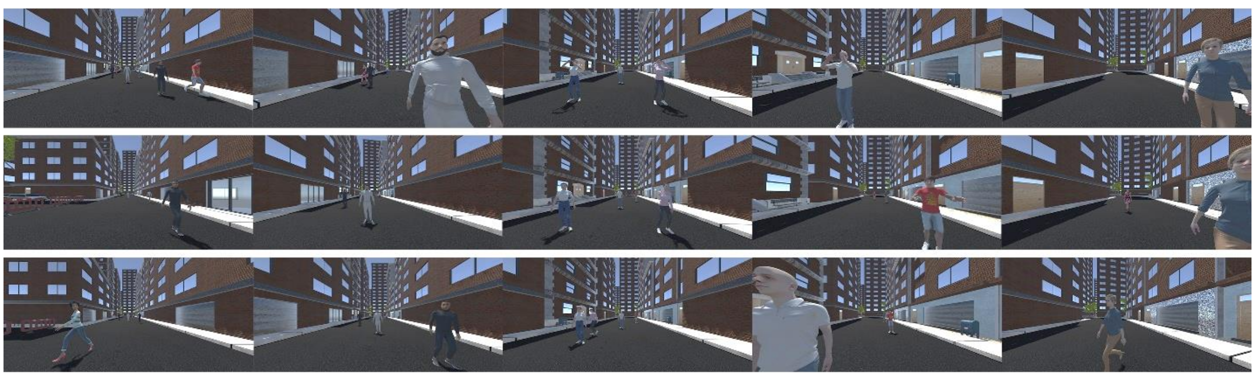


Figure 17. Sequence of each synthesized scene: low target affect (top row), medium target affect (middle row), and high target affect (bottom row).

# CHAPTER 4. USER STUDY AND DATA ANALYSIS

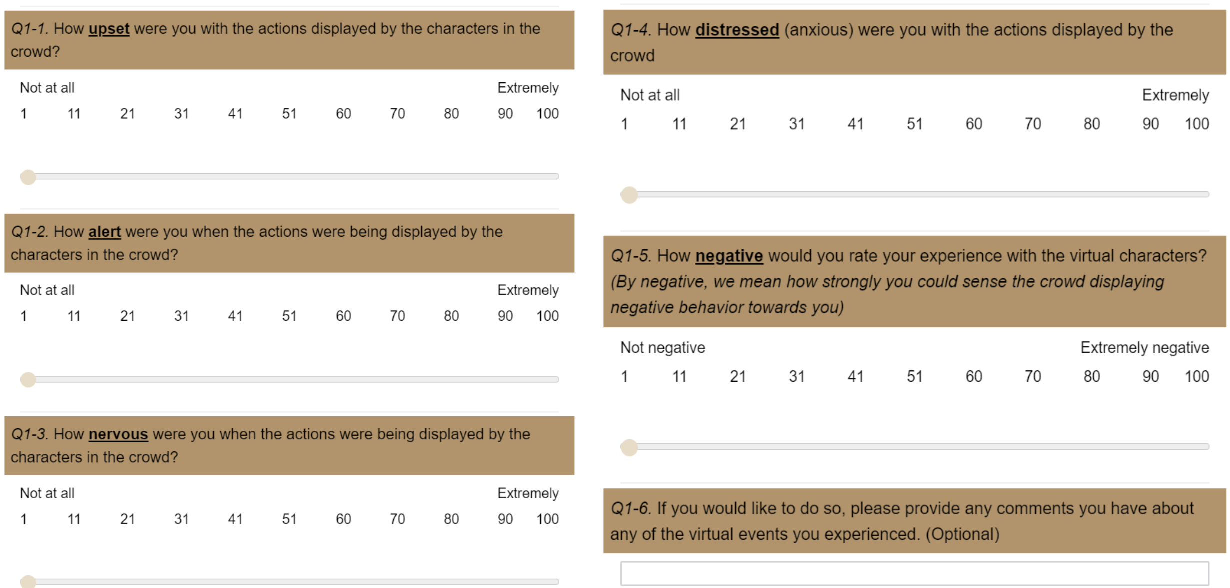
The subsequent sections deal with the final phase of the study, which focuses on user study and data analysis of the three synthesized scenes.

## 4.1 Data Recording and Collection

The data collected, as specified in Section 3.5, will be ordinal in nature and can be translated into a scale (measured between 1 to a 100, where 1 represents "Very slightly or not at all" and 100 represents "Extremely") (Likert, 1932). At the end of the data annotation stage, each affect will have a numerical value associated with it (the values will be an average for all participants of the annotation stage), which will convert the ordinal data to a ratio level data. These ratio data will be used to set the target cost for the optimization algorithm that will help synthesize the experiences and can be scaled up or down to change the intensity of the negative affect of the automatically synthesized and optimized crowd scenarios (for the final stage).

Data collection of the third (final) stage is similar to the first, with the difference being that the crowd will consist of the various behavioral characteristics from the annotation stage, to meet the target intensity cost. Users will again rate the intensity of each affect after each scene, and these values will be compared with the annotated values. This will enable a relationship to be established between the baseline crowd's values, and the optimized values, thus displaying the outcome of synthesis of a crowd that can drive a user's negative affect.

For statistical analysis, a One-Way Repeated Measures Analysis of Variance (ANOVA) was used to explore potential differences across the experimental conditions (synthesized multi-character scenarios with low, mid, and high negative affect) (Lamb, 2003). Figure 18 displays a screen capture of the questions in the questionnaire.


Q1-1. How **upset** were you with the actions displayed by the characters in the crowd?

Not at all — Extremely

1 11 21 31 41 51 60 70 80 90 100

Q1-2. How **alert** were you when the actions were being displayed by the characters in the crowd?

Not at all — Extremely

1 11 21 31 41 51 60 70 80 90 100

Q1-3. How **nervous** were you when the actions were being displayed by the characters in the crowd?

Not at all — Extremely

1 11 21 31 41 51 60 70 80 90 100

Q1-4. How **distressed** (anxious) were you with the actions displayed by the crowd

Not at all — Extremely

1 11 21 31 41 51 60 70 80 90 100

Q1-5. How **negative** would you rate your experience with the virtual characters? *(By negative, we mean how strongly you could sense the crowd displaying negative behavior towards you)*

Not negative — Extremely negative

1 11 21 31 41 51 60 70 80 90 100

Q1-6. If you would like to do so, please provide any comments you have about any of the virtual events you experienced. (Optional)

Figure 18. Questions used for annotating the dataset as well as during the final phase.

