Universitatea Tehnică Gheorghe Asachi din Iaşi,
Facultatea de Automatică şi Calculatoare

Lista de lucrări considerate relevante

1. S.-D. Achirei, M.-C. Heghea, R.-G. Lupu, and V.-I. Manta, "Human Activity Recognition for Assisted Living Based on Scene Understanding," Applied Sciences, vol. 12, no. 21, p. 10743, Oct. 2022, doi: 10.3390/app122110743.

2. O. Zvoris, teanu, S. Caraiman, R.-G. Lupu, N. A. Botezatu, and A. Burlacu, "Sensory Substitution for the Visually Impaired: A Study on the Usability of the Sound of Vision System in Outdoor Environments," Electronics, vol. 10, no. 14, p. 1619, Jul. 2021, doi: 10.3390/electronics10141619.

3. R.-G. Lupu, O. Mitrut, , A. Stan, F. Ungureanu, K. Kalimeri, and A. Moldoveanu, "Cognitive and Affective Assessment of Navigation and Mobility Tasks for the Visually Impaired via Electroencephalography and Behavioral Signals," Sensors, vol. 20, no. 20, p. 5821, Oct. 2020, doi: 10.3390/s20205821.

4. Lupu, R.G., Ungureanu, F., Ferche, O., Moldoveanu, A. (2020). Neuromotor Recovery Based on BCI, FES, Virtual Reality and Augmented Feedback for Upper Limbs. In: Guger, C., Allison, B.Z., Miller, K. (eds) Brain–Computer Interface Research. SpringerBriefs in Electrical and Computer Engineering. Springer, Cham. https://doi.org/10.1007/978-3-030-49583-1_8

5. Moldoveanu, O.-M. Ferche; F. Moldoveanu; R. -G. Lupu; D. Cintez˘a; D. -C. Irimia; C. Toader, "The TRAVEE System for a Multimodal Neuromotor Rehabilitation," in IEEE Access, vol. 7, pp. 8151-8171, 2019, doi: 10.1109/ACCESS.2018.2886271.

6. R.-G. Lupu, D.-C. Irimia, F. Ungureanu, M.-S. Poboroniuc, A. Moldoveanu, "BCI and FES Based Therapy for Stroke Rehabilitation Using VR Facilities", Wireless Communications and Mobile Computing, vol. 2018, Article ID 4798359, 8 pages, 2018. https://doi.org/10.1155/2018/4798359

7. R. G. Lupu, N. Botezatu, F. Ungureanu, D. Ignat and A. Moldoveanu, "Virtual reality based stroke recovery for upper limbs using leap motion," 2016 20th International Conference on System Theory, Control and Computing (ICSTCC), Sinaia, Romania, 2016, pp. 295-299, doi: 10.1109/ICSTCC.2016.7790681.

8. M. Dascalu, A. Moldoveanu, O. Balan, R. G. Lupu, F. Ungureanu and S. Caraiman, "Usability assessment of assistive technology for blind and visually impaired," 2017 E-Health and Bioengineering Conference (EHB), Sinaia, Romania, 2017, pp. 523-526, doi: 10.1109/EHB.2017.7995476.

9. S. Caraiman, A. Stan, N. Botezatu, P. Herghelegiu, R. G. Lupu and A. Moldoveanu, "Architectural Design of a Real-Time Augmented Feedback System for Neuromotor Rehabilitation," 2015 20th International Conference on Control Systems and Computer Science, Bucharest, Romania, 2015, pp. 850-855, doi: 10.1109/CSCS.2015.106.

10. R. G. Lupu, F. Ungureanu and V. Siriteanu, "Eye tracking mouse for human computer interaction," 2013 E-Health and Bioengineering Conference (EHB), Iasi, Romania, 2013, pp. 1-4, doi: 10.1109/EHB.2013.6707244.

applied sciences

MDPI

Article

Human Activity Recognition for Assisted Living Based on Scene Understanding

Stefan-Daniel Achirei *, Mihail-Cristian Heghea, Robert-Gabriel Lupu and Vasile-Ion Manta

Department of Computer Engineering, Faculty of Automatic Control and Computer Engineering, "Gheorghe Asachi" Technical University of Iasi, 700050 Iasi, Romania
* Correspondence: stefan-daniel.achirei@academic.tuiasi.to

Abstract: The growing share of the population over the age of 65 is putting pressure on the social health insurance system, especially on institutions that provide long-term care services for the elderly or to people who suffer from chronic diseases or mental disabilities. This pressure can be reduced through the assisted living of the patients, based on an intelligent system for monitoring vital signs and home automation. In this regard, since 2008, the European Commission has financed the development of medical products and services through the ambient assisted living (AAL) program—Ageing Well in the Digital World. The SmartCare Project, which integrates the proposed Computer Vision solution, follows the European strategy on AAL. This paper presents an indoor human activity recognition (HAR) system based on scene understanding. The system consists of a ZED 2 stereo camera and a NVIDIA Jetson AGX processing unit. The recognition of human activity is carried out in two stages: all humans and objects in the frame are detected using a neural network, then the results are fed to a second network for the detection of interactions between humans and objects. The activity score is determined based on the human–object interaction (HOI) detections.

Keywords: assisted living; home automation; image processing; neural networks; public health care

Citation: Achirei, S.-D.; Heghea, M.-C.; Lupu, R.-G.; Manta, V.-I. Human Activity Recognition for Assisted Living Based on Scene Understanding. *Appl. Sci.* **2022**, *12*, 10743. https://doi.org/10.3390/app122110743

Academic Editor: António J. R. Neves

Received: 20 September 2022
Accepted: 21 October 2022
Published: 24 October 2022

1. Introduction

Technical solutions for ambient assisted living allow continuous monitoring of vital parameters of the elderly and home automation [1,2]. Such solutions for assisted living are particularly important in the context of an accelerated aging European society: it is estimated that by 2060, one in three Europeans will be over 65 years of age, while the ratio of active people to retired people will decrease from 4:1 currently to 2:1 [3]. Moreover, in the coming years, Europe will face more than 2 million vacancies in the health and social care system.

Monitoring the environmental and physiological parameters of elderly people who suffer from chronic diseases or mental disabilities has become a topic of real interest in recent years. Both the scientific community and industry have proposed various assisted living systems for indoor environments.

At the commercial level, countless assistive devices are available, such as those for administering medications, detecting falls, alarm buttons, monitoring vital parameters, as well as complex solutions for recording and tracking environmental parameters [4]. However, the solutions that integrate multiple such devices in assistance applications for autonomy at home have not yet known a commercial spread at the level of end users. Most of the integrative solutions developed so far are prototypes described in the scientific literature [5–9]. The main barrier to building such solutions is that most of these devices come with proprietary applications, use custom communication protocols and do not expose programming interfaces, which makes them impossible to integrate into third-party applications.

The scientific literature provides many examples of complex and innovative systems for monitoring environmental and physiological parameters. For example, Shao et al. [10]

determines the values of physiological parameters (respiratory rate, exhaled air flow, heart rate, pulse) using images from a web camera. Marques and Pitarma [11] describe a system dedicated to environmental monitoring of air temperature, humidity, carbon monoxide and luminosity to assist people inside buildings. For this, a microsensor infrastructure is used for data acquisition, as well as Arduino, ESP8266 and XBee for data processing and transmission using Wi-Fi. The recorded data are evaluated to determine the air quality in the rooms, anticipating technical interventions. A system that analyzes the change of temperature in the environment, as well as physiological indices (temperature, breathing rate, electroencephalogram, electrocardiogram, using Japanese medical equipment from Nihonkohden Co.), is proposed in [12]. With this system, the relationships between ambient temperature, comfort sensations, sleepiness and physiological indicators were determined.

Regarding integrated applications for home assistance using ICT, research efforts have focused on formulating solutions based on the combination of technologies specific to smart home digital platforms, IoT, artificial intelligence and cloud computing. The HABITAT platform [6] incorporates these technologies into everyday household objects: an armchair for monitoring sitting posture, a belt for extracting movement information, a wall panel and mobile phone for user interface. An artificial intelligence component is used for the system to react to various specific events, trigger notifications and receive feedback from the user. However, the radio frequency localization used requires the person to wear a tag, the motion monitoring belt can be uncomfortable and difficult to wear for long periods of time. In addition, the system is not easily scalable, reducing to the implementation of a single usage scenario with specific devices. A significant limitation is the specific character of these integrated systems produced to date—they address limited monitoring and application scenarios.

A very relevant research and development topic is addressing the implementation and interoperability issues encountered in the adoption of life assistance technology. In this sense, possible solutions were formulated for managing the integration of devices and aggregating information from them [7,8]. However, the proposed solutions require significant integration efforts or the involvement of device manufacturers in a common standard alignment process. A relevant effort in this sense is represented by ACTIVAGE [9], a pilot project for intelligent living environments with the main goal of building the first European eco-system. It will have nine implementation locations in seven European countries, reusing and scaling public and private IoT platforms. New interfaces necessary for interoperability are being integrated between these platforms that will implement IoT solutions for active and healthy aging on a large scale. The project will deliver AIOTES (AC-TIVAGE IoT Ecosystem Suite), a set of mechanisms, tools and methodologies for interoperability at different levels between IoT platforms and a free working method for providing their semantic interoperability.

A monitoring system designed for a health care environment has the following components: data exchange between devices, data storage, data processing to determine environmental conditions and physiological aspects, data security and confidentiality, as well as data access to them [1,2]. According to Sanchez [13], there are two types of information: unprocessed data from the sensors and context data used to determine behavior patterns or human activity. The context is important because it helps to evaluate environmental conditions and the health of the person monitored. The system must provide a way to represent the environment. Some solutions only have a 2D map [14,15], while others use 3D models [16,17] from which essential information can be extracted (door and window locations, pieces of furniture, gas sources, etc.) and generate positional alerts. In addition to processing sensor data, a monitoring system must integrate communication technologies in order to provide real-time health services that are aligned with the context and the real needs [18]. In an IoT system, nodes containing sensors are used to monitor the user, to collect data which is subsequently sent to a network of nodes.

The pressure put on the health insurance systems by care services for the elderly or people with special needs is increasing. Therefore, an automated system that can assist a person in his environment would be less expensive. Such a system is based on different types of sensors, independent or within a network. L. Malasinghe, N. Ramzan and K. Dahal in [19] review monitoring systems that use different sensors, such as cardiogram sensors for heart rate, sensors for breath measurements, for blood pressure, and for body temperature,. This paper presents a monitoring system for daily activities based on Visual Scene Understanding.

This paper is organized as follows: in Section 2, different approaches for human activity recognition are presented. In Section 3, we will find a short introduction to the SmartCare system, after which we discuss the integration of the HAR module into the project. In this section are also presented the hardware architecture, the monitoring scenarios for various medical conditions and the results on object detection and HOI detection with the selected models. In Section 4, the datasets used for object detection as well as for HOI detection are presented. In Section 5, we shortly present de HAR module architecture, while in Section 6 the design and implementation of the HAR module is presented; a comparison between the tested object detectors in the context of human–object interaction and the use of the HAR results within the SmartCare are also discussed here. Finally, we draw some conclusions in Section 7.

2. Related Work

An approach to recognizing daily activities based on imaging and neural networks is presented by M. Buzzelli, A. Albé and G. Ciocca in [20]. In the first part, the training and testing subsets are defined and the activity groups are decided based on three properties: duration, type and position. Each group is divided into two: by duration of activities, long or short; by type, dangerous or common; and by position, static or mobile. Based on the above classifications, there is a grouping by status, alarming situations and daily activities. The approach has two steps: detect the person using 'Faster R-CNN' [21] and then recognize the action by using two neural networks: I3D [22] and DeepHAR [23]. The obtained accuracy is 97% on status activities, 83% on alarming situations and 71% on daily activities.

Another image-based approach is described in [24] and uses depth images and thermal images to maintain confidentiality, as these images do not retain details. The dataset was taken from an elderly person's home for one month. The defined classes are: the person sleeps, sits on a bed/chair, stands, walks, uses the nightstand and needs assistance from a caretaker. In addition to the six activities, the background was also noted: when the person is not in the room. Having two streams of images, the authors trained two models and decided to merge them into one [25,26]. The deployed models use the ResNet-34 architecture [27], the average accuracy on the six activities and the background is: 94% on thermal images, 93.2% on depth images, 91.8% using early fusion and 95.8% using delayed fusion.

C. Alexandros et al. [28] presents an activity monitoring system using RGB images or data from an RGB-D, which provides 3D information about the objects in the scene. This system can recognize some basic activities: when a person is standing, sitting on a chair/sofa, walking or falling. The system uses RGB images, and the person's outline is extracted by a background removal algorithm. The system uses a model based on a set of key positions [29] and the most representative body position for each activity class is learned. When the RGB-D sensor data are used, the silhouette of the human body is easier to obtain. The body position is extracted and a generic algorithm is used to select the optimal joints in recognizing activity [30].

In paper [31], the activity monitoring is carried out by means of depth sensors. The images are processed to extract the silhouette and then the human skeleton and its joints. Based on the joints, they compute three features in order to train a hidden Markov model: centroid points, joint distances, and joint magnitude. For each activity, there is a hidden Markov model trained, the obtained results have an accuracy of 84.33%.

V. Vishwakarma, C. Mandal and S. Sural in [32] use a method based on human detection by adaptive background removal and then extract the characteristics by which it will be detected whether a person has fallen. The extracted features are the aspect ratio between the height and the width of the bounding box that fits the person, gradients on X and Y, the fall angle. If the action takes place outside and there are more people, the accuracy is 79%, the sensitivity is 54%, and the specificity is 97%. However, if there is only one person involved, both in outdoor and indoor environments, accuracy, sensitivity, and specificity are 100%.

A different approach in detecting human actions based on videos is presented by Deepmind (Google) in [33,34]. A 700 classes video dataset is used to train the I3D neural network. The model was trained and deployed on a machine with 32 P100 GPUs. The top-1 accuracy on the test set is 57.3%.

A team of researchers from MIT-IBM AI Lab propose a dataset [35,36] of one million short videos for dynamic events which take place over a period of time no longer than three seconds. 339 verbs are associated with over 1000 videos each. Results from three different models are combined resulting in a 31.16% top-1 recognition score.

3. Human Activity Recognition in the SmartCare System

The ICT solution for assisted living proposed in [37] is improving living conditions of the elderly and/or people with chronic disorders through intelligent automation of the environment (home) and monitoring their vital parameters. Home care assistance provides independent living for the categories of people mentioned above. The SmartCare system provides a platform for monitoring, automation, and an alerting protocol in case of life-threatening events. It is a system based on heterogeneous IoT devices which offers access to data services and performs offline analysis using artificial intelligence. The Ambient assisted living system integrates software utilities in order to configure and adapt it to the needs of the beneficiary and make it easily scalable for the designer and the integrator.

To achieve maximum user satisfaction, the SmartCare project [37] is based on a user-centric design that covers all the needs and preferences of the beneficiary, integrator or system administrator.

As described in [37], the ambient assisted living system consists of three main parts:

- Gateway: Implements the communication and interfaces with the installed telemedicine devices, actuators, as well as with the video monitoring component;
- Expert System: Cloud service that implements the intelligent processing of information from the sensor network and other stand-alone components by defining and following predefined monitoring and alerting rules;
- DeveloperUI: Graphical user interface that facilitates the design of solutions for specific home care applications. The patient's needs differ from case to case, according to the medical conditions.

In Figure 1, a conceptual architecture of the SmartCare system is presented with the components mentioned above. Among the devices layer it is integrated as well the human activity recognition component using video monitoring, sub-system which communicates with the Gateway.

The purpose of HAR is to identify the patient's life-threatening events such as falls, fainting or immobility, as well as activities which have to be regularly carried out in order to avoid critical situations, for instance people with diabetes have to serve the meal after a schedule and hydrate continuously. Failure to detect the patient for a longer period of time than a predetermined threshold and the detection of specific activities will generate warnings or notification messages to the caretaker.

The device network is composed of sensors belonging to three main categories: vital parameters (heart rate wristband, glucose meter, blood pressure monitor, body thermometer, pulse oximeter, etc.), home automation (smart switch, smart plug, smart lock, panic button, smart lightbulb, etc.) and physical activity (video monitoring and accelerometer—steps, fall, burned calories). Figure 2 illustrates the human activity recognition module within the bigger SmartCare system.

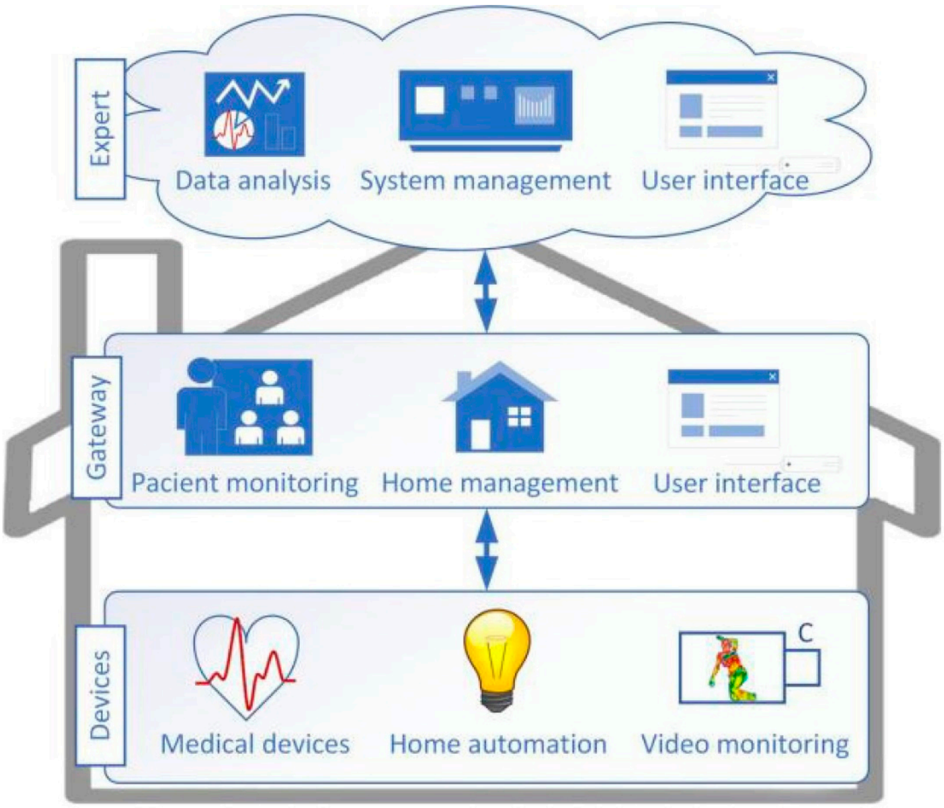
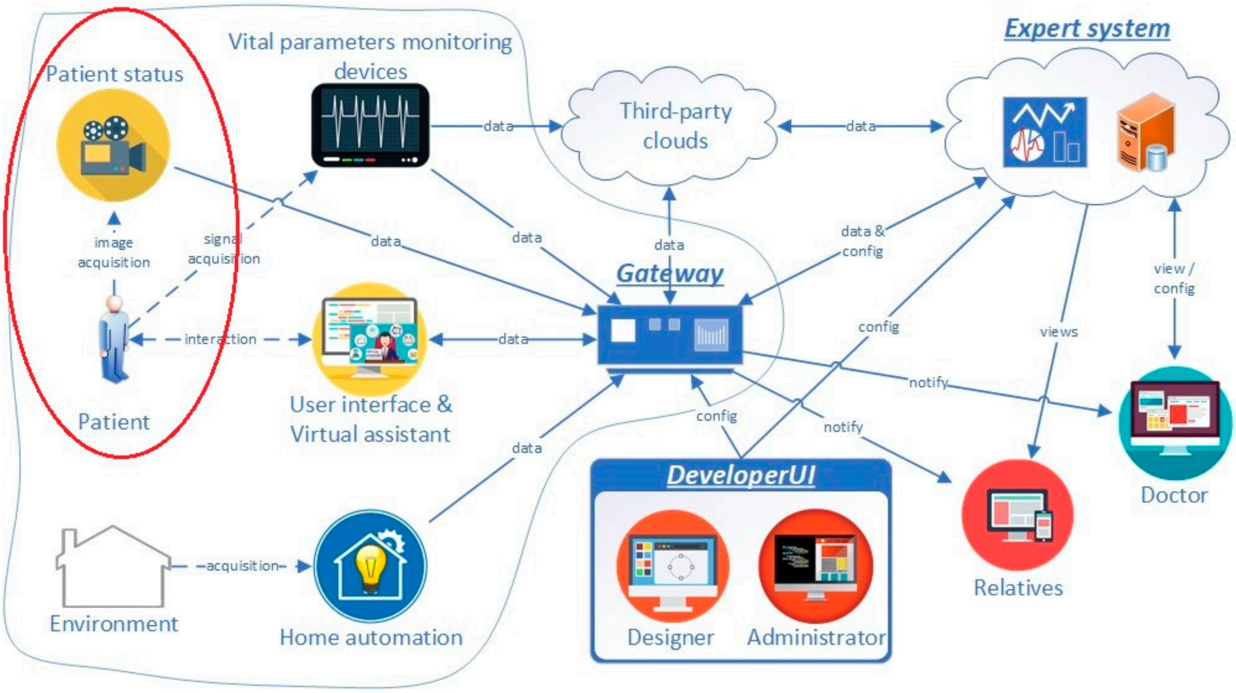


Figure 1. Conceptual architecture of the SmartCare system [37].

Figure 1. Conceptual architecture of the SmartCare system [37].



as falls, in order eal after eriod of generate es: vital ermom- k, panic meter— module

Figure 2. The human activity recognition module integrated in the SmartCare system.

Figure 2. The human activity recognition module integrated in the SmartCare system.

The functional diagram of the system from the perspective of the proposed HAR module is presented in Figure 3: the ZED stereo camera is connected via USB 3.0 to the Jetson unit; the events (activities of interest for the specificity of the monitored person or a life-threatening situation) are sent over MQTT to the Gateway. It is necessary to install the SDK provided in order to make use of all the image post-processing facilities of the ZED sensor. For reasons of modularity but also to relieve the gateway component of the

Figure 2. The human activity recognition module integrated in the SmartCare system.

The functional diagram of the system from the perspective of the proposed HAR module is presented in Figure 3: the ZED stereo camera is connected via USB 3.0 to the Jetson unit; the events (activities of interest for the specificity of the monitored person or a life-threatening situation) are sent over MQTT to the Gateway. It is necessary to install the SDK provided in order to make use of all the image post-processing facilities of the ZED sensor. For reasons of modularity but also to relieve the gateway component of the video processing, this task is implemented on a System-on-Chip platform, Nvidia Jetson Xavier AGX. Object detection, human–object interaction detection and activity detection are also performed on Jetson.

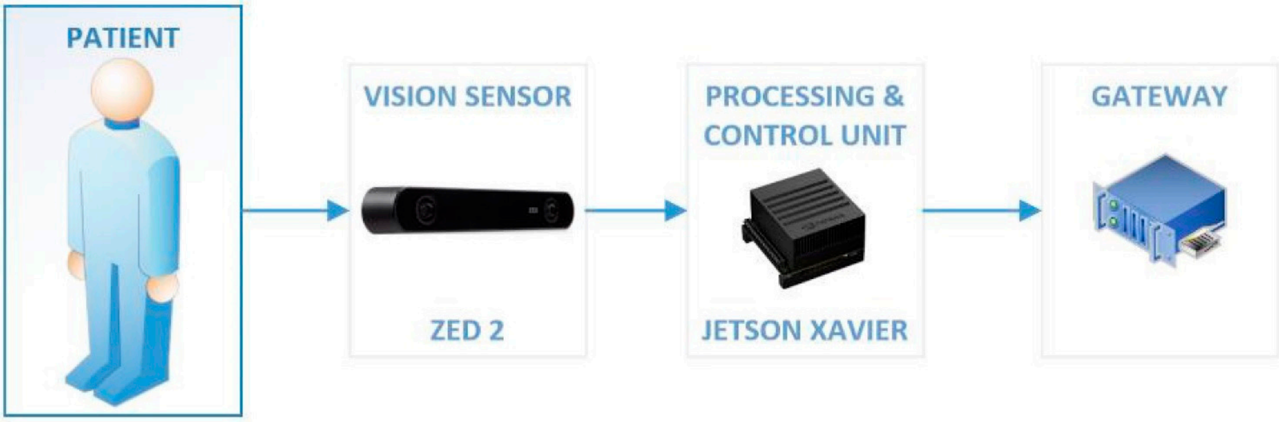


Figure 3. HAR system diagram.

3.1.1. Monitoring Scenarios

In the following three tables are presented the activities that the HAR module is monitoring, according to the specific medical condition of the patient. Table 1 presents the monitoring scenarios for patients with diabetes. For this condition, the focus is on physical activity, the eating schedule and the adequate hydration of the patient. The scenarios for monitoring patients with Alzheimer's disease are presented in Table 2: due to the condition, the person could do an activity repetitively or might put himself/herself in danger by forgetting various household appliances in operation, such as the oven, the cooker or the tap water. Table 3 summarizes the monitoring scenarios for patients with arthritis for which it is recommended to monitor physical activity and eating habits.

Table 1. Monitoring scenarios for diabetes.

Monitored Activity	Object Classes	Verbs	Resulted Triplets <Human, Verb, Object>
The patient is active	Person, chair, couch, bed	To run, to sit, to lay	<human, sits on, chair> <human, lays on, bed>
The patient serves the meal (after a diet recommended by the doctor)	Person, dining table, pizza, banana, apple, sandwich, orange, broccoli, carrot, hot dog, donut, cake, fork, knife, spoon, bowl, plate, oven, microwave, toaster, fridge, chair	To eat, to hold, to cut, to catch, to sit, to carry	<human, eats, sandwich> <human, eats, broccoli> <human, sits on, chair> <human, holds, fork> <human, eats, pizza>
The patient drinks liquids	Person, wine glass, bottle, cup	To drink, to hold, to carry	<human, drinks from, cup> <human, holds, bottle> <human, drinks from, wine glass>

Table 2. Monitoring scenarios for Alzheimer's disease.

Monitored Activity	Object Classes	Verbs	Resulted Triplets <Human, Verb, Object>
The patient serves the meal (too often due to the condition, eats too much)	Person, food (different types of food: pizza, banana, apple, sandwich, hot fog, etc.), dining table, fork, knife, spoon, bowl, oven, toaster, fridge, microwave	To eat, to sit, to hold, to cut, to carry	<human, eats, donut> <human, carry, fork> <human, holds, bowl> <human, cuts with, knife> <human, holds, spoon>
The patient drinks liquids (too often)	Person, wine glass, bottle, cup	To drink, to hold, to carry	<human, drinks from, cup> <human, holds, bottle>
The patient opens the tap (and may forget it opened)	Person, sink	To stand, to hold, to point	<human, stands, -> <human, points at, sink>
The patient opens the gas while cooking (and may forget it opened)	Person, oven, bowl	To hold, to carry, to point	<human, holds, bowl> <human, points at, oven> <human, carry, bowl>

Table 3. Monitoring scenarios for arthritis.

Monitored Activity	Object Classes	Verbs	Resulted Triplets <Human, Verb, Object>
The patient is active	Person, chair, couch, bed	To run, to sit, to lay	<human, runs, -> <human, lays on, bed>
The patient serves the meal (after a diet recommended by the doctor)	Person, dining table, pizza, banana, apple, sandwich, orange, broccoli, carrot, hot fog, donut, cake, fork, knife, spoon, bowl, plate, oven, microwave, toaster, fridge, chair	To eat, to hold, to cut, to catch, to sit, to carry	<human, holds, spoon> <human, holds, bowl> <human, sits at, dining table> <human, eats, apple> <human, eats, cake>
The patient drinks liquids	Person, wine glass, bottle, cup	To drink, to hold, to carry	<human, holds, bottle> <human, carry, cup> <human, drinks from, wine glass>

3.2. Object Detection and Human–Object Interaction Detection in HAR

Due to the fact that the HAR system needs to be deployable at the patient's home on a mobile platform and integrated as a stand-alone sensor, only lightweight models were taken into consideration for SmartCare. Another constraint that must be fulfilled is given by the necessity to detect the interaction of the patient with specific objects of interest (knife, spoon, bottle, oven, chair, etc.); therefore, only solutions based on object detection are suitable. Of all datasets available for human activity recognition used by the solutions presented in Section 2, the ones designed for object detection are the easiest to enhance with new classes. This aspect was also taken into account when the HAR architecture was designed for SmartCare because new classes will be needed in the near future (different type of medicine, wearables such as the insulin pump, fitness wristband, and blood pressure).

Taking into account the limited hardware resources and compatibility restrictions, only the neural networks with a custom architecture for mobile platforms with limited computing power were considered, such as MobileNet v1 and v2 [38,39], Inception [40], or YOLO [41]. The following implementations have been chosen for the four basic architectures:

1. Jetson Inference [42] for SSD-MobileNet v1, SSD-Mobilenet v2 and SSD-Inception V2;
2. Darknet [43] for YOLO v4 and YOLO v4 Tiny.

Jetson Inference [42] is a library written by Nvidia that offers implementations not only for the three neural networks listed above, it provides a wide range of networks for image recognition (ImageNet), object detection (DetectNet), semantic segmentation (SegNet) and pose estimation (PoseNet). The object detection networks are trained on the COCO dataset [44], while the semantic segmentation networks are trained on Cityscapes [45,46], DeepScene [47], Multi-Human [48], Pascal VOC [49] or SUN RGB-D [50].

YOLO v4 and YOLO v4 Tiny [41,43] are not the only open-source network architectures developed by the creators of Darknet [51]. In addition to these, pre-trained models for YOLO v3 [52] or v2 [53] are also available. Just like the Jetson Inference object detector, Darknet uses the COCO dataset.

To understand the context in a scene, we need to recognize how humans interact with objects in the environment. A scene can be further understood by detecting human–object interactions (HOI). Once again, the limited computing power on the used mobile platform restricts us to use the smallest neural network that runs in real time for human–object interaction detection.

In paper [54], the researchers propose a neural network for HOI detection. The model was validated on two public datasets for HOI detection: Verbs in COCO [55] (V-COCO) and Humans Interacting with Common Objects (HICO-DET) [56,57].

We used the V-COCO dataset [55] to evaluate iCAN [54] results for HOI detection. Using the single-thread sequential configuration for the evaluation tests, in Table 4, HOI detection accuracy increases with object detection accuracy. AP is calculated for the Agent (human) and for the Role (object/instrument). Two confidence thresholds were used for the evaluation of HOI detection: 0.5 and 0.2. Gupta Saurabh and Malik Jitendra define in [55] two scenarios for role AP evaluation:

- Scenario 1 (S1): in test cases with missing annotations for role a prediction for agent is correct if the action is correct and the person boxes overlap is >0.5 and the corresponding role is empty. This evaluation scenario is fit for missing roles due to occlusion.
- Scenario 2 (S2): in test cases with missing annotations for role, a prediction for agent is correct if the action is correct and the person boxes overlap is >0.5 (the corresponding role is ignored). This evaluation scenario is fit for cases with roles outside the COCO classes.

Table 4. HOI detection evaluation for agent and role.

Object Detector	Average Agent AP	Average Role AP [S1]	Average Role AP [S2]	Average Agent AP	Average Role AP [S1]	Average Role AP [S2]
	Threshold = 0.2			Threshold = 0.5		
SSD-Mobilenet-v2 [39,42]	25.78%	11.75%	12.99%	26.43%	13.43%	14.80%
YOLO v4 tiny [41,43]	26.62%	13.41%	15.09%	26.79%	14.92%	16.70%
SSD-Inception-v2 [40,42]	29.80%	14.26%	15.94%	30.02%	14.78%	17.56%
YOLO v4 [41,43]	56.68%	37.77%	43.20%	57.05%	40.07%	45.72%

The more objects are detected in the scene and the higher their confidence factor is, the more HOIs are detected with a higher confidence. Figure 4 shows the activities detected by iCAN using different object detectors.

In order to evaluate the HOI detector we keep track of some of the most relevant actions: to eat, to drink, to sit on, to stand, to hold an object/instrument, to lay on. A comparison of accuracies of the actions mentioned above using the four selected object detectors is presented in Table 5.

Table 4. HOI detection evaluation for agent and role.

Object Detector	Average Agent AP	Average Role AP [S1]	Average Role AP [S2]	Average Agent AP	Average Role AP [S1]	Average Role AP [S2]
	Threshold = 0.2			Threshold = 0.5		
SSD-Mobilenet-v2 [39,42]	25.78%	11.75%	12.99%	26.43%	13.43%	14.80%
YOLO v4 tiny [41,43]	26.62%	13.41%	15.09%	26.79%	14.92%	16.70%
SSD-Inception-v2 [40,42]	29.80%	14.26%	15.94%	30.02%	14.78%	17.56%
YOLO v4 [41,43]	56.68%	37.77%	43.20%	57.05%	40.07%	45.72%

Considering the results for human-object interaction detection (Table 4) and the detection accuracy for the most relevant actions (Table 5), the best improvement using YOLO v4 as object detector.

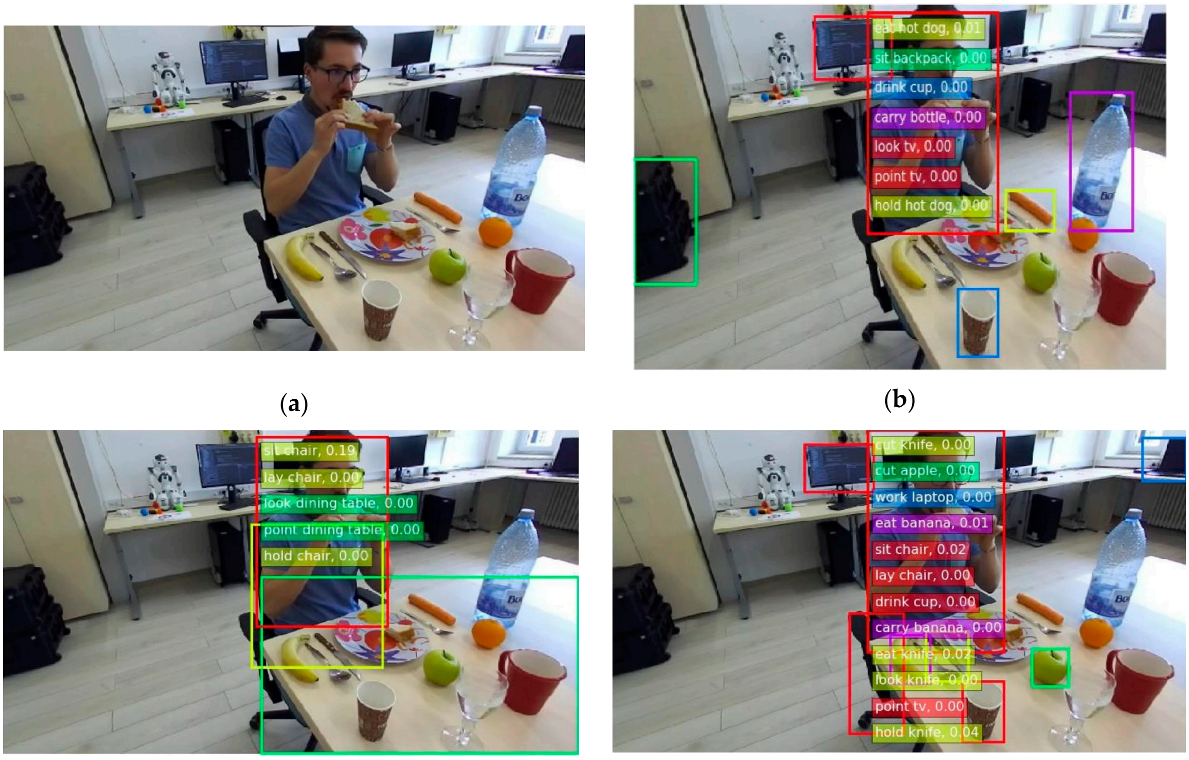


Figure 4. (**a**) Input frame; (**b**) HOI detections using YOLO V4 tiny; (**c**) HOI detections using SSD-MobleNet V2; (**d**) HOI detections using YOLO V4.

Figure 4. (**a**) Input frame; (**b**) HOI detections using YOLO V4 tiny; (**c**) HOI detections using SSD-MobleNet V2; (**d**) HOI detections using YOLO V4.

In order to evaluate the HOI detector we keep track of some of the most relevant actions: to eat, to drink, to sit on, to stand, to hold an object/instrument, to lay on. A comparison of accuracies of the actions mentioned above using the four selected object detectors is presented in Table 5.

Table 5. HOI detection evaluation for the most relevant actions.

Detected Activity	SSD—Mobilenet—v2	YOLO v4 Tiny	SSD—Inception—v2	YOLO v4
To eat	4.24%	7.06%	4.14%	32.11%
To drink	0.75%	5.76%	2.13%	25.69%
To lay on	9.09%	0.32%	13.43%	10.68%
To sit on	17.17%	7.91%	22.59%	34%
To stand	48.88%	53.27%	54.35%	78.51%
To hold an object/instrument	7.85%	9.48%	9.03%	34.78%

4. Used Datasets for Object Detection and HOI Detection

4.1. Common Objects in Context (COCO)

Microsoft developed the COCO dataset [44]. It contains 91 categories of objects and 328,000 images. The majority of the categories have over 5000 annotated objects. The total count of annotated objects is 2,500,000. Although there are other multi-class datasets, such as ImageNet [56], COCO has many more annotated objects in each category. A dataset with fewer classes but more accurate predictions made by the model is preferred over a dataset with more categories but worst predictions. In addition, the COCO dataset contains an average of approximately 8 annotated objects per image, while ImageNet has only 3 or Pascal VOC 2.3 [49]. The used dataset offers annotations for image classification, object detection, and semantic segmentation.

Because a person can interact with different objects, a relevant statistic in choosing the dataset says that only 10% of images contain objects from a single category, while in Pascal VOC approximately 60% of images contain annotations from a single category.

4.2. Verbs in Common Objects in Context (V-COCO)

V-COCO [55] is a dataset that builds on top of COCO for HOI detection. It splits 10,346 images into: 2.533 for training, 2.867 for validation, and 4.946 for testing. In total, there are 16.199 human instances. The dataset considers 26 different action verbs, for a few verbs it includes two types of attributes: instrument and object, resulting in a total number of 29 action classes.

Divided by the existence of an attribute and its type, the verb classes with the associated attributes are as follows:

- <agent, verb>: walk, smile, run, stand;
- <agent, verb, object>: cut, kick, eat, carry, throw, look, read, hold, catch, hit, point;
- <agent, verb, instrument>: surf, ski, ride, talk on the phone, work on computer, sit, jump, lay, drink, eat, hit, snowboard, skateboard.

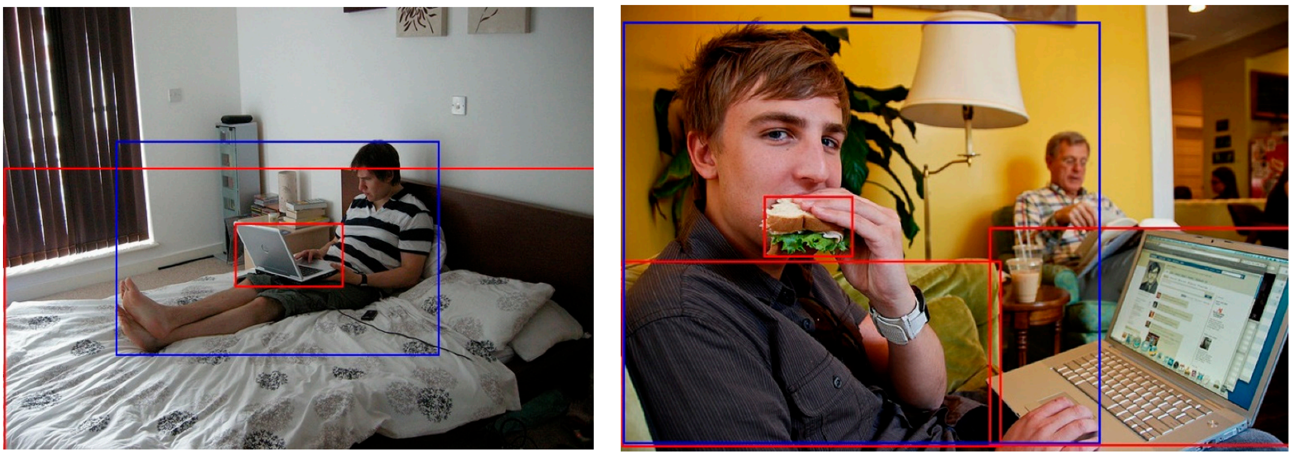
In the context of the SmartCare project, we split the V-COCO verb classes into three categories: essential, util, and unnecessary, according to Table 6.

