INVITED PAPER *Special Section on Life-like Agent and its Communication*

Human Physiology as a Basis for Designing and Evaluating Affective Communication with Life-Like Characters

Helmut PRENDINGER[†a)], ***Nonmember and*** **Mitsuru ISHIZUKA**[††], ***Member***

SUMMARY This paper highlights some of our recent research efforts in designing and evaluating life-like characters that are capable of entertaining affective and social communication with human users. The key novelty of our approach is the use of human physiological information: *first*, as a method to evaluate the effect of life-like character behavior on a moment-to-moment basis, and *second*, as an input modality for a new generation of interface agents that we call 'physiologically perceptive' life-like characters. By exploiting the stream of primarily involuntary human responses, such as autonomic nervous system activity or eye movements, those characters are expected to respond to users' affective and social needs in a truly sensitive, and hence effective, friendly, and beneficial way.

key words: *interface agents, affective communication, markup languages, evaluation, bio-signal, eye tracking*

1. Introduction and Overview

Life-like characters, or animated interface agents, have attracted considerable interest and attention in recent years, mainly for their ability to emulate human–human communication styles that are expected to improve the intuitiveness and effectiveness of human–computer interfaces. Following this user interface paradigm, a considerable number of animated agent (or character) based systems has been developed, ranging from information presentation and online sales to personal assistance, entertainment, and tutoring [1]–[3].

While significant progress has been made in individual aspects of the 'life-likeness' of animated agents, such as their graphical appearance or quality of synthetic voice, research on methods to change entertaining and visually attractive characters into veritable empathic interaction partners is still in its infancy.

The type of affective interaction between humans and characters we envision calls for interface agents that unobtrusively recognize a user's affective state and respond to the user in a sensitive and appropriate manner. This goal involves research challenges on multiple aspects of the affective interaction loop. First of all, life-like characters should be endowed with psychologically adequate models of emotion and social (or socially intelligent) behavior. On the practical side, technologies for easy and intuitive control of affective character behavior should be provided. Moreover, successful affective interaction should have a quantifiable effect on users' affective and attentional state. Finally, characters should be capable of responding to users' affective state or focus of attention in real-time, whenever an emotion is elicited or the focus of attention has changed.

In the remainder of this section, our approach to implementing affective communication with life-like characters as well as related research efforts will be described more thoroughly. In the rest of the paper, we will present some of our recent systems and empirical results that aim at advancing the science of affective interaction with life-like characters.

1.1 Scripting and Modeling Emotion-Based Life-Like Characters

Two core objectives related to designing for affective communication with life-like characters are (i) to develop scripting languages that allow for easy control of the (affective) behavior of synthetic agents; and (ii) to increase the believability of those agents by providing them with affective functions such as emotion, personality, and social intelligence.

In our work, the first objective was achieved by designing a XML-compliant language, the Multi-modal Presentation Markup Language (MPML), which offers easy-to-use tagging structures for the coordination of the verbal and non-verbal behavior of one or multiple characters, affective display, and integration into a web-based environment [4]. Given our practical approach, we largely side-stepped issues of sophisticated animation, and mostly used the Microsoft Agent package as our animation engine [5] (but see also [6]). This package provides high-level controls to animate affective behavior of 2D cartoon-style characters, a text-to-speech engine, and a voice recognizer. Other existing scripting languages put a different focus and usually cover a range of different 'abstraction levels' in a single language. The Character Markup Language (CML) allows specifying high-level concepts such as the emotion "happy" as well as low-level behaviors like "blinking" [7]. The Virtual Human Markup Language (VHML) comprises tags for facial and body animation, speech, gesture, and even dialogues [8]. Scripting languages also differ in their emphasis on a particular competence envisioned for the character. The Behavior Expression Animation Toolkit (BEAT) provides sophisticated synchronization of synthetic speech and non-verbal behavior by parsing textual input [9] and the

Manuscript received June 13, 2005.
[†]The author is with the National Institute of Informatics, Tokyo, 101–8430 Japan.
[††]The author is with the Department of Creative Informatics and the Department of Information and Communication Engineering, Graduate School of Information Science and Technology, University of Tokyo, Tokyo, 113–8656 Japan.
a) E-mail: helmut@nii.ac.jp
DOI: 10.1093/ietisy/e88–d.11.2453

Affective Presentation Markup Language (APML) targets communicative functions. While those approaches permit a more fine-grained control of life-like character animation, they often defeat our goal of creating a character markup language that can be used easily by average users rather than animation experts.

