Revista Romana de Interactiune Om-Calculator 10 (2) 2017, 89-104

Virtual training and testing environments for visually impaired people

Silviu Ivascu, Alin Moldoveanu, Florica Moldoveanu, Maria-Iuliana Dascalu

University POLITEHNICA of Bucharest
E-mail: ivascu.silviu10@gmail.com, alin.moldoveanu@cs.pub.ro, florica.moldoveanu@cs.pub.ro, maria.dascalu@upb.ro

Abstract. There are a lot of problems that visually impaired people face daily. One of them is navigating through unknown environments. The Sound of Vision project (Sound of Vision, n.d.) wants to solve this problem by converting visual stimuli into sound and haptic feedback. This paper has two main parts. In the first part, we offer a general overview of the Sound of Vision project. In the second part, we focus mainly on the Virtual Training Environments (VTE), which were developed for training visually impaired people in using the SOV device and measure their performance for future improvements.

Keywords: Sound of Vision, Virtual Reality - VR, Augmented Reality – AR, Visually impaired

Introduction

Visually impaired people lack one of the human senses, thus sensory substitution is an alternative method for helping them. This can be done by acquiring information about their surroundings and sending it to them encoded into audio or tactile stimuli. Previous experiments proved that, by using this method, visually impaired people can make a spatial cognitive map of the environment almost as sighted people can have (Balan, Moldoveanu, Moldoveanu, & Butean, 2015).

The hearing sense offers essential information regarding direction, duration and intensity of the stimuli, distance and range, information that can be used to create the cognitive spatial map. The information offered by the hearing sense is known as audio cues.

An important aid for navigation, mobility learning and wayfinding is the tactile map. This is used to convey information about spatial components of the environment. Haptic cues are used to create the tactile map of the

environment.

Combining audio and haptic cues results in an audio-tactile multimedia approach that enhances the spatial perception. The Sound of Vision project (Jóhannesson, Balan, Unnthorsson, Moldoveanu, & Kristjánsson, 2016) aims to help visually impaired people to perceive and navigate in almost any kind of environment (indoor/outdoor):

- without the need for any predefined tags/sensors located in the surroundings;
- regardless of the lighting conditions (day/night, natural/artificial light).

The SoV project team has created an original non-invasive, wearable device (including hardware + software components) able to generate a multisensory auditory + haptic representation of the surrounding environment. This representation is created, updated and delivered to the visually impaired users continuously and in real time. Through this representation, the user can perceive, to some extent, the environment, without blocking relevant auditory information (Balan, Moldoveanu, & Moldoveanu, 2015).

In addition to this continuous representation, the system will permanently monitor for discrete events of interest that include a well-defined set of special situations, both of positive nature (e.g. encountering a social acquaintance) and negative nature (e.g. danger to hit an object or fall into a hole in the ground). In the case when any such event is detected, it will be signaled to the user, in a specific way, based on its importance and priority. (Bujacz, et al., 2016)

Related work

In this section we will discuss theories and similar work which guided us in the development of this project. We will present the important aspects of cognitive learning based on visual memory and visual interaction. Also, we will expose the roles of serious games in cognitive learning and the main components of gamification.

2.1 Cognitive Learning Based On Visual Interaction And Visual Memory

This section will present one of the most important parts of the process. We will discuss the relations between cognitive learning and the visual environment. By visual environment we understand all the elements related to visual memory and visual interactions between the user and the environment. We can summarize a small collection of assumptions for the cognitive learning before exploring the approaches into these relations. The collection of cognitive theories is presented in the following list (Vygotsky, 2003):

- Cognitive processes are the focus of study.
- Some learning processes may be unique to human beings.
- Objective observations related to people's behaviour have to be the focus of scientific inquiry, but one can also make inferences about unobservable mental processes.
- Learning involves the formation of mental associations that are not necessarily reflected in behaviour changes.
- Individuals are actively involved in the learning process.
- Learning is a process of relating new information to previously learned information.
- Knowledge is organized.

This collection of assumptions used for general educations implications of cognitive learning has been developed as follows:

- Cognitive processes influence learning.
- People organize the things they learn.
- As children grow, they become capable of more sophisticated thoughts.
- People control their own learning.
- New information is easily retained when people are able to associate it with things they have already known.

