Impact Study of Nonverbal Facial Cues on Spontaneous Chatting with Virtual Humans

Stephane Gobron[1,2], Junghyun Ahn[2] , Daniel Thalmann[3], Marcin Skowron[4], Arvid Kappas[5]

[1] ISIC, HE-Arc, HES-SO, Saint-Imier, Neuchatel, Switzerland
[2] Immersive Interaction Group, EPFL, Lausanne, Switzerland
[3] Division of Computer Science, School of Computer Engineering, NTU, Singapore
[4] Austrian Research Institute for Artificial Intelligence, Vienna, Austria
[5] School of Humanities and Social Sciences, Jacobs University Bremen, Bremen, Germany

Submitted: September 23rd, 2011; reviewed: February 15th, 2012; accepted 5 June 5th, 2013

Abstract

Non-verbal communication (NVC) is considered to represent more than 90 percent of everyday communication. In virtual world, this important aspect of interaction between virtual humans (VH) is strongly neglected. This paper presents a user-test study to demonstrate the impact of automatically generated graphics-based NVC expression on the dialog quality: first, we wanted to compare impassive and emotion facial expression simulation for impact on the chatting. Second, we wanted to see whether people like chatting within a 3D graphical environment. Our model only proposes facial expressions and head movements induced from spontaneous chatting between VHs. Only subtle facial expressions are being used as nonverbal cues –*i.e.* related to the emotional model. Motion capture animations related to hand gestures, such as cleaning glasses, were randomly used to make the virtual human lively. After briefly introducing the technical architecture of the 3D-chatting system, we focus on two aspects of chatting through VHs. First, what is the influence of facial expressions that are induced from text dialog? For this purpose, we exploited an emotion engine extracting an emotional content from a text and depicting it into a virtual character developed previously ([GAS+11]). Second, as our goal was not addressing automatic generation of text, we compared the impact of nonverbal cues in conversation with a chatbot or with a human operator with a *wizard of oz* approach. Among main results, the within group study –involving 40 subjects– suggests that subtle facial expressions impact significantly not only on the quality of experience but also on dialog understanding.

Keywords: Nonverbal communication; Virtual reality; Artificial facial expression impact; Virtual human communication; Avatar; Agent; Wizard of Oz; 3D-Chatting

1 Introduction

Realistic interactions between virtual humans (VH) are crucial for applications involving virtual worlds – especially for entertainment and social network application: inter-character nonverbal communications are a key type of interaction. To address this issue we proposed to study three cases of interactions as illustrated by Figure 1: (a) VH (avatar) "A" chatting with VH (avatar) "B"; (b) "A" chatting with an agent (VH derived by a computer) "C"; (c) "A" believing to chat with an agent but in reality chatting with another VH (avatar) "D", *i.e.* a Wizard of Oz with the acronym (*Woz*).

However, what is the contribution of nonverbal cues to verbal interaction? Despite the communicative importance of such cues is acknowledged in psychology and linguistics [ASN+70] [WDRG72] [GTG+84],

urn:nbn:de:0009-6-38236, ISSN 1860-2037

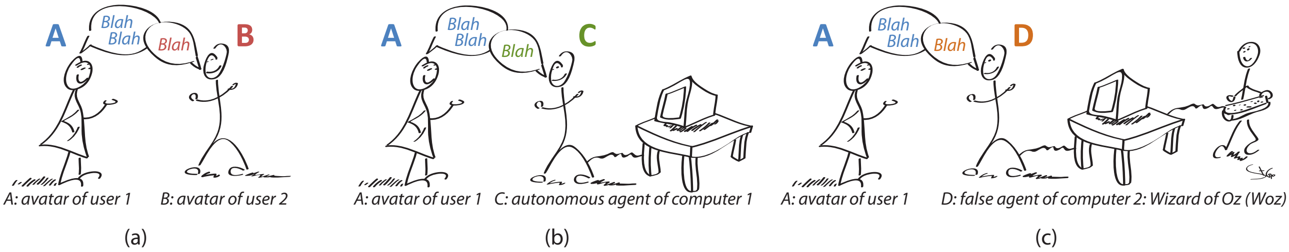

Figure 1: A comic representation of semantic and nonverbal communication in virtual environment, within: (a) two avatars; (b) an avatar and an agent; (c) an avatar and a Wizard of Oz.

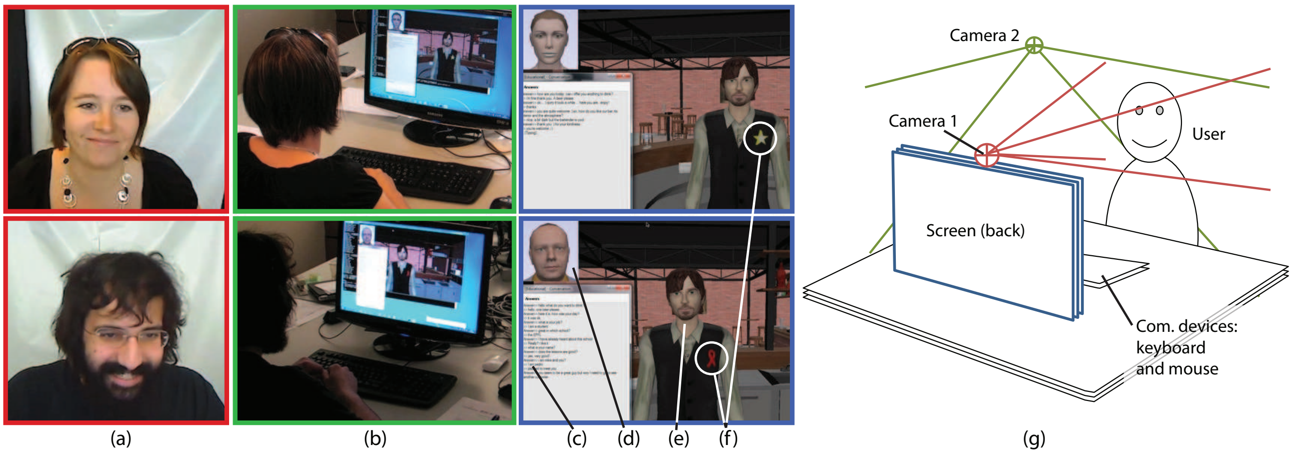

Figure 2: Examples of users performing user-test with different bartender conditions. (a) Users enjoying our chatting system (user face camera 1); (b) Second view at the same moment showing the user interface (global view camera 2); (c) to (f) Close up of the user interface: (c) Message window, where user type input and receive response from agent or *Woz*; (d) Facial expression of user's avatar generated by our graphics engine; (e) Bartender who could be agent or *Woz*; (f) A tag showing different condition of the bartender; (g) Setup during user-test where two cameras were recording the chatting.

there is scarce evidence obtained in laboratory controlled conditions. Experiments studying the contribution of nonverbal cues to natural communication are challenging because non verbal cues occur simultaneously with strictly verbal information like words and paraverbal cues such as voice intonation, pauses, etc. VHs allow for the depiction of expressions and offer an opportunity to study such factors in a controlled laboratory setup. In this paper, we report a study on the impact of nonverbal cues in a 3D virtual environment (VE).

In the virtual reality (VR) domain, communication with emotional computer-driven VHs (called *agents*) is also becoming an important research topic. Indeed, regarding information exchanges during a conversation, semantic or verbal information is only *the tip of the iceberg*. Therefore, to tackle the complex issue of nonverbal communication, we propose to study the impact of automatically generated graphics-based NVC expression on the dialog quality induced from a real-time interactive affective computing system. To avoid the bias of voice intonation this system only uses textual chatting interaction. From this model, we report the impact of nonverbal cues to spontaneous verbal interaction using virtual humans in a laboratory controlled experiment. We show quantitatively that nonverbal cues do indeed contribute to nonverbal communication and demonstrate this technique can be used to control precisely the nonverbal cues associated with the verbal interaction.

The paper is structured as follows: Section 2 describes background information about VE and VR, psychological models, and conversational systems with some of their limitations. Section 3 presents the user-test in terms of material, methods, experimental setup, and the architecture as a recent paper [GAS+11] is dedicated to that aspect. Section 4 details the results and the statistical analysis. Finally, we conclude the paper in Section 5 with a relatively large list of improvement derived from the user-test questions and users comments and suggestions.

