

Leading and following with a virtual trainer

Dennis Reidsma
Human Media Interaction
PO Box 217
7500AE Enschede
Netherlands
d.reidsma@utwente.nl

Eike Dehling
Human Media Interaction
PO Box 217
7500AE Enschede
Netherlands

Herwin van Welbergen
Human Media Interaction
PO Box 217
7500AE Enschede
Netherlands
h.vanwelbergen@utwente.nl

Job Zwiers
Human Media Interaction
PO Box 217
7500AE Enschede
Netherlands
j.zwiers@utwente.nl

Anton Nijholt
Human Media Interaction
PO Box 217
7500AE Enschede
Netherlands
a.nijholt@utwente.nl

ABSTRACT

This paper describes experiments with a virtual fitness trainer capable of mutually coordinated interaction. The virtual human co-exercises along with the user, leading as well as following in tempo, to motivate the user and to influence the speed with which the user performs the exercises. In a series of three experiments (20 participants in total) we attempted to influence the users' performance by manipulating the (timing of the) exercise behavior of the virtual trainer. The results show that it is possible to do this implicitly, using only micro adjustments to its bodily behavior. As such, the system is a first step in the direction of mutually coordinated bodily interaction for virtual humans.

Categories and Subject Descriptors

H5.m. [Information interfaces and presentation (e.g., HCI)]: Miscellaneous

Keywords

Virtual Human, Embodied Conversational Agent, Continuous Interaction, Virtual Trainer, Coordinated Interaction

1. INTRODUCTION

This paper concerns our experiments with a virtual human that is capable of coordinated interaction. The virtual human, a fitness trainer, supervises a user who is doing fitness exercises. The virtual human co-exercises along with the user, leading as well as following in tempo, to motivate the user and to influence the speed with which the user performs the exercises. This paper starts with a discussion of background, motivation and related work, and a short description the architecture of the system. After that, the paper

describes a series of three experiments (20 participants in total) in which we attempted to influence the users' performance by manipulating the exercise behavior of the virtual trainer.

2. BACKGROUND AND MOTIVATION

Interaction between humans, be it talking, negotiating, dancing, rowing a boat or reading poetry to someone, is a joint activity [3]. A major requirement for successfully engaging in joint activities is effective coordination [7]. This coordination may be, e.g., on the semantic level, on the level of the content and behavior selection in the interaction, or even on low level aspects such as the temporal dynamics of the interaction (see, e.g., [8] and [15] for a discussion of the various levels on which this coordination takes place).

Until recently, coordination between humans and computers was a far more simple matter than the complex social processes governing coordination between humans. Computer applications were built for information transfer and processing. Information goes in; processing occurs; information comes out: that was basically the definition of interaction with a computer. Nowadays, through realization of the vision of Ambient Intelligence, our daily environment is instrumented with sensors, interpreters, displays, and computational devices. The environment may observe and support us, but may also pro-actively join in the interactions occurring in it, for example through virtual humans and social robots. In such an environment, the boundaries between human-human interaction and human-machine interaction blur. We get interaction between multiple social actors, some of which may be humans and some of which may be computers.

In this context, there are two ways in which coordinated interaction becomes relevant to human computer interaction. Firstly, ambient intelligence environments are built to detect and interpret as much as possible about the activities of their inhabitants. Interaction with a computer moves from mouse-and-keyboard interaction towards whole body interaction, in which all behavior from humans is potential input to the system [5]. In these environments, mutual coordination between human inhabitants is simply one more

aspect of behavior to be detected and interpreted [12]. Secondly, when the computer itself becomes a social actor [9], it should also be able to employ mechanisms of coordinated interaction similar to what humans do. Interaction with a computer in the form of a social robot or virtual human becomes a joint activity, too, and requires the same types of coordination on the same levels.

This paper describes experiments with a virtual fitness trainer that should behave as much as possible like a human trainer. A good human trainer does not always use words to say “please move faster”, or “move in exactly this way”. Sometimes they lead their pupil by moving together with them. By co-performing the exercise they show the proper form of the movements. By slowly speeding up (or slowing down) in such a way as to lead the pupil along, they can make the pupil move faster (or slower) without any spoken instructions. Such implicit manipulation of a pupil’s behavior can be much more effective, more enjoyable, and less face-threatening, than using verbal comments, criticism, and instructions. However, such leading-and-following behavior is not trivial to implement in a virtual human. It requires appropriate perception of the user’s behavior and coordinated generation of the virtual human’s behavior [10]. Our goal, reported on in this paper, was to experiment with influencing the behavior of the user implicitly, through implementing nonverbal mechanisms for coordinated bodily interaction.

