An Event-Based Architecture to Manage Virtual Human Non-Verbal Communication in 3D Chatting Environment

Stéphane Gobron[1,2], Junghyun Ahn[2], David Garcia[3], Quentin Silvestre[2], Daniel Thalmann[2,4], and Ronan Boulic[2]

[1] Information and Communication Systems Institute (ISIC), HE-Arc, St-Imier, Switzerland
[2] Immersive Interaction Group (IIG), EPFL, Lausanne, Switzerland
[3] Chair of Systems Design (CSD), ETHZ, Zurich, Switzerland
[4] Institute for Media Innovation (IMI), NTU, Singapore

Abstract. Non-verbal communication (NVC) makes up about two-thirds of all communication between two people or between one speaker and a group of listeners. However, this fundamental aspect of communicating is mostly omitted in 3D social forums or virtual world oriented games. This paper proposes an answer by presenting a multi-user 3D-chatting system enriched with NVC relative to motion. This event-based architecture tries to recreate a context by extracting emotional cues from dialogs and derives virtual human potential body expressions from that event triggered context model. We structure the paper by expounding the system architecture enabling the modeling NVC in a multi-user 3D-chatting environment. There, we present the transition from dialog-based emotional cues to body language, and the management of NVC events in the context of a virtual reality client-server system. Finally, we illustrate the results with graphical scenes and a statistical analysis representing the increase of events due to NVC.

Keywords: Affective architecture, Social agents, Virtual reality, Non-verbal communication, 3D-chatting, Avatars.

1 Introduction

Non-verbal communication (NVC) is a wordless process of communication that mainly consists of the following animations: gaze, facial expressions, head and body orientation, and arm and hand movements. One exception to animation is changes of voice tone that are not considered in this paper as we focus on the exchange of text messages. In particular, facial expression plays an important role in the process of empathy [18], and emotional contagion [10], *i.e.* unconscious sharing of the emotions of conversation members. This conscious or unconscious way of communicating influences emotional state of all characters involved in a conversation [2]. NVC is triggered by emotional states, social customs, and personal attributes. In the context of 3D-chatting they should strongly influence character animation, making the conversation alive, the scenarios more consistent, and the virtual world simulation more *ecologically valid*.

Entertainment and industrial applications involving 3D social communication –for instance *Second LIFE* [15,5]– start looking for solutions to simulate this key aspect of

F.J. Perales, R.B. Fisher, and T.B. Moeslund (Eds.): AMDO 2012, LNCS 7378, pp. 58–68, 2012.

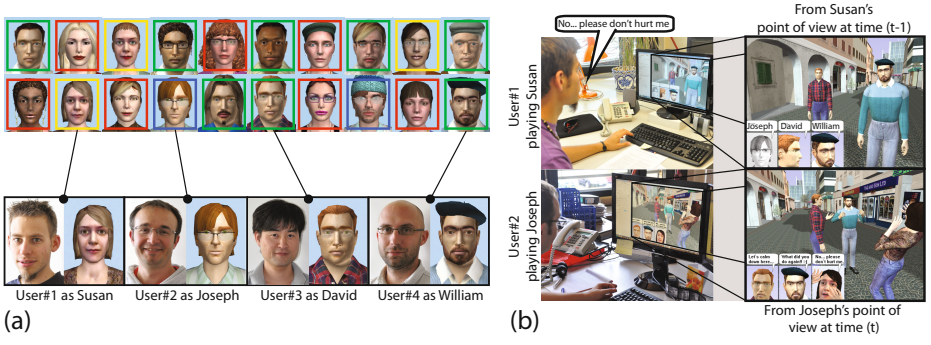

Fig. 1. (a) user's choice of avatar; (b) 3D-chatting scenario involving animated NVC