As for the social cognitive theories, the learning process is related to three variables which work together and help learning occur. The behavioural determinants and the environmental factors represent the two components which the individual’s personal experience converge with.

These three variables and their interrelations are illustrated in Fig 1.

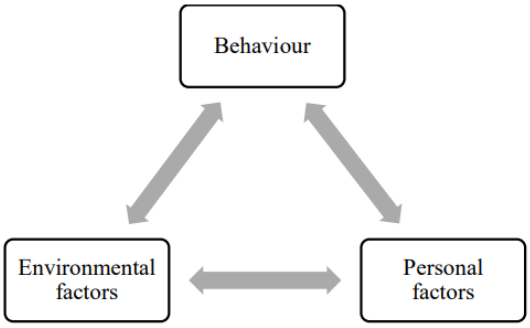

Figure 1. Cognitive Theory Illustration

2.2. Serious Games Used For Cognitive Learning

Games can be defined as being an entertaining activity but can also be defined as a series of activities like the following: mystery, challenges, competition, curiosity, fantasy, goals, skill and rules (Leach & Sugarman, 2006). At its roots, a game is a physical or mental competition. It has a final objective, it is played by one or more persons, according to a set of rules that every player needs to follow (Huizinga, 1995).

Serious games became more and more relevant in educational institutes as the availability of multimedia applications and advanced technology increased as time passed. Moreover, educational centers started recommending practicing serious games in spare time. As a result, the popularity of this games increased. Relying on this principle, the challenging nature of serious games led to a positive reaction towards learning while playing games (Riemer & Schrader, 2015). One of the most important aspects of learning with serious games is the learner's attitude towards this type of games.

The Sound of Vision system overview

The hardware components of the SoV device are (Figure 1): the headset, the haptic belt, the central processing unit and a remote control. The headset includes a Leopard Stereo Camera, an IMU (Inertial measurement unit), a Structure Sensor and a custom designed multi-speaker device (4 speakers/ear).

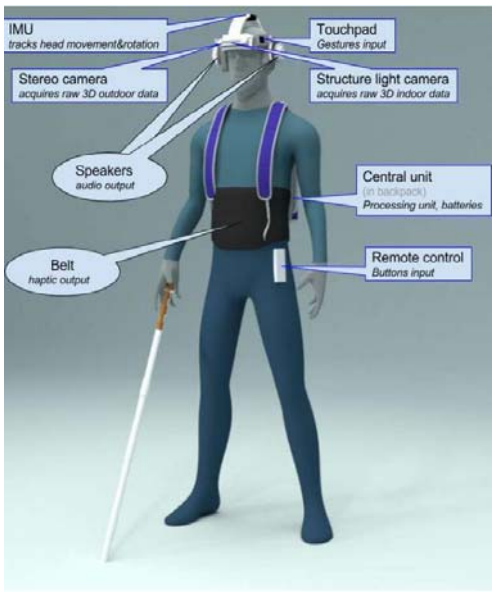

Figure 2. SoV hardware components

The process of assisting visually impaired people that has to be carried out by the SOV system consists of the following processes (Figure 2), performed continuously, as sequential steps in the order that they are listed below, in a real-time loop (enough times each second so that the representation is perceived as dynamic, without delays and in accordance to user movements and environmental changes):

3D information acquisition and 3D Model generation

The SOV system can be used in two different ways:

a) For assisting the user in navigating the real environment; in this case, the 3D information is acquired from the two acquisition systems: the stereo cameras and the structure sensor. The stereo cameras produce RGB images, which are used mostly for outdoor images analysis, while the structure sensor produces depth images that are used for indoor images analysis. The 3D scene model is obtained through more image processing steps: filtering, segmentation and objects labeling.

b) For training purposes; in this case, the 3D information is composed of synthetized frames produced by the VTE (Virtual Training Environments) sub-system.

The 3D model is composed of objects like: walls, doors, stairs, generic objects, holes in ground, texts and others, each of them characterized by specific geometric attributes in the camera space.

Encoding and conveying information to user

The 3D model is encoded using sonification and haptification models, and sent to the audio and haptic rendering devices of the system.

Special situations (missing ground, dynamic objects) are signaled to the user in a distinct way, using audio and haptic modalities, separated from the rendering of the regular information (Balan, et al., 2015).