The second objective was met by basing character behavior on psychological models of emotion and personality, as well as findings from socio-psychological studies. We developed the Scripting Emotion-based Agent Minds (SCREAM) system, a mechanism that allows one to design characters that autonomously generate emotionally and socially appropriate behaviors depending on their mental make-up [10]. In SCREAM, a character's mental state is determined by its goals, beliefs, attitudes, affect-related features of the interlocutor's behavior, and parameters peculiar to the social interaction context. Our system is thus similar to the Em architecture [11] or the Affective Reasoner [12].

1.2 Physiology-Based Evaluation Methods for Life-Like Characters

A feature common to most evaluations of life-like characters is that they are based on questionnaires and focus on the user's experience with the systems hosting them, including questions about their believability, likeability, engagingness, utility, and ability to attract attention [13]. However, as pointed out in [14], the broad variety of realizations of life-like characters and interaction scenarios complicates their comparison, and presumably of more importance, subtle aspects of the interaction, such as whether users follow human–human social interaction protocols when communicating with the character or not, cannot be deduced reliably from self-reports [15].

We therefore propose a different approach to evaluating life-like characters – one that is based on human physiology. *First*, human electrodermal and cardiovascular responses are tracked during the interaction between users and life-like characters. There is a huge body of evidence that physiological signals (or bio-signals) such as skin conductance, muscle tension, and heart rate provide important information regarding the intensity and quality of a person's experience, and can thus be used to infer a user's emotion or affective state [16], [17]. This quality of human–computer interaction is essential since extreme cognitive and affective states, e.g. too high or too low arousal, or cognitive load, can hamper a user's information processing capability [18]. Employing bio-signals allows us to find answers to questions such as:

- Does the interaction with a character produce a state of relaxed happiness or stress in the user?
- Which particular character behaviors cause (or alleviate) stress or frustration in the user?

Physiological variables are useful to gain insight into short-term changes of the user's affective state when interacting with life-like characters that might be hard to measure by other means, e.g. post-experiment questionnaires [19].

Second, eye movements of users interacting with a character-based interface are analyzed. Although gaze point and focus of attention are not necessarily always identical, eye movement data provide rich evidence of the user's visual and (overt) attentional processes [20]. Hence, eye movement information can be used to answer questions such as:

- Can the character's verbal or gestural behavior direct the user's focus of attention?
- Does the user adhere to gaze interaction patterns known from human–human communication?

Eye movement data can offer valuable information for assessing the utility and usability of interfaces employing life-like characters. The tracking of eye movements lends itself to reliably capturing the moment-to-moment experience of users interacting with characters.

1.3 Physiologically Perceptive Life-Like Characters

In the previous section, the analysis of bio-signals and eye movements has been suggested for a *diagnostic* purpose, in particular to evaluate the utility of character-based interfaces. An exciting complementary role for physiological user data is its *interactive* use in order to increase the bandwidth of affective information from user to computer.

'Physiologically perceptive life-like characters' refers to a new generation of interface agents that can be characterized by two features:

- Characters that recognize the user's affective state by analyzing bio-signals;
- Characters that are aware of the user's focus of attention by processing eye movements.

Unlike input modalities that can be easily regulated by humans, such as facial expressions or speech, physiological information is (mostly) involuntary and hence allows us to obtain a more reliable assessment of the user's cognitive and affective state.

Surprisingly little work exists that exploits bio-signal information as an input modality. A notable exception is the upcoming field of 'affective gaming', which refers to research on the evocation and detection of emotion during game play [21]. Most prominently, the "Relax-to-Win" game is a competitive two player racing game where a player's level of arousal, derived from the player's skin conductance, determines the speed of a dragon figure [22]. The more the player relaxes, the faster the dragon will move, and thus win the race. The "Mind Balance" game uses brain waves (Visually Evoked Potentials) in order to assist a graphical tightrope walker to keep its balance [23]. Stevenson [21] uses pupil dilation (with co-recorded electrodermal activity) to measure player arousal during game play.

In our work, we process users' skin conductance, heart rate, and electromyography in real-time to select an appropriate empathic character response [24], [25]. Bio-signal information has also been used to resolve ambiguous dialogue acts [26].

Eye movements, on the other hand, received considerable attention in the human–computer interaction research community. Often, eye movements are investigated as a novel interaction technique, in order to facilitate (interface) object selection, moving of an object (a variation of the 'drag-and-drop' operation) and scrolling of text [27]. Eye-based interaction also plays a key role in visual attentive interfaces [28]. In the realm of life-like character based systems, [29] consider a user's focus of attention (among others) to decide an appropriate response for an educational software, and [30] investigate attentional focus (among others) for a direction-giving task.