Table 6. V-COCO categories in SmartCare.

Category	Verbs with Associated Attribute Type
Essential verbs for SmartCare	walk, eat_object, sit_instrument, lay_instrument, drink_instrument, eat_instrument, hold_object, stand
Util verbs for SmartCare	cut_instrument, cut_object, talk_on_phone_instrument, work_on_computer_instrument, carry_object, smile, look_object, point_instrument, read_object, run, jump_instrument
Unnecessary verbs for SmartCare	surf_instrument, ski_instrument, ride_instrument, kick_object, hit_instrument, hit_object, snowboard_instrument, skateboard_instrument, catch_object

V-COCO has, on average, 1.57 people annotated per image that perform actions: over 7000 images with one annotated person, 2000 with two, 800 with three and the rest with four or more people annotated [55]. Figure 5 illustrates situations in which the person performs multiple actions at the same time.

V-COCO has, on average, 1.57 people annotated per image that perform actions: over 7000 images with one annotated person, 2000 with two, 800 with three and the rest with four or more people annotated. [55] Figure 5 illustrates situations in which the person performs multiple actions at the same time.



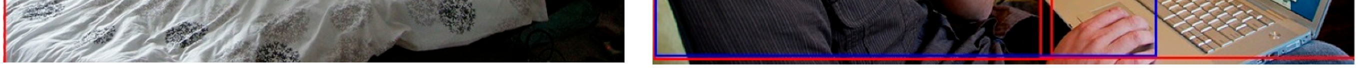
(a) (b)

Figure 5. Examples of annotations in the dataset with a human doing multiple actions at the same time: (**a**) person lying on the bed and working on a computer and (**b**) person sitting on a chair, eating a sandwich and working on a computer [55].

Figure 5. Examples of annotations in the dataset with a human doing multiple actions at the same time: (**a**) person lying on the bed and working on a computer and (**b**) person sitting on a chair, eating a sandwich and working on a computer [55].

5. Human Activity Recognition Module Architecture

From the HAR system's point of view, we have a n-Tier architecture: Zed Camera & Nvidia Jetson, Gateway, System Expert. From the point of view of the application that runs on Nvidia Jetson, we have a multi-level architecture. The hardware architecture is presented in Figure 6. The HAR data flow is presented in Figure 7: the video stream is



(a) (b)

Figure 5. Examples of annotations in the dataset with a human doing multiple actions at the same time: (**a**) person lying on the bed and working on a computer and (**b**) person sitting on a chair, eating a sandwich and working on a computer [55].

5. Human Activity Recognition Module Architecture

5. Human Activity Recognition Module Architecture

From the HAR system's point of view, we have a n-Tier architecture: Zed Camera & Nvidia Jetson, Gateway, System Expert. From the point of view of the application that runs on Nvidia Jetson, we have a multi-level architecture. The hardware architecture is presented in Figure 6. The HAR data flow is presented in Figure 7: the video stream is processed on the Jetson platform, after which an alert is sent when activities of interest are recognized.

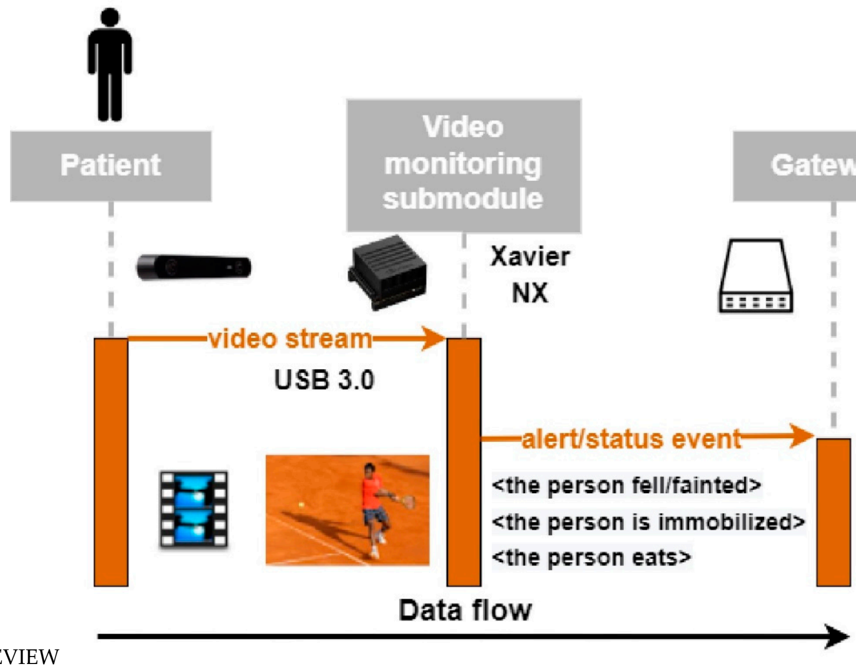


Figure 6. Hardware and software architecture.

Figure 6. Hardware and software architecture.

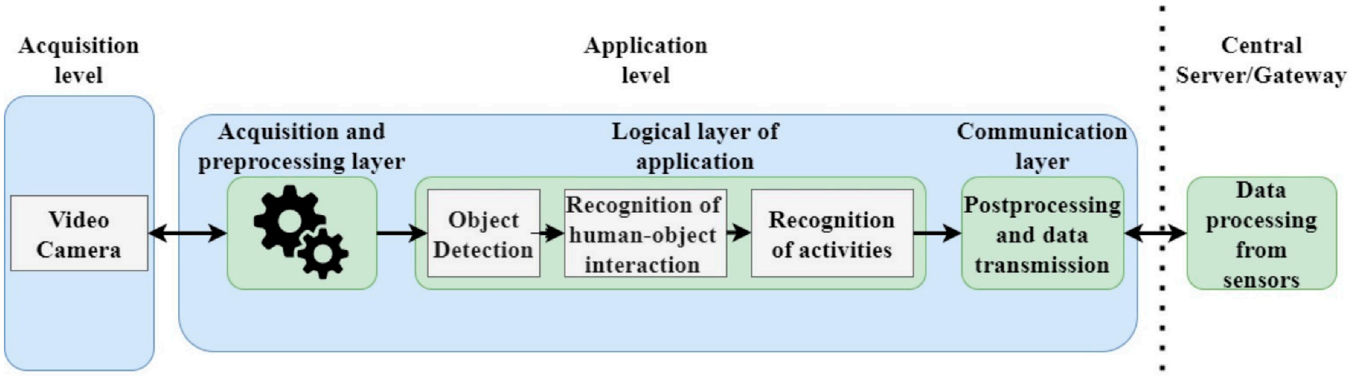


Figure 7. HAR module communication diagram.

Figure 7. HAR module communication diagram.

The ambient images of the patient are acquired by the ZED 2 stereo camera.

At the software level, the application level runs a multi-layer architecture consisting: the image acquisition and preprocessing layer, the logic layer and the communication layer.

- The first layer acquires images and preprocesses them in order to be fed to the logic layer.
- In the logic layer, two neural networks are used: one for human and object detection and another one for human-object interaction detection. Based on the interactions, the activities which are performed by the monitored patient are determined.
- The communication layer prepares the data and transmits it to the server. The communication between the HAR system and the Gateway is carried out over the MQTT protocol. Based on a voting procedure that takes into consideration other sensor results as well, the Gateway decides when notification alerts are sent to the family or the medical doctor.

protocol. Based on a voting procedure that takes into consideration other sensor results as well, the Gateway decides when notification alerts are sent to the family or the medical doctor.

Figure 8 illustrates the proposed pipeline architecture, consisting of two distinct and modular parts: the patient and the environmental objects detection and the detection of the human–object interactions. Based on the interaction between the person and the objects of interest, we can draw conclusions regarding the monitored activities and conditions: falling/fainting, immobility, serving the meal, drinking liquids, etc.

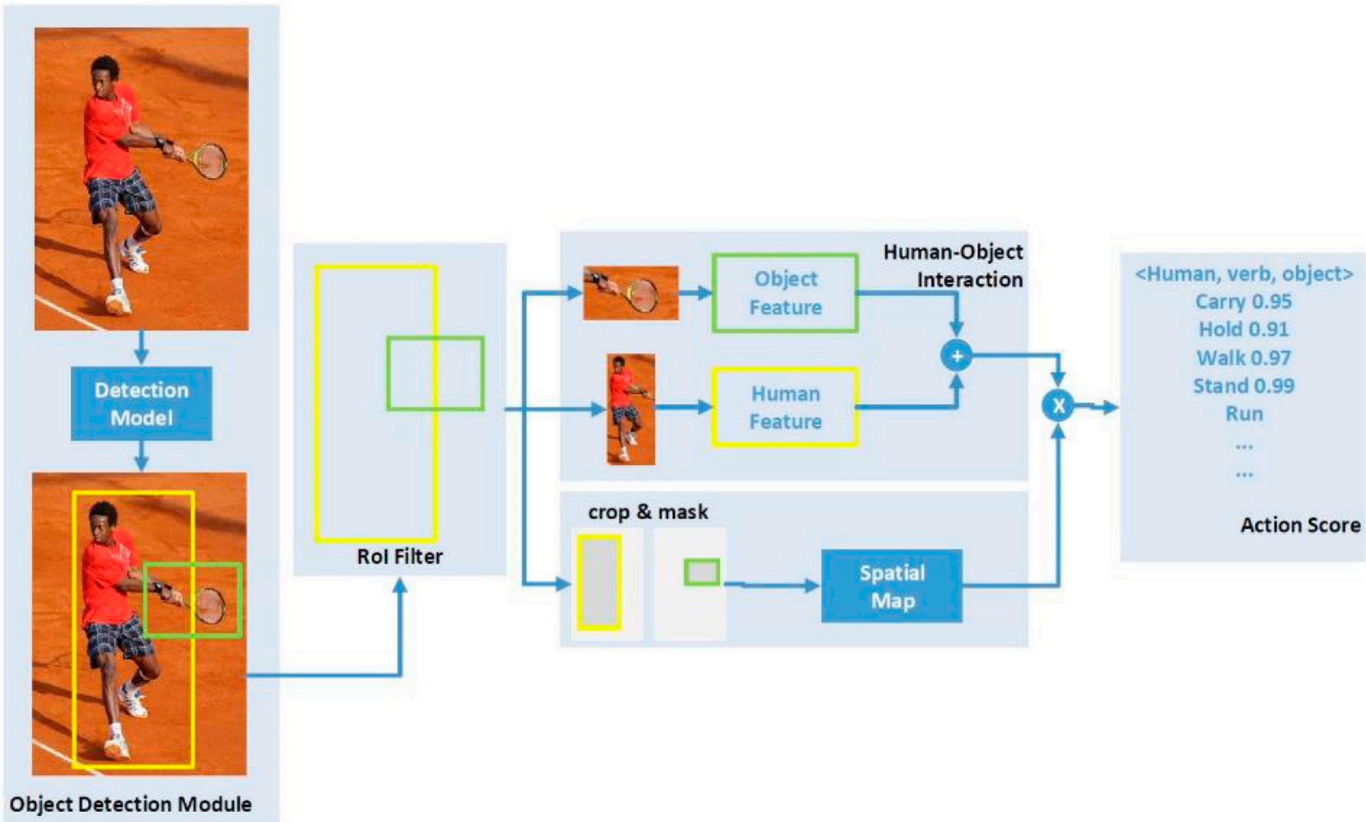


Figure 8. Human–object interaction pipeline architecture.

6. Design and Implementation

In this section, we will discuss the following: class diagram, communication between classes, used threads and finally communication with the Gateway-MQTT server.

As presented in the diagram from Figure 9, the application is modular; if needed, different types of cameras or object detectors can be used:

- CameraAcquisition and ZEDAcquisition implement the functionalities of IAcquisition interface: we considered the option of using a generic camera or the ZED camera (default option);
- JetsonInferenceObjectDetector and DarknetOD implement the functionalities of IObjectDetector interface: two main options were integrated for object detection: YOLO [41,42] (Darknet implementation [51]) and Mobilenet/Inception [38–40] (official Nvidia implementation for Jetson [43]).

For reasons of optimization, threads are used for independent tasks and the communication between them is carried out using priority queues, as follows:

- AcquisitionAndObjectDetect thread is used for image acquisition and object detection;
- queueImageToHOI stores the object detection results;

- HOIDetector thread reads from queueImageToHOI, performs human–object interaction detection;
- queueHOIToMain stores the detected HOIs;
- MQTTConnection thread initiates the connection with the MQTT agent to which the Gateway is connected.

The diagram presented in Figure 10 illustrates the communication between modules and threads for one frame, from acquisition to activity detection.

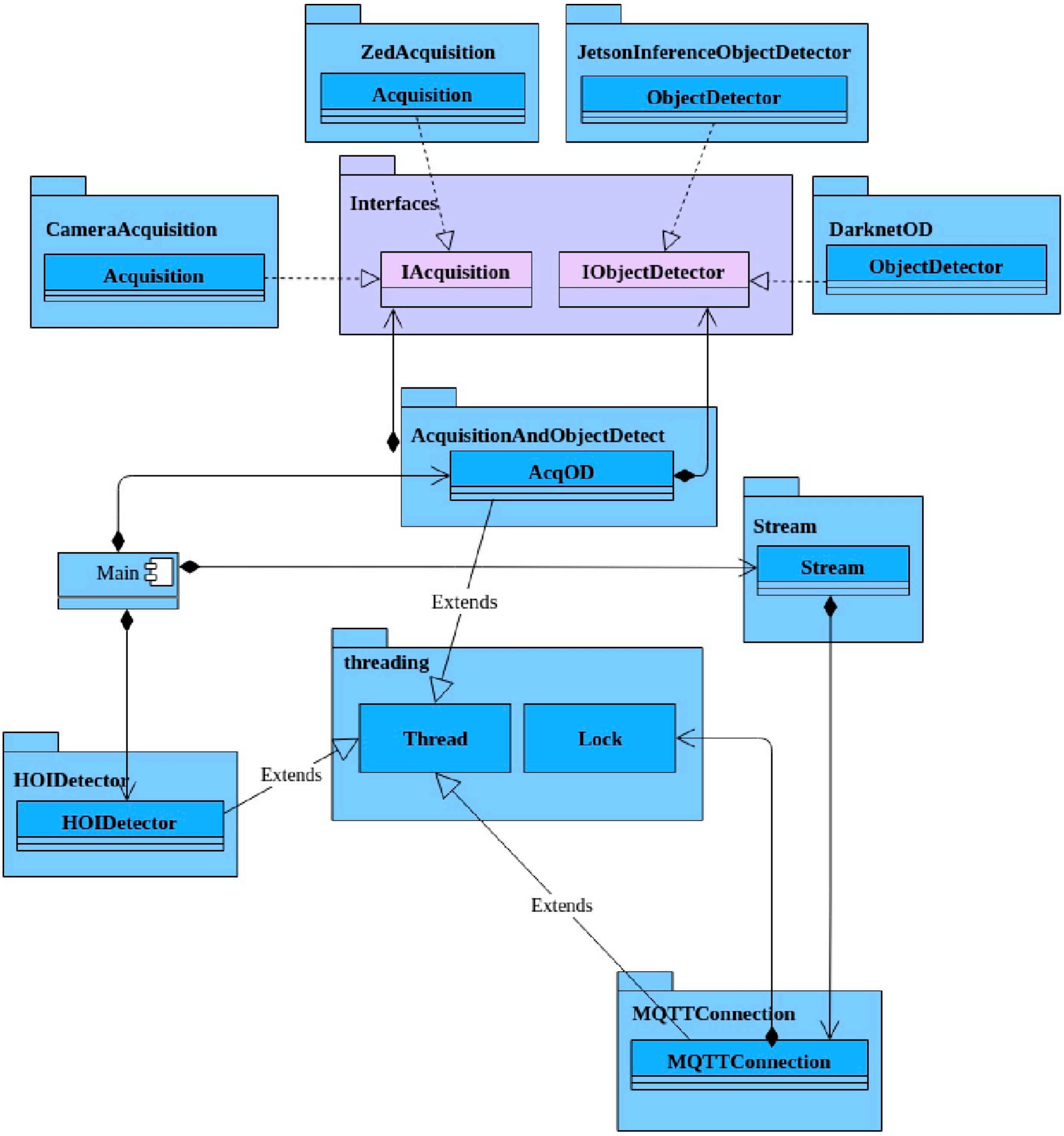


Figure 9. Class diagram.

Figure 9. Class diagram.

For reasons of optimization, threads are used for independent tasks and the communication between them is carried out using priority queues, as follows:

- AcquisitionAndObjectDetect thread is used for image acquisition and object detection;
- queueImageToHOI stores the object detection results;
- HOIDetector thread reads from queueImageToHOI, performs human–object interaction detection;
- queueHOIToMain stores the detected HOIs;
- MQTTConnection thread initiates the connection with the MQTT agent to which the

HOI detector inference time is considerably longer than the time required to detect objects. That is why we decided to test using multiple threads for HOI detections. Figure 11 describes the communication between threads. Priority queues are used for thread communication.

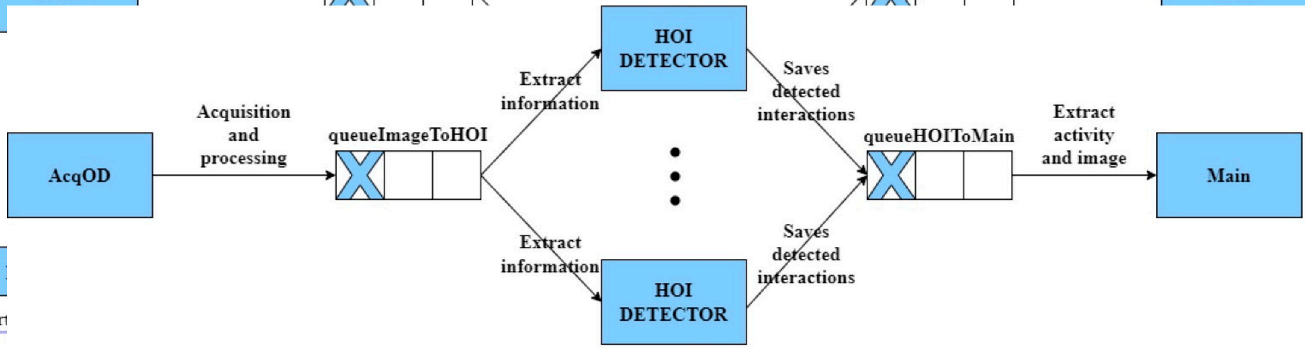


Figure 10. Communication using priority queues.

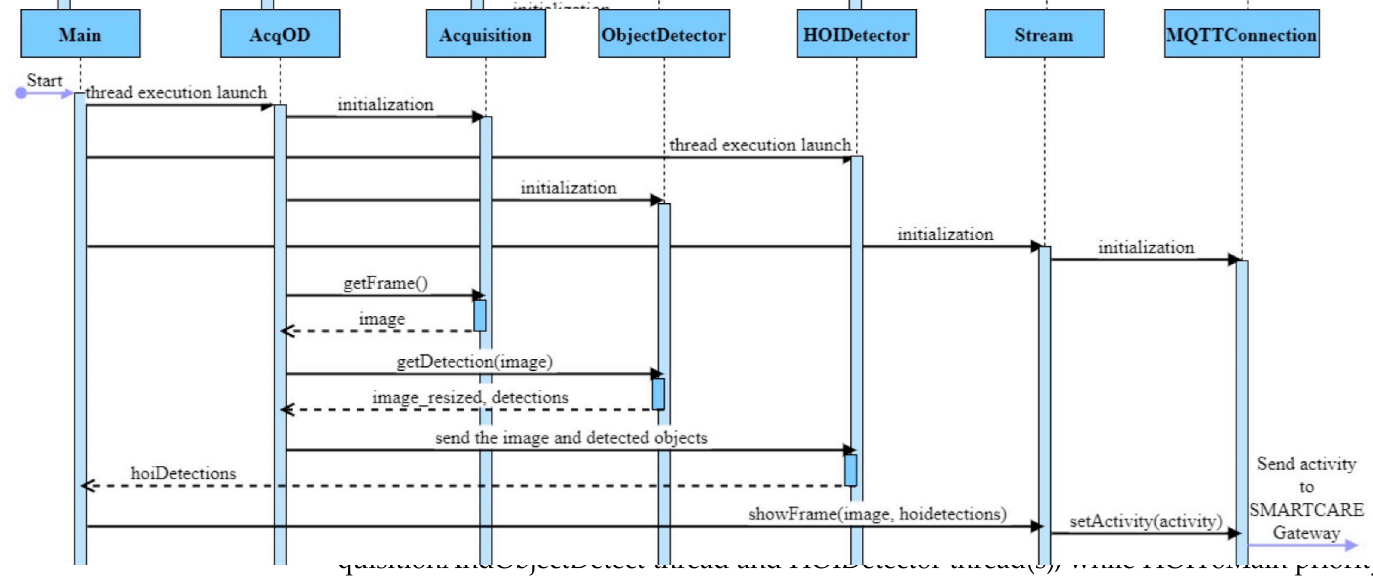


Figure 11. Communication diagram.

There is only one thread for image acquisition and there are object detection, and multiple threads for HOI detection. ImageToHOI priority queue is used between the AcquisitionAndObjectDetect thread and HOIDetector thread(s), while HOIToMain priority queue is used to return the HOI detections to the MQTTConnection thread which updates the current activity and communicates with the MQTT agent.

6.1. Forwarding Detected Activities to the SmartCare Gateway

As described in Section 5, the SmartCare Gateway receives information from several devices, one of which is the HAR system. Communication with the Gateway is carried out using the MQTT protocol by publishing messages. We use serialized JSON objects with information about the current activity of the supervised person.

On initialization, the Gateway subscribes to the following topics:

- gateway/register;
- video-monit/1/activity/response-get;
- video-monit/1/activity/response-async-get.

At start-up, the video monitoring system subscribes to:

- gateway/discover;

- gateway/discover;
- video-monit/1/activity/get;
- video-monit/1/activity/async-get.

It must also make itself known to the Gateway by publishing a serialized JSON object containing system data on the gateway/register topic. The video-monit subtopic is the name of the HAR bridge used for the proposed module. The module communication and initialization process can be followed in Figures 11 and 12, both the HAR system and the SmartCare Gateway subscribe to the MQTT Agent for this connection.

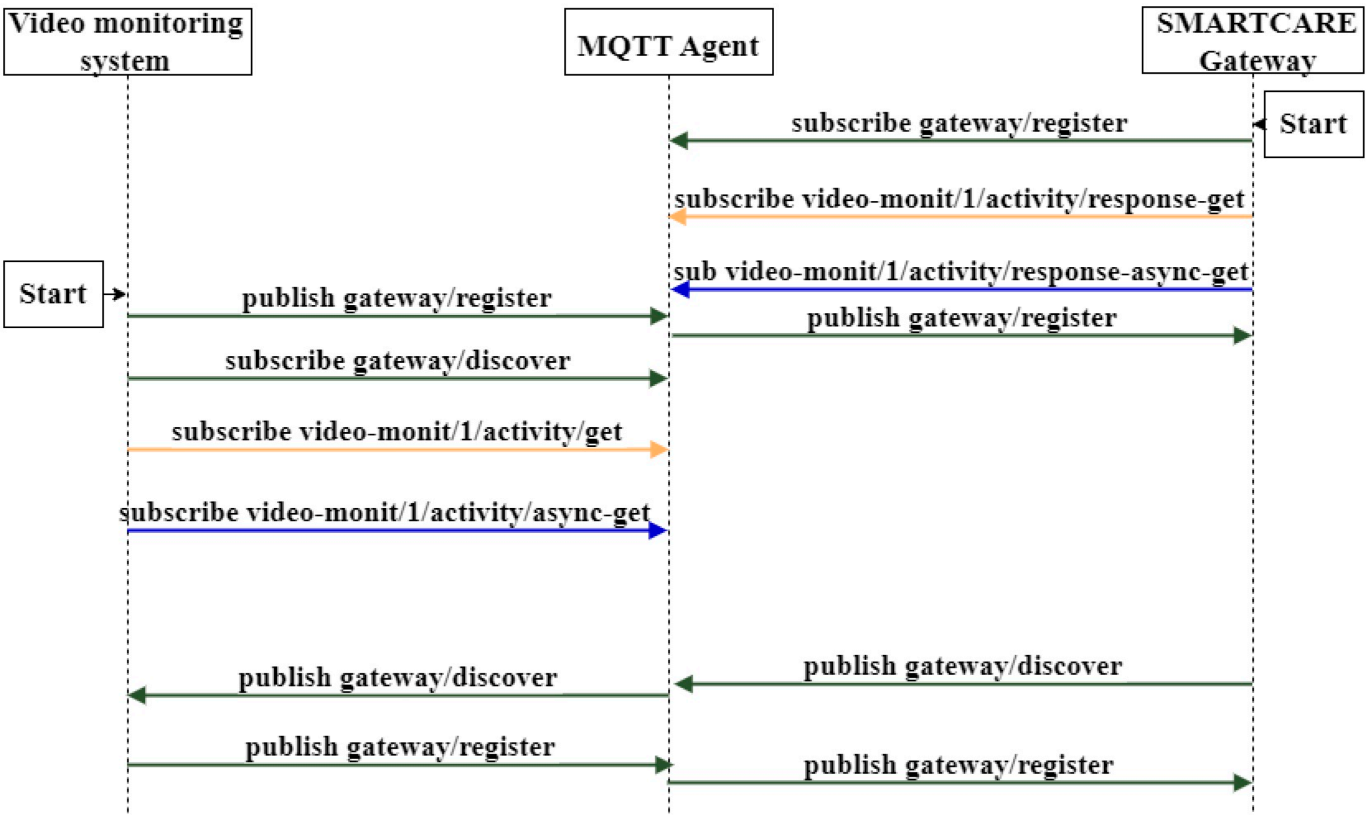


Figure 12. Initialization sequence for the connection between the HAR module and the SmartCare Gateway.

Topic description:

- gateway/register and gateway/discover are used for a handshake protocol: bridges from each device in the network (including the HAR system) subscribe to gateway/discover to receive a "get acquainted" message from the server, after which they publish device details to gateway/register to get recognized by the server.
- video-monit/1/activity/get and video-monit/1/activity/response-get: under the video-monit/bridge name we have *1*/device enrolled with a resource of type activity/. The Gateway publishes an inquiring operation on the *get*/subtopic and the video monitoring system responds on the response-get/subtopic with the patient activity details. This message exchange is represented in Figure 13.
- video-monit/1/activity/async-get and video-monit/1/activity/response-async-get: similar with the above get/response-get topics, an asynchronous behavior is triggered by an 'on'/'off' message. An example of asynchronous communication is illustrated in Figure 14.

Appl. Sci. 2022, 12, x FOR PEER REVIEW 17 of 26
Appl. Sci. 2022, 12, x FOR PEER REVIEW 17 of 26

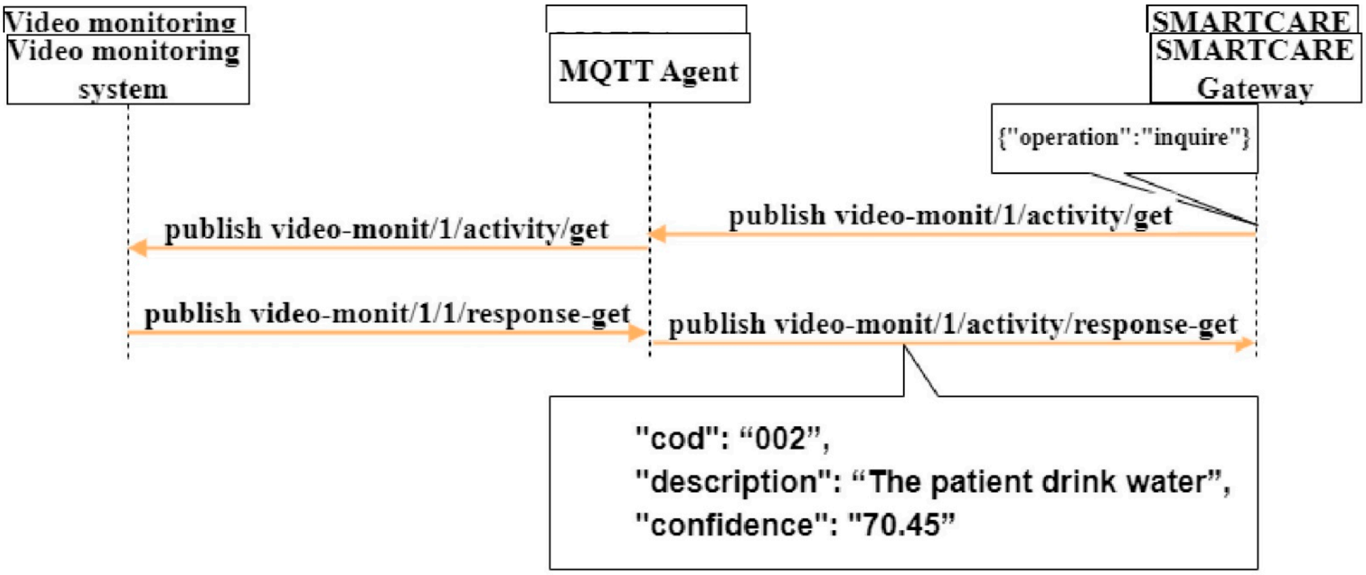


Figure 13. Synchronous communication between the HAR module and the SmartCare system.

Figure 13. Synchronous communication between the HAR module and the SmartCare system.

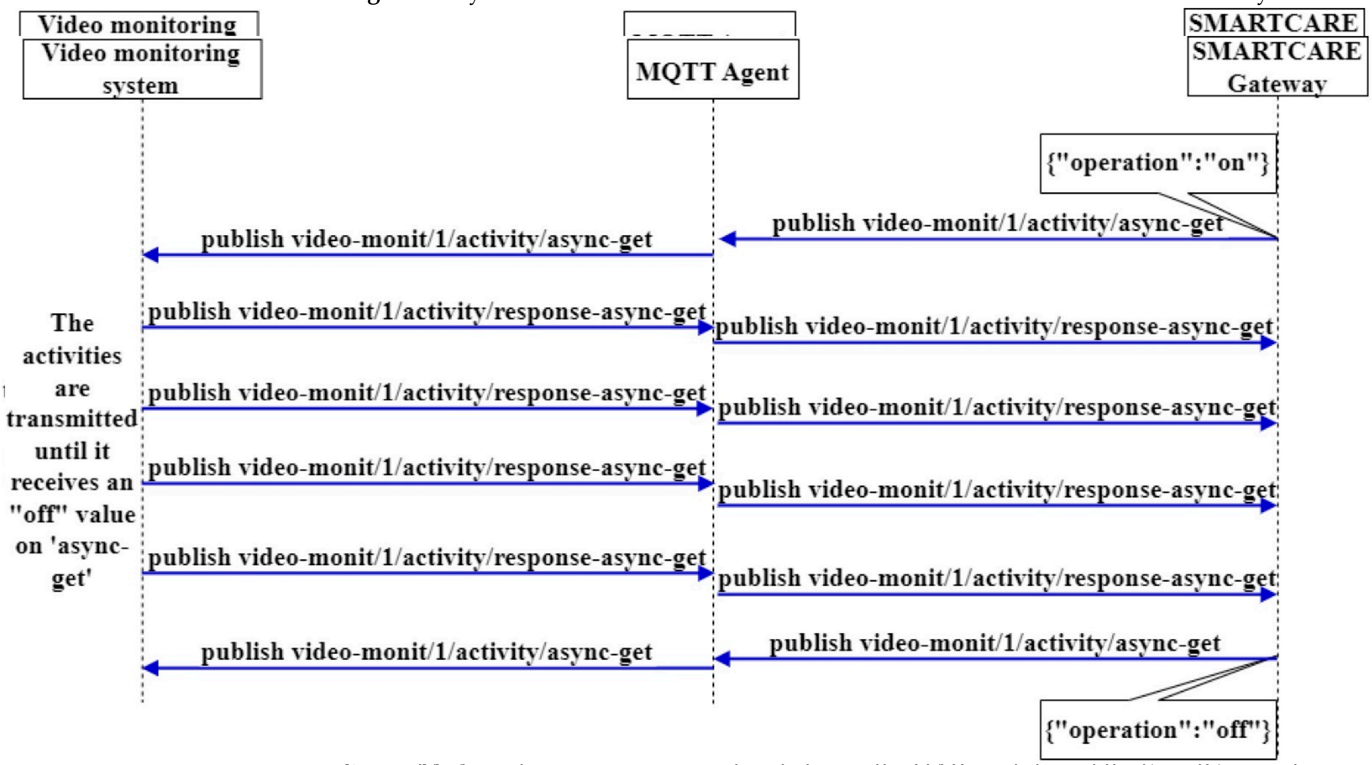


Figure 14. Asynchronous communication between the HAR module and the SmartCare system.

Figure 14. Asynchronous communication between the HAR module and the SmartCare system.

6.2. Thread Experiments

In this subsection, we present the experimental results of configurations with different numbers of threads. One thread is always needed to run the MQTT client which keeps communication with the Gateway open, but we will focus on the threads used by the other components.

6.2.1. Configuration 1: Single Thread for Acquisition, Object Detection, and Activity Detection

All components except MQTTConnection run on a single thread. The entire processing pipeline runs sequentially: first the image is acquired, then object detection is performed,

performed, and finally the HOIs are detected and used to determine the activity performed by the patient. Figure 15 illustrates the single-thread configuration block diagram.

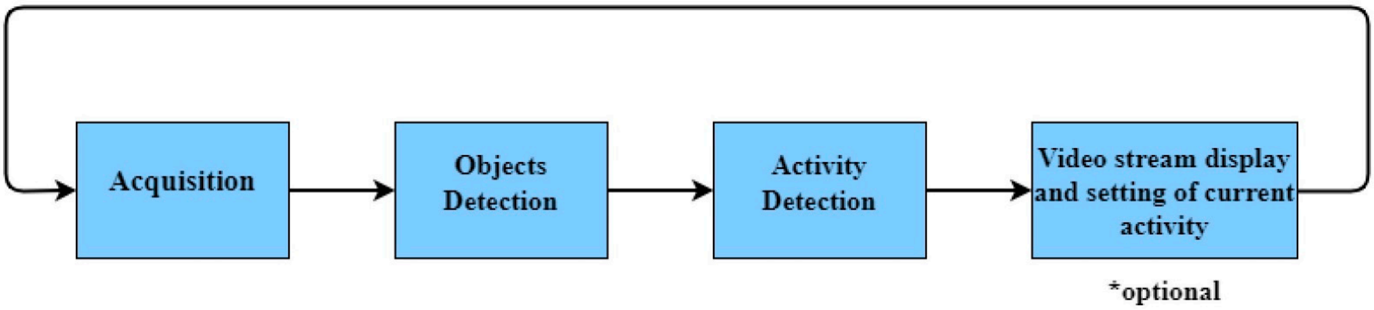


Figure 15. Sequential configuration: single thread for acquisition, object detection and activity detection.

An example of a sequential pipeline timeline is presented in Figure 16: the total processing time is the sum of object detection time and the time needed to detect HOIs. The average frame processing time is t1 + t2; therefore, the result for the third frame is available at T + 3 × (t1 + t2). t1 and t2 are not real measurements; for Figures 16–18, they were chosen to ideally exemplify the tested configurations.

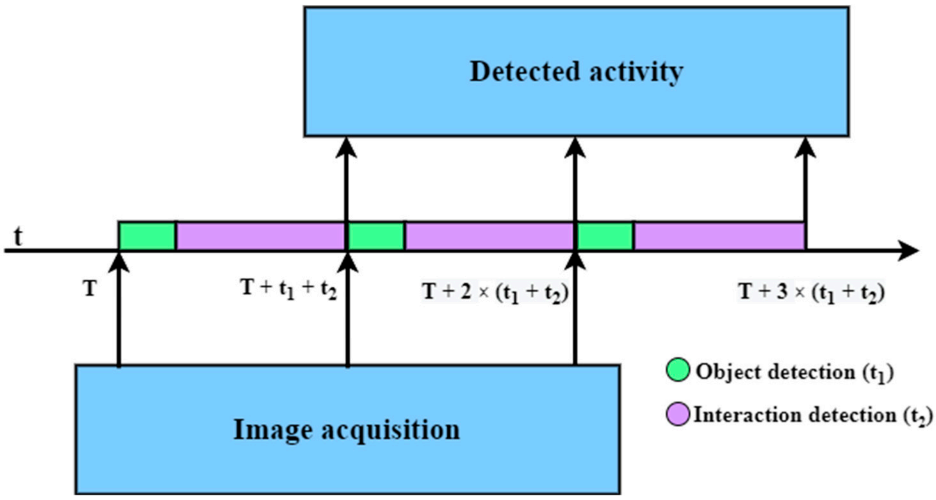
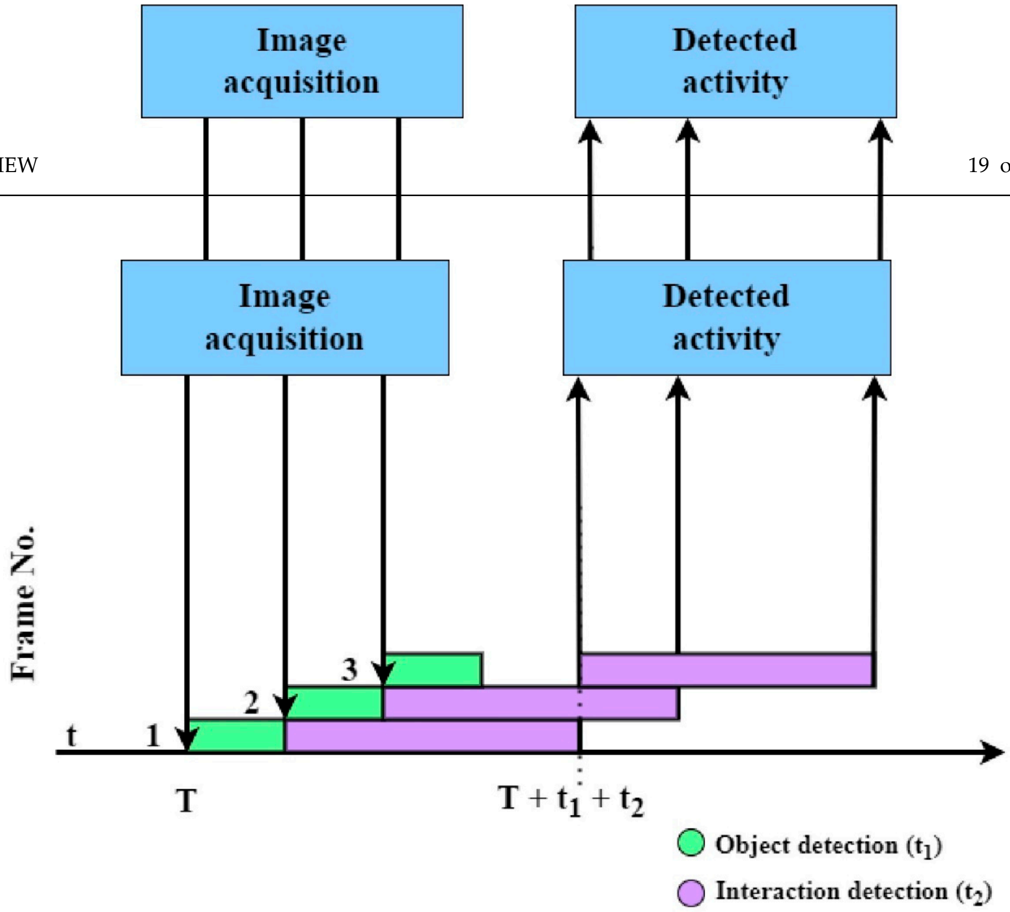


Figure 16. Sequential configuration.