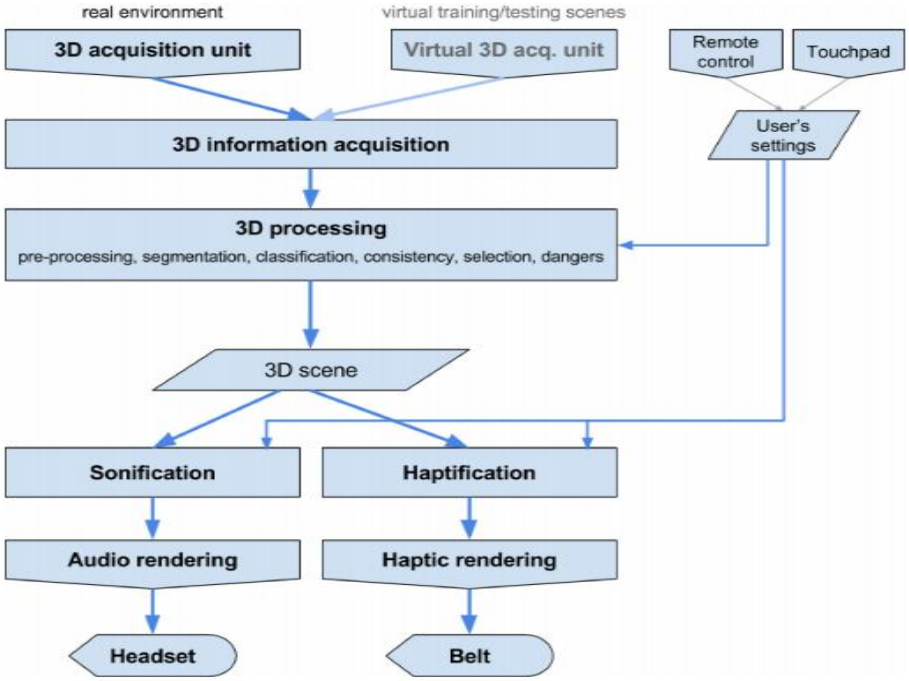

Figure 3. SoV Processing Pipeline

Sound of Vision gives an exceptional attention to training. The system aims to be usable and helpful to some extent even without advanced training, by the targeted visually impaired persons. Yet, it is expected that a good learning and adaptation of the users to its audio+haptic environment representation will require training procedures, to achieve efficiency in

usage.

For this, the device includes a special Training mode, consisting of a collection of serious games based on virtual environments of gradual increasing complexity.

The **SoV Runtime** is the main piece of software that contains the image processing algorithms and the audio and haptic encoding models. It is the core of SoV. The SoV Runtime interface (Figure 3) allows configuring the processing pipeline and displays information about the main modules of the system.

