Natural User Interfaces: A Physical Activity Trainer

Nicolás Jofré*, Graciela Rodríguez, Yoselie Alvarado, Jacqueline Fernández, and Roberto Guerrero

Laboratorio de Computación Gráfica (LCG)
Universidad Nacional de San Luis,
Ejército de los Andes 950
Tel: 02664 420823, San Luis, Argentina
{npasinetti,gbrodriguez,ymalvarado,jmfer,rag}@unsl.edu.ar

Abstract. Despite the known health benefits of regular participation in physical activity people still refuse this practise. Nowadays, Virtual Reality (VR) is a very powerful and compelling computer tool by which humans can interface and interact with computer-generated environments. In this paper, we propose a virtual training system which can be customized for the physical activity level of the user. System provides real time visual action guide and a performance's feedback of users through a Natural User Interface (NUI). We conduct a brief pilot study to evaluate our virtual trainers in which participants' performance doing physical activities is evaluated via our NUI. Initial results indicate that virtual training through a NUI is motivating and entertaining for any kind of user, in particular for individuals with low level of physical activity.

Keywords: Virtual Reality (VR), Embodied Conversational Agent (ECA), Natural User Interface (NUI), Virtual Physical Activities, Kinect.

1 Introduction

Virtual Reality technology has become a very popular technology integrating the newest research achievements in the fields of computer graphics, sensor technology, ergonomics and Human-Computer Interaction (HCI) theory. As a research tool, Virtual Reality provides numerous opportunities of what can be done and seen in a virtual world that is not possible in real world [1, 2].

In HCI area, Embodied Conversational Characters have emerged as an specific type of multimodal interface, where the system is represented as a person conveying information to human users via multiple modalities such as voice and hand gestures, where the internal representation is modality-independent, both propositional and non-propositional. Embodied Conversational Character answers questions and performs tasks through interaction in natural language-style dialogues with users contrasting the traditional view of computers [3].

* Corresponding author

Virtual human research has progressed rapidly over the last 18 years. The first implemented work of this type appeared in 2008 and many works of Embodied Conversational Characters have recently emerged due to interfaces have great potential as smart assistants, travel agents or investment advisors, among others [4–8].

Verbal and non-verbal behavior of virtual characters has become more and more sophisticated as a result of advances in behavior planning and rendering [9]. VR uses these characters for user-environment interaction, where system must collect gestural, positional, sound and biometric user information through use of sophisticated devices [10]. These devices have incremented their simplicity and user-friendliness over time, improving interaction and giving rise to Natural User Interfaces [11]. These interfaces are used to solve problems in many areas such as medicine, robotics, non-verbal communication, among others.

In last years sedentary lifestyle has become an increasing problem for people's health. It is reported to be indeed the cause of several serious illnesses like obesity, diabetes, hypertension and so on. There are many reasons why people do not perform any physical activity: motivational lack, time constraints, difficulties to start, gym membership fees, equipment costs, among others. Physical activity is defined as any bodily movement produced by skeletal muscles that results in energy expenditure.

In order to provide a more natural interaction between humans and machines, HCI works have brought about the wide focus on body motion recognition. This involves the recognition of physical movement of the head, hands, arms, face or body with the aim of conveying kinesic information [12].

In this context, there are some commercial applications like *Nintendo* or *Xbox Games* which provide different entertainment activities [13, 14]. These applications track user's body movements, such as knees, arms, legs, waist, hip, among others. However, the aim is not focused on obtaining a healthier lifestyle and training is not personalized.

There exist many physical activities research works in HCI [15–19]. From previous studies, the existing work technologies for this purpose are smart phone applications, sensor devices and image processing, with a few solutions implementing virtual characters [20–22]. Those last-mentioned share in common a low-cost virtual training computer application for users who need to undertake regular physical exercise or cognitive tasks at home, with identified specific shortcomings in user's interface, interaction's place and evaluating methods.

In this work we present a training system based on an Embodied Conversational Agent as a real-time simulation approach to train and guide people's physical activities through a NUI.

This paper is arranged as: a design guidelines of our system is introduced in Section 2. Our system is described in Section 3. The system architecture is explained and illustrated in Section 4. The evaluation and results are shown in Section 5. We conclude our paper and propose some future works in Section 6.

2 Design guidelines

In 2005, researchers from different Universities of the United States provided a state of the art of pervasive computing in sport technologies, with the aim to encourage new research in this emerging area and to describe how technology can be applied to sports [23]. They identified three areas of application - *athletic performance, entertainment* and *support for referees* - pointing out some research lines for each, such as studying which sensors are more appropriate to monitor the performance of different kinds of athletes and studying the use of games and special equipment to encourage users to exercise harder.