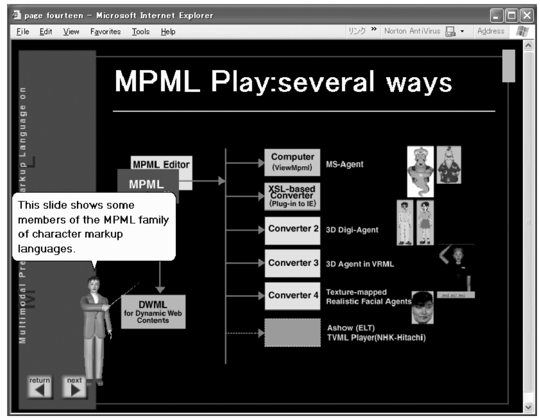

Fig. 1 The "Shima" character presents control technologies for life-like characters in a web-based environment.

2. Scripting and Modeling Characters with Affective Behavior

This section describes our approach to controlling the 'body' and the 'mind' of affective life-like characters.

MPML (Multi-modal Presentation Markup Language) is a language for coordinating affective behavior of life-like characters, specifically designed for non-expert content authors [4]. MPML is a *markup language* compliant with standard XML and hence allows for scripting in a style that is familiar to a broad audience (assuming some background with HTML scripting). MPML is a language designed with the aim of scripting character-based *presentations* that can be viewed primarily in a web browser (see Fig. 1), but also on cellular phones. In order to facilitate the generation of different types of presentations, in particular interactive presentations, MPML provides tagging structures that enable authors to accept various types of input from the user, such as speech or menu selection. Finally, MPML supports the generation of *multi-modal* presentations, i.e. presentations utilizing multiple mechanisms to encode the information to be conveyed, including 2D/3D graphics, audio, and video. Our particular focus has been on modalities specific to humanoid characters; besides affective speech, characters may communicate information by using modalities such as facial displays in order to express emotions, hand gestures including pointing and propositional gestures, head movements, and body posture.

A task complementary to scripting the embodiment of a character is to author the character's mental state and emotional reaction to its environment. We have developed a system called SCREAM (SCRipting Emotion-based Agent Minds) that facilitates scripting a character's affect-related processing capabilities [10]. The system allows to specify a character's mental make-up and to endow it with emotion and personality that are considered to be key features for the life-likeness of characters. A character's mental state can be scripted at many levels of detail ('granularity levels'), from driven purely by (personality) traits to having full awareness of the social interaction situation, including character-specific beliefs and beliefs attributed to others, e.g. the user. The system is designed to generate believable affective character behavior as well as believable transitions between the expression of emotional states. Hence, a character's emotional state will shift from a state of anger to joyfulness with character-specific latency, rather than in an abrupt way as often observed with emotional agents. Spanning a smooth and natural 'emotional arc' of affective behavior is of chief importance as the interaction with a character should be convincing for an extended period of time.

The utility of the SCREAM system has been demonstrated by implementing a virtual casino scenario where the user plays the Black Jack game against an animated dealer and receives suggestions from an animated advisor [10]. The advisor character responds to the user depending on its affective state derived from its mental make-up and the user's game play. Due to sophisticated modulation of emotion expression and attitude change, the advisor character could achieve a high level of naturalness and believability in its responses, as demonstrated e.g. at the First International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2002).

In the following, we will briefly describe the main components of the SCREAM system: generation, regulation, and expression of emotions, and the agent model (for extensive discussion, see [10]). A core activity of an emotion-based character is the appraisal process where events are evaluated regarding their emotional significance for the agent in terms of its beliefs, goals, and attitudes [31]. Generated emotion types include "joy", "distress", "happy for", "sorry for", etc. (twenty-two in total), and have associated intensities. Since more than one emotion type is typically triggered when the character interacts with its environment, dedicated SCREAM modules determine the most dominant emotion and handle the decay process of emotions. The expression of emotions is governed by social and cultural norms that have significant impact on the intensity of their expression [32]. Therefore, we defined regulation parameters and categorized them into ones that constitute a social threat for the agent (social distance and social power), and parameters that refer to the agent's capability of (self-) control (personality, interlocutor personality, and linguistic style). An overall control value then determines the intensity of emotion expression.

The agent model subsumes static features including the agent's personality (agreeableness, extroversion), and dynamic features such as goals, attitudes, and social variables. Of particular importance for believability is that the character's attitude changes as a result of social interaction. We employed the '(signed) summary record' model of dispositional (dis) liking [33], which stores the sign (positive or negative) and intensity of emotions that were induced during interaction. In effect, attitudes not only contribute to the elicitation of emotions by deciding whether the agent has a "sorry for" or "gloat" emotion – but induced emotions may also change the agent's affective state, in particular its attitude and familiarity towards other agents.

3. Evaluating Affective Interactions with Life-Like Characters

We now turn to two of our recent studies that apply physiology-based methods to measure the effect of character behavior on users. The first study tracks bio-signals with the aim of evaluating the impact of an empathic character on the users' arousal and stress levels [34]. The second study analyzes users' eye movements to demonstrate the utility of life-like characters as navigational guides in an e-commerce setting [35].