communication. The main issue is that trying to simulate NVC involves understanding "emotion", which is not an easy task as this concept is shown to be difficult to define even by specialists [8,12]. Another issue arises in a context of virtual world, it is not possible for users to control all the attributes of their avatars [9]. In the real world face to face communication, a large part of the information transmitted is done in a unconscious manner, through cues such as facial expressions or voice intonation. For this reason, users of a virtual world cannot consciously control those communication attributes, and then simulation techniques has to provide a way to fill this gap in virtual communication. In this paper we propose the event-based architecture of a working system –the emotional dynamic model being presented in a companion paper. This system proposes a "walkable 3D world environment enriched with facial and full body animation of every user's avatar consistent with the potential emotion extracted from the exchanged dialog. We believe that this approach simulates the most important NVC attributes, *i.e.*: (a) Virtual human (VH) emotional mind including a dimensional representation of emotions in three axes –*i.e.* valence, arousal, and dominance– and facial animation and "emoMotions" that predefined full-body VH animations corresponding to emotional attitudes –*e.g.* fear (Susan, user 1), anger (William, user 4), and empathy (David, user 3) illustrated Figure 1(b); (b) A virtual reality client-server architecture to manage in real-time events and induced NVC events mainly produced by the emotional dynamic model –see Figure 6; (c) Avatar's internal emotion dynamics by enabling short term emotions, long-term emotions, and emotional memory towards encountered VHs; (d) Automatic gazing, speech target redirection, and breathing rhythm according to arousal level.

2 Related Works

Previous research has explored NVC in different ways. In psychology, non-verbal behavior and communication [20] are widely studied. Virtual reality researches oriented towards psychology give motivations to simulate natural phenomenon of human conversation. NVC contains two different elements: human posture and relative positioning, and they have been analyzed to simulate interpersonal relationship between two virtual

humans [2]. The evaluation of conversational agent's non-verbal behavior has also been conducted in [14].Communication over the internet by various social platforms was also explored. They specified what they learned regarding how people communicate face-to-face in a cyberworld [11].

A number of researches about 3D chatting or agent conversational system have been presented so far. A behavior expression animation toolkit entitled "BEAT" that allows animators to input typed text to be spoken by an animated human figure was also propose by Cassell in [6]. A few years later, emotional dynamics for conversational agent has been presented in [3]. Similarly to our approach, their architecture of an agent called "Max" used an advanced representation of emotions. Instead of a restricted set of emotions a dimensional representation with three dimension *i.e.* v,a,d for *valence-arousal-dominance*. Later, an improved version of agent "Max" has been also presented as a museum guide [13] and as a gaming opponent [4]. A model of behavior expressivity using a set of six parameters that act as modulation of behavior animation has been developed [17] as well. Very recently [16] proposed to study constraint-based approach to the generation of multimodal emotional displays.

For what concerns the cooperation between agents and humans, and in [7] people appreciate to cooperation with a machine when the agent expresses gratitude by means of artificial facial expression were found. For this reason, adding emotional NVC to virtual realities would not only enhance user experience, but also foster collaboration and participation in online communities. An interdisciplinary research was proposed late 2011 in [9] that merges data-mining, artificial intelligence, psychology and virtual reality. Gobron *et al.* demonstrated an architecture of 3D chatting system available only for one to one conversation and their approach did not allow free virtual world navigation.

3 NVC Architecture

Compared to the literature presented in the previous section, our NVC real 3D-chatting approach is original in terms of aims (i.e. NVC enriched 3D-chatting), structure (i.e. building context with events), time related management of events. The process pipeline is especially novel as it enables multiple users chatting with NVC represented on their respective avatars –see Sections 3.1 and 3.4. Different types of events (potential client events, certified server events, secondary server events) play a key role allowing a consistent NVC to be simulated –see Section 3.3 and Figures 2 and 3). As induced NVC events cannot simply be sorted into a FIFO pile, we propse an event management allowing time shifting and forecasting for all types of event –see Section 3.2. The heart of the emotional model, described in details via the formalization of the short term emotional dynamics –this model is proposed in a companion paper. This section details a virtual human [VH] conversation architecture that uses semantic and emotional communication, especially suited for entertainment applications involving a virtual world. Similarly to [9] and [3], our emotional model uses the dimensional representation v,a,d for valence, arousal and dominance that allow any emotion to be represented. The basic idea behind our architecture is that dialogs trigger emotions, emotions and user interruptions trigger events, and events trigger NVC visual output. During the software design of this work, we realized that the key-links between interacting avatars and their

potential emotions were events. Indeed, depending of context, different types of events could, should, or would happen generating waves of emotion, changing avatars' attitudes, and influencing back perceptions and therefore dialogs.

