Proceedings of the Ninth AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment

Creating Playable Social Experiences through Whole-Body Interaction with Virtual Characters

Daniel Shapiro, Josh McCoy, April Grow, Ben Samuel, Andrew Stern, Reid Swanson, Mike Treanor, Michael Mateas
UCSC Center for Games and Playable Media
{dgshapiro, mccoyjo, bsamuel, reid, mtreanor, michaelm}@soe.ucsc.edu, {sterno888, grow.april}@gmail.com

Abstract

This paper describes work towards the goal of enabling unscripted interaction with non-player characters in virtual environments. We hypothesize that we can define a layer of social affordances, based on physical and non-verbal signals exchanged between individuals and groups, which can be reused across games. We have implemented a first version of that substrate that employs whole body interaction with virtual characters and generates nuanced, real-time character performance in response. We describe the playable experience produced by the system, the implementation architecture (based on the behavior specification technology used in Façade, the social model employed in Prom Week, and gesture recognition technology), and illustrate the key behaviors and programming idioms that enable character performance. These idioms include orthogonal coding of attitudes and activities, use of relational rules to nominate social behavior, use of volition rules to rank options, and priority based interleaving of character animations.

Introduction

Social interaction is critical to human experience, but it is rarely addressed in games. Social interaction is difficult to capture in general form, because it requires a model of social state that characters can sense, update, and employ to drive behavior. A few games, like Prom Week (Mccoy et al., 2013) and Versu (Linden Lab, 2013) have built such models, but outside the context of real-time interaction where it could be employed to add a new dimension of realism to game characters. Social interaction also requires generating and recognizing expressions of attitude and style during task conduct. In the context of a game experience, this implies handling a range of player input that is challenging to manage. Façade (Mateas & Stern, 2005) was novel, in part, because it demonstrated the use of a mechanic (a reactive programming language) to address this problem, given textual input. Other work on virtual human interactions faces similar challenges: Gunslinger features movement and exploration of a virtual world via rich multi-modal input but lacks social simulation (Hartholt, Gratch, & Weiss, 2009); some experiences feature deep social interaction with limited or no player body interaction (Babu et al., 2005; Johnson et al., 2004; Kim, Jr, & Durlach, 2009); while other systems enable social interaction with a small cast of virtual characters via partial body interaction (Kistler, André, & Mascarenhas, 2013).

This paper describes an approach to real-time whole-body interaction with social characters that is primarily unscripted and player-driven. The application context is a game for training social interaction skills across language and culture gaps, specifically for soldiers deployed in foreign countries. Here, the central lesson is that good social interaction begets good functional outcome, while bad social interaction has potentially disastrous consequences.

Towards this end, we have isolated a domain of social interaction that is relevant across a wide variety of games, and implemented the first version of a substrate for unscripted interactions of this kind. These interactions concern physically observable social state and signals, exchanged between a player and individual characters or groups. Examples include group entry, exit, group conversation, focus of attention, gaze control, and affect that can be applied to character performance. Our system inputs gestures and speech, and outputs real-time character performance in an immersive environment. Our technical approach augments the A Behavior Language (ABL) engine underlying Façade with the social model, Comme il Faut (CiF), underlying Prom Week, and implements a

Figure 1. Two snapshots of player-group interaction in the lost interpreter scenario: (a) getting the group's attention, and (b) conversing as a member of the group.

behavior authoring architecture using these parts. We call the resulting system a social dynamics simulator.

The Playable Experience

We have employed the social dynamics simulator to create a playable experience that showcases a novel level of social interaction between the player and AI-driven characters. In this demonstration, the player's goal is to find their lost interpreter, in a marketplace within a Mediterranean town that is known to be dangerous. Additionally, there is a significant language barrier as the denizens of the marketplace do not speak English. The physical setup of the social simulator has the player standing in front of, and facing a large screen (which shows a first person perspective of the world to the player), a Microsoft Kinect and a high-definition camera. The player employs a Nintendo Wii Remote to accomplish broad movement around the virtual world. The player is otherwise unencumbered, and can move and gesture as they normally would in everyday social interaction.