2 Background

2.1 Virtual Environment (VE)

Emotional communication in virtual worlds has been a challenging research field over the last couple of decades. A work from [WSS94] has conducted a research concerning the influence of facial expressions for a communication interface. In contrast, our experimental environment offers a wider graphics scene including generic body movement. Cassell *et al.* [CPB+94] proposed a system which automatically generates and animates conversations between multiple human-like agents with appropriate and synchronized speech, intonation, facial expressions, and hand gestures; the whole complex pipeline of VH conversation was unfortunately not presented. Also relative to gesture and expression, [BK09] introduced a research on the relationship between deictic gesture and facial expression. Perlin and Goldberg [PG96] proposed an authoring tool (*Improv*) to create actors that respond to users and to each other in real-time, with personalities and moods consistent with the authors' goals and intentions. Numerous ways to design personality and emotion models for virtual humans were described in “Embodied Conversational Agents” [Cas00]. Coyne and Sproat [CS01] described linguistic analysis and depiction techniques needed to generate language-based 3D scenes, without extraction of emotional parameters and VH mind model. Badler *et al.* [BAZB02] proposed a Parameterized Action Representation (PAR) designed for building future behaviors into autonomous agents and controlling the animation parameters that portray personality, mood, and affect in an embodied agent with a relatively simple emotional architecture. Cassell *et al.* [CVB01] proposed a behavior expression animation toolkit (BEAT) that allows animators to input typed text to be spoken by an animated human face without focusing on visualization of emotional parameters extracted from chat sentence analysis.

Many other approaches exist; here is the non-exhaustive list of most important ones which however, have not the complexity level of our architecture and emotional model. Loyall *et al.* [LRBW04] focused on creating a system to allow rich authoring and to provide automatic support for expressive execution of the content. Su *et al.* [SPW07] predicted specific personality and emotional states from hierarchical fuzzy rules to facilitate personality and emotion control. Park *et al.* [PJRC08] proposed a human-cognition based chat system for virtual avatars using geometric information. Concerning ”empathic” emotions, Ochs *et al.* [OPS08] determined this specific emotions of virtual dialog agents in real time and Becker *et al.* [BNP+05] use a physiological user information enabling empathic feedback through non-verbal behaviors of the humanoid agent Max. Another previous research on non-verbal communication for surgical training in a virtual environment [MWW09] has been presented in 2009. In this research a webcam captures the user's face to express the facial expressions. Unlike this previous research, the proposed framework automatically generates avatar's facial expressions from a given valence and arousal (VA) input, resulting faster simulation in a virtual scene.

Pelachaud [Pel09] developed a model of behavior expressivity using a set of six parameters that act as modulation of behavior animation. Khosmood and Walker [KW10] developed a gossip model for agent conversations with a series of speech-acts controlled by a dialogue manager.

Considerable knowledge has accumulated concerning the encoding and decoding of affective states in humans [Kap03] [RBFD03]. In fields such as VR, computer vision, computer animation, robotics, and human computer interaction, efforts to synthesize or decode facial activity have recently been successful [CK07].

Generating realistic human movement is also important for improving realism. Stone *et al.* [SDO+04] proposed an integrated framework for creating interactive, embodied talking characters. Egges *et al.* [EMMT04] described idle motion generation using principal components analysis. McDonnell *et al.* [MEDO09] investigated human sensitivity to the coordination and timing of conversational body language for virtual characters. Ennis *et al.* [EMO10] demonstrated that people are more sensitive to visual desynchronization of body motions than to mismatches between the characters' gestures and voices.

Eye and head movements help express emotion in a realistic way. Gu and Badler [GB06] developed a computational model for visual attention in multiparty conversation. Masuko and Hoshino [MH07] generated VH eye movements that were synchronized with conversation. Grillon and Thalmann [GT09] added gaze attention behaviors to crowd animations by formulating them as an inverse kinematics problem and then using a solver. Oyekoya *et al.* [OSS09] proposed a saliency model to animate the gaze of virtual characters that is appropriate for multiple domains and tasks. Bee *et al.* [BAT09] implemented an eye-gaze model of interaction to investigate whether flirting helps to improve first encounters between a human and an agent. Recently, Weissenfeld *et al.* [WLO10] proposed method for video realistic eye animation.

From the field of *affective computing* [Pic97] a number of relevant studies have been introduced. Ptaszynski *et al.* [PMD+10] described a way to evaluate emoticons from text-based online communication; Bickmore *et al.* [BFRS10] described an experiment with an agent that simulates conversational touch. Endrass *et al.* [ERA11] proposed a model to validate cultural influence and behavior planning applied to multi-agent system.

2.2 Psychological models

Because nonverbal cues have to be consistent with the verbal content expressed, to decide which nonverbal cues have to be portrayed it is necessary to use some emotional model relating verbal information to nonverbal movements. In fact, different ways to conceive emotions [Iza10] are common in psychological research and all of them have advantages and disadvantages. Three main approaches can be distinguished: (a) specific emotional states are matched with prototypical expressions (here particularly based on the work of Ekman and his colleagues [Ekm04]), (b) the relationship of appraisal outcomes and facial actions (for example based on Scherer's predictions [SE07] or the Ortony, Clore and Collins (OCC) model [OCC88]), and (c) the relationship of two or three dimensional representations of emotions in affective space (for example Russell [Rus97]) to expressions. While approach (a) has an intuitive appeal, it is by now clear that prototypical expressions are not frequent in real life [Kap03] and it is very difficult to extract distinct emotional states from short bits of text (see [GGFO08] [TBP+10]). The appraisal approach is interesting but arguably requires more research to see whether such an appraisal model for the synthesis of conversational nonverbal behavior is viable (*e.g.* [vDKP+08]). Thus, for pragmatic reasons, it is the dimensional approach that appears most relevant and promising. Emotional valence and arousal have been studied in a variety of psychological contexts [RBFD03] and there is reason to believe that the reliability of extracting these two dimensions from text is much higher than that of discrete emotional states [Rus79]. In consequence, the present model is focusing on a dimensional model. It is clear that a multi-modal approach to emotion detection would be desirable and many colleagues are working on such systems. The goal of the present implementation is the very specific case of recreating the richness of emotions in social interactions [MMOD08] from the highly reduced channel of short text messages as they are still common in mediated communication systems. The restriction of the model to two affective dimensions is a consequence of these constraints.

Furthermore, one of the novel aspects of the present model is not only the inclusion of a rudimentary personality system (see also [LA07]) in the guise of affective profiles, but also individual emotion regulation. While emotion regulation is a complex issue in itself [Kap08], the inclusion of "memory" and emotion-regulation makes the present system rather unique and addresses several layers of emotional interaction complexity. Whichever model is underlying the synthesis of nonverbal behavior from text, or context, the degree in which the actual behaviors map reality is critical. In other words, evaluation of models is key, such as Gratch and Marsella [GM05].

2.3 Conversational systems

Research on dialog management, discourse processing, semantic modeling, and pragmatic modeling provides the background for the development of dialog systems for human-computer, natural language based interactions [SH08] [HPB+09, FKK+10]. In some systems enabling users to text chat with a virtual agent such as in [CDWG08], qualitative methods are used to detect affect. The scope of known applications, however, is often focused on closed-domain conditions and restricted tasks such as plane tickets, reservations, and city guidance systems. The main fields of research in dialog management include: finite state-based and frame-based approaches [BR03], plan-based approaches [McG96], and information state-based and probabilistic approaches [WPY05]. In recent years, the development of human-agent interfaces that incorporate emotional behavior has received some interest [Pic97] [PPP07], but as not been yet applied to virtual worlds. Work in this area fo-

	Conversational system	Corresp. symbol	*Woz*	Corresp. symbol
no facial emotion	*Condition 1*	**a yellow star**	*Condition 2*	**a blue medal**
facial emotion	*Condition 3*	**a green tag**	*Condition 4*	**a red ribbon**

Table 1: The four experimental conditions with their respective symbol. Only two users noticed that the virtual bartender had different facial expressions.

cuses on embodied conversational agents [PP01], and VH [GSC+08] [RGL+06]. Even more recently, the design process of intelligent virtual human have been presented for specialized application domain such as clinical skills [RKP11].

Prominent examples of advances in this field are a framework to realize human-agent interactions while considering their affective dimension, and a study of when emotions enhance the general intelligent behavior of artificial agents resulting in more natural human-computer interactions [ABB+04]. Reithinger *et al.* [RGL+06] introduced an integrated, multi-modal expressive interaction system using a model of affective behavior, responsible for emotional reactions and presence of the created VH. Their conversational dialog engine is tailored to a specific, closed domain application, however: football-related games.

3 Material and methods

3.1 Experimental setup

Forty volunteers, 24 men and 16 women aged from 21 to 51 [20˜29: *21*; 30˜39: *13*; 40˜49: *5*; 50˜59: *1*] took part of this user-test (see users shown in front page Figure 2 and Figure 5). They were all unrelated to the field of experiment and were from different educational and social backgrounds. Only 25% were considered non-naive to the field of VR (*i.e.* students who took the VR graduate course).

3.2 From sentence to NVC

Human communication is first of all about social and psychological processes. Therefore, before presenting the general technical architecture of our communication pipeline, we introduce the main motivator of our work: what are *meaning* and *emotion*?