6.2.2. Configuration 2: One Thread for Acquisition and Object Detection, Another One for Activity Detection

In this test scenario, we follow the general diagram from Figure 10: one thread is used for image acquisition and object detection, results go through first queue to the HOI detector which writes the output to the second queue in order to be further processed and extract the patient activity.

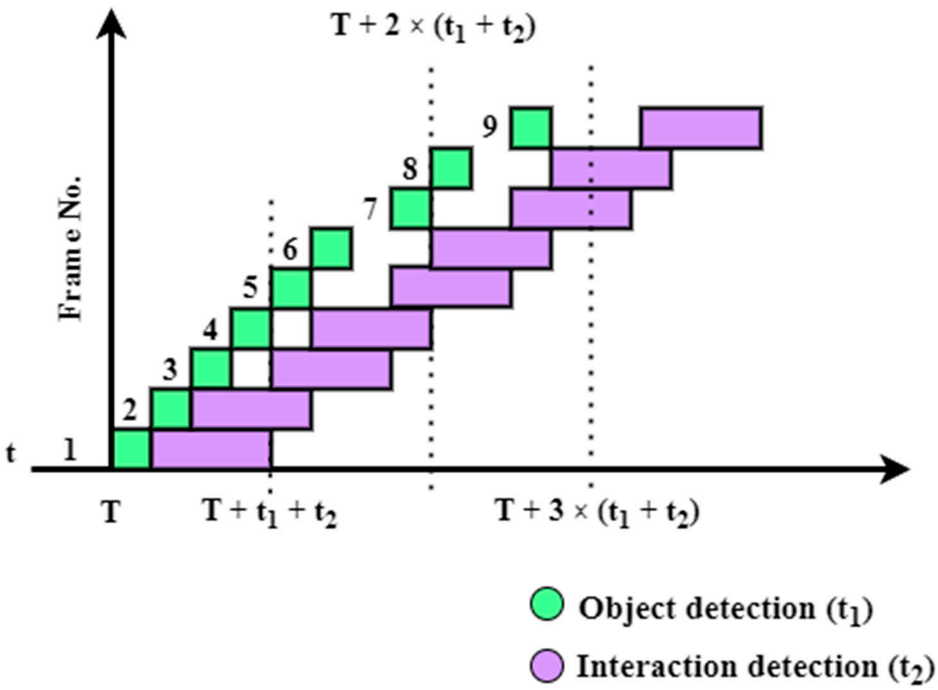
Figure 17 illustrates the result timeline: In addition to the first iteration, which is always affected by the initialization process and is taken out of the time statistics, each time the HOI detection thread finishes the current frame processing, the queue will feed the object detections for the next frame. The average processing time is, in fact, the inference time for HOI detection (t2), the result for the 3rd frame is available at T + t1 + 3 × t2.



ds for

as ex-
even
works,
there should be multiple GPUs. To satisfy this constraint, multiple logical video cards were created by dividing the physical video memory by the number of threads.

Figure 17. Configuration using physical threads.



ne for

used
OI de-
d and

Figure 18. Configuration using three threads.

Figure 17 illustrates the result timeline: In addition to the first iteration, which is always affected by the initialization process and is taken out of the time statistics, each time the HOI detection thread finishes the current frame processing, the queue will feed the

6.2.2. Configuration 2: One Thread for Acquisition and Object Detection, another One for Activity Detection

In this test scenario, we follow the general diagram from Figure 10: one thread is used for image acquisition and object detection, results go through first queue to the HOI detector which writes the output to the second queue in order to be further processed and extract the patient activity.

The advantage of using a single thread for acquisition, object detection and HOI detection is that the time elapsed between frame acquisition and activity detection is the shortest of the three configurations, representing mostly the inference time of the two neural networks (Figure 16). In the second and third configurations, the time elapsed between frame acquisition and activity detection is longer, and a delay is added due to unavailable threads for HOI detection (Figures 17 and 18). An advantage of using dedicated threads in the following configuration: one for acquisition and object detection and one/two for HOI detection, is that the FPS is higher.

Table 7 presents a comparison between different scenarios regarding used threads and queue sizes. Increasing the queue size causes a longer time elapsed between frame acquisition and activity result. In order to have the smallest delay, the queue size is set to be equal to the number of threads used for human–object interaction. Ideally, increasing the number of threads should improve FPS, but, because the memory of the video card is split into multiple virtual GPUs, the processing speed decreases considerably, so that the configuration with two threads for HOI detection is actually slower than the configuration that uses only one thread for this process. The fastest configuration in terms of FPS is the second one-two threads: acquisition and object detection, HOI detection.

Table 7. HOI detection thread experiments.

Acquisition and OD Threads	HOI Threads	Queue Size	FPS
0	0	0	2.17
1	1	1	2.43
1	2	2	2.23

An example of activity detection is shown in Figure 19 and a snapshot of the MQTT message is listed in Figure 20.

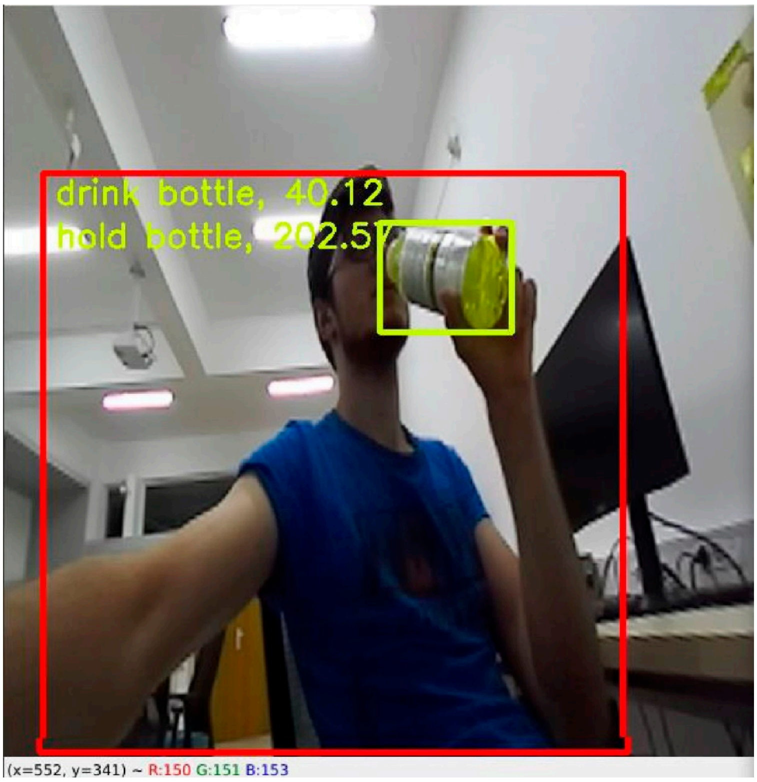
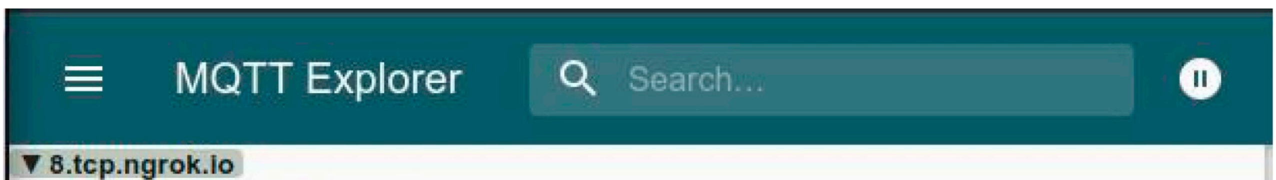


Figure 19. Activity detection result.
Figure 19. Activity detection result.



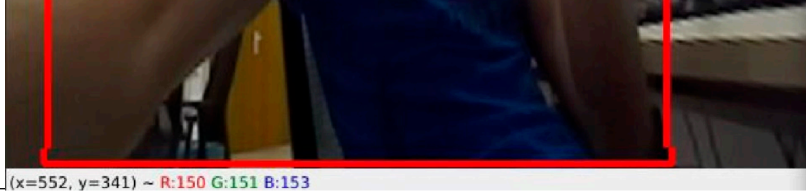


Figure 19. Activity detection result.

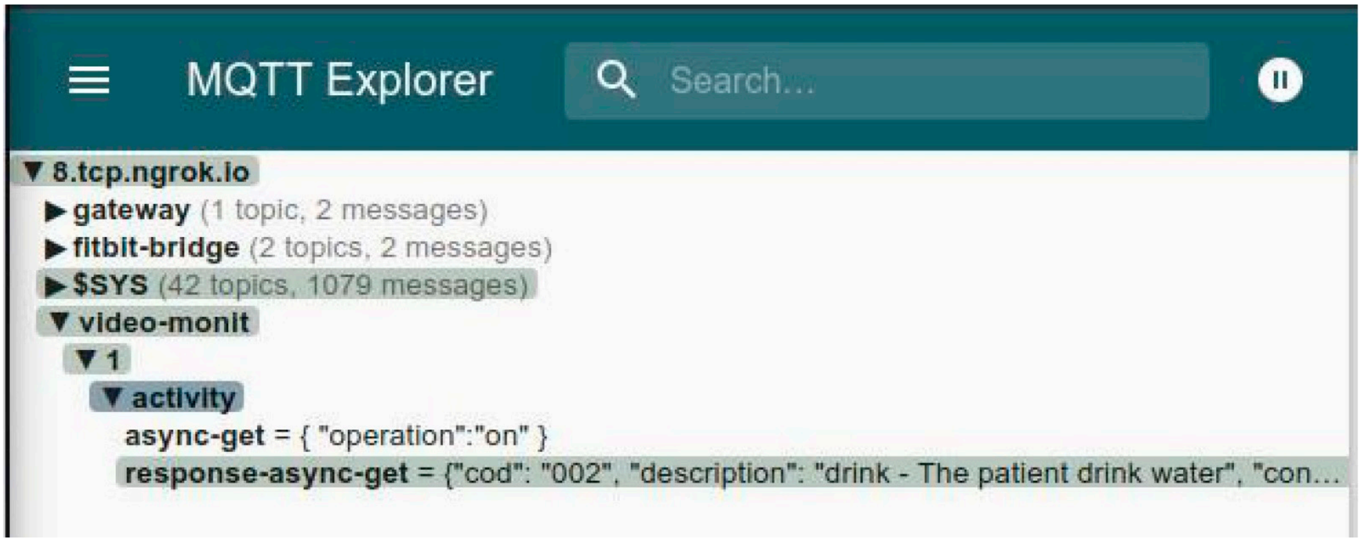


Figure 20. MQTT message sent to the SmartCare Gateway.
Figure 20. MQTT message sent to the SmartCare Gateway.

6.3. Using HAR Results in the SmartCare System

Depending on the monitored patient, the SmartCare system can integrate a wide range of sensors for different aspects, such as

- Vital parameters: heart rate (Fitbit Versa Smartwatch), blood pressure (Omron), blood glucose meter (Contour Plus), etc.
- Home automation: smart dimmer (AD146), smart switch (Fibaro Double Switch 2), valve actuator (Popp Flow Stop 2), smart lock (Danalock V3), waterleak detector (Abus Z-wave SHMW), panic button (Orvibo HS1EB), smart light bulb (Osram Smart A60), ambient temperature, gas sensor, etc.
- Physical activity: steps, fall detection, consumed calories (Fitbit Versa), the video monitoring system.

SmartCare has a predefined set of sensors depending on the health status of the assisted person:

- Alzheimer's disease: mandatory (smart lock, water tap, flood sensor, panic button), recommendation (switch, ambient temperature), nice to have (smart bulb, vibration sensor, humidity, HAR).
- Diabetic: mandatory (blood pressure, blood sugar), recommendation (oxygen saturation), nice to have (panic button, HAR, ambient temperature, humidity).
- Hypertensive: mandatory (blood pressure, heart rate), recommendation (oxygen saturation), nice to have (panic button, HAR, ambient temperature, humidity).
- Obese: mandatory (blood pressure, blood sugar), recommendation (heart rate), nice to have (HAR, ambient temperature).
- etc.

Due to the possibility of a life-threatening event, a rule engine implemented on the Gateway uses information from all connected devices and systems and defines the rules according to the particularities of the monitored person. The WHO and the EU have standardized the use of such rules for the assisted living applications. Accordingly, we define three alert classes: notification (patient), alert (patient, caregiver) and emergency (caregiver, initiate help procedure). Some examples of rules are:

1. If the light intensity is below 300 lux (which is the recommended value for an adult's bedroom), the light is turned on (the dimmer is at a value greater than 0) and the patient is detected (HAR), then the light intensity value is increased step by step up to 300 lux.

2. If the patient is using the sink (HAR) and a leak is detected, then the electricity is turned off (smart switch, smart plug), the tap is closed and the caregiver is alerted.
3. If the patient serves the meal (HAR) and the blood sugar is above the upper limit for diabetes, then the patient and the caregiver are alerted.

The implementation of the HAR system is available at https://github.com/MihaiCHG/VideoMonitoring, accessed on 20 October 2022.

7. Conclusions

In order to reduce the cost of health services provided to an increasingly inactive population, especially the elderly and people with chronic diseases or mental disabilities, the European Union has focused on digital strategies such as e-health and telemedicine. Through programs such as Ageing Well in the Digital World, the EU is financing the development of medical products aiming to maintain the same level of quality in health care services and create a better quality of life for the elderly. An EU pilot project for intelligent living environments worth mentioning is ACTIVAGE. The SmartCare Project follows the European AAL directives being designed to improve living conditions, especially regarding life independence, according to the context of specific needs that people with disabilities may have.

The SmartCare platform in which the human activity recognition system is included has a 'connect and use' architecture that does not require settings to be made. It is designed as a modular platform to which devices can be added or removed in the simplest way possible without affecting the functionality.

To detect activities, this HAR system uses two convolutional neural networks. The first network detects objects in images, including people, while the second one detects human–object interactions based on objects and people received from the previous network. Because the system must run in real time, these networks must have as little processing time as possible. Yolo V4 was chosen from the neural networks for object detection presented in this paper, and iCAN was chosen for the detection of interactions. Based on the detected actions, the monitored activities are detected.

Activities are sent to the gateway via MQTT. The system subscribes to topics that it listens to when it needs to submit activities, and then publishes them on paired topics.

The lab results presented in this paper demonstrate that such a system can be integrated in the SmartCare platform in order to provide information about the activity of the monitored patient, to confirm or deny events detected by other sensors connected to the platform (e.g., IMU). The activity of the person will not be established on the basis of the information received from a single sensor/module, but by merging the information received from other devices.

Future work includes dataset expansion with annotation for specific object classes such as different types of medicine, other objects often found in the patient's environment (home or medical center)—wheelchair, walking crutches, rolling crutches, different types of prostheses or specific wearable technology which also has integration with the SmartCare platform—insulin pump, fitness wristband, blood pressure monitor, panic button, smart switch and smart plug, etc. The object detection model can easily be retrained with an enhanced dataset. The HOI verbs and possible interactions will also be tailored according to the newly added objects.

The recent COVID-19 pandemic has led us to include within SmartCare a stand-alone module in order to monitor the distance and physical interaction between people: patient and patient, patient and caregiver, and patient and medical staff. Future work also implies adapting the HAR output so that the system is compliant with this feature as well.

Author Contributions: Conceptualization, S.-D.A., M.-C.H., R.-G.L. and V.-I.M.; methodology, S.-D.A., M.-C.H., R.-G.L. and V.-I.M.; software, S.-D.A. and M.-C.H.; validation, S.-D.A., M.-C.H., R.-G.L. and V.-I.M.; formal analysis, S.-D.A. and R.-G.L.; investigation, S.-D.A.; resources, S.-D.A.; data curation, S.-D.A. and M.-C.H.; writing—original draft preparation, S.-D.A. and M.-C.H.; writing—review and editing, S.-D.A., M.-C.H. and R.-G.L.; visualization, S.-D.A. and M.-C.H.; supervision, R.-G.L. and V.-I.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Romanian National Authority for Scientific Research (UEFISCDI), Project PN-III-P2-2.1-PTE-2019-0756/2020 Integrative platform for assistance solutions for home autonomy (SMARTCARE).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Mshali, H.; Lemlouma, T.; Moloney, M.; Magoni, D. A Survey on Health Monitoring Systems for Health Smart Homes. *Int. J. Ind. Ergon.* **2018**, *66*, 26–56. [CrossRef]
2. Dang, L.M.; Piran, M.J.; Han, D.; Min, K.; Moon, H. A Survey on Internet of Things and Cloud Computing for Healthcare. *Electronics* **2019**, *8*, 768. [CrossRef]
3. European Commission. *The 2015 Ageing Report: Economic and Budgetary Projections for the 28 EU Member States (2013–2060)*; Publications Office of the European Union: Luxembourg, 2015.
4. Vayyar. Smart Home-Vayyar. Available online: https://vayyar.com/smart-home (accessed on 20 October 2022).
5. Ahmed, A.; Adewale, A.L.; James, A.; Mikail, O.O.; Umar, B.U.; Samuel, E. Human Vital Physiological Parameters Monitoring: A Wireless Body Area Technology Based Internet of Things. *J. Teknol. Dan Sist. Komput.* **2018**, *6*, 13039. [CrossRef]
6. Borelli, E.; Paolini, G.; Antoniazzi, F.; Barbiroli, M.; Benassi, F.; Chesani, F.; Chiari, L.; Fantini, M.; Fuschini, F.; Galassi, A.; et al. HABITAT: An IoT Solution for Independent Elderly. *Sensors* **2019**, *19*, 1258. [CrossRef]
7. Cubo, J.; Nieto, A.; Pimentel, E. A Cloud-Based Internet of Things Platform for Ambient Assisted Living. *Sensors* **2014**, *14*, 14070–14105. [CrossRef]
8. CareStore Project. Available online: https://cordis.europa.eu/project/rcn/105930/factsheet/en (accessed on 20 October 2022).
9. ACTIVAGE Project. Available online: https://cordis.europa.eu/project/rcn/206513/factsheet/en (accessed on 20 October 2022).
10. Shao, D.; Yang, Y.; Liu, C.; Tsow, F.; Yu, H.; Tao, N. Non-contact Monitoring Breathing Pattern, Exhalation Flow Rate and Pulse Transit Time. *IEEE Trans. Biomed. Eng.* **2014**, *61*, 2760–2767. [CrossRef]
11. Marques, G.; Pitarma, R. An Indoor Monitoring System for Ambient Assisted Living Based on Internet of Things Architecture. *Int. J. Environ. Res. Public Health* **2016**, *13*, 1152. [CrossRef]
12. Gwak, J.; Shino, M.; Ueda, K.; Kamata, M. An Investigation of the Effects of Changes in the Indoor Ambient Temperature on Arousal Level, Thermal Comfort, and Physiological Indices. *Appl. Sci.* **2019**, *9*, 899. [CrossRef]
13. Sanchez, L.; Lanza, J.; Olsen, R.; Bauer, M.; Girod-Genet, M. A Generic Context Management Framework for Personal Networking Environments. In Proceedings of the 2006 Third Annual International Conference on Mobile and Ubiquitous Systems: Networking & Services, San Jose, CA, USA, 17–21 July 2006; pp. 1–8. [CrossRef]
14. Höllerer, T.; Hallaway, D.; Tinna, N.; Feiner, S. Steps Toward Accommodating Variable Position Tracking Accuracy in a Mobile Augmented Reality System. In Proceedings of the 2nd International Workshop on Artificial Intelligence in Mobile Systems (AIMS'01), Seattle, WA, USA, 4 August 2001.
15. Tsetsos, V.; Anagnostopoulos, C.; Kikiras, P.; Hadjiefthymiades, S. Semantically enriched navigation for indoor environments. *Int. J. Web Grid Serv.* **2006**, *2*, 453–478. [CrossRef]
16. Chen, L.; Tee, B.K.; Chortos, A.; Schwartz, G.; Tse, V.; Lipomi, D.J.; Wang, H.-S.P.; McConnell, M.V.; Bao, Z. Continuous wireless pressure monitoring and mapping with ultra-small passive sensors for health monitoring and critical care. *Nat. Commun.* **2014**, *5*, 5028. [CrossRef]
17. Lyardet, F.; Grimmer, J.; Mühlhäuser, M. CoINS: Context Sensitive Indoor Navigation System. In Proceedings of the Eigth IEEE International Symposium on Multimedia, San Diego, CA, USA, 11–13 December 2006. [CrossRef]
18. Mshali, H.H. Context-Aware e-Health Services in Smart Spaces. Ph.D. Thesis, Université de Bordeaux, Bordeaux, France, 2017.
19. Malasinghe, L.P.; Ramzan, N.; Dahal, K. Remote patient monitoring: A comprehensive study. *J. Ambient Intell. Human Comput.* **2019**, *10*, 57–76. [CrossRef]
20. Buzzelli, M.; Albé, A.; Ciocca, G. A Vision-Based System for Monitoring Elderly People at Home. *Appl. Sci.* **2020**, *10*, 374. [CrossRef]

21. Zhang, L.; Lin, L.; Liang, X.; He, K. Is Faster R-CNN Doing Well for Pedestrian Detection? In Proceedings of the Computer Vision—ECCV 2016, Amsterdam, The Netherlands, 11–14 October 2016; Springer: Cham, Switzerland, 2016. [CrossRef]
22. Carreira, J.; Zisserman, A. Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 4724–4733. [CrossRef]
23. Luvizon, D.C.; Picard, D.; Tabia, H. 2d/3d pose estimation and action recognition using multitask deep learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 5137–5146. [CrossRef]
24. Luo, Z.; Hsieh, J.-T.; Balachandar, N.; Yeung, S.; Pusiol, G.; Luxenberg, J.; Li, G.; Li, L.-J.; Downing, N.; Milstein, A.; et al. Computer Vision-Based Descriptive Analytics of Seniors' Daily Activities for Long-Term Health Monitoring. *Mach. Learn. Healthc.* **2018**, *2*, 1–18.
25. Karen, S.; Andrew, Z. Two-Stream Convolutional Networks for Action Recognition in Videos. *Adv. Neural Inf. Process. Syst.* **2014**, 568–576. [CrossRef]
26. Lee, R.Y.; Carlisle, A.J. Detection of falls using accelerometers and mobile phone technology. *Age Ageing* **2011**, *40*, 690–696. [CrossRef]
27. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [CrossRef]
28. Chaaraoui, A.A.; Padilla-López, J.R.; Ferrández-Pastor, F.J.; Nieto-Hidalgo, M.; Flórez-Revuelta, F. A vision-based system for intelligent monitoring: Human behaviour analysis and privacy by context. *Sensors* **2014**, *14*, 8895–8925. [CrossRef]
29. Chaaraoui, A.A.; Climent-Pérez, P.; Flórez-Revuelta, F. An Efficient Approach for Multi-view Human Action Recognition Based on Bag-of-Key-Poses. In *Human Behavior Understanding*; HBU 2012; Springer: Berlin/Heidelberg, Germany, 2012. [CrossRef]
30. Chaaraoui, A.A.; Padilla-López, J.R.; Climent-Pérez, P.; Flórez-Revuelta, F. Evolutionary joint selection to improve human action recognition with RGB-D devices. *Expert Syst. Appl.* **2014**, *41*, 786–794. [CrossRef]
31. Kim, K.; Jalal, A.; Mahmood, M. Vision-Based Human Activity Recognition System Using Depth Silhouettes: A Smart Home System for Monitoring the Residents. *J. Electr. Eng. Technol.* **2019**, *14*, 2567–2573. [CrossRef]
32. Vishwakarma, V.; Mandal, C.; Sural, S. Automatic Detection of Human Fall in Video. In Proceedings of the Pattern Recognition and Machine Intelligence, Kolkata, India, 18–22 December 2007. [CrossRef]
33. Carreira, J.; Noland, E.; Hillier, C.; Zisserman, A. A Short Note on the Kinetics-700 Human Action Dataset. *arXiv* **2019**, arXiv:1907.06987. [CrossRef]
34. Smaira, L.; Carreira, J.; Noland, E.; Clancy, E.; Wu, A.; Zisserman, A. A Short Note on the Kinetics-700-2020 Human Action Dataset. *arXiv* **2020**, arXiv:2010.10864. [CrossRef]
35. Monfort, M.; Pan, B.; Ramakrishnan, K.; Andonian, A.; McNamara, B.A.; Lascelles, A.; Fan, Q.; Gutfreund, D.; Feris, R.; Oliva, A. Multi-Moments in Time: Learning and Interpreting Models for Multi-Action Video Understanding. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, 1. [CrossRef] [PubMed]
36. Monfort, M.; Jin, S.; Liu, A.; Harwath, D.; Feris, R.; Glass, J.; Oliva, A. Spoken Moments: Learning Joint Audio-Visual Representations From Video Descriptions. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 14871–14881. [CrossRef]
37. Achirei, S.; Zvoriș teanu, O.; Alexandrescu, A.; Botezatu, N.; Stan, A.; Rotariu, C.; Lupu, R.; Caraiman, S. SMARTCARE: On the Design of an IoT Based Solution for Assisted Living. In Proceedings of the 2020 International Conference on e-Health and Bioengineering (EHB), Iasi, Romania, 29–30 October 2020; pp. 1–4. [CrossRef]
38. Howard, A.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv* **2017**, arXiv:1704.04861. [CrossRef]
39. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.-C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520. [CrossRef]
40. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826. [CrossRef]
41. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934. [CrossRef]
42. Jetson-Inference. Available online: https://github.com/dusty-nv/jetson-inference (accessed on 20 October 2022).
43. Wang, C.-Y.; Bochkovskiy, A.; Liao, H.M. Scaled-yolov4: Scaling cross stage partial network. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021. [CrossRef]
44. Lin, T.; Maire, M.; Belongie, S.; Bourdev, L.; Girshick, R.; Hays, J.; Perona, P.; Ramanan, D.; Zitnick, C.L.; Dollár, P. Microsoft COCO: Common Objects in Context. In Proceedings of the Computer Vision—ECCV, Zurich, Switzerland, 6–12 September 2014. [CrossRef]
45. Cordts, M.; Omran, M.; Ramos, S.; Scharwächter, T.; Enzweiler, M.; Benenson, R.; Franke, U.; Roth, S.; Schiele, B. The Cityscapes Dataset. In Proceedings of the CVPR Workshop on the Future of Datasets in Vision. 2015. Available online: https://www.cityscapes-dataset.com/wordpress/wp-content/papercite-data/pdf/cordts2015cvprw.pdf (accessed on 20 October 2022).

46. Cordts, M.; Omran, M.; Ramos, S.; Rehfeld, T.; Enzweiler, M.; Benenson, R.; Franke, U.; Roth, S.; Schiele, B. The Cityscapes Dataset for Semantic Urban Scene Understanding. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016. [CrossRef]
47. Valada, A.; Oliveira, G.; Brox, T.; Burgard, W. Deep Multispectral Semantic Scene Understanding of Forested Environments Using Multimodal Fusion. In *2016 International Symposium on Experimental Robotics. ISER 2016. Springer Proceedings in Advanced Robotics*; Springer: Cham, Switzerland, 2017; pp. 465–477. [CrossRef]
48. Zhao, J.; Li, J.; Cheng, Y.; Sim, T.; Yan, S.; Feng, J. Understanding Humans in Crowded Scenes: Deep Nested Adversarial Learning and A New Benchmark for Multi-Human Parsing. In Proceedings of the 26th ACM international conference on Multimedia, Seoul, Korea, 22–26 October 2018; pp. 792–800. [CrossRef]
49. Everingham, M.; Van Gool, L.; Williams, C.; Winn, J.; Zisserman, A. The Pascal Visual Object Classes (VOC) challenge. *Int. J. Comput. Vis.* **2010**, *88*, 303–338. [CrossRef]
50. Song, S.; Lichtenberg, S.P.; Xiao, J. SUN RGB-D: A RGB-D scene understanding benchmark suite. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 567–576. [CrossRef]
51. YOLO v4 and YOLO v4 Tiny Implementation by Darknet. Available online: https://github.com/pjreddie/darknet (accessed on 20 October 2022).
52. Joseph, R.; Farhadi, A. Yolov3: An Incremental Improvement. 2018. Available online: https://doi.org/10.48550/arXiv.1804.02767 (accessed on 20 October 2022).
53. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 6517–6525. [CrossRef]
54. Chen, G.; Zou, Y.; Huang, J. iCAN: Instance-centric attention network for human-object interaction detection. In Proceedings of the British Machine Vision Conference (BMVC), Newcastle, UK, 3–6 September 2018. [CrossRef]
55. Gupta, S.; Malik, J. Visual Semantic Role Labeling. *arXiv* **2015**, arXiv:1505.04474. [CrossRef]
56. Chao, Y.; Liu, Y.; Liu, X.; Zeng, H.; Deng, J. Learning to Detect Human-Object Interactions. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, USA, 12–15 March 2018. [CrossRef]
57. Chao, Y.; Wang, Z.; He, Y.; Wang, J.; Deng, J. HICO: A Benchmark for Recognizing Human-Object Interactions in Images. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015. [CrossRef]

electronics

Article

Sensory Substitution for the Visually Impaired: A Study on the Usability of the Sound of Vision System in Outdoor Environments

Otilia Zvorișteanu , Simona Caraiman *, Robert-Gabriel Lupu, Nicolae Alexandru Botezatu and Adrian Burlacu

Faculty of Automatic Control and Computer Engineering, "Gheorghe Asachi" Technical University of Iasi, D. Mangeron 27, 700050 Iasi, Romania; otilia.zvoristeanu@academic.tuiasi.ro (O.Z.); robert-gabriel.lupu@academic.tuiasi.ro (R.-G.L.); nicolae-alexandru.botezatu@academic.tuiasi.ro (N.A.B.); adrian.burlacu@academic.tuiasi.ro (A.B.)

* Correspondence: simona.caraiman@academic.tuiasi.ro

Citation: Zvorișteanu, O.; Caraiman, S.; Lupu, R.-G.; Botezatu, N.A.; Burlacu, A. Sensory Substitution for the Visually Impaired: A Study on the Usability of the Sound of Vision System in Outdoor Environments. *Electronics* **2021**, *10*, 1619. https://doi.org/10.3390/electronics10141619

Academic Editors: Michael Wehner and Daniel Morris

Received: 15 May 2021
Accepted: 28 June 2021
Published: 6 July 2021

Abstract: For most visually impaired people, simple tasks such as understanding the environment or moving safely around it represent huge challenges. The Sound of Vision system was designed as a sensory substitution device, based on computer vision techniques, that encodes any environment in a naturalistic representation through audio and haptic feedback. The present paper presents a study on the usability of this system for visually impaired people in relevant environments. The aim of the study is to assess how well the system is able to help the perception and mobility of the visually impaired participants in real life environments and circumstances. The testing scenarios were devised to allow the assessment of the added value of the Sound of Vision system compared to traditional assistive instruments, such as the white cane. Various data were collected during the tests to allow for a better evaluation of the performance: system configuration, completion times, electro-dermal activity, video footage, user feedback. With minimal training, the system could be successfully used in outdoor environments to perform various perception and mobility tasks. The benefit of the Sound of Vision device compared to the white cane was confirmed by the participants and by the evaluation results to consist in: providing early feedback about static and dynamic objects, providing feedback about elevated objects, walls, negative obstacles (e.g., holes in the ground) and signs.

Keywords: sound of vision; visually impaired; sensory substitution system; outdoor environments; perception; mobility

1. Introduction and Related Work

The World Health Organization (WHO) estimates that at least 2.2 billion persons around the world suffer from blindness or visual impairment. The effects of reduced or absent eyesight have a major impact on the life of a person, e.g., daily routine, school, work. In the last years, several systems were proposed to help visually impaired people to improve their perception and/or navigation in unknown environments. These devices incorporate different technologies and sensors.

Di Mattia et al. [1] proposed a low consumption radar-based system for obstacle avoidance. An acoustic warning is generated every time an obstacle is detected, and the range of detection is within 5 m. To obtain a more complex and accurate estimation of environment objects, a sensor fusion system comprising a low-power millimeter wave (MMW) radar and an RGB-Depth (RGB-D) sensor is described in [2,3]. Using this data fusion, the authors ensured the accuracy and stability of the system under any illumination conditions and expand the object detection range up to 80m. Semantic/non-semantic acoustic feedback is sent to the user by Bluetooth bone conduction headphones. In [4], the authors combined the advantages of an IR sensor with an RGB-D camera. A random sample consensus (RANSAC) segmentation and surface normal vector estimation are used to detect the traversable area. The device is functional in both indoor and outdoor environments and has been tested by eight visually impaired volunteers. The EyeCane [5] is equipped with IR sensors that capture distance information about obstacles and convey it to the user's hand through vibrations.

In the recent years, multiple computer vision based devices for visually impaired have been proposed in the literature. A Raspberry Pi with a camera module was used by Abraham et al. in [6]. The device can identify and locate specific object from the environment, detect text and convey it to speech and also to determine the walkable area. A neural network, i.e., YOLOv3, was used to compute the elements listed before. A module comprising a pi-camera and a controller to move the camera in the required direction was proposed in [7], and this module was integrated into the white-cane. Mask R-CNN is used to detect and classify the objects from the environment. Furthermore, the system estimates the position of obstacles in outdoor environments. Kang et al. [8] proposed a method to detect the risk of collision in a variety of scenarios. The approach effectively locates obstacles at a risk of collision using the shape variation of a grid, called deformable grid. This solution is further improved in [9] by introducing a vertex deformation function to represent the displacement of each vertex in the deformable grid.

User experience understanding is essential to make assistive technology really useful, non-obtrusive and pervasive. Building a technology for the assistance of the visually impaired (VI) requires a deep user study to iteratively assess user satisfaction and then to bring improvements and corrections to the system accordingly. Offline tests and evaluations of the computer vision techniques employed by these assistive systems are truly required to assess technical performance. However, most of the reported contributions are limited to this form of evaluation [10–16], whereas extensive testing with visually impaired users would bring more insight on usability. There are only a few contributions that report taking the system 'in the wild', i.e., in real-life uncontrolled scenarios, to evaluate its performance concerning the technical design and implementation or usability. The obstacle detection system described by Rodriguez et al. [17] has been tested with visually impaired people in real life scenarios consisting of crowded and uncontrolled areas such as a railway station. Experiments in various uncontrolled environments have also been reported by [18,19]. The framework proposed in [20] uses voice messages to alert the user about the presence of obstacles. The system is evaluated with the help of visually impaired subjects and answers to the following aspects: Are the users able to start the application on their own? Can they safely navigate in a novel environment?Is it possible to avoid obstacles using the set of acoustic warnings? Is the system globally useful and can it complement the white cane? In [21], a smartphone camera was used to acquire images from the environment that are further processed on a server. Four visually impaired persons with partial level of visual impairment tested the solution. A questionnaire regarding the overall impression, user interface and experience and alert frequency was collected.

For most of the assistive solutions proposed in the literature, there is a lack of usability assessment. Such evaluations should be based on a more complex feedback provided through visually impaired user experience. Several development loops followed by user evaluations should be employed before reaching a final solution that provides both technical accuracy and user adoption. More visually impaired user evaluations should be designed to assess each component of the system, ranging from the information required to be extracted from the environment to the method of delivering it, but also the various combinations of these components. Furthermore, none of the analyzed systems employ extensive testing in real environments and in uncontrolled settings. New requirements could emerge from these tests, from both technical and user perspective.

2. Purpose of the Study

The Sound of Vision system (SoV) [22–24] is a sensory substitution device (SSD) that allows a visually impaired user to perceive unknown environments and to navigate safely.

It works by permanently scanning the environment, extracting essential features and rendering them to the user through audio and haptic means.

Several design–implementation–evaluation loops have been previously employed for the development of the SoV system. At each development phase, various usability aspects have been carefully assessed: selection of the most appropriate audio and haptic encodings [25–27], the effect of training on performance improvement [28,29], and cognitive and affective assessment of mobility tasks [30]. These previous evaluations have been deployed in controlled laboratory settings.

In contrast, the present study is focused on evaluating the usability of the Sound of Vision system in complex real life environments, outdoors. For this purpose, evaluations to assess user perception and mobility in outdoor environments were devised. The tests were performed in normal lighting (i.e., cloudy to bright sunlight) and weather conditions (i.e., no rain or snow, temperatures above 0 °C). The users are assumed to be familiar with all the encodings and options available in the Sound of Vision (SoV) system, and were encouraged to use their preferred combination in each test.

The tests focused on evaluating the usability of the Sound of Vision system in real world outdoor scenarios. The main research questions addressed were:

1. *Are the visually impaired (VI) users able to perceive the environment (perception)? Are they able to identify obstacles and specific objects (negative obstacles, hanging obstacles, signs, walls) that define the added value of SoV compared to using the white cane? Is the system usable in real life environments and under real life circumstances (outside laboratory setups)?*
2. *Are the VI users able to use the information from the SoV device to guide their interaction with the environment (mobility)? Are they able to move around and avoid obstacles? Are they able to move around and identify targets (e.g., bus stop, corner of a building)? How is their mobility performance with the SoV system compared to traditional assistive devices (i.e., white cane)?*

The results of the evaluations were analyzed using a case study approach, which is more appropriate for understanding the strengths and weaknesses of the system, correlated to the individual user expectations and needs. This allows us to better explain the highly probable individual variance and inconsistencies in user performance and feedback given the small size of the sample involved in the tests vs. the high diversity of the visually impaired community. Further, an analysis of averaged results of all participants was also performed to overcome individual differences. However, statistical significance tests or extensive comparisons between demographic categories were not the purpose of the presented study. An emphasis was laid on acquiring user feedback through interviews and questionnaires.

While several research groups were involved in the multiple experimental and evaluation phases of the Sound of Vision project, each performed a specific study with the system in each phase [25–30]. Overall, almost 50 blind persons were involved in these evaluations throughout the project. However, the replication of the same experiments in real-life environments is difficult and prone to significant variations. In contrast, the experiments involving virtual environments and laboratory settings, were replicated at the premises of four partners in three countries. Furthermore, we do not provide a comparison between the results of the present study and those obtained in the previous usability evaluations with the system. This is mainly justified by the different versions of the SoV prototype used in the evaluations as several design–implementation–evaluation loops were employed during the project. Each loop was followed by improvements on the design and functionality of the system. Moreover, each evaluation phase had a different purpose, specific to the corresponding technological readiness level of the system.

The main contributions of this paper consist in providing a procedure for usability assessment of sensory substitution devices for the visually impaired in complex real-life environments as well as the results of applying it for the evaluation of the Sound of Vision SSD. Besides their intrinsic value for the validation of the multi-sensory feedback employed in the Sound of Vision system, these results also provide further valuable insights for the development of any sensory substitution device for the visually impaired:

- Evaluation of environment perception and understanding based on audio and haptic feedback, in real life usage conditions;
- Evaluation of the usability of multi-sensory feedback of SSDs for mobility tasks in real life environments;
- Importance of training and recommendations for improved protocols and instruments for training with SSDs;
- Evaluation of the interplay between SSDs and traditional assistive instruments (white cane);
- Recommendations for the development of multi-sensory feedback systems for the visually impaired.

To the best of our knowledge, this is the first elaborate study on the usability of a sensory substitution device for the visually impaired in real-life conditions, outdoors.