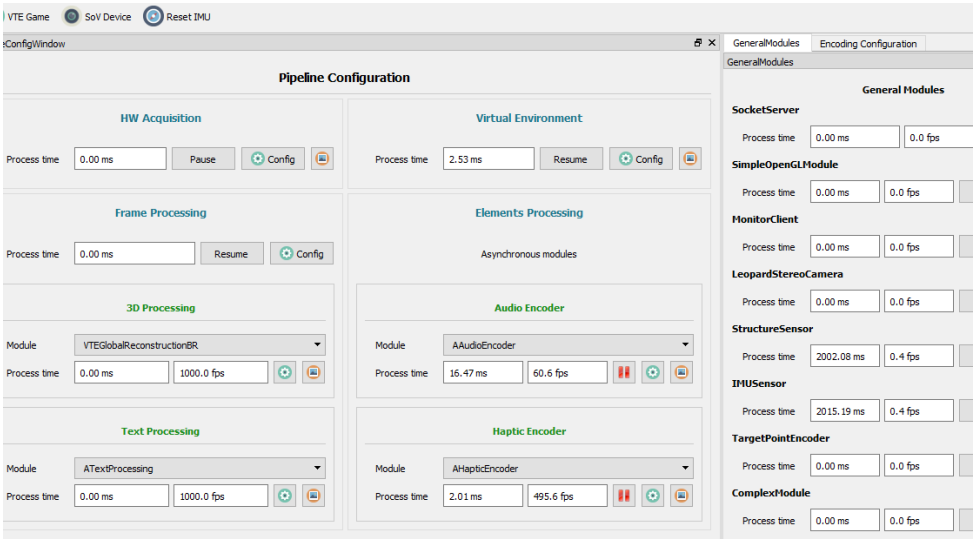

Figure 4. SoV Runtime interface

On the right side of the interface we can see information about each sensor: the Leopard Stereo Camera, the Structure Sensor and the IMU sensor. In the middle section, *Elements Processing*, we have the Audio and Haptic encoders where we can select different audio and haptic models. The models are responsible for encoding the visual stimulus into audio or haptic signals. The *Virtual Environment* section gives information regarding the stream between the VTE and SoV Runtime and we can record testing sessions. The other sections of the SoV runtime will not be discussed here. The VTE will be detailed in the following section.

Virtual Training Environments

The **VTE** is a 3D application developed using Unity 3D engine and its purpose is to train the user for real life situations gradually. The application uses TCP streaming to connect to the SoV Runtime and the configuration for the streamer is stored in an external file.

As external hardware, the VTE uses the SoV device and a joystick for the scenes in which the test taker needs to navigate. The joystick also includes inputs from the user like the buttons needed to answer questions and changing audio and haptic models.

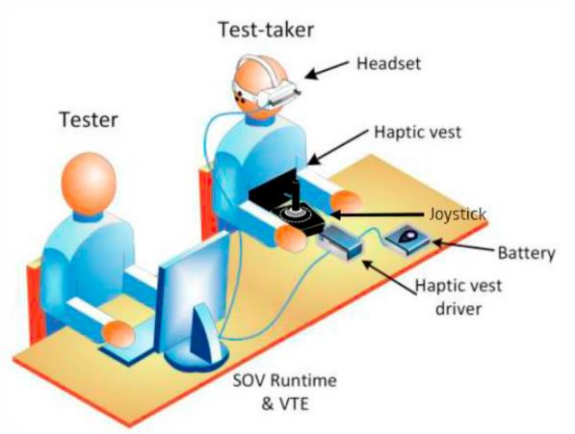

Figure 4. Environment settings tests

4.1. Unity 3D

The Unity 3D game engine (Unity - Game Engine, n.d.) is a viable solution for developing video games, VR and AR applications. The engine was developed by Unity Technologies company. It uses an in-house made rendering engine and the PhysX physics engine developed by NVIDIA. It also uses Mono, an open source implementation of the Microsoft .NET framework. Unity's pros worth mentioning are the following :

- **Documentation**: Unity 3D has a well-made documentation for the whole API in comparison with other game engines such as Unreal engine or Source, the latter having only partial documentation.
- **Developers community**: The online community for this engine is very open and active, offering help for beginners as well as

advanced developers. Another good reason to use this engine is the fact that the developers add features requested by the community.

- **Drag-and-Drop**: Unity's editor has one of the easiest to understand and to use visual interfaces in comparison to the above-mentioned engines. The content is organized in an arborescent hierarchy and it is added to the project in a drag-and-drop manner. Components can be added to each object. Components are scripts written in C#, JavaScript or Boo that should be organized on a behavior based manner. Certain premade components can be added such as physics components, render properties, etc. The scripts can add interactive behavior to the objects, they can create user interfaces or just manage information.
- **Physics & Rendering**: By using physics properties, the developer can add mass, springiness, bounciness and collision detection to objects. The physics properties are simulated using the NVIDIA PhysX engine, an engine used in many commercial games. Rendering properties include shader and texture assignment to objects. They are organized in materials and they use CG Programming as shader specific language. The used wrapper for shaders is ShaderLab. Unity's custom rendering engine uses a simplified shader language which is compiled into DirectX 9 or OpenGL 2.0 shaders depending on the target platform.
- **Multiplatform distribution**: Unity Editor can be used with OSX, Windows and Linux. It can export applications for other platforms such as Android, IOS, Windows, Web-Player etc.
- **Low Cost**: Unity has reduced costs for a complete package of features and compensates being superior to other free game engines on the market. The Personal version of this engine is free but it limits to companies with incomes lower than 100.000 $ yearly. The Plus and Pro versions have a 35$ and 125$ monthly fee per seat and offer other features that increase productivity.

4.2 Training program for mastering SoV

The VTE is designed to gradually train the user with all the audio and haptic models.