As was stated in Section 1, our work's proposal is the study result of several works related to virtual trainers [20–22, 24]. From this, two relevant concepts have been identified: *standard physical activities* (considered by a lot of works) and *learning methods* (used by many similar systems).

2.1 Standard Physical Activities

Levels of participation in physical activity remain low across many age groups, and some strategies to increase activity levels throughout the population are needed. Particularly, is possible to perform a generic levels classification on two groups:

1. Average-to-High Level: people with active jobs or that occasionally play sports and people considering physical activity as regular.
2. Low Level: people with office's work or with relaxing daily life activities.

It is necessary and important to identify and use strategies for making physical activities that are both effective and cost-effective. In general, physical activities can be distinguished in **flexibility exercises** (stretch, do yoga), **muscle fitness exercises** (wall climb, use exercise bands), **vigorous sports and recreation** (play sports, hike, play active games), **vigorous aerobics** (ride a bike, jog, jump rope), **moderate physical activity** (walk, play games, do yard work).

At the same time, these activities can be associated to levels of participation in physical activity of people as follows:

- Average-to-High Level: flexibility exercises, muscle fitness exercises, vigorous sports and recreation, and vigorous aerobics.
- Low Level: flexibility exercises and moderate physical activity.

2.2 Learning Method

Virtual-reality-enhanced interactive learning environments are increasingly common. VR brings together a mixture of virtual and real-life scenarios for a wide range of potential possibilities in teaching and learning.

In this context, *Game Based Learning* (GBL) is a learning method leveraging the power of computer games to captivate and engage end users for a specific purpose, such as to develop new knowledge and skills [25]. From past reviewed studies, the minimum components required in all online games for learning are: *back story and story line*, *game mechanics*, *rules*, *immersive graphical environment*, *interactivity*, *challenge/competition* and *risks* [26]. Additionally, many educators use this method in combination with the following metaphors for an educational game [27]:

- Acquisition: to transfer information from who owns it (the teacher) to a passive receiver (the student).
- Imitation: focuses on imitation of model behaviors through observing the reactions of others to the facts.
- Experimentation: it is applied in the learning of specific activities, complex or dangerous tasks, since it promotes active and contextualized learning processes, mainly related to practical activities and physical abilities to a great extent.
- Participation: the transmitted content by the teacher is taken as a learning stimulus, which occurs naturally and difficult to predict.
- Discovery: learning by discovery can be an individual or social activity; the crucial point is that it creates new contents through the active student's participation.

3 System Description

The basic idea was to simulate an environment where the user can train in a green space accompanied by a virtual trainer.

For user-system interaction, a sport field scenario was modeled and rendered. Scenario was set up with ambient sound, inanimated objects (benches, lights, trees) and animated objects (people). Particularly, people on stage are avatars. The user can navigate through the park, paddle court and other places allowing social interaction. The trainer is a female animated character called ***Sara***. This was developed to provide a visual action guide in real time and an actions assessment after physical users' exercises.

As was mentioned in Section 2.2, a computer educational game requires certain components and can use several metaphors. While our system allows physical training, it is also an intention to add entertainment to interaction. Additionally, as it has seen in Section 2.1, it is possible to perform a generic level classification and to associate certain activities for each level. Thus, some components and metaphors will be indicated along this section.

According to the referred levels, this work established the following particular activities:

- Average-to-High Level: single-leg squats, mountain climbers, tuck jumps and burpees.
- Low Level: squats, lying hip raises, side lunges and frog jumps.

At the interaction's beginning, *Sara*'s face representation welcomes the user and explains him about system's operation (*Acquisition*). The user have to choose one participation level and consequently, the associated level's activities from a *Graphic User Interface* (GUI). The act of choosing a level and an activity implies to enter voice data through a microphone. It is assumed that the user will choose a level according to his/her daily life activities.

After selection, system will show *Sara* in full body to start teaching user selected activity's actions. User's actions are captured and compared to standard *Sara*'s actions (see Figure 1) while giving, in real time, a fitness score representing how well user performs actions (*Challenge*), besides providing an activity completeness percentage. *Sara* will then provide interpreted feedback based on obtained fitness score and how user can improve his/her actions (*Imitation* and *Experimentation*).

Considering that each physical exercise is related to a bodily movements set which have limbs' degrees associated (information provided by real trainer stored in a database), system provides a fitness score (percentage of mistakes 0% - 100%) comparing input movements degrees with stored movements degrees.

