Reusuable Virtual Training and Evaluation Solutions for Sensory Substitution Devices

Alin Moldoveanu
University POLITEHNICA of Bucharest
alin.moldoveanu@cs.pub.ro

Gabriel Ivañică
University POLITEHNICA of Bucharest
gabriel.ivanica@gmail.com

Silviu Ivaşcu
University POLITEHNICA of Bucharest
ivascu.silviu10@gmail.com

Maria-Iuliana Dascălu
University POLITEHNICA of Bucharest
maria.dascalu@upb.ro

Anca Morar
University POLITEHNICA of Bucharest
anca.morar@cs.pub.ro

Oana Bălan
University POLITEHNICA of Bucharest
oana.balan@cs.pub.ro

Iulia Ţănică
University POLITEHNICA of Bucharest
iulia.stanica@upb.ro

Florica Moldoveanu
University POLITEHNICA of Bucharest
florica.moldoveanu@cs.pub.ro

Daniel Flamaropol
University POLITEHNICA of Bucharest
daniel.flamaropol@gmail.com

ABSTRACT

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

CHI'19 Extended Abstracts, May 4-9, 2019, Glasgow, Scotland, UK.
© Position paper submitted to CHI 2019 Workshop on Hacking Blind Navigation. Copyright held by authors.
ACM ISBN 978-1-4503-5971-9/19/05.
DOI: https://doi.org/10.1145/3290607.XXXXXX
This paper presents the reusable virtual training and testing environments and tools developed within the Sound of Vision (SoV) project, which created a wearable sensory substitution device (SSD) that helps the blind and visually impaired people navigate unknown indoor and outdoor environments. Part of Sound of Vision core concept is the importance of intensive, diverse, and efficient training to achieve proficiency with any SSD. Throughout the project, there was also a focus on developing everything as modular and reusable (software) components: for distributed development, to increase adaptability to new hardware and new modalities of substitution, and finally to open source release some of these components, in order to support other SSDs. The complex environments and tools presented in this paper (already used and refined intensively with Sound of Vision) are developed as modular components and can be easily connected to other SSDs, supporting the training and evaluation of their users. By opening our library to the community, we will provide all other SSD researchers and developers with a powerful and efficient tool for training their subjects to achieve best results, as well as testing and assessing users’ performance.

INTRODUCTION

Sound of Vision [11] is a Horizon 2020 research project funded by the European Commission (EC) between 2015 and 2017 that involved partnership from universities, software companies and institutes for blind people from Iceland (coordinator), Romania (technical lead), Poland, Hungary and Italy. SoV received the Tech for Society Award during the ICT 2018 EC event. SoV created a smart device (Figure 1) composed of a headgear, a haptic vest and a wearable control unit. The headgear contains two cameras (structured light for indoor and stereo for outdoor) that continuously 3D scan the environment and transfer data to a processing unit, which applies 3D segmentation algorithms for detecting surrounding objects, ground, walls, stairs, holes, objects at head level, etc. These are presented to the user through multimodal sensory substitution - both auditory and haptic. Directional sound is delivered through ear-speakers, while haptic information is perceived on a belt placed on user’s abdomen, containing a matrix of 60 vibrating motors (6 rows x 10 columns). The audio-haptic rendering is accurately synchronized and conveys, for segmented objects, direction (azimuth and elevation), distance, height, width, and type. Two representation models have been designed – Iterative (objects are rendered sequentially) and Continuous (all objects rendered at once); more information about the models is provided in [1]. The SoV wearable device uploads users’ usage patterns and results to a central server. Substantial attention has been given to training and testing [2] [3] [4]. The system has been tested with blind and visually impaired users in both indoor and outdoor settings, with multiple scenarios ranging from very simple to realistic situations, and during various weather conditions. The SoV training approach is safe, gradual and reliable, as it is based on virtual environments of gradual intensity exposure. Our serious games help the subjects learn the audio and haptic encodings, as well as the modality in which the system works, in a progressive and immersive fashion that maximizes the effects of training [6][7][8].
VTTE offer the following features:

- Support for multiple user profiles
- Easy to configure testing sessions
- Basic on-screen statistics per user & scene
- Automatic full data uploading to server
- Advanced statistics through remote server
- Flexible design, easy to add new scenes, extend or customize the existing ones
- Fast data streaming with compression: stereo view, depth, object labels and unique IDs, player telemetry, any other custom data
- Integration with any local or remote application through specific TCP API/protocol.

In this paper we present the SoV serious games as a modular, extensible and reusable tool, helpful to any other developers of SSDs for blind and visually impaired people, who want to train their end-users and assess their performance. Furthermore, any improvements coming from other contributors may implicitly strengthen the SoV device itself.