3. RELATED WORK

Davis and Bobick combined computer vision technology with recordings of feedback by a real coach [4]. The focus of their system was on the recognition technology and on the choice of appropriate feedback utterances. Babu et al. developed a virtual trainer that can demonstrate exercises, describe and show the user’s mistakes, and praise correct execution of exercises [1]. For perception, they used computer vision to track markers in 3D space. Their aim was to give high quality feedback on exercises, using the system as a virtual physiotherapist. Chua et al. performed experiments to see how effectively users learn Tai-Chi movements when they can see one or more virtual Tai-Chi trainers from different angles [2]. Their trainer did not give feedback on the performance of the user. IJsselsteijn et al. compared training situations with and without a virtual coach that gives feedback, varying the immersiveness of the experience, and evaluated motivational factors, showing that a more immersive application can enhance user performance [6]. Ruttkay et al. collected a corpus of recordings of people performing fitness exercises, and analysed the corpus for the appropriate feedback that should be given for each recording [14]. When giving corrective feedback, these earlier virtual trainers tended to be focused on verbal explanations and ‘demonstrative’ performance of the exercise. The novelty of our work is that we look at how a virtual trainer can *implicitly* influence the *tempo* with which a user performs the exercises using only coordinated bodily interaction. In this, we build upon earlier work with a virtual orchestra conductor that guided human musicians through the performance of a musical piece [11], the main difference being that in the current work we performed structured experiments to investigate the relationship between the users’ performance and the bodily behavior of the trainer.

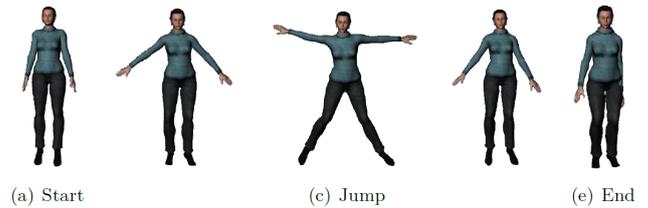


Figure 1: One of the exercise animations used in the virtual trainer

4. ARCHITECTURE

Our virtual trainer was implemented in Elckerlyc, a BML compliant behavior realizer for generating multimodal verbal and nonverbal behavior for virtual humans [16]. We recorded animations for a few (very basic) fitness exercises, such as jumping jacks, squats and side steps, that the trainer should perform together with the user. The performance of the user was measured in two ways: using an *HxM Zephyr Heartrate Monitor*TM to measure the intensity with which the user was performing the exercise, and a *Nintendo Wii Remote*TM to measure the rhythm (tempo) of the performance. The latter worked by taking the average rhythm of the most recent peaks in the acceleration of the Wii Remote. This method served well because the exercises have rhythmic phases (jump, land, extend arms, etc) that produce clearly detectable peaks.

By default, the trainer and the user perform the exercise at the same tempo. By monitoring the heart rate the system determines whether the user is working hard enough. If the heart rate is too low, or too high, the trainer attempts to lead the user to a new tempo by modifying his own performance. To achieve the required modifications, we use Elckerlyc’s flexible plan representation that allows on-the-fly micro adjustments to the content and timing of the virtual human’s behavior [13].

5. EVALUATION

We evaluated the system in several user experiments aimed to find out whether we could influence the speed of the user, and whether this could be done implicitly, i.e., without written or spoken instructions. In this section, we summarize the setup and the results.

We experimented with three different scenarios. In the first scenario, the users were told they had to follow the tempo of the trainer, and the trainer tried to increase the tempo. During the second exercise, the trainer increased or decreased the tempo of its performance to attempt to keep the heart rate of the user inside a target range. During the third scenario, the trainer again tried to increase the tempo, but the participants in this scenario were not told that they had to follow the tempo of the trainer, or that it would change during the exercise.

Twenty subjects participated in the study, about half of them male and half of them female. Most did not have experience with fitness exercises specifically, but did have experience with other sports or with dancing. During the exercise, we recorded the user’s heart rate, the tempo with

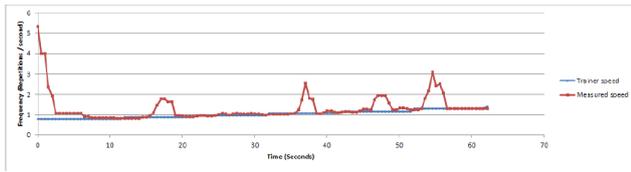


Figure 2: The trainer increases the tempo; the user follows the tempo, as they were told.

