NOTICE: This is a paper that was accepted for publication in the **IEEE Transactions on Visualization and Computer Graphics**. It is not the final published version. The DOI of the definitive version available in IEEE Explore is https://doi.org/10.1109/TVCG.2025.3543740

Introducing Agent Personality in Crowd Simulation Improves Social Presence and Experienced Realism in Immersive VR

Massimiliano Pascoli, Fabio Buttussi, Konstantin Schekotihin, and Luca Chittaro

***Abstract*—Convincing crowd behavior simulation is becoming essential in many application domains, including video games, cinematography, urban planning, safety simulations, and training. In this paper, we propose a novel and lightweight mesoscopic system for personality-based crowd simulation in immersive virtual reality (iVR). We use the Big Five personality framework, also known as OCEAN, to model a synthetic personality for each autonomous agent. Agents can autonomously aggregate in formations using machine learning-based clustering techniques operating on OCEAN. Moreover, agents can also externalize their personality traits by performing peculiar behavioral animations. To choose which animations to perform, we adopt a probabilistic approach that considers each OCEAN dimension as a continuous spectrum with two extremes linked to pairs of animations. Our system is designed to be flexible and suitable for different applications. Flexibility is achieved by using graphs to store agent and map topology data that control how the agents move and behave at runtime. In a within-subjects study with 40 users, we compare our personality-based system against a basic system that does not use personality. Results show that introducing personality into iVR crowd simulation enhances users' social presence and experienced realism. Introducing personality also increases the perceived match between the agents and the virtual environment where the simulation takes place.**

***Index Terms*—Immersive VR, Crowd Simulation, Agent Personality, Group Formation, Behavioral Animation, User Study.**

I. Introduction

SINCE the first research in behavioral animation for emulating flocks of birds and schools of fish [60], the simulation of groups, including crowds, has substantially progressed, as described in the literature surveys by Musse et al. [49] and Yang et al. [81]. Improvements proceeded alongside hardware performance, but a demand for an ever-increasing level of quality from real-world applications accompanied them.

Crowd simulation systems can be used in all fields that require several synthetic autonomous agents, such as:

- video games (more realistic, autonomous non-player characters, with the ultimate goal of making the environment livelier)
- CGI for cinematography (e.g., the well-known application of a custom-made crowd simulation software in the popular fantasy trilogy *The Lord of The Rings* to lower production costs of the battlefield scenes [73])
- military training (as presented by Alexander et al. [2]) or counter-terrorism training in highly crowded locations
- serious games in general (video games to foster serious objectives [86], in which realism is essential to prevent negative training effects)
- mass behavior simulations for sociology studies (to model emerging behaviors resulting from the interactions between agents) and similar fields of application [63]
- architectural and urban construction planning (e.g., city design and evaluation [3])
- safety/riots/evacuation simulations (e.g., Başak et al. [6])
- cultural heritage (e.g., to bring past life into digital cultural heritage [39], [71]).

Given the different demands of these applications and the complexity of simulating several autonomous agents in the same virtual environment (VE), crowd behavior simulation still has several open research problems, as pointed out by Musse et al. [49]: 1) need for improvements to crowd properties and realism using machine learning techniques; 2) inclusion of personality, cultural aspects, and emotions to increase the realism of individual behaviors; 3) need to provide the possibility to zoom in and out in crowd simulations to enable the visualization of crowd details at different levels; 4) application of realistic crowd simulation in immersive virtual reality (iVR) applications; and 5) usage of computer vision methodologies for highly dense crowds.

In this paper, we focus on techniques for crowd behavior simulation in iVR, with particular attention to the personality of the individual autonomous agents in the crowd. To increase social presence when experiencing crowd simulations and to model more realistic and plausible crowds efficiently, we propose a simulation system where agents have a synthetic personality that makes them behave differently at runtime in iVR. The system is based on the OCEAN personality model [44]. The five-dimensional OCEAN tuple is the low-level representation of each agent's personality. Every OCEAN dimension changes how the agent interacts with other agents and the environment. In addition to personality, behavior is affected by external factors, such as the agent's current location and destination.

The proposed system includes a way for agents to dynamically form a group and wander around as a cohesive unit: a formation. This is achieved using machine learning techniques. Moreover, we explore a new way to externalize internal personality traits through agent animations inside the simulation. Destination selection and agent spread in the environment are

Massimiliano Pascoli, Fabio Buttussi and Luca Chittaro are with the Human-Computer Interaction Lab, University of Udine.

Konstantin Schekotihin is with the University of Klagenfurt.

Manuscript received XXXXX XX, XXXX; revised XXXXXX XX, XXXX.

the other components influenced by the interactions between personality and locations in the VE.

We also present the data structures that support a flexible expansion of the set of behaviors and the application of our system to other VEs and use cases. Finally, we report on an evaluation with 40 participants that showed an increased sense of social presence and experienced realism when personality was added to the simulation. The introduction of personality also increases the perceived match between the agents and the VE where the simulation occurs.

The organization of the manuscript is as follows. In Sec. II, we discuss the literature that provided the starting point of this work, focusing on the OCEAN model of personality and simulation of crowds in iVR. We present the technical details of the implementation of our system in Sec. III, with particular regard to data structures and algorithms. The study methodology is illustrated in Sec. IV, and the results obtained from the study in Sec. V. We discuss the implications of the results and the current limitations in Sec. VI. Finally, we outline future work in Sec. VII.

II. Related Work

The research landscape outlined by Musse et al. [49] is an excellent summary of more than 30 years of research in crowd simulation as seen by authoritative experts in crowd simulation, virtual reality, and virtual humans. In addition, as anticipated in Sec. I, Musse et al. highlight various open problems, some of which we aim to address in this paper.

Yang et al. [81] suggest a taxonomy of the several approaches in the literature to handle multiple anthropomorphic agents in the same virtual space. Based on the scope of action of algorithms used to control crowds—single agents as separate entities, whole groups or crowds as single entities, and a mixed approach—the survey organizes the research landscape into three main categories: microscopic, macroscopic, and mesoscopic simulation systems. Macroscopic simulation systems can simulate thousands of people with fewer individual details, microscopic simulation systems are individual-based and fail to manage many agents simultaneously, and mesoscopic simulation systems focus on simulating crowds as collections of independent groups of agents.

Our attention is directed to mesoscopic simulation systems that involve dynamic group behavior described by Yang et al. [81]. For example, Karamouzas and Overmars [36] try to create groups at runtime, providing the basis for dynamic formations in synthetic crowds. Simulation systems that involve dynamic group behavior can also be applied to commercial entertainment scenarios, such as real-time strategy video games [30]. The above-mentioned systems are not used in an iVR setting nor exploit personality. The following two subsections present a brief overview of systems that exploit personality during the simulation and are applied in the iVR setting.

A. OCEAN Personality for Crowd Simulation Yang et al. [81] describe social psychological simulation systems in which personality traits and emotion contagion theories influence agents' behavior. In social psychological simulation systems, emphasis is given to the psychological sphere of the synthetic agent. Personality traits and emotion contagion are considered to model crowd behavior and interactions.

Our system uses the OCEAN personality model [44] to represent agent personality. The OCEAN model, also known as the Big Five personality model, characterizes an individual personality in five dimensions. Each dimension is associated with a number representing the value of that dimension as described in detail in [44] and summarized in the following paragraphs.

a) Openness: A low openness value means that the individual is closed to experience and tends to be traditional in many aspects of social life. The individual favors well-established routines and has few interests. High openness is instead positively correlated with creativity, intelligence, and knowledge. Individuals with high openness tend to be motivated to look for new experiences and show a high level of absorption (disposition for episodes of total attention when performing tasks).

b) Conscientiousness: Conscientiousness is associated with how tasks are generally handled by the individual, e.g., levels of self-organization and how much effort the individual invests in the task. An individual with high conscientiousness desires to perform tasks well and takes obligations toward others seriously. Conscientiousness is also related to self-discipline. Highly conscientious individuals show a consistent level of impulse control. Conscientiousness inhibits spontaneous behavior, thus allowing planned and thoughtful acting. In extreme cases, high conscientiousness can be perceived as stubbornness. Low conscientiousness values are related to flexibility, spontaneity, and procrastination. In extreme cases of low conscientiousness, individuals may become unreliable.

c) Extraversion: Extraversion is related to how an individual relates to groups and behaves in a public context. This dimension can be expressed in numerous ways: from music tastes, [59] to clothing [64]. Generally, high extraversion individuals tend to be outgoing and energetic. They enjoy social interaction and tend to be more prone to boredom when left alone. Introversion (low extraversion values) is manifested with reserved behavior [72]. Introverted individuals are more detached from the social world and hence need much less stimulation from others and more time alone than high extraversion individuals.

d) Agreeableness: The agreeableness dimension focuses on how individuals get along. Individuals with high values of agreeableness are generally perceived as "kind, generous, trusting and trustworthy, helpful, and willing to compromise their interests with others" [28]. Low agreeableness individuals prioritize self-interest and are generally less concerned about others' well-being. Moreover, they can be perceived as suspicious, unfriendly, or uncooperative. They also tend to be competitive.

e) Neuroticism: Neuroticism, also known as emotional instability, indicates how much an individual is emotionally unstable. Neuroticism also refers to the tendency to feel negative emotions [33]. Individuals with high neuroticism values are more likely to experience anxiety, fear, anger, frustration, envy, jealousy, pessimism, guilt, and loneliness [72] and often interpret life situations as threatening. These negative emotions persist longer with high neuroticism. Moreover, individuals with high neuroticism do not usually like changes. Individuals with low neuroticism values tend to be calm, even-tempered, and less likely to feel tensed or stressed.

From now on, we refer to the OCEAN personality as $P = \langle O, C, E, A, N \rangle$. When we refer to one agent in particular, its personality is indicated as $p = \langle o, c, e, a, n \rangle$. A projection of the tuple belonging to a specific agent along a dimension is written as $p(o), p(c), \ldots, p(n)$.

Despite some criticisms [10], [22], [55], the OCEAN personality model remains the most widely known, studied, and used personality model. The OCEAN model gathers essential aspects of personality, which translate into behaviors described in the paragraphs above. Thanks to this, the OCEAN model is used for individual autonomous agents and crowds as described by Chittaro and Serra [14], Durupinar et al. [19], Zhou et al. [84], and Guy et al. [29].

