Elbows higher! Performing, observing and correcting exercises by a Virtual Trainer

Zsófia Ruttkay, Herwin van Welbergen

HMI, Dept. of CS, University of Twente,
POBox 217, 7500AE Enschede, The Netherlands
{z.m.ruttkay, H.vanWelbergen}@ewi.utwente.nl
http://hmi.ewi.utwente.nl

Abstract. In the framework of our Reactive Virtual Trainer (RVT) project, we are developing an Intelligent Virtual Agent (IVA) capable to act similarly to a real trainer. Besides presenting the physical exercises to be performed, she keeps an eye on the user. She provides feedback whenever appropriate, to introduce and structure the exercises, to make sure that the exercises are performed correctly, and also to motivate the user. In this paper we talk about the corpora we collected, serving a basis to model repetitive exercises on high level. Then we discuss in detail how the actual performance of the user is compared to what he should be doing, what strategy is used to provide feedback, and how it is generated. We provide preliminary feedback from users and outline further work.

1 Introduction

In the framework on our ongoing project [11], we aim at developing a Reactive Virtual Trainer (RVT), an intelligent virtual agent (IVA) with the rich functionality of human trainers. We are aiming at a RVT who can act as a stand-in for a real trainer, being capable of:

1. 'Understanding' and *intelligently dealing with a big variety of exercise* tasks, specified by a human expert in a near-natural high-level language.
2. Introducing and *explaining the exercise* to be performed.
3. *Presenting the exercises* in different tempi, accompanied by
   a. rhythmic counting
   b. and/or other acoustics such as music.
4. *Monitoring the user's motion*, and *correcting inaccuracies* in tempo or formation of the exercises.
5. *Monitoring the user's general physical and emotional state,* acknowledge it and possibly adjust the exercises to be followed.
6. *Providing emphatic or other feedback* to keep the user motivated.
7. Keeping track of past sessions of the user, and *evaluating current performance* in that context.

8. *Adjusting his/her own appearance* and, if appropriate, *biomechanical characteristics* to those of the user, in order to get closer to him psychologically and in the intensity and quality of the motions presented.
9. Use *conversation style and feedback strategy* adjusted to the age and personality of the user.

Our envisioned RVT may be used in different scenarios, where a real trainer cannot be present because of financial or other reasons: at home to motivate and supervise regular fitness or rehabilitation training, at work to insist on exercises best suited for people to prevent work-related injuries, see [14] for more.

The above applications are very timely, seeing the increase in age and the 'white collar' jobs in Western societies [5, 18]. A Tai Chi exercise performer was developed to demonstrate exercises using unaltered samples from a motion capture samples DB [4]. Medical and psychological consultancy applications with an empathic IVA have proven to be a success [3, 8]. Different sensors and portable devices [2] are in the focus of industry to be used to augment physical training. Also, currently Wii has been used as instrument to do sport game or other physical training [21].

Babu et al [1] presented the Virtual Human Physiotherapist Framework. In their short account they concentrate on exercise-specific monitoring, based on 3d colour markers placed on the body of the trainee. They provided a proof-of-concept demo of the monitoring mechanism, but have not covered the related issues in the breadths as we do.

The virtual Fitness Trainer of Philips, a companion projected onto an immersive screen in front of the trainee exercising on a home-trainer, evaluates his performance based on heart rate feedback [6, 7, 20]. We expect that in our case, where the RVT has richer (as of types of feedbacks, see above list) and more natural contact (based on visual perception input and subtle feedback strategy) with the trainee, the positive effect will be more significant than what the Philips researchers have reported so far [20].

The idea of vision-based physiotherapy was raised [15]. Sony's successful EyeToy: Kinetic 'game' [16], offers personalised as well as ready-made sequences of fitness exercises, commented on by one of two virtual trainers. In our application we provide more active and situated feedback based on more subtle monitoring of the user.

In a previous paper [14] we explained the above goals and the related technological challenges in depth, and gave account on the first chunk of work we accomplished. Namely, we showed how our animation engine could be used, together with music beat detection, for the major task 3, and 3b. We also described a low-level XML-based scripting language which can serve as a bridge between the authoring language and the animation engine.