## 4.2 Experimental Conditions

For the final phase, three experimental conditions were developed, that would be presented to participants in the form of three scenes made using Unity. These three were low negative affect, medium negative affect, and high negative affect inducing crowds.

1. **Low Negative Affect:** This condition has a value of target negative affect that is considered to be a low value on the scale of 0 to 1. The 10 selected behaviors, that the optimization algorithm provides as an output, had to achieve an overall low total cost (based on the equations described in Section 3.7) of the current scene's configuration. The chosen target values fed to the optimizer algorithm were $\sigma_A = 0.3$ and $\sigma_V = 0.5$. The final behaviors for this scene ended up as (placement of these in the virtual scene were also in the same order):

a. Idle No Look At

b. Yell No Look At (Intimate Space)

c. Point No Look At (Intimate Space)

d. Yell Look At (Personal Space)

e. Walk Towards No Look At (Social Space)

f. Walk Towards No Look At (Personal Space)

g. Walk Across Look At (Intimate Space)

h. Sidestep No Look At (Personal Space)

i. Yell No Look At (Personal Space)

j. Yell Look At (Social Space)

2. **Medium Negative Affect:** This condition has a value of target negative affect that is considered to be a medium value on the scale of 0 to 1. The 10 selected behaviors, that the optimization algorithm provides as an output, had to achieve an overall medium total cost (based on the equations described in Section 3.7) of the current scene's configuration. The chosen target values fed to the optimizer algorithm were $\sigma_A = 0.5$ and $\sigma_V = 0.5$. The final behaviors for this scene ended up as (placement of these in the virtual scene were also in the same order):

a. Yell No Look At (Intimate Space)

b. Yell Look At (Social Space)

c. Walk Towards Look At (Intimate Space)

d. Sidestep No Look At (Social Space)

e. Yell Look At (Intimate Space)

f. Sidestep No Look At (Personal Space)

g. Walk Across Look At (Intimate Space)
h. Walk Towards No Look At (Social Space)
i. Point No Look At (Social Space)
j. Idle No Look At

3. **High Negative Affect:** This condition has a value of target negative affect that is considered to be a high value on the scale of 0 to 1. The 10 selected behaviors, that the optimization algorithm provides as an output, had to achieve an overall high total cost (based on the equations described in Section 3.7) of the current scene's configuration. The chosen target values fed to the optimizer algorithm were $\sigma_A = 0.7$ and $\sigma_V = 0.5$. The final behaviors for this scene ended up as (placement of these in the virtual scene were also in the same order):
    a. Sidestep Look At (Intimate Space)
    b. Point No Look At (Social Space)
    c. Idle Look At (Social Space)
    d. Sidestep No Look At (Intimate Space)
    e. Point Look At (Personal Space)
    f. Walk Towards Look At (Intimate Space)
    g. Walk Across No Look At (Personal Space)
    h. Walk Towards Look At (Personal Space)
    i. Walk Across Look At (Social Space)
    j. Yell No Look At (Intimate Space)


Figure 19. Optimizer Output for low (top row left), medium (top row right), and high (bottom row) conditions.

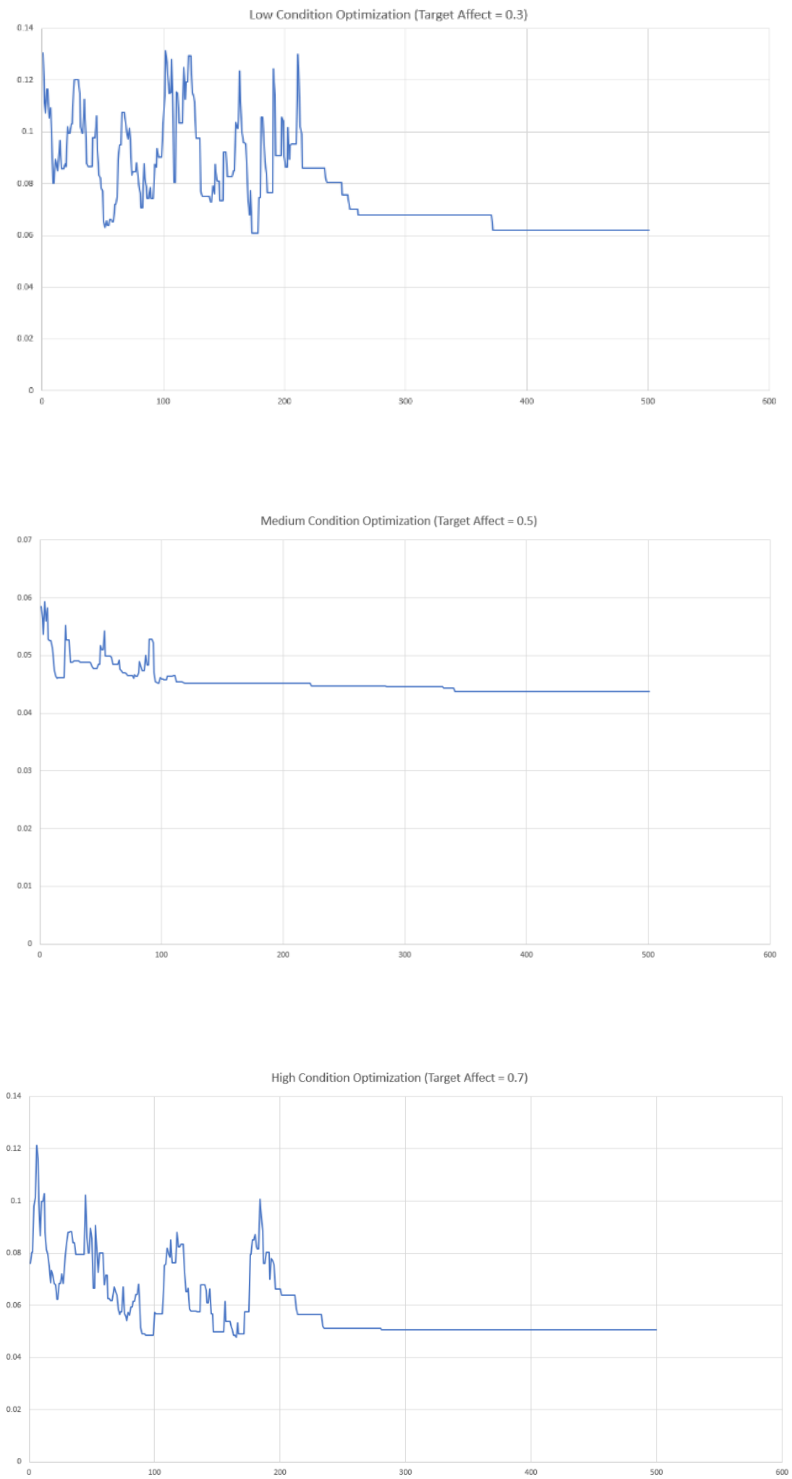



Figure 20. Optimization graphs for low condition (top), medium condition (middle), and high condition (bottom).

