F2FCrowds: Planning Agent Movements to Enable Face-to-Face Interactions

Tanmay Randhavane[+], Aniket Bera, Dinesh Manocha

UNC Chapel Hill

E-mail: `tanmay@cs.unc.edu, ab@cs.unc.edu, dm@cs.unc.edu`

Abstract. The simulation of human behaviors in virtual environments has many applications. In many of these applications, situations arise in which the user has a face-to-face interaction with a virtual agent. In this work, we present an approach for multi-agent navigation that facilitates a face-to-face interaction between a real user and a virtual agent that is part of a virtual crowd. In order to predict whether the real user is approaching a virtual agent to have a face-to-face interaction or not, we describe a model of approach behavior for virtual agents . We present a novel interaction velocity prediction (IVP) algorithm that is combined with human body motion synthesis constraints and facial actions to improve the behavioral realism of virtual agents. We combine these techniques with full-body virtual crowd simulation and evaluate their benefits by conducting a user study using Oculus HMD in an immersive environment. Results of this user study indicate that the virtual agents using our interaction algorithms appear more responsive and are able to elicit more reaction from the users. Our techniques thus enable face-to-face interactions between a real user and a virtual agent and improve the sense of presence observed by the user.

[+] Correspondence concerning this article should be addressed to Tanmay Randhavane, Department of Computer Science, Chapel Hill, NC 27514. E-mail: tanmay@cs.unc.edu

Figure 1: **F2FCrowds:** Our algorithm enables the real user (in blue, wearing an HMD) to have face-to-face interactions with virtual agents. The virtual agents are responsive and exhibit head movements, gazing, and gesture behaviors.

1. Introduction

In many applications, it is important to simulate the behavior of virtual humans and crowds. It is well known that adding virtual agents or avatars into simulated worlds can improve the sense of immersion (Llobera, Spanlang, Ruffini, & Slater, 2010; Musse, Garat, & Thalmann, 1999; Pelechano, Stocker, Allbeck, & Badler, 2008; Slater et al., 2006). The use of virtual characters and associated environments is widely adopted in training and rehabilitation environments (Ulicny & Thalmann, 2001). Other applications include treatment of crowd phobias and social anxiety using VR therapy (Pertaub, Slater, & Barker, 2002), architectural flow analysis and evacuation planning (Cassol, Oliveira, Musse, & Badler, 2016; Haworth et al., 2016), learning a foreign language (Ólafsson, Bédi, Helgadóttir, Arnbjörnsdóttir, & Vilhjálmsson, 2015), etc.

There is considerable work on evaluating the sense of presence and immersion in VR based on the behaviors, interactions, and movements of virtual agents. Many researchers have concluded that the social presence of virtual agents depends on the realism of their behavior (Blascovich et al., 2002) and the nature of their interactions (Guadagno, Blascovich, Bailenson, & Mccall, 2007; Kyriakou, Pan, & Chrysanthou, 2015). Recent advances in artificial intelligence (including natural language processing and computer vision, along with development of embodied conversational agents) are helping to generate realistic interaction scenarios. Other work includes the development of techniques to simulate gazing, collision avoidance movements, head turning, facial expressions, and other gestures (Grillon & Thalmann, 2009; Nummenmaa, Hyönä, & Hietanen, 2009).

One of the main social interactions is *face-to-face (F2F) interaction*, which is typically carried out without the use of any mediating technology (Crowley & Mitchell, 1994). This broad area has been studied in social sciences for more than a century. There is a recent interest in integrating virtual reality technology into social media, and F2F communication is an important component of such a system. Different sensory organs play an important role in these interactions, which may include eye contact or two agents facing or talking in close proximity to each other. As a result, there are many challenges in terms of developing such interaction capabilities between virtual agents.

Previous works have treated F2F conversations as a joint activity involving two or more participants (Ólafsson et al., 2015). Clark (1996) identifies three stages of participation in a joint activity. The *entry, body,* and *exit* of the conversation constitute these stages on a very high level. In this work, we try to model these human behaviors for F2F interactions with virtual agents.

- In order to *enter* an F2F interaction with a virtual agent, a real user must approach the virtual agent and the virtual agent should recognize the real user's intent to interact.
- The virtual agent should respond in a positive way and lead to the next stage, which corresponds to the *body* section of the interaction.

- When the *body* of the interaction is over or if the real user loses interest in the conversation, the virtual agent should recognize this event and *exit* the conversation.

In this work, we propose solutions to the problems of *entry* and *exit* of the virtual agents in an F2F interaction with a real user. The *body* of the conversation usually contains a verbal exchange of information. We do not focus on verbal communications between the user and the agents, though our approach can be combined with such methods.

Main Results: We address the problem of computing the movements or trajectories to enable F2F interactions between a real user and a virtual agent who is part of a virtual crowd. This includes automatically computing collision-free trajectories that enable such agents to come close to each other for F2F communications. Satake et al. (2009) developed a model of approach behavior for robots having F2F interactions with people who are walking. Their model was based on the idea that human interactions can be classified based on social and public distance (Hall, 1966). Motivated by these ideas, we develop a model of approach behavior for virtual agents that models their movement for F2F communication. We present a novel navigation algorithm, *Interaction Velocity Prediction* (*IVP*), which predicts whether the avatar of a real user is trying to approach a virtual agent for F2F interaction. IVP is combined with 2D multi-agent simulation to compute collision-free trajectories. In order to generate plausible full-body simulations, we also integrate the velocity computation with human motion synthesis to generate upper body movements such as gazing and nodding. Overall, our approach (F2FCrowds) can generate smooth and natural-looking trajectories for each agent. We use our algorithms to simulate the movement of tens of virtual agents in complex indoor and outdoor environments at interactive rates.

In order to evaluate the benefits of our algorithms, we performed a user evaluation in an immersive environment where a real user interacted with the virtual agents in four different scenarios. In particular, we compared our algorithm (with and without upper body behaviors) with a baseline crowd simulation algorithm that uses coupled 2D navigation and full-body motion synthesis (Narang, Best, Randhavane, Shapiro, & Manocha, 2016).

We observe a statistically significant preference for our new algorithm. Our algorithm increased the sense of presence felt by the users. When using our algorithms, the virtual agents appeared more responsive and were able to elicit more reaction from the users. Our results for the sense of presence question show that 40% of the participants preferred our algorithm (without upper body behaviors) over the baseline, whereas only 3.33% participants preferred the baseline and the rest remained neutral. Participants felt virtual agents were more aware when using our algorithm (without upper body behaviors) in 60% of the responses, whereas only 3.33% felt that way for the baseline. Our methods (without upper body behaviors) elicited more reaction from the user in 53.33% of the cases, whereas the baseline elicited more reaction in 10% of the cases. The addition of upper body behaviors also showed a significant improvement in the performance.

The rest of the paper is organized as follows. We briefly present prior work on crowd simulation and interactions in Section 2. In Section 3, we provide an overview of our algorithm. We describe the model of approach behavior for virtual agents and the novel velocity computation algorithms in Section 4. In Section 5, we provide the implementation details and highlight our algorithm's performance on different benchmarks. We describe the details of our user evaluation in Section 6.

2. Related Work

In this section, we give an overview of prior work on face-to-face interactions, crowd simulation for VR, and interaction with virtual agents in a virtual environment.

2.1. Face-to-Face Interactions

Face-to-face interactions have been studied in psychology, sociology, and robotics. Satake et al. (2009) presented an algorithm to enable a robot to have F2F interactions with people who are walking. Gonçalves and Perra (2015) studied empirical characteristics of face-to-face interaction patterns and novel techniques to discover mesoscopic structures in these patterns. There is work on investigating F2F interactions in terms of capabilities to understand and generate natural language in combination with non-verbal signals and social management (Bonaiuto & Thórisson, 2008; Cassell, Vilhjálmsson, & Bickmore, 2001; Heylen et al., 2011; Jonsdottir, Thorisson, & Nivel, 2008; Kopp, Stocksmeier, & Gibbon, 2007; Pantic et al., 2011; Vinciarelli, Pantic, & Bourlard, 2009). In this paper, we attempt to provide the users of virtual reality the ability to have F2F interactions with virtual agents in virtual crowds. Our approach provides a platform to implement the aforementioned models in the context of F2F interactions in virtual crowds.

2.2. Crowd and Multi-Agent Simulation

A significant amount of research has been done in multi-agent and crowd simulation. In this paper, we mainly limit ourselves to a class of algorithms that decomposes the trajectory or behavior computation for each agent into two parts: global planning and local navigation (Helbing & Molnar, 1995; Kapadia & Badler, 2013; Ondřej, Pettré, Olivier, & Donikian, 2010; Reynolds, 1999; Van Den Berg, Guy, Lin, & Manocha, 2011). The global planner computes a path for each agent in the environment towards its intermediate goal position. The local navigation algorithms modify these paths so that the agents can avoid collisions with dynamic obstacles or other pedestrians in the environment. Some of these methods also account for a pedestrian's personality (Guy, Kim, Lin, & Manocha, 2011; Pelechano, Allbeck, & Badler, 2007) or use cognitive techniques (Funge, Tu, & Terzopoulos, 1999).