In this paper we report on new work since then, contributing to goals 1, 2, 3a and particularly, 4. In the next section, we discuss our two different corpora we are building up as a reference for designing the motion and communication repertoire for the RVT. Then we outline the architecture of the RVT, and go into details of monitoring the performance of the user and providing feedback. Finally, we account on feedback from users, and outline further work.

2 Modeling physical exercises

In order to design and implement a RVT who resembles in his motion and interaction to a real expert trainer, we need to study practices from real life. For our work, we have established two kinds of corpora. The first one contains exercises performed as a trainee would do, while the second one contains recordings of sessions with a trainer and a trainee. The two corpora are complementary, providing information on motion characteristics and on training sessions and interaction strategy, respectively

2.1 The Exercise Recordings corpus

The Exercise Recording (ER) corpus contains motion-capture + video recording of physical exercises, performed by healthy people or by professional fitness trainers. We selected the persons with the goal of getting the exercises performed correctly. A selected exercise is recorded in different tempi, some by several different performers. In case of expert trainers performing an exercise, we also recorded 'typically wrong' performances of the exercise in question.

We made recordings by using passive optical mocap method in the Vicon 460 optical motion capture laboratory of our university. 41 small white marker balls about 1 cm of diameter are fixed to the body of the user, basically to track the motion of the limb joints, hands, torso and head (See Figure 1.a). Six infra-red cameras trace the light reflected from the markers, and 3d coordinates of each marker are provided automatically, with a 120 Hz frequency, and written to a file for further (off-line) processing.

We have a body of 25 minutes of recordings, of 10 exercises, performed, altogether 101 times in different variations, by 3 subjects. Cardiovascular exercises are to improve general (heart) condition, strengthening ones to develop certain muscles, and stretching ones to warm up and relax before/ after sessions.

This corpus is used to analyze the motion characteristics of exercises, individual variations and correlations between parameters like time and intensity, see further in [17]. We are currently working on making parameterized models for the motion of the 'major joints involved', based on the mocap samples. We wish to use a more advanced, multi-layered animation engine where parameterized procedural motion is layered with physical balancing to improve on quality of the motion performed by the RVT, see [19].

The other type of information on the motion characteristics is to gather knowledge for modeling the 'natural', 'fast' and 'slow' tempi for different exercises. Finally, by analyzing 'wrong' performances, we plan to improve our monitoring and diagnosis process (to be discussed in Section 3).

2.2 The Training Recordings Corpus

The Training Recordings (TR) contains (audio-)video recordings of training sessions. We aim at collecting a big body of recordings systematically, where the trainer, the client and the type of exercises are carefully chosen to end up with a rich collection with samples for all aspects. We also have some exercises from the TR session which were 'performed' (albeit by different people) in the ER corpus.

Currently we have 75 minutes of recordings of 4 complete exercise sequences covering a variety of goal, performed in different postures (standing, seated, …), with the same trainee and a professional physiotherapist trainer.

We annotate the videos by using the ELAN tool [22]. We analyze the video and audio with two goals:

- to learn about the structure of training sessions and the individual exercises, and
- to spot and categorize different types of feedbacks.

See [9] for more on preliminary results.

a)

b)

Fig 1. a) A frame from the MR mocap corpus, with the subject wearing reflective markers in the Vicon lab. b) A frame from the TR video corpus, taken in a real gym hall. The trainer on the right addresses the subject pointing out a correction, while performing the exercise too.

3 The architecture of the RVT

The RVT serves two functions:

- to *author* an exercise sequence;
- to *'make the trainee perform'* this sequence.

The first function is assured by a simple editing tool, which allows compilation of the exercise to be performed Also, the tempo is specified.

The second task is achieved by introducing the exercise, starting it once the user is ready, and then performing a monitoring-performing loop, where the 'performance' involves both spoken feedback (if appropriate) and the presentation of the exercise by the RVT. Figure 2 shows the major modules of the architecture of the RVT.