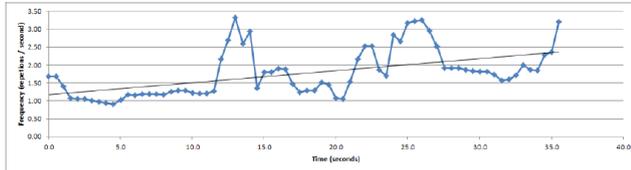


Figure 3: Increasing the tempo every 5 seconds by 5% did not give the user enough time to settle comfortably in the new tempo, and caused wild fluctuations in the user’s tempo.

which the user performed the exercise, and the tempo of the trainer. In the third scenario, we also made video recordings of some of the sessions.

Figure 2 shows the tempo changes of one user for the first scenario, in which the trainer attempted to increase the tempo. After a short warm up period, the trainer raised the tempo by 10% every 10 seconds. After an initial phase in which the user needs to find their base rhythm, they are clearly able to follow the trainer’s tempo.¹ As Figure 3 shows, raising the tempo by 5% every 5 seconds had a drastically different effect.

Figure 4 shows results for a participant in the second scenario, in which the trainer attempted to keep the heart rate of the user inside a target range. After the short warm up period, the trainer would increase or decrease its own tempo by 10% every 10 seconds, if the heart rate of the user warranted it. This experiment achieved the hoped-for results for only two of the six participants who participated in this session; Figure 4 shows one of these two. For the other four participants, the trainer was not even able to get the heart rate of the user inside the target range in the first place.

The third scenario, tested with subjects who did not participate in the other two scenarios, was aimed at finding out whether the influence of the trainer on the users’ tempo would also be present when the users were not told beforehand that the tempo of the trainer would change, or that they were to follow that tempo. The exercise in this scenario started with the trainer saying “jump” for 8 counts. This (implicitly) gave the user a rhythm, linked to the motion. After 8 repetitions, the trainer stopped saying “jump”, but continued the motions in the same tempo for another 8 rep-

¹Visual inspection of the video recordings made during the experiment showed that, although some of the “error peaks” in the graph were caused by the user temporarily losing the tempo, some of them were in fact caused by the tempo measurements erroneously measuring double the actual tempo.

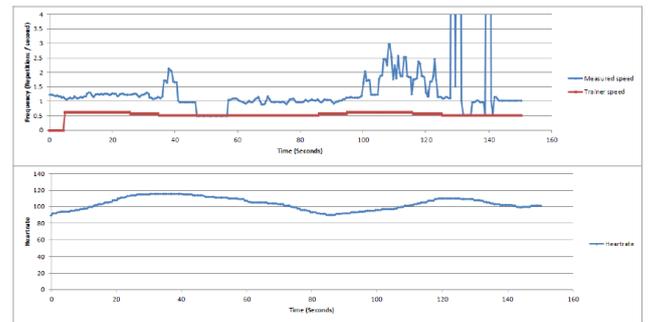


Figure 4: For two of the six participants in the second experiment, the trainer was able to keep the user’s heart rate inside the target range. The graphs show trainer’s tempo, measured user’s tempo, and user’s heartrate, for one of these two users. Goal heartrate was 100. Heartrate rises and falls out of range, but this is corrected implicitly by speed adjustments to the trainers movements.

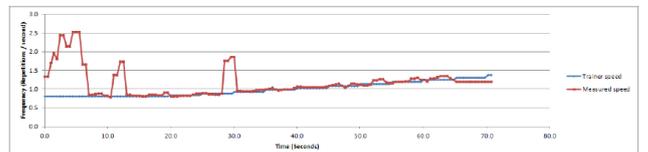


Figure 5: In the third scenario, this user follows the trainer’s tempo very well, even without receiving instructions either beforehand (from the experimenter) or during the exercise (from the virtual trainer).

etitions. Subsequently, the trainer started to (attempt to) influence the user’s tempo. Figure 5 shows how the trainer was able to speed up the performance of the user *implicitly*, *without giving verbal instructions*, solely by making micro adjustments to its own bodily behavior.

In summary, we can say that we have achieved our goal of implementing real coordinated interaction between a virtual human and a user. Building upon these results we can start to create more human like virtual trainers that use more subtle types of feedback.

6. DISCUSSION AND CONCLUSION

Mutually coordinated interaction between humans and virtual humans is a challenging topic that requires new developments in user perception as well as behavior generation for virtual humans. At the same time, such coordinated interaction is an important prerequisite for building virtual humans that are really capable of entering into joint activities with human users. The fitness exercises and sports training scenarios described in this paper are not the only application in which such interaction plays an important role. People sit down *together*, after shaking hands, when they have a meeting. When the meeting is over, likely as not they will indicate they are about to get up and go away, and then proceed to do so in a coordinated fashion. When in a conversation, the breathing of two persons may synchro-

nize; their body sway may display similar rhythms, and in many other ways they will exhibit embodied entrainment. We discussed in this paper our experiments with a virtual trainer that was able to influence the users' performance of a fitness exercise implicitly, using only micro adjustments to its bodily behavior. As such, this system is a first step in the direction of mutually coordinated bodily interaction for virtual humans.