Architecture layers –Human communication is a multi-layered interactive system –see dash line subsets in Figure 4 for our very simplified VH communication model–involving transactions between participants, relying not only on words, but also on a number of paralinguistic features such as facial/body expressions, voice changes, and intonation. These different channels provide rapidly changing contexts in which content and nonverbal features change their meaning. Typically in the interaction process not all information is successfully transmitted and receivers also perceive cues that are not really there [KHS91]. Situational, social and cultural contexts shape further what is being encoded and decoded by the interaction partners [Kap10]. The reduced channel bandwidth available in mediated communication can lead to changes in the conversational process and is often the origin of misunderstandings. It is here that creating redundancy with nonverbal indicators of emotion, as in the present application, could recover some of the robust redundancy of regular face-to-face interaction.

Global algorithm –Affect communication with a machine in a virtual world consists of at least: a human, a user-interface to manipulate his or her associated avatar, a virtual human 3D engine, a dialogue/vocabulary analyzer, an emotional mind model, and a listener framework, also called “conversational system”, playing the role of the agent. Most of the above, except the user, can be viewed in two layers: the factual processes layer, including the agent, the event, and the graphics engines; the emotion processes layer, including the data mining, refinement, and virtual human pseudo-mind engines specialized in the management of emotions called *emoMind*.

The main steps of the model –detailed in [GAS+11] and summarized Figure 4– are as follows:

- An avatar or an agent can start a conversation, and every test utterance is a new event that enters the event engine and is stored in a queue;
- The sentence is analyzed by classifiers to extract potential emotional parameters, which are then refined to produce a multi-dimensional *probabilistic emotional histogram* (PEH) [GAP+10], that define the heart of emotion model for each VH. In our approach, as [BC09] suggested for one of the future research studies, we concentrate to ”wider displays of emotion” driven by valence and arousal $\{v,a\}$ plane, which brings various facial expressions from any given VA coordinate;
- This generic PEH is then *personalized* depending on the character (*e.g.* for an optimist this would trend towards higher valence);
- Then a new current $\{v,a\}$(emotional coordinate axes valence and arousal) state is selected among the VH mind state;
- The emotional $\{v,a\}$ values that result are transmitted to the interlocutor;

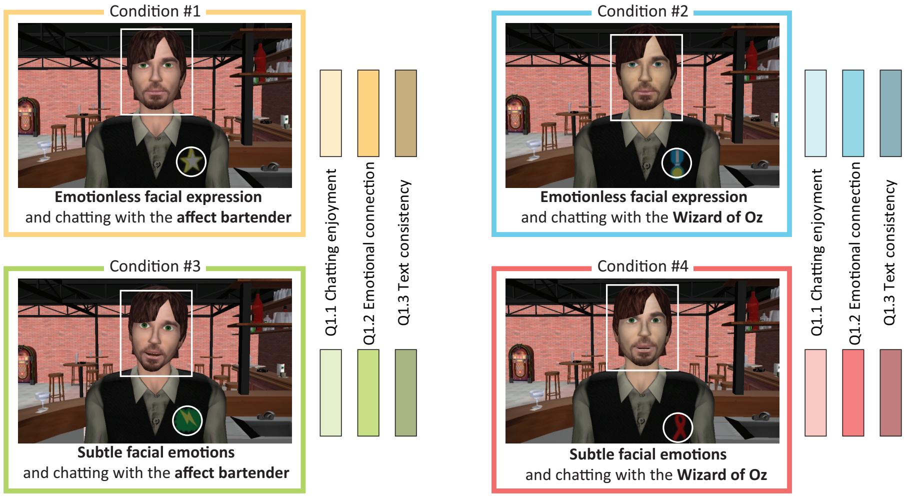

Figure 3: Four conditions of the user-test identified by specific patch textures on the bartender: a yellow star, a blue medal, a green tag, and a red ribbon. As conditions where randomly ordered, users were told to pay attention to these textures and eventually to take notes to help them answering the questionnaire.

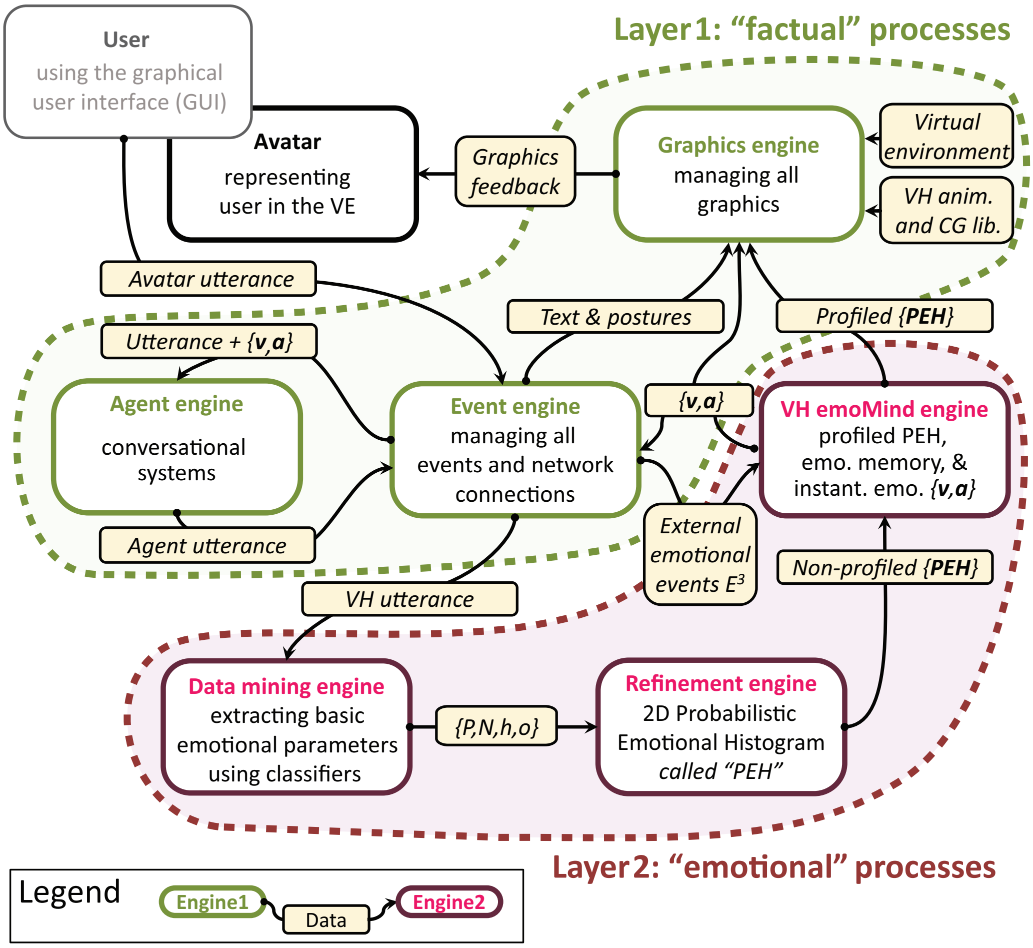

Figure 4: Summary of the general process pipeline where the direct communication layer is represented by the three green engines and the nonverbal engines in red. Arrows represent the data transfer between different engines. Details of each engine can be found in [GAS$^+$11].

- If this is a conversational system (such as the *affect bartender* used in Section 3), then it produces a text response potentially influenced by emotion;
- In parallel, and depending on the events, different animated postures are selected (*e.g.* idle, thinking, speaking motions);
- This process continues until the end of dialog.

Facial expression rendering –Concerning the graphical interpretation of emotion in the facial expression, a *Facial Action Coding System (FACS)* [EF78] was developed. This technique permits an objective description of facial movements based on an anatomical description. We derived our facial expression component from FACS Action Units (AU). Facial expressions were generated with the same algorithm for the avatar and the agent.

3.3 Procedure

All users were asked to repeat the same task four times in a 3D virtual bar, and each of four conditions was executed once but ordered randomly to avoid any sequence bias –which justifies the use of identification texture (see Figure 3). The conditions were described identically to the user as an interaction with a VH bartender. Table 1 and Figure 3 illustrate the four conditions to determine possible impacts of the chatting system and facial emotions:

- Users were sometimes chatting with an agent (*i.e.* a computer) and sometimes an avatar (*i.e.* a *Woz*, a hidden person playing the role of a computer);
- VHs facial emotions were sometimes rendered and sometimes not.

Each chat was about five minutes long and the whole experiment took about one hour per user including the initial description, the execution, and answering the questionnaire. Our user-test team consisted of three persons: the first would present tasks; the second, play the *Woz*, and the third handle the questionnaire. To avoid gaps relative to difference of dialog quality, the person playing the role of *Woz* had the instructions to use simple but natural sentences, and to pay attention not using information that was not contained in the chatting current dialog.