3.1 Monitoring the user

We designed a single-camera vision system based on ParleVision components [10], which can identify 'natural' colour-coded markers the user is wearing (gloves and socks). The *Feature Extractor* uses traditional image processing techniques to compute centre of mass for each marker as current location. Left and right limbs are identified also on the basis of the model of the human body and a short history of the tracking. The current location of the markers is normalised, with respect to the body geometry of the virtual trainer. The *Feature Interpreter* makes the diagnosis about the user's performance, knowing of the (normalized) marker positions. Based on the relative time in performing the exercise, and the structure and tempo of the exercise to be performed, and having access to the motion definition of the simple exercises, the Feature Interpreter generates an *allowed location* for each marker (see Figure 3). Based on comparing the actual and allowed location of the marker, the motion of the limb of the trainee is diagnosed as 'correct', 'slow' or 'fast' or 'wrong'.

3.2 Providing feedback

The diagnosis provided by the Feature Interpreter about the current movement is evaluated with respect to the recent history of the performance. Based on declared feedback policies, the outcome concerning *performance feedback* can be:

- not to say anything and go on,
- confirm that the trainee is doing well (e.g. has corrected a previous error),
- to provide warning and correcting feedback, concerning tempo or formation;
- readjust own tempo, to be in sync with the trainee;
- to abandon the exercise.

Another source of feedback is the *exercise-related feedback*, on structuring the exercises (e.g. last time an exercise is to be performed, half time reached).

The actual feedback is decided on the basis of the events triggered by the performance of the 4 limbs and the relative time of the exercise. *Conflicts* are solved by giving higher priority to recently realized change of tempo of the trainer, and novel anomalies diagnosed with limbs.

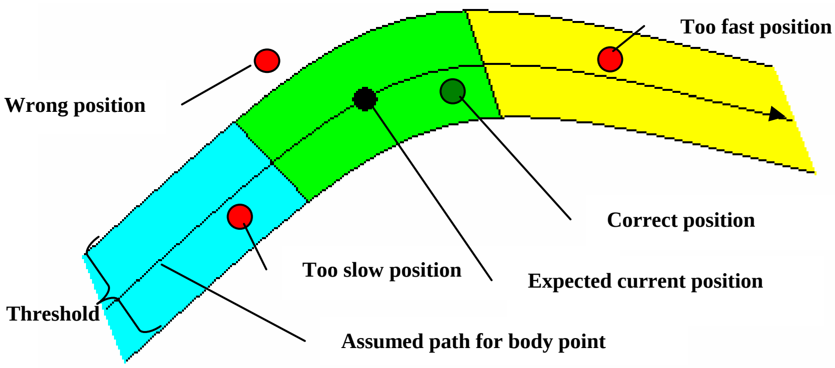

Fig 3. Assumed motion path of a body point, with expected, early, late and wrong positions

3.3 Presenting on time, in sync

The *Synchronization* module receives eventual tempo adjustment instruction for the virtual trainer, and a list of prioritized and time-stamped messages to be conveyed for the user. For different language variants one is selected randomly. The utterances are aligned to beat, or adjusted also to motion tempo (e.g. counting down to indicate the end of the exercise), and the motion of the virtual trainer is generated [13].