The player's primary actions involve approaching groups of characters and acquiring information from them (see figure 1). However, because of the language barrier, this involves carefully observing the social structure of groups and engaging with them in a socially appropriate manner. The experience requires interacting with two groups that have different attitudes, which the player must recognize and consider in order to acquire information about where to find the lost interpreter.

When the player approaches the first group, they will acknowledge her presence by turning and facing her. After performing a greeting, the group will make space for her to stand with them, but the characters clearly cannot understand any verbal inquiries. The player can explore several options to communicate her intent, like displaying a photo of the interpreter, or pointing at possible locations. Members of the group might ask for the photo to get a closer look and pass it around from character to character. Once a character understands that the player is looking for the person in the photo, they will point her toward another group, indicating that they may have more information.

While the first group was friendly and inviting, the second group is hostile and sensitive to the player's gestures and physical distance. These characters are not interested in inviting the player to stand with them, looking at the photo, or generally helping the player at all. They speak harshly to the player and will make sharp gestures. If the player is able to recognize that the members of the group generally defer to a group leader, she is able to get him to help by making a direct, and emotional appeal. At this point, the leader understands the player, and points to a far corner of the marketplace where the lost interpreter can be found. However, when the player moves towards the interpreter, a crowd intervenes and the interpreter disappears. The game continues.

Implementation Architecture

Figure 2 presents the architecture of the social simulator, illustrating the path from a physical performance by a participant through its social interpretation, to the behavior generated by the NPCs.

We use a Kinect to capture player input, and a special purpose recognition system to process that data. The gesture recognition system produces two types of output signals. First it produces labeled events, such as *left arm wave* and *two-handed bow*. Second, it outputs a heat map of recent activity on regions of the skeleton. While the relatively low fidelity of the Kinect is well suited for coarse joint positioning, it lacks the ability to detect