Crowd simulation systems have increasingly incorporated personality traits to enhance the realism of simulated behaviors. The OCEAN model has been used in crowd simulation to capture interactive behaviors representing multiple types of crowds [19], [84]. Personality traits have been integrated into crowd simulation to generate heterogeneous behaviors, enhancing the diversity and complexity of simulated crowds [29].

Durupinar et al. [21] has demonstrated that incorporating the OCEAN model to enhance a previous social-forces crowd simulation framework (HiDAC) [57] influences user perception of crowds. This approach integrates OCEAN personality traits with HiDAC behaviors, effectively avoiding the need for low-level parameter adjustments. The system proficiently manages high-density crowds, exhibiting emergent behaviors related to agent flow and trajectories. The primary focus is indoor virtual environments (VEs), with the design process specifying predetermined agent groups with particular OCEAN parameters. However, the use of animations to convey personality and the ability to control the frequency of specific behaviors to express personality traits were outside the intended scope of their system.

The OCEAN model is widely employed in computational crowd simulations. For example, computer vision can predict OCEAN parameters from trajectories of real crowds by analyzing videos [25], [26].

Researchers explored the application of personality models like OCEAN in various crowd scenarios, such as emergency evacuations, where individual behaviors, emotional changes, and crowd dynamics can be effectively simulated [35], [40], [61]. By incorporating personality traits into crowd simulation, the latter can better represent how different agents within a crowd may react and interact in response to various stimuli, leading to more realistic simulations [35].

Integrating personality traits, mainly through models like OCEAN, in crowd simulation systems offers a valuable approach to create more sophisticated and realistic simulations of crowd behaviors. Researchers can develop accurate representations of crowd dynamics by considering individual personality differences, contributing to improved crowd management strategies and safety protocols.

However, none of the aforementioned papers considers iVR with all the specific challenges that arise, e.g., first-person experience in a real-world scale and performance limitations. Most of them focus on computer vision tasks either to predict psychometric values by performing pedestrian analysis from trajectories or to calculate trajectories of autonomous agents based on personality [25], [26]. With our work, we aim instead at understanding the effects of personality inclusion in simulations for iVR to both enhance the realism of the represented agents and maximize users' sense of presence while experiencing the simulation. In addition, none of the proposed systems considers personality when deciding on an agent destination. Some proposed systems do not even consider choice of destination based on the characteristics of the VE. For example, the goal of agents in [35], [61], [84] is to exit a building or escape the dangerous situation in cases where an evacuation or an emergency is simulated, or agents are wandering in the VE and activating specific behaviors when preconditions are met.

Furthermore, the mesoscopic systems capable of dynamic grouping taking personality into account [19], [29], [40], do not perform grouping operations in an unsupervised manner.

OCEAN is not the only possible model applied to crowd simulations. Other popular personality models used in the literature about emergency crowd simulations are PEN [23] and HEXACO [38]. Regarding emotions and their contagion in crowds, OCC [54], PAD [46], ASCRIBE [11], ESCAPES [75], and SIR [27] are often used.

A. Simulation of Crowds in IVR

The literature is rich in applications of crowd models in iVR, and interactions between a user and autonomous agents in a crowd.

Berton et al. [9] exploit eye-gaze capabilities of HMDs to understand how crowd density in iVR impacts users' eye-gaze activity when evaluating the surroundings to avoid collisions with other agents.

The researchers put particular focus on collisions when modeling crowds. In Berton et al. [8], users were asked to navigate through a dense static crowd with and without haptic rendering of collisions between users' avatar and the agents populating the VE. Collisions were also studied by Yun et al. [82] and Koilias et al. [37] in a non-static crowd focusing on how users adapt their trajectories and behaviors based on expected collision feedback.

Unfortunately, the literature lacks in crowd simulations that consider personality in the iVR setting.

III. Design and Implementation of the Proposed System

This section provided an in-depth explanation of the design choices in our system for the personality-based crowd simulation. The description includes an overview of the logical data structures governing the simulation, the machine learning algorithms implemented, the decision-making process of every agent, and the technique used to express personality for agents.

Code snippets of components internal to the system are available by request to the corresponding author.

A. Graphs

In this section, we describe how persistent data is represented in a way suitable for serialization in multiple formats on files. Our system is based on two fundamental data structures: the agent graph A and the location graph L.

The agent graph A maintains information regarding agents at runtime and the relations between them. The crowd is viewed as a social network of interacting entities represented as a graph. Every node is associated with an agent and its personality p. Edges represent relations between individuals. There is an edge between nodes n, n' in A if associated agents are involved in a relation. These relations can be of various types, depending on the context or the application. To handle formations, we consider the "grouping" relation: two or more agents are involved in the same grouping relation if they are traveling in the same formation. We only have grouping relations created at runtime by the simulation depending on agent distribution in the VE and their p tuples, so no persistent edge data is memorized.

The location graph L stores structured data related to waypoints (used by agents as destinations of the pathfinding algorithm) located in the VE and relations between them. We used polymorphism to create multiple types of nodes that share a basic background. Every node is characterized by an attractiveness value and a radius. These values control how agents spread on the map together with agents' $p(o)$ as described in Sec. III-C. Higher attractiveness for a waypoint means that more agents choose that waypoint as their pathfinding goal.

We distinguished between two types of waypoints, each associated with a different behavior of agents. An agent that arrives at a "stop waypoint" must stay there for some time (the stopping time of an agent is set to be randomly selected in the interval $[10;60[$ seconds in our study). While staying in a stop waypoint, agents can externalize their personality through animations (see Sec. III-D). Furthermore, agents can group dynamically as described in the following Sec. III-B. A "normal waypoint" is used only to reach other waypoints; no particular behavior is implemented there.

From a waypoint, an agent can go to neighbor locations defined by the proximity relation. When referring to the "map topology" (an example is shown in Fig. 1), we mean the location graph L representing the VE.

B. Formations Through Clustering

P is exploited in our system through agent dynamic clustering. When multiple agents are idling at a stop waypoint, agents can form groups and then wander around the VE as a cohesive unit following a formation. This process happens after a random amount of time (set in the interval $[10;30[$ seconds in our study) for each stop location. Our system considers the personality of agents at the same stop waypoint that might eventually form a group. The debate between different

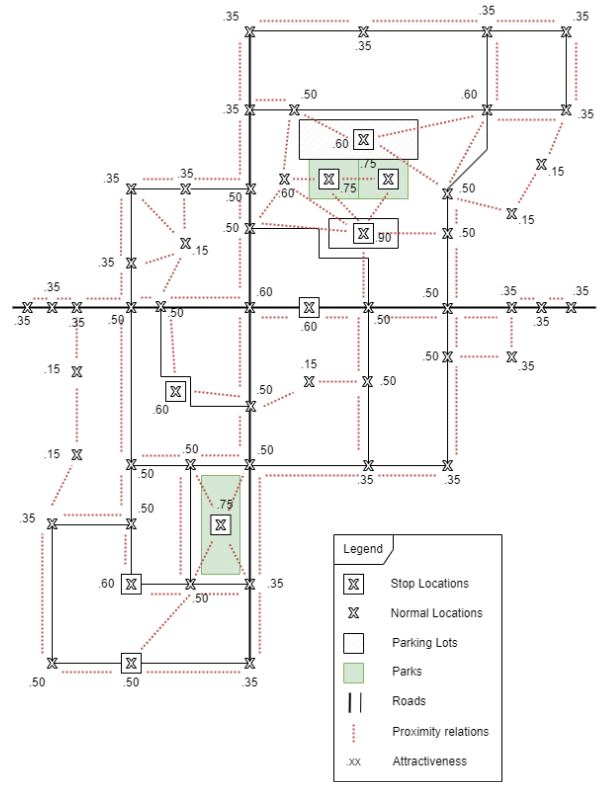


Fig. 1: Map topology used during user testing of our system. The legend depicts some key features present in the VE.

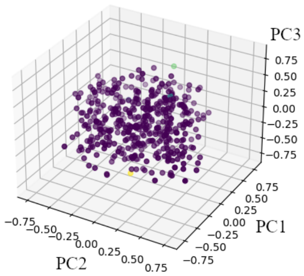
theories of human aggregation is still open in the psychology community. We followed the theory proposed by Izard [31], in which most similar people stay longer in friendship. In our system, similar agents not already in a formation are grouped, and the group lasts longer in time the more its members are similar to one another.

To translate this reasoning to autonomous agents, we used hierarchical clustering. The five-dimensional array p represents a point in five dimensions $p = \langle o, c, e, a, n \rangle, p \in R^5$. The task is to partition all personalities into clusters comprising similar ones. Hierarchical clustering is convenient in this case because there is no need to specify the number of clusters needed. Nonetheless, we need an automatic criterion to cut the resulting clustering dendrogram.

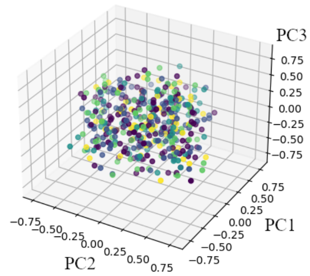
For this reason, we choose cosine similarity (S_C in Eq. (1)) as the measure of distance between points in five-dimensional space, and we adopt an early stopping method for the algorithm. Cosine similarity is commonly used in information retrieval and text mining, usually to find similarity (in terms of subject matter) between documents independently from length [65]. It is also exploited in data mining for clustering cohesion measurement [70]. Other uses of cosine similarity are loss function in deep neural networks, supervised text processing and analysis, and other machine learning applications.

$$S_C(A,B) = \cos(\theta) = \frac{A \cdot B}{\|A\| \cdot \|B\|} \tag{1}$$

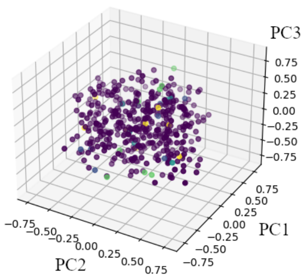
Cosine similarity measures the cosine of the angle between two vectors/points, no matter the magnitude of the vectors. The



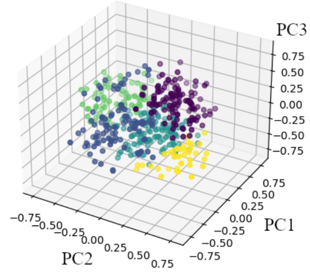
(a) Single linkage clustering.



(b) Complete linkage clustering.



(c) Unweighted average linkage clustering.



(d) Ward's MNVAR linkage clustering.