Boulic (2005) presented a mathematical model to approach a dynamic target with a target orientation but this method does not take into account for proxemic distances to the mobile entities involved. In the robotics community, algorithms have also been proposed to predict humans' intentions and take them into account for robot navigation (Bera, Kim, Randhavane, Pratapa, & Manocha, 2016; Bera, Randhavane, & Manocha, 2017; Bera, Randhavane, Prinja, & Manocha, 2017; Bršččić, Kidokoro, Suehiro, & Kanda, 2015; Park, Ondřej, Gilbert, Freeman, & O'Sullivan, 2016).

2.3. Interaction with Virtual Agents

There is extensive literature on simulating realistic behaviors, movements, and interactions with virtual agents in VR (Magnenat-Thalmann & Thalmann, 2005). In this paper, we restrict ourselves to modeling some of the interactions between real and virtual agents when they are in close proximity. Kyriakou et al. (2015) showed that basic interaction increases the sense of presence, though they did not explicitly model users' intent to participate. Bailenson, Blascovich, Beall, and Loomis (2001) concluded that there is an inverse relationship between gazing and personal space. Pelechano et al. (2008) showed that pushing-based interaction increases the sense of presence in a virtual environment. Bonsch, Weyers, Wendt, Freitag, and Kuhlen (2016) described a gaze-based collision avoidance system and interactions for small-scale virtual environments. Hu, Adeagbo, Interrante, and Guy (2016) presented a system where virtual agents exhibit head turning behavior but do not explicitly model face-to-face interaction. There is also considerable work on individualized avatar-based interactions (Nagendran, Pillat, Kavanaugh, Welch, & Hughes, 2014). Olivier, Bruneau, Cirio, and Pettré (2014) presented a CAVE-based VR platform aimed at studying crowd behaviors and generating reliable motion data. This platform can also be used for studying the interaction between a real user and a virtual human. Our approach to enabling

F2F interactions between a real user and virtual agents is complimentary to most of these methods and can be combined with them.

Ólafsson et al. (2015) proposed a communicative function called the “Explicit Announcement of Presence” to initiate conversations with strangers in a virtual environment. Their approach assumes that the virtual agent does not have any interest in starting a conversation and that the user’s intent for initiating conversations is made known by clicking the mouse. Our approach, on the other hand, enables the virtual agents to learn the user’s intent based on their trajectories and orientation. Pedica and Vilhjálmsson (2009) proposed a reactive framework that allows a group of real users’ avatars to have social interactions with territorial dynamics. This approach focuses mostly on user-controlled avatars, whereas our approach considers only a single real user and a crowd of virtual agents.

3. Overview

In this section, we introduce our notation and give an overview of our approach for crowd simulation to enable F2F interactions.

3.1. Notation

Our approach uses a multi-agent simulation algorithm that computes the trajectory of each agent using a combination of global planning and local navigation. We make following simplifying assumptions:

- The environment consists of one real user, represented by its avatar, and virtual users or agents. The real user is walking in or navigating an immersive setting in an environment with many virtual agents, avoiding collisions, and attempting to interact with the virtual agents. In this work, we do not consider a case where multiple real users share the same virtual space.

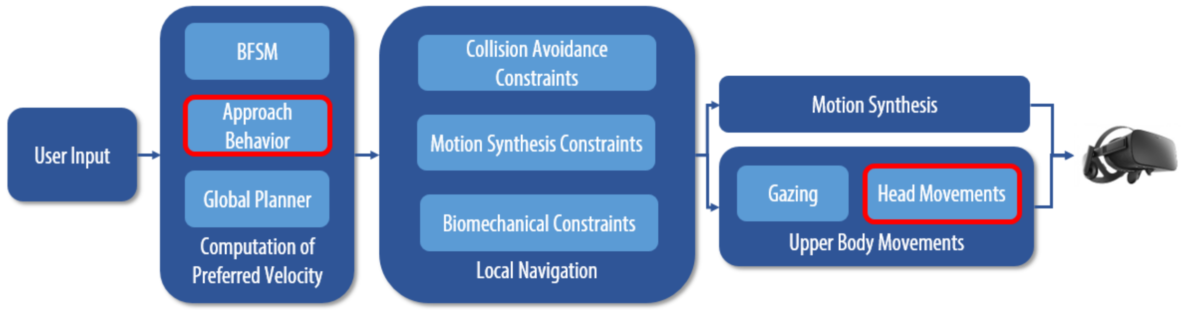

Figure 2: **Overview:** We highlight the various components of our F2F crowd simulation system. The novel algorithmic contributions are highlighted with red boundaries. Our interactive crowd simulation pipeline enables a real user to interact with virtual agents. A model of approach behavior is used to predict whether the avatar of the real user intends to perform F2F interactions. We also simulate upper body movements to increase the realism.

- At any instant, the real user can have a face-to-face interaction with at most one virtual agent. When the real user is interacting with a virtual agent, other virtual agents know that the user is "busy" or not available for such an interaction. This keeps other virtual agents from intruding on the real user's conversation and means that they do not use the IVP algorithm.

One of our goals is to compute collision-free and plausible trajectories for each virtual agent to enable F2F interactions with the user. We model this using a novel model of approach behavior based on Hall's idea of social and public distance (Hall, 1966). This model makes use of the novel IVP algorithm to predict whether the avatar of the real user is trying to approach a virtual agent to perform F2F interactions.

We represent each agent using a high-DOF articulated model and compute upper and lower body motions. The state of an agent i is represented by $\mathbf{q}_i$ and is the union of the position of the root joint and the states of all the joints of the high-DOF character. In terms of 2D multi-agent navigation, we represent an agent i as a circle of radius r_i at a position $\vec{p}_i$, which is the 2D position of the root joint of the agent. At any time, a virtual agent i's current velocity is represented by $\vec{v}_i^c$ and its preferred velocity and preferred orientation are represented by $\vec{v}_i^o$ and o_i^o, respectively. The preferred velocity and orientation are based on the intent of the virtual agent. We use $\mathcal{M}$ to denote the set of available upper body behaviors and m_i to denote the current upper body behavior of agent i. The 3D point on the face of the user avatar at which the virtual agent i is currently gazing is denoted by $\vec{g}_i$ (i.e. the *gazing point*). We represent the user's avatar with the subscript u. Let $\mathcal{S}$ be the simulator state, which is the union of the states of all the entities in the scene, including obstacles and agents.

Figure 2 provides an outline of our interactive crowd simulation pipeline. We use a game engine to gather the user's input, which is then used by our multi-agent simulation system.The 2D multi-agent system uses a layered 2D navigation algorithm. The first layer corresponds to the model of approach behavior and global planning, which computes the preferred velocity $\vec{v}_i^o$ and preferred orientation o_i^o of each virtual agent i. The second layer, local navigation, computes the collision-free velocities $\vec{v}_i^c$ for each virtual agent i. The computed velocity $\vec{v}_i^c$ and upper body behavior motions are passed to the motion synthesis module, which computes the state $\mathbf{q}_i$ for each virtual agent i.

4. Model of Approach Behavior

Hall (1966) studied human behavior and proposed four distance zones in humans: Intimate Distance, Personal Distance, Social Distance, and Public Distance. According to Hall, each of these distance zones facilitates a different type of interaction. At the social distance zone, humans are close enough to communicate and have face-to-face interactions with each other, whereas at the public distance zone, humans are close enough to notice each other but far enough to not be able to have face-to-face interactions with each other. Personal distance and intimate distance are generally reserved for friends and family. Based on these ideas, Satake et al. (2009) proposed a model of approach behavior with which a robot can initiate

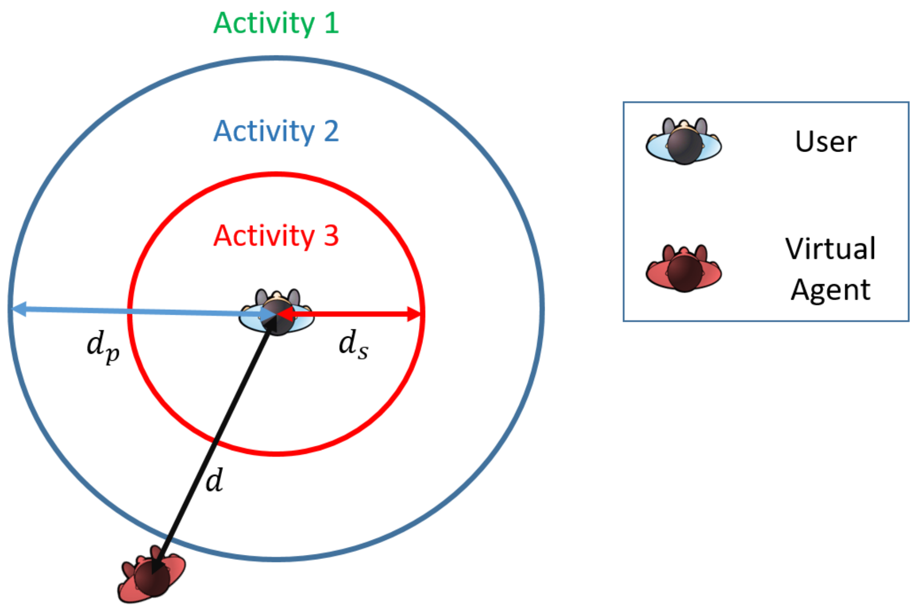

Figure 3: **Model of Approach Behavior**: We define the model of approach behavior for virtual agents as a sequence of three activities based on the distance between the user and the virtual agent (d) : (1) identifying the intent of interaction of the user, (2) approaching at public distance (d_p), and (3) initiating communication at social distance (d_s).