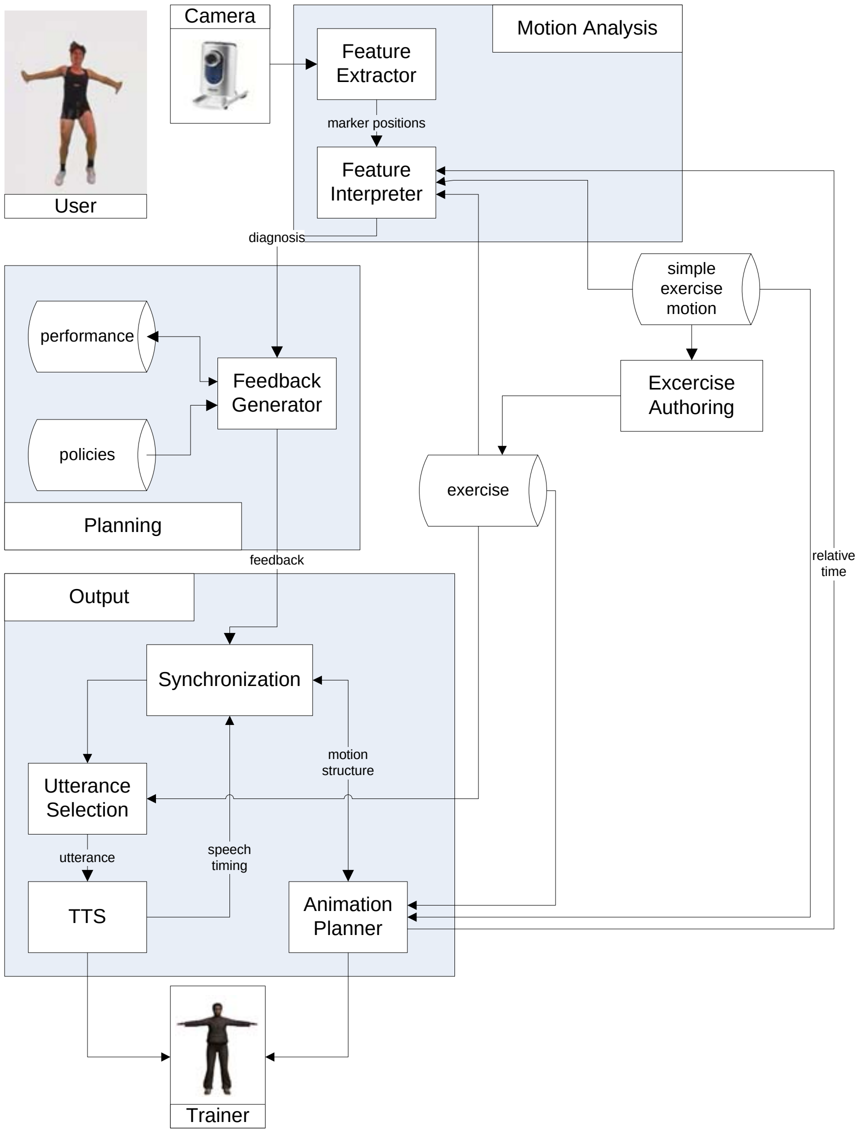

Fig 2. The architecture of the RVT, with major modules and data transfer.

4. Preliminary evaluation

Our current implementation (see Figure 4) was pre-tested with some 10 'casual users' belonging to two major groups: adults around 50, and students around 23, native Dutch speakers. They had to perform a short session (3-5 minutes) introduced and guided by the VT. Afterwards, they were interviewed. The actions of the VT were logged and analyzed after the session.

Most of our users got engaged with the VT, she made several of them sweat. Users saw the VT as someone they would welcome, should she offer more 'serious' exercises, tailored to their needs. This, in general, is a promising feedback, especially that the pre-test had the clear features of a try-out, not of a real application setting.
The critical feedbacks addressed the following issues:

- Too general/misleading 'wrong' feedback.
- Speech quality (Loquendoe TTS) was not engaging.
- Trainer embodiment should be improved.

Fig 4. Screen snapshot showing the Virtual Trainer in action.

Acknowledgement

We thank B. Varga and G. Nagy for their work on recording the corpora, A. Zanderink, F. Zijlstra, M. Klaas, R. Rookhuiszen and R. Leemkuil for implementing some modules of the RVT, and the physiotherapists Dr. M. Horváth and N. van Regteren for giving insight to and contribution with recordings of training practices.

This research has been supported by the GATE project, funded by the Netherlands Organization for Scientific Research (NWO) and the Netherlands ICT Research and Innovation Authority (ICT Regie). The motion capture recordings of the ER corpus were carried out in the framework of a Short Mission supported by the EU Action COST2102.