Fig. 2: Different results of agglomerative hierarchical clustering using several linkage methods and cosine similarity (except for Ward's method, where Euclidean distance is mandatory) on the same uniformly sampled p tuples used in the personality-based simulation. Dimensionality reduction for 3-dimensional plotting is achieved using Principal Component Analysis. Clustering stopping criterion: number of clusters equal to 5.

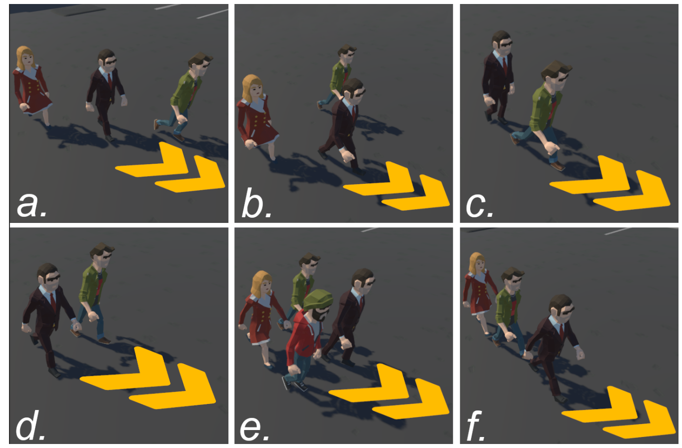


Fig. 3: Possible formation shapes used in the personality-based simulation, created by considering cosine similarity between OCEAN agent personalities in the same stop location. From a. to f.: 3-people-diagonal queue, V-shaped, queue-like couple, horizontal couple, square-shaped, 3-people queue. The yellow arrows show the moving direction of the formations.

measure returns values from the interval $[-1;1]$, as defined by the cosine function. This property is desirable because it provides upper and lower bounds to the similarity between points (an advantage concerning other distance metrics like Euclidean distance). Cosine similarity for personalities will always be in the interval $[0;1]$, due to the bounds that components of tuples p have (each component is a number that lies in $[0;1]$). Here, orthogonal vectors $S_C(A,B) = 0$ mean no similarity of personalities, whereas collinear vectors $S_C(A,B) = 1$ mean the perfect similarity. The measure can be applied to an arbitrary dimensional hyperspace (in our case, the five-dimensional hyperspace).

To fully characterize the clustering algorithm, it is necessary to specify the type of distance between clusters. Hierarchical clustering begins with creating a separate cluster for each sample (personality tuples of agents at the same waypoint). Then, it merges clusters with the smallest distance between them at each pass. After comparing different distances, we selected the Complete Linkage since it shows the best performance on clustering personality tuples with cosine similarity. More precisely, Fig. 2 shows the results for comparing different distances. Since the chosen similarity measure is S_C, we cannot use Ward's MNVAR linkage, even if the clustering is meaningful and clear. This is because Euclidean distance is mandatory for Ward's MNVAR linkage, and thus, cosine similarity cannot be used. Single and average linkages, instead, do not cluster P properly: from the figure, it can be seen that all tuples fall into a single purple cluster, and the algorithm is not able to divide data. It is worth mentioning that Fig. 2 visualizes the clustering results by mapping the five-dimensional space used for p tuples into the three-dimensional space using Principal Component Analysis (PCA), a popular technique for analyzing datasets containing a high number of dimensions/features per observation, preserving the maximum amount of information and enabling the visualization of multidimensional data [56]. Every given sample is mapped by the method to a predefined number of principal components, where each component is a linear combination of features characterizing the sample that explains the most variance in the dataset. On our uniformly sampled dataset of p tuples, PCA shows results similar to other methods, like t-SNE [77] and UMAP [45], which aim to find non-linear mappings from the original to the target space. This fact indicates that distances between data points in the five-dimensional space are irrelevant, and cosine similarity is a valid choice.

The early stopping criterion of the clustering algorithm is given by the number of people in the chosen formation to fill. Formations are taken from the work by Karamouzas and Overmars [36] and represent the most common walking patterns found in their research. In our system, the following formations in Fig. 3 are considered and randomly chosen at the start of the clustering algorithm: 3-people-diagonal queue Fig. 3.a, V-shaped Fig. 3.b, queue-like couple Fig. 3.c , horizontal couple Fig. 3.d, square-shaped Fig. 3.e, 3-people queue Fig. 3.f. Some examples of formations moving in the VE are visible in the video provided as supplemental material.

The clustering algorithm stops when, after a merge pass, at least one cluster in the cluster list has enough members to fill all the formation slots. If more than one cluster meets the criteria, the system chooses the first in the list. If there are more members than needed, we pick them in chronological merge order.

Finally, after clustering, we use the within-cluster cosine similarity (defined as the average cosine similarity between members of the cluster) to decide how long a formation will hold. The heuristic used to disband formations is based on [31]: the more friends have in common, the stronger and durable the bonds are. For instance, if the average cosine similarity of a group cluster tends to 1, personalities in that cluster will be very similar; hence the time the formation will hold is maximum (the cohesion time of a formation is proportional to the average similarity of the formation, and it is selected in interval $[90;300[$ seconds in our study).

C. Agents' Decision-Making and Navigation

Every agent uses deterministic finite state machines (DFSMs) to make decisions. Whenever required, agents use their DFSMs to determine the next state to reach, knowing the current one and the options they have at their disposal.

Movement and pathfinding in our system are managed by default A* algorithm with RVO [76] collision avoidance. Locations influence how agents move and interact. Agents account for an attraction parameter representing a weight in the random choice of a location. The algorithm makes every agent follow the graph topology of the VE based on an "attractiveness decay" process to avoid recently visited waypoints. When an agent chooses a new pathfinding goal, the process starts by getting a set H of available waypoints/destinations selected based on the context and using one of the two strategies.

a) Random strategy: . A waypoint in the map topology is randomly selected, ignoring proximity relations described in the location graph L, among all the possible nodes of L. The chosen waypoint is the new pathfinding goal for the considered agent. The chances that the random strategy is applied are based on the openness value associated with each agent ($p(o)$). Here, $p(o)$ is used as the probability to perform the random selection, and H becomes the set of all the possible nodes of L, with attractiveness lowered based on the decay function.

b) Greedy strategy: . The attractiveness of neighbor waypoints is lowered based on how recently the agent has visited them. Hence, this strategy lowers the chances for an agent to start looping or going back. To keep track of the recently visited waypoints in the VE, we used a last appearance record (LAR) data structure. The LAR maintains a time window of hops through the topology of the map of fixed length $|LAR|$. Waypoints visited more than $|LAR|$ hops ago are not influenced by the attractiveness decay procedure. A waypoint X's actual attractiveness (a_X) after decay is computed as:

$$a'_X = a_X \cdot (1 - f\_decay(l_X))$$

where f_decay (decay function) can be modified and represents the relative amount to subtract from a location's basic attractiveness value (definition domain of f_decay is $[0;1]$), l is a fraction of $|LAR|$ representing the destination's last appearance index in the LAR divided by $|LAR|$. In our system, f_decay follows an exponential trend as in Fig. 4. The agent uses the newly computed destination weights to choose the next pathfinding goal among the neighbors.

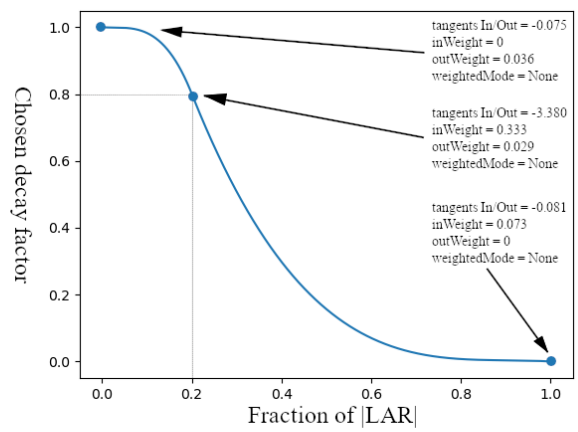


Fig. 4: Attractiveness decay curve used in our system (drawn and defined using the animation curve window in Unity; tangents and weights data are indicated for the three keyframes composing the curve for reproducibility). The curve is evaluated for $|LAR|$ fractions i.e., the destination's last appearance indexes divided by $|LAR|$. The evaluation returns the relative amount to subtract from a location's basic attractiveness value.

When referring to a waypoint choice, the algorithm considers the radius of a waypoint. Waypoints are treated as circular areas in which agents can uniformly sample a coordinate as the actual goal of their pathfinding. This is to avoid several agents standing very close to the center, causing a visually unrealistic effect, especially in stop waypoints. The problem of uniformly sampling points inside a circular shape is resolved using the inverse transform sampling technique. Inverse transform sampling is a method for pseudo-random number sampling from any probability distribution given its cumulative distribution function.

D. Externalizing Personality Through Behaviors

The system translates personality into visible behaviors in the simulation. We associated the five OCEAN dimensions $P(O),\ldots,P(N)$ with behaviors visualized through specific animations. Animations convey a current mood, but we assume that a personality is characterized by a set of behaviors [1] that recur in time. Based on how frequently and with what magnitude an individual expresses a certain behavior, one can deduce a personality that suits the individual. For example, an agent with high $p(n)$ (neuroticism or emotional instability) tends to be more prone to negative emotions such as stress, anxiety, or depression [33].

The proposed method is implemented by relying on weighted randomness and has a direct correspondence to the agent's p tuple. After a random amount of time (set in interval $[5;15[$ seconds in our study), the system decides how to externalize the internal personality for every agent idling at a destination. A pair of animations for each dimension of P represents the opposite behaviors corresponding to lower/higher extremes of an OCEAN dimension score. When the externalization happens, from a personality tuple p of a considered

Dimension polarity	Description
$P(C)$ - High	[*Looking*, Fig. 5.a] Looks in the distance, paying attention to the surroundings in order to give an impression of trustworthiness.
$P(C)$ - Low	[*Searching pockets*, Fig. 5.b] Searches for lost items in the pockets in order to give an impression of unreliability.
$P(E)$ - High	[*Waving*, Fig. 5.c] Tries to get someone's attention by shaking hands in the air. Would like to be noticed.
$P(E)$ - Low	[*Texting while standing*, Fig. 5.d] Uses a mobile phone. Types frequently, even when in a group with other people and does not participate in the social context, showing an enclosing behavior in order to convey intraversion.
$P(A)$ - High	[*Agreeing strong*, Fig. 5.e] Uses body language to convey an amenable attitude. Stretches out and raises the arms, with open hands, to agree with someone.
$P(A)$ - Low	[*Angry*, Fig. 5.f] Looks angry and keeps the arms crossed while rapidly stamping a foot on the ground.
$P(N)$ - High	[*Sad idle*, Fig. 5.g] Looks at the ground, swinging a leg and the whole body in order to give a sad, tired, or depressed impression. The body is slightly hunchbacked.
$P(N)$ - Low	[*Happy idle*, Fig. 5.h] The head is high and lightly swings in a similar way when listening to music to give an impression of liveliness or joy. The body is overall relaxed and straight.