conversation with people who are walking. In order to have an F2F interaction, a robot should find a person with whom to talk, start approaching that person at a public distance, and initiate the conversation at a social distance. Therefore, Satake et al. (2009) defined "approach behavior" as a sequence of the following activities: (1) selecting a target, (2) approaching the target at public distance, and (3) initiating conversation at social distance. We use similar ideas and propose a model of approach behavior for virtual agents (Figure 3) that models how the virtual agent should approach the user. The model is a sequence of the following activities: (1) identifying the intent of interaction of the user, (2) approaching at public distance (d_p), and (3) initiating communication at social distance (d_s).

4.1. Identifying the Intent of Interaction of the User

A user may wish to interact with a virtual agent that is currently farther than public distance (d_p). In order to have an interaction, the user will attempt to be within the social distance

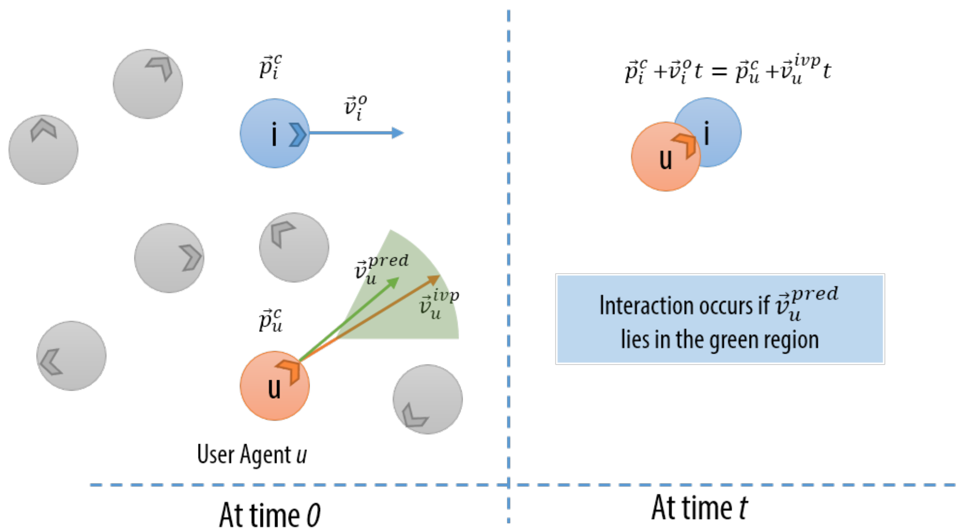

Figure 4: **Interaction Velocity Prediction:** Given the current position ($\vec{p}_i^c$) and preferred velocity of the virtual agent ($\vec{v}_i^o$) and the current position of the user agent ($\vec{p}_u^c$), our IVP algorithm predicts the velocity ($\vec{v}_u^{ivp}$) of the user agent to intercept the virtual agent in time t_{min}. If the user's predicted velocity $\vec{v}_u^{pred}$ satisfies the constraint $\vec{v}_u^{ivp} \bullet \vec{v}_u^{pred} \geq \theta_v$, it will result in F2F communication.

(d_s) of the virtual agent by moving towards the virtual agent. The virtual agent then must be able to identify the user's intent of interaction in order to have an F2F interaction. To achieve this, we use a novel algorithm, **Interaction Velocity Prediction (IVP)**, that predicts whether or not a user is trying to interact with a virtual agent.

Given the current position $\vec{p}_i^c$ and preferred velocity $\vec{v}_i^o$ of a virtual agent i and the current position of the user agent $\vec{p}_u^c$, our IVP algorithm $\mathbf{IVP}_i : \mathbb{R}^2 \times \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}^2 \times \mathbb{R}$ determines the velocity ($\vec{v}_u^{ivp}$) that the real user should follow to intercept the virtual agent in time t_{min}. If the public distance is given by d_p, then the time of interception t can be given as:

$$\left\| \vec{p}_u^t - \vec{p}_i^t \right\| \leq d_p. \tag{1}$$

Assuming that the user agent has the velocity $\vec{v}_u^{ivp}$ and the virtual agent has the average velocity $\vec{v}_i^o$,

$$\left\| (\vec{p}_u^c + \vec{v}_u^{ivp} t) - (\vec{p}_i^c + \vec{v}_i^o t) \right\| \leq d_p, \tag{2}$$

$$\left\| \vec{v}_u^{ivp} - (\vec{v}_i^o - \frac{(\vec{p}_u^c - \vec{p}_i^c)}{t}) \right\| \leq \frac{d_p}{t}. \tag{3}$$

We solve the above equation for interaction velocity $\vec{v}_u^{ivp}$, i.e. when t is minimized. We also take into account motion and dynamic constraints of the agent and put a limit on the maximum speed:

$$\left\|\vec{v}_u^{ivp}\right\| \leq v_{max}, \tag{4}$$

where v_{max} is the maximum speed of the user agent. Simplifying these two equations leads to a 4^{th} order polynomial. Therefore, we calculate $\vec{v}_u^{ivp}$ such that the center of the circular virtual agent and circular user agent coincide, i.e. :

$$\vec{v}_u^{ivp} = \vec{v}_i^o - \frac{(\vec{p}_u^c - \vec{p}_i^c)}{t}. \tag{5}$$

Substituting this expression in Equation 4 results in

$$\left\|\vec{v}_i^o - \frac{(\vec{p}_u^c - \vec{p}_i^c)}{t}\right\| \leq v_{max} \tag{6}$$

$$(\vec{v}_{ix}^o t - (\vec{p}_{ux}^c - \vec{p}_{ix}^c))^2 + (\vec{v}_{iy}^o t - (\vec{p}_{uy}^c - \vec{p}_{iy}^c))^2 \leq v_{max}^2 t^2. \tag{7}$$

We simplify Equation 7 as $at^2 + bt + c \leq 0$, where

$$a = (\vec{v}_{ix}^o)^2 + (\vec{v}_{iy}^o)^2 - v_{max}^2, \tag{8}$$

$$b = -2((\vec{p}_{ux}^c - \vec{p}_{ix}^c)\vec{v}_{ix}^o + (\vec{p}_{uy}^c - \vec{p}_{iy}^c)\vec{v}_{iy}^o), \tag{9}$$

$$c = (\vec{p}_{ux}^c - \vec{p}_{ix}^c)^2 + (\vec{p}_{uy}^c - \vec{p}_{iy}^c)^2. \tag{10}$$

We assume that the agent's speed will be lower than the user's speed (otherwise the user agent will never be able to intercept the virtual agent), $a \leq 0$. Since $c > 0$, t_{min} is the larger root of the equation $at^2 + bt + c = 0$ and the interaction velocity is:

$$\vec{v}_u^{ivp} = \vec{v}_i^o - \frac{(\vec{p}_u^c - \vec{p}_i^c)}{t_{min}} \tag{11}$$

4.1.1. Computation of Preferred Velocity We use IVP (Figure 3) to compute the interaction velocity $\vec{v}_u^{ivp}$ that the user will have in order to have F2F interactions with a virtual agent i at time t_{min}. Based on the user's position from the past few frames, we can predict the velocity of the user based on some motion model and denote it as $\vec{v_u^{pred}}$. In this case, the virtual agent i will have F2F interaction with the user if:

$$\vec{v_u^{ivp}} \bullet \vec{v_u^{pred}} \geq \theta_v, \tag{12}$$

where θ_v is a pre-determined threshold.

The preferred velocity $\vec{v_i^o}$ for a virtual agent i is then computed as follows:

$$\vec{v_i^o} = v_{pref} * \frac{(\vec{p_u^c} - \vec{p_i^c})}{\|(\vec{p_u^c} - \vec{p_i^c})\|}, \tag{13}$$

where $\vec{p_i^c}$ and $\vec{p_u^c}$ are the current positions of the virtual agent i and user agent u, respectively, and v_{pref} is the preferred natural speed of the virtual agent.