3.1 The Impact of Empathic Character Response on Users' Level of Arousal

In the study, a life-like character acts as a quizmaster of a mathematical game[†]. This application was chosen as a simple, and for the sake of the experiment, highly controllable, instance of human–computer interaction. As skin conductance increases with a person's overall level of arousal or stress [18], the impact of empathic response was measured by comparing the skin conductance readings of subjects that received empathic feedback with the readings of subjects that did not.

A quiz game was implemented where subjects are instructed to sum up five consecutively displayed numbers and are then asked to subtract the *i*-th number of the sequence ($i \leq 4$). The quizmaster is a 2D cartoon-style animated agent called "Shima" that was developed by [36] and is controlled by MPML [4]. The Shima character is designed with gestures of a typical Japanese professional and hence the character's actions are familiar to Japanese people and easily understood. For instance, one animation shows the agent bowing, a gesture which Japanese people readily perceive as a signal of the interlocutor's apology (see Fig. 2).

Subjects compete for the best score in terms of correct answers and time. They were told that they would interact with a prototype interface that may still contain some bugs. This warning was essential since in some quiz questions, a delay was inserted before showing the 5th number. The delay was assumed to induce frustration as the subjects' goals of giving the correct answer and achieving a fast score are

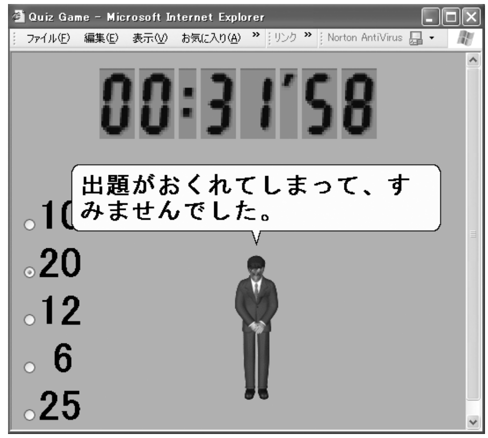

Fig. 2 Shima character: "I apologize that there was a delay in posing the question."

thwarted. In the version of the game using the empathic agent, an apology as depicted in Fig. 2 was shown to subjects, while in the non-affective version the occurrence of the delay was simply ignored by the animated instructor.

In order to demonstrate the effect of the character's empathic behavior on the physiological state of subjects, we have been interested in two specific segments: (i) the DELAY segment refers to the period after which the character suddenly stops activity while the question is not completed until the moment when the agent continues with the question; (ii) the RESPONSE-TO-DELAY segment refers to the period when the character expresses empathy concerning the delay, or ignores the occurrence of the delay, which follows the character's response to the subject's answer (regarding its correctness). The main results of the study are as follows:

- If a character shows empathy to a deliberately frustrated user (RESPONSE-TO-DELAY period), then the user's skin conductance is significantly lower than when the character does not display empathy, as compared to the period of induced frustration (the DELAY period).
- A character with affective behavior has a positive effect on users' perception of task difficulty, i.e. users perceive the game as less difficult.

If the level of skin conductance is interpreted as the user's level of stress or frustration, the result indicates that empathic feedback may *undo* some of the user's negative emotions. On the other hand, affective behavior could not be shown to have an impact on users' performance in playing the quiz game, an outcome that is in line with comparable research [14].

[†]This section builds on collaborative work with Junichiro Mori from the University of Tokyo.

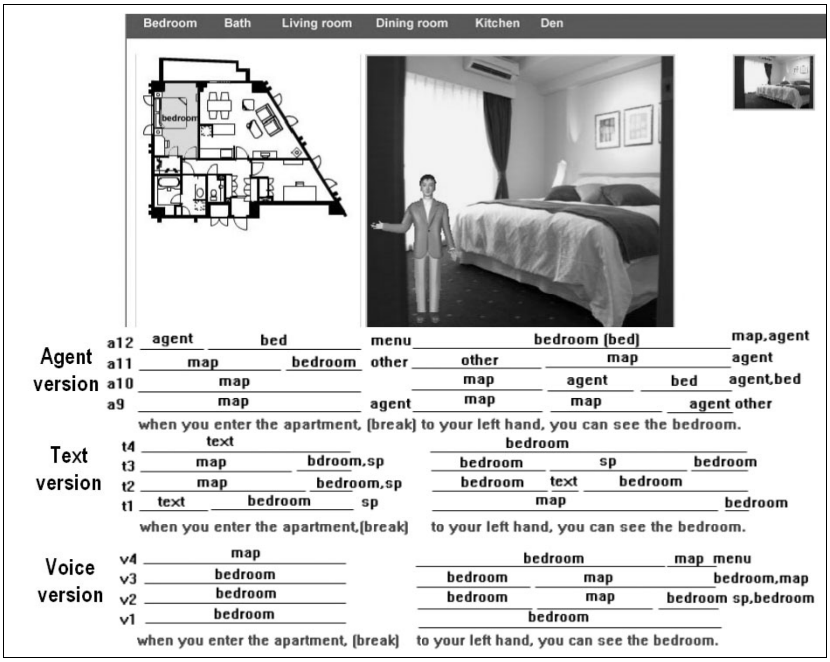

Fig. 3 Effect of deictic reference on eye movement. Underlining indicates the time period a user spends for attending to the respective interface object. (The abbreviation“sp” refers to the small picture to the top-right.)