3. Materials and Methods

3.1. The Sound of Vision System

The Sound of Vision system (SoV) is a wearable sensory substitution device (SSD) for visually impaired persons. The SSD aims to help them to understand the surrounding environment (perception) and to improve their mobility in unknown, indoor and outdoor environments (navigation). The SoV prototype used in the present study integrates custom and complex software and hardware solutions that enable real-time operation of the device. The system works by permanently acquiring environmental information through a fusion of cameras and sensors, extracting essential features and providing real-time feedback to the user by conveying the information through audio signals and haptics (vibrations)—Figure 1. Moreover, the system is designed to work in both indoor and outdoor environments and irrespective of the illumination conditions. A presentation of the SoV system can be accessed at https://youtu.be/6QRiwykp_bM (accessed on 5 July 2021).

Figure 1. The Sound of Vision system. (**Left**) The prototype used in the usability evaluations. (**Right**) Various parts of the system.

Environment sensing and reconstruction is performed based on data acquired with different imaging and inertial sensors [31]. A structured light camera provides depth information in indoor environments and in low light or in the dark. In outdoor environments with normal and bright lighting conditions, the depth information is provided by a stereo vision system. This combination of sensors ensures the environment sensing in any conditions. The 3D reconstruction and recognition of the various elements of interest in the environment is obtained with specific processing pipelines, tailored for the particular input used [23].

The 3D reconstruction pipeline used in the performed tests is based on stereo vision and is specifically tailored for outdoor environments. The details of the technical solution and the reconstruction algorithms for outdoor environments are presented in [24]. In this configuration, the system is able to identify various types of elements of interest and their properties (width, height and position with respect to the camera): generic obstacles, walls, negative obstacles (e.g., holes in the ground, stairs down), signs and texts.

Sound of Vision offers to the user several ways of perceiving the surrounding scene: 2 full scene encoders, plus tools useful for specific situations, and danger rendering. Changing between these and adjusting their audio/haptic options and volumes is easy to perform in real-time, using a remote control.

The full scene encoders are independent modes of encoding the information about all the objects in the scene (i.e., segmented and identified by the 3D module): iterative (renders the objects in a loop—one by one, in increasing order of distance from user, similar to an expanding sphere); continuous (renders all the objects simultaneously). In both cases, the rendering of each individual object has both audio and haptic outputs, which are carefully synchronized. Furthermore, both of the full scene encoders provide several options regarding the way that individual objects are rendered. For example, in the iterative mode, for audio, the user can choose between stimuli of two types: bar impact sounds or bubble sounds, while for haptics, the user can choose between the projection of shapes on the belt, or just of the closest points of the objects in the scene. The properties of the sound stimuli (e.g., pitch, duration, amount of oscillation, etc.) generated for each object intuitively encode its width, height, distance to the user and elevation from the ground [25]. For example, when encoded with impact sounds, the closer an obstacle is to the user, the louder its sound will be. The wider the obstacle, the deeper (lower) its sound will be.

The tools are designed in order to help the user in certain situations by providing simpler information; e.g., by reducing the number of objects in the scene that are encoded and rendered. The 'flashlight' tool provides a very simple encoding of the distance from camera to the first object touched by an imaginary line going straight out from the camera. It is helpful to carefully explore a scene and allows accurate perception of distances to objects' surfaces, and especially of their margins. The 'Best free space' tool gives the user a simple indication of the open space where he/she can navigate. Other tools offer the functionality of texts detection and reading, signs detection and encoding, and TTS scene description. Another important feedback provided to the user is represented by the notification for Dangers'. The system renders to the user, through acute, hard-to-miss stimuli, the potentially dangerous elements in the scene—i.e., head-level obstacles or holes on the ground that are located on collision course, in a specified range.

3.2. *Virtual Training and Testing Environment (VTE)*

Before testing the SoV final prototype in real-world scenarios, every user participated in a short training and testing session using the Virtual Training and Testing Environment (https://youtu.be/hBay25-KN10, accessed on 5 July 2021) [26,28]. The VTE offers a series of 3D scenes meant to train, evaluate and improve user skills and their familiarity with the SoV system. The VTE integrates three operating modes: learning—tasks are presented individually and the user can switch between various tasks, audio and haptic models, and through this mode the user can learn the audio/haptic feedback associated with the task; practice—the user pre-tests the learned information and receives feedback (correct/wrong) as well as additional details if required; and testing—the users test their ability to correctly answer the task, no feedback involved.

The main goals of the usability evaluation included in the present study concern the ability of visually impaired persons to use the system in the "wild", in outdoor environments. These represent the most complex environments for the SoV sensory substitution device and its users. Thus, prior to performing any training or testing in outdoor environments, the users required to have a minimal experience with the system in the Virtual Testing and Testing Environment and indoor environments:

- The already trained users participated in one VTE and indoor training session, where they became familiar with the latest updates on the SoV system (changes or additions to the encodings, using the remote control, etc.).
- The new participants went through full training and testing in VTE (Single attributes, Frontal Pickups, Passing between, Treasure hunt—Boxes) and indoor (Frontal Pickups, Passing between, Treasure hunt—Boxes).

During the VTE and indoor training session(s), the users were presented with all the audio and haptic encodings available in the SoV system. After that, they were able to select the models according to their preferences and even switch between them during the training and testing sessions. This feature is available using the remote control by the visually impaired or by the trainer. The outdoor sessions contained both training and testing exercises. The goal of training was to make the users familiar with using the SoV system to identify real-world obstacles (both generic and special objects), both in ego-static scenarios and mobility scenarios. The training and testing scenarios contained both static and dynamic obstacles in natural outdoor scenes (in the parking lot, on the sidewalk, etc.) in usual environmental noise conditions. Each scenario started with training ego-static perception of the presented scene and continued with training the use of the system while moving in that scene.

Each outdoor session started with training in predefined scenarios and ended with a set of tests related to the scenarios trained in that session. For each scenario, the user trained the perception of the environment (i.e., ego-static), followed by training the mobility in that specific environment (i.e., ego-dynamic). For most of the scenarios, the egostatic training started with an environment containing only static objects, followed by adding dynamic objects, too. The dynamic obstacles were at first represented by moving persons (SoV team members). Uncontrolled environments were gradually added to the training and testing where dynamic obstacles were represented by any person, bicycle, or car passing by. The testing scenarios were different from the trained ones to minimize the effect of learning the environment. At most, it was acceptable to use a trained setup, but starting with a different position and orientation for the user. Figure 2 illustrates a map of the University campus and downtown Iasi where the locations of the training and testing environments are marked.

Figure 2. Geographical localization of the environments used for training and testing.

The outdoor training time was minimal, it had small variations between users, depending on their skills with the system. There was a number of 4 sessions of 2 h planned for training (TR) and testing with each user (Table 1). However, due to weather conditions (temperatures approaching 0 °C making some of the testing equipment work inappropriately) and users'availability, some sessions were shorter than 2 h, and thus some users were invited for one extra session.

Table 1. Overview of the training and testing sessions in outdoor environments (TR—training session, PT—perception testing session, MT—mobility testing session).

Session	Training	Testing
1	Training in the parking lot—generic objects, static and dynamic (environment TR1); training with elevated objects—tree branches (environment TR 4)	Testing PT I
2	Training with walls (environment TR 4); training with negative obstacles—holes in the ground (environments TR2 and TR3)	Testing PT II
3	Training in the parking lot—generic objects, static and dynamic(TR1), on the sidewalk, finding the bus stop (environments TR6)	Testing MT I
4	Training mobility on the sidewalk, finding the bus stop (environment TR5)	Testing MT II

3.3. Data Collected During the Experiments

For each test, the following data was collected: use of encodings, electrodermal activity (Galvanic Skin Response, Heart Rate)—only during mobility tests, video footage, and user feedback. The electrodermal activity data was collected using a device (GRS + Shimmer) that is mounted on the fingers and hand. This setup incapacitates the use of the remote control with the respective hand. The other hand is used to hold the white cane in the corresponding tests. Thus, it was decided that the users would not change the selected encodings during a test. However, they are able to choose their favorite combination before each test, after the task is explained to them. The video footage was processed offline to extract the completion time and accuracy metrics for each test. After each test, the users responded to the task related questions. The users also responded to the system general questions, once after the finalization of the Perception Tests (PT) and once after the finalization of the Mobility Tests (MT).

3.4. Testing Equipment

The equipment used in the testing sessions was composed of the following devices:

- SoV system (using as main input the video feed from the stereo camera);
- GRS + Shimmer for electrodermal activity recording;
- SoV Test Utility for automatic recording of encodings usage and time (application running on the SoV device);
- Tablet for remote connection to the SoV device (used for inspection of system status, for starting/stopping the recording of the test session using the SoV Test Utility, for starting/stopping the recording of electrodermal activity data);
- Wi-Fi bundle to ensure connection between the SoV device and the tablet.

3.5. Study Design

The evaluation tests are divided in two categories: Perception Tests (PT)—evaluation of the SoV prototype and assessment of its usability compared to the white cane for perception; Mobility Tests (MT)—evaluation of SoV usability for mobility.

Some mobility testing scenarios were performed by the users in three conditions: (1) with the white cane only, (2) with SoV and white cane, and (3) with SoV only. This allowed a comparison between the performance in the three use cases. To this end, some testing scenarios (i.e., particular environments) were carefully selected such that they contained the same structure of the environment in all cases. Still, outdoor environments are highly dynamic in change. Thus, testing in such scenarios can also pose a high uncontrollable variance between different tests and users, even if following the same course in the same environment. To account for the variance, the analysis of averaged results across users was performed by weighting the results based on the difficulty of the course. The difficulty level was assigned based on the total number of obstacles. In order to collect meaningful data, some testing scenarios were "fabricated". However, they still contained real-world objects and resembled for as much as possible situations that users can encounter in the "wild". Operating in such semi-controlled environments also helped ensure the safety of the visually impaired participants. For safety reasons, during all training and testing sessions, the participants were closely assisted by at least one sighted test assistant. The role of the test assistant was to take care of the safety of the test-taker and stop him/her before running into any dangerous situations.

The perception tests (Table 2) were aimed at evaluating how the visually impaired participants perceive the environment in ego-static scenarios (i.e., the user is standing in a fixed position). They were set in realistic outdoor environments, with various noise conditions, containing static and dynamic elements (cars, poles, trees, people, buildings), as well as generic and special (walls, holes in the ground, signs) types of objects that can be identified and signaled by the SoV device. For safety reasons, the training and testing sessions with negative obstacles (holes in the ground) were performed without the user moving in the scene. The identification of negative obstacles was only evaluated in the tests in the perception category and not in the mobility one. The distance and direction for dynamic objects were not evaluated (a dynamic object is an object that moves in the scene, changing its direction and/or distance to the user).

Table 2. Overview of the outdoor tests designed for evaluation of environment perception with the SoV device (PT—perception test).

Setting	Test ID	Scenario Details	Tasks	Performance Metrics
PT I—complex scenes with generic objects (static, dynamic, hanging)	PT I-1	Generic static objects (car, person, bush, tree, pole)		
	PT I-2	Generic static and dynamic objects	The user is asked to identify:	
	PT I-3	Generic static and elevated objects	How many objects are present in the scene	
	PT I-4	Generic static, dynamic and elevated objects		Accuracy
PT II—complex scenes with generic objects and special objects (static, dynamic, negative obstacles, walls)	PT II-1	Generic static and special objects (negative obstacles: hole in the ground)	Type of objects	Completion time (static, dynamic, special)
	PT II-2	Generic static, dynamic and special objects (negative obstacles: hole in the ground)	Localization of objects (distance, direction)	
	PT II-3	Generic static and special objects (wall)	Elevation of objects	

The mobility tests (ego-dynamic scenarios) were aimed at evaluating how well the users can guide their interaction with the environments based on the information received from the SoV device. Two types of scenarios were considered: semi-controlled environments (Table 3 MT I) and uncontrolled environments (Table 3 MT II). The users were asked to walk on predefined routes fulfilling specific tasks. By semicontrolled environments, we denote natural areas, usually containing short testing routes (15–30 m long), with no or light traffic, for which the testing team could control the structure of the scene. This ensured the presentation of the same scene and tasks to all participants. Specific static and/or dynamic obstacles (i.e., people) were purposely and systematically introduced in some testing scenarios. By uncontrolled environments we understand public areas, with varying, uncontrollable traffic. This only allows for qualitative (as opposed to quantitative) performance evaluations and comparison of performance between users and modalities (i.e., white cane, SoV + cane, SoV). Still, it better reflects the usability of the system in the targeted real-life environments.

Table 3. Overview of the tests designed for evaluation of mobility in outdoor environments with the SoV device (MT—mobility test).

Setting	Test ID	Scenario Details	Tasks	Performance Metrics
MT I—semi-controlled environments (static, dynamic, hanging objects, walls)	MT I-1	Walking by a wall (30 m path): static obstacles on the path	The user is asked to: Identify the wall; Walk along the wall at a certain maximum distance from it;	Accuracy Time to completion Collisions
	MT I-2	Walking by a wall (15 m path): hanging obstacles on the path	Identify and avoid the obstacle(s) on the path; Identify the corner of the wall.	
	MT I-3	Walking in a parking lot (25 m path): parked cars, dynamic obstacles on the path	The user is asked to: Walk along parked cars at a certain maximum distance from them; Identify the last parked car on the left/right of the course; Identify and avoid dynamic obstacles on the path.	
MT II—uncontrolled environments (static, dynamic, hanging objects, signs)	MT II-1	Walking on the sidewalk (45 m path): static and dynamic obstacles on the path, bus stop	The user is asked to: Walk on the sidewalk; Identify and avoid static and dynamic obstacles; Identify and stop at the bus stop	
	MT II-2	Walking on the sidewalk (250 m path): static and dynamic obstacles on the path	The user is asked to: Walk on the sidewalk; Identify and avoid static and dynamic obstacles;	

3.6. Description of the Perception Experiments

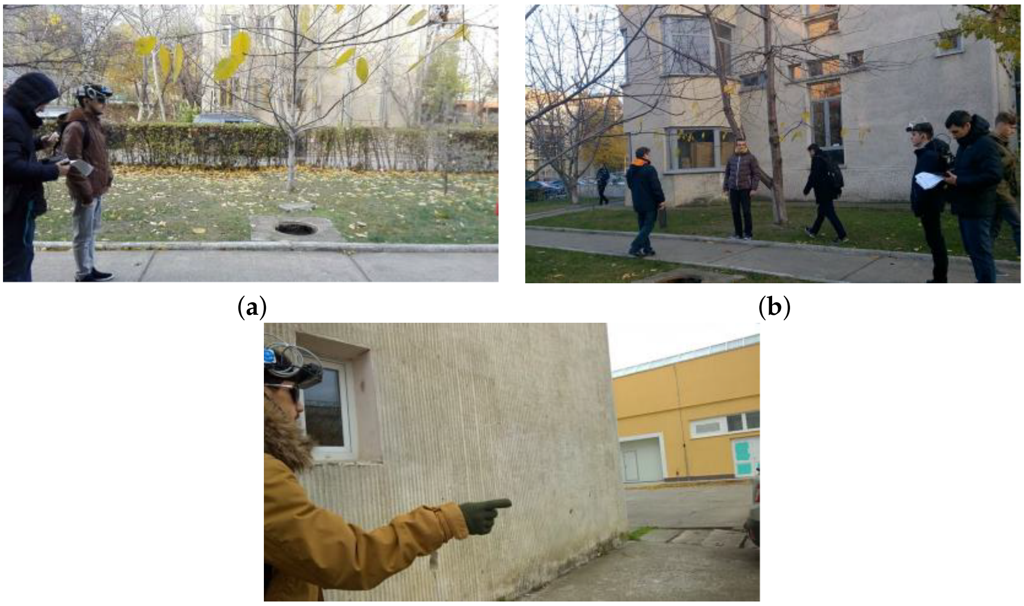
Only two types of scenarios were selected for evaluation: complex scenes with generic objects (Table 2 PT I) and complex scenes with generic and special objects (Table 2 PT II). These are ego-static tests, where the visually impaired user is standing at a fixed point within the real-world environment and interprets the varying scenes presented to him/her. The user is only relying on the SoV system. An overview of the scenes selected for testing is presented in Table 2. Even though the user is not changing their position, he/she can look around in the environment. This is important, as head turning is a fundamental part of orientation for sighted people, with the role of expanding the visual field. The SoV training program encourages visually impaired people to use the device in the same way. Moreover, unlike the tests performed previously in Virtual Testing Environment [28] (VTE) and indoor real-world, the outdoor tests also included the presence of dynamic obstacles. The main goals of the perception tests were to evaluate whether the visually impaired participants:

1. Are able to perceive the environment using the SoV device from a stationary position.
2. Are able to identify obstacles and recognize specific objects in static environments.
3. Are able to identify both static and dynamic obstacles and recognize specific objects in real-life environments.

The first type of perception tests (PT I) included only generic objects. These objects were represented by people, cars and poles (Figure 3). The tests were performed in an order of gradually increasing difficulty, starting from a scene with static objects (PT I-1, Figure 3a), and then adding a dynamic object (PT I-2, Figure 3b). In the next step, the perception of elevated objects was tested, first in a static scene (PT I-3, Figure 3c), then in a scene with dynamic objects (PT I-4, Figure 3d). A piece of cardboard was held by one test assistant at head level to represent the elevated object. The second type of perception tests (PT II) included generic and special objects. The special objects considered were walls and negative obstacles, i.e., a hole in the ground (Figure 4). The hole in the ground was represented by a sewer for which the cap was removed for the purpose of the tests.

(**a**) (**b**) (**c**) (**d**)

Figure 3. Examples of testing scene setups for perception tests with generic objects: (**a**) Scene with generic static objects (PT I-1); (**b**) scene with generic static and dynamic objects (PT I-2); (**c**) scene with generic static and elevated objects (PT I-3); (**d**) scene with generic static, dynamic and elevated objects (PT I-4).



(**a**) (**b**) (**c**)

Figure 4. Examples of testing scene setups for perception tests with special objects: (**a**) scene with generic static objects and hole in the ground (PT II-1); (**b**) scene with generic static, dynamic objects and hole in the ground (PT II-2); (**c**) scene with generic static objects and wall (PT II-3).

The system was used with a fixed distance range of 5 m. The scenes contained 3 to 5 objects in this range, with varying distance and orientation to the user. Some tests were performed in the same physical location (e.g., parking lot). However, the user position and orientation were changed to maximize the difference between consecutive test scene layouts. The system output was paused while preparing the testing setup or switching between consecutive testing scenarios. The system output was turned on only after the testing setup was prepared and the user was positioned and oriented accordingly.

Before starting the test, users are instructed:

- To stand still and analyze the presented scene by scanning with the head.
- To inform the testers immediately in case the representation stops or the audio sounds are distorted, etc. In that case, the users are asked to stop, and the testers pause the completion time while fixing the problem. In case it can be fixed directly, the user is asked to proceed with the scene, and the time measurement continues.
- For each tested scenario, the user is informed about the context and the tasks he/she is supposed to perform:

  You will be presented with a scene including several objects at the same time, each with varying distance, direction, size and quantity. You will hear (feel) how the scene is represented with sounds (vibration patterns). Your task is to try to understand the scene, the relation between the different objects, trying to get an inner picture of its composition. You will be asked the following questions for each scene in random order, one at a time. As soon as one of these questions is asked, you should answer it verbally:

  1. *How many objects do you perceive?*
  2. *Out of the perceived objects, how many are special objects (i.e., wall, hole in the ground)?*
  3. *If any special objects, specify their type.*
  4. *Out of the perceived objects, how many are dynamic?*
  5. *For each static object, indicate: the distance to the object, the direction of the object, the elevation of the object.*

- To choose their favorite encoding for audio and haptic, which they are not be able to change during the test.

3.7. Description of the Mobility Experiments

The mobility tests included two types of scenarios: semi-controlled environments (MT I) and uncontrolled environments (MT II). These are ego-dynamic tests, where the VI user is asked to walk on a predefined route fulfilling specific tasks. An overview of the scenes selected for testing is presented in Table 3.

The main goals of the mobility tests were to evaluate:

1. *Whether the visually impaired participants are able to use the information from the SoV device to guide their interaction with the environment.*
2. *Whether the visually impaired participants are able to move around and avoid obstacles using the SoV device.*
3. *The efficiency of navigation with SoV device compared to using only the cane.*

For two testing scenarios, i.e., MT I-1 and MT II-2, the course was performed in three conditions: (1) only white cane, (2) SoV + white cane, (3) only SoV. Scenarios MT I-3 and MT II-1 were performed in conditions (2) and (3). Using only the white cane cannot solve scenarios involving the detection of hanging objects or signs. Thus, in testing these scenarios, the condition of using only the white cane was omitted.

The first type of mobility tests (MT I) were performed in semi-controlled environments (Figure 5). The tests evaluated the usability of the system on short paths in natural identical set-ups for all users. For scenario MT I-3, a dynamic obstacle was introduced in the scene by having a SoV team member walk towards the user on their path. The second type of mobility tests (MT II) were also performed on the same courses for all users and for all modalities. However, the environment was not controlled in any way (Figure 6). The difficulty of the course with respect to the number of obstacles varied between different instances of the tests. To account for this variability when comparing the performance between users and between different modalities on the same course (white cane, SoV and white cane, SoV), the performance metrics (time and collisions) were adjusted based on the number of obstacles present on the course (as described in the Performance Metrics section below).

Figure 5. Examples of testing scene set ups for mobility tests in semi-controlled environments: (**a**) Walking by a wall (30 m path), static obstacles on the path (MT I-1); (**b**) Geolocation of MT I-1 testing scenario; (**c**) Walking by a wall (15 m path), hanging obstacles on the path (MT I-2); (**d**) Geolocation of MT I-2 testing scenario; (**e**) Walking in a parking lot (25 m path), parked cars, dynamic obstacles on the path (MT I-3); (**f**) Geolocation of MT I-3 testing scenario.

The system output was paused while preparing the testing setup or switching between consecutive testing scenarios. The system output was turned on only after the testing setup was prepared and the user was positioned and oriented accordingly.

Before starting the test, users are instructed:

- To avoid collisions with obstacles as well as with walls on any sides.
- To inform the testers immediately in case the representation stops or the audio sounds distorted, etc. In that case, the users are asked to stop, and the testers pause the time while fixing the problem. In case it can be fixed directly, the user is asked to proceed with the scene and the time measurement continues.
- To choose their favorite encoding for audio and haptic, which they are not able to change during the test.
- For each tested scenario, the user is informed about the context and the tasks he/she is supposed to perform:
    1. *MT I-1: You are standing in close vicinity of a building. Please identify its wall and walk along the wall at a comfortable distance to it, no more than 2m. There might be static and/or dynamic obstacles on the path that you should avoid. Please indicate verbally when you have reached the corner of the building.*
    2. *MT I-2: Same as for MT I-1.*
    3. *MT I-3: You are standing in a parking lot. There are parked cars to your left/right. Walk along them until you identify the last parked car on the left/right of the course. There*

might also be other static and/or dynamic obstacles on the path that you should avoid. Indicate verbally when you have reached the last car.

4. *MT II-1: You are standing on the sidewalk. Walk on the sidewalk in front of you until you reach the bus stop. There might be static and/or dynamic obstacles on the path that you should avoid. Indicate verbally when you have reached the bus stop and point to the direction of the bus stop sign. You will be closely assisted by two test assistants who will stop you in case of any unsafe situation.*
5. *MT II-2: You are standing on the sidewalk. Walk on the sidewalk in front of you until you are told to stop (about 250 m). There will be static and/or dynamic obstacles on the path that you should avoid. There are two small side streets that you will cross, which are not equipped with traffic lights or pedestrian crossings. Report if you think it is safe to cross, otherwise stop. You will be closely assisted by two test assistants who will stop you in case of any unsafe situation.*

The users were NOT informed about the scene layout, meaning there were no instructions on the number, location and size of objects or about the length of the course they are supposed to follow.

Figure 6. Examples of testing scene set ups for mobility tests in uncontrolled environments: (**a**) Walking on the sidewalk (45 m path), static and dynamic obstacles on the path, bus stop (MT II-1); (**b**) walking on the sidewalk (250 m path), static and dynamic obstacles on the path (MT II-2); (**c**) geolocation of MT II-1 testing scenario; (**d**) geolocation of MT II-2 testing scenario.

3.8. Performance Metrics

The following metrics were used to assess the performance for the perception tests:

- The accuracy is computed based on correct or incorrect answers given by the user to each of the addressed questions in each tested scenario.
- Completion time is computed as the time between the moment when the SoV system is turned on at the beginning of the test (this is also when the user starts perceiving the environment) and when the answer to the last question is provided by the user. That is, the completion time is measured per test and not per task.

A perception test is considered as completed with success if the average accuracy over all the tasks (questions) in the test is greater than 85%.

The metrics used to assess the performance for the mobility tests are:

- The accuracy for a task is reflected by identification of the elements of interest (e.g., wall, corner of the building, hanging object, bus stop), and/or leaving the

testing area (i.e., distancing from the indicated shoreline, leaving the sidewalk by walking to the street).
- The number of collisions, where only major collisions with obstacles are considered. By major collisions, we mean contacts with objects the participant was completely surprised about, while minor collisions are considered when the user brushes the obstacles.
- The number of cane hits was accounted for in some scenes, considering only the hits on obstacles.
- Completion time is measured between the moment when the SoV system is turned on at the beginning of the test (this is also when the user starts perceiving the environment) and when the task is finalized with success.

A mobility test is considered as completed with success if the elements of interest have been identified and the user has not left the testing area, irrespective of the number of collisions or completion time. Due to the high variance of the number of obstacles on the MT II-2 course between the tests performed with different users and between tests with different modalities for each user, a supplemental measure was introduced: *Adjusted Completion Time*. This measure allows for comparing between different instances of the same test, by weighting the actual measured time taken to complete the course with the number of extra obstacles found on the course. We consider that the course contains a fixed number of obstacles in all instances (represented by trees, poles, benches, bushes), besides which, a number of extra obstacles, static or dynamic, were present on the course (people, parked cars, dynamic cars, bicycles). Thus, we define the *Adjusted Completion Time (ACT)* for scenario MT II-2, to be:

$$ACT_{MTII-2} = CT \times \frac{O_a}{O_{crt}}, \quad (1)$$

where CT is the measured completion time, O_a is the average number of extra obstacles, over all test instances of all users in mobility scenario MT II-2, and O_{crt} is the number of extra obstacles in the current test instance. Thus, if the number of extra obstacles in one test is higher than the average number of extra obstacles over all MT II-2 tests, the ACT value will be lower than the measured completion time (CT).

3.9. Ethical Aspects

All the investigations of the present study that involved tests with human subjects were carried out following the rules of the Declaration of Helsinki of 1975 (https://www.wma.net/what-we-do/medical-ethics/declaration-of-helsinki/, accessed on 5 July 2021), revised in 2013. According to point 23 of this declaration, an approval from the Research Ethics Committee of the "Gheorghe Asachi" Technical University of Iasi, Romania, was obtained before undertaking the research: research ethics assent no. 13582/05.07.2016 for the activities of the project no. 643636 (H2020), Sound of Vision—Natural Sense of Vision Through Acoustics and Haptics. Before starting any testing (e.g., questionnaire or "first hands on" period), the participants were informed about the project aims, the general function of the device and the aim of the tests. Further, the participants were informed about the test methods, their tasks and about possible risks involved in the testing but also the project teamș efforts to minimize them. Additionally, the participants could agree on being filmed for the project purposes. Afterwards, the participants confirmed that they were informed and agreed on participating by signing the informed consent. According to the ethical approval requirements, that restrict reporting of the individual data, in the following, all users will be referred to as male, independently of their gender. Further, the age of the users will not be reported individually but they will be referred to as young (20–30), middle-aged (31–39) and older users (40–50).

4. Results of the Usability Experiments

Four visually impaired users, aged between 20 and 42, one female and three male, participated in training and testing sessions which took place at the Technical University of Iasi. An overview of relevant information regarding the user profiles is presented in Table 4. Three of the testers were already familiar with the SoV system, since they were also invited to the training and testing sessions with an early SoV prototype. However, because improvements have been made (hardware, software, audio and haptic models) since the last sessions of training/testing, these users undergone one training session in VTE and indoor in order to refresh their memory and to present them the new features of the system. Furthermore, a new user, without any previous experience with the SoV system participated in the training/testing sessions. The newcomer followed the procedure of training and testing in VTE and indoor before outdoor sessions. He had to familiarize with the system and to learn its operating modes (audio and haptic). The users belonged to categories 4 and 5 of visual impairment as defined by the World Health Organization (WHO). Two users belonged to category 4, a category of blindness, meaning that visual acuity is less than 10% (FC at 1 m) and equal to or better than light perception. Two users belonged to category 5, total blindness, meaning no light perception. None of the 4 users use echolocation to guide their interaction with the environment. Users of category 4 were NOT blindfolded during the training and testing sessions. The main purpose of the tests was to evaluate the usability of the system in the "wild". Thus, blindfolding the users would be in contradiction with the realistic way in which they would use the system.

Due to light conditions, weather and also being restricted by users' time to participate both in training and testing in outdoor scenarios, the results detailed below are obtained with a minimum training time. With proper training, we assume that users will be able to improve and will obtain better or even excellent results. Thus, it is important to keep in mind that the results listed are based on minimal training in outdoor scenarios.

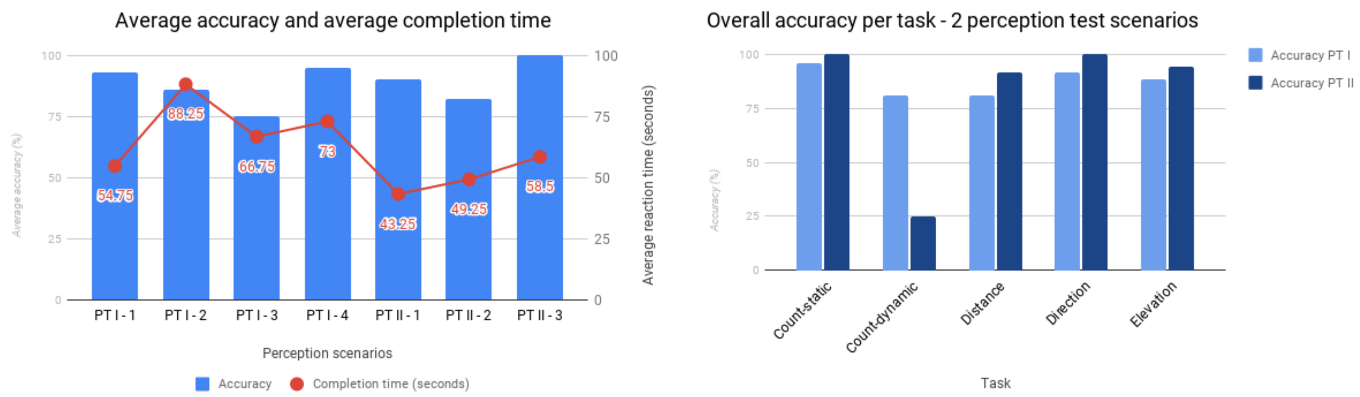
Average accuracy and completion times in the perception tasks are presented in Figure 7a,b. The most time consuming test was PT I-2 (first complex scene in which dynamic objects were introduced), while the smallest accuracy was recorded for the PT I-3 scene (first complex scene in which elevated objects were introduced). However, perceiving elevation of objects was not the most difficult task. Moreover, the perception of almost all object properties improved from PT I to PT II thanks to the additional training time with the system between the testing sessions. The only exception is represented by the dynamicity of objects. Due to the powerful and distracting sound made by the negative obstacles, 3 out of 4 users were not able to correctly identify if the scene contains a dynamic object or not, when holes in the ground were present. However, users ranked above 80% accuracy regarding special objects: they obtained 100% accuracy in counting them, 83% accuracy in indicating the correct distance and 91.6% accuracy in indicating the correct direction.

The perception tests conducted in ego-static scenarios helped us to conclude that the participants using the SoV device are able to perceive the environment even if they are not moving, and all users obtaining an average accuracy greater or equal to 85% in both types of scenes. The overall average accuracy in the perception tests was 89.3%.

Thanks to the improvements in audio models (danger mode and to different encodings for special objects) users correctly identified obstacles (possible dangers) and special objects (hole, wall) in static environments. As a result, all of them correctly identified the special objects (100% accuracy) and three out of four users correctly pointed to the direction and approximated the distance to the special objects. As we have mentioned before, in complex scenes, where both negative obstacles and dynamic obstacles are present, participants had difficulties in identifying the presence of a dynamic object in the scene (mostly because their attention is directed to the negative obstacle), but the accuracy of identifying and localizing static objects increased even with a small amount of training. Moreover, the perception of dynamic objects was the least trained aspect in the VTE sessions, even no training at all was performed by User 4.

Table 4. Description of the SoV users.

	User 1	**User 2**	**User 3**	**User 4**
age: young, middle-age, old	middle-age	old	young	young
visual impairment category (according to WHO)	4	5	5	4
travelling in unknown environments	accompanied by sighted person	accompanied by sighted person	accompanied by sighted person	accompanied by sighted person
white cane user	uses the white cane but is not an experienced user	uses the white cane but is not an experienced user	no	no
level of experience	training and testing beginning with the first prototype	training and testing beginning with the first prototype	training and testing with the prototype before the final one	no previous training and testing
preferred audio model	the expanding sphere model with the impact sounds	the expanding sphere model with the bubble sounds and the flashlight audio encoding	the expanding sphere model with the bubble sounds, the bubblestream and flashlight	the expanding sphere model with the impact sounds
preferred haptic model	closest point	closest point	closest point	closest point
others	likes to sing, passionate about smartphones and IT technologies	active member of the local blind community, participates in different competitions and social activities dedicated to visually impaired	studies foreign languages	loves music, plays 7 instruments (e.g., flute, pan, oboe), studies pan flute at the Arts University, participates in various cultural activities



(**a**) (**b**)

Figure 7. (**a**) Average accuracy and completion time over all users in perception tests (**b**) Average per task accuracy over all users in both perception scenarios.

Another important aspect addressed in the evaluation was the comparison between SoV and the white cane in mobility scenarios. To this end, scenarios MT I (walking by a wall) and MTII-2 (walking on the sidewalk) were employed with three modalities: using the SoV device together with the white cane, only using the white cane, only using the SoV device. As expected, the average time for completion was the lowest when using the white cane in both scenarios (Figures 8 and 9). With the SoV system, users can walk at a reasonable distance from the wall in MT I, without the need to hit it with the cane (average of 23 cane hits when using the white cane only, compared to 3.5 hits when using the white cane and SoV) and could easily detect the corners of the buildings. Users with no experience of either of the two devices tend to walk slower when using both modalities than only with either of them, as the amount of information coming from both devices can be overwhelming. This is also confirmed by the results of the MT II-2 test: the average time (adjusted based on the average number of additional obstacles on the course in each modality) to complete the course in the conditions with a single assistive device (cane only and SoV only) was very similar, but much less than in the condition with both devices.

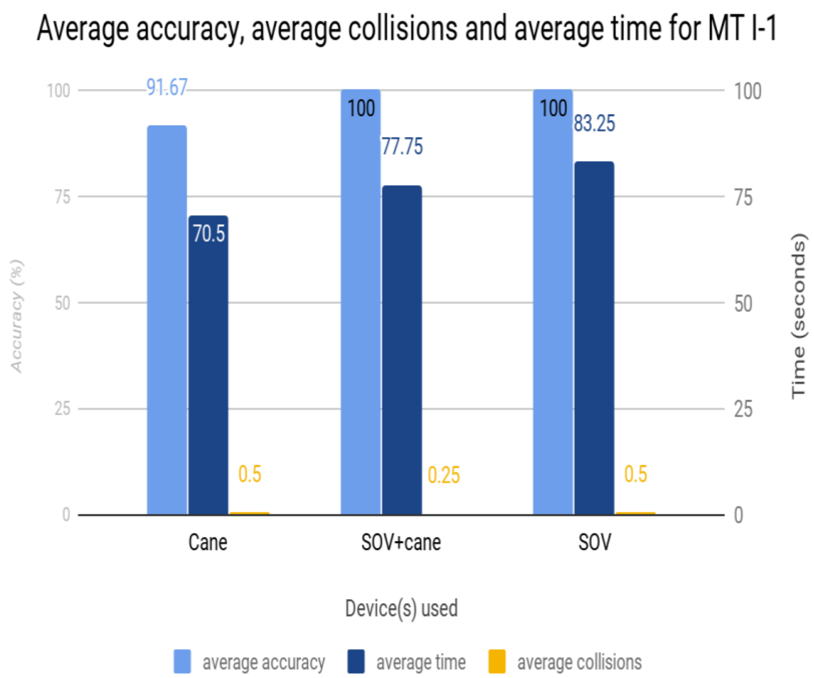


Figure 8. Average accuracy, collisions and completion time over all users in mobility scenario MT I-1 (walking by a wall) when using the SoV system alone, SoV and white cane, white cane alone.

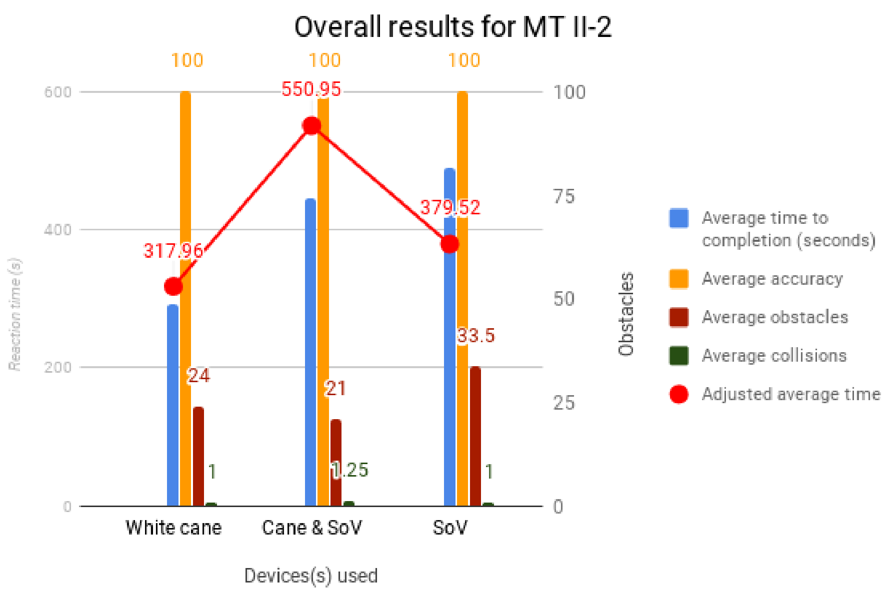


Figure 9. Average results over all users for the MT II-2 scenario (walking on the sidewalk) with the 3 modalities: SoV, white cane, SoV and white cane.

An overall measure for expressing the completion rate of the tasks with a system is the *effectiveness*. It can be used to evaluate the current version of the SoV device, being a measure that embeds accuracy and task completion. Considering the following notations N-the number of the scenarios, R-the number of users, n_{ij} the result of coming through scenario i by respondent j, if the user successfully completes the task, then $n_{ij} = 1$, if not, then $n_{ij} = 0$, t_{ij} the time spent by user j to complete task i. If the task is not successfully completed, then time is measured until the moment the user quits the task; the overall product effectiveness E can be computed using:

$$E = \frac{\sum_{j=1}^{R} \sum_{i=1}^{N} n_{ij}}{R \times N} \times 100\%, \tag{2}$$

its statistic error being:

$$\sigma = \sqrt{\frac{E \times (100 - E)}{R}} \tag{3}$$

The data used for computing the effectiveness of the SoV system was obtained by merging all tests (perception and mobility). A number of 65 scenarios for all four users were considered. If a task i is performed by user j with accuracy greater than 85%, then $n_{ij} = 1$. Otherwise, $n_{ij} = 0$. Based on this data, the effectiveness of the SoV system is $E = 88.85\%$, with a statistic error $\sigma = 15.74\%$.