4.2. Approaching at Public Distance

At public distance, the virtual agent and the user can acknowledge each other. The virtual agent achieves this by slowing down and gazing at the user. We define a boolean function $approach_p()$ to denote the conditions when a virtual agent i decides to approach the real user at public distance:

$$approach_p() = (d_s < \|\vec{p_u^c} - \vec{p_i^c}\| < d_p) \tag{14}$$

$$\wedge \, (\vec{o_u^c} \bullet \frac{\vec{p_u^c} - \vec{p_i^c}}{\|\vec{p_u^c} - \vec{p_i^c}\|} > o_{thresh}), \tag{15}$$

where $\vec{o_u^c}$ is the 2D orientation vector of the user, o_{thresh} is a pre-determined threshold, and d_s and d_p are the social and public distances, respectively.

When $approach_p()$ evaluates to true, the virtual agent slows down to allow a friendly approach and its preferred velocity is given by:

$$\vec{v_i^o} = k * v_{pref} * (\vec{p_u^c} - \vec{p_i^c}). \tag{16}$$

Here, $0 < k < 1$ is a pre-determined constant. Notice that the virtual agent's speed is directly proportional to the distance between the agent and the user to slow down the virtual agent as it approaches the user.

4.2.1. Gazing In addition to computing the appropriate velocities, it is also important to exhibit appropriate upper body movements and behaviors for F2F communications. Gazing plays an important role in conveying the intent of interaction and it is important for virtual agents to maintain eye contact with the user while approaching. Therefore, the virtual agents gaze at the eyes of the user's 3D avatar (Figure 6) whenever $approach_p()$ evaluates to true. We do this by setting the gazing point g_i of the virtual agent i to the position of the eye of the user's 3D avatar.

4.3. Initiating Communication at Social Distance

Social distance is the distance at which humans typically have face-to-face interactions in social scenarios (Hall, 1966). Therefore, when the distance between the real user and the virtual agent is less than social distance, the virtual agent stops and attempts to have a communication with the user 5 as denoted by the boolean function $approach_s()$:

$$approach_s() = (\|\vec{p_u^c} - \vec{p_i^c}\| < d_s) \tag{17}$$

$$\wedge\ (\vec{o_u^c} \bullet \frac{\vec{p_u^c} - \vec{p_i^c}}{\|\vec{p_u^c} - \vec{p_i^c}\|} > o_{thresh}), \tag{18}$$

where $\vec{o_u^c}$ is the 2D orientation vector of the user, o_{thresh} is a pre-determined threshold, and d_s is the social distance.

4.3.1. Head Movements Head movements play an important part in F2F interactions (McClave, 2000). Therefore, our virtual agents exhibit head movements like nodding, shaking, and tossing their heads to communicate with the user (Figure 6). During the

Figure 5: **F2F Communications:** Our approach enables F2F communications between the avatar of the real user and the virtual agent.

communication, the virtual agent performs head movements at randomized time intervals ranging from $6 - 10$ seconds (based on Hadar, Steiner, and Clifford Rose (1985)). The head movement is chosen at random from the set of motions $\mathcal{M} = \{nod, toss, shake\}$. Since nod implies a positive intent of interaction, the first head movement is always chosen to be a nod. The virtual agents pursue a conversation only until the user's attention is on the agent. The virtual agent concludes that the communication is over when $approach_s()$ evaluates to false and continues its goal-directed navigation in the scene.

4.4. Navigation

The model of approach behavior discussed so far determines whether or not a virtual agent is a part of an interaction and then calculates its preferred velocity. All the other agents that are not part of any interaction follow goal-directed behavior. We use algorithms described in Narang, Randhavane, Best, Shapiro, and Manocha (2016) to plan the movement of these agents and determine their preferred velocity. We use constraint modeling from Narang, Randhavane, et al. (2016) to model the collision avoidance constraints and modify the preferred velocity of each agent i to get the current velocity $\vec{v}_i^c$.

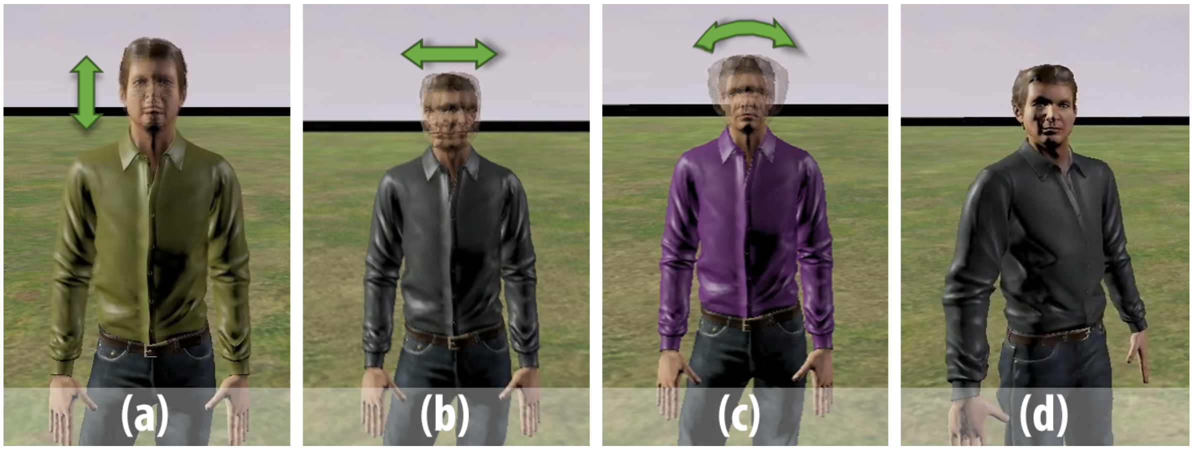

Figure 6: **Gestures:** Our virtual agents exhibit head movements and gazing. Appropriate head movements are chosen from the set of movements including (a) nod (vertical head movement), (b) shake (horizontal head movement), and (c) toss (sideways head movement). (d) Virtual agents also gaze at the user agent to establish eye contact.

5. Results

5.1. Implementation and Performance

We have implemented our system on a Windows 10 desktop PC with Intel Xeon E5-1620 v3 in parallel on 4 cores and 16 GB of memory. We use Menge (Curtis, Best, & Manocha, 2016) as our multi-agent simulation library that computes the 2D trajectory for each agent. We have modified the global planning and local navigation algorithms based on the components described by Narang, Randhavane, et al. (2016). A game engine (Unreal Engine 4) serves as an interface to the user and as a rendering framework. We use Smartbody (Shapiro, 2011) to synthesize motion of virtual agents and to provide the joint angles to simulate the motions corresponding to various gestures or upper body movements. Though the proxemic distances vary from person to person, we used a value of $d_s = 3.6m$ for social distance and $d_p = 7.6m$ for public distance (Hall, 1966). Other parameters that can be controlled include the thresholds $\theta_v = 1.41$, $o_{thresh} = \frac{\pi}{12}$, and the multiplier $k = 0.1$.
We compared our algorithms with a baseline crowd simulation algorithm, PedVR (Narang, Best, et al., 2016). Table 1 highlights the performance of our system on the following

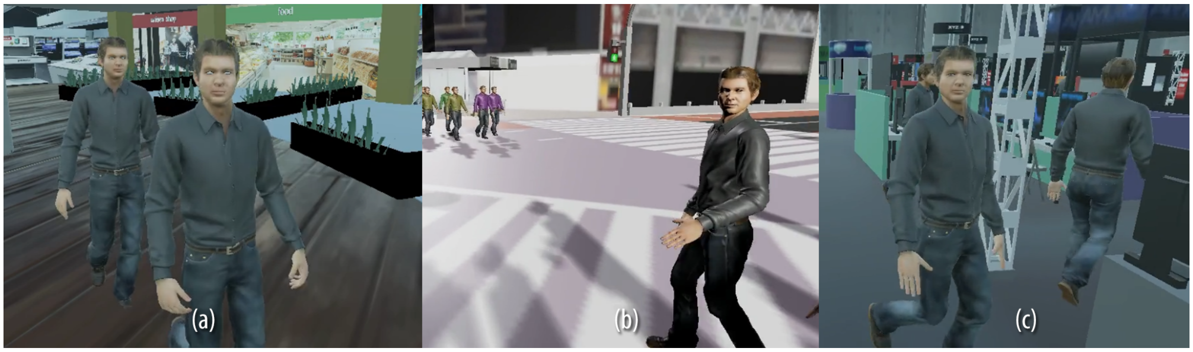

Figure 7: **Benchmarks:** We highlight the performance of our algorithm on three benchmarks. (a) A shopping mall scene shows virtual agents walking in a mall. (b) The user agent travels in a crossing scenario with multiple virtual agents who gaze at the user's avatar. (c) Virtual agents explore a tradeshow scenario and acknowledge the user avatar's presence with eye contact. We are able to simulate tens of agents at interactive rates and evaluate the benefits of F2F interactions.

benchmark scenarios (Figure 7):

- **Shibuya Crossing** A busy crossing scenario. We initialize the agents at different positions of the intersection. The goal positions are assigned using a probability distribution. After reaching the goals, each agent waits for a few seconds and then moves towards the next goal.
- **Shopping Mall** Virtual agents walk in a shopping mall. They walk to the shops and exhibit head movements (nod or shake for approval or disapproval, respectively) and gazing behaviors at the shops.
- **Tradeshow** Virtual agents walk up to the booths in a tradeshow and exhibit head movements.

