Replay 1 (3) / 2016; I. CEEGS 2015 Proceedings; *Articles / Artykuły*
http://dx.doi.org/10.18778/2391-8551.03.04

Stefan Stavrev
TRI Soft

Natural User Interface for Education in Virtual Environments

Introduction

Over the last five years we have witnessed a rise in natural human-computer interactions (Dimitropoulos, Manitsaris & Mavridis, 2008, Bontchev & Vassileva, 2006, Whitton, & Moseley, 2012, Freitas & Liarokapis, 2011, Galland, Gaud, Demange & Koukam, 2014, Gaming and simulations, 2011, Wagner, 2012). This tendency allows researchers to experiment with different devices on the market, such as the Microsoft Kinect and Hololens, Asus's Xtion sensor, Leap motion, Google glasses, Oculus rift, Playstation VR. These devices can also be used as input for controlling three-dimensional avatars in simulated environments, such as computer games. Playing has always been a natural way of examining the world around us, especially in our early years of development (Dawley & Dede, 2014, Gaming and simulations, 2011). It was our motivation to build a virtual platform that enables young children to learn in an easy and natural way. The domain we focused on – safety of street crossing and movement – is both interesting and applicable to meet young children's needs. The virtual platform uses Microsoft's Kinect sensor for capturing gesture input. It then turns these gestures into interactions within the virtual environment. There are two types of interactions / gestures – the directional movement of a 3D avatar and manipulation of the in-game menu. The software that we used is C# and the Unity game engine.

The main purpose of this research is to show that a VR environment can be used for teaching young children vital habits, regulations, and skills for moving safely on and near the streets. As we discuss in this study, the main advantage of VR is that it is safer when compared to real-life learning. We also show that learning in a VR is accepted by children and that they can perform and interact well in it via a natural interface. The rest of the paper is organized as follows. Firstly, we describe the role serious games play in education. Secondly, we make an overview of existing input sensors and motivate our choices of one. Thirdly, we describe in detail the platform that we developed. Fourthly, we discuss the empirical evaluation process during the testing of the platform and discuss the results. Finally, we draw conclusions and mention future aspects of this research.

List of terms:

- Natural user interface (NUI) – it is a user interface that is effectively invisible and remains invisible as the user continuously learns increasingly complex interactions.
- Physical world – the world that we, humans, live in
- Virtual world – a computer-based simulated environment. It is usually populated by one or more users.
- Avatar – it is a representation of the user in a virtual world. The avatars can be textual, two, or three-dimensional graphical representations. In this paper we refer to an 'avatar' as a three-dimensional graphical model of the user.
- FPS – frames per second

Serious games in education

Some researchers refer to the virtual educational environments as "serious gaming". The term is used for the first time by Clark C. Abt in his 1970 book *Serious Games*. He includes in the term both board games and digital games. Serious games serve a clear educational purpose and are not intended as entertainment. That does not mean they cannot be entertaining. The modern understanding of the term "serious games" originates from *Origins of Serious Games* by Sawyer and Rejeski, which manages to inspire a great interest in the idea that public and private educational institutions should utilize modern communication and information technologies to produce video games and simulations that target and research real-world problems (Sawyer & Rejeski, 2002). The most widely accepted definition of the term belongs to one of its creators – Mike Zyda. The formal definition is – "Serious game: a mental contest, played with a computer in accordance with specific rules, that uses entertainment to further government or corporate training, education, health, public policy, and strategic communication objectives" (Zyda, Mayberry, Mccree & Davis, 2005).