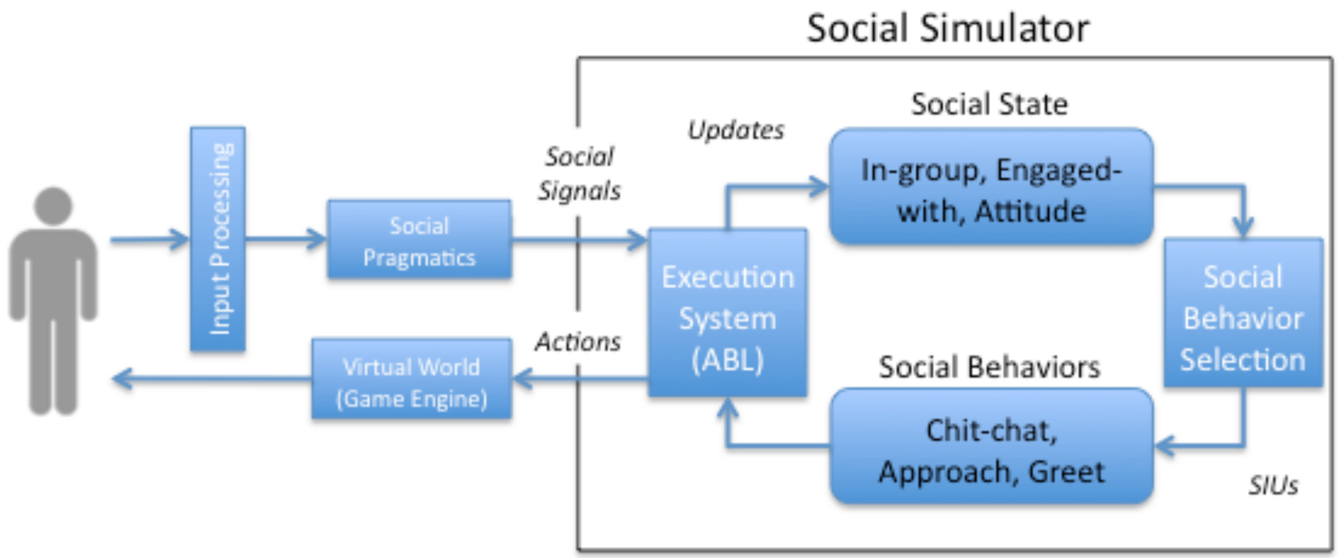

Figure 2. Social Simulator Architecture

nuanced differences in facial expressions that carry rich emotional and social meaning. To capture these types of expressions we employ a high-definition camera that tracks the player's head as it moves through the field of view. We also utilize a microphone to listen for audio and perform basic speech to text using an off-the-shelf component.

The output of these recognition components is input into a *social pragmatics* module that generates multiple social interpretations of a physical signal. These interpretations can be mutually exclusive or overlapping, but they are all passed to the simulator, which makes an appropriate selection in the context of ongoing behaviors. The module employs a set of hand-authored rules implemented in a production rule engine.

The signals generated by social pragmatics are further interpreted in the context of ongoing behaviors to update social state. The system represents social state in terms of character statuses and relations between characters, e.g., that the player is engaged with a group, that characters are members of a group, or possess a given attitude. The representation adapts and generalizes the social model of CiF into a relational language that is embedded in ABL.

Social behavior selection chooses behaviors based on current social state, and social signals obtained from the world. Behaviors are authored in ABL, and their execution supplies real-time interactive performance. Subsequent player input drives the trajectory of the social dynamics simulation as a whole.

Idioms

The goal of the simulator is to create an unscripted interaction that dynamically responds to the signals produced by the player in an impactful and socially appropriate way. This is a complex undertaking that involves several interdependent processes working together. At a minimum, the virtual characters must perform the functionally appropriate behavior in response to the player in order to complete the desired action given the context. However, to create a socially believable environment the embodied performance of these actions by the NPCs should reflect their individual attitudes and the implied norms of the virtual culture we are simulating. Finally, the experience must be dramatically compelling so that players are personally invested in the success or failure of the interaction.

We have implemented a number of programming idioms in ABL to support authoring. These consist of social behaviors (called social interaction units, or SIUs), an behavior selection process that chooses SIUs, a performance manager that interleaves SIU execution, a "wrap-on" mechanism that tailors SIU performance to reflect character attitude, and continuous behaviors that maintain a sense of character aliveness.

Social Interaction Units are akin to beats in Façade (Mateas & Stern, 2005) and social exchanges in Prom Week. An SIU is designed to be as small as possible to accomplish a single, discrete social goal, such as a reaction to the player's approach, exchanging an object, or momentarily discussing a question. It typically lasts from 5 to 10 seconds.

An SIU has a primary initiator and one or more responders, one of whom may be a primary responder. For example, the Greet SIU has a greeter (initiator), and one or more characters being greeted (responders). SIUs may take parameters, such as a character to add or remove from a group formation, a gesture to mirror, or a question being asked. SIUs can be built to allow either the player or an NPC to be the initiator or a responder.

SIUs are nominated by decision rules that react to stored social state, plus interpretations of social signals obtained from other characters (typically the player). SIUs can also be launched autonomously as self-motivated behavior, e.g., to maintain a group formation.

Each NPC is typically involved in multiple simultaneous active SIUs at the same time. For example, ManageFormation, ChitChat, StudyOthersObject, and ReactToQuestion may all be active simultaneously.

The behavior selection process performs social reasoning by deciding which SIUs to execute and when. The mechanism employs three types of production rules, added to ABL for this purpose. These rules leverage the existing pattern matching capabilities of ABL, but provide new functionality for asserting and processing new working memory elements. The first are *nomination* rules. These rules assert an SIU decision object into working memory when its preconditions hold in the social context. Nomination rules execute whenever a new social interpretation is introduced to the simulator through the input channel, or a specified amount of time elapses. This is the first step of the behavior selection process. At this stage, the resulting SIU decision objects only indicate that the associated behaviors are possible given the social context, but they are not immediately executed.