Taking into account that for successful task completion a high accuracy threshold of 85% was considered, the value of the obtained effectiveness underlines a very good overall performance of the SoV system.

An exploration of the individual performance of each of the four users is provided as Supplementary Material for the paper.

5. Discussion

To summarize, the initial questions we posed were all addressed with the performed tests. In the following, we provide a discussion of the combined results with respect to each aspect that was evaluated.

Are the visually impaired (VI) users able to perceive the environment (perception)? Are they able to identify obstacles and specific objects (negative obstacles, hanging obstacles, signs, walls) which define the added value of SoV compared to using the white cane? Is the system usable in real life environments and under real life circumstances (outside laboratory setups)?

Identifying the objects in the environment and their individual properties (position, size, elevation) is important for both perception and mobility. Perception of the environment with the SoV system was therefore evaluated in both ego-static and ego-dynamic scenarios. Evaluation of perception was specifically addressed in ego-static scenarios in the virtual training environment and the tests performed in real-world scenarios. The perception tests revealed impressive accuracy scores in counting the objects in real-world complex scenes (97.5%). The participants could also identify their distance well (86.45% accuracy) and direction (95.8% accuracy). Detecting the elevation of objects is important for avoiding head-height obstacles. As users noted, it is sufficient to be aware of the presence of the obstacle, not necessarily its exact elevation. Still, understanding this property with the help of the SoV device appeared to be easy, given the 91% accuracy obtained in real-world outdoor tests.

Identification of the presence of special objects in the scene was performed with very good results. Localizing walls proved to be easy to perform with the SoV system (100% accuracy obtained in the tests). This was also the case for identifying holes in the ground, represented by missing sewer caps in real-world outdoor tests (100% accuracy in identifying their presence in the scene, 87.5% accuracy in determining their direction and 75% for their distance). Another added value of the SoV device compared to using the white cane lies in the identification of signs. This aspect was tested as part of a mobility scenario involving walking on the sidewalk and identifying a bus stop. Remarkably, this task was completed with a 100% success rate and was considered very easy to perform by all the participants.

The perception of dynamic obstacles was also evaluated. It seems that identifying the dynamicity of objects (i.e., whether an object is moving or not) is more difficult to perform when special objects (e.g., holes in the ground) are also present in the scene. As the participants themselves explained, the negative obstacles are represented with a powerful sound that draws most of their attention. Aspects that require more concentration can therefore be misinterpreted. Still, it is important to note here that the perception of dynamic objects was the least trained aspect, so there is room for improvement.

The tests performed outside laboratory setups revealed that the individual properties of objects in complex scenes can still be perceived, even in natural outdoor scenes in the presence of high environmental noise. The participants in the tests obtained an overall 89.3% success rate in identifying object properties in such scenarios. The easiest to understand were the number and direction of objects, while the most difficult aspect was whether an object was moving or not (dynamicity).

Are the VI users able to use the information from the SoV device to guide their interaction with the environment (mobility)? Are they able to move around and avoid obstacles? Are they able to move around and identify targets (e.g., bus stop, corner of a building)? How is their mobility performance with the SoV system compared to traditional assistive devices (i.e., white cane)?

Mobility in outdoor environments was evaluated in the tests performed. The main aim was to assess the system usability in real-life environments and also its added value compared to the white cane. The results of these tests show that the SoV system offers the clear advantage of informing users about the presence of objects which could not be otherwise detected with the cane (head level objects, signs) or that could be missed by it (holes in the ground). Moreover, the system provides a good solution for detecting walls, which are frequently used by VI people as a shoreline during navigation. With the SoV system, users can walk at a reasonable distance from the wall, without the need to hit it with the cane and could easily detect the corners of the buildings. Head level obstacles were identified and avoided with 100% success rate in the tests. The primary reason for failure was when the participants walked too fast while "looking" down, so that the head-height obstacles (i.e., tree branches) were out of the camera's field of view until they were very close. This result also emphasizes the need to train the VI users to hold their head in positions similar to those of the sighted and scan the environment with the cameras like the sighted do with their eyes.

It is important to note that, while most of the participants had some level of experience with the system in VTEs and indoors, their outdoor training time was minimal (average of 2 h).

The SoV system offers a rich perception of the environment, which is not by far available with the white cane. White cane users have no to very little information about the environment especially in static scenes (where they do not move). They can perceive only the objects in their immediate proximity, as far as they can scan the scene with the white cane, no information is available further than the length of the white cane. On the other hand, the SoV system offers the possibility to acquire information about surrounding objects further away from the user (5–10 m), so one can have a better, more complex and early understanding about the environment. Furthermore, even when users cannot successfully identify if an object is hanging, they are still able to perceive its presence, while with white cane it is almost impossible to detect a hanging object without bumping into it.

With minimal training on the SoV prototype outdoors, the users could perceive and navigate in the testing environments with very good accuracy. An overall performance analysis revealed an effectiveness of 88.85% for the SoV device. The effectiveness of a system is a measure that indicates the completion rate of the tasks with the system. The value was obtained considering all perception and mobility tasks performed by the users, where a task was considered successfully completed if the accuracy per task was higher than 85%. That is, the users were able to complete 88.85% of the tasks with an accuracy greater than 85%. The most difficult task for the users was to identify the dynamicity of objects (i.e., to count the dynamic objects), especially in ego-static scenarios and when negative obstacles were also present Figure 7b. The users reported that, in these scenes, they mainly focused on the negative obstacles due to the powerful and distracting sound associated with it. Still, the perception of dynamic objects was the least trained aspect in the VTE and indoor sessions.

While the elevation of objects was not identified with maximum accuracy in perception tests, the users considered that avoiding hanging obstacles was easy to perform in mobility tasks. This feedback is also confirmed by users' performance results in task MT I-2, where all participants were 100% accurate in identifying and avoiding the tree branches.

The tests with special objects indicate that they could easily be perceived. The VI users could identify the presence of walls, holes in the ground and bus stop signs with 100% accuracy. They indicated the correct distance to such objects in 83% of the tests, and the correct direction in 91.6%.

With minimal training in using the SoV system outdoors, the users could perform the mobility tasks with very good accuracy. When comparing SoV with the white cane, we found out that, for our sample of rather inexperienced white cane users, mobility with the SoV device was accomplished with performance comparable to one with the white cane, and sometimes better. Analyzing the average walking speed over all users and all scenarios in which each modality was used revealed that, while using the white cane is the fastest, the SoV system has the advantage of reducing the average number of collisions.

User feedback was collected through questionnaires containing task specific questions, general questions about the system as well as individual user comments and suggestions. All items were designed to fit an answer format of a 5-point Likert-like scale, where 1 corresponds to strong disagreement and 5 corresponds to strong agreement. The feedback on how the system helps the users in accomplishing the tasks is summarized below as AVG (STD) values over all perception and mobility tests and all users:

T1—I found it easy to do this task with the device. Perception tests—4.39 (0.74), Mobility tests—4.70 (0.47)

T2—The device provides a good solution to problems I encounter in this task. Perception tests—4.54 (0.64), Mobility tests—4.80 (0.41)

T3—I am satisfied with the amount of time it took to complete this task. Perception tests—4.50 (0.75), Mobility tests—4.85 (0.37)

All users liked the device, for both perception and mobility. None of them believes it is unnecessarily complicated. They found it rather easy to use (to operate and switch between modes). They all disagree regarding the possible inconsistencies of the system and believe it works similarly in both the virtual training and testing environment and the real world. They are confident that most people would learn to use the device quickly. They don't find the device cumbersome to use, and although they are expecting design improvements for its commercial version, they were satisfied with the shape and functionality of the tested prototype. They were confident when using the device, and the confidence grew even more after the mobility tests. The visually impaired participants felt safe when using the device, and this feeling was even more emphasized for the mobility scenarios. Half of the participants thought that the device was comfortable, while the other two suggested several improvements. They all agreed that the SoV device would enhance their capacity for leisure activities, especially after going through the outdoor mobility sessions.

6. Conclusions

With minimal training in using the SoV system outdoor, the perception and mobility in the environments were achieved with very good accuracy. The tests revealed 88.85% effectiveness (task completion rate) of the SoV system. Regarding the perception, the system could be successfully used for perception (89.3% average accuracy) in noisy outdoor environments, without the environment sounds to have an obvious effect on the performance of perceiving the environment. The hear-through feature of the SoV headphones was found very useful by the users. Perceiving the dynamicity of obstacles can pose difficulties, especially in the presence of negative obstacles, which are signaled by the system with a very powerful sound. In complex environments, the perception of individual obstacle elevation can pose difficulties to the visually impaired users. However, they all considered this to not be a major issue, since they can perceive the presence of these obstacles with the normal encoding and further have a distinctive feedback with the danger mode when approaching them closer than 1m. Even with a small amount of additional training time, perception of the environment improved from one testing session to a subsequent one.

With minimal training, the system could be successfully used in outdoor real-life environments to perform various mobility tasks. The visually impaired participants reported that performing mobility tasks with the SoV device was easier than building a detailed perception of complex scenes. They were also more satisfied about the solution provided by the SoV system for these tasks and by the time it took to complete them than for the perception tasks.

For inexperienced white cane users, mobility with the SoV device was accomplished with performance comparable to using the cane, and sometimes better. Users more experienced in using the white cane tend to rely more on the cane than on the system, when provided with both assistive devices. Less skilled white cane users chose to rely more on the SoV system. When using both modalities, users walk slower than only with the SoV system, as the amount of information coming from both devices can overwhelm users inexperienced with either of them.

The added value of SoV compared to the white cane was confirmed by the participants to consist in: providing early feedback about static and dynamic objects, providing feedback about elevated objects, walls, negative obstacles and signs.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/electronics10141619/s1.

Author Contributions: Conceptualization, S.C.; methodology, S.C. and R.-G.L.; software, O.Z., N.A.B. and S.C.; validation, S.C., A.B. and R.-G.L.; formal analysis, S.C.; investigation, A.B., S.C., O.Z., N.A.B. and R.-G.L.; data curation, O.Z., R.-G.L. and A.B.; writing—original draft preparation, O.Z.; writing—review and editing, all; supervision, S.C.; project administration, S.C.; funding acquisition, S.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the European Union's Horizon 2020 research and innovation program under grant agreement no 643636 "Sound of Vision" and by CNCS-UEFISCDI project PN-III-P2-2.1-PTE-2019-0810.

Institutional Review Board Statement: Rsearch ethics assent no. 13582/05.07.2016 for the activities of the project no. 643636 (H2020), Sound of Vision—Natural Sense of Vision Through Acoustics and Haptics, issued by the Research Ethics Committee of the "Gheorghe Asachi" Technical University of Iasi, Romania.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

SoV	Sound of Vision
VTE	Virtual Testing Environment
CT	Completion Time
TR	Training
PT	Perception Test
MT	Mobility Test

References

1. Di Mattia, V.; Caddemi, A.; Cardillo, E.; Manfredi, G.; Leo, A.; Russo, P.; Scalise, L.; Cerri, G. A Feasibility Study of a Compact Radar System for Autonomous Walking of Blind People. In Proceedings of the 2016 IEEE 2nd International Forum on Research and Technologies for Society and Industry Leveraging a better tomorrow (RTSI), Bologna, Italy, 7–9 September 2016; [CrossRef]
2. Long, N.; Wang, K.; Cheng, R.; Yang, K.; Weijian, H.; Bai, J. Assisting the visually impaired: Multitarget warning through millimeter wave radar and RGB-depth sensors. *J. Electron. Imaging* **2019**, *28*, 1. [CrossRef]
3. Long, N.; Wang, K.; Cheng, R.; Weijian, H.; Yang, K. Unifying obstacle detection, recognition, and fusion based on millimeter wave radar and RGB-depth sensors for the visually impaired. *Rev. Sci. Instrum.* **2019**, *90*, 044102. [CrossRef] [PubMed]
4. Yang, K.; Wang, K.; Hu, W.; Bai, J. Expanding the Detection of Traversable Area with RealSense for the Visually Impaired. *Sensors* **2016**, *16*, 1954. [CrossRef] [PubMed]
5. Maidenbaum, S.; Hanassy, S.; Abboud, S.; Buchs, G.; Chebat, D.R.; Levy-Tzedek, S.; Amedi, A. The 'EyeCane', a new electronic travel aid for the blind: Technology, behavior & swift learning. *Restor. Neurol. Neurosci.* **2014**, *32*, 813–824. [CrossRef] [PubMed]
6. Abraham, L.; Mathew, N.S.; George, L.; Sajan, S.S. VISION- Wearable Speech Based Feedback System for the Visually Impaired using Computer Vision. In Proceedings of the 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI) (48184), Tirunelveli, India, 15–17 June 2020; pp. 972–976. [CrossRef]
7. Rohit, P.; Vinay Prasad, M.S.; Ranganatha Gowda, S.J.; Krishna Raju, D.R.; Quadri, I. Image Recognition based SMART AID FOR VISUALLY CHALLENGED PEOPLE. In Proceedings of the 2019 International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 17–19 July 2019; pp. 1058–1063. [CrossRef]
8. Kang, M.C.; Chae, S.H.; Sun, J.Y.; Yoo, J.W.; Ko, S.J. A novel obstacle detection method based on deformable grid for the visually impaired. *IEEE Trans. Consum. Electron.* **2015**, *61*, 376–383. [CrossRef]
9. Kang, M.C.; Chae, S.H.; Sun, J.Y.; Lee, S.H.; Ko, S.J. An enhanced obstacle avoidance method for the visually impaired using deformable grid. *IEEE Trans. Consum. Electron.* **2017**, *63*, 169–177. [CrossRef]
10. Peng, E.; Peursum, P.; Li, L.; Venkatesh, S. A Smartphone-Based Obstacle Sensor for the Visually Impaired. In *Ubiquitous Intelligence and Computing*; Yu, Z., Liscano, R., Chen, G., Zhang, D., Zhou, X., Eds.; Springer: Berlin/Heidelberg, Germany, 2010; pp. 590–604.
11. José, J.; Farrajota, M.; Rodrigues, J.; Buf, J.M.H.D. The SmartVision local navigation aid for blind and visually impaired persons. *Int. J. Digit. Content Technol. Appl.* **2011**, *5*, 362–375.
12. Costa, P.; Fernandes, H.; Martins, P.; Barroso, J.; Hadjileontiadis, L. Obstacle Detection using Stereo Imaging to Assist the Navigation of Visually Impaired People. *Procedia Comput. Sci.* **2012**, *14*, 83–93. [CrossRef]
13. Tapu, R.; Mocanu, B.; Bursuc, A.; Zaharia, T. A Smartphone-Based Obstacle Detection and Classification System for Assisting Visually Impaired People. In Proceedings of the 2013 IEEE International Conference on Computer Vision Workshops, Sydney, Australia, 2–8 December 2013; pp. 444–451. [CrossRef]
14. Caldini, A.; Fanfani, M.; Colombo, C. Smartphone-Based Obstacle Detection for the Visually Impaired. In *Image Analysis and Processing—ICIAP 2015*; Murino, V., Puppo, E., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 480–488.
15. Mocanu, B.; Tapu, R.; Zaharia, T. An outdoor cognition system integrated on a regular smartphone device. In Proceedings of the 2015 E-Health and Bioengineering Conference (EHB), Iasi, Romania, 19–21 November 2015; pp. 1–4. [CrossRef]
16. Mocanu, B.; Tapu, R.; Zaharia, T. Seeing Without Sight—An Automatic Cognition System Dedicated to Blind and Visually Impaired People. In Proceedings of the IEEE International Conference on Computer Vision Workshops, Venice, Italy, 22–29 October 2017; pp. 1452–1459. [CrossRef]
17. Rodríguez, A.; Yebes, J.J.; Alcantarilla, P.; Bergasa, L.M.; Almazán, J.; Cela, A. Assisting the Visually Impaired: Obstacle Detection and Warning System by Acoustic Feedback. *Sensors* **2012**, *12*, 17476–17496. [CrossRef] [PubMed]
18. Valencia, G.; Diego, J. A Computer-Vision Based Sensory Substitution Device for the Visually Impaired (See ColOr). Ph.D. Thesis, University of Geneva, Geneva, Switzerland, 2014. Available online: https://nbn-resolving.org/urn:nbn:ch:unige-345689 (accessed on 4 July 2021)
19. Mocanu, B.; Tapu, R.; Zaharia, T. An Obstacle Categorization System for Visually Impaired People. In Proceedings of the 2015 11th International Conference on Signal-Image Technology Internet-Based Systems (SITIS), Bangkok, Thailand, 23–27 May 2015; pp. 147–154. [CrossRef]
20. Mocanu, B.; Tapu, R.; Zaharia, T. When Ultrasonic Sensors and Computer Vision Join Forces for Efficient Obstacle Detection and Recognition. *Sensors* **2016**, *16*, 1807. [CrossRef] [PubMed]
21. Lin, B.S.; Lee, C.C.; Chiang, P.Y. Simple Smartphone-Based Guiding System for Visually Impaired People. *Sensors* **2017**, *17*, 1371. [CrossRef] [PubMed]
22. The Sound of Vision Project. Available online: https://soundofvision.net (accessed on 4 July 2021).
23. Caraiman, S.; Morar, A.; Owczarek, M.; Burlacu, A.; Rzeszotarski, D.; Botezatu, N.; Herghelegiu, P.; Moldoveanu, F.; Strumillo, P.; Moldoveanu, A. Computer Vision for the Visually Impaired: The Sound of Vision System. In Proceedings of the 2017 IEEE International Conference on Computer Vision Workshops (ICCVW), Venice, Italy, 22–29 October 2017; pp. 1480–1489. [CrossRef]
24. Caraiman, S.; Zvoristeanu, O.; Burlacu, A.; Herghelegiu, P. Stereo Vision Based Sensory Substitution for the Visually Impaired. *Sensors* **2019**, *19*, 2771. [CrossRef] [PubMed]

25. Csapo, A.; Spagnol, S.; Herrera Martinez, M.; Bujacz, M.; Janeczek, M.; Ivanica, G.; Wersényi, G.; Moldoveanu, A.; Unnthorsson, R. Usability and effectiveness of auditory sensory substitution models for the visually impaired. In Proceedings of the 142nd Audio Engineering Society International Convention 2017, Berlin, Germany, 20–23 May 2017; Paper 9801. Available online: http://www.aes.org/e-lib/browse.cfm?elib=18675 (accessed on 4 July 2021).
26. Dascalu, M.; Moldoveanu, A.; Balan, O.; Lupu, R.G.; Ungureanu, F.; Caraiman, S. Usability assessment of assistive technology for blind and visually impaired. In Proceedings of the 2017 E-Health and Bioengineering Conference (EHB), Sinaia, Romania, 22–24 June 2017; pp. 523–526. [CrossRef]
27. Spagnol, S.; Hoffmann, R.; Herrera Martínez, M.; Unnthorsson, R. Blind wayfinding with physically-based liquid sounds. *Int. J. Hum. Comput. Stud.* **2018**, *115*, 9–19. [CrossRef]
28. Moldoveanu, A.; Mitruț, O.; Dascalu, M.; Stanica, I.C.; Bodea, C.N.; UnnÓRsson, R.; Moldoveanu, F. Sound of vision 3D virtual training environments-a gamification approach for visual to audio-haptic sensory substitution. *Rev. Roum. Sci. Tech. Ser. Electrotech. Energetique* **2018**, *63*, 112–117.
29. Hoffmann, R.; Spagnol, S.; Kristjánsson, R.; Unnthorsson, R. Evaluation of an Audio-haptic Sensory Substitution Device for Enhancing Spatial Awareness for the Visually Impaired. *Optom. Vis. Sci. Off. Publ. Am. Acad. Optom.* **2018**, *95*, 757. [CrossRef] [PubMed]
30. Lupu, R.G.; Mitruț, O.; Stan, A.; Ungureanu, F.; Kalimeri, K.; Moldoveanu, A. Cognitive and Affective Assessment of Navigation and Mobility Tasks for the Visually Impaired via Electroencephalography and Behavioral Signals. *Sensors* **2020**, *20*, 5821. [CrossRef] [PubMed]
31. Botezatu, N.; Caraiman, S.; Rzeszotarski, D.; Strumillo, P. Development of a versatile assistive system for the visually impaired based on sensor fusion. In Proceedings of the 2017 21st International Conference on System Theory, Control and Computing (ICSTCC), Sinaia, Romania, 19–21 October 2017; pp. 540–547. [CrossRef]

sensors

MDPI

Article

Cognitive and Affective Assessment of Navigation and Mobility Tasks for the Visually Impaired via Electroencephalography and Behavioral Signals

Robert-Gabriel Lupu [1], Oana Mitruț [2,*], Andrei Stan [1], Florina Ungureanu [1], Kyriaki Kalimeri [3] and Alin Moldoveanu [2]

1 Faculty of Automatic Control and Computer Engineering, Gheorghe Asachi Technical University of Iași, 700050 Iași, Romania; robert-gabriel.lupu@academic.tuiasi.ro (R.-G.L.); andrei.stan@academic.tuiasi.ro (A.S.); fungurea@tuiasi.ro (F.U.)

2 Department of Computers, Faculty of Automatic Control and Computers, University Politehnica of Bucharest, 060042 Bucharest, Romania; alin.moldoveanu@cs.pub.ro

3 Institute of Scientific Interchange, Via Chisola 5, 10126 Torino, Italy; kyriaki.kalimeri@isi.it

* Correspondence: oana.balan@cs.pub.ro

Received: 15 September 2020; Accepted: 13 October 2020; Published: 15 October 2020

check for updates

Abstract: This paper presented the assessment of cognitive load (as an effective real-time index of task difficulty) and the level of brain activation during an experiment in which eight visually impaired subjects performed two types of tasks while using the white cane and the Sound of Vision assistive device with three types of sensory input—audio, haptic, and multimodal (audio and haptic simultaneously). The first task was to identify object properties and the second to navigate and avoid obstacles in both the virtual environment and real-world settings. The results showed that the haptic stimuli were less intuitive than the audio ones and that the navigation with the Sound of Vision device increased cognitive load and working memory. Visual cortex asymmetry was lower in the case of multimodal stimulation than in the case of separate stimulation (audio or haptic). There was no correlation between visual cortical activity and the number of collisions during navigation, regardless of the type of navigation or sensory input. The visual cortex was activated when using the device, but only for the late-blind users. For all the subjects, the navigation with the Sound of Vision device induced a low negative valence, in contrast with the white cane navigation.

Keywords: sensory substitution; cognitive load; brain activity; navigation; multimodal; audio; haptic

1. Introduction

At the world level, approximately 2.2 billion people have a vision impairment or suffer from blindness, caused primarily by uncorrected refractive errors, cataracts, age-related macular degeneration, and glaucoma. The majority of people with vision impairments are over 50 years old, originating especially from low and middle-income countries [1]. The purpose of the Sound of Vision project (SoV) [2] was to develop an assistive system for the blind and visually impaired users that would facilitate navigation and obstacle detection. In this paper, we presented a study of cognitive load assessment and brain activation evaluation during an experiment in which eight visually impaired subjects performed various object detection and navigation activities while using both the white cane (a navigation aid they use on a daily basis) and the SoV device, which provided three types of sensory input—*audio cues* delivered through headphones, *haptic cues* delivered as vibrations applied on a vest that was placed on the user's abdomen, and a combination of both audio and haptic information, called the *multimodal sensory input*. We performed a metrics analysis, cognitive load, working memory assessment, brain activity, visual cortex evaluation, and the identification of emotions during navigation in the real-world environment.

Section 2 presents an overview of mobility assistive devices, Section 3 introduces the biophysical signals and cognitive load, Section 4 describes the Sound of Vision device, Section 5 details the method, Section 6 presents the results and a discussion, and finally, Section 7 provides the final conclusions and future research directions.

2. Overview of Mobility Assistive Aids

The absence of visual information in the case of blind individuals can be substituted by conveying auditory and tactile stimuli, separately or simultaneously, through specialized assistive devices.

2.1. Auditory Vision Sensory Substitution

Auditory vision sensory substitution (AVSS) devices [3] map the image "seen" by the camera into a matrix of active audio sources. The diversity of AVSSs is very large, ranging from optophone-like systems [4] to devices that use spatialized 3D sounds. The *optophone* (or the piano transform device) scans the image from left to right and converts the detected input into sound cues. The most well-known optophone is the vOICe [5], where the pixels' vertical position is mapped to frequency, and their brightness is mapped to sound intensity. In other AVSSs [6,7], height is correlated to frequency distribution in the horizontal plane to binaural loudness, and brightness is encoded as sound intensity. In *pitch transform systems* [7,8], distance is related to sound frequency, while in *verbal transform systems* [7,9], objects are assigned to synthetic voice recordings. A problem of most optophone-like modern systems is that they overwhelm the users with too much output information, increasing cognitive load, effort, and concentration. This drawback can be overcome by reducing scene complexity, by maintaining only the salient characteristics and relevant objects, and by applying an effective sonification technique in order to provide the users a natural, effective, and easy to understand environmental representation. Modern AVSSs use binaural 3D sounds spatialized with generic (prerecorded, stored in large databases) or individualized head-related transfer functions (HRTFs). Individualized HRTFs are preferable for creating 3D sound as they are more accurate and fit the user's auditory characteristics.

2.2. Tactile Visual Sensory Substitution

Tactile visual sensory substitution (TVSS) systems use a matrix of controllable elements that provide spatial and temporal environmental information on the skin, either through kinesthetic or cutaneous sensations. In this type of device, a camera is used to acquire visual input that is consequently transformed into a tactile rendering via the multi-dimensional pin array, facilitating reading, shape, and face recognition [10–12]. One advantage of using TVSS devices is that the tactile sense, contrary to the auditory one, is less used and demanded in everyday activities. Thus, the user can receive cutaneous (awareness and stimulation of the outer surface of the body [13]) and kinesthetic (awareness of the limb position and displacement [10]) cues, without hampering locomotion or auditory perception at all [14]. On the other hand, a drawback lies in the fact that the capacity of the tactile channel is restrained to a limited maximum number of actuators and patterns to be applied. In addition, such devices cannot be used to a large extent because they are tiring and uncomfortable.

2.3. Auditory Tactile Visual Substitution Devices

When the scene is too difficult to be mapped onto the tactile array, the auditory channel is additionally recruited in order to enhance environmental representation, creating ***auditory tactile visual substitution*** (ATVS) devices. The first multi-sensory device was Nomad [15]. Tactile cues are delivered through a touch-sensitive tablet, and the auditory information consists of synthesized voice recordings. The Heard and Felt Vision Effects (HiFiVE) [16,17] system uses moving speech-like sounds (tracers—area tracers and shape tracers), binaural panning, and tactile effects in order to map visual images to an audio-tactile representation.

3. Biophysical Signals and Cognitive Load

3.1. Electroencephalography

Electroencephalography (EEG) can provide neurophysiological markers of cognitive-emotional processes induced by stress and indicated by changes in brain rhythmic activity [18,19]. EEG signal processing techniques play a significant role in quantifying cognitive load [20–24]. Bos et al. [20] showed that cognitive load was an indicator of the learning progress. Berka et al. [22] extracted features from EEG signals for monitoring cognitive workload and task engagement. Nilsson et al. [23] showed learning outcomes from the subjects when they navigated a hypermedia environment. Scott et al. [24] also showed that a navigational map could create significantly more germane or extraneous cognitive load. Therefore, EEG/ Electrodermal Activity (EDA) signals are used to measure cognitive load and affective responses, and the overall process is explained in the following section.

Cognitive load and affective responses may impact the learning progress. The detection of reliable cognitive load and affective responses would improve the design of emotional intelligent mobility systems for the visually impaired people (VIPs). The complexity of the tasks is quantified in terms of cognitive load index and affective index, considering two well-established metrics in the scientific literature—the *event-related desynchronization (ERD)/event-related synchronization (ERS) index* and the *left-right asymmetry index*.

Affective responses directly influence the processes of cognitive learning. However, the challenges of learning can evoke negative affective responses [25]. Emotion assessment is a challenging and demanding task because people are not always able to express their emotions verbally [26]. Bos [20] showed that cognitive load could indicate changes in the learning process. He proposed an approach to determine the optimal placement of a limited number of electrodes, and then these electrodes were placed in an experiment aimed at determining arousal and valence. Left frontal inactivation is an indicator of a withdrawal response, which is often linked to negative emotion. On the other hand, right frontal inactivation is a sign of an approach response or positive emotion. High alpha activity (8–12 Hz in the EEG frequency band) is known to be an indicator of low brain activity.

Researchers have addressed the intertwining role of affective responses, learning, and cognitive load. Bower et al. [27] introduced the following hypothesis to study learning patterns: (1) a positive emotion usually increases the learning process through attention and motivation, (2) a positive emotion improves learning by enhancing cognitive load, and (3) a negative emotion decreases the learning process. Cattaneo et al. [28] employed the cognitive load theory for the understanding of the perceptual and neurocognitive mechanisms; however, there are still many open questions on how emotion and cognitive load can ease the learning process of the visually impaired people.

3.2. Electrodermal Activity and Heart Rate

Electrodermal activity (EDA) is a well-known indicator of physiological arousal and stress activation in affective computing [29,30]. It is more sensitive to emotion-related arousal variations as opposed to physical stressors, which can be better reflected by heart rate (HR) measurements. Blood volume pulse (BVP) patterns can also reflect transient arousal and cognition processes [31]. Two outdoor mobility studies from the early 1970s have suggested that some form of psychological rather than physical stress is responsible for visually impaired people’s increased HR versus sighted pedestrians [32,33]. However, certain mobility tasks (for example, stairs climbing) may result in an interactive psychological stress effect and momentary physical workload; thus, cardiovascular measures may be less suitable than EDA.

4. The Sound of Vision Device

Spatial navigation is a category of spatial cognition related to performing tasks, such as following paths, detecting obstacles, and reaching targets. It is based on developing, maintaining, and recalling an internal representation of the environment [34,35]. This internal representation depends on the spatial relation between entities and on the subject's position, being classified into two categories: egocentric—the navigator is in the center of the coordinate system, and allocentric—the reference external to the navigator.

4.1. Technical Description

Sound of Vision is a wearable device that allows a visually impaired user to perceive and navigate the environment. It works by permanently scanning the environment, extracting essential features, and rendering them to the user through audio and haptic means.

The Sound of Vision final prototype includes an integrated custom hardware solution and a complex software solution, supporting the real-time operation of the device, as well as training tools and materials.

The hardware components of the system are:

- a headgear, including a 3D acquisition unit (depth camera for indoor or low light outdoor conditions, stereo camera for outdoor or bright light conditions, head and body inertial measurement unit (IMU) for body orientation) and an audio rendering unit (mounted on the head);
- a haptic belt with a matrix of 60 vibrating motors (six rows and 10 columns, placed on the abdomen);
- a processing unit: a small laptop with powerful CPU and GPU units (in a backpack);
- a wireless remote control (in the pocket).

When scanning the environment, the user can select from two different models for both audio and haptic: the *discrete model* (which renders the objects sequentially) and the *continuous model* (which provides real-time information at once about all the objects in the field of view). They are divided into sub-models and have different variations, as well as additional features for safe and reliable navigation: *Danger mode*—alerts about proximity objects and prevents collisions, *Flashlight*—enables the rendering of an object's distance in front of the user, *texts and special signs detection* and *best free space*—which indicates a navigable opening between surrounding objects.

The user movement is guided solely by the Sensory Substitution Device (SSD) with no additional feedback from the assistant or from other sources (i.e., maps from Google or GPS coordinates from a GPS device). The SSD device generates audio and haptic signals that are an encoded representation of the environment in the proximity of the user. Through intensive training, the user gains fluency in understanding the audio and haptic encoded feedback issued by the SSD device, and then he/she can make proper decisions for further movement in the environment.

Like any other person, the VIP wants to walk in the direction of the sound source. The SoV device scans the environment, detecting the obstacles and their features, and sends audio or/and haptic stimuli to users. These stimuli help the VIP to avoid obstacles and to find a secure path to the sound source. Obviously, the real scene and the stimulation are changed/updated according to the VIP's route, like in a maze. Depending on their perception, the VIPs may choose different paths.

4.2. The Focus of the Study

In line with the SoV project overall goals, the aim of this study was to evaluate the VIPs' cognitive load and emotional stress in real-world mobility experiments, in two cases: navigation with white cane and navigation relying on the SoV prototype with audio, haptic, or audio and haptic (multimodal) codification. The research questions we pursued were based on the following comparative assessments of cognitive load:

- when using the Sound of Vision device with audio vs. haptic vs. multimodal input;
- when using the Sound of Vision device vs. white cane during a navigation task in the real-world environment.

Based on some achievements presented in scientific papers and on the valuable previous experience [36–39], the experiments were oriented to collect EEG and physiological (EDA and HR) signals in five different mobility tasks in order to highlight the VIPs' cognitive load and stress in correlation with some events (collisions or total confusion) captured from the recorded videos.

As presented above, it has been proven by many studies that EEG is a promising and common approach to measuring cognitive load (CL), working memory load, emotional states, and any cerebral signals denoting cortex responses to specific stimuli. CL was an effective real-time index of task complexity backed up by behavioral evidence. Complementary, the peripherical physiological measurements reflected arousal and stress activation (EDA) and transient processes in arousal and cognition (HR). The mobility tasks did not include a consistent physical effort, so HR could be also considered in a multimodal approach.

An important aspect of this work was related to the VIPs' preference and long-term accommodation to navigating using the white cane, as they were educated to use it from the moment they lost sight or from childhood. Although the SoV device offered much more information about the environment (the number of nearby objects, their position and properties, the presence of specific objects, and so on), it was expected that for the first tests, the VIPs would have lower cognitive load and better performances during the benchmark task with the white cane than with audio or haptic stimuli. Obviously, the length of the accommodation period with the SoV device depended on each VIP's education and ability to learn. It was important for this study to understand how easily the SoV stimuli were perceived and processed by a VIP and which navigation modality (audio, haptic, or audio and haptic) was less stressful and more quickly accepted. We expected that audio mobility would outperform the haptic and fusion mobilities, knowing that blind people generally have a well-developed hearing sense.

Additionally, the VIPs' visual cortex excitation by audio and haptic stimuli during navigation was investigated, expanding the existing literature [40,41] that has reported brain activity in the visual cortex during EEG measurements for blind people who have received visual information through sensory substitution devices (SSDs).

To our best knowledge so far, this paper was the first one to present a comparative study regarding VIPs' stress, cognitive load, and visual cortex excitation while navigating in the real-world using a common white cane vs. a sensory substitution device.

5. Materials and Methods

This experiment has been carried out using the Acticap EEG device with 16 electrodes, provided by Brain Products GmbH from Germany and Shimmers Multisensory provided by the Shimmer Sensing company from Dublin, Ireland.

5.1. Experimental Setup

The aim of the experiment was to obtain a dataset, as large as possible, with EEG and physiological signals during the trials designed for traveling in fixed scenarios with the help of the SoV device [42]. There were two user setups. The first one was the **virtual training environment (VTE)** setup that was used to train the subjects and accommodate them with the audio and haptic encoding models prior to using the system in real-life scenarios [43,44]. The second setup (Figure 1) was the **real-world (RW)** setup used in an indoor controlled environment and outdoors.

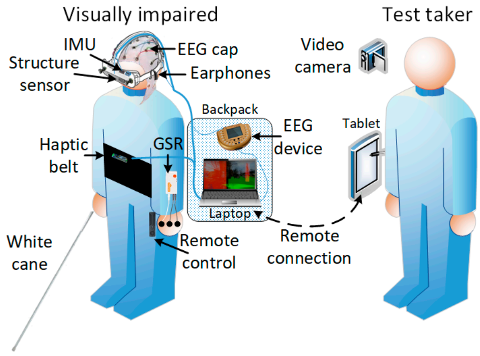


Figure 1. Real-world setup. IMU- Inertial Motion Unit, EEG–Electroencephalography, GSR–Galvanic Skin Response.

The difference between the VTE and RW setups consisted mainly of the video streaming sources that feed the processing and control unit. For VTE, the video stream was provided by a virtual reality serious game in which the VIP navigated using the keyboard or a joystick. In the VTE tests, the VIP wore the headset consisting of a structure sensor stereo video camera and an IMU sensor [45,46], but only the IMU signal was used in order to orient in the virtual environment by head movements. For the RW setup, the video stream was provided by the structure sensor (indoor use of the system) or stereo video camera (outdoor use of the system).

The SoV system used in the RW experiments is depicted in Figure 2 and consisted of two computing systems: the first one was the processing and control unit (PCU) attached to the VIP, and the second one (the tablet from Figure 2) was used by the assistant who controlled how the trial was performed. Via a remote connection with the PCU, the assistant could adjust parameters, select scenarios, and enable physiological signals recording. The PCU ran the SoV runtime application, which sensed the environment and provided audio and haptic stimuli. It also recorded physiological data from the user who performed the navigation tasks. The IMU signals were used to determine the user's body and head orientation that was further used to render the audio and haptic output in accordance with the RW scene. The VIP had the opportunity to select the most appropriate audio or haptic encoding by using a remote control connected to the PCU. During navigation, there was no communication between the VIP and the assistant. The VIP walked autonomously based only on the stimulation provided by the SSD. The test taker's task during tests was to ensure that the system was working and that the VIP received correct clues. He did this by using a tablet connected wirelessly to the VIP's laptop.

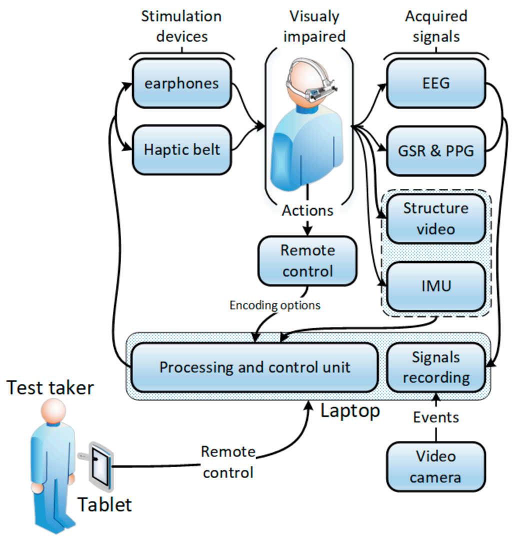
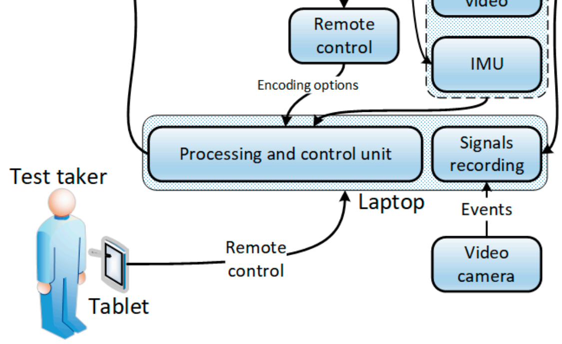


Figure 2. The system architecture for the real-world (RW) setup. PPG-plethysmography.



We used a lightweig[…] r to cool down the laptop, which are part of the SoV[…] erimental system (laptop, battery, fans, and EEG dev[…] lid not feel uncomfortable due to the weight and/or heat. Actually, a lighter version of the SoV system—more energy-efficient and low cost—is under development.

Figure 2. The system architecture for the real-world (RW) setup. PPG-plethysmography.

A print screen from the serious game called **"treasure hunt" (TH)** is presented in Figure 3. The VIP had to use the joystick in order to position himself in the virtual scene, exactly where the sound source originates. In RW, the user had to navigate through an indoor environment in order to reach a target sound source while avoiding cardboard box obstacles of various dimensions (Figure 4).