The average frame update time is almost the same for both PedVR and F2FCrowds indicating that our IVP algorithm does not add a significant overhead. The addition of head movements and gazing behaviors adds an overhead of 20% and, overall, our system can simulate 30+ agents at approximately 40-60 FPS.

Benchmark	Agents	Average Frame Update Time (ms)		
		PedVR	F2F	F2F+G
Mall	24	19	20	24
Shibuya	32	18	18	20
Tradeshow	30	21	22	26
Circle	8	9	9	11
Bidirectional	8	8	8	10

Table 1: **Average frame update time:** In the absence of upper body movements, F2FCrowds with IVP does not have significant overhead over PedVR. F2FCrowds with gesture can simulate 30+ virtual agents at 40-60 FPS.

5.2. User Evaluation

In this section, we describe our user study, which was conducted to evaluate our new algorithms that enable F2F interactions. We performed a within-users study showing the advantages of our model of approach behavior and upper body motion generation.

5.2.1. Study Goals and Expectations We designed our study based on the prior work of Pelechano et al. (2008); Garau, Slater, Pertaub, and Razzaque (2005); and Narang, Best, et al. (2016), which evaluated the level of presence based on the behavior and interactions of a user with the virtual agents within a crowd. In these works, *Presence* has been defined as the extent to which people respond realistically to virtual events and situations and we use a similar criterion. Our study was aimed at measuring the advantage of our model of approach behavior for virtual agents over a baseline interactive system. We expected to find that participants felt it easier to have F2F interactions with our algorithm and that these interactions also benefited from the addition of head movements and gazing behaviors. In particular, we propose the following hypotheses:

- Hypothesis 1: Addition of model of approach behavior and upper body motion generation increases the sense of presence felt by the user.
- Hypothesis 2: Users do not have to make extra effort to avoid the virtual agents after

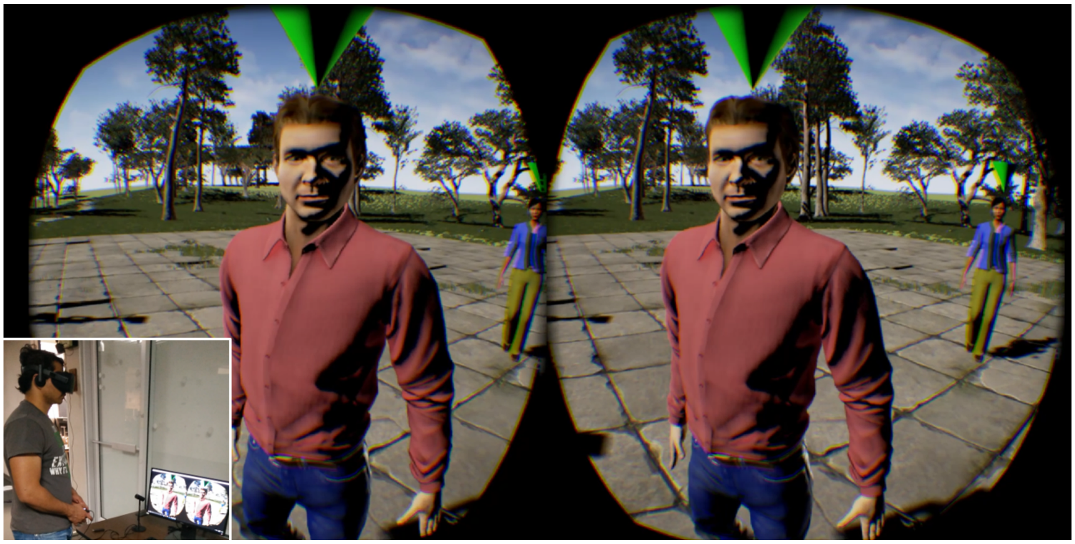

Figure 8: **User Interacting with the Virtual Agent:** Participants approached virtual agents and attempted to have an F2F interaction.

the addition of model of approach behavior and upper body motion generation.

- Hypothesis 3: Virtual agents appear more aware of the user after the addition of model of approach behavior and upper body motion generation.
- Hypothesis 4: Addition of model of approach behavior and upper body motion generation elicits more response from the users.
- Hypothesis 5: Virtual agents appear more responsive after the addition of model of approach behavior and upper body motion.

5.2.2. Experimental Design A within-users study was performed in which the participants (N = 15 - 20) were asked to participate in several scenarios using an Oculus Rift head mounted display. Participants were standing up and used a joystick for movement in the virtual world (Figure 8). A training scenario was also presented to familiarize the participants with the movement. The participants performed three trials of each scenario in randomized order and answered a questionnaire at the end of each scenario.

Evaluated Methods : In the study, participants compared three different interaction enabling algorithms:

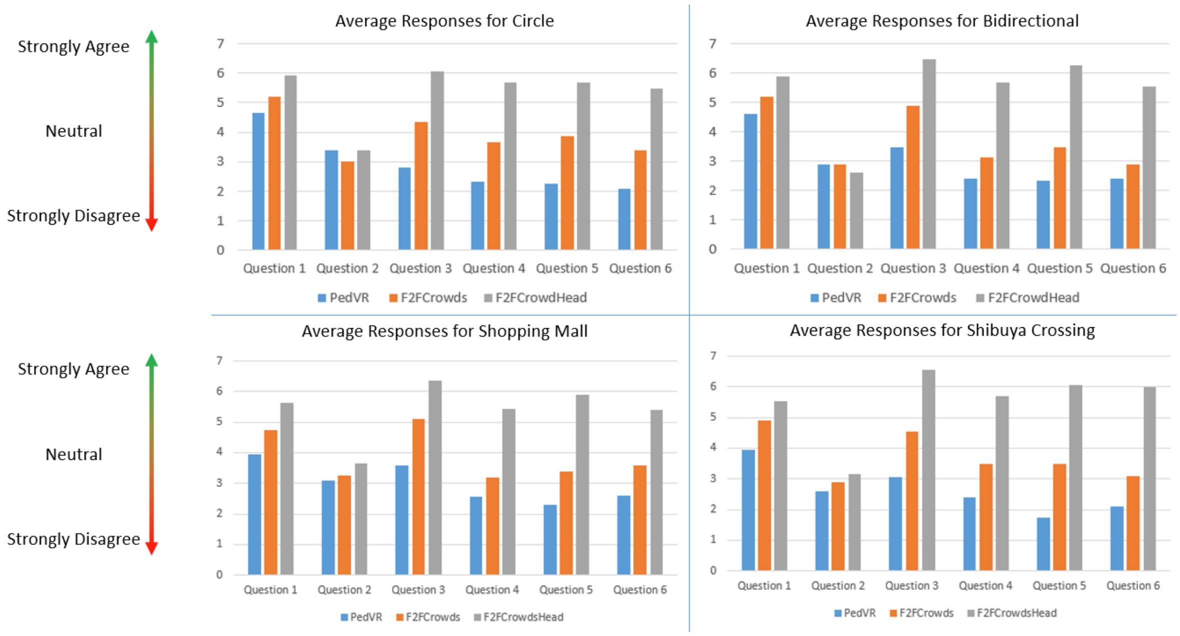

Figure 9: **Average Responses:** Participants experienced a higher amount of social presence for F2FCrowds compared to PedVR as observed from higher average responses to question 1. Responses to question 2 are almost similar for the three methods, indicating that participants had to make a similar amount of effort to avoid collisions across the three methods. Responses to questions 3, 4, and 5 indicate that participants felt that our model of approach behavior was beneficial in making the characters responsive to participants' attempts to interact.

- **PedVR** : We used the coupled crowd simulation method PedVR as the baseline (Narang, Best, et al., 2016). Gazing and head movements were not included in this algorithm.
- **F2FCrowds** : Our model of approach behavior without gazing and head movements.
- **F2FCrowdsHead** : In addition to the approach behavior, virtual agents also communicated using gazing and head movements.

Task: The participants were asked to approach any virtual agent and were informed that when they felt that it was possible to have an F2F interaction with the virtual agent, they should press a button and the agent in front of them would be highlighted for two seconds.

Scenarios : The following scenarios were presented to the participants. The participants performed three trials of each scenario (45 seconds each) corresponding to each method and answered a questionnaire after each scenario.