## 4.3 Measurement Methods

Participant's affective ratings were collected using the four items from the PANAS scale (check Figure 14). The questions were targeted towards four specific emotions: upset, alert, nervous, and distressed. The questions were presented in the form of a 100-point visual scale, which went from 1 to 100. Here, 1 denoted "Not at all" and 100 denoted "Extremely".

In the response for the annotation phase, these questions were used to annotate the data in the behavior dataset. In the user study phase, the fifth question which directly asked participants to rate the negative affect perceived by them and that was used to directly measure the extent of negative affect sensed by the participants. Finally, a designated space was provided in the survey to allow participants to enter their feedback and comments about the virtual reality application and the conditions they were exposed to.

## 4.4 Procedures

The procedure for data collection was the same throughout the entire process, with the only elements being different are the scenes that the annotation participants and the user study participants got to experience. First task when the participants arrived at the approved lab space, they first were asked to sanitize their hands with the prescribed hand-sanitizers, to minimize the risk of the spread of the COVID-19 virus. The participants and the researcher wore proper apparel (masks) for preventive measures and appropriate physical distancing methods were undertaken whenever necessary. The participants were provided with an Institutional Review Board (IRB) approved consent form and were advised to give a quick read through. At the end of the consent form, the participants were asked to confirm their age to be either 18 years or older, and whether they gave consent to take part in the study. If they answered that with a “Yes”, a demographics questionnaire was shown with questions related to gender, age, average hours of week playing video games, and prior exposure to Virtual Reality. Once they answered that, they were briefed on the study’s process. First, they were familiarized with the headset in question (the Oculus Quest) and were asked to take some time to adjust it to their comfort level. They were informed that they would be experiencing three virtual reality scenarios, and all of them were observation based i.e., they only had to observe the behavior of the characters in the crowd. At the end of each scene, they were asked to input their feedback in the questionnaire based on what they just experienced and rate the negative affect of the virtual population on a scale of 1 to 100. This was repeated twice more (total of three times for three scenes). The study was a between-groups study; thus all the participants experienced all three conditions mentioned in Section 4.2. The order of the three conditions was counterbalanced by a Latin Square (Grant, 1948) for controlling any negative carry over effects. The total duration of the study, for each participant, lasted for around 30 minutes.

## 4.5 Results

### *4.5.1 Annotation Results*

Once the dataset of behaviors was annotated with the respective affect values, it resulted in each behavior having a normalized value between 0 and 1, ready to be fed into the optimization algorithm. Figure 21 displays the average negative affect of each behavior derived from the annotation phase.

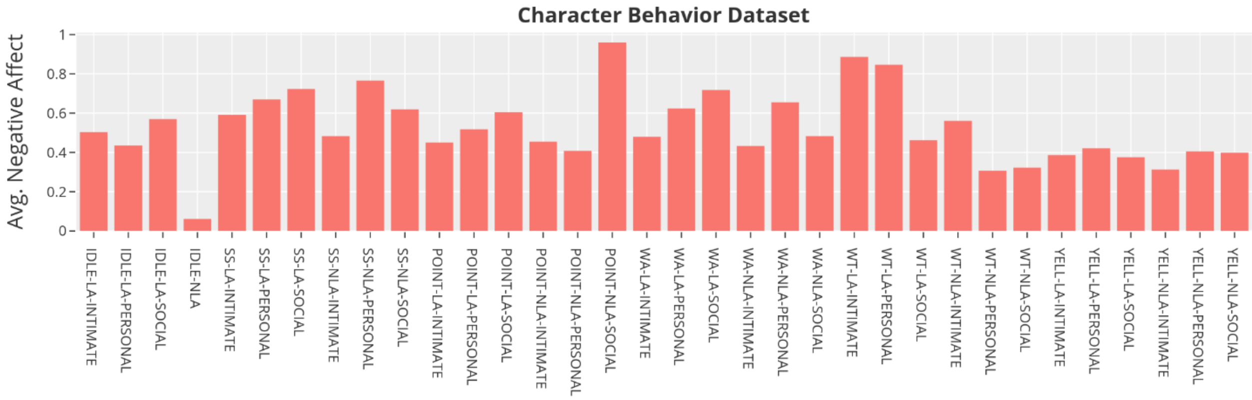


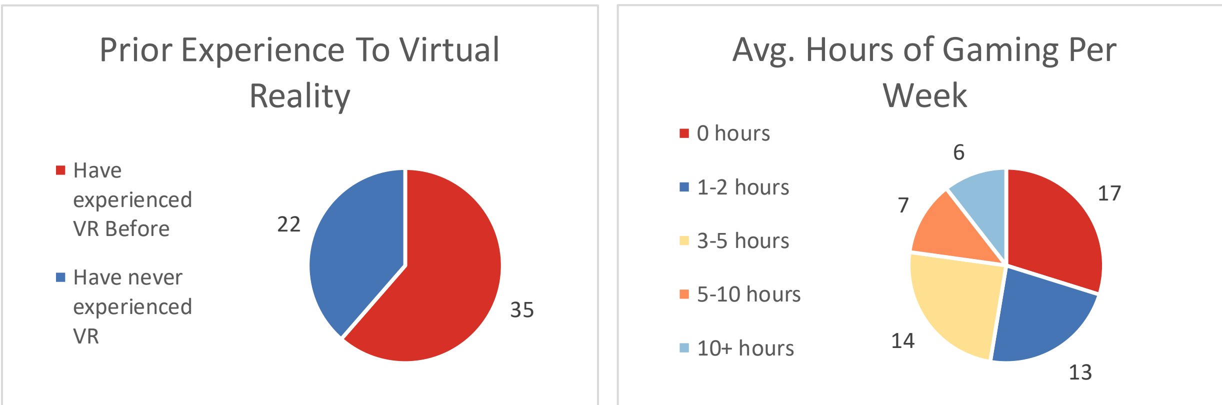

Figure 21. Average negative affect of each behavior derived from the annotation phase (LA: Look At; NLA: No Look At; IDLE: Idle Behavior; SS: Sidestep; POINT: Point Behavior; WA: Walk Across Behavior; WT: Walk Toward; YELL: Yell Behavior; INTIMATE, PERSONAL, and SOCIAL: Proxemics Zones).