A print screen from the serious game called **"treasure hunt" (TH)** is presented in Figure 3. The VIP had to use the joystick in order to position himself in the virtual scene, exactly where the sound source originates. In RW, the user had to navigate through an indoor environment in order to reach a target sound source while avoiding cardboard box obstacles of various dimensions (Figure 4).

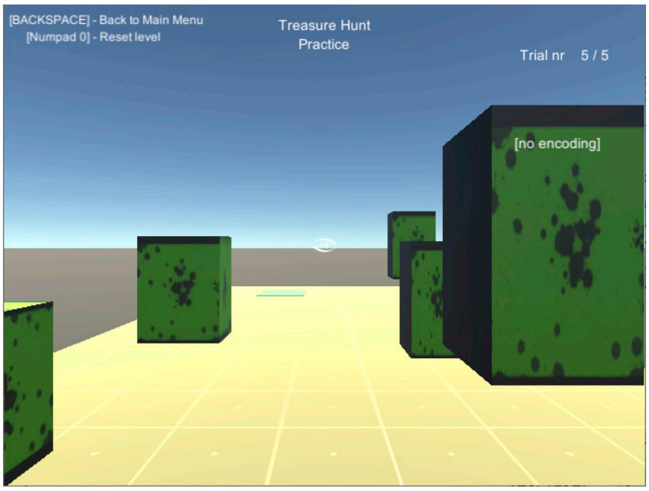


Figure 3. Treasure hunt in a virtual training environment (VTE).

Figure 4. Treasure hunt in RW.

We conducted EEG and EDA/HR recordings during experimentation in RW, under two conditions: **white cane only** (for the participants who used this mobility aid on a regular basis) and **SoV device only**. The scenes were tested using the *audio encoding, the haptic encoding, and both the audio and haptic encodings (multimodal)* with the SoV device, 5 trials each. In order to minimize the required testing resources, the users had the opportunity to choose the sonification model and tactile stimulation [47–49] that best suited their level of perception and understanding.

The **static scenes (1R)** were tested with the *discrete model*, while the **dynamic scenes (TH)** were tested with the continuous *model* in both virtual and real-world environments.

In the *discrete (or iterative) model*, the scene was rendered in a loop, one by one. A sphere was constantly expanding its radius until 5.25 m with a speed of 2 m/sec. Auditory and tactile stimuli were provided when this sphere intersected scene objects, allowing distance detection, as well as comparing the distance between objects. The continuous model rendered an entire scene at once, providing instantaneous information via audio and haptic.

VTE tests were recorded automatically by the SoV system, while RW tests were recorded manually by testing assistants. Furthermore, each trial in a test was videotaped for annotation purposes. The time needed to finish every trial and the accuracy were saved: number of collisions between user and obstacles, number of cane contacts with objects, time duration, together with path length followed.

Each test consisted of 5 trials, and each one was assigned to a fixed path/boxes arrangement for TH, as it is presented in Figures 5–9.

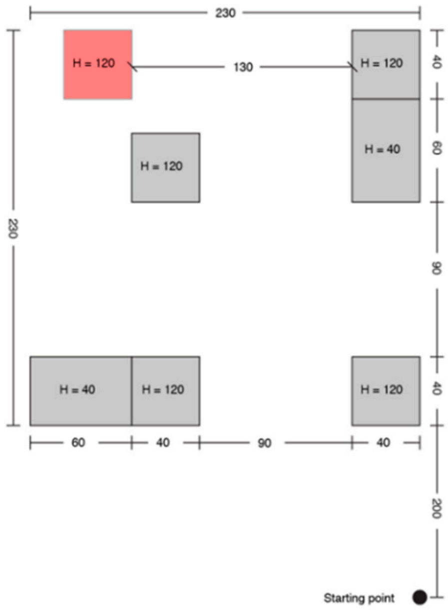


Figure 5. Treasure hunt (TH) configuration A.

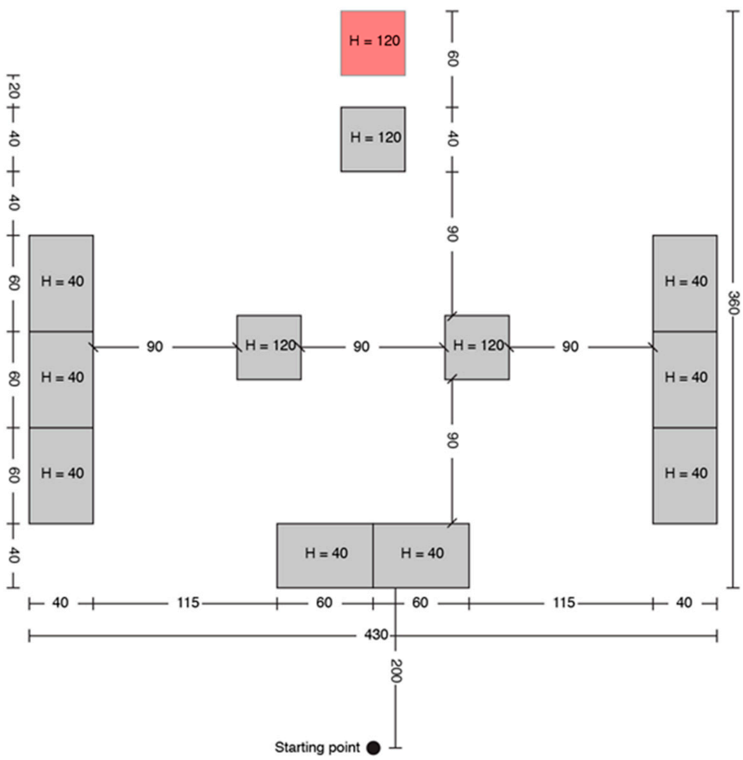


Figure 6. TH configuration B.

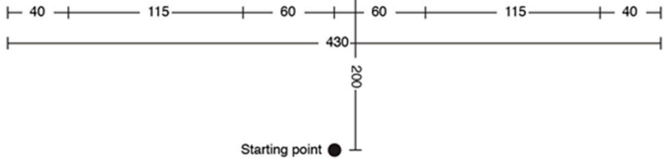
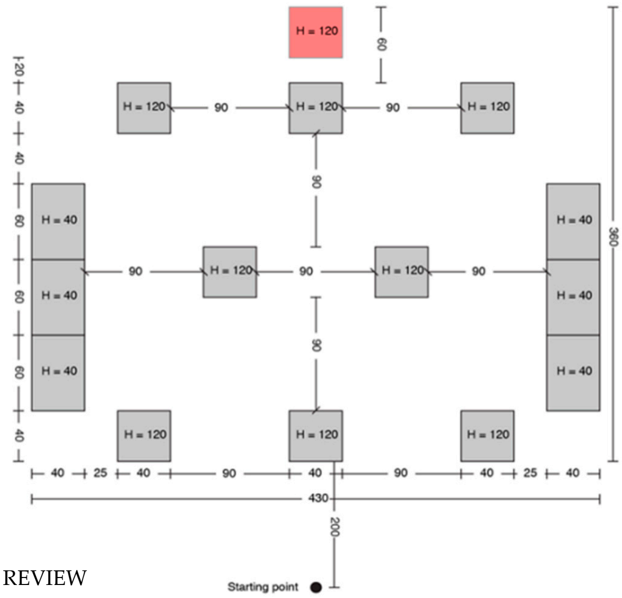


Figure 6. TH configuration B.



Sensors **2020**, *20*, x FOR PEER REVIEW 10 of 31

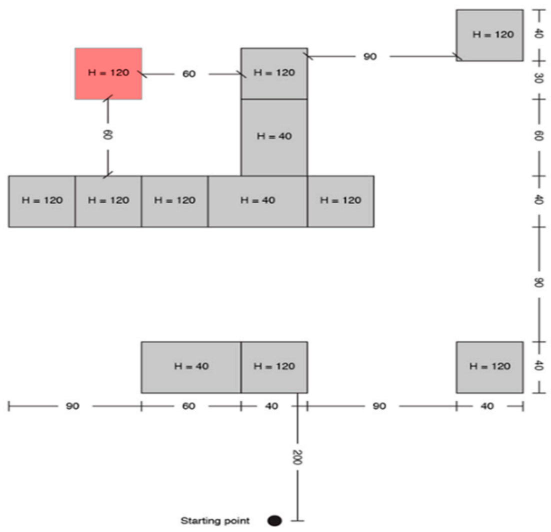


Figure 8. Treasure hunt configuration D.

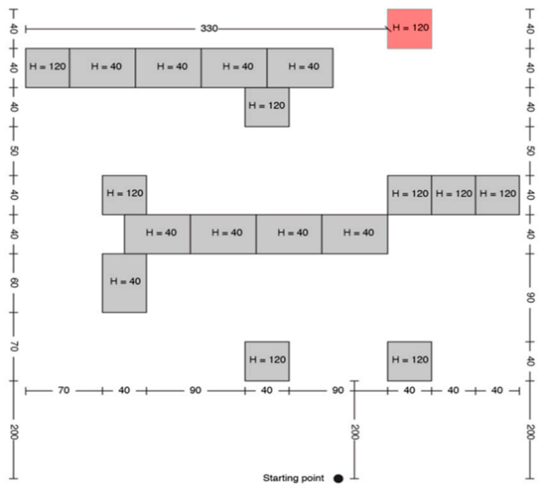


Figure 9. Treasure hunt configuration E.

Figure 9. Treasure hunt configuration E.

5.2. Data Collection

5.2. Data Collection

In this study, 15 VIPs were involved in training and testing tasks in the virtual environment and real-world settings, using the SoV prototype and different releases of the SoV runtime. After each testing stage, important improvements were made to the hardware and software resources, including audio and haptic encodings based on the VIPs' feedback. Finally, only 8 complete datasets

5.2. Data Collection

In this study, 15 VIPs were involved in training and testing tasks in the virtual environment and real-world settings, using the SoV prototype and different releases of the SoV runtime. After each testing stage, important improvements were made to the hardware and software resources, including audio and haptic encodings based on the VIPs' feedback. Finally, only 8 complete datasets (corresponding to 3 females and 5 males, aged 20–42) with fully validated data were retained and subsequently analyzed. All the participants provided informed consent approved by the research ethics committee of the institutions involved in the project (Approval number: 9083/15.05.2017). One hour before performing the tests, the VIPs did not drink coffee nor black tea, and also smoking was forbidden before or during the experiments.

For EEG and additional physiological measurements, the following equipment was used:

- a BrainProducts V-Amp 16 amplifier and an EasyCap helmet with 19 sintered Ag/AgCl miniaturized passive electrodes for EEG signal acquisition with a sampling rate of 512 Hz;
- a Shimmer3 GSR+ unit sensor for measuring electrodermal activity/galvanic skin response (EDA/GSR) and continuous HR;
- a video camera or smartphone for video recording in real-time.

The acquisition procedure used 16 electrodes, namely Fp1, F7, F3, C3, P3, P7, C4, O1, O2, Fp2, F8, F4, C4, P4, T8, P8, and an ear reference, placed according to the 10–20 international system. The sampling rate was 512 Hz, and the AFz electrode was connected to the ground. To ensure reliable EEG raw data, the impedance of each electrode was maintained below 5 kΩ, by using a good abrasive gel. The OpenVibe open source software was used for EEG acquisition. The OpenVibe server acquired the EEG signals, and the OpenVibe client saved or sent the raw data as a stream. The Shimmer GSR unit sent the acquired data via Bluetooth.

The data acquisition process is outlined in Figure 10. An important part of the acquisition process is the usage of the lab streaming layer (LSL) protocol [50] so that each data component had to provide a stream of data as output. For the components that do not natively provide LSL output streams, simple adaptors had to be designed, as in the case of the data provided by the EDA/GSR device. The EEG data was available as an LSL stream provided by the OpenVibe application. The applications developed in this project (VTE and SoV runtime) provided LSL streams to source events that were internally generated and of interest for later analysis. The data was stored in a tabular format inside hdf5 files [51], together with a timestamp provided by the LSL.

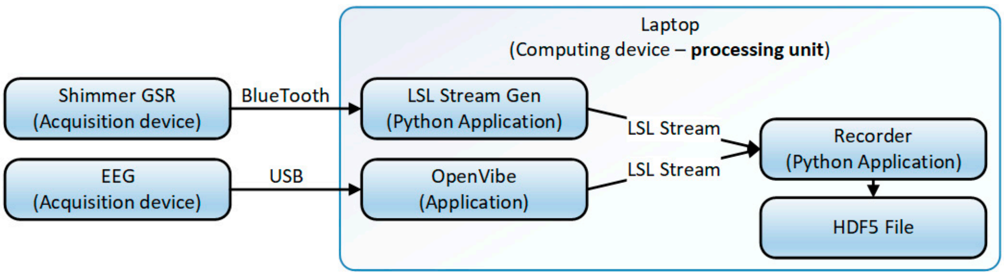


Figure 10. The data acquisition process. LSL, lab streaming layer.

In what concerns video recording and annotations, the mp4 video files recorded during the tests were annotated with Chronoviz [52]. The events: *Start* (beginning of the recording), *Collision*, *Find* (find the source/treasure), *Lost* (lost control), *TouchCane* (only for the test with the white cane), *Stop* (end of the recording) were considered. For each mp4 file, a CSV file with annotated events and corresponding Unix timestamps was generated.

Alternatively, if Chronoviz could not be used due to system constraints (Chronoviz needs a system with Mac OS X 10.6 or later), a Python script was designed to synchronize the data streams acquired by the processing unit with the video recordings of the experiment. The script aligned the timestamps of the samples in the acquired data streams (provided as csv files) with the timing information found in the video recording. The application ExifTool [53] was used for gathering timing information from the movie files. The data streams were trimmed or padded in order to fit the movie length. The script detected and reported any timing misalignments and provided means to the processing unit with the video recordings of the experiment. The script aligned the timestamps of the samples in the acquired data streams (provided as csv files) with the timing information found in the video recording. The application ExifTool [53] was used for gathering timing information from the movie files. The data streams were trimmed or padded in order to fit the movie length. The script detected and reported any timing misalignments and provided means to fine-tune the synchronization process. The resulting adjusted data streams and movies could be annotated later in a similar way as in Chronoviz. The synchronization process is presented in Figure 11.

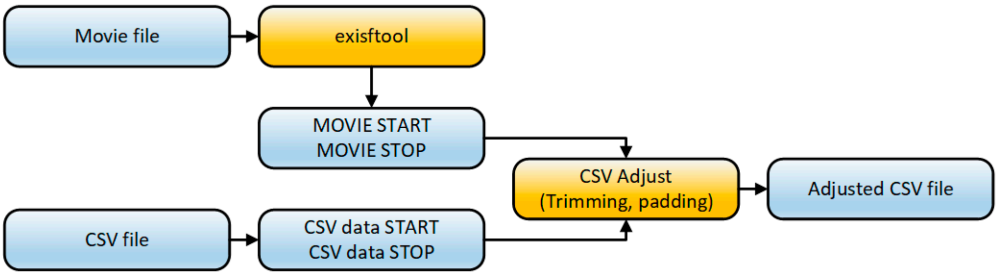


Figure 11. RW events synchronization process.

5.3. Data Acquisition and Preprocessing

The acquired brain waves were pre-processed. We applied a band-pass filter for 0.5–100 Hz, a notch filter to remove power line contamination at 50 Hz, and a band-pass filter to obtain frequency bands of interest (delta, theta, alpha, beta, and gamma). The artifacts (involuntary eye blinks, muscle movements, brief amplifier saturations) presented in the EEG signals were removed using an online Savitzky-Golay filter. The EEG data obtained after pre-processing were baseline-normalized by subtracting for each participant and for each channel the mean of the resting state recordings (recorded in the laboratory during the VTE sessions).

The Shimmer software for EDA acquisition could not be efficiently used in the SoV setup (Figure 1), and therefore the streams from the sensor were acquired over Bluetooth at 16 Hz. The skin resistance values (y, µS) were computed from the Shimmer ADC values with the following linear function:

$$y = p1 \times x + p2 \tag{1}$$

where p1 and p2 are parameters specific to the range setting and can be selected from the datasheet of the sensor. If the electrodes are not tightly attached and lose contact with the skin, motion artifacts (high-frequency noise) can be present in the acquired signals. A low pass filter was applied to remove high-frequency noise, which can be attributed to movement artifact and other noise components. A cutoff frequency of as low as 1–5 Hz could be used without affecting the data of interest due to the slowly varying nature of the EDA responses.

5.4. Data Analysis

Within the broader framework of the SoV project, the aim of this study was to explore the VIPs' brain activity during navigation tasks with the help of an SSD based on audio, haptic, and multimodal encoding, compared to white cane navigation. The research was focused on assessing cognitive load, visual cortex excitation, and emotions evaluation during RW navigation. For each exploration, the EEG signals were selected according to the analyzed brain lobes and the *power spectrum, and the asymmetry between the two cortex hemispheres was calculated.*

Usually, CL is investigated in the channels corresponding to the frontal lobe, which reflect the activity of short-term memory and consists of calculating **frontal-asymmetry**, meaning the difference between the logarithms of the power spectrum of the left and right hemispheres divided by the logarithm of the total power spectrum of both hemispheres. There is no single standard way to calculate asymmetry, and some authors use the difference or the ratio between the spectral powers of the signals on the right and the correspondents in the left hemisphere. Anyhow, *higher asymmetry reflects a strong workload, while lower asymmetry reflects avoidance and relaxation* [54].

CL is strongly related to emotional well-being states. The "feeling good" aspect of well-being deals with the balance of positive emotions vs. negative emotions. Well-being reflects a person's ability to identify and respond to the challenges of everyday life, even painful and unpleasant events [55]. Hawthorne presented an extensive study on how feeling good might contribute to cognitive load in different ways [55].

Certain states can be more accurately investigated if the EEG waves are analyzed in the five specific bands: *delta, theta, alpha, beta,* and *gamma*. The *delta* waves reflect the unconscious states, and it is usually recorded in deep dreamless sleep. The *theta* waves are typically associated with the subconscious mind, sleeping, dreaming, meditation, or even artistic creation. The *alpha* waves are visible in all the cortex lobes and give valuable information regarding brain activation and the relaxed (but yet aware) mental state. High alpha activity has been correlated to brain inactivation. The *beta* waves are correlated to high mental activity, more prominent in the frontal cortex but visible over other lobes as well. The alpha and beta waves are the most used to classify workload using EEG. The *gamma* waves (>30 Hz) reflect hyper brain activity and have become more and more studied as the sampling frequency of the acquisition systems has increased [56].

Regarding **visual cortex (VC) excitation**, it must be specified that it was not known before the year 2000 whether the visual cortex could receive input from other sensory modalities besides the eyes through the lateral geniculate nuclei. Afterward, the EEG measurements have revealed that the VC activity is higher for blind subjects during rest or auditory/tactile tasks than in normal control. Without a certain demonstration, Sadato et al. suggested that in blind subjects, the cortical areas normally reserved for vision might be activated by other sensory modalities [57]. In 2003, Burton reviewed various brain imaging studies, which investigated the visual cortex activity of VIPs during nonvisual tasks, such as hearing messages, Braille reading, or even sensory discriminations of tactile or auditory stimuli, and concluded that the loss of vision did not lead to a permanent inactivation of the visual cortex [58]. A scientific report from Georgetown University Medical Center concluded in 2010 that "people who have been blind from birth make use of the visual parts of their brain to refine their sensation of sound and touch" [59]. In recent years, several studies have highlighted enhanced auditory processing in blind persons to partially compensate their impairment, with greater sensitivity of the other senses. It has been proved that the VC plasticity allows this cortical lobe to be colonized by the auditory and somatosensory systems in the case of congenitally blind persons. The study conducted by Campus et al. revealed that the occipital activation to sound was strong in sighted persons and much lower in blind persons [60]. Another valuable conclusion was that the occipital lobe of sighted subjects played a major role in the reconstruction of the environmental spatial metrics and that vision loss blocked this process. Obviously, it is expected to remark differences in VC excitation between the people who are blind from birth and those who lost their sight later and know what color, distance, or shape mean. For this analysis, the O1 and O2 electrodes are the most important, but also the Oz and the electrodes from parietal lobes should be considered in an extensive study.

In terms of evaluating emotions, it is well known that the **amygdala** is responsible for the perception of emotions, such as anger, fear, and sadness. The pre-frontal cortex and the hippocampus (located in the medial region of the temporal lobe) are highly correlated to emotional activity [56,61]. Because the right hemisphere is associated with negative emotions (i.e., fear or disgust), and the left hemisphere is highly activated by positive emotions and motivation (i.e., happiness and satisfaction), the EEG asymmetries in the frontal and parietal lobes are relevant for valence and arousal assessment [56]. A thorough evaluation can be performed if the signal analysis is performed on the EEG frequency bands of alpha, beta, and gamma. According to these findings and based on some other studies related to efficient EEG channels selection for emotion recognition, Zhang and his coworkers recommended the following set of electrodes: Fp1, Fp2, F7, F8, C3, FC5, FC2, AF4 (frontal lobe), T7, T8 (temporal lobe), O1, Oz (occipital lobe), and P3, P4, Pz, PO4 (parietal lobe) [62]. For emotions assessment in this study, only the channels C3, C4, T7, T8, P3, P4, F3, and F4 were considered due to the limited number of electrodes of the EasyCap helmet. The asymmetry in the pre-frontal lobe was presented in CL evaluation, and O1 was not considered because the standard list refers to sighted people, and, in our approach, the visual cortex was subjected to special attention.

6. Results and Discussion

6.1. Navigation Metrics Analysis

As mentioned above, the EEG, HR, and EDA (GSR) signals were acquired for the treasure hunt tests, using the white cane or the SoV device with three spatial information encodings—audio, haptic, and multimodal (audio and haptic). Besides the video recordings and the files containing the data obtained during the experiments, important metrics regarding navigation were collected for each user involved in the study: the *time required to accomplish a trial, the length of the path, the number of major or minor collisions,* and also *the numbers of white cane contacts with the obstacles*. All these data are summarized in Table 1 and reflect the cumulative performance of all the users for each scenario type.

Table 1. Cumulative experimental data for the treasure hunt (TH) tasks—navigation with the Sound of Vision (SoV) device and white cane.

Codification	Scenario Type	Collisions Total Number	Path Total Distance (m)	Total Time (s)
Audio	A	12	33.69	293
	B	10	44.36	299
	C	22	42.8	261
	D	18	41.8	306
	E	20	62.05	409
Haptic	A	13	28.3	179
	B	11	39.1	221
	C	17	41	261
	D	20	44.1	255
	E	34	54	440
Audio and Haptic	A	7	33.35	236
	B	14	39	267
	C	15	47.3	327
	D	12	51.6	271
	E	24	56.5	345
White cane	A	4	23.27	204
	B	5	28.8	205
	C	6	27.8	213
	D	2	30.16	206
	E	6	37.77	255

Figures 12–14 present the averages of time duration, number of collisions, and traveled distances for RW navigation with the help of the white cane and SoV device, in case of all the five obstacle arrangements (A to E). A+H stands for audio and haptic (multimodal). In the case of the short and easy routes, the walking durations were very similar for audio stimulation and cane traveling, while the haptic and multimodal stimulation required less time than the white cane. Only for the most complicated test scenario (E), the cane and multimodal tasks were performed in a shorter time than with haptic and audio input. It is known that the VIPs usually walk slowly, and it was encouraging that the SoV device did not slow down the movement of the users.

metrics for the white cane navigation were better because the VIPs were accustomed to using it on a daily basis. As a particular conclusion, *the VIPs' navigation performance with the SoV device was better in the case of the multimodal encoding, in terms of duration and number of collisions*. On the other hand, no general conclusion could be drawn because the number of VIPs involved in the experiments was small, and also a VIP could have learned the scenes during the first trials and performed better during the last trial, even if the experiments were randomly conducted.

daily basis. As a particular conclusion, *the VIPs' navigation performance with the SoV device was better in the case of the multimodal encoding, in terms of duration and number of collisions*. On the other hand, no general conclusion could be drawn because the number of VIPs involved in the experiments was small, and also a VIP could have learned the scenes during the first trials and performed better during the last trial, even if the experiments were randomly conducted.

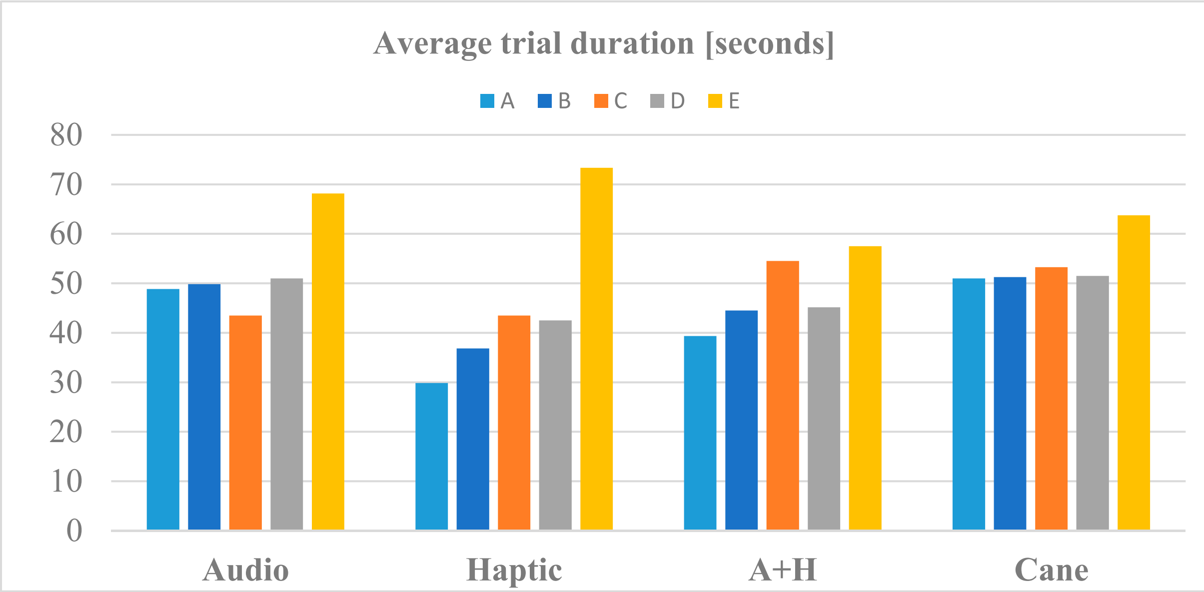


Figure 12. The average time to perform navigation tasks.

Figure 12. The average time to perform navigation tasks.

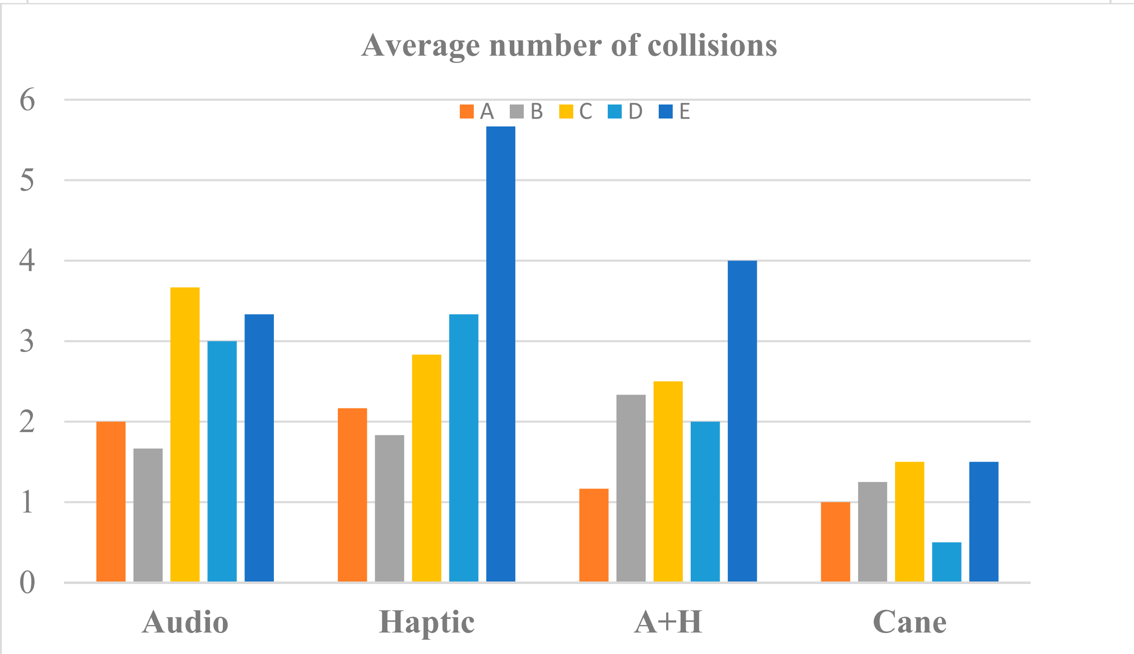


Figure 13. The average number of collisions during navigation tasks.

Figure 13. The average number of collisions during navigation tasks.

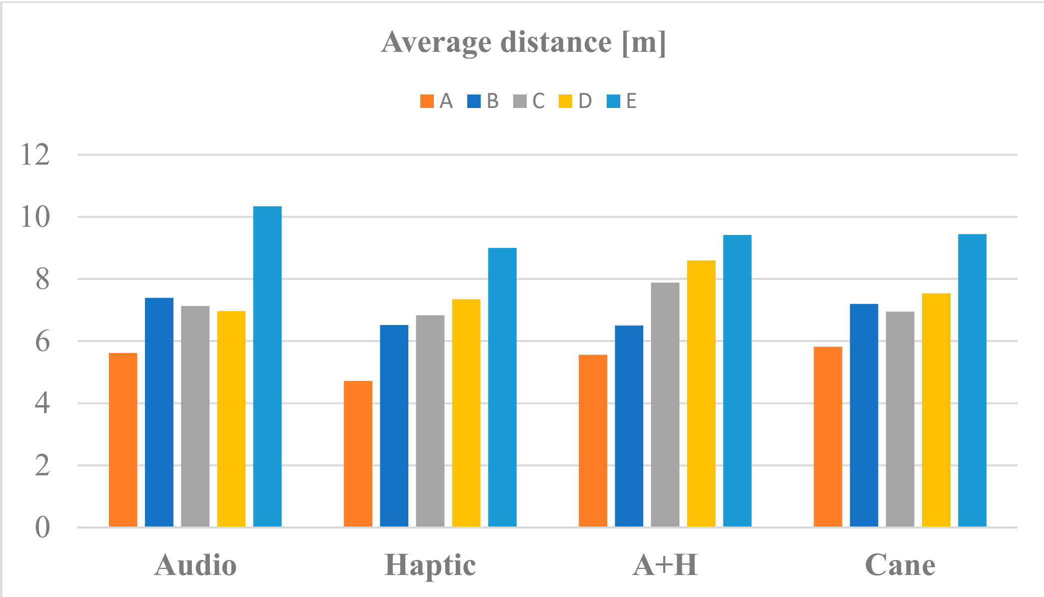


Figure 14. The average distances of navigation tasks.

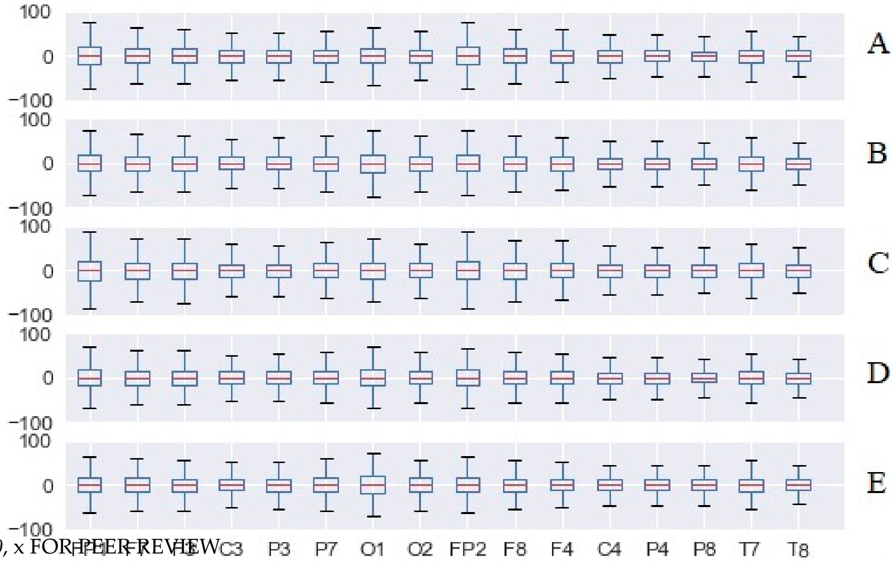
We noticed a higher number of collisions when using the SoV device in comparison to the white cane navigation. This fact was expected because usually, the VIPs touch the objects with the cane along their path, avoiding the great majority of collisions. The average distances did not differ too much between cane and SoV navigation, except for the SoV audio mode, for which the routes were significantly longer regardless of the testing scenario.

From Table 1 and Figures 12–14, it can be concluded that the required time, the length of the path chosen by each VIP according to his/her perception of the SoV stimulation, and the number of collisions depended on the complexity of the scene and on the user's training and ability to adapt to a new navigation aid. Obviously, the time, length of the path, and the number of collisions were much higher for the scenes C, D, and F. Some VIPs had better results with the audio mode and others with the haptic mode, but the number of collisions was higher for the haptic mode. As expected, the metrics for the white cane navigation were better because the VIPs were accustomed to using it on a daily basis. As a particular conclusion, *the VIPs' navigation performance with the SoV device was better in the case of the multimodal encoding, in terms of duration and number of collisions*. On the other hand, no general conclusion could be drawn because the number of VIPs involved in the experiments was small, and also a VIP could have learned the scenes during the first trials and performed better during the last trial, even if the experiments were randomly conducted.

6.2. Cognitive Load Analysis

Figures 15 and 16 present the total cognitive load for all the validated experiments and the five test scenarios, in the case of both the SoV device and white cane navigation. In the case of SoV navigation, we computed the average of the audio, haptic, and multimodal stimulations. Regardless of the difficulty of the test scenario (A is the easiest, and E is the most difficult), high values of CL were observed for the electrodes related to the frontal cortex (O1 especially, in the vision area) if the SoV device was used. The increase of frontal cortical activity was expected, but the activity of the visual cortex (VC) was worth being investigated because it supports some previous opinions about VC activation in the case of the VIPs who received various environmental sensory stimulation. It should be noted that the brain activity corresponding to the O1 channel was significantly higher than for the O2 channel for both types of navigation, resulting in an increased emotional state.

were observed for the electrodes related to the frontal cortex (O1 especially, in the vision area) if the SoV device was used. The increase of frontal cortical activity was expected, but the activity of the visual cortex (VC) was worth being investigated because it supports some previous opinions about VC activation in the case of the VIPs who received various environmental sensory stimulation. It should be noted that the brain activity corresponding to the O1 channel was significantly higher than for the O2 channel for both types of navigation, resulting in an increased emotional state.



Sensors **2020**, *20*, x FOR PEER REVIEW 17 of 31

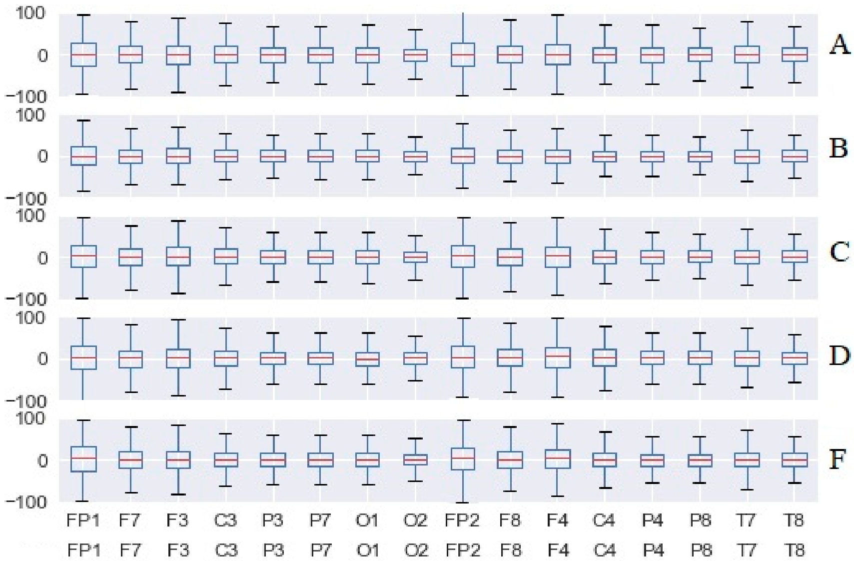


Figure 16. The total cognitive load of all electrodes for Sound of Vision (SoV) navigation.
Figure 16. The total cognitive load of all electrodes for Sound of Vision (SoV) navigation.

In Figure 17, we present the total CL for the scenario treasure hunt (TH), configuration C. The main conclusion was that there was a *significant increase in the CL index* (indicated as a negative fluctuation according to the CL index definition) *in the case of using the SoV device with audio, haptic, and multimodal stimulation in comparison to white cane navigation*. The conclusion was similar in the case of the other testing scenarios (TH, configurations A, B, D, and E).

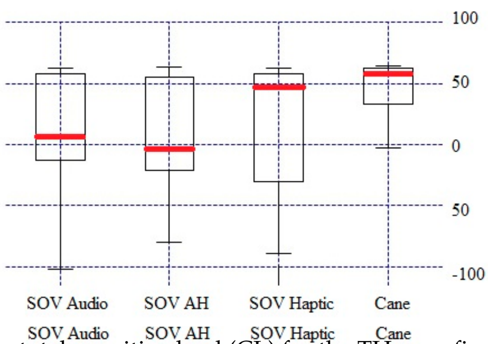


Figure 17. The total cognitive load (CL) for the TH—configuration C task.
Figure 17. The total cognitive load (CL) for the TH—configuration C task.

In contrast to the CL values presented in Figures 15 and 16 for each electrode, the global CL index was calculated on average for all brain waves for all users, aiming to have a general representation of the brain activity. The short box related to cane traveling meant that the data consistently hovered around the center value, denoting a similar effort for all users, and the whiskers

In contrast to the CL values presented in Figures 15 and 16 for each electrode, the global CL index was calculated on average for all brain waves for all users, aiming to have a general representation of the brain activity. The short box related to cane traveling meant that the data consistently hovered around the center value, denoting a similar effort for all users, and the whiskers indicated a quite limited distribution as well. In the case of using the SoV device with audio, haptic, and multimodal stimulation, the taller boxes indicated more variable data, and the whiskers showed a wider distribution, namely more scattered data. The different ways in which the users perceived the haptic and sonification models could explain this conclusion, which anyway was in accordance with the plots depicted in Figures 12–14. The consistency of the experimental data and the preprocessing accuracy were proved by the lack of outliers in the total CL indexes. Although the median value for haptic stimulation was closer to the median for cane walking, the distribution of global CL was the widest one. It could be observed that multimodal stimulation had the effect of reducing the spread of the global CL index. The tactile and auditory stimuli were processed by distinct lobes of the cerebral cortex with significant differences in CL, and this should explain the negative skewness in the case of cane and haptic stimulation and the positive skewness in the case of audio and multimodal stimulation. The conclusion was similar for the other testing configurations (A, B, D, and F) of the TH scenario.

6.3. Brain Activity Analysis

Besides this general evaluation of cognitive load, it was relevant to explore how the VIPs' cortical lobes were activated during walking on certain routes with the white cane or guided by the SoV device using the three input encodings. Particular reactions were expected, depending on the type of visual impairment and on the users' training or education. For this, the analysis of the individual frequency bands was performed according to the literature guidelines. First of all, *the alfa waves* were investigated, especially in the *frontal lobe*, taking into account that there is an *inverse relationship between alpha power and cortical activity;* namely, more brain activity (engagement) means less alpha power [63]. A more detailed analysis should be done if the alpha-1 (lower alpha, 7–10 Hz) and the alpha-2 (higher alpha, 10–13 Hz) frequencies were considered because it is well known that *alpha-1 is related to response inhibition and attentional demands,* and *alpha-2 reflects task performance in terms of speed, relevance, and difficulty* [64]. It has been proved that people with relatively *increased left-frontal alpha activity are more motivated and focused in a positive way,* and their related emotions are joy or anger. In contrast, *the increase of right-frontal activity denotes a more negative motivation accompanied by fear, sadness, and disgust* [63,65]. The asymmetry was calculated based on the difference between the logarithms of the spectral powers from the left and right brain hemispheres.