3.2 The Effect of a Character's Deictic Gestures on Users' Focus of Attention

For the eye tracking study, a presentation of an apartment located in Tokyo has been prepared using a web page based interface†. Views of each room of the apartment are shown during the presentation, including pictures of some parts of the room and close-up pictures. Three versions of the apartment show have been designed: (i) *Agent (& Speech)* version: the Shima character presents the apartment using synthetic speech and deictic facial and hand gestures; (ii) *Text (& Speech)* version: the presentation content of each scene is displayed by a text box and read out by Microsoft Reader; (iii) *Voice (only)* version: speech is the only medium used to comment on the apartment. The same type and speed of (synthetic) voice was used in all versions.

For data analysis, the recorded video data of a presentation were first divided into individual scenes. A scene is a presentation unit where a referring entity (agent, text box, or voice) describes a reference object (an item of the apartment). For each scene, the following screen area categories were defined: (i) a (visible) referring entity: the agent or the text box (the agent area is further subdivided into face and body areas); (ii) the reference object: the object currently described; (iii) the apartment layout area (a designated, permanent reference object); (iv) other screen areas.

Results were distilled from applying both *spatial (cumulative)* and *spatio-temporal* analysis methods [35]. Spatial analysis counts the gaze points that fall within certain screen areas and hypothesizes users' attentional focus. Our findings of the cumulative analysis include e.g. that users are looking mostly at the character's face, indicating users interact socially with agents. While a spatial analysis can indicate where attention is spent, it cannot reveal how users traverse the interface when watching a presentation. In order to address those more complex aspects of character-based interfaces, we also performed a spatio-temporal analysis.

In Fig. 3, the referring entity (agent, text box, voice) is intended to direct the user's attention to the map area that depicts the bedroom. It is important to notice that the word "bedroom" in the sentence is ambiguous with regard to its reference object: "bedroom" might refer to either the specified area in the map to the left or to the picture of the bedroom to the right. In the Agent version only, subjects mostly direct their attention to the intended reference object, the map. Although subjects in the Voice version eventually attend to the map, subjects in the Agent version (mostly) do so from the beginning. This kind of user behavior is seemingly effected by the agent performing a deictic gesture (to its right) shortly before starting the utterance. In the Text version, subjects resolve the reference object only with considerable latency.

Our findings also shed new light on *auditory language processing*, claiming that people who simultaneously listen to speech and are shown a visual object featuring elements that are semantically related to the spoken information, tend to focus on the elements that are most closely related to the meaning of the currently heard spoken language [37]. We observed that auditory language processing has less latency

†The study described in this section has been conducted jointly with Chunling Ma from the University of Tokyo.

when accompanied by according deictic gestures. Our observations also advance the discussion about the believability of life-like characters in that they contribute quantifiable evidence on the character's ability to direct the user's focus of attention to objects of interest.

4. Affective Interaction with Physiologically Perceptive Life-Like Characters

This section reports on our first experiences with physiologically perceptive characters. Currently, only bio-signals are considered as an input modality.

4.1 An Empathic Companion

In the virtual job interview scenario, one character takes the role of a (virtual) interviewer, whereas a second character, the *Empathic Companion*, offers the user (as interviewee) companionship during the interview [24]. The Empathic Companion character assesses the user's emotion during the questioning and provides (positive) empathic feedback, if the user is detected to be a negative affective state. Emotion recognition is based on the two-dimensional model of [38], which claims that all emotions can be characterized in terms of judged valence (positive or negative) and arousal (calm or aroused). As skin conductance increases with a person's level of overall arousal, and electromyography correlates with negatively valenced emotions [18], named emotions can be identified in the arousal–valence space proposed by [38]. E.g. the "frustrated" emotion is located in the 'arousal=medium-high'&'valence=negative' segment. Those values are used by a decision network that outputs an appropriate response for the Empathic Companion. If the user's emotion is recognized as frustrated, the character will show empathy, e.g. by saying "I am sorry that you seem to feel bad about this question". Else, if the character detects that the user is aroused without experiencing a negative emotion (and gave a satisfying answer to the interviewer), the character will congratulate the user.