### *4.5.2 User Study Results*

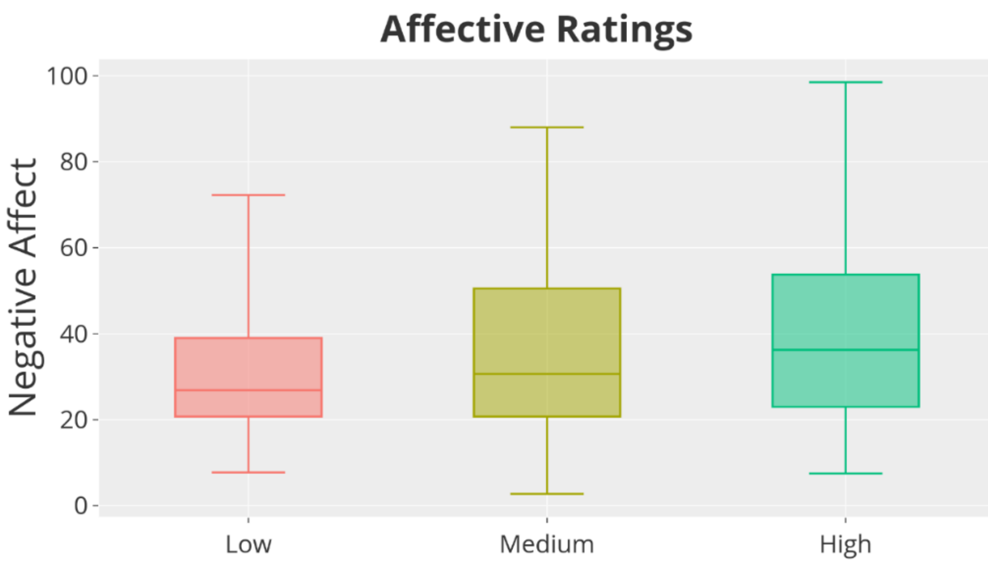
After the user study, a few notable demographics information were found. First, the distribution of the people with prior exposure to VR was 35 and 22 who were having their first VR experience. This was important in the final results, as it gave us the idea that the perception of the emotion and intensity was perceived as expected by new and experienced users alike. Another measured quantity was the amount of time participants played video games. This provided us with an idea to how comfortable participants were with virtual characters and identifying the behaviors of virtual characters, as well as being exposed to virtual scenarios. Figure 22 provides details about both sets of data.

The analysis of the results from the user study phase was done using one-way repeated measures analysis of variance (ANOVA) using the three conditions as independent variables and the questionnaire responses as dependent variables. The internal validity of the questionnaire's scale was measured using Cronbach's alpha coefficient. With sufficient scores ($0.75 < \alpha < 0.81$), a cumulative score was used. Removal of items would not enhance the reliability measures. The normality assumption of the ratings was evaluated with Shapiro-Wilk tests at the 5% level and with the Q-Q plots of the residuals. Post hoc comparisons were conducted using Bonferroni corrected estimates. A $p < .05$ value was deemed statistically significant.

Figure 22. Prior exposure to virtual reality and video game experiences.




The analysis revealed significant results [$\Lambda = 0.726$, $F(2;\ 52) = 9.928$, $p < 0.001$, $\eta_p^2 = 0.274$]. Post hoc comparisons showed that the low negative affect condition ($M = 31.71$, $SD = 16.81$) was rated lower than that of the medium negative affect condition ($M = 36{:}73$, $SD = 21.13$) at the $p = .041$ and high negative affect condition ($M = 40.68$, $SD = 22.50$) at $p = .001$. Moreover, the medium negative affect condition was rated lower than that of the high negative effect condition at the $p = .040$. Boxplots of our results are shown in Figure 23.




\

Figure 23. Boxplots of final user study results.

In conjunction with the self-reported data, participants were also asked to provide comments and feedback based on their experience after every scene. Notable responses pointed towards the fact that the apparent hostile nature of the virtual crowd was perceived well by the participants. In addition to that, participants did report the fact that the negativity of an experience was heightened when virtual characters invaded their spatial zones, confirming that using the proxemics with behaviors is an acceptable design guideline that can be considered when trying to induce a negative feeling in a participant. Some notable comments are provided in Table 3:

Table 3. Participant ID and their comments (with respect to the scenes).

| Participant ID | Comment | Scene Context |
|---|---|---|
| P8 | This particular event was the most threatening, as the virtual characters made hostile remarks, causing a feeling of being uncomfortable and anxious. | With respect to the high negative affect |
| P17 | Characters felt like they didn't like me. They didn't feel threatening, exactly, but they made me feel unwelcome. | With respect to the high negative affect |
| P20 | This scenario the crowd seemed less active in their movements, and less aggressive as a result. | With respect to low negative affect crowd |
| P26 | I did not feel like my personal space was being invaded. | With respect to low negative affect crowd |
| P34 | The first couple simulations were more negative than the last one and they got progressively more positive. The first one was the most negative as all the virtual characters were looking at me and following me as I walked. The last one was the most positive as most of the characters seemed to act more naturally. Whenever a character didn't move out of the way even as they saw me approaching them it seemed very rude to me. | With respect to affect changing from high to medium to low |
| P44 | Yes, the amount of hostility shown was different in each of the cases. | With respect to all the three scenes |

## 4.6 Discussions

The experimental study was conducted to confirm that it is indeed possible to synthesize affective multi-character experiences that induces a certain amount of negative affect in the users. This fundamentally also provides a measure to evaluate whether the proposed pipeline would be effective to synthesize affect driven virtual reality experiences that would be perceived by any users in an expected way. From the results and the self-reported ratings by the participants, conclusive results had been achieved which point towards the synthesized experiences' crowds affected participants in an expected way: the low negative affect displaying crowd was rated to be lower than a medium affect displaying crowd, which itself was rated lower than a high affect displaying crowd.

# CHAPTER 5. CONCLUSIONS AND FUTURE WORK

## 5.1 Conclusion

The study presented a method for synthesizing multi-character experiences that can help incite and elicit a certain amount of negative affect to users who are in a virtual reality scenario, which is also confirmed by the conducted user study. The it can be concluded that the proposed algorithm is successful in optimizing multi-character scenarios and generating a set of characters in the form of a virtual crowd, that are able to induce a specific level of negative affect.