The goal of this experiment was twofold. First, we wanted to compare impassive and emotion facial expression simulation for impact on the chatting. Second, we wanted to see whether people like chatting within a 3D graphical environment. To access each aspect of these goals two types of questions were asked: (a) comparative and (b) general.

In type (a) questions, we were especially interested in identifying cues about the influence of subtle emotional rendering on facial expression. To do so, users were asked to remember each of the four experimentations. And since the sequence of conditions was ordered randomly, to have the minimum effect on users' perception, they were advised to identify each condition with the symbol (ribbon) shown on the bartender's (see Figure 3) shirt and to take notes for the questionnaire. Here are the three questions comparing the *enjoyment* to chat, the *emotional* connection, and the *quality* of dialog:

Questions comparing conditions (answers given on a 6-point Likert scale), *type (a)*:

- Q1.1 How did you enjoy chatting with the virtual human?
- Q1.2 Did you find a kind of "emotional connection" between you and the virtual human?
- Q1.3 Did you find the dialog with the virtual human to be realistic?

The six type (b) questions were related to the general experience users had in chatting in a 3D graphical environment: other VHs and self appreciation of facial expression; 3D graphics and chatting; contrast between a self identification (that we knew would be poor) and appreciation of repeating or continuing the experimentation.

Questions relative to the general chatting experience (five answers also given on a 6-point and one on a 12-point Likert scale), *type (b)*:

- Q2.1 Did you enjoy seeing the other VH emotions?
- Q2.2 Did you enjoy seeing your own avatar emotions?
- Q2.3 Do you think that 3D graphics enhance the chatting?
- Q2.4 Would you like to repeat a similar experiment?
- Q2.5 How well did you identify yourself with your avatar?
- Q2.6 Do you think that 20 minutes of conversation was long enough?

3.4 Data Analysis

Users were asked to answer the nine questions directly on an anonymous Microsoft Excel data sheet. From the questionnaire responses, statistical results were automatically computed using available tools (ANOVA). Two initial aspects motivated this study: first, proving whether relatively subtle facial emotional expression had impact on how the chatting dialog was perceived; and second, determining the general appreciation of users towards chatting in a 3D virtual environment.

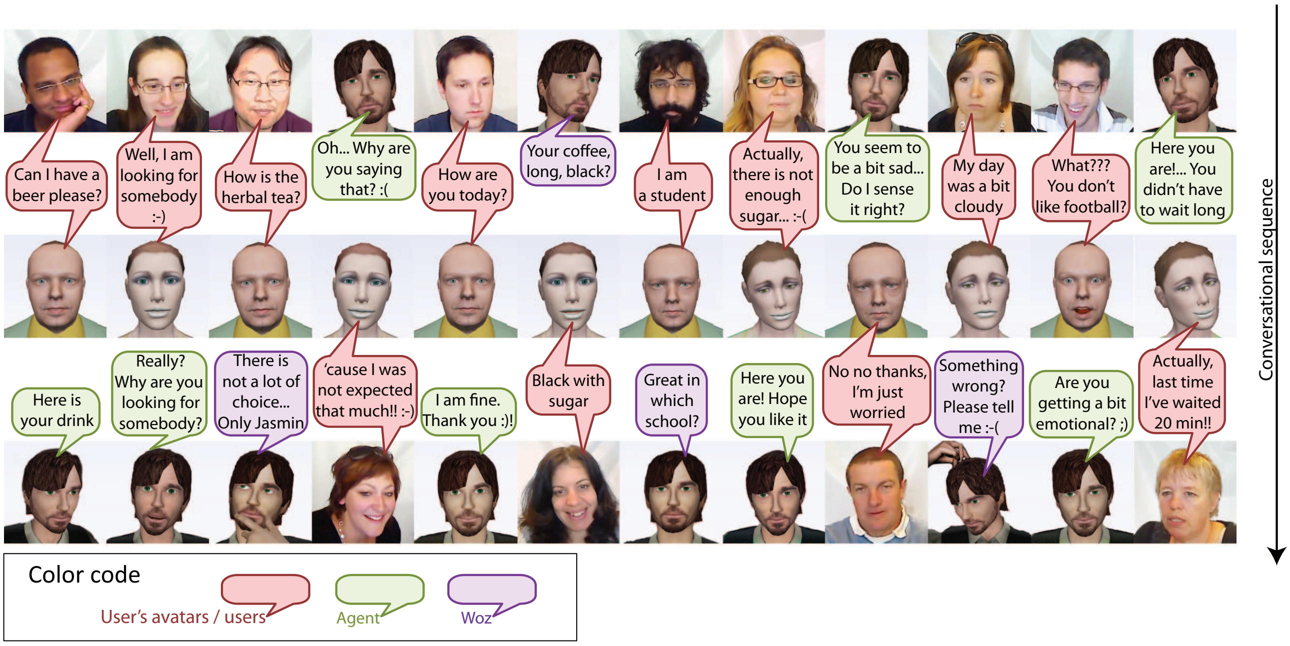

Figure 5: Sample of conversation between real users and agents through their respective avatars. Red text bubbles are sentences expressed by users and therefore by their avatar in the virtual world; the green ones are from agent and blue from *Woz*.

Results of this data analysis are presented in the following section.

4 Results

Figure 6 and Figure 9 summarize the results of type (a) and type (b) questions.

Facial emotion and nonverbal communication: Comparing the emotionless and emotion facial expression conditions (especially the one employing the *Woz*, *e.g.* row 1, column 2 and 4), even with subtle facial emotion, users preferred (average increase of 10%) chatting with emotional feedback, and graphical attributes strongly influenced their perception of dialog quality.

Conversational system vs. *Woz*: As shown, *Woz* (*i.e.* human taking the place of conversation system) chatting was more appreciated in terms of dialog and emotion, nevertheless the difference is not large (5%). Furthermore, all participants were surprised (sometimes even shocked) to learn that two out of four experiments were performed with a human rather than a machine. We even observed two users criticizing *the machine* for "having an unrealistic conversational algorithm" when they were referring to a real human.

From the first three comparative questions and the results in Figure 6, we performed the comparisons shown in Figures 7 and 8 by comparing two aspects: (a) the fact to have a present or missing NVC (depicted by facial expressions and head movements); (b) the influence between chatting with an artificial system (agent) or a human playing the role of an artificial machine (*Woz*).

The effect of facial emotion is illustrated on the first row of Figure 7. There is a clear positive influence on feelings of enjoyment, emotional connection and even text quality, by adding facial emotion when chatting with a human. In contrast, the second row demonstrates almost no effect when chatting with a computer. Second, Figure 8 shows the difference between chatting with a computer or a human: The first row indicates that chat without facial emotions is more enjoyable and with a better emotional connection if it is with a computer. This relatively surprising result can be explained by the fact that the chatting system algorithm always deviates the conversation back to the user as it cannot chat about itself (*i.e.* the conversation is not understood by the machine, only interpreted). On the second row, we can only deduce the sensation of emotional connection is increased if the dialog occurs with a human. In this case, we believe that the source of this observation is the difference of text quality.

The following observations are based on Figure 9:

3D environment: People enjoyed the 3D environ-

urn:nbn:de:0009-6-38236, ISSN 1860-2037

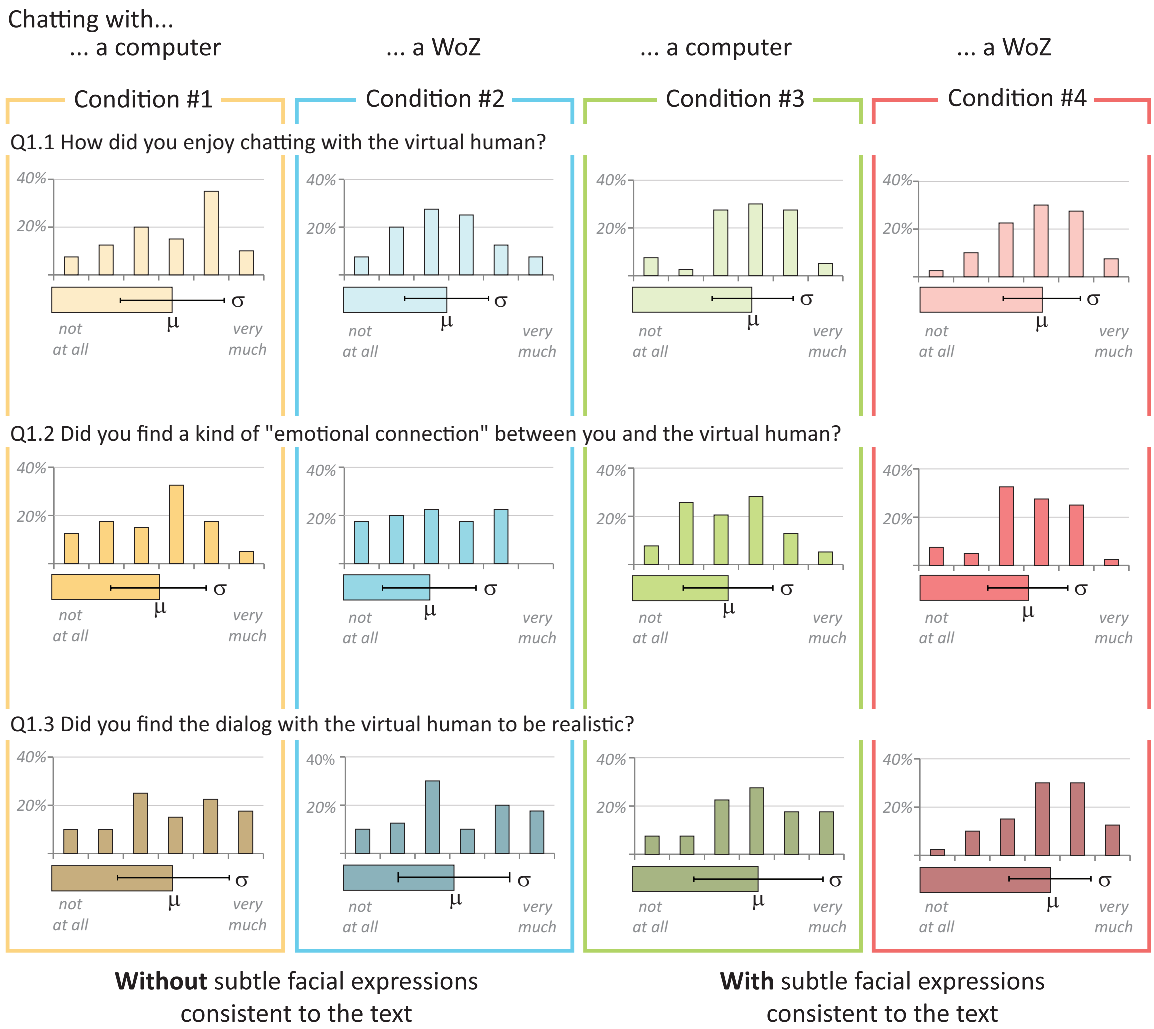

Figure 6: User-test results, comparing the four conditions: chatting appreciation (row 1), emotional connection (row 2), and dialog consistency (row 3). The corresponding statistical analysis is presented in the Figure 7.

ment for chatting and seeing the emotion of the VH chat partner.

VR 3D chatting appreciation: A large majority of users would like to repeat a similar experiments and found 20 minutes to be not long enough.

In post interview experimental, we asked all participants if they noticed any difference between the four conditions, especially in terms of graphics or animation of VHs. None of them noticed explicitly the presence of nonverbal cues, and most were surprised when showing rendered NVC cues. This suggests the kind of expression portrayed was subtle enough to contribute to communication despite their perception was unconscious. However, further improvements in the experimental methodology should confirm this fact.

A video demonstration and commented results of our user-test can be found at:

`http://go.epfl.ch/spontaneous-chatting`.

5 Discussion

In this paper we presented an impact study of nonverbal cues using a chatting system in a 3D environment focussing on one-to-one VH chatting. After briefly introduced the general architecture detailed in [GAS+11], we presented a user-test involving 40 participants performed with four conditions: with or without facial emotion and with a conversational system or a *Woz*. This test demonstrated that the influence of 3D graphical facial expression enhances the perception of the chatting enjoyment, the emotional connection, and even the dialog quality when chatting with another human. This implies that nonverbal communication simulation is promising for commercial ap-

	Condition ♯1			Condition ♯2			Condition ♯3			Condition ♯4		
	μ	σ	σ_M	μ	σ	σ_M	μ	σ	(σ_M)	μ	σ	σ_M
Q1.1	56%	1.45	3.83%	48%	1.34	3.52%	55%	1.22	3.22%	57%	1.19	3.14%
Q1.2	49%	1.41	3.72%	42%	1.44	3.80%	46%	1.32	3.48%	53%	1.19	3.15%
Q1.3	55%	1.56	4.12%	53%	1.58	4.17%	57%	1.66	4.38%	60%	1.25	3.29%

Table 2: Averages (μ) in percentage, standard deviations (σ) over 6 shown Figure 6, and standard error of the mean (σ_M) shown in black font Figures 7 and 8.

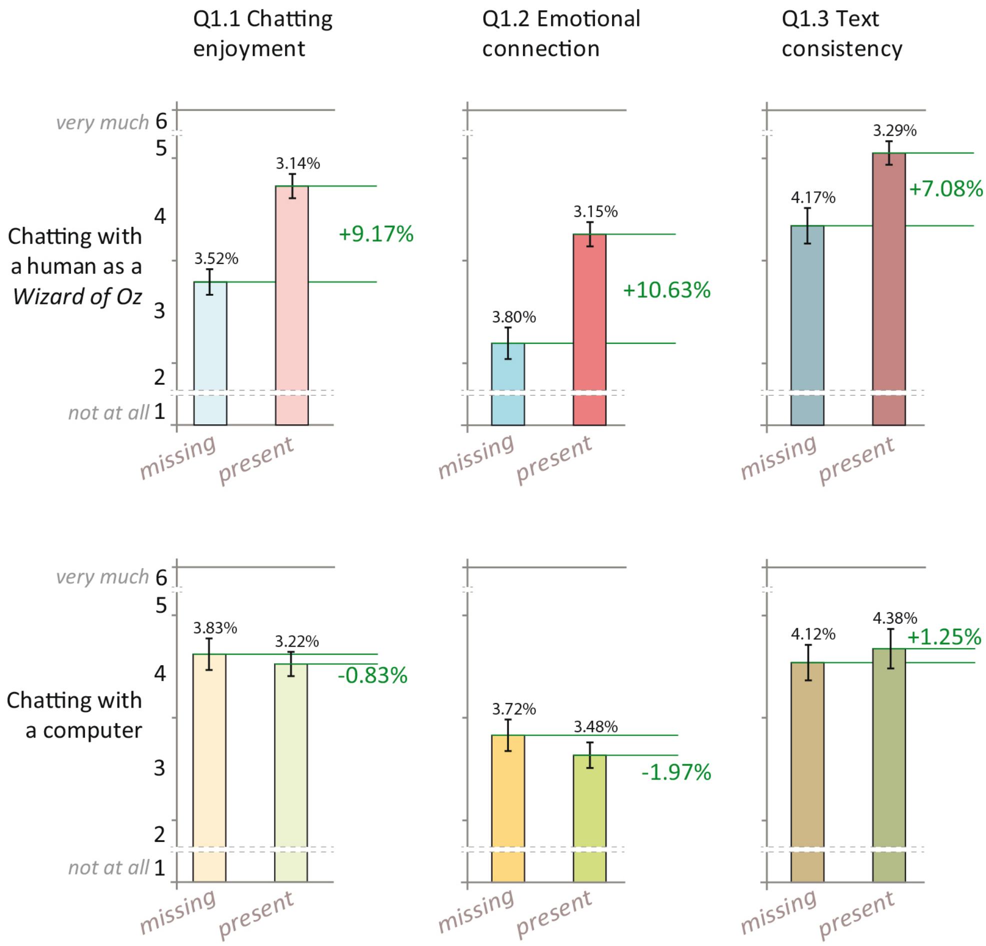

Figure 7: First comparison of user appreciation and influence of the subtle facial expressions when *missing* or *present*. Comparison of average (μ) and standard error of the mean (σ_M) percentages are presented respectively in green and black. Only graphs of the first row (representing conversations between humans) provide statistically significant enough differences for a conclusion: when chatting with facial expressions (even subtle compared to random limb animation), the chatting enjoyment, the emotional connection, and even the text consistency are perceived by users to be improved.

plications such as advertisements and entertainment. Furthermore, the test setup that was created provides a solid foundation: first, for further experiments and statistical evaluations at individual level (*i.e.* one to one