VIRTUAL TRAINING AND TESTING ENVIRONMENTS

The Virtual Training and Testing Environments (VTTE) (Figure 2), implemented in Unity 3D C#, were useful tools during system refinement and evaluations with blind and visually impaired people. They will be released as an independent open-source product, including a software library, pre-compiled applications, and documentation, to support other SSDs researchers and developers. The license will ensure free availability for non-commercial purposes (research, charity, individual needs) and availability through a custom fee for commercial purposes.

The reusable design covers two aspects. Firstly, the VTTE can easily integrate with any SSD or application through the Streamer component (all virtual 3D data and game information can be streamed to any kind of application that implements protocol defined by the library). Secondly, the overall architecture allows to easily add new scenes and testing scenarios or adjust existing ones [5].

The VTTE integrates facilities to monitor and save user performance data, in real-time. All the information is stored into a local database and at the end of the session it is automatically uploaded to a central remote database. The following data is recorded: general events, answers to questions, user commands, in-game telemetry (position, body and head rotation), testing information such as timing, tasks, trials, completeness. A user-dashboard shows in real-time basic statistics and can be extended with custom charts or visualizations, while advanced statistics are available from server.

Virtual scenes can be instantiated into 3 different modes: Learning, Practice or Testing. For Learning, the training tasks are presented together with voice guidance and explanations, and the user can control some aspects of the scene and observe the effects. Practice trials are meant for pre-testing the knowledge learned previously and receive correct/wrong feedback, as well as additional details if necessary. The Testing is similar with the Practice trials, with the difference that no correct/wrong feedback is provided. In all 3 modes, user performance data is recorded.

The virtual scenes are grouped into 6 categories. First, the Encodings Scenes (with either 1 or 2 objects for which the user can control all properties: width, height, distance, direction, elevation, and placement) allow the user to learn how the encodings work. Then, the Single Attributes Scenes are intended to train and test the correct identification of a single object position, dimensions, and type. There are 7 scenes in total (Width, Height, Distance, Direction, Elevation, Quantity, Type) and each one of them has all the 3 modes activated: Learning, Practice and Testing. Then, the Static Scenes are intended to test multiple object properties during the same instance [9]. There are 3 types of scenes: 1 Random (same as the one from Encodings Scenes, but randomly generated), Pre-Complex and Complex. Further, the Basic Navigation Scenes provide tasks such as: Frontal pickups.
(the user has to move towards an object to collect it and many objects are spawned one after another), **Passing by an object** (the user must navigate by an object without hitting it), **Passing between objects** (2 or 3 spawned objects are placed on a horizontal line in front of the user). Then, the **Advanced Navigation Scenes** represent complex navigation tasks. The user must navigate the scene to a target area signaled by an auditory icon without colliding with obstacles. There are two types of scenes: **Boxes** (Figure 3, rectangular scenes randomly filled with obstacles) and **Corridor** (a corridor randomly filled with obstacles), each with 3 levels of difficulty.

Finally, the last set includes **3 gamified modules** – a **Realistic realm** (Figure 4), a **Fantasy realm**, and a collection of **Mini-games** (Figure 5). The realistic realm is a 3D environment which simulates a real-world setting, consisting of multiple scenes: the interior of a house, the exterior (part of a city), an office building interior, two universities and a mall. The user must complete a series of tasks, with increasing levels of complexity and difficulty. The environment is highly realistic, including real-size accurately modelled objects, sounds, and animations (such as walking people, cars, wind, rain, etc.).

The fantasy realm puts the user in a fantasy setup with attractive and ludic elements, where he/she has to complete a sequence of quests, with strong emotional immersion (letting the users take decisions and face the consequences), but actually designed to train user’s skills with the SSD. Finally, the 9 mini-games gamify aspects of object recognition and avoidance, with various levels of complexity and difficulty [10].

If needed, new scenes can be added, by extending a base class that controls integration in the system, user interaction and data collection. After the 3D elements and logic of the new scenes are defined, they are registered in a database, configured with various tasks and number of repetition trials for the 3 modes, Learning, Practice and Testing. Also, the main menu of the application can be easily configured by editing an XML file, which represents a hierarchy of sections and buttons. The **VTTE statistics server** is a component designed to analyze VTTE user progress in detail, through advanced statistics and visualizations. The interface has a user centric design. For each user an information dashboard is presented where detailed information about testing history, individual scenes scores, timing, events and encodings is presented, numerically and graphically (Figure 6).

**CONCLUSIONS**

During the development of the Sound of Vision system, the consortium focused on the importance of sharing software resources within the scientific community that is preoccupied with the design of assistive devices for the visually impaired people. Thus, we presented here the reusable training and testing virtual environments utilized during our research and our plan to publish the API as open-source, in order to encourage and stimulate the area of SSDs development. Fellow researchers in this area will thus have access to our software for users’ training, testing and assessment, with any SSD, adapt it to their needs and contribute with new features to its continuous development.
REFERENCES