The second type of rules integrated into ABL represent desirability, or *volition*. A given volition rule matches an SIU decision object plus additional state data. The action side of the rule increments or decrements a preference value for the execution of that SIU. The matching rules produce a single aggregated score for each SIU, and any SIUs above a static threshold are marked for execution. Authors tailor the volition rules to produce an acceptable number of candidates.

The third type of rule matches a set of relations on the left hand side and either asserts or deletes any number of standard ABL working memory elements on the right hand side. This is done for all matching variable bindings. These rules generalize the trigger functionality in CiF and are fired at a specific moment in the decision process. A separate parallel process continuously monitors SIU objects marked for execution, and launches the corresponding behavior.

The ABL execution system maintains a tree of active behaviors (the ABT) that refines SIUs into their component subgoals and primitive actions. This system processes the ABT to produce behavior (Mateas and Stern, 2002).

The performance manager interleaves SIU execution using an author-defined priority assigned to each SIU. Priority tiers are heuristically tuned to favor more important performance requests such as reacting quickly to a gesture from the player, over less urgent performance such as maintaining group formation or chitchatting.

If the Performance Manager suspends a less important ongoing performance to allow a more important request to perform immediately, the suspended performance will typically re-run when the interrupting performance has completed. Since each performance request typically lasts no more than a few seconds, re-running interrupted performances often maintains coherence and believability.

Future work includes annotating performance requests to allow for more sophisticated, dynamic reasoning about the relative importance of competing requests, as well as animation blending of performance requests to achieve multiple performances simultaneously.

The wrap-on mechanism addresses the problem of encoding character performance given a wide variety of behaviors and attitudes. A naïve approach to capturing these variations quickly leads to an intractable amount of authoring, so we focus, instead, on building reusable social affordances that can be authored independently and recombined dynamically to produce the desired performance range. The wrap-on mechanism does this by layering expressions of attitude onto behavior.

In more detail, we author the core functionality of each SIU in an affect-neutral way, comprised of very short functional performances. For example, ExchangeObject has the initiator holding out the object, waiting for the responder to reach for it, which finally completes the exchange. Next, we write behaviors that can perform these low-level actions with a variety of attitudes. This can be as simple as defining attitudinal variations of individual animations, such as holding out an object in a friendly, aggressive or hesitant way. The wrap-on manager acts by inserting these animations before, after, and/or in place of the core performance, such as impatient sighs, frustrated gestures, or nervous glances, using the meta-behavior capabilities in ABL to alter the structure of the ABT.

Finally, **continuous behaviors** execute when a character is not currently engaged in a social interaction with the player or another NPC. These are simple behaviors that breathe life into the characters, like small talk between NPCs, or expressions of their current emotional state such as stamping their feet to illustrate impatience.

Domain Content

We have employed the architecture and idioms discussed above to express the suite of social signals, social state, SIUs, wrap-ons and continuous behaviors underlying the lost interpreter scenario. We illustrate these, below.

The social pragmatics module maps the verbal and non-verbal behavior of the player into one or more of the following social signals: Approach, Greet, Agree, Refuse, Thank, ShowObject, OfferObject, TakeObject, RequestObject, PointAtObject, StudyObject, Beckon, PointInDirection, EmotionalAppeal, MoveRequest and Dismiss. Examples include the player extending their right hand in front of her body being interpreted as ShowObject.

Domain specific social state labels are represented and manipulated in the social simulator to keep track of the social aspects of the virtual world. These labels include Engaged, which represents the player being an active participant in a group of characters and attitudes such as Hostile that are used in SIU selection and wrap-on configuration.