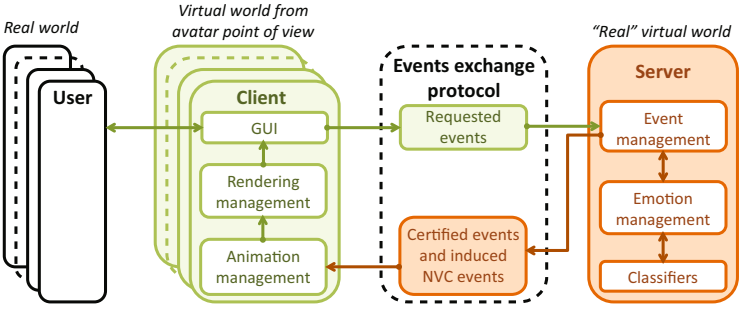

Fig. 2. Global interaction between users, associated clients, and the unique server

This is why, as shown in Figures 2 and 3, we propose an event-based architecture system to manage NVC for 3D-chatting. A relatively large number of aspects have to be presented to cover such virtual reality simulation. In the following subsections, we present how user commands influence NVC graphics in a context of client-server architecture; next, we describe the management of events; and finally, we propose relationships between client and server.

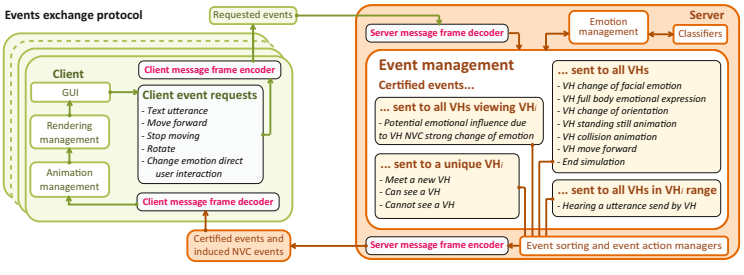

Fig. 3. Events exchange protocol between clients and server: events requested by clients (representing VHs) produce direct and indirect events of different nature such as induced-events

3.1 Building a Context

The Figure 4 details the 3D-chatting simulation with NVC. As shown, users interact in the virtual world through their individual client environment with a graphical user interface (GUI). User's input and commands to the 3D chatting system are collected via the GUI using keyboard and mouse on two windows. The first one presenting the virtual world from the point of view of the avatar (1st person view). The second window being the representation of emotional states of the avatar including also interruption

command buttons for immediate change of emotion. The user can input text for verbal communication, he can move forward, rotate while standing still, and, at will, adjust the avatar emotion manually. The server puts the received events in a history list according to their arrival time. When the server executes an event, it sends the corresponding message to the concerned clients. As a VH travels in the virtual world, it meets unknown VHs (*encounters*) and/or looses sight (see Figure 7(a)) some others, which affect the avatar emotional memories and the GUI in such way:

- Meeting a new VH implies: the creation of an emotional object in the avatar memory; the rendering of a new window–at the bottom left of the main GUI window; and the change of gazing at least for a very short period;
- (b) if a VH is not anymore in the view range of an avatar, it is kept in memory but not animated and displayed in black and white;
- (c) the facial windows are sorted from most recently seen in the left and their sizes are inversely proportional to the number of people met.

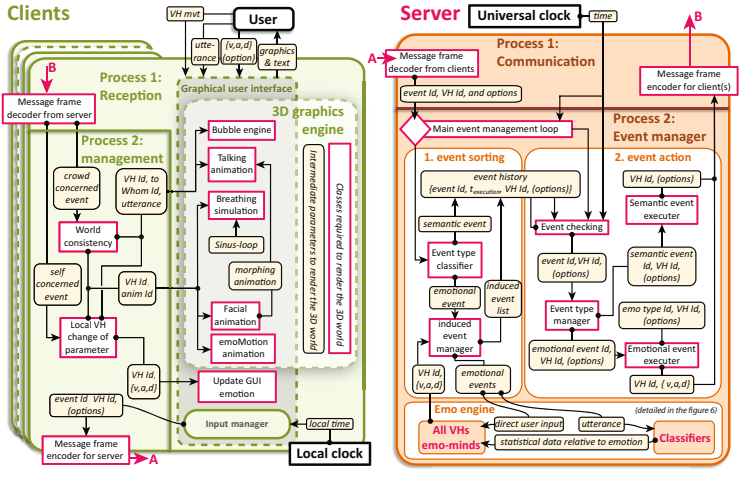

Fig. 4. This data flow chart details the system architecture where parallel processes on both server and clients are necessary to manage NVC events