Besides exploiting bio-signal information to provide appropriate character responses, we also performed an exploratory study to assess the impact of the Empathic Companion on the user. As a control condition a 'non-empathic' companion was used. The results demonstrate that in the empathic condition, users show significantly lower arousal (stress) levels while being questioned by the interviewer character [24].

4.2 Affective Gaming

We recently conducted a follow-up study that measures user emotion (derived from skin conductance and electromyography) in an interactive cards game where the user plays the "Skip-Bo" against a virtual opponent [25]. Here, users' physiological responses can be more directly attributed to the affective behavior of the character, since the user interacts with only one character (unlike the Empathic Companion system). The game partner is based on Max, a 3D character [39] that uses the emotion simulation system described in [40]†. As Skip-Bo is a competitive game, the impact of two types of empathy on the user could be implemented. In the *positive empathic* version, the character displays happiness if the user is detected to be in happy or relaxed affective state. In the *negative empathic* version, on the other hand, the character will display e.g. gloating joy if the user is recognized to be negatively aroused. In both cases, the character will also display self-centered emotions, such as being happy about its own successful game move. As control conditions, the character will either display only self-centered emotions or no emotions at all. First results of this study indicate that the absence of negative empathy is conceived as stressful (derived from skin conductance) and irritating, as it might also be experienced when playing against a human player. A complementary result is that negative emphatic behavior induces negatively valenced emotions (derived from electromyography) in the user.

5. Conclusions and Outlook

The paper describes our recent efforts towards the implementation of affective communication between human users and life-like characters. Our primary goal is to design agents that recognize physiological information in real-time, and appropriately address users' affective state and attentional focus. In this paper, we presented key challenges in designing a new generation of agents that we call physiologically perceptive life-like characters. The challenges include the control, mind design, evaluation, and novel perceptive capabilities of life-like characters.

MPML is currently being extended to support authors in the creation of more personalized, powerful, and flexible presentation and interactive content. Specifically, we intend to integrate concepts from virtual storytelling [41] and chatbot systems [42]. Its most recent version, MPML-HR, may even control the Honda humanoid robot ASIMO [6].

Inspired by work on gaze behavior [30], a new system architecture is being developed that tracks and analyzes both bio-signals and eye movements, such that characters will be capable of responding to the user's focus of attention and affective interest. Eye tracking technology will also allow us to investigate users' pupil dilation that has been shown as an index for the affective states of "confusion" and "surprise" [43]. Furthermore, we plan to combine multiple input modalities in order to improve the reliability and robustness of affect recognition [44]. Besides physiology, affect sensing from text appears as a promising approach [45].

The salient feature of our approach to affective communication with characters is that we use physiological information for both *diagnostic* and *interactive* roles. In its diagnostic role, physiology provides us with moment-by-

†The following description is based on joint research with Christian Becker from the University of Bielefeld, performed during his stay as a JSPS Pre-Doctoral Fellow at the National Institute of Informatics, Tokyo.

moment information about a character's effect on the user's affective and attentional state. Our empirical results enable us to design more natural and effective characters. In its interactive role, physiology allows to react to users in an appropriate and sensitive way and thus create the illusion of characters that are aware of users, emphasize with them, and address their social and affective needs.

Eventually, we expect physiologically perceptive life-like characters to be of chief importance and utility in e-learning, product presentation, and edutainment applications and environments.

Acknowledgement

This research was supported by the Research Grant (FY1999–FY2003) for the Future Program of the Japan Society for the Promotion of Science (JSPS), by a JSPS Encouragement of Young Scientists Grant (FY2005–FY2007), and a NII Joint Research Grant (FY2005).