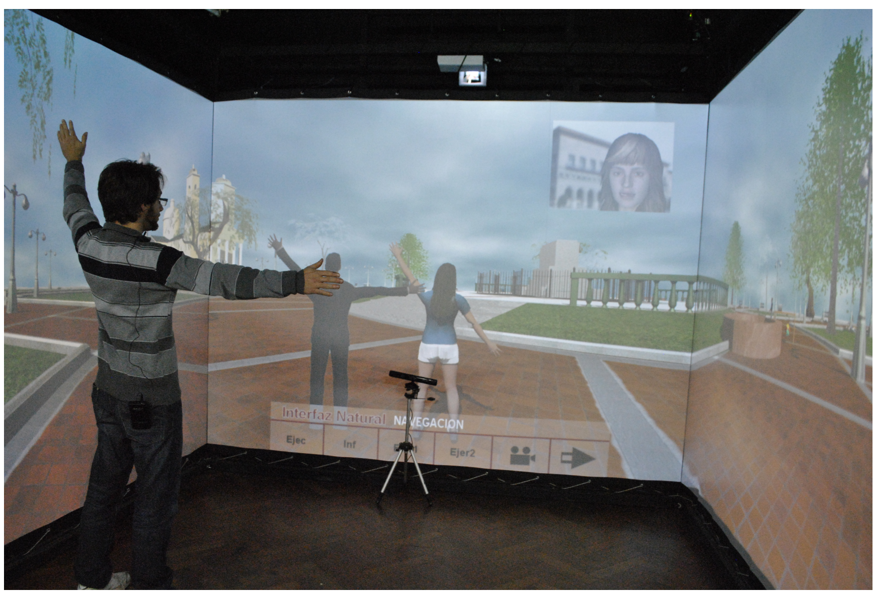

Fig. 1. System Operation.

The system was developed to work on a computing platform for immersive collaborative 3D virtual world visualization (*Immersive Graphical CAVE-like environment*). This platform allows the user to play an appropriate role (through a character) during system's explanation, physical activity and feedback according with conversational aspects of virtual trainer (which results in a *Verbal Natural User Interface*).

User's actions are captured by a motion sensing input device [28]. The motion device is able to detect the shape of a human body by using an RGB camera and depth sensor (*Skeleton Tracking*). User and bodily sensor must be located inside CAVE-like environment for an enhanced interaction (*interactivity*) and a better recognition of user movements during physical activity (resulting in a *Gestural Natural User Interface*).

4 System Architecture

The implemented training system is an integrated system consisting of three subsystems working in an independent parallel way: a *CAVE-like Subsystem*, a *Conversational Character Subsystem* and a *Motion-NUI Subsystem*. The system comprises the hardware and software necessary to gather the information obtained during the interaction between the user and the training system: via a motion sensing device, sound system, microphone, screen/projection surfaces and projectors, among others.

In the *CAVE-like Subsystem* (immersive environment), the user's movements are captured by the *Motion-NUI Subsystem* which sets in motion the avatar animation according with the user movements and notifies to *Conversational Character Subsystem* a percentage of mistakes done by body area. According to the obtained percentage the character advises users how well they are doing the exercises. Figure 2 shows an overview of training system.

CAVE-like SubSystem will provide the necessary structure for attributes definition, rendering and collaborative multi-visualizations, as well as the needed interactive resources. *Cake-like multi-VRmedia Subsystem* is an approach on a computing platform for immersive collaborative visualization of 3D and dynamic system proposed by [29].

Conversational Character SubSystem is composed by a Conversational Character gifted with human figure animation and verbal communication skills, like natural language processing (by *Automatic Speech Recognition mechanism* (ASR)) and speech recognition and synthesis (by *Text-To-Speech mechanism* (TTS)). This subsystem, named *CAVE-VOX* was presented and evaluated in [7, 30] and it was based on previous work [31]. This particular Conversational Character has a collection of responses relevant to a physical activities topic.

Motion-NUI SubSystem is a NUI based on gestures that corresponds with bodily movements. This subsystem obtains, processes and replies information about user's movements. User data is collected thought a motion device, which uses a depth sensor and obtains a 3D input based on user location and posture. The 3D input is utilized for two aspects: user movements analysis and animation data matching. User movements analysis calculates the difference between 3D input and data stored from a predefined data base, and animation data matching calculates the correspondence between 3D input (user body) and user's virtual