Notification of any VH NVC have to be sent to every client but the users are aware of such change only if their coordinates and orientation allow it. Similarly, whenever a VH is animated for any reason, walking, re-orientating or communicating non-verbally, all clients are also notified in order to animate the corresponding VH if it is in the client field of view. Furthermore, the emotion influence is also sent to all the VHs that see the concerned VH which can lead to multiple emotion exchanges without any exchange of word. However, as messages are instantaneous, when a VH communicates a text utterance, only the ones that can "hear", *e.g.* the ones that are in the field of view, will receive the message.

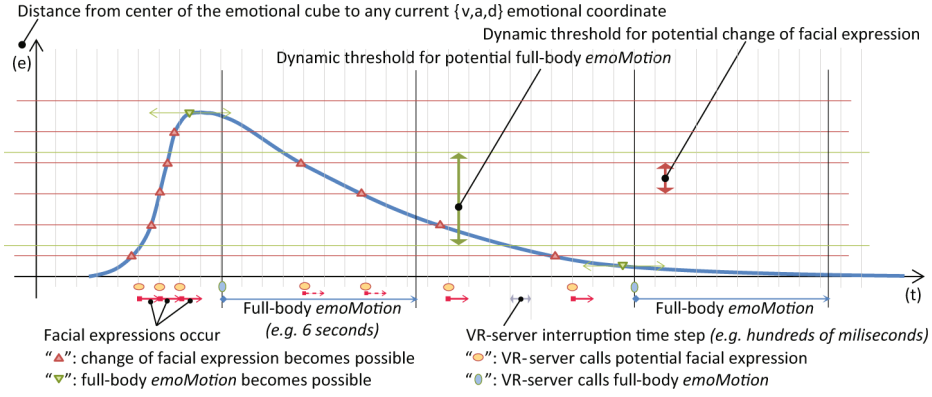

Fig. 5. Example of events chain reaction that an emotional peek (blue wave) that the VR-server produce which involve new facial expressions and full-body *emoMotions*

3.2 From Dialog to NVC Motions

Collected data is transmitted from the client to the server which processes it using the event manager. In a context of virtual reality client-server, one major difficulty is probably to animate VH corresponding to their potential emotions without sending too many events between VR-server and VR-clients. For that purpose, we defined different thresholds (see Figure 5) for synchronizing and differentiating simultaneous changes of facial expressions and full-body movements (mainly arms movements, torso posture, and head orientation). Another way to relax the streaming and to avoid computational explosion at the server level, is to make the assumption that minor changes of emotion can be simulated only by facial expression (low cost NVC–in term of interruption delay), and full-body emotional motions (high cost NVC) occur only as major change of emotional events. The figure 5 illustrates how an emotional peak (computed by the short term emotional model) is interrupted in term of indirect events at server's level. These emotional events will produce other interruptions at client's level for VH facial and full-body emotional animations. In that figure, the blue line represents a sudden change of emotion, triangles indicate that the VR-server identifies a potential graphical action relative to emotion (*i.e.* an event is created and stored inside a dynamic chain of events depending of its time priority), and ellipses represent actual emotional animation orders from server to clients.

3.3 VR Event Manager

In the context of simulation involving virtual worlds, avoiding a break in presence is an absolute priority. Considering that NVC related events occurring before triggered verbal events is a serious threat to the simulation consistency. Therefore, we have designed an event-driven architecture at the server level. The *event manager* is then probably the most important part of the system as it guarantees: first, the correctness of time sequences; second, the coherence production of non-verbal induced-events; and third, the information transmission to clients (*e.g.* dialog, movements, collisions, self emotional states, others visible emotional change of emotion).