7. ACKNOWLEDGMENTS

This research has been supported by the GATE project, funded by the Dutch Organization for Scientific Research (NWO). The authors would like to thank the participants in the experiments for their participation as well as for their helpful comments afterwards.

References

- [1] S. Babu, C. Zambaka, J. Jackson, T. Chung, B. Lok, M. C. Shin, and L. F. Hodges. Virtual human physiotherapist framework for personalized training and rehabilitation. In *Proc. Graphics Interface 2005*, May 2005.
- [2] P. T. Chua, R. Crivella, B. Daly, N. Hu, R. Schaaf, D. Ventura, T. Camill, J. Hodgins, and R. Pausch. Training for physical tasks in virtual environments: Tai chi. In *VR '03: Proceedings of the IEEE Virtual Reality*, page 87. IEEE Computer Society, 2003.
- [3] H. H. Clark. *Using Language*. Cambridge University Press, 1996.
- [4] J. W. Davis and A. F. Bobick. Virtual PAT: A virtual personal aerobics trainer. In *Proceedings of Workshop on Perceptual User Interfaces (PUI'98)*, pages 13–18, New York, 1998. IEEE.
- [5] D. England, editor. *Whole Body Interaction*. Human-Computer Interaction Series. Springer-Verlag, London, May 2011.
- [6] W. IJsselsteijn, Y. de Kort, J. H. D. M. Westerink, M. de Jager, and R. Bonants. Fun and sports: Enhancing the home fitness experience. In M. Rauterberg, editor, *Third International Conference on Entertainment Computing*, volume 3166 of *Lecture Notes in Computer Science*, pages 46–56. Springer, 2004.
- [7] G. Klein, P. J. Feltovich, J. M. Bradshaw, and D. D. Woods. *Common Ground and Coordination in Joint Activity*, chapter 6, pages 139–178. Wiley Series in Systems Engineering and Management. John Wiley and sons, Hoboken, New Jersey, 2004.
- [8] S. Kopp. Social resonance and embodied coordination in face-to-face conversation with artificial interlocutors. *Speech Communication*, 52(6):587–597, 2010.
- [9] C. Nass, J. Steuer, and E. R. Tauber. Computers are social actors. In *Proceedings of the SIGCHI conference on Human factors in computing systems: celebrating interdependence*, CHI '94, pages 72–78, New York, NY, USA, 1994. ACM.
- [10] A. Nijholt, D. Reidsma, H. van Welbergen, H. J. A. op den Akker, and Z. M. Ruttkay. Mutually coordinated anticipatory multimodal interaction. In *Nonverbal Features of Human-Human and Human-Machine Interaction*, volume 5042 of *LNCS*, pages 70–89, Berlin, 2008. Springer Verlag.
- [11] D. Reidsma, A. Nijholt, and P. Bos. Temporal interaction between an artificial orchestra conductor and human musicians. *Computers in Entertainment*, 6(4):1–22, 2008.
- [12] D. Reidsma, A. Nijholt, W. Tschacher, and F. Ramseyer. Measuring multimodal synchrony for human-computer interaction. In D. England, J. Sheridan, and B. Crane, editors, *Proceedings of the CHI Workshop on Whole Body Interaction*, 2010.
- [13] D. Reidsma, H. v. Welbergen, and J. Zwiers. Multimodal plan representation for adaptable BML scheduling. In *Proc. of IVA 2011*, 2011.
- [14] Z. M. Ruttkay and H. van Welbergen. Elbows higher! Performing, observing and correcting exercises by a virtual trainer. In H. Prendinger, J. C. Lester, and M. Ishizuka, editors, *Proceedings of the 8th International Conference on Intelligent Virtual Agents*, volume 5208 of *Lecture Notes in Artificial Intelligence*, pages 409–416, Berlin, Heidelberg, 2008. Springer-Verlag.
- [15] K. R. Thórisson. Natural turn-taking needs no manual: Computational theory and model, from perception to action. In B. Granström, D. House, and I. Karlsson, editors, *Multimodality in Language and Speech Systems*, pages 173–207. Kluwer Academic Publishers, Dordrecht, The Netherlands, 2002.
- [16] H. van Welbergen, D. Reidsma, Z. M. Ruttkay, and J. Zwiers. Elckerlyc: A BML realizer for continuous, multimodal interaction with a virtual human. *Journal on Multimodal User Interfaces*, 3(4):271–284, 2010.