Fundamentals of Verbal and Nonverbal Communication and the Biometric Issue
A. Esposito et al. (Eds.)
IOS Press, 2007

Nonverbal and Bodily Interaction in Ambient Entertainment

Anton NIJHOLT[1], Dennis REIDSMA, Zsofia RUTTKAY, Herwin van WELBERGEN, Pieter BOS
University of Twente, Center of Telematics & Information Technology, the Netherlands

Abstract. In future Ambient Intelligence environments we assume intelligence embedded in the environment and its virtual, sometimes visualized agents (virtual humans). These environments support the human inhabitants or visitors in their activities and interactions by perceiving them through their sensors. In this paper we look at our research on bodily and gestural interaction with environments equipped with simple sensors, application-dependent intelligence, and an embodied virtual agent employed in the display of reactive and pro-active activity. The virtual humans we discuss play roles such as dance partner, conductor or trainer. All of them require the perception and the generation of bodily activity and other display of nonverbal communication. The role of affect and persuasion in these ambient entertainment environments is touched upon.

Keywords. Ambient Intelligence, entertainment, nonverbal communication, embodied agents, bodily interaction, behavioral processes, affect

1. Introduction

In future Ambient Intelligence (AmI) environments we assume intelligence embedded in the environment, its objects (furniture, mobile robots) and its virtual, sometimes visualized agents (virtual humans). These environments support the human inhabitants or visitors of these environments in their activities and interactions by perceiving them through their sensors (proximity sensors, cameras, microphones, etc.). Support can be reactive, but also and more importantly, pro-active and unobtrusive, anticipating the needs of the inhabitants and visitors by sensing their behavioral signals and being aware of the context in which they act [8].

Health, recreation, sports and playing games are among the needs inhabitants and visitors of smart environments will have. Sensors in these environments can detect and interpret bodily activity and can give multimedia feedback to invite, stimulate, guide and advise on bodily activity. Rather than aiming at improving user task efficiency, in the environments we investigate the aim is to improve physical and mental health (well-being) through exercise and through play. Exercises can be done in order to improve fitness, to prevent certain injuries (e.g., RSI), or to recover from an accident (e.g., physiotherapy exercises). Other exercises may aim at improving certain capabilities related to a profession (ballet, etc.), some kind of recreation (juggling, etc.),

[1] Corresponding Author: University of Twente, Center of Telematics & Information Technology, PO Box 217, 7500 AE Enschede, the Netherlands, e-mail: A.Nijholt@ewi.utwente.nl.

or sports (fencing, etc.). Fun, just fun, achieved from interaction (e.g. dancing or physical gaming) can be another aim of such environments.

In this paper we look at our research on bodily and gestural interaction with environments equipped with some simple sensors (cameras, microphones, dance pads), some application-dependent intelligence (allowing reactive and pro-active activity), and an embodied virtual agent employed in the display of reactive and pro-active activity. The activities we look at are dance movements (both from a virtual and a human dancer), music conducting (a virtual conductor conducting human musicians) and fitness, aerobics or physiotherapy exercises (to be performed by a human trainee or patient and to monitored and demonstrated by a virtual trainer).

The organization of this paper is as follows. In section 2 we look at application domains of our research on nonverbal and bodily interaction in smart environments. Section 3 is devoted to some of our virtual agents performing in these application domains. We show both practical and theoretical design considerations. In section 4 we look at shortcomings of our applications and topics of future research.

2. Bodily Interaction: Entertainment, Health, and Others

Entertainment, health, sports, and leisure applications using information and communication technology often require and encourage physical body movements and often applications are designed for that reason. In our research we look at bodily and gestural interaction with game and leisure environments that are equipped with sensors (cameras, microphones, touch, and proximity sensors) and application-dependent intelligence (allowing reactive and proactive activity). Interpretation of the bodily interaction, requiring domain-dependent artificial intelligence, needs to be done by the environment and the agents that maintain the interaction with the human partner. In the display of reactive and pro-active activity embodied virtual agents play an important role. Virtual agents can play the role of teacher, coach, partner or buddy. One underlying assumption is that emphasis on activities in which the experience rather than the result will guide the design of social and intelligent systems that will become part of ambient intelligence home environments [7].

3. Inviting and Maintaining Interaction by Virtual Humans

Three applications have been designed in which our ideas about nonverbal and bodily interaction have been implemented. The implementations are available, but they are certainly not final. We will have a look at a virtual dancer that invites a visitor to her environment to dance with her, a conductor that guides musicians in its environment to play according the score designed by a composer, and a virtual trainer (e.g. in the role of fitness trainer or physiotherapist) that knows about exercises that need to be performed by a user or patient. In all these applications there is a continuous interaction between embodied agent and its human partner. Moreover, rather than have the more traditional verbal interaction supported by nonverbal communication, here the main interaction that takes place is nonverbal, and speech and language, when present at all, take the supporting role. External signals like music being played can also have a role in addition to the multimodal communication.