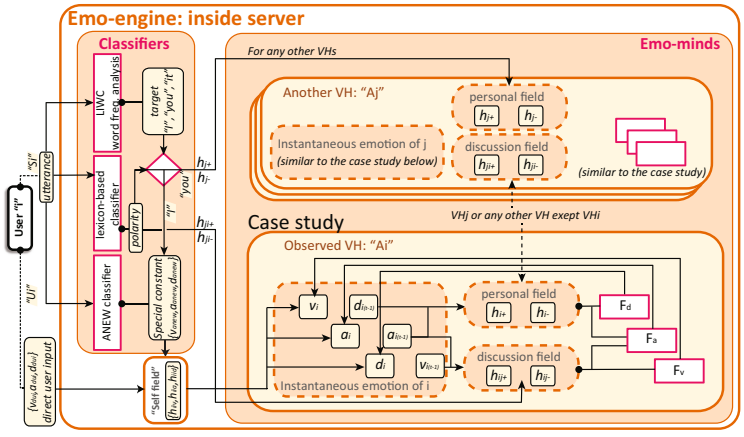

Fig. 6. Data flow chart presenting the core of the emotional model for NVC: the short term emotions model based from [19]; details of this model are proposed in a companion paper [1]

In this system, an event is anything that can happen in the virtual world. It can be a VH movement as well as a semantic transmission (*i.e.* text utterance) or a user emotional interruption command. Events can be sorted into two categories, events requested by the user input at GUI level and events generated by the server. The first category represents a request from a VR-client, not what will happen for sure. For instance, the "move forward" user command is not possible if a wall is obstructing the avatar's path. The second category represents factual consequences produced by the server that occur in the virtual world. Notice that from a single client-event, multiple induced-events can occur. For instance, a user inputs a text utterance: specific personal emotional cues can be identify that will imply changes of facial and body expression. Then, depending on the avatar world coordinate and orientation, multiple VHs can receive the change of emotion (at $t + \epsilon$, *i.e.* as fast as server can handle an event) and the text utterance (at $t + \delta t$, *i.e.* slightly later so that emotions are presented before possible semantic interpretation, *e.g.* 350 milliseconds). Naturally, the body behavior of an avatar has also consequences on the emotional interpretation of any VH that can see it –*i.e.* purpose of NVC– which will produce other induced-events.

Multiple threads are needed to simulate the above mentioned effects. The event manager stores induced events in a dynamic history list depending on their time occurrence (present or future). Simultaneously, the event manager also un-stacks all events that are stored in the past. In both cases, the process consists of defining from what could happen and what should happen in the virtual world. This management of events must then be centralized which is why the VR-server level represent the *reality* and the VR-client levels its *potential projections* only.

3.4 VR Client-Server Tasks

As seen in previous paragraphs, the server defines the reality of the virtual world as it is the only element that dynamically forecasts events. It runs two threads: one for

the event management, connecting to database and emotional engines and the other to execute events, communicating to specific clients corresponding information that runs animation and rendering engines (lips, facial expression, motion capture emotional full-body sequences, interactive text bubbles, etc.). VR-clients also run two parallel processes: one for the communication with the server and the other for the GUI. The data flow chart Figure 4 left depicts main processes of the VR-client which basically is: (a) communicating user requests and context to server, and (b) execute server semantical and induced events valid from the point of view of the local avatar. Every VR-client receives all information relative to physical world and only part of events relative to utterance. One part of these data is stored in the crowd data structures, the other part in the local VH data structure, but both are needed to animate and render the scene from the local VH point of view.

4 Results

Figure 1(b) and Figure 7 present experimental setups with four users in the same conversation. Two aspects of the results are illustrated: emotionally triggered animations, and a statistical analysis representing increase of events due to NVC.

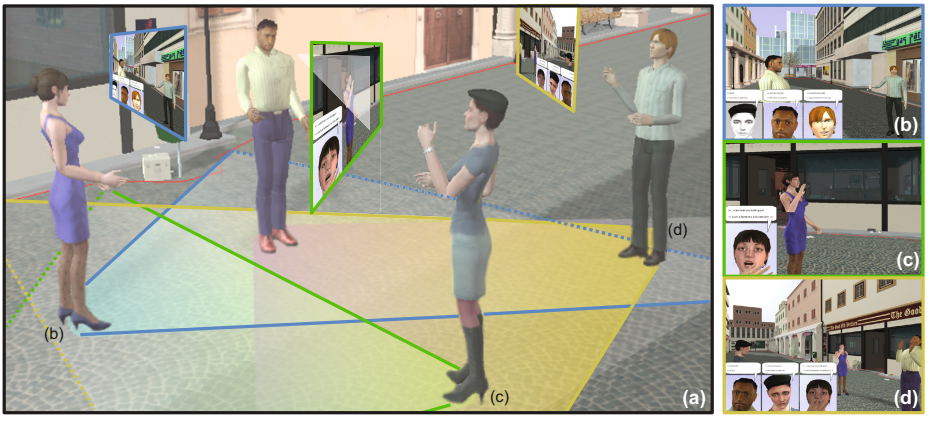

Fig. 7. User-test: (a) visualization of communication areas have been artificially added (Illustrator) in blue for view *Linda*'s (b), green for *Patricia*'s view (c), and yellow for *Joseph*'s view (d), to rendered scenes of the four participants