References

1. S. Babu, C. Zanbaka, J. Jackson, T-O. Chung, B. Lok, M. C. Shin, and L. F. Hodges: *Virtual Human Physiotherapist Framework for Personalized Training and Rehabilitation*, Short Paper Graphics Interface 2005, Victoria, British Columbia, Canada, 2005.
2. F. Buttussi, L. Chittaro, D. Nadalutti:, *Bringing mobile guides and fitness activities together: a solution based on an embodied virtual trainer*, Proceedings of MOBILE HCI 2006: 8th International Conference on Human-Computer Interaction with Mobile Devices and Services, ACM Press, New York, September 2006, pp. 29-36
3. T. Bickmore, R. Picard: *Towards caring machines*, Proc. of CHI, 2004.
4. S-P.Chao, C-Y Chiu, S-N, Yang, T-G. Lin: *Tai Chi synthesizer: a motion synthesis framework based on key-postures and motion instructions*. Computer Animation and Virtual Worlds, Vol. 15. pp. 259-268, 2004.
5. Cybertherapy URL: http://www.cybertherapy.info/pages/main.htm
6. HomeLab Philips, http://www.research.philips.com/technologies/misc/homelab/
7. W. IJsselsteijn, Y. de Kort, J. Westerink, M. De Jager, R. Bonants: *Fun and Sports: Enhancing the Home Fitness Experience*, Proc. of ICEC 2004.
8. S. Marsella, L. Johnson, and C. LaBore: *Interactive Pedagogical Drama for Health Interventions*, AIED 2003, 11th International Conference on Artificial Intelligence in Education, Australia, 2003.
9. G. Nagy: *Modelling Physical Exercises*, Engineering Design Report. Pazmany Peter Catholic University, Dept. of Information Technology, Budapest, 2007. (In Hungarian),
10. Parlevision showcase at HMI, University of Twente, http://hmi.ewi.utwente.nl/showcases/parlevision
11. Reactive Virtual Trainer showcase at HMI, University of Twente, http://hmi.ewi.utwente.nl/showcase/The%20Reactive%20Virtual%20Trainer
12. H. S. Reichel, R. Groza-Nolte: *Fizioterápia.* Budapest: Medicina Kiadó, 2001
13. D. Reidsma, H. van Welbergen, R.W. Poppe, P. Bos and A. Nijholt *Towards Bi-directional Dancing Interaction*, in *Proc. of 5th International Conference on Entertainment Computing*, LCS 4161, Springer Verlag, Berlin, pp. 1-12, 2006.
14. Z.M. Ruttkay, J. Zwiers, H. van Welbergen and D. Reidsma: *Towards a Reactive Virtual Trainer*, in Proceedings of the 6th International Conference on Intelligent Virtual Agents, J. Gratch, M. Young, R. Aylett, D. Ballin and P. Olivier (eds), Lecture notes in Computer Science, volume 4133, Springer Verlag, Berlin, pp. 292-303,
15. S. Shafaei, M. Rahmati: *Physiotherapy Virtual Training by Computer Vision Approach*, Third International Workshop on Virtual Rehabilitation, 2004. http://www.iwvr.org/2004/
16. Sony: EyeToy: Kinetic, http://www.us.playstation.com/Content/OGS/SCUS-97478/Site/ and http://www.eyetoykinetic.com
17. B. Varga: Movement modeling and control of a Reactive Virtual Trainer, Master Thesis, Pazmany Peter Catholic University, Dept. of Information Technology, Budapest, 2008.
18. Virtual Rehabilitation 2007, http://www.aristea.com/iwvr2007/
19. H. Welbergen, Zs. Ruttkay, *Multilayered Motion Modelling for Physical Exercises*, Submitted to MiG 2008, Utrecht, The Netherlands
20. J. Westerink, M. de Jager, M., Y. de Kort, W., IJsselsteijn, R., Bonants, J. Vermeulen, J. van Herk, M. Roersma: *Raising Motivation in Home Fitnessing: Effects of a Virtual Landscape and a Virtual Coach with Various Coaching Styles*. ISSP 11th World Congress of Sport Psychology, 15 - 19 August 2005, Sydney, Australia.
21. Wii Sports, http://nlbe.wii.com/software/02/
22. P. Wittenburg, H. Brugman, A. Russel, A. Klassmann, and H. Sloetje: *Elan: a professional framework for multimodality research*, Proc. of LREC 2006.

Web addresses were visited on 10.06.2008.