There are various examples of virtual reality research. One such example is the work on distance education on the web, where the emphasis is on running the virtual environment in real-time across large number of clients (Dimitropoulos et al., 2008). Another is the work of Stephen Bronack and Robert Sanders. They are developing an educational platform called "AET Zone". According to the authors (Bronack, Sanders, Cheney, Riedl, Tashner & Matzen, 2008), learning occurs in several parts of our brain and is thus distributed. The researchers show that learning in immersive virtual worlds can lead to the activation of one or all of these areas and is advantageous when compared with traditional educational means. Others (Dawley & Dede, 2014, Active Words, 2008) research learning in VE from more general settings. Virtual collaboration rooms for college students are researched by (Bronack et al, 2008). Some researchers prefer to use a well-known client-server platform *Second Life* (Aebersold, Tschannen & Bathish, 2012) because it has a simple and easy frame to work with. On the other hand, it lacks the immersive element when it comes to real-time integration with sensors. Some researchers prefer to build their own framework from the ground up (Barbosa, 2014, Dawley & Dede, 2014, Galland et al., 2014, Avancini, 2012). Although this is a more resource-consuming method, it gives them the possibility of implementing special customization features that are otherwise unavailable in *Second Life* and other popular platforms.

Virtual platform description

Game and level Design

To meet our goal, we developed a 3-dimensional urban environment (Bhattacharya, Paul & Goon, 2012, How to use 3D content in simulations, 2014). The VE does not correspond directly to any actual place on the map. It was designed to meet specific goals –situations diversity, obeying traffic rules, giving appropriate signals on and near the road.

The environment itself is a small city "sandbox". It is comprised of urban buildings – private houses, block apartments, administrative public service buildings – schools, kindergartens, libraries, a sports center and a shopping mall. In addition, we put together a working and realistic traffic system. Furthermore, we developed a simple vehicle artificial intelligence (AI), which contributes to the immersion and realism of the simulation. Finally, we designed a player's avatar and the way it interacts with that environment.

The visual style that we picked is not a realistic-looking one. It is rather cartoon-looking because our target are young children and we wanted them to have fun and feel more like playing a game, rather than training in a virtual environment. To achieve this goal we not only had to use cartoonish color palettes, but also design the 3D objects and the avatar to look attractive to children in terms of size and proportions. For instance, the avatar's head is unusually large compared to its body. This serves an additional purpose as well. While the user is looking left and right in front of a crosswalk for incoming traffic, the classroom and teachers are able to better distinguish which way the avatar's head is rotating. This feature allows for a better understanding of the child's behavior and timely corrections of any misbehavior. Brain research scientists (Knickmeyer et al., 2008, Johnson, 2005) have shown how important learning how to properly check for traffic is in the early years of one's development – 85% of the human brain develops during the first 5 years of our lives (National Research Council and Institute of Medicine, 2000).

We have designed and implemented a total of 10 gameplay situations (levels) including a tutorial level for using the NUI interface. The gameplay levels are modeled as possible scenarios of a child's everyday life – getting from school home, going to the library to get a book, waiting for the bus at the bus stop, cycling to the gym, etc. Each learning situation has a different level of difficulty. We begin with simple and short routes with a walking avatar and move to a cycling avatar and more complex routes. In the main game menu each level is represented by a number (from 1 to 10). In addition, there is a different number of star icons next to each level indicating its difficulty. The notation is: the more stars, the more complex the level is. Another indication on the main menu is whether the level is a walking or a cycling one – we have put cycling icons next to the bicycle levels.

Navigation

The navigation system we implemented inside the game is pretty straightforward. We got ideas from other games (such as GTA), Google maps, and GPS systems. The user follows in-game markers [Figure 1] and on reaching each of them the user is shown the next one on the route. Since the goal was to make the game appealing to children, the last marker is in the form of a league cup [Figure 2]. In order to provide additional insight of where the user was in the level, we implemented a mini-map showing the user their exact location and the outline of the whole route. The mini-map is accessible via the pause in-game menu [Figure 3].

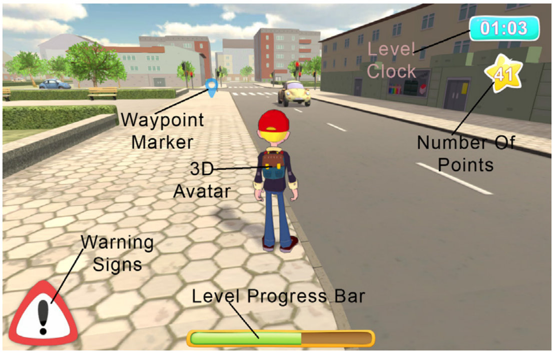


Figure 1. User interface and explanation of gameplay elements

Figure 2. End game marker is a league cup!