Encountered VHs and Memory – In the participant views of Figure 7 (three pictures on the right) encountered avatars faces are shown with a different rendering: sometimes colored and animated, sometimes black and white and freezed. For instance, in (b) three other VHs have been encountered but only two are currently within the view area, therefore *Patricia* is depicted in black and white. **Relationship between Time, Event, and NVC –** Figures 1(b) and 7 illustrate the effect of NVC with respect to time with corresponding changes of emotional behavior. In the first figure two user points-of-view are shown at initial time $(t_{-1}$ and emotional peek time (t). In the second result figure

time remains identical but we can see the global view and the different communication range of each involved avatar. **Architecture Testing –** We have tested the computational effect of adding non-verbal events in a 3D-chatting virtual world environment. In terms of computational capability, the entire architecture is running at 60 fps on nowadays classical PCs. We produced 11 tests to compute the additional expected computational cost due to the enriched non-verbal aspect. As illustrated in Figure 8, the increase on the VR-server in input is less than 10 percent. Generation of induced-events (*e.g.* emotional changes), increases output around 70 percent. Two aspect can be concluded: first, the total increase remains small compared to the computer capabilities, and second, the increase factor is not dependant of the number of participants but relative to the user number chatting within each group –which usually is two persons and rarely larger than four. Video demonstrations are also available online for "changes of facial expression" and "test of the general architecture".

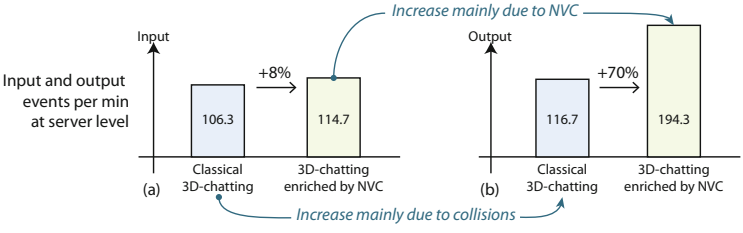

Fig. 8. Comparing input and output events occurring at server level between a classical 3D-chatting and a 3D-chatting enriched by NVC

5 Conclusion

We have presented a 3D virtual environment conversational architecture enriched with non-verbal communication affecting virtual humans movements. This approach adds a new dimension to animation applicable to virtual worlds as resulting conversations enable to create more natural exchanges. Whereas a companion paper details the emotional model and correlation to emotional animations, this paper focuses on the global event-based architecture and corresponding effect on motion-based NVC. One consequence to this event-based management of emotion is the different aspect of motion that makes VHs to look more natural: change of facial expression, breathing rhythm due to stress level, and full body emotional motions occurring at intense change of emotion. We have also shown that the increase of messages between server and clients due to NVC is not a threat when finding a compromise between visible change of expression and time to react to NVC events. The next phase of this study is a large scale user-test focusing on how users would react in this new way to experience virtual world. This study would also help defining a good parametrization for the management of events.

Acknowledgements. The authors wish to thank O. Renault and M. Clavien for their hard work and collaboration in the acting, motion capture, and VH skeleton mapping of *emoMotion*. We thanks J. Llobera, P. Salamin, M. Hopmann, and M. Guitierrez for

participating in the multi-user testings. This work was supported by a European Union grant by the 7th Framework Programme, part of the CYBER**EMOTIONS** Project (Contract 231323).