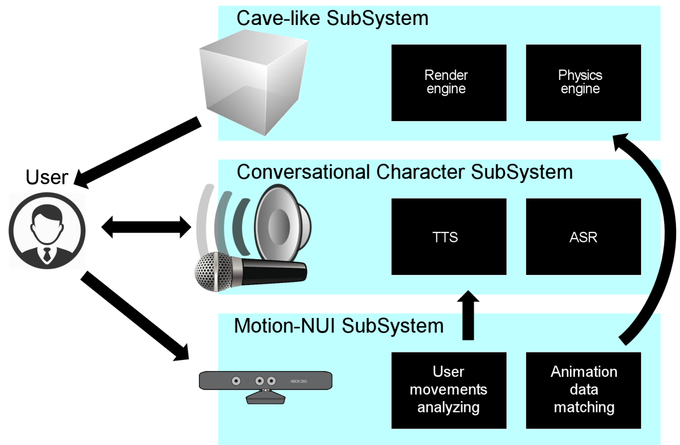

Fig. 2. System Architecture.

avatar (virtual body). After animation data matching an avatar animation is set up for reply the user movements.

5 System Evaluation

A pilot test was applied to a twenty-two participants group between 16 and 60 years old. At first, the system functions were explained and demonstrated to the users undertaking the testing. Every user was then required to explore every function and try it out by him/herself.

As was mentioned in Section 3, participants must be classified in some level. From the total participants, 8 chose the Average-to-High Level and 14 the Low Level.

During the test, each participant is required twice to execute some corresponding level's activity. Each activity consisted of three sets of exercises: a set of 15 repetitions (first set), a set of 30 repetitions (second set) and a set of 40 repetitions (third set).

Each execution was evaluated in different days (at least two days in between) due to it is not recommended anyone exercise more than once in two days [32]. According with each execution an average percentage of completeness by set were recorded (See Table 1 and 2).

For Low Level, the percentage of completeness increased significantly as increasing the execution allowing to achieve an incremental learning. For Average-to-High Level, the percentage of completeness gradually increased with each execution, however, the difference among third set and the others was not overwhelming.

Table 1. Completeness' averages for Low Level

attempt	first set	second set	third set
1	82%	73%	66%
2	96%	86%	89%

Table 2. Completeness' averages for Average-to-High Level

attempt	first set	second set	third set
1	95%	87%	71%
2	97%	96%	94%

6 Conclusions and Future Works

This paper involved the development and evaluation of a training system based in an ECA. We described the design and learning aspects, along with the standard physical activities considered for the implemented system. We introduced the general and most important issues of the training system like, *CAVE-like*, *Conversational Character*, and *Motion-NUI SubSystem*.

The obtained results show that people doing regular physical exercises had no difficulties with virtual exercising but they had an appreciable period of adjustment during first execution. On the other hand, people in Low Level had considerable difficulties and a major period of adjustment on all testing. As regards period of adjustment, we considered that it is necessary to make usability tests and include interviews with children.

Future works will be oriented to:

- improve user's experience, adding new challenges related to physical activities,
- develop additional evaluations of the existing virtual training system,
- extend the pilot test.

References

1. W. Sherman and A. Craig. *Understanding virtual reality: Interface, application, and design*. Elsevier, 2002.
2. J. Blascovich and J. Bailenson. *Infinite reality: Avatars, eternal life, new worlds, and the dawn of the virtual revolution*. William Morrow & Co, 2011.
3. M. Mancini and C. Pelachaud. Generating distinctive behavior for embodied conversational agents. *Journal on Multimodal User Interfaces*, 3(4):249–261, 2009.
4. J. Cassell, T. Bickmore, L. Campbell, H. Vilhjálmsson, and H. Yan. Embodied conversational agents. chapter Human Conversation As a System Framework: Designing Embodied Conversational Agents, pages 29–63. MIT Press, Cambridge, MA, USA, 2000.
5. M. Mancini. *Multimodal Distinctive Behavior for Expressive Embodied Conversational Agents*. Universal Publishers, 2008.