The study started with developing a pipeline for said synthesis and optimization of affect driven crowds. The first step here was to develop the first stage of the pipeline: an annotated dataset that would contain certain behaviors that could be displayed by entities in a crowd, and have users give feedback on that, to associate them with a negative affect value. This would be helpful in understanding how users perceive certain actions when in multi-character scenarios. Using that dataset, a virtual scene was generated and optimized to a specified target negative affect value, and finally a user study was conducted on the synthesized levels, to validate whether the algorithm would be successful in its task of generating virtual crowds, with a specified behavior that induces a certain level of negative affect in the user inside the virtual environment.

Thus, from the work presented here, it is evident that using annotated behavioral datasets for automatically synthesizing virtual crowds with emotions is a viable method of world-building as well as conveying and inducing emotions in the user, thus skipping over the tedious step of manually having to build a virtual crowd.

## <u>5.2 Future Work</u>

An important aspect while formulating the study was to use it as a starting point for future studies to build up on this and the dataset that would be developed and add onto the conditions by making the scenarios and experiences more complex. Some of the future directions that can be taken are listed as follows:

1. Include more behaviors in the dataset, which would help increase the realism factor of the crowds due to varied behaviors.
2. Enhance the simulation more by adding sounds and reactive audial components to the scenes.
3. Use techniques such as galvanic skin response (GSR) and electroencephalogram (EEG) to collect the responses from participants, rather than using self-reported data. This will aid in providing a more accurate response to the stimuli, and a better idea of how the affect inducing crowd would impact the user.
4. Develop more complex scenarios which involve more levels of interactions (such as virtual reality games).
5. Another aspect to be considered is include more design decisions (also known as cost terms) in the scenarios, as well as include techniques such as artificial intelligence, and behavior trees to author event-centric and affect-driven multi-character narratives.
6. Since this study primarily focused on establishing a baseline of affect-driven multi-character scenarios in virtual reality, there is definitely a lot that can be done to expand and add more dimension to this study.

# REFERENCES


Affect. (2020). In Oxford Online Dictionary. Retrieved from https://www.lexico.com/en/definition/affect

Ağıl, U., & Güdükbay, U. (2018). A group-based approach for gaze behavior of virtual crowds incorporating personalities. *Computer Animation and Virtual Worlds*, *29*(5), pp. 1–26. Article Number: e1806. https://doi.org/https://doi.org/10.1002/cav.1806

Albi, G., Bellomo, N., Fermo, L., Ha, S.-Y., Kim, J., Pareschi, L., … Soler, J. (2019). Vehicular traffic, crowds, and swarms: From kinetic theory and multiscale methods to applications and research perspectives. *Mathematical Models and Methods in Applied Sciences*, *29*(10), pp. 1901–2005. https://doi.org/10.1142/s0218202519500374

Amiri, Z., & Sekhavat, Y. A. (2019). Intelligent Adjustment of Game Properties at Run Time Using Multi-armed Bandits. *The Computer Games Journal*, *8*(3-4), pp. 143–156. https://doi.org/10.1007/s40869-019-00083-3

Bailenson, J. N., Blascovich, J., Beall, A. C., & Loomis, J. M. (2003). Interpersonal Distance in Immersive Virtual Environments. *Personality and Social Psychology Bulletin*, *29*(7), 819–833. https://doi.org/10.1177/0146167203029007002

Brogan, D. C., & Hodgins, J. K. (1997). Group Behaviors for Systems with Significant Dynamics. *Autonomous Robots*, *4*(1), pp. 137–153. https://doi.org/10.1007/978-1-4757-6451-2_8

Brooks, S. (1998). Markov chain Monte Carlo method and its application. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *47*(1), 69–100. https://doi.org/10.1111/1467-9884.00117

Carretero, M. R., Peters, C., & Qureshi, A. (2014). Modelling emotional behaviour in virtual crowds through expressive body movements and emotion contagion (pp. 95-98). In M. Obaid, D. S. Sjölie, E. Sintorn, M. Fjeld (Eds.), SIGRAD 2014: Proceedings of SIGRAD 2014, Visual Computing; Göteborg, Sweden, June 12-13, 2014. Linköping, Sweden: Linköping Electronic Conference Proceedings. https://doi.org/10.1145/2628257.2628266

Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., … Wu, H. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, *351*(6280), pp. 1433–1436. https://doi.org/10.1126/science.aaf0918

Chib, S., & Greenberg, E. (1995). Understanding the Metropolis-Hastings Algorithm. *The American Statistician*, *49*(4), pp. 327–335. https://doi.org/10.2307/2684568

Colombi, A., Scianna, M., & Tosin, A. (2016). Moving in a crowd: Human perception as a multiscale process. *Journal of Coupled Systems and Multiscale Dynamics*, *4*(1), pp. 25–29. https://doi.org/10.1166/jcsmd.2016.1093

Dickinson, P., Gerling, K., Hicks, K., Murray, J., Shearer, J., & Greenwood, J. (2018). Virtual reality crowd simulation: effects of agent density on user experience and behaviour. *Virtual Reality*, *23*(1), pp.19–32. https://doi.org/10.1007/s10055-018-0365-0

Durupinar, F., Allbeck, J., Pelechano, N., and Badler, N. I. (2008). Creating crowd variation with the OCEAN personality model (Vol. 3, pp. 1217-1220). In L. Padgham, D. Parkes, J. Muller, S. Parsons (Eds.), AAMAS '08: The 7th International Joint Conference on Autonomous Agents and Multiagent Systems, Estoril, Portugal, May 12-16, 2008. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems.

Durupınar, F., Güdükbay, U., Aman, A., & Badler, N. I. (2015). Psychological parameters for crowd simulation: From audiences to mobs. *IEEE transactions on visualization and computer graphics*, *22*(9), pp. 2145-2159. https://doi.org/10.1109/TVCG.2015.2501801

Durupınar, F., Güdükbay, U., Aman, A., & Badler, N. I. (2016). Simulation of Collective Crowd Behavior with Psychological Parameters. In N. Pelechano, J. Allbeck, M. Kapadia, & N. I. Badler (Eds.), *Simulating Heterogeneous Crowds with Interactive Behaviors* (1st Ed., pp. 131-152). Florida: CRC Press.