For a more accurate assessment of brain activity related to users with different perceptions and visual impairments, the envelopes of the alpha1 and alpha2 bands asymmetries were depicted for the considered navigation tasks, and the collisions annotated with Chronoviz were marked with black dots. It must be mentioned that the acquired signals for navigation with the cane or with the SoV device had different lengths, according to the time required to perform the task and the path chosen by the user, as it is presented in Table 1. The brain activity exploration was oriented towards analyzing the late visual impaired users in a group and the subjects who were born blind in another group.

In Figures 18 and 19, the envelopes of the asymmetries depicted are related to a user who was born blind (early-blind). He usually navigates using the cane, and he took part in all the training sessions in the virtual environment and ego-static real-world tests. *A significant difference between navigation with the cane and the SoV device was observed only for the audio encoding, in terms of response inhibition and attentional demands*. Although SoV is a completely new device that implies a different way of navigation, however, the consistent training in the virtual environment and the ego-static real-world tests helped a lot the user to accommodate to the encodings. *The greater attentional demand (reflected by the alpha1 waves) was evident for audio stimulation,* and it could be assumed that this was due to the fact that the VIPs usually rely heavily on the environmental noise when they navigate. They also try to perceive natural noises when the SSD sonification is conveyed to them. From the perspective of *alpha2 frequencies, meaning speed and task difficulty, for this user, for all the encodings, the values obtained for navigation with the SoV device were significantly higher than those obtained in the case of using the white cane*. Anyhow, this conclusion was expected, considering the novelty of the SoV system for the users and the fact that the VIPs walked relying on the white cane in a natural style for a long time. The collisions (marked with dark dots) were well correlated with the inflection points of the envelopes' variations.

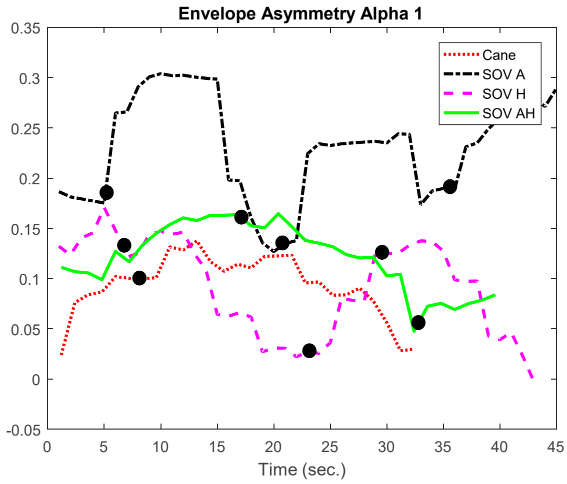


Figure 18. The envelopes of the alpha1 asymmetries for a user who was born blind.

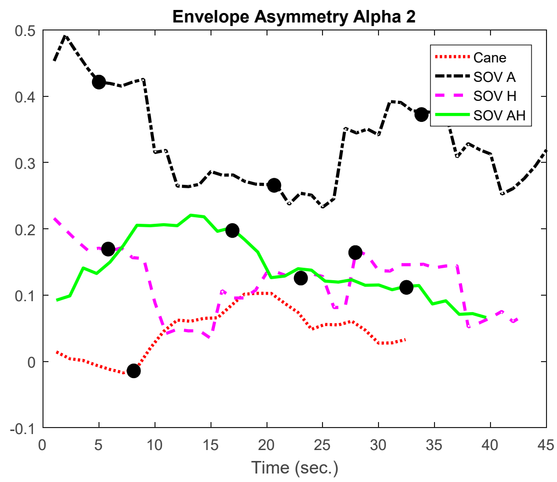


Figure 19. The envelopes of the alpha2 asymmetries for a user who was born blind.

In Figures 20 and 21, the envelopes presented correspond to a user from the late-blind group. He lost his sight at 17, has a good education, and usually navigates accompanied by a family member, without using the white cane. He quickly got used to the SoV device and got good scores in the training sessions. In this study case, the alpha1 asymmetry values (Figure 20) were higher for the cane navigation (even the necessary time was shorter), compared to those obtained for SoV navigation, regardless of how the environmental information was encoded. This demonstrated *a higher concentration for the cane navigation and good and fast accommodation with the SoV device. The alpha2 asymmetries (Figure 21) highlighted increasing difficulties for the audio and multimodal encoding tasks*. However, ***the range variations of the alpha1 and alpha 2 asymmetries were similar for the two users considered***. This suggested that ***an early and a late VIP had the same cognitive load, but there were differences between the navigation tasks: cane* vs. *SSD and between the different types of encodings (audio, haptic, and multimodal).***

Besides the two particular cases presented above, an overview of the CL analysis is presented in Figures 22–25. *By averaging the results for all the VIPs (early- and late-blind), in case of the most difficult trial (scenario E), it could be concluded that the alpha1 asymmetries for audio and multimodal codifications* *were a little higher (~0.2) compared to the cane and haptic modality (~0.1).* The variations within the whole asymmetries data set are displayed in the whisker plot from Figure 24.

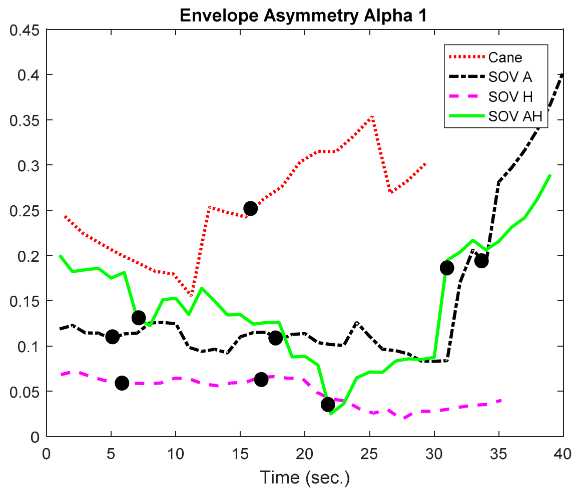


Figure 20. The envelopes of the alpha1 asymmetries for a late-blind user.

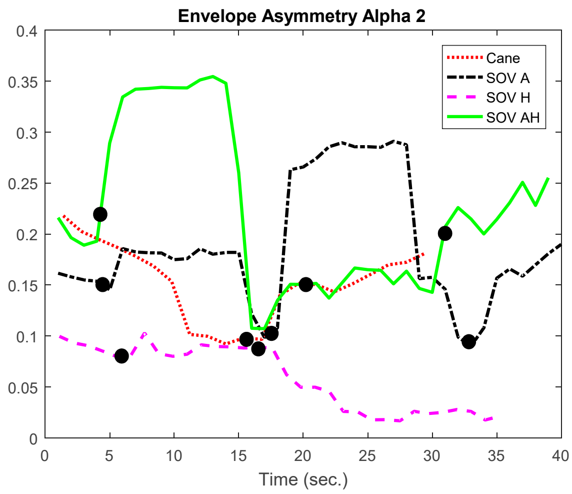


Figure 21. The envelopes of the alpha2 asymmetries for a late-blind user.

Besides the two pa … analysis is presented in Figures 22–25. *By avera… case of the most difficult trial (scenario E), it coul… multimodal codifications were a little higher (~0.2)* … iations within the whole asymmetries data set ar…

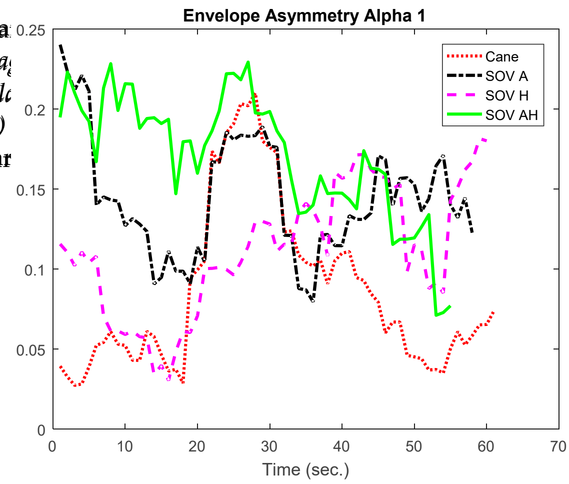
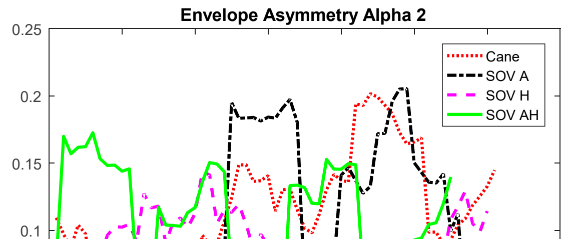


Figure 22. The average of the alpha1 asymmetry for all the users.



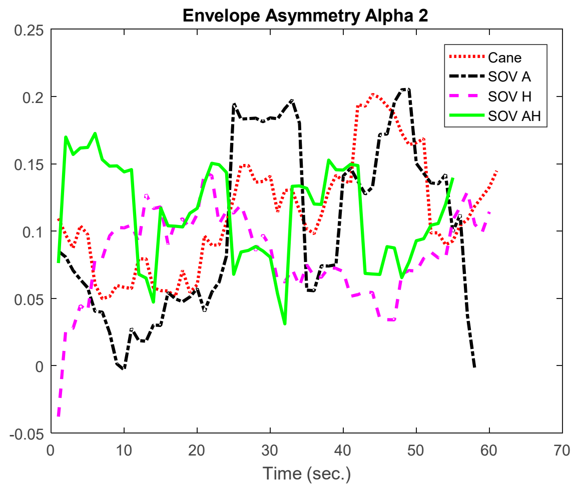


Figure 23. The average of the alpha2 asymmetry for all the users.

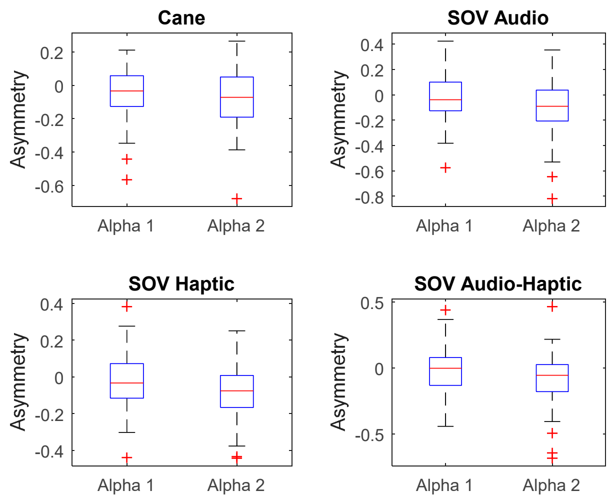


Figure 24. The envelopes for all the users, for scenario E.

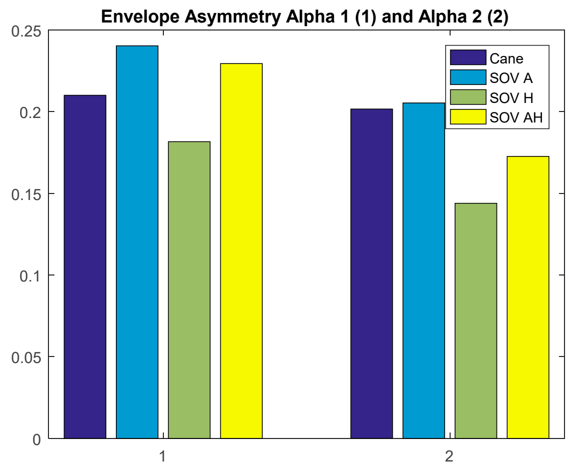


Figure 25. Average envelopes' asymmetries for all users, for scenario E.

6.4. Visual Cortex Activation Analysis

Some previous studies have revealed the presence of visual cortex activity in the case of the VIPs if a sensory substitution system creates an "information map" of the environment. Therefore, the total cognitive load (TCL) for electrodes O1 and O2 was investigated. Preliminary investigations

Figure 25. Average envelopes' asymmetries for all users, for scenario E.

6.4. Visual Cortex Activation Analysis

Some previous studies have revealed the presence of visual cortex activity in the case of the VIPs if a sensory substitution system creates an "information map" of the environment. Therefore, the total cognitive load (TCL) for electrodes O1 and O2 was investigated. Preliminary investigations of the experimental data showed that there was a major difference between the VC activity of the late-blind persons and of those who were born blind. Thus, a general conclusion regarding all VIPs could be drawn.

The asymmetries of TCL in the visual cortex for UserA (early-blind) and UserB (late-blind) were calculated. In Figures 26 and 27, the upper envelopes for TCL values are represented for UserA and UserB. *UserA was a born blind person, and, in his case, the asymmetry of the cane task was greater than the asymmetries of SoV tasks.* **In line with some previous research, it is possible to associate the VC activity of the cane task with the fact that parts of his visual cortex were activated to refine his sensations and usual activities.** In contrast, *the late-blind person's VC activity in SoV tasks was much higher than in the cane task (which was negative) and more than five times higher than UserA's visual activity.* **The limited number of VIPs from each group (five early- and three late-blind users) did not permit to obtain valuable statistical results, but** *for all the late-blind persons guided by audio and haptic stimuli, the average asymmetry of VC was around six times greater than that of the persons born blind,* **as can be seen in Figures 28 and 29. Another important observation was that** *VC asymmetry was lower in the case of multimodal stimulation than in the case of separate stimulation (audio or haptic).* **It must be emphasized that there was** *no correlation between visual cortical activity and the number of collisions during navigation, regardless of the type of navigation or sensory input.*

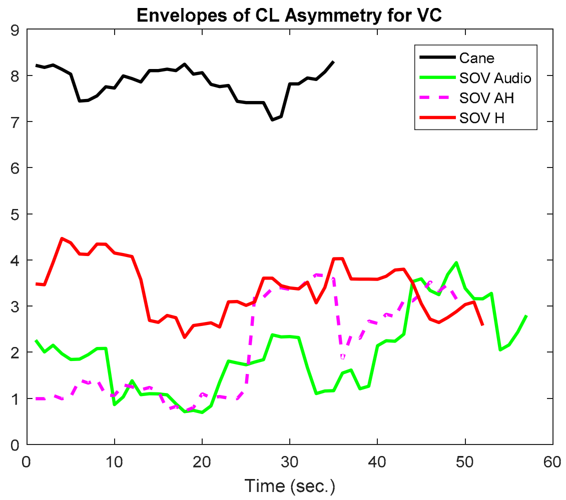


Figure 26. Visual cortex asymmetry for UserA (early-blind).

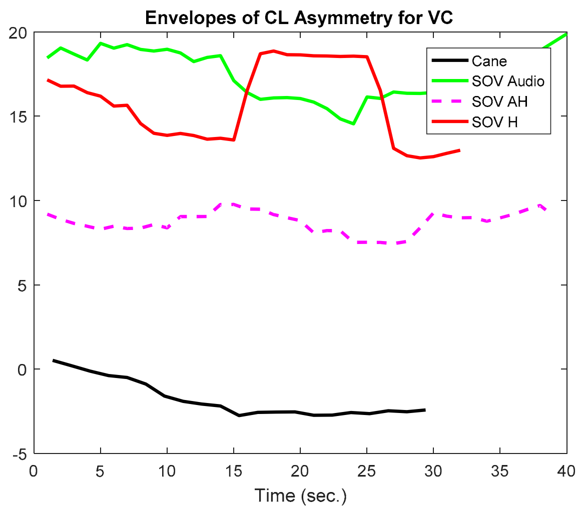
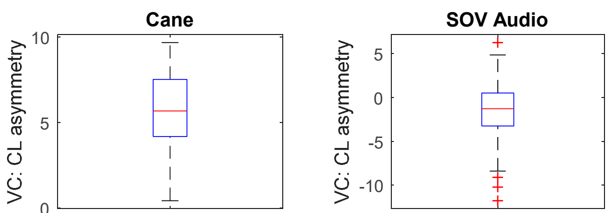


Figure 27. Visual cortex asymmetry for UserB (late-blind).



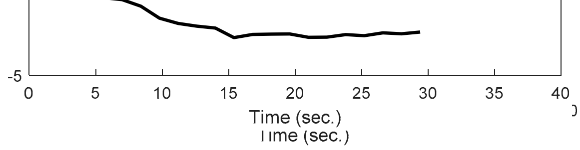


Figure 27. Visual cortex asymmetry for UserB (late-blind).

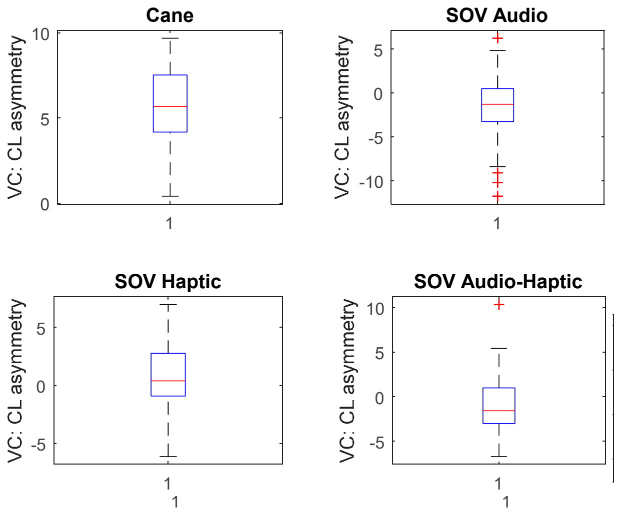


Figure 28. Visual cortex activity for UserA (early-blind).

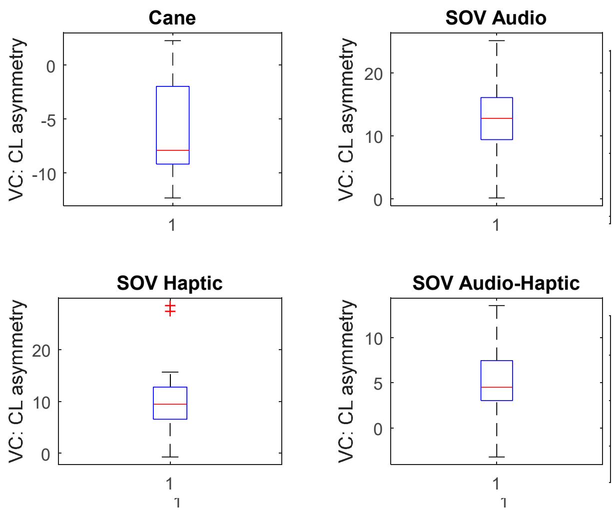


Figure 29. Visual cortex activity for UserB (late-blind).

6.5. Emotions Assessment During Real-World Navigation

As presented in the previous chapter, the emotional influence is a vast topic on which, from the dimensional perspective, the valence and arousal dimensions are advocated by Russell [66]. *Arousal* expresses calmness or excitement, whereas *valence* expresses a negative or positive effect. According to the comprehensive literature, *left frontal inactivation is an indicator of a withdrawal response*, which is often linked to a negative emotion, and *right frontal inactivation is a sign of an approach response* or positive emotion. Therefore, the **ratio of right and left asymmetry (valence state—VS**) was computed with the equation:

$$VS = log\left(\frac{PS_R}{PS_L}\right) \quad (2)$$

where PS_R and PS_L are the power spectrum values of the right and the corresponding left hemisphere channels in a specific frequency band. The channels T7-T8, which are considered the most relevant for emotions assessment, but also C3, C4, P3, P4, F3, and F4, were considered based on the theoretical statements from the previous chapter [62].

In addition, the HR and EDA signals acquired during the tests were processed according to the standard approaches described in the literature [65]. The root mean square of the successive differences (RMSSD) values were calculated for the HR recorded using the Shimmer sensor. A low RMSSD value means a high HR, denoting a strong concentration, emotion, or physical effort, whereas a high RMSSD value corresponds to resting or to a relaxing activity. The HR values in the resting state for the involved users were different according to their age and personal rhythm. Therefore, the percentage of variation of RMSSD compared to the resting state was calculated. By pre-processing the signals acquired with Shimmer, the EDA signals (µS) were obtained. Then, the deconvolution performed using the Ledalab software provided the phasic and tonic components and skin conductance responses (SCRs)—abrupt increases in the conductance of the skin, measured in µS, were calculated.

Firstly, the global VS of all the users and all the performed tests was calculated for the chosen pairs of electrodes. The result is graphically depicted in Figure 30. The C3 and C4 electrodes were considered because their waves could be associated with hippocampus activity, together with T8-T7, which obviously are the most relevant for assessing emotions [65]. The VS calculated using the T8-T7 pair (T8 in the right hemisphere denotes negative emotions and lack of motivation, in contrast to T7) indicated a *low positive valence for cane navigation and a negative valence for SoV navigation*, in accordance with the cumulative time, distance, and the number of collisions from Table 1. For the *parietal lobe, the P4-P3 pair indicated a moderate VS for cane navigation and low VS for SoV navigation*. The pair C4-C3 reflected a low positive valance, close to the neutral state. The same observation applied for the F4-F4 pair, for which the low negative valence in the SoV navigation using the audio encoding must be taken into account.

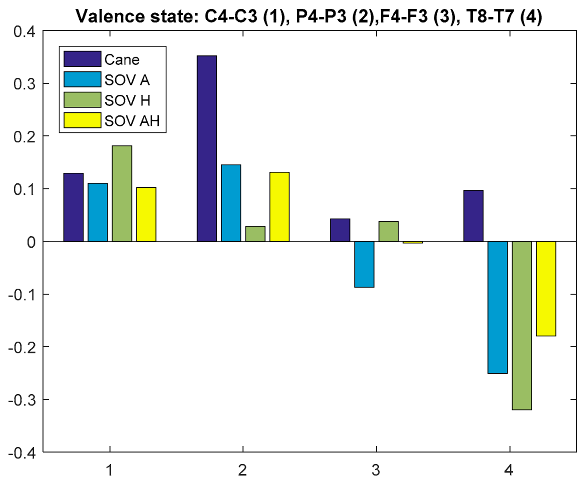


Figure 30. Average valence state (VS) for all visually impaired people (VIP).

The global percentage variation rate of RMSSD compared to the resting state decreased with: 21% for cane navigation, 39% for SoV navigation using the audio encoding, 44% for SoV navigation using the haptic encoding, and 41% for SoV navigation using the multimodal encoding. By computing the global SCR index, the following values were obtained: 0.18 for white cane navigation, 0.61 for SoV navigation using the audio encoding, 0.76 for SoV navigation using the haptic encoding, and 0.71 for SoV navigation using the multimodal encoding.

The trials were performed randomly within the same navigation type (cane or SoV) and for the same user, and no significant differences of the RMSSD values were remarked between the trials, even if they had different durations. Moreover, a slight increase (corresponding to an HR decrease) was observed towards the end of most of the tests. The users were not subjected to intense physical activity because they walked on the plain ground; however, the average of HR values was a little bit increased in comparison to the VTE tests. On the other hand, the EDA signals were sensitive to most of the collisions, especially in the case of the SoV navigation.

In Figures 31 and 32, we present the VS values for the two special users (UserA and UserB). In the case of UserA, who was intensively trained in the VTE and in the RW, the *valence was positive but very close to the neutral state for all the navigation types, with the remark that the valence for SoV navigation was a little bit higher than for cane navigation*. His RMSSD and SCR values did not differ significantly between the navigation modes.

him, navigating using the white cane induced a neutral to low negative valence, and navigating using the SoV device gave him more security, satisfaction, and a more comfortable state. This was underlined by a decrease of the RMSSD percentages and an increase of SCRs for all navigation modes.

In Figures 33 and 34, the evolution in time of the VS for the T8-T7 pair is represented for UserA and UserB in order to highlight the slow evolution of valence during a trial. In general, the collisions did not essentially affect the valence changes, as in the case of the cognitive load assessment.

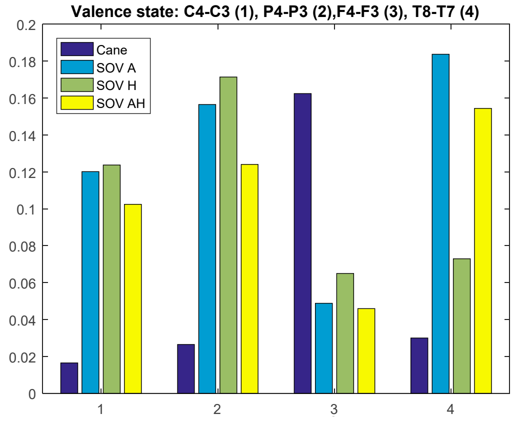


Figure 31. Average VS of UserA.

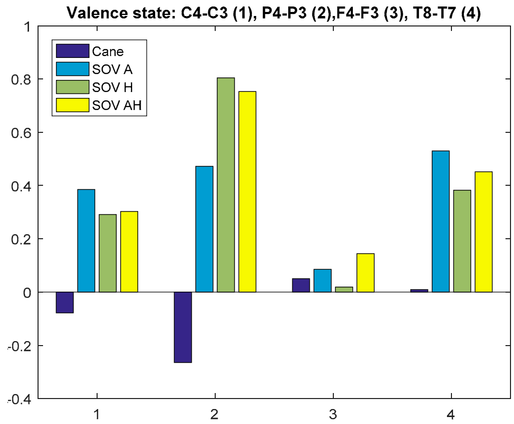


Figure 32. Average VS of UserB.

Analyzing the results for UserB, who is usually guided by a family person, it is obvious that for him, navigating using the white cane induced a neutral to low negative valence, and navigating using the SoV device gave him more security, satisfaction, and a more comfortable state. This was underlined by a decrease of the RMSSD percentages and an increase of SCRs for all navigation modes.

In Figures 33 and 34, the evolution in time of the VS for the T8-T7 pair is represented for UserA and UserB in order to highlight the slow evolution of valence during a trial. In general, the collisions did not essentially affect the valence changes, as in the case of the cognitive load assessment.

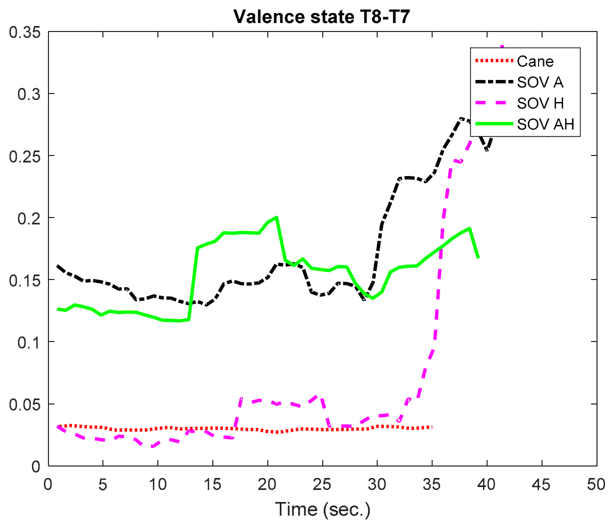


Figure 33. VS evolution in time for UserA.

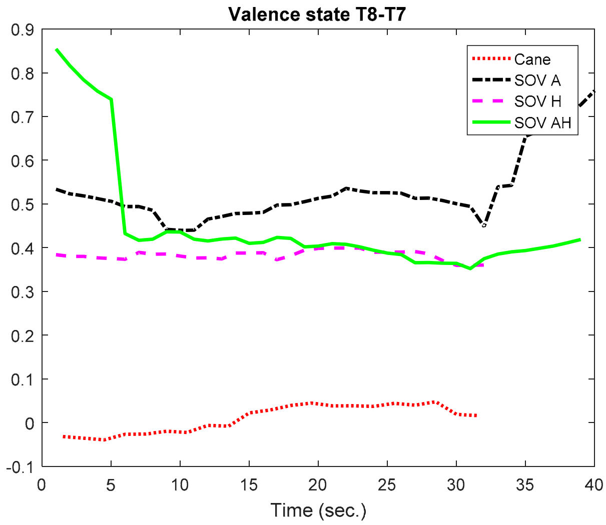


Figure 34. VS evolution in time for UserB.

6.6. Limitation of This Study

The limitations of the current study reside in the fact that we performed the tests with a small number of users for both categories (early-blind and late-blind). Although the results were interesting and in line with the existing literature, a more thorough evaluation should be realized. In addition, the EEG recordings were performed using a limited number of electrodes. As future directions, we plan to use a more advanced EEG recording device, with a higher number of electrodes, and to improve the SoV device so that it would be lighter and more comfortable to be worn.

The results of real-world experiments were strongly influenced by a consistent training period, similar to all users, which requires a great deal of time. Future work can extend the realistic scenarios of RW traveling for enhancing the impact of the study.

7. Conclusions

This paper presented an experimental framework and a study based on EEG, HR, and GSR signal analysis, aiming to assess the brain cortex activation and affective reactions of the visually impaired persons to the stimuli provided by a sensory substitution device used for navigation in real-world scenarios, compared to the white cane navigation. The study was focused on the evaluation of working memory load, visual cortex activation, and emotional experience when the VIPs perceived audio, haptic, and multimodal stimuli during a navigation task in five different types of scenarios.

The choice of the Brain-Computer Interface (BCI) equipment proved to be inspired because its characteristics allowed a good acquisition of EEG signals simultaneously with the use of the SoV device. The same BCI equipment has been employed successfully in other studies of our own concerning multimodal neuromotor rehabilitation [67,68]. An important feature of the experimental setup is the ability to synchronize the data streams and to align the acquired signals with the events extracted from the video recordings. The training performed in the VTE and the ego-static tests performed indoors had an essential role in preparing the users to perceive distances, positions, and object dimensions only by means of the audio, haptic, and multimodal stimuli, giving confidence to all of them in using the SoV device. The aim was to provide all users the ability to automatically understand the complexity of a scene. Besides, during the VTE training, multiple resting sessions were recorded for all volunteers, which had an important role in establishing a baseline.

The perception of audio and haptic stimuli using the SoV device was assessed in terms of cognitive load, pleasantness, excitement, and events, for all the visually impaired users, as well as for the specific categories (early-blind or late-blind). *All in all, the haptic stimuli appeared to be less intuitive than the audio stimuli.*

The analysis showed that *navigating with the SoV device increased the cognitive load and the working memory (lower accuracy and longer response times).* The analysis of the EEG data revealed *the usage of verbal working memory in the posterior parietal cortices*. The obtained results indicated that the left-right asymmetry of the prefrontal cortex had distinguishable characteristics when the VIPs were navigating in real-world *environments with a wide range of obstacles*.

The visual cortex exploration revealed *a significant activation when using the SoV device, only for the late VIPs*. The low VC activity of congenitally blind persons during SoV navigation could be related to brain plasticity, which allows the auditory and somatosensory systems to extend their functionality in that part of the cortex.

Finally, we assessed the valence state of the users when navigating in unfamiliar indoor environments based on mobile monitoring and a fusion of EEG and physiological (EDA and HR) signals. *For the generic VIP population, the use of the SoV device induced a low negative valence in contrast with cane usage*. But the findings differed for the specific categories of sight loss (early- and late-blind), pointing out the particular needs/difficulties faced by each category of VIP.

This study proved once more that sensory substitution is an alternative method, which helps the blind people to acquire information about the surrounding space and to navigate independently in unknown real-world environments, safely and comfortably, after substantial training.

The findings hopefully empower the knowledge of how the visually impaired persons are stressed and emotionally affected by SSD navigation and contribute to the development of the intelligent navigation devices, aiming for the VIPs' safety and well-being. The results of our work can inspire researchers working in the field of IoT devices comprising sensors, antennas, and Bluetooth, which have created navigation rules based on a fuzzy controller [69], GPS embedded in a stick with voice recognition for obstacles detection [70], computer vision-based assistants [71], or assistive systems relying on wearable smart glasses and mobile applications [72].

Valuable research projects have investigated the efficiency of intelligent sensory substitution devices [73], and, in this context, our research brought an important contribution by analyzing EEG and physiological signals in order to assess the cognitive effort and emotional state of users in real-world navigation.

Author Contributions: Conceptualization, F.U. and R.-G.L.; methodology, A.S., F.U., K.K., O.M. and A.M.; software, R.-G.L. and A.S.; validation, O.M., F.U., K.K. and A.M.; formal analysis, F.U.; investigation, O.M., F.U. and K.K.; resources, R.-G.L., A.S. and O.M.; data curation, F.U.; writing—original draft preparation, R.-G.L., F.U., O.M. and A.M.; writing—review and editing, A.M.; visualization, F.U., O.M., R.-G.L., A.S. and K.K.; supervision, A.M.; project administration, A.M.; funding acquisition, A.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 643636 "Sound of Vision". Financial support from the Competitiveness Operational Program 2014–2020, Action 1.1.3: Creating synergies with RDI actions of the EU's HORIZON 2020 framework program and other international RDI programs, MySMIS Code 108792, Acronym project "UPB4H", financed by contract: 250/11.05.2020.

Acknowledgments: The authors wish to thank the volunteer visually impaired subjects for their collaboration and involvement. This work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 643636 "Sound of Vision". Financial support from the Competitiveness Operational Program 2014–2020, Action 1.1.3: Creating synergies with RDI actions of the EU's HORIZON 2020 framework program and other international RDI programs, MySMIS Code 108792, Acronym project "UPB4H", financed by contract: 250/11.05.2020 is gratefully acknowledged.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. World Health Organization—Blindness and Vision Impairment. Available online: https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment (accessed on 10 September 2020).
2. Sound of Vision. Available online: https://soundofvision.net/ (accessed on 10 September 2020).

3. Väljamäe, A.; Kleiner, M. Spatial sound in auditory vision substitution systems. *Audio Eng. Soc. Conv.* **2006**, *120*, 6795.
4. Fournier d'Albe, E.E. The Optophone: An instrument for reading by ear. *Nature* **1920**, *105*, 295–296.
5. Meijer, P.B.L. The Voice Technology. Available online: http://www.seeingwithsound.com (accessed on 10 September 2020).
6. Fish, R.M. Audio Display for Blind. *IEEE Trans. Biomed. Eng.* **1976**, *23*, 144–154. [CrossRef]
7. Sanz, P.R.; Mezcua, B.R.; Sánchez Pena, J.M. A sonification proposal for safe travels of blind people. In Proceedings of the 18th International Conference on Auditory Display, Atlanta, GA, USA, 18–21 June 2012.
8. Milios, E.; Kapralos, B.; Kopinska, A.; Stergiopoulos, S. Sonification of range information for 3-D space perception. *IEEE Trans. Neural Syst. Rehab. Eng.* **2003**, *11*, 416–421. [CrossRef]
9. Bestpluton World Cie. The "Mini-Radar", Your Small Precious Companion that Warns You Obstacles in a Spoken Way, and that Helps You to Walk Straight. Available online: http://aweb1208.free.fr/EnglishMiniRadar.htm (accessed on 10 September 2020).
10. Pun, T.; Roth, P.; Bologna, G.; Moustakas, K.; Tzovaras, D. Image and video processing for visually handicapped people. *J. Image Video Process.* **2007**, *5*, 1–12.
11. Kaczmarek, K.A.; Bach-y-Rita, P. Tactile displays. In *Virtual Environments and Advanced Interface Design*; Barfield, W., Furness, T.A., Eds.; Oxford University Press Inc.: Oxford, UK, 1995.
12. Wall, S.A.; Brewster, S. Sensory substitution using tactile pin arrays: Human factors, technology, and applications. *Signal Process.* **2006**, *86*, 3674–3695. [CrossRef]
13. Loomis, J.M.; Lederman, S.J. Tactual perception. In *Handbook of Perception and Human Performance: Cognitive Processes and Performance*; Boff, K.R., Kaufman, L., Thomas, J.P., Eds.; John Wiley & Sons: New York, NY, USA, 1986.
14. Cancar, L.; Díaz, A.; Barrientos, A.; Travieso, D.; Jacobs, D.M. Tactile-sight: A sensory substitution device based on distance-related vibrotactile flow. *Int. J. Adv. Robot. Syst.* **2013**, *10*, 272. [CrossRef]
15. Parkes, D. Nomad: An audio-tactile tool for the acquisition, use and management of spatially distributed information by visually impaired people. In Proceedings of the 2nd International Symposium on Maps and Graphics for Visually Impaired People, London, UK, 20–22 April 1988; pp. 24–29.
16. Dewhurst, D. Creating and accessing audio-tactile images with HFVE vision substitution software. In *Proceedings of the Third Interactive Sonification Workshop*; KTH: Stockholm, Sweden, 2010; pp. 101–104.
17. Dewhurst, D. Using imprints to summarise accessible images. In Proceedings of the Interactive Sonification Workshop ISon, 4th Interactive Sonification Workshop, Fraunhofer IIS, Erlangen, Germany, 10 December 2013.
18. Hosseini, S.A.; Naghibi-Sistani, M.B. Classification of emotional stress using brain activity. In *Applied Biomedical Engineering*; Gargiulo, G.D., McEwan, A., Eds.; InTech: Rijeka, Croatia, 2011; pp. 313–336.
19. Jena, S.K. Examination stress and its effect on EEG. *Int. J. Med. Sci. Public Health* **2015**, *11*, 1493–1497. [CrossRef]
20. Bos, D.O. *EEG-Based Emotion Recognition: The Influence of Visual and Auditory Stimuli*; Department of Computer Science, University of Twente: Enschede, The Netherlands, 2006.
21. Antonenko, P.; Pass, F.; Grabner, R.; van Gog, T. Using Electroencephalography to Measure Cognitive Load. *Educ. Psychol. Rev.* **2010**, *22*, 425–438. [CrossRef]
22. Berka, C.; Levendowski, D.J.; Lumicao, M.N.; Yau, A.; Davis, G.; Zivkovic, V.T.; Olmstead, R.E.; Tremoulet, P.D.; Craven, P.L. EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviat. Space Environ. Med.* **2007**, *78*, B231–B244.
23. Nilsson, R.M.; Mayer, R.E. The effects of graphics organizers giving cues to the structure of a hypertext document on users' navigation strategies and performance. *Int. J. Hum. Comput. Stud.* **2002**, *57*, 1–26. [CrossRef]
24. Scott, B.M.; Schwartz, N.H. Navigational spatial displays: The role of metacognition as cognitive load. *Learn. Instr.* **2007**, *17*, 89–105. [CrossRef]
25. Pekrun, R. The impact of emotions on learning and achievement: Towards a theory of cognitive/motivational mediators. *Appl. Psychol.* **1992**, *41*, 359–376. [CrossRef]
26. 26. Soleymani, M.; Pantic, M.; Pun, T. Multimodal emotion recognition in response to videos (Extended abstract). In Proceedings of the International Conference on Affective Computing and Intelligent Interaction (ACII), Xi'an, China, 21–24 September 2015; pp. 491–497.