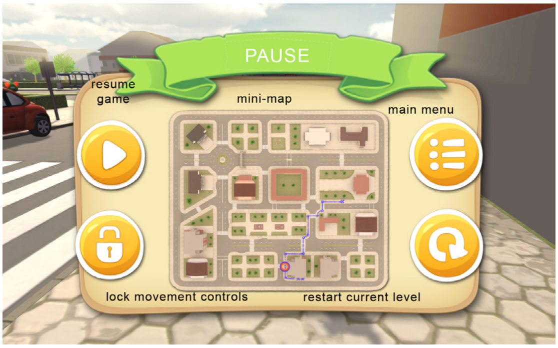
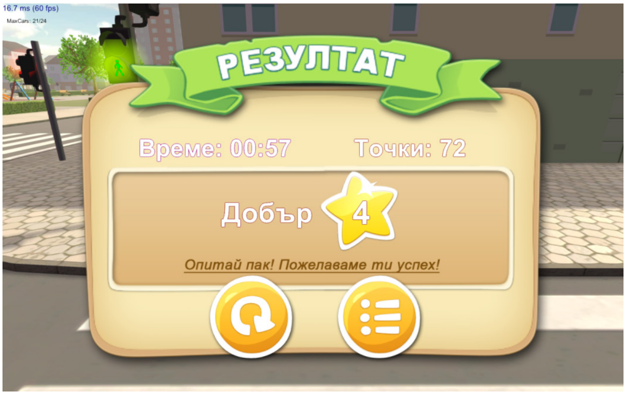


Figure 3. In-game pause menu. The mini-map is drawn at the center of the menu, showing also the player's position and route

Scoring system

For evaluating children's total performance we developed a point-based scoring system. Every time the user behaves in a safe and appropriate manner, their actions are rewarded with positive points. Misbehavior and dangerous actions are discouraged by taking points away. The total accumulated score a user can get by the end of a level equals 100. a score of 80 and more is considered a pass, i.e. the child is capable of moving safely in an urban traffic environment. If a child's actions are done by the book – they have collected all the rewards and no negative points were assigned – the child receives a special reward, applause, at the end of the level.

Visually, the points can be seen in the upper-right corner of the game screen. To make them more appealing to young children, we put them inside a star. At the end of a level, another screen appears [Figure 4] – the so called *end game screen*. It summarizes the performance of the user throughout the level and shows the final score.



16.7 ms (60 fps)

MaxCars: 21/24

РЕЗУЛТАТ

Време: 00:57 Точки: 72

Добър 4

Опитай пак! Пожелаваме ти успех!

Figure 4. End game screen and total score

Artificial Intelligence

In order to make the game more realistic, we added a simple artificial intelligence to the street vehicles. Vehicles are able to "see" some meters in front of them. They react to traffic lights, crosswalks, or other vehicles in front of them. If, for instance, an avatar is on a crosswalk, an approaching vehicle will decelerate and stop by the crosswalk [Figure 5]. On the other hand, if the avatar crosses spontaneously or on red lights, an approaching vehicle will still stop but will force its engine and honk its horn as a symbol of disapproval of the avatar's actions. Even though such vehicle behaviors are hard-coded, they add to the overall user experience and bring realism to the simulation.