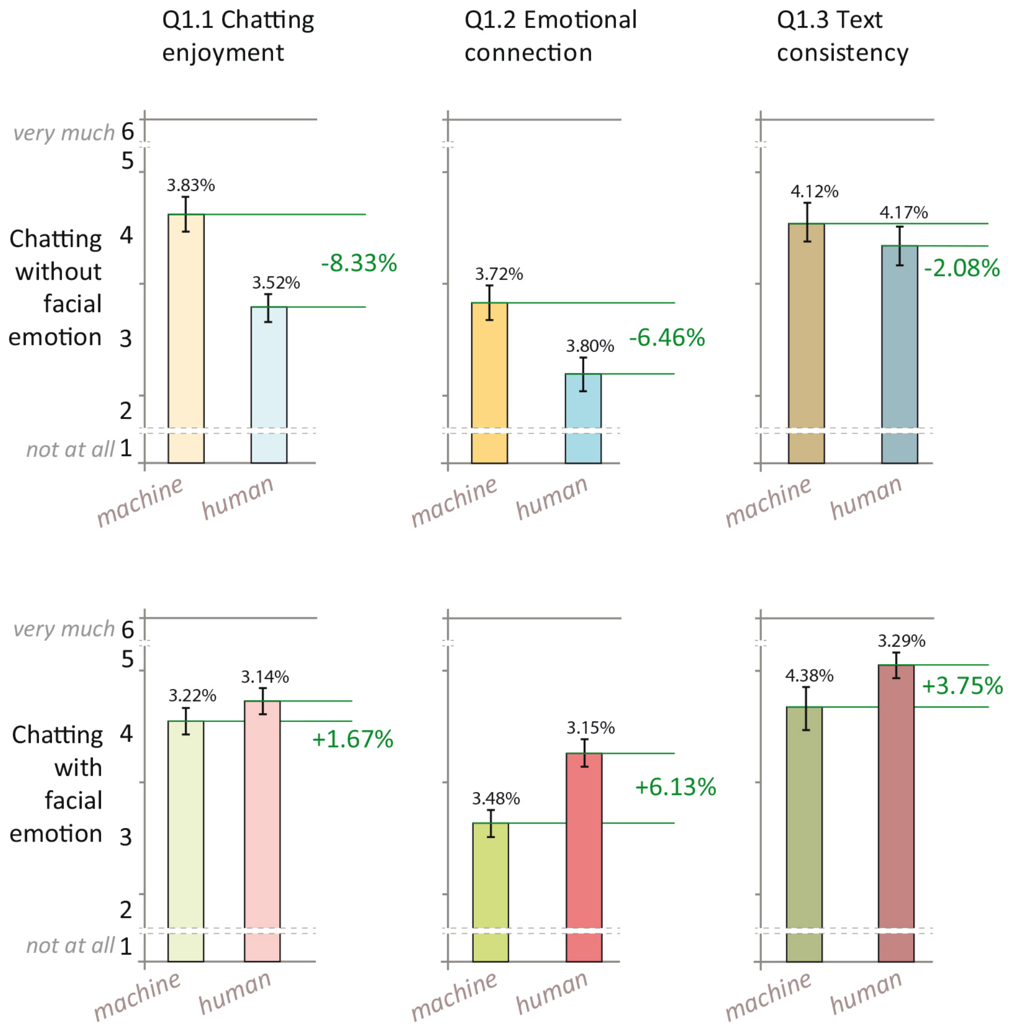

Figure 8: Second comparison of user appreciation: when chatting with a machine or a human (*Woz*). Comparison of average (μ) and standard error of the mean (σ_M) percentages are presented respectively in green and black. Three of the six graphs provide sufficient statically significant differences for a conclusion: when chatting without facial expressions, chatting is more enjoyable and emotional connection seems improved with a machine than with a human; and if facial expressions are rendered, users have a stronger emotional connection when chatting with a human.

VH communication); second, for extension to crowd interaction, a kind of "virtual society" level, *e.g.* MMO applications including realistic emotional reaction of user's avatars and AI system's agents.

Current works –We are currently working on multiple VHs chatting system; the following related improvements will soon be presented with two companion papers [GAG+12] and [AGG+12]:

a. We are developing a VH's emotional mind enriched with an additional emotion axis, *dominance*. This third axis partially solves the emotional ambiguities that can drastically change the VH behavior –*e.g.* rendering of the facial expression, body animation, choice of answer, etc.;

b. We parameterized facial movements directly according to the $\{v,a,d\}$ values;

c. As subjects had difficulties to see simultaneously the graphical and text windows, we developed an interactive text bubble that seems to facilitate the use 3D chatting system;

d. In addition to facial animation, conversational body movements such as arm movements are now adapted based on the emotional values.

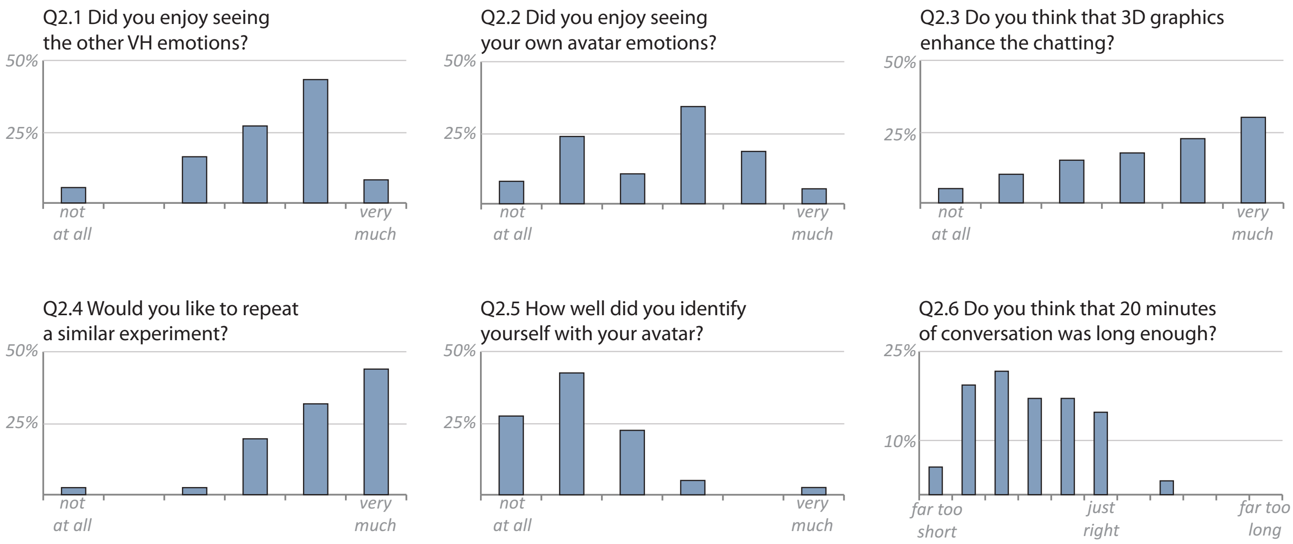

Figure 9: General questions on the user-test.

Future works –The following improvements are currently being parameterized and tested:

a. We first plan to make a large scale user-test including preliminary result of the above mentioned current work;

b. According to the user-test, some people mentioned that our system can be improved by using speech input instead of typing sentences. Allowing speech input will also reduce the time of our experiment;

c. Finally, VR equipment such as CAVE and motion capture system can be used to enrich our conversational environment. Emotional attributes could be extracted and should influence avatar's emotional behaviors."

Acknowledgements –Authors would like to thanks Dr. Joan Llobera for the proofreading of the manuscript. The research leading to these results has received funding from the European Community, Seventh Framework Programme (FP7/2007-2013) under grant agreement no.231323.

Citation
Stephane Gobron, Junghyun Ahn, Daniel Thalmann, Marcin Skowron, and Arvid Kappas, *Impact Study of Nonverbal Facial Cues on Spontaneous Chatting with Virtual Humans*, Journal of Virtual Reality and Broadcasting, 10(2013), no. 6, December 2013, urn:nbn:de:0009-6-38236, ISSN 1860-2037.

References

[ABB+04] Y. Arafa, L. M. Botelho, A. Bullock, P. Figueiredo, P. Gebhard, K. Hook, E. H. Mamdani, A. Paiva, P. Petta, P. Sengers, and M. Vala, *Affective Interactions for Real-time Applications: the SAFIRA Project*, Special issue of KI-Journal: Embodied Conversational Agents, KI **18** (2004), no. 1, 30, ISSN 0933-1875.

[AGG+12] J. Ahn, S. Gobron, D. Garcia, Q. Silvestre, D. Thalmann, and R. Boulic, *An NVC Emotional Model for Conversational Virtual Humans in a 3D Chatting Environment*, Articulated Motion and Deformable Objects: 7th International Conference, AMDO 2012, Port dAndratx, Mallorca, Spain, July 11-13, 2012, Lecture Notes in Computer Science, vol. 7378, 2012, pp. 47–57, ISBN 978-3-642-31566-4.

[ASN+70] M. Argyle, V. Salter, H. Nicholson, M. Williams, and P. Burgess, *The communication of inferior and superior attitudes by verbal and non-verbal signals*, British journal of social and clinical psychology **9** (1970), no. 3, 222–231, ISSN 2044-8260.

[BAT09] N. Bee, E. André, and S. Tober, *Breaking the Ice in Human-Agent Communication: Eye-Gaze Based Initiation of Contact with an Embodied Conversational Agent*, IVA 2009: Proceedings of the 9th international conference on Intelligent Virtual Agents (Berlin, Heidelberg), Springer-Verlag, 2009, pp. 229–242, ISBN 978-3-642-04379-6.