3.1 A Virtual Dancer

Our Virtual Dancer [10] is an interactive dancing agent that dances together with the user to the beat of the music (cf. [13] for related work). The dancer adapts its performance to whatever the human user is doing. The moves of the virtual dancer, to be chosen from a database obtained from motion-capturing, are aligned to the beats of the music. These beats are detected real-time. The system observes the movements of the human dancer using computer vision software. The system extracts global characteristics about the movements of the human dancer, such how much (s)he moves around or how much (s)he waves with the arms. Thee characteristics are used to select moves from the database that are in some way "appropriate" to the dancing style of the human dancer. Finally, there is a dance pad that registers feet activity. By alternating patterns of following the user with taking the lead with new moves, the system attempts to achieve a mutual dancing interaction where both human and virtual dancer influence each other. Finding the appropriate nonverbal interaction patterns that allow us to have a system that establishes rapport with its visitors is one of the longer term issues that is being addressed in this research.

3.2 A Virtual Conductor

We have designed and implemented a virtual conductor [2] that is capable of leading, and reacting to, live musicians in real time. The conductor possesses knowledge of the music to be conducted, and it is able to translate this knowledge to gestures and to produce these gestures. The conductor extracts features from the music and reacts to them, based on information of the knowledge of the score. The reactions are tailored to elicit the desired response from the musicians.

Clearly, if an ensemble is playing too slow or too fast, a (human) conductor should lead them back to the correct tempo. She can choose to lead strictly or more leniently, but completely ignoring the musicians' tempo and conducting like a metronome set at the right tempo will not work. A conductor must incorporate some sense of the actual tempo at which the musicians play in her conducting, or else she will lose control. If the musicians play too slowly, the virtual conductor will conduct a little bit faster than they are playing. When the musicians follow him, he will conduct faster yet, till the correct tempo is reached again.

The input of the virtual conductor consists of the audio from the human musicians. From this input volume and tempo are detected. These features are evaluated against the original score (currently stored in MIDI) to determine the conducting style (lead, follow, dynamic indications, required corrective feedback to musicians, etc) and then the appropriate conducting movements of the virtual conductor are generated. Computer vision has not yet been added to the system. That is, musicians can only interact with the conductor through their music. In a future implementation we can look at the possibility to have the conducting behavior directed to (the location of) one or more particular instruments and their players.

3.3 A Virtual Trainer

A Virtual Trainer is a virtual human capable of presenting physical exercises that are to be performed by a user, monitoring the user's performance and providing feedback accordingly at different levels [1,3]. Our virtual trainer [11] fulfills most of the functions of a real trainer: not only performs the exercises to be followed, provides professionally and psychologically sound, human-like coaching. Depending on the motivation and the application context, the exercises may be general ones of fitness to improve the user's physical condition, special exercises to be performed from time to time during work to prevent for example RSI, or physiotherapy exercises with medical indications.

Currently, the trainer is in its design phase, but parts of its intended behavior are already implemented. In the design the focus is on the reactivity of the trainer, manifested in natural language comments on readjusting the tempo, pointing out mistakes or rescheduling the exercises. When choosing how to react, the static and dynamic characteristics of the user and the objectives to be achieved are taken into account, and evaluated with respect to biomechanical knowledge and psychological considerations of real experts. Hence if the user is just slowing down, the trainer will urge him in a friendly way to keep up with the tempo, and acknowledge with cheerful feedback good performance and introduce small talk every now and then to keep the user motivated.

Figure 1. Virtual dancer, virtual conductor, and virtual trainer.

3.4 Aligning Interaction with Goals and Events

In Figure 1 we have illustrated our three applications. In our applications there is a strong alignment or integration to external channels or events over which our embodied agents and the other partners in the interaction hardly have control. For example, whatever the bodily and nonverbal interaction is that takes place between virtual dancer and human dancer, the virtual dancer is programmed to follow the music and the human dancer is assumed to cooperate. Hence, their dance movements focus on the alignment of the dance behavior to the music. Similar observations can be made about the virtual conductor and the virtual trainer or therapist. The conductor has the score of the music as basis of his conducting behavior and its behavior aims at making the musicians follow the music. Our therapist has not yet been fully implemented, but one of the aims is to have its gestures, body movements, and verbal expressions aligned with music that supports the fitness, aerobics and physiotherapy exercises it offers to its human partners. It should be mentioned that similar kinds of alignment need to be observed in every application where we have rules and constraints on interaction, not only in dancing, conducting and training applications, but also in sports (tennis, fencing, chess) and in dialogues or conversations. For example, also in a dialogue there are constraints, there are dialogue dynamics and there often is an aim that guides the progress of the dialogue and the conversational partners align their conversation (and turn taking) behavior to conversational principles and progress towards a joint goal or towards individual goals.