Tab. I: Associations between OCEAN dimensions and agents' animations. Animations were downloaded from https://www.mixamo.com/. The name of the animation, as shown on the website, is displayed in brackets.

agent at a stop waypoint, the i-th OCEAN component is selected with uniform probability to be externalized. Based on the score $p(i)$, the process uses weighted random sampling to decide whether to externalize the lower or higher extreme of i with the related animation.

Table I shows the associations between animations and OCEAN dimension extremes used in the system. $P(O)$ is used for VE navigation as described in Sec. III-C and is not mapped to any pair of animations. All the animations are shown in Fig. 5 and in the video we provide as supplementary material.

For example, if the system selects $P(N)$ (neuroticism component) as the OCEAN dimension to externalize for an agent with a value $p(n) = 0.8$, this score means that the agent is prone to stress, anxiety, or depression. More precisely, we interpret the number as the probability, in every moment, of finding the person depressed. Hence, for our heuristic interpretation of this value, with probability 0.8, the agent shows sad/depressed/anxious behavior, and with 0.2, a happy one (opposite of the spectrum) when the externalization happens. In this case, using weighted random sampling with weights $w = \langle 0.8, 0.2 \rangle$ that sum up to 1, it is equivalent to checking if a random number in the interval $[0,1[$ is less than $p(n)$ to play the sad idling animation, or else play the happy one.

We acknowledge that this way of externalizing a component of OCEAN is a simplification. In fact, when externalizing the chosen i-th component, we assume every other component to be average or, at least, to have a negligible impact on the manifestation of the i-th dimension. For example, we could be externalizing $p(n)$ with a happy animation, but the happy animations of an agent with low or high $p(e)$ might be different.

IV. User Study

The study aims to understand if the proposed crowd simulation system with personality contributions is better than a basic version of the same system that does not use personality. More precisely, the personality-based simulation is implemented as described in Sec. III. The basic simulation is a control condition that ignores all contributions coming from the OCEAN model. The basic simulation works by reading the map topology from the same files but requires only a subset of the persistent data utilized by the personality-based one. Relations between agents are ignored, and clustering using p tuples is not active. Agents always follow the map topology (they cannot "random select") since their $p(o)$ is not considered. Agents during the basic simulation stop at stop locations, but all of them play the same generic idle animation when they are not walking. Walking animations are the same for every agent in both simulations.

The study followed a within-subjects design: all participants tested both conditions, i.e., they experienced the crowd simulation in iVR with personality contributions to agents' behavior (P condition) and with no personality contribution (NP condition) in a counterbalanced order.

The study proposal was approved by the Institutional Review Board of the Department of Mathematics, Computer Science, and Physics of the University of Udine.

A. Hypotheses

Since the goal of the proposed system is to simulate a more plausible and realistic crowd behavior overall, we selected a set of measures that indicate the fulfillment of the goal. Since the system aims to provide more realistic and plausible agents, we tested the hypothesis of an increase in the plausibility of the agents when participants try the personality-based experience (H1). Molina et al. [47] find that, when humanoid characters are used, a variety of animations leads to more realism than using only locomotion animations. In our study, the personality-based simulation uses a variety of animations (including locomotion), while the basic simulation uses only locomotion ones. Therefore, an increased experienced realism (H2) could be expected. If plausibility and/or experienced realism are increased, social presence may be enhanced (H3). It is known that differences in presence (experienced realism is a component of presence [62]) can lead to differences in the emotional state [34], [43] and emotional affect can be higher when agents perform animations [80]. Therefore, differences in users' emotional valence, arousal, and/or intensity of positive and negative emotions between the personality-based simulation and basic simulation are expected (H4).

Fig. 5: Agents' behavioral animations to externalize OCEAN personality; from left to right by couples: animations represent high and low values for conscientiousness, extraversion, agreeableness, and neuroticism.

B. Materials

The evaluation of our system used the same iVR setup for every user, and every test was performed at the "SMACT 3 - M15 - Lab Village Module" of the University of Udine. The personality-based and the basic simulations were implemented in Unity, version 2022.3.6f1. The iVR setup used a PC equipped with a 2.50 GHz Intel i9-11900 processor, 32 GB RAM, an NVidia GTX 3090 graphic card, and a Meta Quest Pro headset connected through AirLink.

All the animations for behavior externalization (see Tab. I), idle, and walking were downloaded from Mixamo (https://www.mixamo.com/). The VE and the agents (https://assetstore.unity.com/packages/3d/environments/urban/polygon-city-low-poly-3d-art-by-synty-95214) are the same for both the personality-based and basic simulations.

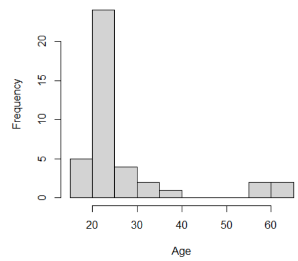
All questionnaires are translated into the language spoken by the participants (Italian) and are administered using the PsyToolkit [68], [69] online platform.

Code snippets of components internal to the system will be available by request to the corresponding author.

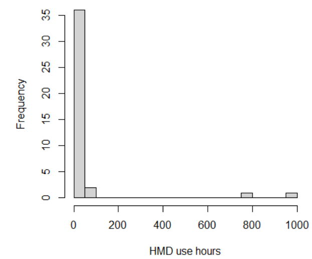
C. Participants

An a priori power analysis was conducted using G*Power version 3.1.9.7 [24] to determine the minimum sample size required to test the study hypotheses. Results indicated that the required sample size to achieve 80% of power in detecting a medium effect, at a significance criterion of $\alpha = 0.05$, was $N = 34$ for paired T-test. Thus, the obtained sample size of $N = 40$ is adequate to test the study hypotheses in case of a medium or larger effect size.

To get general data regarding the participants, a demographic questionnaire (gender, age, and HMD usage in hours) was used, which helped characterize our sample of 40 participants (25M, 15F). The two inclusion criteria were 1) being at least 18 years old (adult age in Italy to sign the participation consent document personally) and 2) not suffering from epilepsy. The age of the participants ranged from 18 to 65 years old ($M = 26.95, SD = 12.28$). Most participants (33/40) had less than 5 hours of experience in total with iVR headsets. The age of the participants and usage of iVR headsets were found to be distributed according to the histograms shown in Fig. 6.



(a) Age frequency.



(b) Hours of experience with iVR headsets.

Fig. 6: Frequencies of the measured data in the demographic questionnaire.

D. Measures

a) Plausibility: The plausibility of agents is measured by administering the VHPQ – Virtual Human Plausibility Questionnaire to participants. VHPQ [42] is used to measure if the rendered agents in the VE are plausible from the point of view of their behavior and appearance and if their behavior matches the visual style of the VE. VHPQ comprises 11 items divided into 2 subscales: appearance behavior plausibility (7 items) and match with the VE (4 items). Each item is associated with a 7-point Likert scale; all values lie in the interval $[1;7]$. In VHPQ, both subscales follow the mapping that assigns higher values to higher plausibility or coherence of the virtual human/crowd.

b) Experienced realism: Experienced realism for each participant is measured by administering the IPQ – Igroup Presence Questionnaire. IPQ [62] is a widely used questionnaire to evaluate perceived presence in a VE. It includes one single item directly related to general presence and three subscales: spatial presence (5 items), experienced realism (4 items), and involvement (4 items) for a total of 14 items, each one associated with a 7-point Likert scale with values in the range $[0;6]$. The general presence item and every subscale follow the mapping that assigns higher values to higher levels of the represented construct (higher general presence, spatial presence, experienced realism, or involvement). Furthermore, IPQ includes a scale that yields a total score for presence with values in the range $[0;6]$; this scale also follows the mapping that assigns higher values to higher levels of its represented construct. We will measure all (sub)scales of presence with IPQ (not only experienced realism) because Pelechano et al.

[58] have found that presence can be used to assess the performance of an immersive VR crowd system provided that egocentric features are present (in our system we do have Shaking, Continuous movement, Overlapping and Pushing for both simulations and Communication only in the personality-based one).

c) Social presence: Social presence is measured by administering participants the SPS – Social Presence Survey [4]. It comprises 5 items, each one associated with a 7-point Likert scale. The only scale of this questionnaire, social presence, can assume values in the interval $[5;35]$ and is directly mapped to its represented construct: higher values translate into higher perceived social presence.

d) Emotional affect: Emotional affect is measured by administering participants two questionnaires: SAM and PANAS-S. SAM – Self Assessment Manikin [12] measures participants' pleasure, arousal, and dominance after each experience. SAM is composed of 3 items, each one corresponding to a construct: pleasure (also known as valence), arousal, and dominance. This image-based questionnaire measures the three constructs of the affective reaction to stimuli. Each item value lies in the interval $[1;9]$. Higher values of the first item correspond to negative valence of perceived emotions, and higher values of the second item correspond to less arousal. Higher values of the third scale correspond to higher emotion control. PANAS-S – Positive and Negative Affect Schedule (Short) [16] is a questionnaire used to measure the tendency of an individual to experience positive and negative emotions (PA positive affect, NA negative affect) in response to an experience. Unlike SAM, PANAS-S can be used to assess self-reported positive and negative affect. It comprises 20 items, each related to a specific emotion and associated with a 5-point Likert. In this questionnaire, two subscales are measured by 10 items each: negative affect and positive affect, with values in the interval $[5;50]$. Both subscales follow the mapping that assigns higher subscale values to higher positive affect or negative affect depending on the subscale considered.