Training starts with basic understanding of the encoding models. The

first set of tests starts with a single object in the scene and the user is trained for basic understanding of the object, detecting width, distance, direction and elevation of a single object. Succeeding these tests are scenes with more than one object (3-5 objects). The user is asked to recognize properties of the scene like number of objects, the widest object and the closest object. The final tests in the current implementation are training the user for basic navigation, avoiding obstacles of different sizes to reach a target. Each set of the three main types of tests have *Learning*, *Practice* and *Testing* modes (Balan, Moldoveanu, Moldoveanu, & Dascalu, 2014), (Balan, Moldoveanu, Moldoveanu, & Morar, 2016), (Balan, et al., 2017).

The levels of training are organized in four categories:

- Single attribute
- Static scenes
- Basic navigation
- Advanced navigation

Firstly, the *Single attribute* scenes train the user on how the device models individual object properties like width, height, distance, direction, elevation and quantity. Each scene has three modes: learning, practice and test.

While in learning mode the test taker is given feedback for the current scene, in practice mode, the system asks questions regarding the scene and the user answers. After answering each question, the test taker receives feedback if the answer was correct or not. In testing mode, the user receives only questions without feedback. While the user trains in testing mode, statistics are saved.

In each of the *Single attribute* scenes the test taker is trained with one attribute at a time. The attributes tested are the following:

- Width: 40cm, 80cm or 120cm
- Height: 40cm, 80cm or 120cm
- Distance: 1-5m
- Direction: 30° left, 15° left, center, 15° right, 30° right
- Elevation: ground level, head level
- Quantity of objects: 3-5

For the *Static scenes,* the difficulty increases. We have two types of static scenes:

- Random
- Complex

In a Random scene, the test taker is presented with a single object with random properties. Similar to the previous scenes it has learning, training and testing modes. This scene trains the user to identify multiple object properties at once.

A Complex scene generates multiple objects with random properties. The user should identify some properties like the number of objects in the scene, which is the closest object and the widest object.

The *Basic navigation* tests come with situations closer to real situations. At this point the user is considered to be capable of avoiding objects and reaching a goal. The following scenes are included in the Basic navigation category:

- Reaching an object

An object is generated in front of the test taker. The goal is to get close to the object without touching it.

- Passing by an object

The object is generated in the user's view. The goal is to pass on either side of the object without touching it.

- Circling an object

The object is generated in the test taker's view. The goal is to circle the object on either side and reach the point where he first started.

- Passing between objects

Two to four objects are generated in the user's view. The goal is to pass between any two objects of the user's choice.

- Slalom

A number of three to five objects will be generated straight in front of the user. The goal is to pass alternatively between them until he reaches the last object.

- Pickups

Objects are spawned randomly in the scene and the user needs to collect the objects. This scene comes with three levels of difficulty depending on where the pickups spawn: in the view, in front of the player or all around the player.

- Walking by a wall

The user needs to walk by a wall to reach the goal located at the end of the wall maintaining a fixed distance from the wall.

- Room sides navigation:

The goal for this scene is to walk on the sides of a square room and return to the point where he started.

The last part of the training environment, the *Advanced navigation* scenes, uses the knowledge and training of the user's past experiences to accomplish specific tasks like finding an object, avoiding static obstacles and combining the audio and haptic models as he wishes.

The last three scenes are the following:

- Boxes navigation
- Frogger
- Asteroids

The Boxes navigation scene comes with three levels of difficulty. The obstacles are randomly generated and the level of difficulty consists of the space between obstacles.

Frogger is a scene in which the user constantly walks forward on three lanes from which two always have obstacles. The user needs to find the clear lane.

In the Asteroids scene, the user has a fixed position in the center and objects constantly generate around him. The difficulty gradually increases similar to the pickups scene. The goal is to target and "shoot" the obstacles that come towards him.

4.3 Training Scenes

Random scene

In a random scene (Figure 5), the user is presented with one object that has the following properties:

- Width: 40, 80 or 120 cm
- Height: 80 cm
- Direction: 30° left, 15° left, Center, 15° Right, 30° Right
- Distance: 1m, 2m, 3m, 4m, 5m from the user position
- Elevation: Ground level, Head level

In the *Learning* phase, the user can press enter and receive audio feedback for the object's properties. The user stays in this scene for a minute (configurable from the settings menu) before continuing to the

Practice mode.