- **Circle** : This scenario consisted of 8 virtual agents starting on the perimeter of a circle. Their target positions were selected randomly on the perimeter of the circle and the simulation resulted in a high-density area at the center of the circle. The participants started from a position inside the circle.
- **Bidirectional** : 8 virtual agents started from opposite ends of a hallway, with half the agents at either end of the hallway, and traveled between the two ends. The participant (i.e. the real user) started at the middle of the hallway.
- **Shopping Mall** : 8 virtual agents explored a shopping mall scenario. The background of the scene visually resembled a shopping mall. The participant started at the center of the mall.
- **Shibuya Crossing** : 8 virtual agents walked in the crossing scenario, which looks like Shibuya crossing in Tokyo. The virtual agents started at the ends of the crosswalks and the participant started at the center of the scene.

Questionnaire : The aim of the user study was to show the benefits of our model of approach behavior for virtual agents and upper body movement. We used a modified version of a well-established questionnaire for social presence (Garau et al., 2005). In particular, we used a subset of the original questions and asked additional questions regarding the participant's interaction with the virtual agents. The questions were of an Agree/Disagree type and participants noted their preference using a seven-level Likert scale with values labeled "Strongly disagree", "Disagree", "Slightly disagree", "Neutral", "Slightly agree", "Agree", and "Strongly agree". For analysis, we convert the participant responses to a scale of 1 (Strongly Disagree) - 7 (Strongly Agree). We list the questionnaire in Table 2.

5.2.3. Discussion In this section, we present and analyze the participant responses (Figure 9) to the three interaction simulation algorithms described previously. For each scenario, the

Question 1: I had a sense of being in the same space as the characters.
Question 2: I had to make an effort to avoid the characters.
Question 3: The characters seemed to be aware of me.
Question 4: I felt that I should talk/nod/respond to the characters.
Question 5: The characters seemed to respond to my attempts of interaction.
Question 6: The characters seemed to respond even if I did not attempt to interact.

Table 2: **Questionnaire:** Participants were asked to answer the above questions on a seven-level Agree/Disagree Likert scale.

	Circle		Bidirectional		Shopping Mall		Shibuya Crossing	
	χ^2	p	χ^2	p	χ^2	p	χ^2	p
Question 1	10.516	0.005	10.759	0.005	14.731	0.001	12.34	0.002
Question 2	0.167	0.92	0.2	0.905	1.2	0.549	2.311	0.315
Question 3	16.642	0	23.787	0	21.522	0	25.552	0
Question 4	16.51	0	23.231	0	20.338	0	27.193	0
Question 5	16	0	22.37	0	16.794	0	25.618	0
Question 6	13.192	0.001	17.796	0	8.211	0.016	14.389	0.001

Table 3: **Results of a Friedman Test**: We present the test statistic (χ^2) value and the significance level (p) of a Friedman test performed to test for differences between the responses for the three algorithms.

simulation algorithm is the independent variable and the participant response is the dependent variable. Since our dependent variable is ordinal, we used the **Friedman test** to test for differences between the responses for the three algorithms. Post hoc analysis with **Wilcoxon signed-rank tests** was conducted with a Bonferroni correction applied, resulting in a significance level set at $p < 0.017$. We tabulate the test statistic (χ^2) value and the significance level (p) in Table 3. We also tabulate the Z statistic and the significance level (p) for the post hoc test in Tables 4-7. In all questions except Question 2, the Friedman test revealed a significant difference in the participant responses depending on the algorithm used. Since the results of the Friedman test for Question 2 were not statistically significant, we did not run a post hoc test for this question. We discuss the results for each question below:

- **Question 1:** Question 1 asked whether participants felt a sense of presence in the virtual environment. In the Wilcoxon signed-rank test for the *PedVR / F2FCrowds*

	PedVR vs F2FCrowds		F2FCrowds vs F2FCrowdsHead		PedVR vs F2FCrowdsHead	
	Z	p	Z	p	Z	p
Question 1	-1.807	0.071	-2.456	0.014	-2.395	0.017
Question 3	-2.291	0.022	-2.618	0.009	-2.713	0.007
Question 4	-2.62	0.009	-2.424	0.015	-2.726	0.006
Question 5	-2.625	0.009	-2.647	0.008	-3.097	0.002
Question 6	-2.155	0.031	-2.369	0.018	-2.747	0.006

Table 4: **Post hoc test for the Circle scene**: We present the Z statistic and the significance level (p) of a post hoc analysis with a Wilcoxon signed-rank test.

	PedVR vs F2FCrowds		F2FCrowds vs F2FCrowdsHead		PedVR vs F2FCrowdsHead	
	Z	p	Z	p	Z	p
Question 1	-2.46	0.014	-1.549	0.121	-2.388	0.017
Question 3	-2.549	0.011	-3.019	0.003	-3.33	0.001
Question 4	-1.652	0.098	-3.315	0.001	-3.315	0.001
Question 5	-1.839	0.066	-3.191	0.001	-3.453	0.001
Question 6	-0.855	0.393	-3.075	0.002	-2.917	0.004

Table 5: **Post hoc test for the Bidirectional scene**: We present the Z statistic and the significance level (p) of a post hoc analysis with a Wilcoxon signed-rank test.

	PedVR vs F2FCrowds		F2FCrowds vs F2FCrowdsHead		PedVR vs F2FCrowdsHead	
	Z	p	Z	p	Z	p
Question 1	-2.537	0.011	-2.14	0.032	-2.782	0.005
Question 3	-2.348	0.019	-2.548	0.011	-3.179	0.001
Question 4	-1.517	0.129	-3.308	0.001	-3.247	0.001
Question 5	-1.687	0.092	-2.913	0.004	-3.098	0.002
Question 6	-1.294	0.196	-1.622	0.105	-2.103	0.035

Table 6: **Post hoc test for the Shopping Mall scene**: We present the Z statistic and the significance level (p) of a post hoc analysis with a Wilcoxon signed-rank test.

comparison, there was no significant difference in the Circle and Shibuya Crossing scenes, but significant difference was observed in the Bidirectional and Shopping Mall scenes. For the *F2FCrowds / F2FCrowdsHead* comparison, significant difference was observed only in the Circle scene. For the *PedVR / F2FCrowdsHead* comparison, significant difference was observed in all the scenes. This proves the **Hypothesis 1**,

	PedVR vs F2FCrowds		F2FCrowds vs F2FCrowdsHead		PedVR vs F2FCrowdsHead	
	Z	p	Z	p	Z	p
Question 1	-1.608	0.108	-2.033	0.042	-2.436	0.015
Question 3	-2.479	0.013	-3.152	0.002	-3.64	0
Question 4	-2.555	0.011	-3.26	0.001	-3.656	0
Question 5	-3.023	0.003	-3.106	0.002	-3.782	0
Question 6	-1.491	0.136	-2.682	0.007	-2.68	0.007

Table 7: **Post hoc test for the Shibuya scene**: We present the Z statistic and the significance level (p) of a post hoc analysis with a Wilcoxon signed-rank test.

which suggests that the users feel a sense of presence in the virtual environment after the addition of both the model of approach behavior and upper body motion generation.

- **Question 2:** Question 2 evaluated the effort required to avoid collisions. Participants reported no difference between the three algorithms for both the scenarios as indicated by the Friedman test. This proves the **hypothesis 2** that when approaching the virtual agents, the users do not have to make extra effort to avoid collisions with the virtual agents after the addition of model of approach behavior and upper body motion generation. To ascertain that the users do not have to make extra effort to avoid collisions with the virtual agents when performing a goal-directed task, we performed another user evaluation. Instead of asking the participants to consciously approach the virtual agents, we asked the participants to follow a goal-directed behavior and answer Question 2 for the Shopping Mall and Shibuya Crossing scenarios. The Friedman test revealed no statistically significant difference for both the scenes (Shopping Mall, $\chi^2(2) = 3.138, p = 0.208$; Shibuya Crossing, $\chi^2(2) = 1.922, p = 0.383$). This proves that the users do not have to make an extra effort to avoid collisions with the virtual agents either while performing a goal-directed task or while consciously approaching virtual agents.

- **Question 3:** Question 3 evaluated whether the participants felt that the virtual agents were aware of the participant. In the Wilcoxon signed-rank test for the *PedVR / F2FCrowds* comparison, there was no significant difference in the Circle and Shopping Mall scenes, but significant difference was observed in the Bidirectional and Shibuya Crossing scenes. For the *F2FCrowds / F2FCrowdsHead* and *PedVR / F2FCrowdsHead* comparisons, significant difference was observed in all the scenes proving the **Hypothesis 3**, which suggests virtual agents appear more aware of the user after the addition of model of approach behavior and upper body motion generation.