6. Q. Chen, P. Torroni, S. Villata, J. Hsu, and A. Omicini. *PRIMA 2015: Principles and Practice of Multi-Agent Systems: 18th International Conference, Bertinoro, Italy, October 26-30, 2015, Proceedings*. Lecture Notes in Computer Science. Springer International Publishing, 2015.
7. N. Jofré, G. Rodriguez, Y. Alvarado, J. Fernández, and R. Guerrero. Virtual humans conversational character for a cave-like environment. In *XX Congreso Argentino de Ciencias de la Computación*, volume 20, pages 937–946, 2014.
8. J. Pérez, E. Cerezo, and F. Serón. E-vox: A socially enhanced semantic eca. In *Proceedings of the International Workshop on Social Learning and Multimodal Interaction for Designing Artificial Agents*, DAA '16, pages 2:1–2:6, New York, NY, USA, 2016. ACM.
9. S. Kopp and I. Wachsmuth. Model-based animation of co-verbal gesture. In *Computer animation, 2002. proceedings of*, pages 252–257. IEEE, 2002.
10. G. Burdea and P. Coiffet. Virtual reality technology. *International Journal of e-Collaboration*, 2(1):61–64, 2006.
11. J. Blake. *Natural User Interfaces in . Net*. Manning Pubs Co Series. Manning Publications Company, 2012.
12. J. Koenemann, F. Burget, and M. Bennewitz. Real-time imitation of human whole-body motions by humanoids. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2806–2812, May 2014.
13. D. Sheff. *Game over: How Nintendo conquered the world*. Vintage, 2011.
14. J. Andrews and N. Baker. Xbox 360 system architecture. *IEEE micro*, 26(2):25–37, 2006.
15. K. Miller, B. Adair, A. Pearce, C. Said, E. Ozanne, and M. Morris. Effectiveness and feasibility of virtual reality and gaming system use at home by older adults for enabling physical activity to improve health-related domains: a systematic review. *Age and ageing*, 43(2):188–195, 2013.
16. R. Mellecker and A. McManus. Active video games and physical activity recommendations: A comparison of the gamercize stepper, xbox kinect and xavix j-mat. *Journal of science and medicine in sport*, 17(3):288–292, 2014.
17. D. Nathan, D. Huynh, J. Rubenson, and M. Rosenberg. Estimating physical activity energy expenditure with the kinect sensor in an exergaming environment. *PloS one*, 10(5):e0127113, 2015.
18. C. Campos and H. del Castillo. The benefits of active video games for educational and physical activity approaches: A systematic review. *Journal of New Approaches in Educational Research*, 5(2):115, 2016.
19. N. Jofré, G. Rodríguez, Y. Alvarado, J. Fernández, and R. Guerrero. El uso de la realidad virtual inmersiva en terapias motricess. In *XIX Workshop de Investigadores en Ciencias de la Computación*, 2017.
20. B. Li, M. Maxwell, D. Leightley, A. Lindsay, W. Johnson, and A. Ruck. Development of exergame-based virtual trainer for physical therapy using kinect. In *Games for Health 2014*, pages 79–88. Springer, 2014.
21. M. Pedraza-Hueso, S. Martín-Calzón, F. Díaz-Pernas, and M. Martínez-Zarzuela. Rehabilitation using kinect-based games and virtual reality. *Procedia Computer Science*, 75:161–168, 2015.
22. L. Shaw, R. Tourrel, B. Wunsche, C. Lutteroth, S. Marks, and J. Buckley. Design of a virtual trainer for exergaming. In *Proceedings of the Australasian Computer Science Week Multiconference*, page 63. ACM, 2016.
23. E. Chi, G. Borriello, G. Hunt, and N. Davies. Guest editors' introduction: Pervasive computing in sports technologies. *IEEE Pervasive Computing*, 4(3):22–25, 2005.

24. R. Cózar-Gutiérrez and J. Sáez-López. Game-based learning and gamification in initial teacher training in the social sciences: an experiment with minecraftedu. *International Journal of Educational Technology in Higher Education*, 13(1):2, 2016.
25. P. Tan, S. Ling, and C. Ting. Adaptive digital game-based learning framework. In *Proceedings of the 2nd international conference on Digital interactive media in entertainment and arts*, pages 142–146. ACM, 2007.
26. A. Derryberry. Serious games: Online games for learning. 2013.
27. P. Simons and M. Ruijters. Varieties of work related learning. *International Journal of Educational Research*, 47(4):241–251, 2008.
28. Zhengyou Zhang. Microsoft kinect sensor and its effect. *IEEE multimedia*, 19(2):4–10, 2012.
29. Y. Alvarado, N. Moyano, D. Quiroga, J. Fernández, and R. Guerrero. *Augmented Virtual Realities for Social Developments. Experiences between Europe and Latin America*, chapter A Virtual Reality Computing Platform for Real Time 3D Visualization, pages 214–231. Universidad de Belgrano, 2014.
30. Y. Alvarado, G. Claudia, V. Gil Costa, and R. Guerrero. *Computer Science & Technology Series XXII Argentine Congress of Computer Science-Selected Papers,*, chapter ARENA Simulation Model of a Conversational Characters Speech System, pages 119–129. Universidad de La Plata, 2017.
31. F. Serón, C. Bobed, and P. Latorre. *Augmented Virtual Realities for Social Developments. Experiences between Europe and Latin America*, chapter The Vox System, pages 130–159. Universidad de Belgrano, 2014.
32. World Health Organization et al. Recomendaciones mundiales sobre actividad física para la salud. 2010.