The *Practice* mode of the scene has several trials (also configured in the settings menu) in which the player is presented with one object and the properties are randomized for each trial. In this mode, he receives the following questions:

- What is the object's width?
- What is the object's distance?
- What is the object's direction?
- Is the object on the ground or at head level?

After answering each question, the user receives audio feedback on the answer.

In *Testing* mode, the user needs to answer the same set of questions, he does not receive feedback on his answers and statistics of his performance are saved for later analysis.

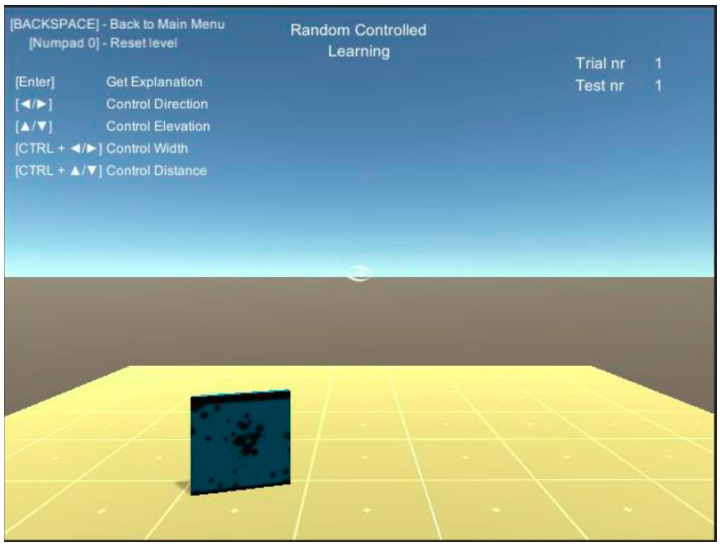

Figure 5. Random Scene

Complex scene

The scene (Figure 6) is generated using two to five objects with different dimensions that do not overlap with one another and are in close range to the user (1-5 m).

In *Learning* mode, the player is presented with several randomized scenes. By pressing <ENTER> he receives audio feedback regarding the

scene. The user stays in this scene for a minute before continuing to the Practice mode.

The *Practice* mode requires the user to answer the following set of questions:

- How many objects are in the scene?
- Which object is the closest?
- Which object is the widest?

This mode has a set number of trials and after answering the questions he gets feedback on his answer.

In *Testing* mode, the user needs to answer the same set of questions, he does not receive feedback on his answers and statistics of his performance are saved for later analysis.

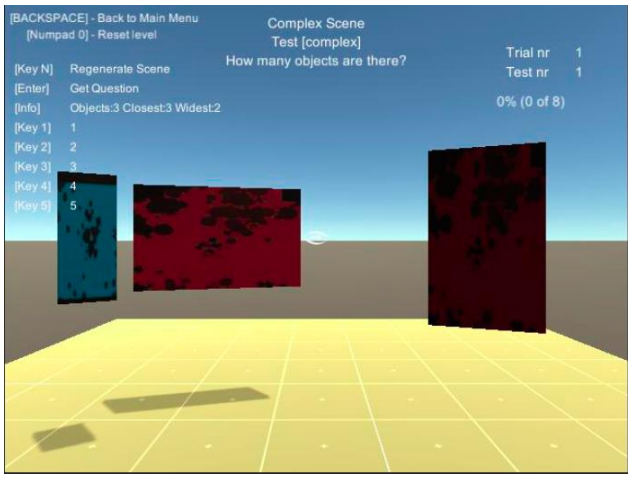

Figure 6. Complex Scene

Boxes scene

The Boxes scene (Figure 7) has a broader area, this being the *only scene* in which the player can move. The purpose of this scene is for the player to reach the target (seen in the background as a green object on the ground). By pressing <SPACE> the user hears a localized beeping sound coming from the target.

This scene only has *Practice* and *Testing* modes. The practice mode simply lets the user navigate freely to get used to the test that comes after a fixed number of successful trials (configured from the settings menu). During the testing mode statistics are stored.