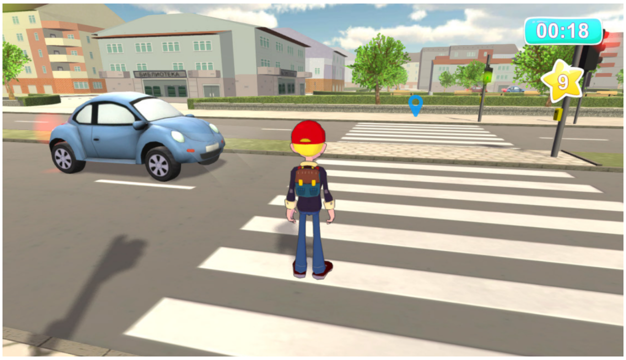


Figure 5. Vehicle AI – a car stops while the avatar is on the crosswalk

Sensory inputs

Humans are not designed to work with devices such as a mouse and a keyboard. Yes, we have learned and adapted but it still feels “unnatural” to most of us. Fortunately, in the last 5–7 years, both proprietary companies and independent researchers are pushing our interactions with machines further towards creating a more natural and human-like interfaces. That “push” has led to the appearance of new devices that tend to shorten the gap in human-computer interactions. Examples of heads-up-displays that help create virtual and augmented realities are Microsoft Hololens, Google glasses, Oculus Rift, and Sony’s Project Morpheus. There are some other sensors that focus more on capturing and processing human gestures. Such devices are Microsoft Kinect, Asus’s Xtion sensor, and Leap motion. Although these sensors employ different approaches (in terms of how they work), their general purpose is the same.

A decision was made to use Kinect as the sensor for processing natural user gestures. This choice of technology was not an arbitrary one. The capture range of Kinect is from 2 to 6 meters (Avancini, 2012, Kinect – Windows app development, 2016), which is far greater compared to the Xtion and Leap motion sensors. In addition, Kinect has a vertical tilt motor, which allows for automatic calibration and an additional degree of freedom, compared to the Xtion. Although Xtion has a slightly better infrared depth camera, its API is not as widely supported and integration with a gaming platform was not going to be trivial. On the other hand, Kinect had a plug-in that can be directly integrated into the game engine we used.

The data that Kinect provided could not be directly used for our avatars – we could not simply map the skeletal tracking data to the 3D rigged skeleton of the avatar.

A different approach was chosen – instead of directly moving the avatar's skeleton, we decided to use pre-recorded avatar animations and 'trigger' them at the appropriate time. For instance, walking on the spot triggers a 'walk' animation; taking a step backwards triggers 'a step backwards' animation and so on. The only notable exception was when we implemented the looking left and right before a crosswalk. For that to work, we had to use the raw position of the user's head because it was changing very rapidly. Another tweak we did was to cap the avatar's head rotation to a certain degree, since the avatar's head was able to rotate to a greater angle than the user's. That trick saved us from visually abrupt and unnatural 3D motions.

In addition to the core movement gestures, we added several helper ones. Pausing the game, for instance, is achieved by making a "safe haven" gesture [Figure 6]. Our empirical research has shown that this is a very common notion among youngsters. Another feature we added helps physically challenged children. People on a wheelchair that are able to move their hands and head are still able to play in the simulator and learn. From the in-game options menu, one could 'lock' part of the controls, so that only the upper-body skeletal data is captured and processed. Moving the avatar forward and backwards in this mode is achieved by a regular keyboard. Unfortunately, we did not get the opportunity to field test this mode on any physically-challenged students. We only tested it in a controlled environment by ourselves, sitting in a chair.

Coming up with different gestures was a challenging task but at the end we did manage to cover all possible in-game interactions.

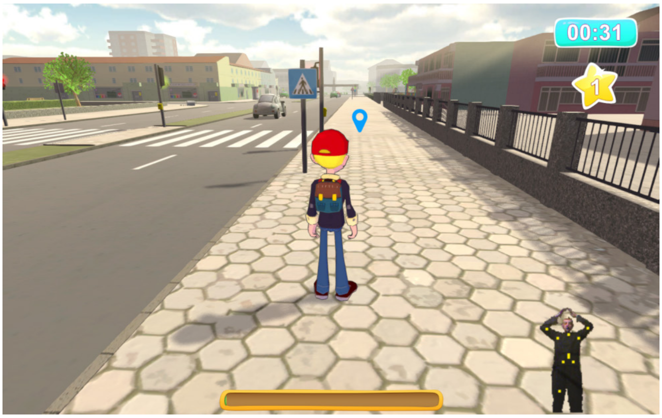


Figure 6. Pause game gesture – safe haven

Frame-rate optimization

The simulation was needed to run in 25–30 fps on a low-end machine. Unity 3D by itself is able to achieve such rendering speeds but adding a Kinect sensor into the setup made that task a challenging one. The Kinect device is bound to 30 fps and even lower (15 fps) in low-lighting conditions (Kinect – Windows app development, 2016). We had to optimize not only the capture, drawing on the screen, and skeletal data tracking but also the visual style and the quality of the scene.