E. Experimental Procedure

The experimental procedure consists of the following steps for each user:

1) Participants are informed that the study involves testing a crowd simulation system. They are assured that they can stop the experiment at any time if they feel uncomfortable or choose not to continue, without needing to provide a reason. The experimenter also explains that the study requires participants to complete some questionnaires, which are anonymized employing an alphanumeric anonymous ID to protect their privacy. After providing written informed consent, participants completed the demographic questionnaire.
2) The experimenter explains the task the participant has to perform in the VE. The task is to explore the VE by following a predefined path and looking around. The path is indicated by a semi-transparent yellow arrow, low-floating on the floor: the arrow always points to the next destination belonging to the path. Destinations are represented as semi-transparent yellow semi-spheres with a diameter of approximately 3 meters and a yellow semi-transparent human-shaped figure inside (see the provided video as supplemental material). When the participant reaches the inner area of the semi-sphere, a sound is played, and the arrow disappears. At this point, the participant has some time to observe the crowd. When the participant hears the same sound again and the arrow reappears, he or she is invited to move to the next observation point following the direction pointed by the arrow. The experimenter tells the participant that the experience will last 10 minutes.
3) The experimenter describes the environment and the interaction techniques used during the experience. When the application is started, the participant is located in a city. Initially, a virtual keyboard is drawn in front of the participant. The keyboard will be used to type the participant ID. The experimenter explains what button to press on the controllers to type on the keyboard. To move into the virtual city, the participant uses teleportation (see the provided video as supplemental material). The experimenter makes sure to show the participant what button to press in order to activate teleportation. It is made clear that the participants can rotate around themselves, bend, and lean over to have a better view.
4) The experimenter assists the participant in wearing the iVR headset and starts the application. The experimenter also warns the participant to look at their left wrist where a virtual watch is located and tells the participant that the watch tells how much time is remaining. At the end of the countdown, the participant hears a sound and is asked to take off the iVR headset to fill out the questionnaire.
5) The participant is free to familiarize with the teleportation technique in the keyboard scene. When ready, the participant can start the actual experience by typing the identifier and pressing the "ENTER" button on the keyboard.
6) The participant explores the VE for 10 minutes. The task of following the path is given to keep interest in the simulation up for the duration of the single experience. Half of the participants experience the basic simulation; the other half experience the personality-based simulation.
7) At the end of the 10 minutes, the experimenter helps the participant remove the iVR headset. Then, questionnaires are proposed in this order: SAM, PANAS-S, IPQ, VHPQ, and SPS.
8) The experimenter assists the participant in putting on the iVR headset a second time and starts the application. The participant explores the VE for 10 minutes for a second time experiencing the simulation not already performed at point 11. The path in the VE to follow is the same path as the first simulation but reversed.
9) At the end of the 10 minutes, the experimenter helps the participant take off the iVR headset.
10) The participant is invited to fill out the same questionnaires of point 7.

V. RESULTS

Following Norman [52], we analyzed the data from SPS, VHPQ, PANAS, and IPQ questionnaires as well as Spatial Presence, Experienced Realism, and Involvement subscales of IPQ using paired T-tests. For the single items of the SAM questionnaire and the general presence item of the IPQ, we instead used the Wilcoxon signed rank test. All statistical tests were computed using the software JASP [32].

Data in Tab. II summarizes all the results obtained with the questionnaires, organized by subscale, showing the results of the paired T-test or the Wilcoxon signed ranks test performed on the data, and the computed effect size as measured by Cohen's d or r in accordance with Coolican [17].

The resulting differences are not statistically significant for any of the items in the SAM questionnaire and the subscales of PANAS-S questionnaires ($p > 0.05$).

The difference on the subscale IPQ-Experienced Realism reached statistical significance $t(39) = -2.76, p = 0.009$, and Cohen's $d = -0.44$ indicated a medium effect size. IPQ-Experienced Realism reports a higher value for the P condition ($M = 2.33, SD = 1.36$) with respect to the NP condition ($M = 2.01, SD = 1.11$). For all other (sub)scales of IPQ and the general presence item, the resulting differences are not statistically significant ($p > 0.05$).

The difference on the subscale VHPQ-Match to VE reached statistical significance $t(39) = -2.81, p = 0.008$, and Cohen's $d = -0.44$ indicates a medium effect size. VHPQ-Match to VE reports a higher value for the P condition ($M = 5.76, SD = 0.88$) with respect to the NP condition ($M = 5.39, SD = 1.16$). For VHPQ-Plausibility, the resulting differences are not statistically significant ($p > 0.05$).

The difference on the scale SPS-Social Presence reached statistical significance $t(39) = -2.33, p = 0.025$, and Cohen's $d = -0.37$ indicates a medium effect size. SPS-Social Presence reports a higher value for the P condition ($M = 3.85, SD = 1.44$) with respect to the NP condition ($M = 3.52, SD = 1.28$).

VI. DISCUSSION

Considering our hypothesis about plausibility (H1) we discuss the results obtained from VHPQ, where VHPQ-Plausibility is relevant. This subscale measures the plausibility of the appearance and behavior of the rendered agents in the VE and exhibits the expected negative differences (NP-P), although we did not reach a statistically significant result. The introduction of OCEAN personality in crowds with our system may have only a small perceivable effect on the plausibility of appearance and behavior. VHPQ-Plausibility considers both these factors; however, the appearance of agents does not change between the two simulations. The component of VHPQ-Plausibility focusing on appearance can interfere with the behavior one, and only the latter is influenced by the addition of OCEAN in the personality-based simulation of our system. Considering the specific items about behavior plausibility (1, 5, and 11), Wilcoxon signed rank test shows a significant difference only in the scores for item 1 (i.e., "The behavior of the virtual character seemed to be plausible to me" $z = -3.209, p = 0.001$) with a higher score for the personality-based simulation ($M = 4.375, SD = 1.750$) than the basic one ($M = 3.575, SD = 1.693$). The other subscale, VHPQ-Match to VE, showed a statistically significant difference between the two simulations, with a medium effect size. This shows a higher matching between agents and VE in the personality-based simulation. Hence, we conclude that in the personality-based simulation, autonomous agents matched better with the environment they were in, but we cannot confirm H1.

Considering our hypothesis about experienced realism (H2) we discuss the results obtained from IPQ. According to Molina et al. [47], since we have a variety of animations in the personality-based simulation, we expect an increase in experienced realism when autonomous agents show personality. Experienced realism can be measured with the appropriate subscale of IPQ. We reached statistical significance in IPQ-Experienced Realism with a medium effect size. In this subscale, results show a significant increase in experienced realism in the personality-based simulation. Therefore, we can confirm H2. The results of all the other IPQ subscales show a slightly lower value for the personality-based simulation. The difference is never statistically significant, which is desirable because the inclusion of personality does not break the sense of presence, both general and spatial, nor lowers user's involvement. For the latter one, this construct is influenced by interactions between VE and user [18]. Since the task and all the interactions were the same in both simulations, IPQ-Involvement was expected to be the same. Considering the hypothesis about social presence (H3), we reached a statistically significant difference in SPS with a medium effect size. SPS-Social Presence shows an increase in the personality-based simulation as hypothesized: participants felt more socially present in the same VE with agents showing personality. Therefore, we can confirm H3.

Trivedi and Mousas [74] focus on avoidance proximity and find it does not influence social presence but influences experienced realism. Interestingly, the system by Durupinar et al. [21] changes avoidance proximity based on personality but, unfortunately, does not assess social presence and realism. We found differences in both social presence and experienced realism, but avoidance proximity for virtual agents was the same in both simulations. This indicates that our personality-based system should have influenced the differences in social presence and experienced realism.

Considering our hypothesis about emotional affect (H4), we discuss the results obtained from SAM and PANAS-S since both questionnaires measure emotional affect in response to stimuli: according to Jicol et al. [34], since we found significant differences in IPQ-Experienced Realism, that is related to presence, we can expect some differences in valence or arousal of emotions. The SAM questionnaire did not reach significance in any of its items. Still, the item SAM-Arousal shows a lower score (higher arousal) in the personality-based simulation. Focusing on the SAM-Dominance item, results show that in the personality-based simulation, the value is lower than the other simulation on average and close to significance ($p = 0.083$). Lower dominance values correspond to less control over the emotional state. Personality-based agents

Quest.	Subscale	Mean		Std. Dev.		Paired Differences (NP-P)					t/z	p	d/r
		NP	P	NP	P	Mean	Std. Dev.	S.E. Mean	95% C.I. of differences Lower	Upper			
SAM	Pleasure	3.15	2.88	1.53	1.68	0.28	1.41	0.22	-0.18	0.73	1.09	0.271	0.12
	Arousal	5.03	4.80	1.86	2.03	0.23	1.39	0.22	-0.22	0.67	1.05	0.285	0.12
	Dominance	6.33	5.85	1.86	1.86	0.48	1.55	0.25	-0.02	0.97	1.72	0.083	0.19
PANAS-S	Pos. Affect	30.40	31.27	9.05	8.92	-0.88	5.15	0.81	-2.52	0.77	-1.07	0.289	-0.17
	Neg. Affect	48.17	48.20	2.83	2.20	-0.03	2.04	0.32	-0.68	0.63	-0.08	0.939	-0.01
IPQ	General Pres.	4.40	4.17	1.43	1.34	0.23	1.12	0.18	-0.13	0.58	1.35	0.168	0.15
	Spatial Pres.	4.23	4.19	1.18	1.17	0.04	0.92	0.14	-0.25	0.34	0.31	0.758	0.05
	Realism	2.01	2.33	1.11	1.36	-0.32	0.73	0.12	-0.55	-0.08	-2.76	**0.009**	-0.44
	Involvement	3.65	3.62	1.28	1.24	0.03	1.14	0.18	-0.33	0.39	0.17	0.863	0.03
	Tot. Presence	3.44	3.49	0.96	1.05	-0.05	0.65	0.11	-0.26	0.16	-0.48	0.640	-0.08
VHPQ	Plausibility	3.86	4.10	1.26	1.40	-0.23	0.97	0.15	-0.54	0.08	-1.52	0.137	-0.24
	Match to VE	5.39	5.76	1.16	0.88	-0.36	0.82	0.13	-0.62	-0.10	-2.81	**0.008**	-0.44
SPS	Social Pres.	17.57	19.25	6.38	7.22	-1.68	4.54	0.72	-3.13	-0.22	-2.33	**0.025**	-0.37

Tab. II: Results of the questionnaires. Paired T-test or paired Wilcoxon signed-ranks test on questionnaire subscales or items respectively. In every case, the reported differences are the subscale values after the basic simulation (NP) minus the subscale values after the personality-based one (P). For every T-test, $df = 39$. Column "d/r" shows effect sizes either as Cohen's d for T-tests or as r for Wilcoxon signed-ranks tests. Statistically significant results are flagged in bold.

may be the cause of lower SAM-Dominance in the personality-based simulation because of agents' increased realism. Agents could have been perceived as more similar to real humans, and the control we can exercise on our emotions caused by another human is limited: we cannot know how the other person will act or react to our actions or stimuli. This is in line with previous work showing that social-cognitive factors such as beliefs about control influence the emotional regulation of the individual [51], [53].