As discussed previously, SIUs are a core representational construct that enables characters to perform a range of behaviors consistent with a social purpose. The behavior of each participant within an SIU is controlled by a tree of subgoals that are executed in parallel. Each performance behavior typically consists of one or two animations, lasting a few seconds. In addition, every active SIU requests and continuously vies for performance control over its NPC participants. The lost interpreter scenario contains several SIUs that involve an initiating character and multiple responding characters: ManageFormation, ChitChat, Approach, Greet, MirrorGesture, StudyOthersObject, ReactToQuestion, ReactToAppeal, DiscussQuestion and AnswerQuestion. One-on-one SIUs are: RequestObject, ExchangeObject, StudyOwnObject and PersonalSpaceViolation. PassObjectAround is a hierarchical SIU that sequences other SIUs together.

We have implemented a number of wrap-ons for expressing character attitudes. Characters can be hostile, suspicious, friendly, sympathetic, frustrated, happy, and nervous, and these attitudes can each be applied to multiple SIUs. More exactly, the object of a wrap-on is a fine-grained behavior within an SIU that invokes an animation, specified by its class. Each wrap-on behavior pair has multiple realizations, expressed by insertions of a preamble step, a postlude, and/or a specific instantiation of the animation class. For example, a hostile instance of returning the picture could involve a pause (preamble), a rapid extension of the hand (the instantiation), and a shake (postlude), or a cough, a sneer plus a slow extension, and a rapid retraction of the hand in sequence. Wrap-ons apply to behavior instances, so an NPC can greet one character with hostility, and another in a friendly way.

The continuous behaviors in the social simulator keep the characters active to give them the impression of an inner life. The major categories of continuous behaviors are group arrangement, attention management, gaze control and formation. The non-player characters in many games and interactive experiences lack these capabilities. In more detail, groups of varying size perform acts of chit-chat, or small-talk, in order to enliven background behaviors. Within a group, a single character speaks and demands the attention of others in their circle. Speaking characters gesture more emphatically along with speaking dialogue, while listeners are more subdued. When a speaker is done, they pass the speaking torch to another character in their circle, and the cycle begins anew. Side-conversations between two adjacent characters will emerge periodically, expressed through close leaning and focused attention diverted away from the group's primary speaker. Previously described attitude wrap-ons supply a huge amount of additional variation: jovial groups gesticulate more wildly and are more agreeable, while nervous participants shuffle in-place, avert their gaze, and make few open gestures.

All characters can sense the world without restriction, so attention management is a means of expressing concerns internal to the character without words. As briefly mentioned when describing chit-chat, a character's primary focus is on their group's current speaker or side-conversationalist. If a character is currently speaking, they split their attention between all other group members. When the player intrudes upon a group, the player holds the attention of every member in the group except during side conversation interruptions, or when an object holds the characters' attention.

Characters primarily use gaze and eye contact to express their attention to the player. If attention is split between multiple sources, periodic glances between the multiple objects occur. If an NPC is nervous and avoiding eye contact, the orientation of their body still expresses where their gaze would be. Because the player is a subject of interest for NPCs, they often glance in the player's direction. Wrap-ons also provide variation here: suspicious groups will all periodically turn to gaze toward the player outside their group before immediately converging for a quick, hushed version of chit-chat. Standard groups will simply glance in the player's direction.

Discussion

The goal of this research is to develop a layer of social competence for virtual characters that can be reused across games. We have demonstrated the feasibility of producing one such substrate but it is too early to show generality, as that would require us to capture a much wider range of physical social phenomena and utilize them across gaming contexts (and cultural settings for the training application). However, we can identify three qualitatively new player experiences that the social dynamics simulator enables by way of evaluation:

- whole body interaction with virtual characters,
- use of physical social signals to communicate with individuals and groups, and
- nuanced, real-time interaction with characters that pursue multiple social objectives in parallel.

We can trace the source of power behind these experiences to key elements of the architecture. It utilizes inference to transform gestures into socially meaningful signals, an explicit social model to capture social state and operators, a reactive language to encode real-time character performance, and idioms within that language to support the pursuit of simultaneous objectives, as well as the semi-independent specification of attitudes and activities.