[BAZB02] N. Badler, J. Allbeck, L. Zhao, and M. Byun, *Representing and Parameterizing Agent Behaviors*, CA '02: Proceedings of the Computer Animation (Washington, DC, USA), IEEE Computer Society, 2002, pp. 133–143, ISBN 0-7695-1594-0.

[BC09] R. Beale and C. Creed, *Affective interaction: How emotional agents affect users*, International Journal of Human-Computer Studies **67** (2009), no. 9, 755–776, ISSN 1071-5819.

[BFRS10] T. W. Bickmore, R. Fernando, L. Ring, and D. Schulman, *Empathic Touch by Relational Agents*, IEEE Transactions on Affective Computing **1** (2010), no. 1, 60–71, ISSN 1949-3045.

[BK09] A. L. Baylor and S. Kim, *Designing nonverbal communication for pedagogical agents: When less is more*, Computers in Human Behavior **25** (2009), no. 2, 450–457, ISSN 0747-5632.

[BNP+05] C. Becker, A. Nakasone, H. Prendinger, M. Ishizuka, and I. Wachsmuth, *Physiologically interactive gaming with the 3d agent max*, Intl. Workshop on Conversational Informatics, in cooperation with JSAI 2005 (The 19th Annual Conference of JSAI), Kitakyushu, Japan, 2005, http://www.becker-asano.de/-PhysiologicallyInteractiveGamingMax.-pdf, pp. 37–42.

[BR03] D. Bohus and A. Rudnicky, *RavenClaw: Dialog Management Using Hierarchical Task, Decomposition and an Expectation Agenda*, In Proc. of the Eurospeech'03, 2003, pp. 597–600.

[Cas00] J. Cassell, *Embodied conversational agents*, MIT Press, Cambridge, MA, USA, 2000, ISBN 0-262-03278-3.

[CDWG08] S. D. Craig, S. DMello, A. Witherspoon, and A. Graesser, *Emote Aloud during learning with AutoTutor: Applying the Facial Action Coding System to cognitive-affective states during learning*, Cognition and Emotion **22** (2008), no. 5, 777–788, ISSN 0269-9931.

[CK07] J. F. Cohn and T. Kanade, *Automated facial image analysis for measurement of emotion expression*, Series in Affective Science, pp. 222–238, Oxford University Press, Oxford, 2007.

[CPB+94] J. Cassell, C. Pelachaud, N. Badler, M. Steedman, B. Achorn, T. Becket, B. Douville, S. Prevost, and M. Stone, *Animated conversation: rule-based generation of facial expression, gesture & spoken intonation for multiple conversational agents*, SIGGRAPH'94 (New York, NY, USA), ACM, 1994, pp. 413–420, ISBN 0-89791-667-0.

[CS01] B. Coyne and R. Sproat, *WordsEye: an automatic text-to-scene conversion system*, SIGGRAPH'01: Proceedings of the 28th annual conference on Computer graphics and interactive techniques (New York, NY, USA), ACM, 2001, pp. 487–496, ISBN 1-58113-374-X.

[CVB01] J. Cassell, H. H. Vilhjálmsson, and T. Bickmore, *BEAT: the Behavior Expression Animation Toolkit*, SIGGRAPH '01: Proceedings of the 28th annual conference on Computer graphics and interactive techniques (New York, NY, USA), ACM, 2001, pp. 477–486, ISBN 1-58113-374-X.

[EF78] P. Ekman and W. V. Friesen, *Facial Action Coding System*, Consulting Psychologists Press, Palo Alto, 1978.

[Ekm04] P. Ekman, *Emotions revealed*, Henry Holt and Company, LLC, New York, 2004, ISBN 9780805075168.

[EMMT04] A. Egges, T. Molet, and N. Magnenat-Thalmann, *Personalised Real-Time Idle Motion Synthesis*, PG '04: Proceedings of the Computer Graphics and Applications, 12th Pacific Conference (Washington, DC, USA), IEEE Computer Society, 2004, pp. 121–130, ISBN 0-7695-2234-3.

[EMO10] C. Ennis, R. McDonnell, and C. O'Sullivan, *Seeing is believing: body motion dominates in multisensory conversations*, ACM Trans. Graph. **29** (2010), no. 4, ISSN 0730-0301, article no. 91.

[ERA11] B. Endrass, M. Rehm, and E. Andre, *Planning Small Talk behavior with cultural influences for multiagent systems*, Computer Speech & Language **25** (2011), no. 2, 158–174, ISSN 0885-2308, Language and speech issues in the engineering of companionable dialogue systems.

[FKK+10] R. Fernández, Y. Katagiri, K. Komatani, O. Lemon, and M. Nakano, *Proceedings of the sigdial 2010 conference*, Association for Computational Linguistics, Tokyo, Japan, 2010, http://www.sigdial.org/workshops/-workshop11/proc/.

[GAG+12] S. Gobron, J. Ahn, D. Garcia, Q. Silvestre, D. Thalmann, and R. Boulic, *An Event-based Architecture to Manage Virtual Human Non-verbal Communication in 3D Chatting Environment*, Articulated Motion and Deformable Objects: 7th International Conference, AMDO 2012, Port dAndratx, Mallorca, Spain, July 11-13, 2012, Lecture Notes in Computer Science, vol. 7378, 2012, pp. 58–68, ISBN 978-3-642-31566-4.

[GAP+10] S. Gobron, J. Ahn, G. Paltoglou, M. Thelwall, and D. Thalmann, *From sentence to emotion: a real-time three-dimensional graphics metaphor of emotions extracted from text*, The Visual Computer **26** (2010), no. 6-8, 505–519, ISSN 0178-2789.

[GAS+11] S. Gobron, J. Ahn, Q. Silvestre, D. Thalmann, S. Rank, M. Skowron, G. Paltoglou, and M. Thelwall, *An Interdisciplinary VR-architecture for 3D Chatting with Non-verbal Communication*, EGVE/EuroVR (Sabine Coquillart, Anthony Steed, and Greg Welch, eds.), Eurographics Association, 2011, pp. 87–94, ISBN 978-3-905674-33-0.

[GB06] E. Gu and N. I. Badler, *Visual Attention and Eye Gaze During Multiparty Conversations with Distractions*, IVA 2006: Proceedings of the 6th international conference on Intelligent Virtual Agents, Springer, 2006, pp. 193–204, ISBN 3-540-37593-7.

[GGFO08] A. J. Gill, D. Gergle, R. M. French, and J. Oberlander, *Emotion rating from short blog texts*, In Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems, 2008, pp. 1121–1124, ISBN 978-1-60558-011-1.

[GM05] J. Gratch and S. Marsella, *Evaluating a Computational Model of Emotion*, Autonomous Agents and Multi-Agent Systems **11** (2005), no. 1, 23–43, ISSN 1387-2532.

[GSC+08] P. Gebhard, M. Schroeder, M. Charfuelan, Ch. Endres, M. Kipp, S. Pammi, M. Rumpler, and O. Tuerk, *IDEAS4Games: Building Expressive Virtual Characters for Computer Games*, In Proceedings of the 8th International Conference on Intelligent Virtual Agents, Lecture Notes in Computer Science, vol. 5208, Springer, 2008, pp. 426–440, ISBN 978-3-540-85482-1.

[GT09] H. Grillon and D. Thalmann, *Simulating gaze attention behaviors for crowds*, Comput. Animat. Virtual Worlds **20** (2009), no. 2-3, 111–119, ISSN 1546-4261.

[GTG+84] A. J. Giannini, D. Tamulonis, M. C. Giannini, R. H. Loiselle, and G. Spirtos, *Defective response to social cues in Mobius syndrome*, Journal of Nervous and

Mental Disease **172** (1984), no. 3, 174–175, ISSN 0022-3018.

[HPB+09] Patrick Healey, Roberto Pieraccini, Donna Byron, Steve Young, and Matthew Purver, *Proceedings of the sigdial 2009 conference*, Association for Computational Linguistics, London, UK, 2009, http://sigdial.org/workshops/workshop10/-proc/.

[Iza10] C. E. Izard, *The Many Meanings/Aspects of Emotion: Definitions, Functions, Activation, and Regulation*, Emotion Review **2** (2010), no. 4, 363–370, ISSN 1754-0739.

[Kap03] A. Kappas, *What facial expressions can and cannot tell us about emotions*, pp. 215–234, Kluwer Academic Publishers, 2003.

[Kap08] A. Kappas, *Psssst! Dr. Jekyll and Mr. Hyde are actually the same person! A tale of regulation and emotion*, pp. 15–38, Blackwell Publishing, Malden, MA, 2008.

[Kap10] A. Kappas, *Smile When You Read This, Whether You Like it or Not: Conceptual Challenges to Affect Detection*, IEEE Transactions on Affective Computing **1** (2010), no. 1, 38–41, ISSN 1949-3045.