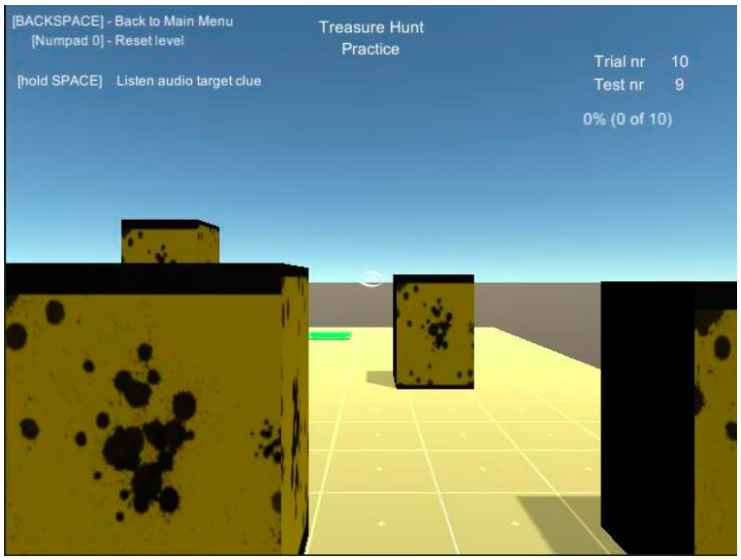

Figure 7. Boxes Scene

Conclusions

Developing virtual training environments proved to be very efficient and a safe step towards training the visually impaired people for using the SOV device. It is also efficient in tracking user progress during the training period. As it is useful regarding the previous aspects, it lacks a good reward system to keep the user entertained and, as a result, users tend to need time to rest after a few testing sessions. For future developments, we need to focus more on the audio feedback and other training exercises for example, a good and simple solution for keeping the user entertained is to give a sound effect periodically, whenever a test or training session has finished. Another solution could be adding a system that promotes competitivity between users.

ACKNOWLEDGEMENT

This work was supported by the European Union's Horizon2020 Research and Innovation Programme under grant agreement No 643636 "Sound of Vision".

References

(n.d.). Retrieved from Sound of Vision: https://soundofvision.net/category/publications/

(n.d.). Retrieved from Unity - Game Engine: https://unity3d.com/

Balan, O., Moldoveanu, A., & Moldoveanu, F. (2015, April 21). Navigational audio games: an effective approach toward improving spatial contextual learning for blind people. *ISSN*, pp. 110-114.

Balan, O., Moldoveanu, A., Moldoveanu, F., & Butean, A. (2015, July 12-16). Auditory and haptic spatial cognitive representation in the case of visually impaired people. *ICSV22*, pp. 3-6.

Balan, O., Moldoveanu, A., Moldoveanu, F., & Dascalu, M.-I. (2014, November 17-19). Audio Games – A novel approach towards effective learning in case of visually-impaired people. *ICERI*, pp. 6543-6546.

Balan, O., Moldoveanu, A., Moldoveanu, F., & Morar, A. (2016, April 21-22). From game design to gamification and serious gaming – how game design principles apply to educational gaming. *The 12th International Scientific Conference eLearning and Software for Education Bucharest*, pp. 335-338.

Balan, O., Moldoveanu, A., Moldoveanu, F., Nagy, H., Wersényi, G., & Unnþórsson, R. (2017, March). Improving the Audio Game-Playing Performances of People with Visual Impairments Through Multimodal Training. *Journal of Visual Impairment & Blindness*, pp. 148-152.

Balan, O., Moldoveanu, A., Nagy, H., Wersényi, G., Botezatu, N., Stan, A., & Lupu, R.-G. (2015, July). Haptic-auditory perceptual feedback based training for improving the special acoustic resolution of the visually impaired people. *ICAD*, pp. 22-23.

Bujacz, M., Kropidlowski, K., Ivanica, G., Moldoveanu, A., Saitis, C., Csapo, A., . . . Witek, P. (2016, July 6). Sound of vision-spatial audio output and sonification approaches. *ICCHP*, pp. 202-209.

Jóhannesson, Ó. I., Balan, O., Unnthorsson, R., Moldoveanu, A., & Kristjánsson, Á. (2016, 6 20). The Sound of Vision Project: On the Feasibility of an Audio-Haptic Representation of the Environment, for the Visually Impaired. *Brain Sci*, pp. 12-15.

.