The results did not show significance in any of the subscales of the PANAS-S questionnaire either. We expected significant differences in PANAS-S subscales due to possibly different emotional affect between simulations. Means for subscale PANAS-S-Positive Affect show a slight increase for positive emotions in the personality-based simulation. This is in accordance with SAM-Pleasure: participants showed a more positive valence during the latter (a lower value for this subscale means higher pleasure/positive valence of emotions). PANAS-S-Negative Affect values are roughly the same for both simulations, with a negligible increment (0.03) when experiencing the personality-based simulation. Since neither the SAM nor the PANAS-S subscales showed a significant difference, we cannot confirm H4, despite agents in our personality-based simulation could perform animations for both positive and negative emotions (e.g., for low and high neuroticism as described in Tab. I), while only idle and walking was displayed in the basic simulation, and animated characters led to higher emotional affect compared to static characters in Wu et al. [80]. A possible explanation for this could be the lack of facial expressions in our agents since facial expressions play an important role in conveying emotions [5].

Overall, we can conclude that adding personality to the simulation enhanced social presence, experienced realism, and matching with the VE of the crowd represented.

A. Limitations of the System

The following main limitations concerning our crowd simulation system were identified.

First, the test was conducted comparing a suite of personality-based enhancements. At the moment of writing, we cannot know what functionality was the most important or the most effective.

Second, our VE is low-polygonal and stylized, with low agent variation in terms of appearance. Graphics can impact the participant experience by varying the sense of presence and possibly other factors, including emotional affect [78]; thus, the overall results might differ in a photorealistic setting. It is known in the literature that the materials, drawing style, and shapes of humanoid agents influence realism, appeal, eeriness, familiarity, and expression intensity [83], [85]. In stylized settings like ours, it is reasonable to assume that the user might be keener on accepting lower levels of simulation accuracy. Hence, the stylized graphic style may have influenced the perception of agents' low variety and behavioral animations.

The third limitation concerns the way personality is externalized. We did not include other modalities of expressing OCEAN that might have more impact on the participant experience. Several examples of previous attempts to externalize agent personality can be found in the literature, such as dialogues [41], facial expressions [13], hand motion [79], body motion [20], or combinations of the previous [67], but all of them concern a single agent. All mentioned work could be extended to crowds in future research. We used animations and tried to relate them to an underlying personality. Moreover, the mapping between personality and behavioral animations in the simulation is obtained from some possible characterizations of OCEAN dimensions, but it might not be the best possible. Thus, a different mapping can potentially change the outcomes of the study. Moreover, in our system, we did not consider possible relations between different OCEAN dimensions.

Fourth, the user study presented in Sec. IV was conducted on people living in Italy, who may perceive and react to personality differently with respect to people living in other cultures. Therefore, to generalize the results, the study should be repeated with additional, more diverse population samples. The population specificity also concerns the agents: in both simulations, they represented a typical Western country population. Users might expect different behaviors from agents representing people from other areas of the world.

Fifth, the user study employed self-assessment questionnaires to measure the psychological constructs that were useful for verifying the hypotheses. While this approach is well-established and commonly applied in user studies within the human-computer interaction setting, it inherently relies on subjective measurements. Such reliance can present challenges, including potential variability in responses due to participant fatigue or lack of patience during the study. To examine this possibility, we conducted a supplemental analysis. In particular, for every measured subscale in each questionnaire and for both simulations (P and NP) separately, we extracted the questionnaire subscale values for users who experienced a specific simulation first and the subscale values for users who experienced the same simulation second. Then, statistical tests were performed as independent samples T-tests or Mann-Whitney-U-tests based on normality and homogeneity tests (Shapiro-Wilk and Levene tests, respectively). This analysis found no statistically significant differences ($p \geq 0.05$) due to the order in which conditions were tried (and possible accumulated fatigue), except for IPQ-GeneralPres for simulation NP and IPQ-SpatPres for simulation NP. Notably, for these subscales of IPQ possibly affected by fatigue, no statistically significant difference was found in the comparison between P and NP. Hence, no sufficient statistical evidence exists showing that fatigue influenced the results. Nevertheless, subjective measures remain a limitation, as they may not fully capture the complexity of psychological constructs. A promising direction for future work is to complement self-assessment questionnaires with indirect, objective measures, such as physiological signals. For instance, ECG signals have been explored in limited work on the sense of presence in virtual environments [66]. However, employing such methods presents significant challenges due to the inherent difficulty in interpreting physiological responses in virtual contexts. Addressing these challenges will be an important step in enhancing the robustness and generalizability of future studies, as outlined in the subsequent section.

All the aforementioned limitations point to the need for new studies, which are discussed in the next section and outline crucial and interesting perspectives for future research.

B. *Performance Considerations*

It is challenging to replicate social and real-world interactions between humans. Both are complex to simulate due to their variety and their belonging to a social context that cannot be easily modeled. This is even more evident in an iVR setting. Computational constraints limit current VR technology; therefore, the performance of the tested system is of concern. Thanks to the data structures implemented, the personality-based simulation and the basic one are comparable from the performance point of view, with both simulations averagely running at more than 70 frames per second with the hardware described in Sec. IV-B. Indeed, the necessary rendering passes impact performance far more than personality simulation. All the functions described for personality-based simulation can be made asynchronous (especially computationally heavy ones like clustering), and we did not apply asynchronous pathfinding. The latter is another option to boost performances even further, combined with ambient occlusion of both agents and VE. It is important to mention that asynchronous code execution, in this case, would negate the possibility of having reproducibility of a session: load on the system coming from other background processes, also external to the application, would influence the start and termination of functions and methods. We did not use asynchronous code for this specific reason during the user study.

VII. Conclusion and Future Work

In this paper, we described the design and evaluation of a system for crowd simulation exploiting a well-known model in psychology to represent the personality of individual agents. We evaluated our system on a sample of 40 participants. We can conclude that personality-based simulation was perceived as an effective enhancement of social presence and realism. The study also shows that introducing personality increases the matching between virtual agents and the VE they were in. Our system is extensible to other relations and behaviors. IVR is a powerful tool for studying how users react to artificial stimuli that mimic the real ones. At the same time, it is known that virtual crowds can be used to study real human behavior [7], [50] and iVR can help recreate realistic scenarios. Our study shows that the described approach to personality-based crowd simulation is a viable solution to increase social presence and realism when shifting the setting from a single-agent to a multi-agent VE, considering the computational constraints of iVR technology. Future work can include:

- Addition of persistent relations in agent graph A as "facts" already known by the simulation. These persistent relations can influence the evolution of agents together with runtime-generated relations. For example, plausible relationships could be family relations, colleagues, friends, etc.
- Improving the personality externalizer system by:
  - Including interactions between the users and the agents to increase the ways in which personality can be externalized instead of passive observation of the crowd (see also Sec. VI-A).
  - Introducing personality-influenced facial expressions and more realistic animations to convey personality, also when walking, or to account for relations between different OCEAN dimensions when externalizing personality.
  - Focusing on the most visible and perceivable characteristics when externalizing personality. This would be an opportunity to include generative AI techniques

that correlate a personality to different ways of expressing it (e.g., clothing [48]).

- Experiment with different mappings between personality models and behaviors (possibly different from the ones considered in this study) in the simulation to test which one is the most effective in representing convincing crowds.
- Explore relations between graphics and perception of crowds in order to understand the most salient changes of agents that lead to differences in perceiving them. This would be useful when computational resources are scarce and the application demands an appropriate level of variety.
- Further validate the positive results found by this study on larger and more diverse samples of users since our sample was composed entirely of people living in Italy and perception of personality may vary between different cultures. Additionally, different settings can be tested, specifically regarding graphical realism and agent variety. Hence, the present work could be extended by systematically exploring different important directions:
  - Employing representative characters from other populations and testing on users living in other countries.
  - Conducting an in-depth examination of how individuals' personality dimensions specifically influence user perceptions of agents based on their cultural background.
  - Assessing the system's effectiveness in a more photorealistic setting and with higher agent visual variability (also in relation to the fourth bullet point).
- Implement more objective measurements of the psychological constructs used in this study. The study employed self-assessment questionnaires to measure the psychological constructs, as described in Sec. IV-D. Since self-assessment questionnaires record subjective measurements, a future possibility is to employ indirect measures to obtain more objective results of the psychological constructs considered in our study. Implementing such measures involves performing several studies focusing on the complex interpretation and relations between physiological signals or similar indirect measurements and constructs classically measured via self-assessment questionnaires.
- Apply personality-based crowd simulation in emergency scenarios training, where personality could influence the reaction of individual autonomous agents to the emergency. For example, this could be applied to extend terror attack simulations such as those proposed by Sioni and Chittaro [15], also to train learners in managing a crowd during a disaster.

References

[1] APA Dictionary of Psychology. https://dictionary.apa.org/.

[2] A. L. Alexander, T. Brunyé, J. Sidman, and S. A. Weil. From Gaming to Training: A Review of Studies on Fidelity, Immersion, Presence, and Buy-in and Their Effects on Transfer in PC-Based Simulations and Games. 2005.

[3] G. Aschwanden, Gideon, J. Halatsch, Jan, G. Schmitt, and Gerhard. Crowd Simulation for Urban Planning. *Architecture in Computro [26th eCAADe Conference Proceedings / ISBN 978-0-9541183-7-2] Antwerpen (Belgium) 17-20 September 2008, pp. 493-500*, Jan. 2008.

[4] J. Bailenson, J. Blascovich, A. Beall, and J. Loomis. Interpersonal Distance in Immersive Virtual Environments. *Personality & social psychology bulletin*, 29:819–33, Aug. 2003. doi: 10.1177/0146167203029007002

[5] L. F. Barrett, M. Lewis, and J. M. Haviland-Jones. *Handbook of Emotions*, p. 469. Guilford Publications, 2016.