- **Question 4:** Question 4 evaluated whether the participants felt that they should talk/nod/respond to the characters. For the *PedVR / F2FCrowds* comparison, post hoc tests did not reveal a significant difference for the Bidirectional and Shopping Mall scenes, but significant difference was observed for the Circle and Shibuya scenes. Significant difference was observed for all the scenes for both *F2FCrowds / F2FCrowdsHead* and *PedVR / F2FCrowdsHead* comparisons. Thus, the results prove the **Hypothesis 4**, which suggests that a combination of the model of approach behavior and upper body motion generation is necessary to elicit a response from the users.
- **Question 5:** Question 5 evaluated whether the virtual agents seemed responsive. For the *PedVR / F2FCrowds* comparison, post hoc tests did not reveal a significant difference for the Bidirectional and Shopping Mall scenes, but significant difference was observed for the Circle and Shibuya Crossing scenes. For the *F2FCrowds / F2FCrowdsHead* and *PedVR / F2FCrowdsHead* comparisons, significant difference was observed in all the scenes. Thus, the results prove the **Hypothesis 5**, which suggests the addition of model of approach behavior and upper body motion generation made the virtual agents appear more responsive.
- **Question 6:** We also asked the participants to report if they felt that the virtual agents

responded even if the participant did not attempt to interact. Significant difference was not observed for the *PedVR / F2FCrowds* comparison in all the scenarios. Thus, the addition of the model of approach behavior does not make the virtual agents appear more responsive when the user does not attempt to interact. For the *F2FCrowds / F2FCrowdsHead* comparison, significant difference was observed for the Bidirectional and Shibuya Crossing scenes, but no significant difference was revealed for the Circle and Shopping Mall scenes. For the *PedVR / F2FCrowdsHead* comparison, all the scenes except the Shopping Mall scene showed significant difference. Thus, in most cases, the combination of model of approach behavior and upper body motion generation makes the virtual agents appear more responsive when the user does not attempt to interact. Some participants reported after the experiment that this made the virtual agents appear more "friendly" but further investigation is necessary.

Effect of Density The four scenes used in user evaluation also had varying conditions of pedestrian densities. The Circle scene included an area of high crowd density near the center. The Bidirectional scene had two groups of virtual agents starting from opposite ends of a hallway and areas of high density were formed when the agents crossed each other. The Shopping Mall scene had a smaller walking area than the other scenes and had a high density, whereas the Shibuya Crossing scene had a low density. Despite the variations in density, in all the four scenarios, we observed that the addition of the approach algorithm and the upper body behaviors contributed to the quality of face-to-face interactions.

6. Conclusion, Limitations, and Future Work

In this paper, we have presented techniques to compute the movements and trajectories of virtual agents to enable face-to-face interactions as part of a crowd. This includes an automatic approach for interaction velocity prediction, which we use to compute a collision free velocity. We further augment the approach by simulating many upper body behaviors and movements. Our approach can simulate crowds with tens of agents at interactive rates, with support for F2F communications between the real user and virtual agents. We also performed a user study and concluded that our new algorithms increase the sense of social presence in virtual environments. The virtual agents using our algorithms also appeared more responsive and were able to elicit more reaction from the users.

Our approach has some limitations. In particular, our criteria to trigger F2F interactions do not take into account the agent's personality or emotions or the social norms. Furthermore, we only support a limited number of upper body movements or gestures. It would be useful to support verbal communication or conversations between the agents to increase the level of interaction. We would also like to model social signals like turn-taking and backchanneling, which are an important part of F2F interactions. We would like to evaluate our approaches in more complex scenarios, and compare with real-world scenarios. We use Oculus Rift to take user input. Since the walking area of Rift is limited, the users have to use a joystick or a keyboard, which puts constraints on the realism of face-to-face interactions. We would like to use a wide area tracker framework to allow the real user to walk large distances in the physical world. In this work, we do not consider the case where multiple real users share the same virtual space; ideas from our approach can be combined with work from Pedica and Vilhjálmsson (2009) to handle this case. At any instant, the real user in our approach can have a face-to-face interaction with at most one virtual agent. A modified version of our approach can enable face-to-face interactions with more than one virtual agent and we plan to implement that as part of a future work. Implementing it would involve implementing group formation and mathematical modeling of social and psychological ideas about group behaviors (He, Pan, Wang, & Manocha, 2016; Knowles & Bassett, 1976). We assume that the real user is not familiar with the virtual agents and treats them as strangers. In applications like games, the user may know the virtual character and he/she may approach them in a different manner (e.g., calling them by their name). For this paper, we do not allow the real users to express their intention to interact verbally. In many cases, it may be possible that the system allows for verbal input and the user can just announce their intent to a virtual agent verbally. Our approach can still be implemented in these systems as a supporting/complementary mechanism.

Our approach uses gaze as a mechanism to acknowledge the user's presence and to initiate a face-to-face interaction. This mechanism may or may not be compatible with other behaviors that also use gazing (e.g., avoiding another person). In such a case, it might be better to have verbal communication (e.g., saying "hi" or "hello") along with gazing behavior. The virtual agents will pursue an F2F interaction only until the user's attention (as denoted by his/her orientation) is on it. If the user's orientation changes above a certain threshold, the virtual agent will conclude the face-to-face conversation to be over and continue its goal-directed navigation in the scene. This is a simplified approach to deduce the end of a conversation and further literature from psychology/sociology (Alterman & Garland, 2001; Bangerter, Clark, & Katz, 2004) can be used to design a more advanced strategy to recognize the end of a conversation.

Acknowledgements

This work was supported by ARO contract W911NF-16-1-0085.

References

Alterman, R. & Garland, A. (2001). Convention in joint activity. *Cognitive Science*, *25*(4), 611–657.

Bailenson, J. N., Blascovich, J., Beall, A. C., & Loomis, J. M. (2001). Equilibrium theory revisited: Mutual gaze and personal space in virtual environments. *Presence: Teleoperators and Virtual Environments*, *10*(6), 583–598.

Bangerter, A., Clark, H. H., & Katz, A. R. (2004). Navigating joint projects in telephone conversations. *Discourse Processes*, *37*(1), 1–23.

Bera, A., Kim, S., Randhavane, T., Pratapa, S., & Manocha, D. (2016). GLMP-realtime pedestrian path prediction using global and local movement patterns. In *Robotics and Automation (ICRA), 2016 IEEE International Conference on* (pp. 5528–5535). IEEE.

Bera, A., Randhavane, T., & Manocha, D. (2017). Aggressive, Tense or Shy? Identifying Personality Traits from Crowd Videos. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17* (pp. 112–118). doi:10.24963/ijcai.2017/17

Bera, A., Randhavane, T., Prinja, R., & Manocha, D. (2017). SocioSense: Robot Navigation Amongst Pedestrians with Social and Psychological Constraints. In *Intelligent Robots and Systems (IROS), 2017 IEEE/RSJ International Conference on*. IEEE.

Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., & Bailenson, J. N. (2002). Immersive virtual environment technology as a methodological tool for social psychology. *Psychological Inquiry*, *13*(2), 103–124.

Bonaiuto, J. & Thórisson, K. R. (2008). Towards a Neurocognitive Model of Realtime Turntaking in Face-to-Face Dialogue. In *Embodied Communication in Humans and Machines*. Oxford University Press.

Bonsch, A., Weyers, B., Wendt, J., Freitag, S., & Kuhlen, T. W. (2016, March). Collision avoidance in the presence of a virtual agent in small-scale virtual environments. In *2016 IEEE Symposium on 3D User Interfaces (3DUI)* (pp. 145–148). doi:10.1109/3DUI.2016.7460045

Boulic, R. (2005). Reaching oriented targets with funnelling trajectories. *Proceedings of V-CROWDS*, 24–25.

Bršćić, D., Kidokoro, H., Suehiro, Y., & Kanda, T. (2015). Escaping from Childrens Abuse of Social Robots. In *10th ACM/IEEE International Conference on Human-Robot Interaction HRI 2015*.

Cassell, J., Vilhjálmsson, H. H., & Bickmore, T. (2001). BEAT: The Behavior Expression Animation Toolkit. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques* (pp. 477–486). SIGGRAPH '01. New York, NY, USA: ACM. doi:10.1145/383259.383315

Cassol, V., Oliveira, J., Musse, S. R., & Badler, N. (2016, March). Analyzing egress accuracy through the study of virtual and real crowds. In *2016 IEEE Virtual Humans and Crowds for Immersive Environments (VHCIE)* (pp. 1–6). doi:10.1109/VHCIE.2016.7563565

Clark, H. H. (1996). *Using language*. Cambridge university press.

Crowley, D. J. & Mitchell, D. (1994). *Communication theory today*. Stanford University Press.

Curtis, S., Best, A., & Manocha, D. (2016). Menge: A modular framework for simulating crowd movement. *Collective Dynamics*, *1*, 1–40.

Funge, J., Tu, X., & Terzopoulos, D. (1999). Cognitive modeling: knowledge, reasoning and planning for intelligent characters. In *SIGGRAPH* (pp. 29–38). ACM Press.

Garau, M., Slater, M., Pertaub, D.-P., & Razzaque, S. (2005, February). The Responses of People to Virtual Humans in an Immersive Virtual Environment. *Presence: Teleoperators and Virtual Environments*, *14*(1), 104–116. doi:10.1162/1054746053890242

Gonçalves, B. & Perra, N. (2015). *Social phenomena: From data analysis to models*. Springer.

Grillon, H. & Thalmann, D. (2009). Simulating gaze attention behaviors for crowds. *Computer Animation and Virtual Worlds*, *20*(2-3), 111–119.