References

[1] J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, eds., Embodied Conversational Agents, The MIT Press, Cambridge, MA, 2000.
[2] H. Prendinger and M. Ishizuka, eds., Life-Like Characters. Tools, Affective Functions, and Applications, Cognitive Technologies, Springer Verlag, Berlin Heidelberg, 2004.
[3] T. Rist, E. André, S. Baldes, P. Gebhard, M. Klesen, M. Kipp, P. Rist, and M. Schmitt, "A review of the development of embodied presentation agents and their appication fields," in Life-Like Characters. Tools, Affective Functions, and Applications, ed. H. Prendinger and M. Ishizuka, Cognitive Technologies, pp.377–404, Springer Verlag, Berlin Heidelberg, 2004.
[4] H. Prendinger, S. Descamps, and M. Ishizuka, "MPML: A markup language for controlling the behavior of life-like characters," J. Visual Languages and Computing, vol.15, no.2, pp.183–203, 2004.
[5] Microsoft, Developing for Microsoft Agent, Microsoft Press, Redmond, WA, 1998.
[6] K. Kushida, Y. Nishimura, H. Dohi, M. Ishizuka, J. Takeuchi, and H. Tsujino, "Humanoid robot presentation through multimodal presentation markup language MPML-HR," AAMAS-05 Workshop on Creating Bonds with Humanoids, 2005.
[7] Y. Arafa, K. Kamyab, and E. Mamdani, "Towards a unified scripting language. Lessons learned from developing CML & AML," in Life-Like Characters. Tools, Affective Functions, and Applications, ed. H. Prendinger and M. Ishizuka, Cognitive Technologies, pp.39–63, Springer Verlag, Berlin Heidelberg, 2004.
[8] A. Mariott and J. Stallo, "VHML—Uncertainties and problems. A discussion," Proc. AAMAS-02 Workshop on Embodied Conversational Agents—let's specify and evaluate them!, 2002.
[9] J. Cassell, H. Vilhjálmsson, and T. Bickmore, "BEAT: the Behavior Expression Animation Toolkit," in Life-Like Characters. Tools, Affective Functions, and Applications, Cognitive Technologies, ed. H. Prendinger, and M. Ishizuka, pp.163–185, Springer Verlag, Berlin Heidelberg, 2004.
[10] H. Prendinger, S. Descamps, and M. Ishizuka, "Scripting affective communication with life-like characters in web-based interaction systems," Applied Artificial Intelligence, vol.16, no.7-8, pp.519–553, 2002.
[11] W.S.N. Reilly, Believable Social and Emotional Agents, Ph.D. Thesis, Carnegie Mellon University, 1996. CMU-CS-96-138.
[12] C. Elliott, The Affective Reasoner. A process model of emotions in a multi-agent system, Ph.D. Thesis, Northwestern University, Institute for the Learning Sciences, 1992.
[13] J.C. Lester, S.A. Converse, S.E. Kahler, S.T. Barlow, B.A. Stone, and R.S. Bhogal, "The Persona effect: Affective impact of animated pedagogical agents," Proc. CHI-97, pp.359–366, ACM Press, 1997.
[14] D.M. Dehn and S. van Mulken, "The impact of animated interface agents: A review of empirical research," Int. J. Human-Computer Studies, no.52, pp.1–22, 2000.
[15] R.E. Nisbett and T.D. Wilson, "Telling more than we know: Verbal reports on mental processes," Psychological Review, vol.84, pp.231–259, 1977.
[16] R.W. Picard, Affective Computing, The MIT Press, Cambridge, MA, 1997.
[17] R.W. Levenson, "Autonomic specifity and emotion," in Handbook of Affective Sciences, ed. R.J. Davidson, K.R. Scherer, and H.H. Goldsmith, pp.212–224, Oxford University Press, Oxford, 2003.
[18] J.L. Andreassi, Psychophysiology. Human Behavior & Physiological Response, 4 ed., Lawrence Erlbaum Associates, Mahwah, NJ, 2000.
[19] J. Schreirer, R. Fernandez, J. Klein, and R.W. Picard, "Frustrating the user on purpose: A step toward building an affective computer," Interacting with Computers, vol.14, pp.93–118, 2002.
[20] A.T. Duchowski, Eye Tracking Methodology: Theory and Practice, Springer, London, UK, 2003.
[21] C. Stevenson, "Affective gaming: Using pupil dilation to determine player arousal," The 2005 International Conference on Changing Views: Worlds in Play, 2005.
[22] D. Bersak, G. McDarby, N. Augenblick, P. McDarby, D. McDonnell, B. McDonald, and R. Karkun, "Intelligent biofeedback using an immersive competitive environment," Online Proc. Ubicomp 2001 Workshop on Designing Ubiquitous Computer Games, 2001.
[23] R. Burke, "Breathing new life into accessibility," CodeZone Magazine, Dec. 2003.
[24] H. Prendinger and M. Ishizuka, "The Empathic Companion: A character-based interface that addresses users' affective states," Int. J. Applied Artificial Intelligence, vol.19, no.3, pp.267–285, 2005.
[25] C. Becker, H. Prendinger, M. Ishizuka, and I. Wachsmuth, "Evaluating affective feedback of the 3D agent Max in a competitive cards game," Proc. First International Conference on Affective Computing and Intelligent Interaction (ACII-05), LNCS 3784, pp.466–473, Springer, 2005.
[26] W. Bosma and E. André, "Exploiting emotions to disambiguate dialogue acts," Proc. 9th International Conference on Intelligent User Interfaces (IUI-04), pp.85–92, ACM Press, New York, 2004.
[27] R.J.K. Jacob, "The use of eye movements in human-computer interaction techniques: What you look at is what you get," ACM Trans. Information Systems, vol.9, no.3, pp.152–169, 1991.
[28] T. Selker, "Visual attentive interfaces," BT Technology Journal, vol.22, no.4, pp.146–150, 2004.
[29] L. Qu, N. Wang, and W.L. Johnson, "Pedagogical agents that interact with learners," AAMAS-04 Workshop on Balanced Perception and Action in ECAs, 2004.
[30] Y.I. Nakano, G. Reinstein, T. Stocky, and J. Cassell, "Towards a model of face-to-face grounding," Proc. Association for Computational Linguistics (ACL-03), pp.553–561, 2003.
[31] A. Ortony, G.L. Clore, and A. Collins, The Cognitive Structure of Emotions, Cambridge University Press, Cambridge, 1988.
[32] H. Prendinger and M. Ishizuka, "Social role awareness in animated agents," Proc. 5th International Conference on Autonomous Agents (Agents-01), pp.270–277, ACM Press, New York, 2001.
[33] A. Ortony, "Value and emotion," in Memories, thoughts, and emotions: Essays in the honor of George Mandler, ed. W. Kessen, A. Ortony, and F. Craik, pp.337–353, Erlbaum, Hillsdale, NJ, 1991.
[34] H. Prendinger, J. Mori, and M. Ishizuka, "Using human physiology to evaluate subtle expressivity of a virtual quizmaster in a mathematical game," Int. J. Human-Computer Studies, vol.62, no.2, pp.231–245, 2005.
[35] H. Prendinger, C. Ma, J. Yingzi, A. Nakasone, and M. Ishizuka,