Related Work

Core to the social simulation are A Behavior Language (ABL) and Comme il Faut (CiF). ABL is a reactive planning language that does not require a commitment to a formal domain model (Mateas & Stern, 2002). ABL serves as the glue for our integrated agent. Different competencies within an agent can make use of distinct problem solving techniques, and communicate with one another through ABL's working memory. The social AI system, CiF, weaves concepts from the social sciences with existing AI, computer science, and software engineering knowledge to enable the possibility of creating a richly realized interactive media experience while taming the complexities of authoring socially believable characters (McCoy et al., 2010). Dramaturgical analysis (Goffman, 1974) is the basis for extending the power of SIUs from their initial form in *Façade* (Mateas & Stern, 2005), where their variation in performance was implicitly encoded in behaviors and were not reusable between characters, to their current, reusable form that explicitly supports performance variation.

The history of interactive experiences with social characters is short, but it heavily influenced our work. The social AI system, CiF, from Prom Week (Mccoy et al., 2013), served as the inspiration for the next evolution of ABL, our behavior authoring language. The Sims 3 and Versu are commercial games that feature a high level of social play in either abstract or turn-based fashions. Though the internal social model is lighter, SOLVE (Habonneau, Richle, Szilas, & Dumas, 2012) and FearNot! (R. S. Aylett, Louchart, Dias, Paiva, & Vala, 2005) enable play in certain social venues.

Several systems that have been created or proposed provide support for non-interactive social behavior in virtual characters (Endrass, Damian, Huber, Rehm, & André, 2010; Marsella, Pynadath, & Read, 2004; Osborne, 2011; Vilhjalmsson & Marsella, 2005). In contrast, our social simulator features both a rich social model and embodied player interaction.

The social simulator is one of the first examples of real time physical interaction with social characters, but prior work has been done in a similar vein. For example, Aylett, et al. (R. Aylett, Vannini, Andre, & Paiva, 2009) employ Dancepads and a Wii remote to support player navigation and gesture in a virtual space, in a game focused on learning the foreign culture of an alien race in an attempt to save their planet. The elements of this interaction form the basis of our work, which we are extending in the dimensions of gesture recognition, social modeling, and character performance.

Future Work

This work makes the claim that behaviors surrounding body language cover a wide range of social interactions that will be useful in games. These behaviors are well-studied in the social sciences (Knapp & Hall, 2010), but the utility for games is an open question. We plan to test that hypothesis in future work by extending the social dynamics substrate to address social timing, incremental engagement in social activities, and to enhance our methods for recognizing affect in player motion, and generating affect in synthetic characters. This work will involve several planned extensions to the ABL language that capture character choice points and decision criteria. We will apply these capabilities to explore new types of player experience.

Acknowledgements

We gratefully acknowledge the contribution to this work of our collaborators at BBN Raytheon Technologies, SRI International at Princeton, and SoarTech. This work was sponsored by the DARPA Strategic Social Interaction Modules program, under U.S. Army Research Office contract W911NF-12-C-0002. The contents of this paper are the responsibility of the authors, and do not necessarily reflect the position or the policy of the US Government. No official endorsement should be inferred.

References

Aylett, R. S., Louchart, S., Dias, J., Paiva, A., & Vala, M. (2005). FearNot!: An experiment in emergent narrative. In T. Panayiotopoulos, J. Gratch, R. Aylett, D. Ballin, P. Olivier, & T. Rist (Eds.), *Proceedings of the Fifth International Conference on Intelligent Virtual Agents* (Vol. 3661, pp. 305–316). Springer Berlin Heidelberg.

Aylett, R., Vannini, N., Andre, E., & Paiva, A. (2009). But that was in another country: agents and intercultural empathy. *... on Autonomous Agents ...*, 10–15.

Babu, S., Schmugge, S., Inugala, R., Rao, S., Barnes, T., & Hodges, L. F. (2005). Marve: A Prototype Virtual Human Interface Framework for Studying Human-Virtual Human Interaction. In Proceedings of Intelligent Virtual Agents (pp. 120–133).