Evans L., & Rzeszewski, M. (2020) Hermeneutic Relations in VR: Immersion, Embodiment, Presence and HCI in VR Gaming (pp. 23-38). In X. Fang (Ed.), HCII 2020: Proceedings of the 22nd HCI International Conference (HCI in Games), Copenhagen, Denmark, July 19-24, 2020. Part of Lecture Notes in Computer Science (Vol. 12211, pp. 23-38). Cham: Springer. https://doi.org/10.1007/978-3-030-50164-8_2

Ekman, P. (1999). Basic Emotions. In T. Dalgleish, M. Power (Eds.). *Handbook of Cognition and Emotion* (1st ed, pp. 45-60). Sussex, UK: John Wiley & Sons. https://doi.org/10.1002/0470013494.ch3

Falk, A., & Heckman, J. J. (2009). Lab Experiments Are a Major Source of Knowledge in the Social Sciences. *Science*, *326*(5952), pp. 535–538. https://doi.org/10.1126/science.1168244

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), pp. 1149-1160. https://doi.org/10.3758/BRM.41.4.1149

Fisher, B. S., & Nasar, J. L. (1992). Fear of Crime in Relation to Three Exterior Site Features: Prospect, Refuge, and Escape. *Environment and Behavior*, *24*(1), pp. 35–65. https://doi.org/10.1177/0013916592241002

Grant, D. A. (1948). The latin square principle in the design and analysis of psychological experiments. *Psychological Bulletin*, *45*(5), pp. 427–442. https://doi.org/10.1037/h0053912

Hall, E. T., Birdwhistell, R. L., Bock, B., Bohannan, P., Diebold,, A. R., Durbin, M., … Vayda, A. P. (1968). Proxemics [and Comments and Replies]. *Current Anthropology*, *9*(2/3), pp. 83–108. https://doi.org/10.1086/200975

Helbing, D., Farkas, I., & Vicsek, T. (2000). Simulating dynamical features of escape panic. *Nature*, *407*(6803), pp. 487–490. https://doi.org/10.1038/35035023

Hendrikx, M., Meijer, S., Velden, J. V. D., & Iosup, A. (2013). Procedural content generation for games. *ACM Transactions on Multimedia Computing, Communications, and Applications*, *9*(1), pp. 1–22. https://doi.org/10.1145/2422956.2422957

Henig, R. M. (2007, March 4). Darwin's God. *The New York Times*. Retrieved September 27, 2020, from https://www.nytimes.com/2007/03/04/magazine/04evolution.t.html

Hsieh, F. Y., Bloch, D. A., & Larsen, M. D. (1998). A simple method of sample size calculation for linear and logistic regression. *Statistics in Medicine*, *17*(14), pp. 1623–1634. https://doi.org/10.1002/(sici)1097-0258(19980730)17:14<1623::aid-sim871>3.0.co;2-s

Huang, P.-H., & Wong, S.-K. (2018). Emotional virtual crowd on task completion in virtual markets. *Computer Animation and Virtual Worlds*, *29*(3-4). Article number: e1818. https://doi.org/10.1002/cav.1818

Iachini, T., Coello, Y., Frassinetti, F., & Ruggiero, G. (2014). Body Space in Social Interactions: A Comparison of Reaching and Comfort Distance in Immersive Virtual Reality. *PLoS ONE*, *9*(11), pp. 1-7. https://doi.org/10.1371/journal.pone.0111511

Joshi, A., Kale, S., Chandel, S., & Pal, D. (2015). Likert Scale: Explored and Explained. *British Journal of Applied Science & Technology*, *7*(4), pp. 396–403. https://doi.org/10.9734/bjast/2015/14975

Kapadia, M., Pelechano, N., Allbeck, J., & Badler, N. (2015). Virtual Crowds: Steps Toward Behavioral Realism. *Synthesis Lectures on Visual Computing*, *7*(4), pp. 1–270. https://doi.org/10.2200/s00673ed1v01y201509cgr020

Karpouzis, K., & Yannakakis, G. N. (2018). Emotion-Driven Level Generation. In K. Karpouzis & G. N. Yannakakis (Eds.), *Emotion in Games Theory and Praxis* (Vol. 4, pp. 155–166). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-41316-7_9

Kelly, C., Bernardet U., & Kessler K. (2020) A Neuro-VR toolbox for assessment and intervention in Autism: Brain responses to non-verbal, gaze and proxemics behaviour in Virtual Humans (pp. 565-566). In L. O' Connor (Ed.), IEEEVR 2020: Proceedings of IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), Atlanta, GA, USA, March 22-26 2020. New York: IEEE. https://doi.org/10.1109/VRW50115.2020.00134

Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. *science*, *220*(4598), pp. 671-680. http://www.jstor.org/stable/1690046

Koilias, A., Mousas, C., & Anagnostopoulos, C. N. (2020). I feel a moving crowd surrounds me: Exploring tactile feedback during immersive walking in a virtual crowd. *Computer Animation and Virtual Worlds*, *31*(4-5), pp. 1–16. https://doi.org/10.1002/cav.1963

Koilias, A., Nelson, M., Gubbi, S., Mousas, C., & Anagnostopoulos, C. N. (2020). Evaluating Human Movement Coordination During Immersive Walking in a Virtual Crowd. *Behavioral Sciences*, *10*(9), Article number: 130. https://doi.org/10.3390/bs10090130

Koilias, A., Nelson, M. G., Anagnostopoulos, C. N., & Mousas, C. (2020). Immersive walking in a virtual crowd: The effects of the density, speed, and direction of a virtual crowd on human movement behavior. *Computer Animation and Virtual Worlds*. Advance Online Publication. Retrieved October 17, 2020. https://doi.org/10.1002/cav.1928

Kolkmeier J., Vroon J., & Heylen D. (2016). Interacting with Virtual Agents in Shared Space: Single and Joint Effects of Gaze and Proxemics (pp. 1-14). In D. Traum, W. Swartout, P. Khooshabeh, S. Kopp, S. Scherer, A. Leuski (Eds.), IVA 2016: Proceedings of 16th International Conference on Intelligent Virtual Agents, Los Angeles, CA, USA, September 20-23, 2016. Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-47665-0_1

Kraayenbrink, N., Kessing, J., Tutenel, T., Haan, G. de, Marson, F., Musse, S. R., & Bidarra, R. (2012). Semantic crowds: reusable population for virtual worlds. *Procedia Computer Science, 15*, pp. 122–139. https://doi.org/https://doi.org/10.1016/j.procs.2012.10.064

Krogmeier, C., & Mousas, C. (2020). Eye fixations and electrodermal activity during low-budget virtual reality embodiment. *Computer Animation and Virtual Worlds, 31*(4-5), pp. 1–12. https://doi.org/10.1002/cav.1941