References

1. Ahn, J., Gobron, S., Garcia, D., Silvestre, Q., Thalmann, D., Boulic, R.: An NVC Emotional Model for Conversational Virtual Humans in a 3D Chatting Environment. In: Perales, F.J., Fisher, R.B., Moeslund, T.B. (eds.) AMDO 2012. LNCS, vol. 7378, pp. 47–57. Springer, Heidelberg (2012)
2. Becheiraz, P., Thalmann, D.: A model of nonverbal communication and interpersonal relationship between virtual actors. In: Proceedings of Computer Animation 1996, pp. 58–67 (June 1996)
3. Becker, C., Kopp, S., Wachsmuth, I.: Simulating the Emotion Dynamics of a Multimodal Conversational Agent. In: André, E., Dybkjær, L., Minker, W., Heisterkamp, P. (eds.) ADS 2004. LNCS (LNAI), vol. 3068, pp. 154–165. Springer, Heidelberg (2004)
4. Becker, C., Nakasone, A., Prendinger, H., Ishizuka, M., Wachsmuth, I.: Physiologically interactive gaming with the 3d agent max. In: Intl. Workshop on Conversational Informatics, pp. 37–42 (2005)
5. Boellstorff, T.: Coming of Age in Second Life: An Anthropologist Explores the Virtually Human. Princeton University Press (2008)
6. Cassell, J., Vilhjálmsson, H.H., Bickmore, T.: Beat: Behavior expression animation toolkit. In: SIGGRAPH 2001, pp. 477–486 (2001)
7. de Melo, C.M., Zheng, L., Gratch, J.: Expression of Moral Emotions in Cooperating Agents. In: Ruttkay, Z., Kipp, M., Nijholt, A., Vilhjálmsson, H.H. (eds.) IVA 2009. LNCS, vol. 5773, pp. 301–307. Springer, Heidelberg (2009)
8. Ekman, P.: Emotions revealed. Henry Holt and Company, LLC, New York (2004)
9. Gobron, S., Ahn, J., Silvestre, Q., Thalmann, D., Rank, S., Skoron, M., Paltoglou, G., Thelwall, M.: An interdisciplinary vr-architecture for 3d chatting with non-verbal communication. In: EG VE 2011: Proceedings of the Joint Virtual Reality Conference of EuroVR. ACM (September 2011)
10. Hatfield, E., Cacioppo, J.T., Rapson, R.L.: Emotional Contagion. Current Directions in Psychological Science 2(3), 96–99 (1993)
11. Kappas, A., Krämer, N.: Face-to-face communication over the Internet. Cambridge Univ. Press (2011)
12. Kappas, A.: Smile when you read this, whether you like it or not: Conceptual challenges to affect detection. IEEE Transactions on Affective Computing 1(1), 38–41 (2010)
13. Kopp, S., Gesellensetter, L., Krämer, N.C., Wachsmuth, I.: A Conversational Agent as Museum Guide – Design and Evaluation of a Real-World Application. In: Panayiotopoulos, T., Gratch, J., Aylett, R.S., Ballin, D., Olivier, P., Rist, T. (eds.) IVA 2005. LNCS (LNAI), vol. 3661, pp. 329–343. Springer, Heidelberg (2005)
14. Krämer, N.C., Simons, N., Kopp, S.: The Effects of an Embodied Conversational Agent's Nonverbal Behavior on User's Evaluation and Behavioral Mimicry. In: Pelachaud, C., Martin, J.-C., André, E., Chollet, G., Karpouzis, K., Pelé, D. (eds.) IVA 2007. LNCS (LNAI), vol. 4722, pp. 238–251. Springer, Heidelberg (2007)
15. Michael, R., Wagner, J.A.: Second Life: The Official Guide, 2nd edn. Wiley Publishing (2008)
16. Niewiadomski, R., Hyniewska, S.J., Pelachaud, C.: Constraint-based model for synthesis of multimodal sequential expressions of emotions. IEEE Transactions on Affective Computing 2(3), 134–146 (2011)

17. Pelachaud, C.: Studies on gesture expressivity for a virtual agent. Speech Commun. 51(7), 630–639 (2009)
18. Prestonand Frans, S.D., de Waal, B.M.: Empathy: Its ultimate and proximate bases. The Behavioral and Brain Sciences 25(1), 1–20 (2002)
19. Schweitzer, F., Garcia, D.: Frank Schweitzer and David Garcia. An agent-based model of collective emotions in online communities. The European Physical Journal B - Condensed Matter and Complex Systems 77, 533–545 (2010)
20. Weiner, M., Devoe, S., Rubinow, S., Geller, J.: Nonverbal behavior and nonverbal communication. Psychological Review 79, 185–214 (1972)