[6] A. E. Başak, U. Güdükbay, and F. Durupınar. Using real life incidents for creating realistic virtual crowds with data-driven emotion contagion. *Computers & Graphics*, 72:70–81, May 2018. doi: 10.1016/j.cag.2018.02.004

[7] B. Bendiksen, N. Lin, J. Kim, and F. Durupinar. Assessing Human Reactions in a Virtual Crowd Based on Crowd Disposition, Perceived Agency, and User Traits. *ACM Trans. Appl. Percept.*, p. 3658670, Apr. 2024. doi: 10.1145/3658670

[8] F. Berton, F. Grzeskowiak, A. Bonneau, A. Jovane, M. Aggravi, L. Hoyet, A.-H. Olivier, C. Pacchierotti, and J. Pettre. Crowd Navigation in VR: Exploring Haptic Rendering of Collisions. *IEEE Trans. Visual. Comput. Graphics*, 28(7):2589–2601, July 2022. doi: 10.1109/TVCG.2020.3041341

[9] F. Berton, L. Hoyet, A.-H. Olivier, J. Bruneau, O. Le Meur, and J. Pettre. Eye-Gaze Activity in Crowds: Impact of Virtual Reality and Density. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 322–331, Mar. 2020. doi: 10.1109/VR46266.2020.00052

[10] J. Block. A contrarian view of the five-factor approach to personality description. *Psychological Bulletin*, 117:187–215, 1995. doi: 10.1037/0033-2909.117.2.187

[11] T. Bosse, R. Duell, Z. A. Memon, J. Treur, and C. N. van der Wal. Agent-Based Modeling of Emotion Contagion in Groups. *Cogn Comput*, 7(1):111–136, Feb. 2015. doi: 10.1007/s12559-014-9277-9

[12] M. M. Bradley and P. J. Lang. Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1):49–59, Mar. 1994. doi: 10.1016/0005-7916(94)90063-9

[13] S. Chin, C. Y. Lee, and J. Lee. An automatic method for motion capture-based exaggeration of facial expressions with personality types. *Virtual Reality*, 17:219–237, 2013.

[14] L. Chittaro and M. Serra. Behavioral Programming of Autonomous Characters based on Probabilistic Automata and Personality.

[15] L. Chittaro and R. Sioni. Serious games for emergency preparedness: Evaluation of an interactive vs. a non-interactive simulation of a terror attack. *Computers in Human Behavior*, 50:508–519, Sept. 2015. doi: 10.1016/j.chb.2015.03.074

[16] L. A. Clark and D. Watson. The PANAS-X: Manual for the Positive and Negative Affect Schedule - Expanded Form, 1994. doi: 10.17077/48vt-m4t2

[17] H. Coolican. *Research Methods and Statistics in Psychology*. Routledge, 7 ed., Nov. 2018.

[18] H. Desurvire and C. Wiberg. User Experience Design for Inexperienced Gamers: GAP – Game Approachability Principles. pp. 131–147. Dec. 2010. doi: 10.1007/978-1-84882-963-3_8

[19] F. Durupinar, U. Gudukbay, A. Aman, and N. I. Badler. Psychological Parameters for Crowd Simulation: From Audiences to Mobs. *IEEE Trans. Visual. Comput. Graphics*, 22(9):2145–2159, Sept. 2016. doi: 10.1109/TVCG.2015.2501801

[20] F. Durupinar, M. Kapadia, S. Deutsch, M. Neff, and N. I. Badler. PERFORM: Perceptual Approach for Adding OCEAN Personality to Human Motion Using Laban Movement Analysis. *ACM Trans. Graph.*, 36(1):1–16, Feb. 2017. doi: 10.1145/2983620

[21] F. Durupinar, N. Pelechano, J. M. Allbeck, U. Gudukbay, and N. I. Badler. How the Ocean Personality Model Affects the Perception of Crowds. *IEEE Comput. Grap. Appl.*, 31(3):22–31, May 2011. doi: 10.1109/MCG.2009.105

[22] H. J. Eysenck. Four ways five factors are not basic. *Personality and Individual Differences*, 13(6):667–673, June 1992. doi: 10.1016/0191-8869(92)90237-J

[23] S. B. G. Eysenck and H. J. Eysenck. The place of impulsiveness in a dimensional system of personality description. *British Journal of Social and Clinical Psychology*, 16(1):57–68, 1977. doi: 10.1111/j.2044-8260.1977.tb01003.x

[24] F. Faul, E. Erdfelder, A.-G. Lang, and A. Buchner. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2):175–191, May 2007. doi: 10.3758/BF03193146

[25] R. M. Favaretto, L. Dihl, S. Raupp Musse, F. Vilanova, and A. Brandelli Costa. Using Big Five Personality Model to Detect Cultural Aspects in Crowds. In *2017 30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pp. 223–229, Oct. 2017. doi: 10.1109/SIBGRAPI.2017.36

[26] R. M. Favaretto, P. Knob, S. R. Musse, F. Vilanova, and Â. B. Costa. Detecting personality and emotion traits in crowds from video sequences. *Machine Vision and Applications*, 30(5):999–1012, July 2019. doi: 10.1007/s00138-018-0979-y

[27] L. Fu, W. Song, W. Lv, and S. Lo. Simulation of emotional contagion using modified SIR model: A cellular automaton approach. *Physica A: Statistical Mechanics and its Applications*, 405:380–391, July 2014. doi: 10.1016/j.physa.2014.03.043

[28] L. R. Goldberg. The structure of phenotypic personality traits. *American Psychologist*, 48:26–34, 1993. doi: 10.1037/0003-066X.48.1.26

[29] S. J. Guy, S. Kim, M. C. Lin, and D. Manocha. Simulating heterogeneous crowd behaviors using personality trait theory. In *Proceedings of the 2011 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pp. 43–52. ACM, Vancouver British Columbia Canada, Aug. 2011. doi: 10.1145/2019406.2019413

[30] W. He, G. Pan, X. Wang, and J. X. Chen. Real-time crowd formation control in virtual scenes. *Simulation Modelling Practice and Theory*, 122:102662, Jan. 2023. doi: 10.1016/j.simpat.2022.102662

[31] C. Izard. Personality similarity and friendship. *The Journal of Abnormal and Social Psychology*, 61(1):47–51, 1960. doi: 10.1037/h0040056

[32] JASP Team. JASP (Version 0.18.3)[Computer software], 2024.

[33] B. F. Jeronimus, H. Riese, R. Sanderman, and J. Ormel. Mutual reinforcement between neuroticism and life experiences: A five-wave, 16-year study to test reciprocal causation. *Journal of Personality and Social Psychology*, 107(4):751–764, 2014. doi: 10.1037/a0037009

[34] C. Jicol, C. H. Wan, B. Doling, C. H. Illingworth, J. Yoon, C. Headey, C. Lutteroth, M. J. Proulx, K. Petrini, and E. O'Neill. Effects of Emotion and Agency on Presence in Virtual Reality. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–13. ACM, Yokohama Japan, May 2021. doi: 10.1145/3411764.3445588

[35] M. Kapadia, A. Shoulson, F. Durupinar, and N. I. Badler. Authoring Multi-actor Behaviors in Crowds with Diverse Personalities. In S. Ali, K. Nishino, D. Manocha, and M. Shah, eds., *Modeling, Simulation and Visual Analysis of Crowds: A Multidisciplinary Perspective*, The International Series in Video Computing, pp. 147–180. Springer, New York, NY, 2013. doi: 10.1007/978-1-4614-8483-7_7

[36] I. Karamouzas and M. Overmars. Simulating and Evaluating the Local Behavior of Small Pedestrian Groups. *IEEE Trans. Visual. Comput. Graphics*, 18(3):394–406, Mar. 2012. doi: 10.1109/TVCG.2011.133

[37] A. Koilias, C. Mousas, and C.-N. Anagnostopoulos. 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):e1963, 2020. doi: 10.1002/cav.1963

[38] K. Lee and M. C. Ashton. The HEXACO Model of Personality Structure. In V. Zeigler-Hill and T. K. Shackelford, eds., *Encyclopedia of Personality and Individual Differences*, pp. 1932–1936. Springer International Publishing, Cham, 2020. doi: 10.1007/978-3-319-24612-3_1227

[39] C. K. Lim, K. L. Tan, A. A. Zaidan, and B. B. Zaidan. A proposed methodology of bringing past life in digital cultural heritage through crowd simulation: A case study in George Town, Malaysia. *Multimed Tools Appl*, 79(5-6):3387–3423, Feb. 2020. doi: 10.1007/s11042-019-07925-2

[40] T. Liu, Z. Liu, M. Ma, T. Chen, C. Liu, and Y. Chai. 3D visual simulation of individual and crowd behavior in earthquake evacuation. *SIMULATION*, 95(1):65–81, Jan. 2019. doi: 10.1177/0037549717753294

[41] F. Mairesse and M. Walker. PERSONAGE: Personality Generation for Dialogue. *TAP*, 13(2):496–503, 2007, June.

[42] D. Mal, E. Wolf, N. Döllinger, M. Botsch, C. Wienrich, and M. E. Latoschik. Virtual Human Coherence and Plausibility – Towards a Validated Scale. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 788–789. IEEE, Christchurch, New Zealand, Mar. 2022. doi: 10.1109/VRW55335.2022.00245

[43] A. Mazumdar and C. Mousas. Synthesizing affective virtual reality multicharacter experiences. *Computer Animation and Virtual Worlds*, 32(3-4):e2004, 2021. doi: 10.1002/cav.2004

[44] R. R. McCrae and P. T. Costa Jr. The five-factor theory of personality. In *Handbook of Personality: Theory and Research, 3rd Ed.*, pp. 159–181. The Guilford Press, New York, NY, US, 2008.