Kyriakou, M., Pan, X., & Chrysanthou, Y. (2016). Interaction with virtual crowd in Immersive and semi-Immersive Virtual Reality systems. *Computer Animation and Virtual Worlds, 28*(5). Article number: e1729. https://doi.org/10.1002/cav.1729

Lamb, G. D. (2003, February). *Understanding "within" versus "between" ANOVA Designs: Benefits and Requirements of Repeated Measures* [Paper presented in meeting]. Annual Meeting of the Southwest Educational Research Association, San Antonio, Texas, USA, February 13-15, 2003. https://files.eric.ed.gov/fulltext/ED474271.pdf

Lee, J., Li, T., & Padget, J. (2013). Towards polite virtual agents using social reasoning techniques. *Computer Animation and Virtual Worlds*, *24*(3-4), pp. 335-343. https://doi.org/10.1002/cav.1517

Li, W., Xie, B., Zhang, Y., Meiss, W., Huang, H., & Yu, L.-F. (2020). Exertion-aware path generation. *ACM Transactions on Graphics*, *39*(4). pp. 1-14. Article number: 115. https://doi.org/10.1145/3386569.3392393

Likert, R. (1932). A technique for the measurement of attitudes. In R. S. Woodworth (Ed.), Archives of Psychology (Vol. 22, Ser. 140, pp. 5-55). New York: The Science Press. OCLC Number: 812060

Lin, J., Zhu, R., Li, N., & Becerik-Gerber, B. (2020). Do people follow the crowd in building emergency evacuation? A cross-cultural immersive virtual reality-based study. *Advanced Engineering Informatics*, *43*. Article number: 101040. https://doi.org/10.1016/j.aei.2020.101040

McAndrew, F. T. (1993). *Environmental psychology*. Pacific Grove, California: Thomson Brooks/Cole Publishing Co.

McFadden, C. (2019, July 17). Why Are We so Attracted to Abandoned Places? *Interesting Engineering*. Retrieved September 27, 2020, from https://interestingengineering.com/why-are-we-so-attracted-to-abandoned-places

McCrae, R.R, Costa Jr., P.T. (1996). Chapter 3: Toward a New Generation of Personality Theories: Theoretical Contexts for the Five-Factor Model. In J.S. Wiggins (Ed.), The Five-Factor Model of Personality: Theoretical Perspectives (1st Ed., pp. 51—87). New York: The Guildford Press

Magyar-Moe, J. L. (2009). Chapter 3 - Positive Psychological Tests and Measures. In J. L. Magyar-Moe (Ed.), *Therapist's Guide to Positive Psychological Interventions* (1st Ed., pp. 43—72). Cambridge. Massachusetts: Elsevier/Academic Press

Makled, E., Abdelrahman, Y., Mokhtar, N., Schwind, V., Abdennadher, S., & Schmidt, A. (2018). I like to Move it: Investigating the Effect of Head and Body Movement of Avatars in VR on User's Perception (pp. 1-6). In R.L. Mandryk, M. Hancock, M. Perry & A. L. Cox (Eds.), CHI EA '18: Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal, Quebec, Canada, April 21-26, 2018. Paper Number: LBW099. New York: Association for Computing Machinery. https://doi.org/10.1145/3170427.3188573

Maij, David L. R., Van Schie, Hein T., & Van Elk, Michiel. (2019). The boundary conditions of the hypersensitive agency detection device: An empirical investigation of agency detection in threatening situations. *Religion, Brain & Behavior*, *9*(1), pp. 23-51. https://doi.org/10.1080/2153599X.2017.1362662

Mousas, C., Koilias, A., Anastasiou, D., Rekabdar, B., & Anagnostopoulos, C.-N. (2019). Effects of Self-Avatar and Gaze on Avoidance Movement Behavior (pp. 726-734). In M. Haley & Junction Publishing (Comp.), Proceedings of IEEEVR 2019: Proceedings of the 26th IEEE Conference on Virtual Reality and 3D User Interfaces, Osaka, Japan, March 23-27, 2019. New York: IEEE. https://doi.org/10.1109/vr.2019.8798043

Moussaïd, M., Kapadia, M., Thrash, T., Sumner, R. W., Gross, M., Helbing, D., & Hölscher, C. (2016). Crowd behaviour during high-stress evacuations in an immersive virtual environment. *Journal of The Royal Society Interface*, *13*(122), Article number: 20160414. https://doi.org/10.1098/rsif.2016.0414

Mossberg, A., Nilsson, D., & Wahlqvist, J. (2020). Evacuation elevators in an underground metro station: A Virtual Reality evacuation experiment. *Fire Safety Journal*. Article In Press. Article Number: 103091. https://doi.org/10.1016/j.firesaf.2020.103091

Nelson, M., Koilias, A., Gubbi, S., & Mousas, C. (2019). Within a Virtual Crowd: Exploring Human Movement Behavior during Immersive Virtual Crowd Interaction (pp. 1-10). In S. N. Spencer (Ed.), VRCAI 2019: Proceedings of The 17th International Conference on Virtual-Reality Continuum and Its Applications in Industry; Brisbane, Queensland, Australia, November 14-16, 2019. New York: Association for Computing Machinery. https://doi.org/10.1145/3359997.3365709

Novelli, D., Drury, J., Reicher, S., & Stott, C. (2013). Crowdedness Mediates the Effect of Social Identification on Positive Emotion in a Crowd: A Survey of Two Crowd Events. *PLoS ONE, 8*(11). Article number: e78983. https://doi.org/10.1371/journal.pone.0078983

Paris, S., Pettré, J., & Donikian, S. (2007). Pedestrian Reactive Navigation for Crowd Simulation: a Predictive Approach. *Computer Graphics Forum*, *26*(3), pp. 665–674. https://doi.org/10.1111/j.1467-8659.2007.01090.x

Pettré, J., Ondřej, J., Olivier, A.-H., Cretual, A., & Donikian, S. (2009). Experiment-based modeling, simulation and validation of interactions between virtual walkers (pp. 189-198). In D. Fellner & S. Spencer (Eds.), SCA '09: Proceedings of the 8th Annual ACM SIGGRAPH/Eurographics Symposium on Computer Animation, New Orleans, Louisiana, USA, August 1-2, 2009. New York: Association for Computing Machinery. https://doi.org/10.1145/1599470.1599495

Picard, R. W. (2000). In R. W. Picard (Ed.). *Affective computing*. Massachusetts: MIT Press.