The first thing we noticed was that the Kinect plugin we used draws the user on the game screen together with the overlaid skeletal tracking data. We tried out several techniques to speed up the user gesture data rendering. One of our initial moves was to just cut-out the Kinect rendering window at all. That did improve the frame rate but the users lost a feedback system to show what gestures they have performed. This led us to try a different approach. Instead of eliminating the Kinect rendering window, we decided to show only the skeletal tracking data, drawn with the help of connected yellow lines [Figure 7]. That trick did improve rendering performance but again the feedback we fed to the user was not good enough. Finally, we decided to keep the user image and draw only the skeletal points that the Kinect sensor captures [Figure 6]. Reducing the resolution of the Kinect drawing window was another optimization that we were able to do – the user did not need to see his captured image in high detail. Moreover, we only draw every second captured frame in the Kinect rendering window.

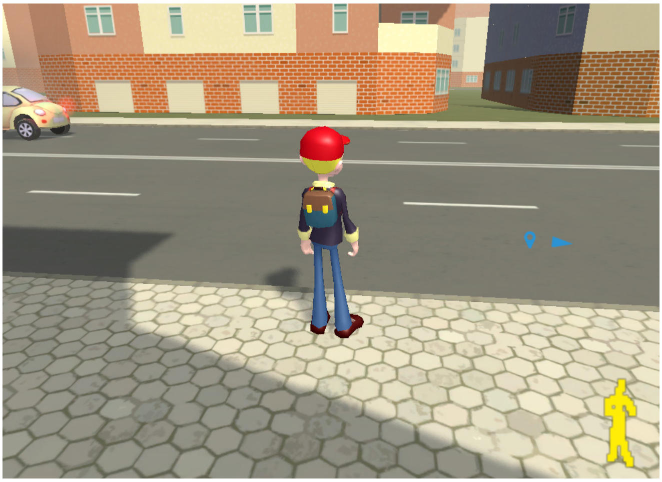


Figure 7. Optimizing Kinect skeletal data drawing window

Optimizing rendering features was another challenge. Enabling antialiasing, for instance, did drop a few fps. Controlling special effects, such as depth of field, drawing (view) distance, motion blur, bloom, etc. allowed us to get a better handling of the rendering pipeline. We were forced to use low-detail level geometry and compensate with better-looking textures. In general, we did not use complex shaders for the scene – the visual style was cartoonish and so the emphasis was not on realism. With all the mentioned optimizations implemented, we were able to boost the frame-rate to 30–32 fps on a low-end machine (CPU: Intel Core 2 Duo @ 2.6 GhZ, RAM: 2GB DDR 3, Video: 512 MB, OS: Windows 7).

Results

The game was initially shipped and tested in more than 50 schools and kindergartens in Bulgaria. These production tests allowed us to gather additional feedback data for the game. The research was conducted on children from 6 to 10 years old of both genders – M(ale) and F(emale). The evaluation criteria were several: how easy it is to control the mouse and the avatar, and how many points the child was able to gather at the end of the level (game score is form 0 to 100 points). The last column shows whether the child scored over 80 points, at which point the level is considered a 'pass'. An excerpt of the evaluation results is presented in Table 1.

Age	Gender	Controlling the avatar	Controlling the mouse cursor	Game Score (1–100)	Level Pass (80+ points)
6	M	Difficult	Difficult	86	Yes
6	F	Difficult	Difficult	73	No
6	F	Difficult	Difficult	100	Yes
7	M	Difficult	Difficult	96	Yes
7	F	Easy	Easy	83	Yes
7	M	Difficult	Easy	90	Yes
8	M	Difficult	Difficult	84	Yes
9	F	Easy	Easy	100	Yes
10	M	Easy	Easy	100	Yes
10	F	Easy	Easy	96	Yes