"Understanding the effect of life-like interface agents through eye users' eye movements," Proc. Seventh International Conference on Multimodal Interfaces (ICMI-05), ACM Press, 2005.
[36] "Hottolink Inc.," URL: http://www.hottolink.com
[37] R.M. Cooper, "The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing," Cognitive Psychology, vol.6, pp.84–107, 1974.
[38] P.J. Lang, "The emotion probe: Studies of motivation and attention," American Psychologist, vol.50, no.5, pp.372–385, 1995.
[39] S. Kopp, B. Jung, N. Lessmann, and I. Wachsmuth, "Max — A multimodal assistant in virtual reality construction," KI Zeitschift (German Magazine of Artificial Intelligence), Special Issue on Embodied Conversational Agents, vol.4, pp.11–17, 2003.
[40] C. Becker, S. Kopp, and I. Wachsmuth, "Simulating the emotion dynamics of a multimodal conversational agent," Proc. Tutorial and Research Workshop on Affective Dialogue Systems (ADS-04), LNAI 3068, pp.154–165, Springer, Berlin Heidelberg, 2004.
[41] U. Spierling, D. Grasbon, N. Braun, and I. Iurgel, "Setting the scene: Playing digital director in interactive storytelling and creation," Comput. Graph., vol.26, pp.31–44, 2002.
[42] "A.L.I.C.E AI Foundation," URL: http://www.alicebot.org.
[43] H. Umemuro and J. Yamashita, "Detection of user's confusion and surprise based on pupil dilation," Japanese Journal of Ergonomics, vol.39, no.4, pp.153–161, 2003.
[44] A. Kapoor, R.W. Picard, and Y. Ivanov, "Probabilistic combination of multiple modalities to detect interest," Proc. International Conference on Pattern Recognition, pp.969–972, 2005.
[45] H. Liu, H. Lieberman, and T. Selker, "A model of textual affect sensing using real-world knowledge," Proc. International Conference on Intelligent User Interfaces (IUI-03), pp.125–132, 2003.

Helmut Prendinger is associate professor at the National Institute of Informatics. Previously, he held positions as a research associate and JSPS post-doctoral fellow at the Univ. of Tokyo. Earlier he worked as a junior specialist at the Univ. of California, Irvine. He received his M.A. and Ph.D. degrees from the Univ. of Salzburg, Dept. of Logic and Philosophy of Science and Dept. of Computer Science. His research interests include artificial intelligence, affective computing, and human–computer interaction, in which areas he has published more than 60 papers in international journals and conferences. He is a co-editor (with Mitsuru Ishizuka) of a book on Life-Like Characters that appeared in the Cognitive Technologies series of Springer.

Mitsuru Ishizuka is a professor of the Graduate School of Information Science and Technology, Univ. of Tokyo. Previously, he worked at NTT Yokosuka Laboratory and Institute of Industrial Science, Univ. of Tokyo. During 1980–81, he was a Visiting Assoc. Professor at Purdue University. He received his B.S., M.S. and Ph.D. degrees in electronic engineering from the Univ. of Tokyo. His research interests are in the areas of artificial intelligence, multimodal media with lifelike agents, and intelligent WWW information space. He is a member of IEEE, AAAI, Japanese Society for AI (currently, president), IPS Japan, etc.