Pelechano, N., Allbeck, J. M., & Badler, N. I. (2008). Virtual Crowds: Methods, Simulation, and Control. *Synthesis Lectures on Computer Graphics and Animation*, *3*(1), pp. 1–176. https://doi.org/10.2200/s00123ed1v01y200808cgr008

Porras Garcia, Bruno, Ferrer Garcia, Marta, Olszewska, Agata, Yilmaz, Lena, González Ibañez, Cristina, Gracia Blanes, Mireia, . . . Gutiérrez Maldonado, José. (2019). Is This My Own Body? Changing the Perceptual and Affective Body Image Experience among College Students Using a New Virtual Reality Embodiment-Based Technique. *Journal of Clinical Medicine*, *8*(7), pp. 925 - 938. https://doi.org/10.3390/jcm8070925

Proxemics. (2020). In Oxford Online Dictionary. Retrieved from https://www.lexico.com/en/definition/proxemics

Short, T., Adams, T. (2017). Algorithms and Approaches. In T. Short & T. Adams (Eds.), *Procedural Generation in Game Design* (1st ed., pp. 271-298). New York: A K Peters/CRC Press. https://doi.org/10.1201/9781315156378

Snowden, R. J., & Freeman, T. C. A. (2004). The visual perception of motion. *Current Biology*, *14*(19), pp. 1694–1790. https://doi.org/10.1016/j.cub.2004.09.033

Sohre, N., Mackin, C., Interrante, V., Guy, S. J. (2017). Evaluating collision avoidance effects on discomfort in virtual environments (pp. 25-29). In L. Ming, A. H. Olivier, J. Pettre (Comp.), IEEE VHCIE: 2nd Workshop of Virtual Humans and Crowds for Immersive Environments, Los Angeles, CA, USA, March 19, 2017. New York: IEEE. https://doi.org/10.1109/VHCIE.2017.7935624

Stamatopoulou, I., Sakellariou, I., & Kefalas, P. (2012). Formal Agent-Based Modelling and Simulation of Crowd Behaviour in Emergency Evacuation Plans.(pp. 1133-1138). In Juan E. Gurrero (Ed.), ICTAI 2012: Proceedings of the 24th International Conference on Tools with Artificial Intelligence; Athens, Greece, November 7-9, 2012. New York City: IEEE Computer Society CPS. https://doi.org/10.1109/ictai.2012.161

Still, P. (2019, February 13). Crowd Safety and Crowd Risk Analysis. Retrieved November 01, 2020, from https://www.gkstill.com/Support/crowd-density/CrowdDensity-1.html

Sun, Z., Yu, W., Zhou, J., & Shen, M. (2017). Perceiving crowd attention: Gaze following in human crowds with conflicting cues. *Attention, Perception, & Psychophysics*, *79*(4), pp. 1039–1049. https://doi.org/10.3758/s13414-017-1303-z

Taber, K. S. (2017). The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. *Research in Science Education*, *48*(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Tao, J., Tan, T. (2005). Affective Computing and Intelligent Interaction (pp. 981-995). In T. Tao, T. Tan, R. W. Picard (Eds.), ACII 2005: The First International Conference on Affective Computing and Intelligent Interaction; Beijing, China, October 22-24, 2005. Berlin, Germany: Springer-Verlag Berlin Heidelberg. Part of Lecture Notes in Computer Science book series (Vol. 3784). https://doi.org/10.1007/11573548

Thompson, D. E., Aiello, J. R., & Epstein, Y. M. (1979). Interpersonal distance preferences. *Journal of Nonverbal Behavior, 4*(2), pp. 113–118. https://doi.org/10.1007/bf01006355

Vater, C., Klostermann, A., Kredel, R., & Hossner, E.-J. (2020). Detecting motion changes with peripheral vision: On the superiority of fixating over smooth-pursuit tracking. *Vision Research*, *171*, pp. 46–52. https://doi.org/10.1016/j.visres.2020.04.006

Volonte, M., Hsu, Y.-C., Liu, K.-Y., Mazer, J. P., Wong, S.-K., & Babu, S. V. (2020). Effects of Interacting with a Crowd of Emotional Virtual Humans on Users' Affective and Non-Verbal Behaviors (pp. 293 – 302). In L. O' Connor (Ed.), IEEEVR 2020: Proceedings of IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops, Atlanta, GA, USA, March 22-26, 2020. New York: IEEE. https://doi.org/10.1109/vr46266.2020.00049

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*(6), pp. 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Wilcox, L. M., Allison, R. S., Elfassy, S., & Grelik, C. (2006). Personal space in virtual reality. *ACM Transactions on Applied Perception (TAP)*, *3*(4), pp. 412–418. https://doi.org/10.1145/1190036.1190041

Wurm, Moritz F., & Schubotz, Ricarda I. (2018). The role of the temporoparietal junction (TPJ) in action observation: Agent detection rather than visuospatial transformation. *NeuroImage*, *165*, pp. 48-55. https://doi.org/10.1016/j.neuroimage.2017.09.064

Xu, M. L., Jiang, H., Jin, X.-G., & Deng, Z. (2014). Crowd Simulation and Its Applications: Recent Advances. *Journal of Computer Science and Technology*, *29*(5), pp. 799–811. https://doi.org/10.1007/s11390-014-1469-y

Xu, M., Xie, X., Lv, P., Niu, J., Wang, H., Li, C., … Zhou, B. (2019). Crowd Behavior Simulation With Emotional Contagion in Unexpected Multihazard Situations. IEEE *Transactions on Systems, Man, and Cybernetics: Systems*. Advance Online Publication. Retrieved October 17, 2020. https://doi.org/10.1109/tsmc.2019.2899047

Yannakakis, G. N., & Togelius, J. (2011). Experience-Driven Procedural Content Generation. *IEEE Transactions on Affective Computing*, *2*(3), pp. 147–161. https://doi.org/10.1109/t-affc.2011.6

Zhang, Y., Xie, B., Huang, H., Ogawa, E., You, T., & Yu, L. F. (2019). Pose-guided level design (pp. 1-12). In (Eds.), CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, Scotland, UK, May 4-9, 2019. New York: Association for Computing Machinery. https://doi.org/10.1145/3290605.3300784

Zibrek, K., Kokkinara, E., & Mcdonnell, R. (2018). The Effect of Realistic Appearance of Virtual Characters in Immersive Environments - Does the Character's Personality Play a Role? *IEEE Transactions on Visualization and Computer Graphics*, *24*(4), pp. 1681–1690. https://doi.org/10.1109/tvcg.2018.2794638