Table 1. Evaluation test results

Analysis and discussion

What we found out was that younger children had difficulty to walk and rotate at the same time. They stop, rotate on the spot and continue in the new direction. This is most likely due to the fact that their fine muscle coordination is not fully developed yet. Older participants show that they are better at rotating while moving. We plan to solve this by integrating a VR headset (such as Oculus Rift) so that younger participants can rotate more naturally, i.e. with their head instead with their head and shoulders. We also noticed that some gestures are more intuitive and easy to perform than others. For instance, the gesture for "safe haven" is easier to perform than the gesture for "pushing a button". In addition, we noticed that the use of simpler gestures is preferred over more complex ones. We plan to address those issues in the future development of the virtual platform. Another shortcoming of our approach was the fact that younger children had difficulty to steadily move their hands while selecting options from the menu. As a result of this feedback, we added additional menu navigation options, such as holding one's hand still over the desired option for five seconds in order for the option to get selected. Other than the issues addressed above, we observed no difficulties for the children interacting with the virtual platform.

Future work

The platform presented in this paper is not perfect. Future upgrades may include adding new gameplay levels that further improve the students' behavior near and on the road. We also thought about using different sensors (such as Oculus Rift) to increase the immersion depth of the user's reality perception. In addition, further research may show that some gestures are generally preferred over others. The choice of the "button activation" gesture, for example, seems difficult to perform for most children. On the other hand, we've noticed that the Kinect sensor recognizes some gestures better than others. This may be partly because of the low resolution of the Kinect' RGB camera (640 x 480 px). We would like to see how the Kinect 2.0 performs with the current implementation of the virtual platform, since it has a better resolution of the RGB camera (1920 x 1080 px). The other reason why we plan to use the Kinect 2.0 in the future is that version 1.0 will no longer be supported or manufactured.

Conclusion

In this paper we have presented a successful proof-of-concept that a NUI can be integrated as part of game design in order to produce an immersive virtual educational environment. Furthermore, we showed that VE can be a safer way for children to learn when compared to learning in the physical world. As we demonstrated, gaming in education is a fast-morphing field of study. We saw that games present a natural environment for learning and that children can be beneficial for interacting in such virtual worlds. In addition to physical safety, the virtual platform is a time and energy saver for a teacher. That being said, we believe that the future of education lies in simulation environments and NUI sensors.

Acknowledgements

The research presented in this paper was commissioned by our company – TRI Soft Ltd. It was assigned to us and funded by a private Bulgarian company – Didasko Ltd., specializing in education for safety on the roads. I would like to thank our development team for the dedication they put into this project.

References

Active Words. (2008, September 18). Retrieved March 10, 2016, from http://www.activeworlds.com

Aebersold, M., Tschannen, D., & Bathish, M. (2012). Innovative Simulation Strategies in Education. *Nursing Research and Practice*, 2012, 1–7.

Avancini, M. (2012). *Using Kinect to emulate an Interactive Whiteboard* (Unpublished master's thesis). Università Degli Studi Di Trento.

Barbosa, H. (2014). Delivering 3D, Simulation and Serious Gaming For Education and Training. *3D Visualization World Magazine*.

Bhattacharya, A., Paul, P. S., & Goon, S. (2012). History and comparative study of modern game engines. *IJAMS International Journal of Advanced Mathematical Sciences*, 3(2), 245–249. Retrieved from ISSN.

Bontchev, B., & Vassileva, D. (2006). Software architecture of a self-adaptive hypermedia e-learning system. In *Proc. of 3rd Int. Conf. on Computer Science*. Istanbul, Turkey.

Bronack, S., Sanders, R., Cheney, A., Riedl, R., Tashner, J., & Matzen, N. (2008). Presence Pedagogy: Teaching and Learning in a 3D Virtual Immersive World. *International Journal of Teaching and Learning in Higher Education*, 20(1), 59–69.

Dawley, L., & Dede, C. (2014). Situated learning in virtual worlds and immersive simulations. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), T*he Handbook of Research for Educational Communications and Technology* (4th ed.). 723–734.

Dimitropoulos, K., Manitsaris, A., & Mavridis, I. (2008). Building Virtual Reality Environments for Distance Education on the Web: a Case Study in Medical Education. *International Journal Of Social Sciences*, 2(1).