Endrass, B., Damian, I., Huber, P., Rehm, M., & André, E. (2010). Generating Culture-specific Gestures for Virtual Agent Dialogs. In J. Allbeck, N. Badler, T. Bickmore, C. Pelachaud, & A. Safonova (Eds.), *Intelligent Virtual Agents* (Vol. 6356, pp. 329–335–335). Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-642-15892-6

Goffman, E. (1974). *Frame Analysis*. (K. Brown, Ed.) (p. 586). Harvard University Press. doi:10.1002/mar.4220020104

Habonneau, N., Richle, U., Szilas, N., & Dumas, J. (2012). 3D simulated interactive drama for teenagers coping with a traumatic brain injury in a parent. *Interactive Storytelling*, 174–182.

Hartholt, A., Gratch, J., & Weiss, L. (2009). At the virtual frontier: Introducing Gunslinger, a multi-character, mixed-reality, story-driven experience. Intelligent Virtual Agents. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-04380-2_62

Johnson, W. L., Beal, C., Fowles-Winkler, A., Lauper, U., Marsella, S., Narayanan, S., … Vilhjálmsson, H. (2004). Tactical Language Training System: An Interim Report. In 7th International Conference on Intelligent Tutoring Systems (pp. 336–345).

Kim, J., Jr, R. H., & Durlach, P. (2009). BiLAT: A Game-Based Environment for Practicing Negotiation in a Cultural Context. International Journal of Artificial Intelligence in Education., 19(3), 289–308. Retrieved from http://www.iaied.org/pub/1296/file/19_3_03_Kim_.pdf

Kistler, F., André, E., & Mascarenhas, S. (2013). Traveller: An Interactive Cultural Training System Controlled by User-Defined Body Gestures. Human-Computer …. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-40498-6_63

Knapp, M. L., & Hall, J. A. (2010). *Nonverbal communication in human interaction* (p. 496). Cengage Learning. Retrieved from http://books.google.com/books?hl=en&lr=&id=j5HIIfRUPm0C&pgis=1

Linden Lab. (2013). Versu. Retrieved from http://lindenlab.com/products/versu

Marsella, S., Pynadath, D., & Read, S. (2004). PsychSim: Agent-based modeling of social interactions and influence. *Proceedings the International Conference on Cognitive Modeling*. Pittsburg.

Mateas, M., & Stern, A. (2002). A behavior language for story-based believable agents. *IEEE Intelligent Systems*, *17*(4), 39–47. doi:10.1109/MIS.2002.1024751

Mateas, M., & Stern, A. (2005). Structuring Content in the Façade Interactive Drama Architecture. *Artificial Intelligence and Interactive Digital Entertainment (AIIDE 2005)* (Vol. 3). Marina del Rey, CA.

Mccoy, J., Treanor, M., Samuel, B., Reed, A. A., Mateas, M., & Wardrip-fruin, N. (2013). Prom Week : Designing past the game / story dilemma. *Proceedings of Foundations of Digital Games (FDG 2013)*.

McCoy, J., Treanor, M., Samuel, B., Tearse, B., Mateas, M., & Wardrip-fruin, N. (2010). Comme il Faut 2 : A fully realized model for socially-oriented gameplay. *Proceedings of Foundations of Digital Games (FDG 2010) Intelligent Narrative Technologies III Workshop (INT3)*. Monterey, California.

Osborne, F. (2011). A New Approach to Social Behavior Simulation: The Mask Model. In M. Si, D. Thue, E. Andre, J. Lester, J. Tanenbaum, & V. Zammitto (Eds.), *Interactive Storytelling 4th International Conference on International Digital Storytelling* (pp. 97–108). Springer Berlin Heidelberg.

Vilhjalmsson, H., & Marsella, S. (2005). Social performance framework. *Proceedings of the Workshop on Modular Construction of Human-Like Intelligence at the 20th National AAAI Conference on Artificial Intelligence*. Pittsburgh, PA.