[45] L. McInnes, J. Healy, and J. Melville. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction, Sept. 2020. doi: 10.48550/arXiv.1802.03426

[46] A. Mehrabian. Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in Temperament. *Current Psychology*, 14(4):261–292, Dec. 1996. doi: 10.1007/BF02686918

[47] E. Molina, A. Ríos, and N. Pelechano. The Impact of Animations in the Perception of a Simulated Crowd. In N. Magnenat-Thalmann, V. Interrante, D. Thalmann, G. Papagiannakis, B. Sheng, J. Kim, and M. Gavrilova, eds., *Advances in Computer Graphics*, vol. 13002, pp. 25–38. Springer International Publishing, Cham, 2021. doi: 10.1007/978-3-030-89029-2_2

[48] W. Moody, P. Kinderman, and P. Sinha. An exploratory study: Relationships between trying on clothing, mood, emotion, personality and clothing preference. *Journal of Fashion Marketing and Management: An International Journal*, 14(1):161–179, Jan. 2010. doi: 10.1108/13612021011025483

[49] S. R. Musse, V. J. Cassol, and D. Thalmann. A history of crowd simulation: The past, evolution, and new perspectives. *Vis Comput*, 37(12):3077–3092, Dec. 2021. doi: 10.1007/s00371-021-02252-w

[50] M. Nelson, A. Koilias, S. Gubbi, and C. Mousas. Within a Virtual Crowd: Exploring Human Movement Behavior during Immersive Virtual Crowd Interaction. In *Proceedings of the 17th International Conference on Virtual-Reality Continuum and Its Applications in Industry*, pp. 1–10. ACM, Brisbane QLD Australia, Nov. 2019. doi: 10.1145/3359997.3365709

[51] K. Niven, P. Totterdell, and D. Holman. A classification of controlled interpersonal affect regulation strategies. *Emotion*, 9(4):498–509, 2009. doi: 10.1037/a0015962

[52] G. Norman. Likert scales, levels of measurement and the "laws" of statistics. *Adv in Health Sci Educ*, 15(5):625–632, Dec. 2010. doi: 10.1007/s10459-010-9222-y

[53] I. Nyklíček, A. Vingerhoets, and M. Zeelenberg. Emotion Regulation and Well-Being: A View from Different Angles. In I. Nyklíček, A. Vingerhoets, and M. Zeelenberg, eds., *Emotion Regulation and Well-Being*, p. 32. Springer, New York, NY, 2011. doi: 10.1007/978-1-4419-6953-8_1

[54] A. Ortony, G. L. Clore, and A. Collins. *The Cognitive Structure of Emotions*. Cambridge University Press, Cambridge, 1988. doi: 10.1017/CBO9780511571299

[55] S. V. Paunonen and D. N. Jackson. What Is Beyond the Big Five? Plenty! *Journal of Personality*, 68(5):821–835, 2000. doi: 10.1111/1467-6494.00117

[56] K. Pearson. LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572, 1901. doi: 10.1080/14786440109462720

[57] N. Pelechano, J. Allbeck, and N. Badler. *Controlling Individual Agents in High-Density Crowd Simulation*, vol. 2007. Aug. 2007. doi: 10.1145/1272690.1272705

[58] N. Pelechano, C. Stocker, J. Allbeck, and N. Badler. Being a Part of the Crowd: Towards Validating VR Crowds Using Presence. *AAMAS*, pp. 136–142, May, 12-16, 2008.

[59] P. J. Rentfrow and S. D. Gosling. The do re mi's of everyday life: The structure and personality correlates of music preferences. *Journal of Personality and Social Psychology*, 84(6):1236–1256, 2003. doi: 10.1037/0022-3514.84.6.1236

[60] C. W. Reynolds. Flocks, herds and schools: A distributed behavioral model. In *Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '87, pp. 25–34. Association for Computing Machinery, New York, NY, USA, Aug. 1987. doi: 10.1145/37401.37406

[61] L. Saifi, A. Boubetra, and F. Nouioua. An approach for emotions and behavior modeling in a crowd in the presence of rare events. *Adaptive Behavior*, 24(6):428–445, Dec. 2016. doi: 10.1177/1059712316674784

[62] T. Schubert, F. Friedmann, and H. Regenbrecht. The Experience of Presence: Factor Analytic Insights. *Presence: Teleoperators & Virtual Environments*, 10(3):266–281, June 2001. doi: 10.1162/105474601300343603

[63] D. Schweingruber. A Computer Simulation of a Sociological Experiment. *Social Science Computer Review*, 13(3):351–359, Oct. 1995. doi: 10.1177/089443939501300305

[64] R. S. Sharma. Clothing behaviour, personality, and values: A correlational study. *Psychological Studies*, 25(2):137–142, 1980.

[65] A. Singhal. Modern Information Retrieval: A Brief Overview.

[66] M. Slater, B. Lotto, M. M. Arnold, and M. V. Sanchez-Vives. How we experience immersive virtual environments: The concept of presence and its measurement. 40.

[67] S. Sonlu, U. Güdükbay, and F. Durupinar. A Conversational Agent Framework with Multi-modal Personality Expression. *ACM Trans. Graph.*, 40(1):1–16, Feb. 2021. doi: 10.1145/3439795

[68] G. Stoet. PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods*, 42(4):1096–1104, Nov. 2010. doi: 10.3758/BRM.42.4.1096

[69] G. Stoet. PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, 44(1):24–31, 2017. doi: 10.1177/0098628316677643

[70] P.-N. Tan, M. Steinbach, and V. Kumar. *Introduction to Data Mining*. Pearson Addison Wesley, 2006.

[71] D. Thalmann and S. R. Musse. *Crowd Simulation*. Springer Science & Business Media, Oct. 2012.

[72] E. R. Thompson. Development and Validation of an International English Big-Five Mini-Markers. *Personality and Individual Differences*, 45(6):542–548, Oct. 2008. doi: 10.1016/j.paid.2008.06.013

[73] K. M. Thompson. Scale, Spectacle and Movement: Massive Software and Digital Special Effects in The Lord of The Rings. In *From Hobbits to Hollywood*, chap. From Hobbits to Hollywood, pp. 283–299. Brill, Jan. 2006. doi: 10.1163/9789401201513_021

[74] H. Trivedi and C. Mousas. Human-virtual crowd interaction: Towards understanding the effects of crowd avoidance proximity in an immersive virtual environment. *Computer Animation and Virtual Worlds*, 34(3-4):e2169, 2023. doi: 10.1002/cav.2169

[75] J. Tsai, N. Fridman, E. Bowring, M. Brown, S. Epstein, G. Kaminka, S. Marsella, A. Ogden, I. Rika, A. Sheel, M. Taylor, X. Wang, A. Zilka, and M. Tambe. ESCAPES - Evacuation Simulation with Children, Authorities, Parents, Emotions, and Social comparison.

[76] J. Van Den Berg, Ming Lin, and D. Manocha. Reciprocal Velocity Obstacles for real-time multi-agent navigation. In *2008 IEEE International Conference on Robotics and Automation*, pp. 1928–1935. IEEE, Pasadena, CA, USA, May 2008. doi: 10.1109/ROBOT.2008.4543489

[77] L. Van der Maaten and G. Hinton. Visualizing data using t-SNE. *Journal of machine learning research*, 9(11), 2008.

[78] M. Volonte, S. V. Babu, H. Chaturvedi, N. Newsome, E. Ebrahimi, T. Roy, S. B. Daily, and T. Fasolino. Effects of Virtual Human Appearance Fidelity on Emotion Contagion in Affective Inter-Personal Simulations. *IEEE Trans. Visual. Comput. Graphics*, 22(4):1326–1335, Apr. 2016. doi: 10.1109/TVCG.2016.2518158

[79] Y. Wang, J. E. F. Tree, M. Walker, and M. Neff. Assessing the Impact of Hand Motion on Virtual Character Personality. *ACM Trans. Appl. Percept.*, 13(2):1–23, Mar. 2016. doi: 10.1145/2874357

[80] Y. Wu, S. V. Babu, R. Armstrong, J. W. Bertrand, J. Luo, T. Roy, S. B. Daily, L. C. Dukes, L. F. Hodges, and T. Fasolino. Effects of Virtual Human Animation on Emotion Contagion in Simulated Inter-Personal Experiences. *IEEE Transactions on Visualization and Computer Graphics*, 20(4):626–635, Apr. 2014. doi: 10.1109/TVCG.2014.19

[81] S. Yang, T. Li, X. Gong, B. Peng, and J. Hu. A review on crowd simulation and modeling. *Graphical Models*, 111:101081, Sept. 2020. doi: 10.1016/j.gmod.2020.101081

[82] H. Yun, J. L. Ponton, A. Beacco, C. Andujar, and N. Pelechano. Exploring the Role of Expected Collision Feedback in Crowded Virtual Environments. In *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pp. 472–481. IEEE, Orlando, FL, USA, Mar. 2024. doi: 10.1109/VR58804.2024.00068

[83] E. Zell, C. Aliaga, A. Jarabo, K. Zibrek, D. Gutierrez, R. McDonnell, and M. Botsch. To stylize or not to stylize?: The effect of shape and material stylization on the perception of computer-generated faces. 34(6):1–12. doi: 10.1145/2816795.2818126

[84] R. Zhou, Y. Ou, W. Tang, Q. Wang, and B. Yu. An Emergency Evacuation Behavior Simulation Method Combines Personality Traits and Emotion Contagion. *IEEE Access*, 8:66693–66706, 2020. doi: 10.1109/ACCESS.2020.2985987

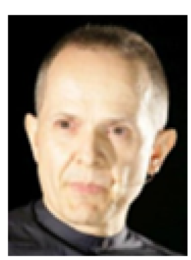
[85] K. Zibrek, S. Martin, and R. McDonnell. Is Photorealism Important for Perception of Expressive Virtual Humans in Virtual Reality? 16(3):1–19. doi: 10.1145/3349609

[86] M. Zyda. From visual simulation to virtual reality to games. *Computer*, 38(9):25–32, Sept. 2005. doi: 10.1109/MC.2005.297

Massimiliano Pascoli received the Master's Degree in Artificial Intelligence and Cybersecurity from the University of Udine, Udine, Italy, and the University of Klagenfurt, Klagenfurt, Austria. He is currently a Ph.D. Student with the Department of Mathematics, Computer Science, and Physics, University of Udine. His major research interests include human-computer interaction, virtual reality, and machine learning applied to crowd simulation.

Fabio Buttussi received the Ph.D. degree in computer science from the University of Udine, Udine, Italy. He is currently an associate professor with the Department of Mathematics, Computer Science, and Physics, University of Udine. His major research interests include human-computer interaction, virtual reality, serious games, and their applications in health and safety education.

Konstantin Schekotihin is a professor of intelligent systems at the University Klagenfurt, Austria. His research focuses mainly on various aspects of hybrid systems integrating knowledge representation and reasoning with machine learning methods to solve complex industrial problems such as scheduling and planning, configuration, recommender systems, acquisition and maintenance of knowledge graphs, or retrieval-augmented generation. Dr. Schekotihin authored over 80 papers published in prestigious AI conferences and leading international journals.



Luca Chittaro is full professor of Human-Computer Interaction (HCI) in the Department of Mathematics, Computer Science, and Physics of the University of Udine, Italy, where he heads the HCI Lab (http://hcilab.uniud.it). He has authored or co-authored over 200 international academic publications, and he has been an ACM Distinguished Speaker. His major research interests are in virtual reality, serious games, persuasive technology, mobile HCI, and their applications in health and safety. He has received research grants from organizations such as the US Federal Aviation Administration (FAA) and the European Union (EU), and companies such as the Benetton Group and the Intesa Sanpaolo banking group.