To define the alignment of the multimodal interaction channels (like speech, gesture, motion and sounds) with each other and events like the beat of the music, we make use of our own extension of the multimodal synchronization language BML [6]. To make behavior, like body movement, fit external events, we might have to stretch or skew its timing. Currently we are researching which parameters can be used to speed up and slow down movement [12]. We are analyzing captured motion of humans to develop models of modifying movements. These models will be used to drive a virtual human, and the resulting motion will be evaluated by user tests. We made use of a time warping technique similar to [14] to align dance or conducting movement to the beat of music. However, this technique allows only relatively small variations in timing. Movements that are played back too slow look like a slow-motion movie. Playing motion too fast produces physically impossible movement. Our preliminary motion capture analysis indicates that amplitude reduction (in our test defined as the maximum amount of distance between the hands in a clapping motion) of a movement can be used as a speed up mechanism.

4. Shortcomings: Intelligence, Social Influence and Persuasion

A serious shortcoming of our agents (conductor, physiotherapist, and dancer) is an explicit modeling of their intelligence and their reasoning behavior. We certainly need to look at BDI&E (Beliefs, Desires, Intentions & Emotions) agents in order to model both the aims and the knowledge of our virtual humans as well as their ability to plan interaction behavior. This planning requires reasoning capabilities that take into account the role of emotions and persuasive strategies. The latter aim at inducing changes in actions and behavior of the partners of the virtual humans.

It is well-known that people respond to computers as if they are social actors [9]. This is even truer if the computer is hidden behind a virtual human that has been designed to be a social actor in its interactions with a human partner. Such a virtual human can exercise social influence and therefore it can persuade and motivate.

Persuasion plays a role in our applications. We do want interaction between trainer and user, between conductor and musicians and between virtual and human dancer. For example, the virtual dancer is not really subtle in inviting you to dance with her and she is really disappointed when you leave. In [4] a comprehensive survey of persuasive technologies is presented. Persuasive agents are discussed in [5]. It is certainly useful, but the length of this paper does not allow this, to look at our applications from the point of view of persuasion. That is, analyze our designs and implementations, hopefully including evaluations, and redesign them from the point of view of computers as social actors that can persuade and from the point of view of persuasive environments where we can simulate desired, expected and inspiring behavior with the aim to improve physical performance or enhance entertaining experience.

References

[1] S. Babu et al. Virtual Human Physiotherapist Framework for Personalized Training and Rehabilitation. Proceedings Graphics Interface 2005, Victoria, Canada, 2005.

[2] P. Bos, D. Reidsma, Z. Ruttkay & A. Nijholt. Interacting with a Virtual Conductor. Proceedings International Conference on Entertainment Computing (ICEC 2006), Cambridge, UK, LNCS 4161, Springer, 2006, 25-30.
[3] J.W. Davis & A.F. Bobick. Virtual PAT: A Virtual Personal Aerobics Trainer. MIT Media Laboratory, TR 436, 1998.
[4] B.J. Fogg. *Persuasive Technology.* Morgan Kaufmann Publishers, San Francisco, CA, 2003.
[5] M. Guerini. Persuasion Models for Multimodal Message Generation. Ph.D. thesis, University of Trento, Trento, Italy, March 31, 2006.
[6] S. Kopp, B. Krenn, S. Marsella, A.N. Marshall, C. Pelachaud, H. Pirker, K.R. Thórisson & H.H. Vilhjálmsson. Towards a Common Framework for Multimodal Generation: The Behavior Markup Language. Proceedings Intelligent Virtual Agents (IVA 2006), Marina del Rey, USA, LNAI 4133, Springer, 2006, 205-217.
[7] A. Nijholt. Google Home: Experience, Support and Re-experience of Social Home Activities. Information Sciences. Elsevier, Amsterdam, 2007, to appear.
[8] M. Pantic, A. Pentland, A. Nijholt & T. Huang. Human Computing and Machine Understanding of Human Behavior: A Survey. ACM SIGCHI Proceedings of the Eight International Conference on Multimodal Interfaces (ICMI'06), Banff, Alberta, Canada, ACM, New York, NY, 239-248.
[9] B. Reeves & C. Nass. *The Media Equation.* Cambridge University Press, Cambridge, MA., 1996.
[10] D. Reidsma, H. van Welbergen, R. Poppe, P. Bos, & A. Nijholt. Towards Bi-directional Dancing Interaction. Proceedings International Conference on Entertainment Computing (ICEC 2006), Cambridge, UK, LNCS 4161, Springer, 2006, 1–12.
[11] Z. Ruttkay, J. Zwiers, H. Welbergen van, & D. Reidsma. Towards a Reactive Virtual Trainer. Proceedings Intelligent Virtual Agents (IVA 2006), Marina del Rey, USA, LNAI 4133, Springer, 2006, 292-303.
[12] Z. Ruttkay & H. van Welbergen. On the timing of gestures of a Virtual Trainer. Central European Multimedia and Virtual Reality Conference 2006, 6-8 November 2006, Eger, Hungary.
[13] T. Shiratori, A. Nakazawa & K. Ikeuchi. Dancing-to-Music Character Animation. EUROGRAPHICS 2006, E. Gröller and L. Szirmay-Kalos (Guest Editors), Volume 25 (2006), Number 3, to appear.
[14] A. Witkin & Z. Popovic. Motion warping. Proceedings of the 22nd annual conference on Computer Graphics and Interactive Techniques (SIGGRAPH '95), ACM Press, 1995, 105-108.