[KHS91] A. Kappas, U. Hess, and K. R. Scherer, *Voice and emotion*, Fundamentals of nonverbal behavior (Cambridge) (R. S. Feldman and B. Rimé, eds.), Cambridge University Press, 1991, pp. 200–238, ISBN 0-521-36388-8.

[KW10] F. Khosmood and M. Walker, *Grapevine: a gossip generation system*, FDG '10: Proceedings of the Fifth International Conference on the Foundations of Digital Games (New York, NY, USA), ACM, 2010, pp. 92–99, ISBN 978-1-60558-937-4.

[LA07] M. Y. Lim and E. Aylett, *A new approach to emotion generation and expression*, 2nd International Conference on Affective Computing and Intelligent Interaction (Lisbon), 2007, http://www.di.uniba.it/intint/DC-ACII07/Lim.pdf.

[LRBW04] A. B. Loyall, W. S. N. Reilly, J. Bates, and P. Weyhrauch, *System for authoring highly interactive, personality-rich interactive characters*, SCA '04: Proceedings of the 2004 ACM SIGGRAPH/Eurographics symposium on Computer animation (Aire-la-Ville, Switzerland, Switzerland), Eurographics Association, 2004, pp. 59–68, ISBN 3-905673-14-2.

[McG96] S. McGlashan, *Towards multimodal dialogue management*, In Proceedings of Twente Workshop on Language Technology, 1996.

[MEDO09] R. McDonnell, C. Ennis, S. Dobbyn, and C. O'Sullivan, *Talking bodies: Sensitivity to desynchronization of conversations*, ACM Trans. Appl. Percept. **6** (2009), no. 4, ISSN 1544-3558, Article No. 22.

[MH07] S. Masuko and J. Hoshino, *Head-eye Animation Corresponding to a Conversation for CG Characters*, Computer Graphics Forum **26** (2007), no. 3, 303–312, ISSN 1467-8659.

[MMOD08] Ali Mazalek, Basil Mironer, Elijah O'Rear, and Dana Van Devender, *The TViews Table Role-Playing Game*, Journal of Virtual Reality and Broadcasting **5** (2008), no. 8, ISSN 1860-2037, urn:nbn:de:0009-6-14913.

[MWW09] S. Marks, J. Windsor, and B. Wünsche, *Enhancing Virtual Environment-Based Surgical Teamwork Training with Non-Verbal Communication*, Proceedings of GRAPP 2009, 2009, pp. 361–366.

[OCC88] A. Ortony, G. L. Clore, and A. Collins, *The cognitive structure of emotions*, New York: Cambridge University Press, 1988, ISBN 0-521-35364-5.

[OPS08] M. Ochs, C. Pelachaud, and D. Sadek, *An empathic virtual dialog agent to improve human-machine interaction*, AAMAS '08: Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems (Richland, SC), International Foundation for Autonomous Agents and Multiagent Systems, 2008, pp. 89–96, ISBN 978-0-9817381-0-9.

[OSS09] O. Oyekoya, W. Steptoe, and A. Steed, *A saliency-based method of simulating visual attention in virtual scenes*, VRST '09: Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology (New York, NY, USA), ACM, 2009, pp. 199–206, ISBN 978-1-60558-869-8.

[Pel09] C. Pelachaud, *Studies on gesture expressivity for a virtual agent*, Speech Commun. **51** (2009), no. 7, 630–639, ISSN 0167-6393.

[PG96] K. Perlin and A. Goldberg, *Improv: a system for scripting interactive actors in virtual worlds*, SIGGRAPH'96: Proceedings of the 23th annual conference on Computer graphics and interactive techniques (New York, NY, USA), ACM, 1996, pp. 205–216, ISBN 0-89791-746-4.

[Pic97] R. W. Picard, *Affective Computing*, MIT Press, Cambridge, MA, 1997, ISBN 0-262-16170-2.

[PJRC08] S.-H. Park, S.-H. Ji, D.-S. Ryu, and H.-G. Cho, *A new cognition-based chat system for avatar agents in virtual space*, VRCAI '08: Proceedings of The 7th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry (New York, NY, USA), ACM, 2008, pp. 1–6, ISBN 978-1-60558-335-8.

[PMD+10] M. Ptaszynski, J. Maciejewski, P. Dybala, R. Rzepka, and K. Araki, *CAO: A Fully Automatic Emoticon Analysis System Based on Theory of Kinesics*, IEEE Transactions on Affective Computing **1** (2010), no. 1, 46–59, ISSN 1949-3045.

[PP01] C. Pelachaud and I. Poggi, *Towards believable interactive embodied agents*, Fifth Int. Conf. on Autonomous Agents workshop on Multimodal Communication and Context in Embodied Agents, 2001.

[PPP07] A. Paiva, R. Prada, and R. W. Picard, *Affective computing and intelligent interaction, second International Conference, ACII 2007, Lisbon, Portugal, September 12-14, 2007*, Springer, Berlin, 2007, ISBN 978-3-540-74888-5.

[RBFD03] J. A. Russell, J.-A. Bachorowski, and J.-M. Fernandez-Dols, *Facial and vocal expressions of emotion*, Annual Review of Psychology, vol. 54, 2003, pp. 329–349.

[RGL+06] N. Reithinger, P. Gebhard, M. Loeckelt, A. Ndiaye, and N. Pfleger nd N. Klesen, *VirtualHuman - Dialogic and Affective Interaction with Virtual Characters*, ICMI '06 Proceedings of the 8th international conference on Multimodal interfaces, 2006, pp. 51–58, ISBN 1-59593-541-X.

[RKP11] A. S. Rizzo, P. Kenny, and T. D. Parsons, *Intelligent Virtual Patients for Training Clinical Skills*, Journal of Virtual Reality and Broadcasting **8** (2011), no. 3, ISSN 1860-2037, urn:nbn:de:0009-6-29025.

[Rus79] J. A. Russell, *Affective space is bipolar*, Journal of Personality and Social Psychology **37** (1979), no. 3, 345–356, ISSN 0022-3514.

[Rus97] J. A. Russell, *Reading emotions from and into faces: Resurrecting a dimensional-contextual perspective*, The psychology of facial expression (Cambridge), Cambridge University Press, 1997, pp. 295–320.

[SDO+04] M. Stone, D. DeCarlo, I. Oh, C. Rodriguez, A. Stere, A. Lees, and C. Bregler, *Speaking with hands: creating animated conversational characters from recordings of human performance*, ACM Trans. Graph. **23** (2004), no. 3, 506–513, ISSN 0730-0301.

[SE07] K. R. Scherer and H. Ellgring, *Are facial expressions of emotion produced by categorical affect programs or dynamically driven by appraisal?*, Emotion **7** (2007), no. 1, 113–130, ISSN 1528-3542.

[SH08] D. Schlangen and B. A. Hockey, *Proceedings of the 9th SIGdial Workshop on Discourse and Dialogue*, Association for Computational Linguistics, Columbus, Ohio, 2008, ISBN 978-1-932432-17-6.

[SPW07] W.-P. Su, B. Pham, and A. Wardhani, *Personality and Emotion-Based High-Level Control of Affective Story Characters*, IEEE Transactions on Visualization and Computer Graphics **13** (2007), no. 2, 281–293, ISSN 1077-2626.

[TBP+10] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, and Arvid Kappas, *Sentiment strength detection in short informal text*, Journal of the American Society for Information Science and Technology **61** (2010), no. 12, 2544–2558, ISSN 1532-2890.

[vDKP+08] K. van Deemter, B. Krenn, P. Piwek, M. Klesen, M. Schröder, and S. Baumann, *Fully generated scripted dialogue for embodied agents*, Artificial Intelligence Journal **172** (2008), no. 10, 1219–1244, ISSN 0004-3702.

[WDRG72] M. Weiner, S. Devoe, S. Rubinow, and J. Geller, *Nonverbal behavior and nonverbal communication*, Psychological Review **79** (1972), no. 3, 185–214, ISSN 0033-295X.

[WLO10] A. Weissenfeld, K. Liu, and J. Ostermann, *Video-realistic image-based eye animation via statistically driven state machines*, The Visual Computer **26** (2010), no. 9, 1201–1216, ISSN 0178-2789.

[WPY05] J. Williams, P. Poupart, and S. Young, *Partially observable markov decision processes with continuous observations for dialogue management*, In Proceedings of the 6th SigDial Workshop on Discourse and Dialogue, 2005, http://www.sigdial.org/workshops/-workshop6/proceedings/pdf/12-williams2005continuous.pdf, pp. 25–34.

[WSS94] J. H. Walker, L. Sproull, and R. Subramani, *Using a human face in an interface*, Proceedings of the SIGCHI conference on Human factors in computing systems: celebrating interdependence, CHI ’94, 1994, pp. 85–91, ISBN 0-89791-650-6.