Freitas, S. D., & Liarokapis, F. (2011). Serious Games: a New Paradigm for Education? *Serious Games and Edutainment Applications*, 9–23.

Galland, S., Gaud, N., Demange, J., & Koukam, A. (2014). Multilevel Model of the 3D Virtual Environment for Crowd Simulation in Buildings. *Procedia Computer Science*, 32, 822–827.

Gaming and simulations: Concepts, methodologies, tools and applications (2011). Hershey PA, NY: Information Science Reference.

How to use 3D content in simulations for teaching and learning (n.d.). Retrieved September 23, 2014, from https://www.jisc.ac.uk/blog/3d-in-teaching-and-learning-23-sep-2014

Johnson, M. H. (2005). Sensitive periods in functional brain development: Problems and prospects. *Dev. Psychobiol. Developmental Psychobiology*, 46(3), 287–292.

Kinect – Windows app development (n.d.). Retrieved March 10, 2016, from https://www.microsoft.com/en-us/kinectforwindows/

Knickmeyer, R. C., Gouttard, S., Kang, C., Evans, D., Wilber, K., Smith, J. K., Gilmore, J. H. (November 19, 2008). A Structural MRI Study of Human Brain Development from Birth to 2 Years. *Journal of Neuroscienc*e, 28(47), 12176–12182.

National Research Council and Institute of Medicine (2000). *From Neurons to Neighborhoods: The Science of Early Childhood Development*. Washington, D.C.: National Academy Press.

Sawyer, B., Rejeski D. (2002). *Serious Games: Improving Public Policy through Game--Based Learning and Simulation*. Washington, DC: Woodrow Wilson International Center for Scholars

Stavrev S., Terzieva T. (2015). Virtual environment simulator for educational safety crossing. *Proceedings of the 11th Annual International Conference on Computer Science and Education in Computer Science (CSECS)*, June 4th to June 7th, 2015, Boston, MA, USA, pp. 92–98, ISSN 1313–8624.

Wagner, R. (2012). Learning with Digital Games: a Practical Guide to Engaging Students in Higher Education – By Nicola Whitton. *Teaching Theology & Religion*, 15(1), 91–92.

Whitton, N., & Moseley, A. (2012). Authentic contextual games for learning. In Whitton, N., & Moseley, A. (Eds.), *Using games to enhance learning and teaching: a beginner's guide*, U.S.A.

Zyda, M., Mayberry, A., Mccree, J., & Davis, M. (2005). From Viz-Sim to VR to Games: How We Built a Hit Game-Based Simulation. *Organizational Simulation Rouse/ Organizational*, 553–590.

Ludography

Rockstar North (2013), *GTA V* (GOTY edition) [videogame][DVD-ROM][Microsoft Windows], Rockstar Games

TRI Soft (2015), *Virtual Platform for Safety Crossing* [videogame][DVD-ROM][Microsoft Windows], Didasko Ltd.

Summary

Education and self-improvement are key features of human behavior. However, learning in the physical world is not always desirable or achievable. That is how simulators came to be. There are domains where purely virtual simulators can be created in contrast to physical ones. In this research we present a novel environment for learning, using a natural user interface. We, humans, are not designed to operate and manipulate objects via keyboard, mouse or a controller. The natural way of interaction and communication is achieved through our actuators (hands and feet) and our sensors (hearing, vision, touch, smell and taste). That is the reason why it makes more sense to use sensors that can track our skeletal movements, are able to estimate our pose, and interpret our gestures. After acquiring and processing the desired – natural input, a system can analyze and translate those gestures into movement signals.

Keywords: simulators, virtual environments, 3D environment, serious gaming, sensors, education with Kinect.

Stefan STAVREV, MSc, part-time assistant professor in FMI, Plovdiv University Paisii Hilendarski; C.E.O. of TRI Soft Ltd, 33 “Petko D. Petkov” Str, Plovdiv 4000, Bulgaria, ceo@tri-soft.net. Major fields of scientific research: Artificial Intelligence, Reinforcement Learning, Serious Gaming.