Guadagno, R. E., Blascovich, J., Bailenson, J. N., & Mccall, C. (2007). Virtual humans and persuasion: The effects of agency and behavioral realism. *Media Psychology*, *10*(1), 1–22.

Guy, S. J., Kim, S., Lin, M. C., & Manocha, D. (2011). Simulating Heterogeneous Crowd Behaviors Using Personality Trait Theory. In *Proceedings of the 2011 ACM SIGGRAPH/Eurographics Symposium on Computer Animation* (pp. 43–52). SCA '11. Vancouver, British Columbia, Canada: ACM. doi:10.1145/2019406.2019413

Hadar, U., Steiner, T. J., & Clifford Rose, F. (1985). Head movement during listening turns in conversation. *Journal of Nonverbal Behavior*, *9*(4), 214–228. doi:10.1007/BF00986881

Hall, E. T. (1966). The hidden dimension .

Haworth, B., Usman, M., Berseth, G., Khayatkhoei, M., Kapadia, M., & Faloutsos, P. (2016, March). Using synthetic crowds to inform building pillar placements. In *2016 IEEE Virtual Humans and Crowds for Immersive Environments (VHCIE)* (pp. 7–11). doi:10.1109/VHCIE.2016.7563566

He, L., Pan, J., Wang, W., & Manocha, D. (2016). Proxemic group behaviors using reciprocal multi-agent navigation. In *Robotics and Automation (ICRA), 2016 IEEE International Conference on* (pp. 292–297). IEEE.

Helbing, D. & Molnar, P. (1995). Social force model for pedestrian dynamics. *Physical Review E*, *51*(5), 4282.

Heylen, D., Bevacqua, E., Pelachaud, C., Poggi, I., Gratch, J., & Schröder, M. (2011). Generating Listening Behaviour. In *Emotion-Oriented Systems: The Humaine Handbook* (pp. 321–347). Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-642-15184-2_17

Hu, R., Adeagbo, M., Interrante, V., & Guy, S. (2016, March). Virtual Human Head Turning with Crowd Simulation. In *2016 IEEE Virtual Humans and Crowds for Immersive Environments (VHCIE)*.

Jonsdottir, G. R., Thorisson, K. R., & Nivel, E. (2008). Learning Smooth, Human-Like Turntaking in Realtime Dialogue. In *Intelligent Virtual Agents: 8th International Conference, IVA 2008, Tokyo, Japan, September 1-3, 2008. Proceedings* (pp. 162–175). Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-540-85483-8_17

Kapadia, M. & Badler, N. I. (2013). Navigation and steering for autonomous virtual humans. *Wiley Interdisciplinary Reviews: Cognitive Science*.

Knowles, E. S. & Bassett, R. L. (1976). Groups and crowds as social entities: Effects of activity, size, and member similarity on nonmembers. *Journal of Personality and Social Psychology*, *34*(5), 837.

Kopp, S., Stocksmeier, T., & Gibbon, D. (2007). Incremental Multimodal Feedback for Conversational Agents. In *Intelligent Virtual Agents: 7th International Conference, IVA 2007 Paris, France, September 17-19, 2007 Proceedings* (pp. 139–146). Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-540-74997-4_13

Kyriakou, M., Pan, X., & Chrysanthou, Y. (2015). Interaction with virtual agentsComparison of the participants' experience between an IVR and a semi-IVR system. In *2015 IEEE Virtual Reality (VR)* (pp. 217–218). IEEE.

Llobera, J., Spanlang, B., Ruffini, G., & Slater, M. (2010). Proxemics with multiple dynamic characters in an immersive virtual environment. *ACM Transactions on Applied Perception (TAP)*, *8*(1), 3.

Magnenat-Thalmann, N. & Thalmann, D. (2005). *Handbook of virtual humans*. John Wiley & Sons.

McClave, E. Z. (2000). Linguistic functions of head movements in the context of speech. *Journal of Pragmatics*, *32*(7), 855–878.

Musse, S. R., Garat, F., & Thalmann, D. (1999). Guiding and interacting with virtual crowds. *Computer Animation and Simulation*, 23–33.

Nagendran, A., Pillat, R., Kavanaugh, A., Welch, G., & Hughes, C. (2014). A unified framework for individualized avatar-based interactions. *Presence: Teleoperators and Virtual Environments*, *23*(2), 109–132.

Narang, S., Best, A., Randhavane, T., Shapiro, A., & Manocha, D. (2016). PedVR: Simulating Gaze-Based Interactions between a Real User and Virtual Crowds. In *Proceedings of the 22nd ACM Symposium on Virtual Reality Software and Technology*. VRST ’16.

Narang, S., Randhavane, T., Best, A., Shapiro, A., & Manocha, D. (2016). Fbcrowd: Interactive multi-agent simulation with coupled collision avoidance and human motion synthesis. Technical report, Department of Computer Science, UNC Chapel Hill.

Nummenmaa, L., Hyönä, J., & Hietanen, J. K. (2009). I’ll walk this way: Eyes reveal the direction of locomotion and make passersby look and go the other way. *Psychological Science*, *20*(12), 1454–1458.

Ólafsson, S., Bédi, B., Helgadóttir, H. E., Arnbjörnsdóttir, B., & Vilhjálmsson, H. H. (2015). Starting a Conversation with Strangers in Virtual Reykjavik: Explicit Announcement of Presence. In *Proceedings of 3rd European Symposium on Multimodal Communication*.

Olivier, A.-H., Bruneau, J., Cirio, G., & Pettré, J. (2014). A virtual reality platform to study crowd behaviors. *Transportation Research Procedia*, *2*, 114–122.

Ondřej, J., Pettré, J., Olivier, A.-H., & Donikian, S. (2010). A synthetic-vision based steering approach for crowd simulation. In *ACM Transactions on Graphics (TOG)* (Vol. 29, *4*, p. 123). ACM.

Pantic, M., Cowie, R., D'Errico, F., Heylen, D., Mehu, M., Pelachaud, C., … Vinciarelli, A. (2011). Social Signal Processing: The Research Agenda. In *Visual Analysis of Humans: Looking at People* (pp. 511–538). London: Springer London. doi:10.1007/978-0-85729-997-0_26

Park, C., Ondřej, J., Gilbert, M., Freeman, K., & O'Sullivan, C. (2016). HI Robot: Human intention-aware robot planning for safe and efficient navigation in crowds. In *Intelligent Robots and Systems (IROS), 2016 IEEE/RSJ International Conference on* (pp. 3320–3326). IEEE.

Pedica, C. & Vilhjálmsson, H. (2009). Spontaneous avatar behavior for human territoriality. In *Intelligent Virtual Agents* (pp. 344–357). Springer.

Pelechano, N., Allbeck, J. M., & Badler, N. I. (2007). Controlling individual agents in high-density crowd simulation. In *Proceedings of the 2007 ACM SIGGRAPH/Eurographics symposium on Computer animation* (pp. 99–108). Eurographics Association.

Pelechano, N., Stocker, C., Allbeck, J., & Badler, N. (2008). Being a part of the crowd: towards validating VR crowds using presence. In *Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems-Volume 1* (pp. 136–142).

Pertaub, D.-P., Slater, M., & Barker, C. (2002, February). An Experiment on Public Speaking Anxiety in Response to Three Different Types of Virtual Audience. *Presence: Teleoperators and Virtual Environments*, *11*(1), 68–78. doi:10.1162/105474602317343668

Reynolds, C. W. (1999). Steering behaviors for autonomous characters. *Game Developers Conference*.

Satake, S., Kanda, T., Glas, D. F., Imai, M., Ishiguro, H., & Hagita, N. (2009, March). How to approach humans?-strategies for social robots to initiate interaction. In *2009 4th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 109–116). doi:10.1145/1514095.1514117

Shapiro, A. (2011). Building a Character Animation System. In *Motion in Games* (Vol. 7060, pp. 98–109). Lecture Notes in Computer Science. Springer Berlin / Heidelberg.

Slater, M., Guger, C., Edlinger, G., Leeb, R., Pfurtscheller, G., Antley, A., … Friedman, D. (2006). Analysis of physiological responses to a social situation in an immersive virtual environment. *Presence: Teleoperators and Virtual Environments*, *15*(5), 553–569.

Ulicny, B. & Thalmann, D. (2001). Crowd simulation for interactive virtual environments and VR training systems. In *Computer Animation and Simulation 2001* (pp. 163–170). Springer.

Van Den Berg, J., Guy, S. J., Lin, M., & Manocha, D. (2011). Reciprocal n-body collision avoidance. In *Robotics research* (pp. 3–19). Springer.

Vinciarelli, A., Pantic, M., & Bourlard, H. (2009). Social signal processing: Survey of an emerging domain. *Image and Vision Computing*, *27*(12), 1743–1759. Visual and multimodal analysis of human spontaneous behaviour:

doi:http://dx.doi.org/10.1016/j.imavis.2008.11.007