27. Bower, G.H.; Forgas, J.P. *Mood and Social Memory*; Lawrence Erlbaum Associates Publishers: Mahwah, NJ, USA, 2001.
28. Cattaneo, Z.; Vecchi, T.; Cornoldi, C.; Mammarella, I.; Bonino, D.; Ricciardi, E.; Pietrini, P. Imagery and spatial processes in blindness and visual impairment. *Neurosci. Biobehav. Rev.* **2008**, *32*, 1346–1360. [CrossRef]
29. Healey, J.A.; Picard, R.W. Detecting stress during real-world driving tasks using physiological sensors. *IEEE Trans. Intell. Transp. Syst.* **2005**, *6*, 156–166. [CrossRef]
30. Setz, C.; Arnrich, B.; Schumm, J.; La Marca, R.; Tröster, G. Discriminating stress from cognitive load using a wearable EDA device. *IEEE Trans. Inf. Technol. Biomed.* **2010**, *14*, 410–417. [CrossRef]
31. Peper, E.; Harvey, R.; Lin, I.M.; Tylova, H.; Moss, D. Is There More to Blood Volume Pulse Than Heart Rate Variability, Respiratory Sinus Arrhythmia, and Cardiorespiratory Synchrony? *Biofeedback* **2007**, *35*, 54–61.
32. Peake, P.; Leonard, J.A. The use of heart rate as an index of stress in blind pedestrians. *Ergonomics* **1971**, *14*, 189–204. [CrossRef]
33. Wycherley, R.J.; Nicklin, B.H. The heart rate of blind and sighted pedestrians on a town route. *Ergonomics* **1970**, *13*, 181–192. [CrossRef]
34. Urgen, B.A.; Plank, M.; Ishiguro, H.; Poizner, H.; Saygin, A. EEG theta and Mu oscillations during perception of human and robot actions. *Front. Neurorobotics* **2013**, *7*, 19. [CrossRef]
35. Gallistel, C.R. *The Organization of Learning*; MIT Press: Cambridge, MA, USA, 1993.
36. Pinquart, M.; Pfeiffer, J.P. Psychological well-being in visually impaired and unimpaired individuals: A meta-analysis. *Br. J. Vis. Impair.* **2011**, *29*, 27–45. [CrossRef]
37. Kalimeri, K.; Saitis, C. Exploring Multimodal Biosignal Features for Stress Detection during Indoor Mobility. In Proceedings of the 18th ACM International Conference on Multimodal Interaction, Tokyo, Japan, 12–16 November 2016.
38. Saitis, C.; Parvez, M.Z.; Kalimeri, K. Cognitive Load Assessment from EEG and Peripheral Biosignals for the Design of Visually Impaired Mobility Aids. *Wirel. Commun. Mob. Comput.* **2018**, *2018*. [CrossRef]
39. Saitis, C.; Kalimeri, K. Identifying urban mobility challenges for the visually impaired with mobile monitoring of multimodal biosignals. In Proceedings of the International Conference on Universal Access in Human-Computer Interaction, Toronto, ON, Canada, 17–22 July 2016.
40. Amedi, A.; Stern, W.; Camprodon, J.; Bermpohl, F.; Merabet, L.; Rotman, S.; Hemond, C.; Meijer, P.; Pascual-Leone, A. Shape conveyed by visual-to-auditory sensory substitution activates the lateral occipital complex. *Nat. Neurosci.* **2007**, *10*, 687–689. [CrossRef]
41. Naue, N.; Rach, S.; Struber, D.; Huster, R.J.; Zaehle, T.; Korner, U.; Herrmann, C.S. Auditory Event-Related Response in Visual Cortex Modulates Subsequent Visual Responses in Humans. *J. Neurosci.* **2011**, *31*, 7729–7736. [CrossRef] [PubMed]
42. Sound of Vision Outputs. Available online: https://soundofvision.net/outputs/ (accessed on 10 September 2020).
43. Moldoveanu, A.D.B.; Ivascu, S.; Stanica, I.; Dascalu, M.-I.; Lupu, R.; Ivanica, G.; Morar, A. Mastering an advanced sensory substitution device for visually impaired through innovative virtual training. In Proceedings of the IEEE 7th International Conference on Consumer Electronics-Berlin (ICCE-Berlin), Berlin, Germany, 3–6 September 2017; pp. 120–125. [CrossRef]
44. Dascalu, M.; Moldoveanu, A.; Balan, O.; Lupu, R.G.; Ungureanu, F.; Caraiman, S. Usability assessment of assistive technology for blind and visually impaired. In Proceedings of the 2017 E-Health and Bioengineering Conference (EHB), Sinaia, Romania, 22–24 June 2017; pp. 523–526. [CrossRef]
45. Botezatu, N.; Caraiman, S.; Rzeszotarski, D.; Strumillo, P. Development of a versatile assistive system for the visually impaired based on sensor fusion. In Proceedings of the 2017 21st International Conference on System Theory, Control and Computing (ICSTCC), Sinaia, Romania, 19–21 October 2017; pp. 540–547. [CrossRef]
46. Kristjánsson, A.; Moldoveanu, A.; Jóhannesson, O.I.; Balan, O.; Spagnol, S.; Valgeirsdóttir, V.V.; Unnthorsson, R. Designing sensory-substitution devices: Principles, pitfalls and potential. *Restor. Neurol. Neurosci.* **2016**, *34*, 769–787. [CrossRef] [PubMed]
47. Moldoveanu, A.; Taloi, B.; Balan, O.; Stanica, I.; Flamaropol, D.; Dascalu, M.I.; Moldoveanu, F.; Mocanu, I. Virtual Mini-Games—A Serious Learning Tool for Sensory Substitution Devices. In Proceedings of the EduLearn Conference, Barcelona, Spain, 3–5 July 2017.

48. Spagnol, S.; Wersényi, G.; Bujacz, M.; Bălan, O.; Martínez, M.H.; Moldoveanu, A.; Unnthorsson, R. Current Use and Future Perspectives of Spatial Audio Technologies in Electronic Travel Aids. *Wirel. Commun. Mob. Comput.* **2018**, *2018*. [CrossRef]
49. Caraiman, S.; Morar, A.; Owczarek, M.; Burlacu, A.; Rzeszotarski, D.; Botezatu, N.; Moldoveanu, A. Computer Vision for the Visually Impaired: The Sound of Vision System. In Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCVW), Venice, Italy, 22–29 October 2017; pp. 1480–1489. [CrossRef]
50. Lab Streaming Layer (LSL), Multi-Modal Time-Synched Data Transmission over Local Network. Available online: https://github.com/sccn/labstreaminglayer (accessed on 10 September 2020).
51. HDF5 Data Model, Library, and File Format for Storing and Managing Data. Available online: https://support.hdfgroup.org/HDF5/ (accessed on 10 September 2020).
52. ChronoViz, Visualization of Time-Based Multimodal Data. Available online: https://chronoviz.com/ (accessed on 10 September 2020).
53. ExifTool. Available online: https://exiftool.org/ (accessed on 10 September 2020).
54. iMotions. Available online: https://imotions.com/ (accessed on 10 September 2020).
55. Hawthorne, B.; Vella-Brodrick, D.; Hattie, J. Well-Being as a Cognitive Load Reducing Agent: A Review of the Literature. *Front. Educ.* **2019**, *4*, 121. [CrossRef]
56. Alarcao, S.M.; Fonseca, M.J. Emotions Recognition Using EEG Signals: A Survey. *IEEE Trans. Affect. Comput.* **2019**, *10*, 374–393. [CrossRef]
57. Sadato, N.; Pascual-Leone, A.; Grafman, J.; Ibañez, V.; Deiber, M.P.; Dold, G.; Hallett, M. Activation of the primary visual cortex by Braille reading in blind subjects. *Nature* **1996**, *380*, 526–528. [CrossRef]
58. Burton, H. Visual cortex activity in early and late blind people. *J. Neurosci.* **2003**, *23*, 4005–4011. [CrossRef]
59. Georgetown University Medical Center. People Blind from Birth Use Visual Brain Area to Improve Other Senses: Can Hear and Feel with Greater Acuity. *ScienceDaily*. 10 October 2010. Available online: www.sciencedaily.com/releases/2010/10/101006131203.htm (accessed on 1 September 2020).
60. Campus, C.; Sandini, G.; Amadeo, M.B.; Gori, M. Stronger responses in the visual cortex of sighted compared to blind individuals during auditory space representation. *Sci. Rep.* **2019**, *9*, 1935. [CrossRef] [PubMed]
61. Bălan, O.; Moise, G.; Petrescu, L.; Moldoveanu, A.; Leordeanu, M.; Moldoveanu, F. Emotion Classification Based on Biophysical Signals and Machine Learning Techniques. *Symmetry* **2020**, *12*, 21. [CrossRef]
62. Zhang, J.; Chen, M.; Zhao, S.; Hu, S.; Shi, Z.; Cao, Y. Relief F-Based EEG Sensor Selection Methods for Emotion Recognition. *Sensors* **2016**, *16*, 1558. [CrossRef] [PubMed]
63. iMotions, Frontal Asymmetry 101—How to Get Insights on Motivation and Emotions from EEG. Available online: https://imotions.com/blog/frontal-asymmetry-101-get-insights-motivation-emotions-eeg/ (accessed on 10 September 2020).
64. Ray, W.J.; Cole, H.W. EEEG alpha activity reflects attentional demands, and beta activity reflects emotional and cognitive processes. *Science* **1985**, *228*, 750–752. [CrossRef] [PubMed]
65. Saitis, C.; Kalimeri, K. Multimodal Classification of Stressful Environments in Visually Impaired Mobility Using EEG and Peripheral Biosignals. *IEEE Trans. Affect. Comput.* **2018**. [CrossRef]
66. Russell, J.A. Core affect and the psychological construction of emotion. *Psychol. Rev.* **2003**, *110*, 145–172. [CrossRef]
67. Ferche, O.; Moldoveanu, A.; Cinteză, D.; Toader, C.; Moldoveanu, F.; Voinea, A.; Taslitchi, C. From neuromotor command to feedback: A survey of techniques for rehabilitation through altered perception. In Proceedings of the E-Health and Bioengineering Conference (EHB), Iasi, Romania, 19–21 November 2015; pp. 1–4.
68. Moldoveanu, A.; Ferche, O.; Moldoveanu, F.; Lupu, R.; Cinteză, D.; Irimia, D.C.; Toader, C. The TRAVEE System for a Multimodal Neuromotor Rehabilitation. *IEEE Access* **2015**, *7*, 8151–8171. [CrossRef]
69. Tayyaba, S.; Ashraf, M.W.; Alquthami, T.; Ahmad, Z.; Manzoor, S. Fuzzy-Based Approach Using IoT Devices for Smart Home to Assist Blind People for Navigation. *Sensors* **2020**, *20*, 3674. [CrossRef]
70. Mala, N.S.; Thushara, S.; Subbiah, S. Navigation gadget for visually impaired based on IoT. In Proceedings of the 2nd International Conference on Computing and Communications Technologies (ICCCT), Chennai, India, 23–24 February 2017; pp. 334–338.
71. Islam, M.M.; Sadi, M.S.; Zamli, K.Z.; Ahmed, M.M. Developing Walking Assistants for Visually Impaired People: A Review. *IEEE Sens. J.* **2019**, *19*, 2814–2828. [CrossRef]

72. Chang, W.; Chen, L.; Chen, M.; Su, J.; Sie, C.; Yang, C. Design and Implementation of an Intelligent Assistive System for Visually Impaired People for Aerial Obstacle Avoidance and Fall Detection. *IEEE Sens. J.* **2020**, *20*, 10199–10210. [CrossRef]
73. Jicol, C.; Lloyd-Esenkaya, T.; Proulx, M.J.; Lange-Smith, S.; Scheller, M.; O'Neill, E.; Petrini, K. Efficiency of Sensory Substitution Devices Alone and in Combination with Self-Motion for Spatial Navigation in Sighted and Visually Impaired. *Front. Psychol.* **2020**, *11*, 1443. [CrossRef] [PubMed]

Neuromotor Recovery Based on BCI, FES, Virtual Reality and Augmented Feedback for Upper Limbs

Robert Gabriel Lupu, Florina Ungureanu, Oana Ferche, and Alin Moldoveanu

Abstract Recently investigated rehabilitative practices involving Brain-Computer Interface (BCI) and Functional Electrical Stimulation (FES) techniques provided long-lasting benefits after short-term recovering programs. The prevalence of this revolutionary approach received a boost from virtual reality and augmented reality, which contribute to the brain neuroplasticity improvement and can be used in neurorehabilitation and treatment of motor/mental disorders. This work presents a therapy system for stroke rehabilitation based on these techniques. The novelty of the proposed system consists of including an eye tracking device that detects the patient's vigilance during exercises and warns if patient is not focused on the items of interest from the virtual environment. This additional feature improves the level of user involvement and makes him/her conscious of the rehabilitation importance and pace. Moreover, the system architecture is reconfigurable, and the functionalities are specified by software. The laboratory tests have validated the system from a technical point of view, and preliminary results from the clinical tests have highlighted the system's quick accommodation to the proposed therapy and fast progress for each user.

1 Introduction

Rehabilitation is an important part of recovery and helps the patient to become more independent after a stroke or a motor/mental disorder. In the last decade, the Brain-Computer Interface (BCI), the Virtual Reality (VR) and the Functional Electrical Stimulation (FES) techniques are widely used in more complex and efficiently

R. G. Lupu (✉) · F. Ungureanu
Faculty of Automatic Control and Computer Engineering, Computer Engineering Department, "Gheorghe Asachi" Technical University of Iasi, Dimitrie Mangeron 27, 700050 Iasi, Romania
e-mail: lupu.robert@gmail.com

O. Ferche · A. Moldoveanu
Computer Engineering Department, "Politehnica" University of Bucharest, Bucharest, Romania

C. Guger et al. (eds.), *Brain–Computer Interface Research*,
SpringerBriefs in Electrical and Computer Engineering,
https://doi.org/10.1007/978-3-030-49583-1_8

systems aiming to bolster the rehabilitation process. In this context, different specific devices became affordable, and many research groups and health institutions are focused on motor, cognitive or speech recovery after stroke (Stroke Centre from Johns Hopkins Institute0, ENIGMA-Stroke Recovery, StrokeBack) [1–3].

In this paper, we present an affordable system for recovery of patients with neuromotor impairments following strokes, traumas or brain surgery. It relies on a brand new idea—recovery through augmented and magnified feedback—that creates new, distinct possibilities to overcome block stages typical to early recovery, to stimulate recovery through neuroplasticity. The system was customised and tested for upper-limb recovery but can be tailored for any other particular purpose. Another own idea of our approach is that the tasks and guidance are provided by a virtual therapist—a new concept in the field of rehabilitation and considered extremely promising by the healthcare professionals. Besides others research projects dedicated to upper limb recovery (RETRAINER, NIHR) [4, 5] or very recent published works [6, 7], our solution makes use of an eye-tracking method to provide a warning if the patient stops concentrating during exercises.

The purpose of the proposed recovery system is to help in fulfilling the causal chain/loop of recovery, consisting mainly of three steps: motor act is performed or attempted, by the patient, with or without external help; the patient observes sensations and results (visually, haptic or proprioceptive); the patient's cortex associates the motor act with the observations and gradually learns and perfects the motor act. Most techniques and systems for neuromotor recovery only pay attention to the motor act performance, neglecting the essentiality of observation. The system handles the whole recovery causal chain in a unified way. Previous versions and facilities of presented recovery system were designed and implemented in the framework of TRAVEE project [8] and are presented in a comprehensive manner in some published papers [9–12].

From a user's point of view, the system has two main components: one that is dedicated to the patient that undergoes the rehabilitation process after stroke, and one that is dedicated to the therapist—the clinician that guides the rehabilitation session [8]. The complex system dedicated to stroke rehabilitation involves devices and software that immerse the patient in a Virtual Environment to identify themselves with the presented avatar, as well as devices dedicated to support his/her movements and providing complex feedback during the exercises. The component for the therapist is aimed mostly at providing intuitive tools for configuring the rehabilitation session composition and the devices used for each exercise, as well as to monitor the activity of the patient.

2 Materials and Methods

The system is designed to support three main features: patient monitoring, patient training and stimulation and data analysis and processing, Fig. 1a. Devices for the first two features are each optional "plugin" components of the system. Hence, the

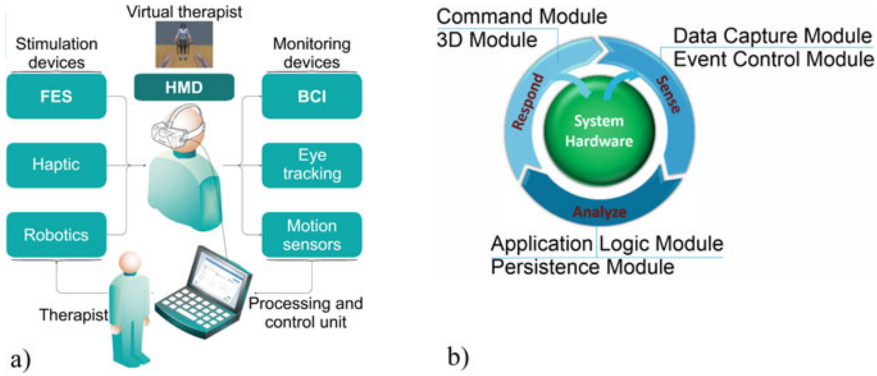


Fig. 1 **a** The system architecture; **b** Sense—analyse—respond mapping functions

system can be tailored to use all of the devices together, but not all of them are mandatory. Results of processed raw data from monitoring devices are used to trigger the stimulation devices with respect to rehabilitation exercise [11]. The software has an event driven architecture to manage needs like real-time data processing, communication and security, data access/storage, patient condition and working conditions, interoperability [9]. The running processes are managed through the sense-analyse-respond approach shown in Fig. 1b. In the “sense” component, monitoring devices capture and process data in real-time. “Analyse” refers to continuous evaluation of the processing results (when clause of the when-then rules) in order to decide to “respond” by executing the then clause.

From the first category, the used hardware devices manage the system functions of continuous patient monitoring during the exercises and the movement and stimulation of the upper limb that needs to be rehabilitated. The processing and control unit (PCU) determines the correctness of the exercise performed by the patient based on information received from used monitoring devices. The same information is used to update the patient avatar from the virtual environment in which the patient is immersed through the use of VR glasses (HMD). If more than one monitoring device is used, then the system aggregates and synchronizes the gathered information to interpret the status and actions of the patient. The following monitoring devices have been tested: g.tec gUSBamp & gBSanalyze, Kinect V2, Leap Motion, video cam + ArUco markers, Myo armband, EMG, DGTech glove.

The stimulation devices are used to restore and to maintain muscle tone and/or to assist the patient when performing the recovery exercises. The processing and control device synchronize all events and decisions to allow the system to act as a whole. Both hardware equipment and software components are selected to fulfil the system constraints regarding the performance and operational safety. The following stimulation devices have been tested for the best setup and configuration: Oculus Rift/HTC Vive, headphones, Motionstim 8, robotic glove, wireless sticky vibrating motors.

3 Results and Discussion

The system functions dedicated to the therapist are related to patient configuration (search, add and edit), session configurations (patient profile, session content and length, selecting the devices used by each exercise and their configuration), session supervision (through graphs that represent the essential parameters regarding the session in real time) and session history [10], Fig. 2a.

The doctors in the TRAVEE project consortium selected the available exercises and included the most common rehabilitation exercises. These include the Finger Flexion-Extension, the Palm Flexion-Extension, the Forearm Flexion-Extension and the Arm Adduction-Abduction movements. For each selected exercise, the therapist must configure the exercise (the number of repetitions, the duration of repetition and body side left or right), add support (Visual Augmentation, Vibrations, FES), and add monitoring devices (BCI, glove, motion sensor, kinect, leap motion). Every option

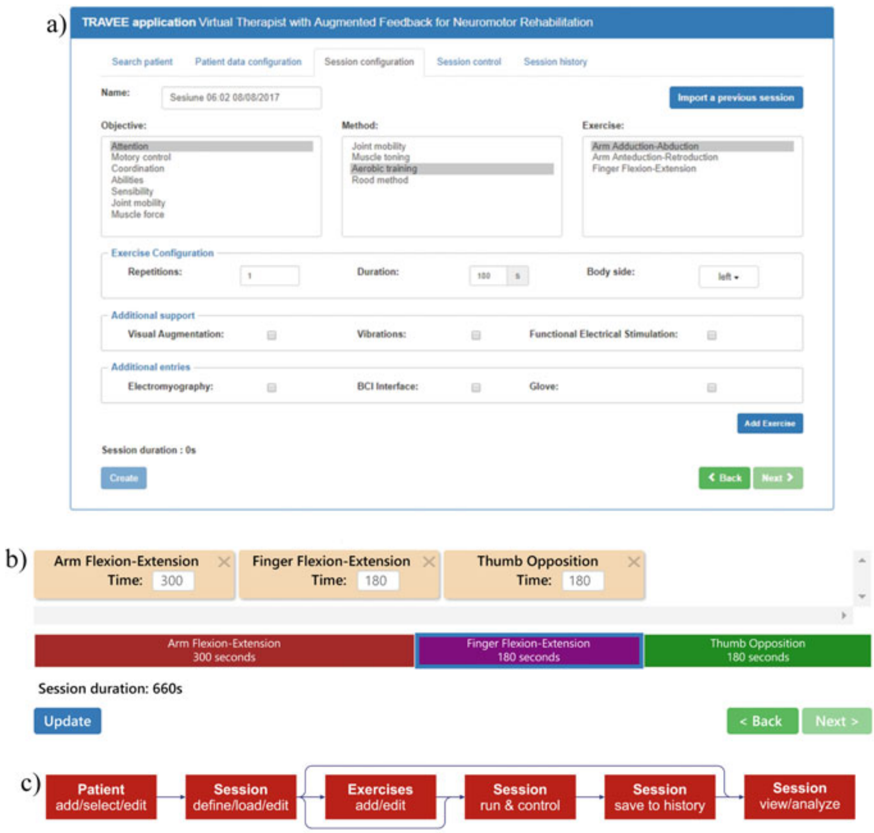


Fig. 2 Patient, exercise and session configuration/control

Fig. 3 Virtual environment (patient and world view): **a** patient facing the therapist, **b** patient and therapist facing a mirror, **c** serious game

and potential addition is on the bottom of the session configuration page, Fig. 2b and the flowchart is briefly presented in Fig. 2c. The therapist may choose to edit, run or view/analyse a saved session.

The main features of the system are dedicated to the patient—the subject of the upper limb rehabilitation process. Figure 3 shows that the patient is immersed in a Virtual Environment that includes two avatars (both 3D humanoids). One avatar represents the therapist, which demonstrates the movement that the patient needs to try to reproduce in the real environment. The second avatar represents the user from a first-person point of view that mimics the real-life movements of the patient. The patient may face the therapist Fig. 3a, c or sit next to him/her, both facing a mirror like in a dance room.

There are two minimum recommended configurations: the so called BCI-FES and motion sensor configurations to which other devices can be added. The first configuration consists of a 16 channels biosignal amplifier g.USBamp and an 8-channel neurostimulator Motionstim8. The 12/16 acquired EEG signals are collected from the sensorimotor areas according to the 10–20 International System. The number of EEG signals may vary because four channels may be used, in differential mode, to acquire EOG signals to determine whether the patient is paying attention to the virtual therapist. The 256 Hz sampled EEG signals are preprocessed (filtered with 50 Hz notch filter and 8–30 Hz band pass filter), fed to an algorithm to execute Common Spatial Patterns (CSP) [13–15] spatial filtering, and classify the output as left or right hand movement with Linear Discriminant Analysis (LDA) [16]. For CSP and LDA, the class Common Spatial Patterns 2 from BCI MATLAB&Simulink model provided by g.tec have been used together with g.BSanalyze software (g.tec) for offline data analysis.

As for EOG, the 256 Hz sampled signals are filtered with a moving average filter of 128 samples and then fed to a Simulink block that contains a custom developed algorithm for EOG signal processing. The output of the algorithm is the x-y (HEOG—VEOG) gaze normalized coordinates (Fig. 4) and the number of trigonometric quadrants or centre of the image where the patient is looking on the VR glasses. This is needed to determine whether the patient is dozing off or otherwise ignoring the virtual therapist. If so, the system warns the patient to concentrate/focus on the exercise and pay attention to the virtual therapist.

For the BCI-FES configuration to provide VR feedback based on the patient's imagined movement, the system needs to create a set of spatial filters and classifiers. This is done by recording 4 runs of training data with 20 left and 20 right motor imagery trials in random order [13]. Each 8-second trial consists of:

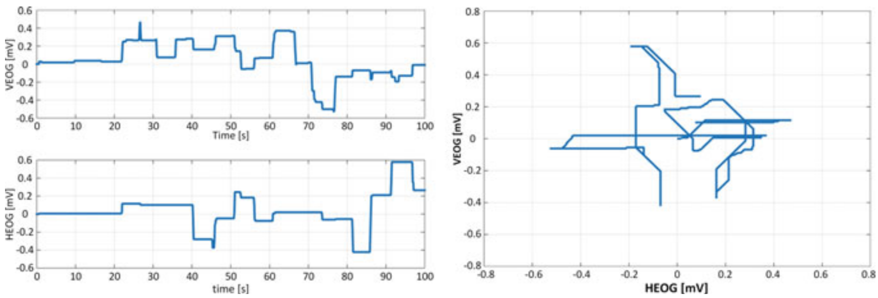


Fig. 4 Eye tracking: **a** HEOG and VEOG, **b** gaze position

(1) A beep at second 2 informing the patient about the upcoming cue;
(2) The cue to perform left or right motor imagery, which is presented from second 3 until the end (second 8) through both audio (left or right) and video (left or right red arrow and left or right therapist hand movement) and indicates that the patient needs to start imagining the corresponding movement; and
(3) Visual feedback (in form of the patient avatar moving its hand) begins at second 4.25. At the same time, the neurostimulator starts to trigger the patient's hand movement corresponding to the virtual therapist's cue.

The first two runs are used to build the spatial filters and classifiers. For the following two runs, each sample classification result is compared with the presented cue to calculate the error rate for that session as follows:

$$\text{Err} = \left(1 - \left(\frac{\text{Tcc}}{N}\right)\right) \cdot 100$$

where N represents the number of trials and Tcc the number of trials correctly classified. From the obtained array of 40 error values in the feedback phase, the mean and minimum error are obtained.

In Fig. 5, an example of output of the LDA classifier can be seen during feedback phase. Each trial classification output is represented with dotted lines (right-blue, left-green) and the corresponding average classification output with the solid lines.

Table 1 shows that the mean and minimum classification errors in the feedback phase are smaller with VR than during the session in which the patient instead received the visual feedback from a screen. This is because the patient was more cognitively involved during the exercise using VR, since the VR environment shielded him from real-world distractions and, in the VR environment, he is no longer a disabled person. Table 1 contains the mean and minimum classification errors for seven subjects. The first four subjects (S1–S4) received the visual feedback on a screen in front of them and the following three subjects (S5–S7) received the visual feedback through VR glasses.

For the second minimum recommended configuration—which use motion sensors—a device like Kinect (V1/V2), LeapMotion, video cam and ArUko markers,

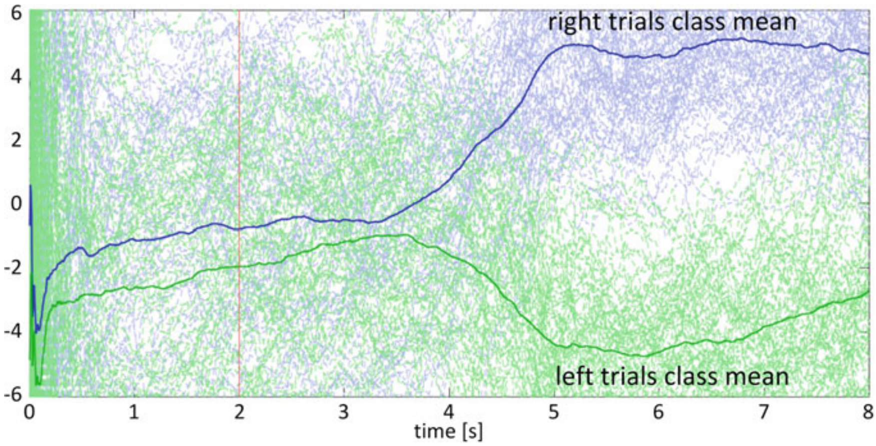


Fig. 5 LDA classification output

Table 1 Mean and minimum classification errors in the feedback phase without (left) and with (right) VR

Subject	Session	Mean Err (%)	Min Err (%)
S1	1	20.62	5.48
	2	22.34	7.11
	3	26.48	19.7
S2	1	23.96	11.97
	2	24.6	14.1
	3	28.83	21.1
S3	1	33.56	22.78
	2	37	21.35
	3	35.58	29.51
S4	1	32.58	24.77
	2	31.54	24.61
	3	37.21	26.22
Mean values	29.53	19.06	
S5	1	18.5	7.36
	2	19.72	10.72
	3	20.8	9.45
S6	1	19.2	6.37
	2	19.25	7.68
	3	19.58	1.95
S7	1	28.19	15
	2	25.53	13.56
	3	21.91	5.13
Mean values	21.41	8.58	

and IMU is used to monitor the patient's hand/arm movement. The neuromotor rehabilitation is divided in three session types: "mirror;" augmented and magnified feedback; and real feedback. All sessions relied on VR glasses to immerse the patient in a virtual environment to receive professional guidance, encouragement, feedback and motivation. The mirror session type was designed to be used immediately after the stroke or brain surgery, when the patient is not able to move the impaired arm or hand. The patient is told to imagine/try/execute the exercise with both arms/hands. The system tracking sensors are set to track only the healthy arm or hand but update both arms/hands of the patient avatar in VR. This way, the patient can see his both arms/hands working just like the therapist instructed, and realizes that s/he can move the hand/arm at will. This visual feedback is very important because it activates the mirror neurons that intermediate learning and closes the causal chain specific to recovery.

If the patient can perform small hand movements of the impaired arm, enough to be detected by the motion sensors, then all executed movements are augmented (session type two). The patient can see a much larger movement than s/he actually executes. The amplification factor decreases from a maximum set value (when the patient's movements are barely detected) to the value of one (when the movement is complete and correctly executed). The augmented feedback transforms the received visual information into knowledge.

The exercises of the third session type should be used after the patient regains partial/total control over the impaired arm and needs motivation to continue therapy by proposing different scenarios and tasks. The difficulties and challenges of the exercises can be adjusted to each patient's condition and progress.

To remind the user about the system and its benefits, each recovery session starts with the therapist and patient sitting in a chair facing each other. The real therapist explains to the patient what s/he will see, hear and must do. Specifically, the virtual therapist will move its left/right hand/arm to demonstrate the exercise to the patient. At the same time, an arrow will appear on the left/right side of the screen, followed by a corresponding left or right audio cue. The patient is instructed to imagine the left or right motor act and do his/her best to execute it. If BCI is used, the patient will receive visual feedback only while s/he is correctly imagining that movement. After this explanation, the VR glasses are mounted and the recovery exercises may begin (see Fig. 6).

Because the prevalence of post stroke spasticity is around 38% [17], some patients with hand spasticity need to perform special exercises to reduce spasticity with a therapist's help [18] before using the TRAVEE. To meet these needs, a second working group on stroke recovery from Technical University of Iasi led by Prof. Poboroniuc designed, build and added a module to TRAVEE to be used especially for despasticisation as well as for recovery exercises for the upper limbs. It consists of a distal exoskeleton glove that can copy the finger movements of the healthy hand by using another glove equipped with bending sensors. It can also actively assist flexion/extension movements of all fingers or each individual finger. The module uses an FES system for better and faster results. This hybrid approach can replace the recovery therapist who usually assists the FES induced movements and can

Fig. 6 Patients using the system

copy the movements of the healthy hand. This mirror-like therapy induces cortical reorganisation and motivates the patient.

The glove (left hand) is made from leather with tendons (metal wires) clamped on the top (dorsal side) and bottom (root) of each finger, as shown in Fig. 7. The right hand is using a textile/leather glove with bending sensor insertion for each finger. Figure 8 presents the hardware architecture of the despasticisation module. The FES module consists of a MotionStim8 neurostimulator that uses two channels for stimulating both the interosseous and extensor digitorum muscles.

The module is not used just to reduce spasticity. It is also integrated in the TRAVEE system, where the therapist can select it as stimulation device and/or as motion sensor based on the type of exercise.

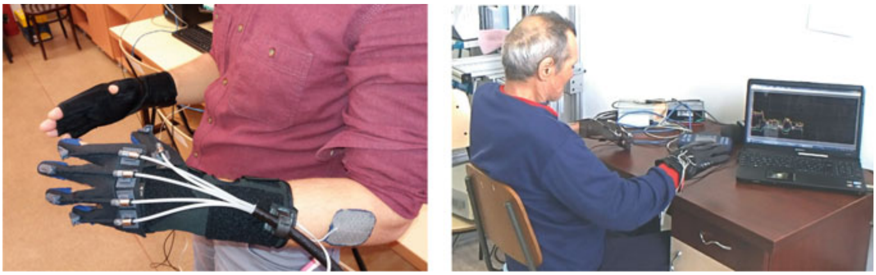


Fig. 7 The distal exoskeleton glove (left hand) and bending sensor glove (right hand)

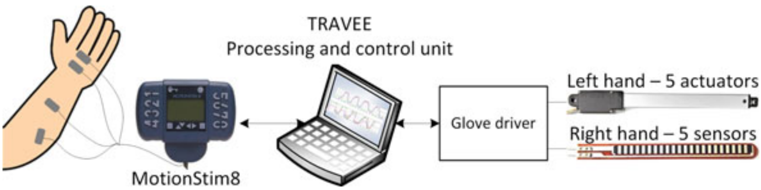


Fig. 8 The hardware architecture of the despasticisation module

The whole system was first tested on three healthy people. Next, we performed some fine tuning based on their suggestions to improve accuracy and validate system repeatability. Each patient signed an informed consent and an authorization for videos and photographs. The experiments with patients were approved by the institutional review board of the National Institute of Rehabilitation, Physical Medicine and Balneoclimatology from Bucharest, Romania. Patients were women and men with ages between 52 and 79, with post stroke central neuromotor syndrome and stable neurological status, stable consciousness, state, sufficient cognitive functions to allow learning, communication ability, and sufficient physical exercise tolerance.

4 Conclusions

This work presents a BCI-FES system for stroke rehabilitation with the unique combination of BCI and EOG devices to supervise how exercises are performed and monitor patient commitment. The Oculus rift headset increases the patient's immersion in VR. The system must be seen as a software kernel that allows users to define/run a series of rehabilitation exercises using a series of "plugin" devices. By using VR, the patient is not distracted by the real environment and is more cognitively involved during recovery exercises. The patient is focused most of the time, but if s/he loses concentration, the eye tracking system detects this problem and provides a warning. For the BCI-FES configuration, the use of VR makes it possible to provide neurofeedback in one or (rarely) two training sessions.

To our knowledge, the proposed neuromotor recovery system is the only one that includes an eye-tracking device for assessing patient concentration during exercises, enhancing engagement and effectiveness.

Technical performance was validated by testing the system on healthy persons with good knowledge in assistive technologies. The healthy people achieved low control error rates relative to those reported in the literature.

There are two patents pending:

- System, method and software application for automated augmented, gradual and naturalistic representation of human movements 00814/2017, OSIM patent pending.
- Mechatronic glove-neuroprosthesis hybrid system with knitted textile electrodes for hand rehabilitation for patients with neuromotor disabilities 00072/2017, OSIM patent pending.

Acknowledgements This work was supported by the Romanian National Authority for Scientific Research (UEFISCDI), Project 1/2014 Virtual Therapist with Augmented Feedback for Neuromotor Recovery (TRAVEE).

References

1. ENIGMA-Stroke Recovery. http://enigma.ini.usc.edu/ongoing/enigma-stroke-recovery/. Last visit October 2018
2. Johns Hopkins Institute—Strock Centers. www.hopkinsmedicine.org/neurology_neurosurgery/centers_clinics/cerebrovascular/stroke/. Last visit December 2018
3. StrokeBack Project. http://www.strokeback.eu/project.html. Last visit December 2018
4. NIHR—A practical, yet flexible functional electrical stimulation system for upper limb functional rehabilitation, Centres for Health Sciences Research, 2014–2017. https://www.salford.ac.uk/research/health-sciences/research-groups/human-movement-technologies/a-practical,-yet-flexible-functional-electrical-stimulation-system-for-upper-limb-functional-rehabilitation. Last visit December 2018
5. RETRAINER. http://www.retrainer.eu/start/. Last visit December 2018
6. C.M. McCrimmon, C.E. King, P.T. Wang, S.C. Cramer, Z. Nenadic, A.H. Do, Brain-controlled functional electrical stimulation for lower-limb motor recovery in stroke survivors, in *36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 1247–1250, 2014
7. M. Sun, C. Smith, D. Howard, L. Kenney, H. Luckie, K. Waring, P. Taylor, E. Merson, S. Finn, FES-UPP: a flexible functional electrical stimulation system to support upper limb functional activity practice. Front Neurosci. **12**, 449 (2018)
8. O. Ferche, A. Moldoveanu, F. Moldoveanu, The TRAVEE system for neuromotor recovery: Architecture and implementation, in *2017 E-Health and Bioengineering Conference (EHB), Sinaia*, 2017, pp. 575–578. https://doi.org/10.1109/EHB.2017.7995489
9. S. Caraiman, A. Stan, N. Botezatu, P. Herghelegiu, R.G. Lupu, A. Moldoveanu, Architectural design of a real-time augmented feedback system for neuromotor rehabilitation, in *2015 20th International Conference on Control Systems and Computer Science*, Bucharest, 2015, pp. 850–855. https://doi.org/10.1109/cscs.2015.106
10. R.G. Lupu et al., Virtual reality system for stroke recovery for upper limbs using ArUco markers, in *2017 21st International Conference on System Theory, Control and Computing (ICSTCC)*, Sinaia, 2017, pp. 548–552, https://doi.org/10.1109/icstcc.2017.8107092
11. R.G. Lupu, N. Botezatu, F. Ungureanu, D. Ignat, A. Moldoveanu, Virtual reality based stroke recovery for upper limbs using Leap Motion, in *2016 20th International Conference on System Theory, Control and Computing (ICSTCC)*, Sinaia, 2016, pp. 295–299. https://doi.org/10.1109/icstcc.2016.7790681
12. R.G. Lupu, D.C. Irimia, F. Ungureanu, M.S. Poboroniuc, A. Moldoveanu, BCI and FES based therapy for stroke rehabilitation using VR facilities. Wireless Commun. Mob. Comput. (2018)
13. D.C. Irimia, M.S. Poboroniuc, R. Ortner, B.Z. Allison, C. Guger, Preliminary results of testing a BCI-controlled FES system for post-stroke rehabilitation, in *Proceedings of the 7th Graz Brain-Computer Interface Conference 2017*, September 18th–22nd, Graz, Austria, 2017
14. D.C. Irimia, R. Ortner, M.S. Poboroniuc, B.E. Ignat, C. Guger, High classification accuracy of a motor imagery based brain-computer interface for stroke rehabilitation training. Front. Rob. AI **5**, 130 (2018)
15. S. Lemm, B. Blankertz, T. Dickhaus, K.-R. Müller, Introduction to machine learning for brain imaging. NeuroImage **56**(2), 387–399 (2011)
16. J. Müller-Gerking, G. Pfurtscheller, H. Flyvbjerg, Designing optimal spatial filters for single-trial EEG classification in a movement task. Clin. Neurophysiol. **110**(5), 787–798 (1999)
17. C.L. Watkins, M.J. Leathley, J.M. Gregson, A.P. Moore, T.L. Smith, A.K. Sharma, Prevalence of spasticity post stroke. Clin. Rehab. (2002). https://doi.org/10.1191/0269215502cr512oa
18. D.A. De Silva, N. Venketasubramanian, A. Jr. Roxas, L.P. Kee, Y. Lampl, Understanding Stroke—A Guide for Stroke Survivors and Their Families, 2014. http://www.moleac.com/ebook/Understanding_Stroke_-_Guide_for_Stroke_Survivors.pdf

Received October 10, 2018, accepted November 21, 2018, date of publication January 11, 2019, date of current version January 23, 2019.

Digital Object Identifier 10.1109/ACCESS.2018.2886271

The TRAVEE System for a Multimodal Neuromotor Rehabilitation

ALIN MOLDOVEANU[1], OANA-MARIA FERCHE[1], FLORICA MOLDOVEANU[1], ROBERT GABRIEL LUPU[2], DELIA CINTEZĂ[3], DANUT CONSTANTIN IRIMIA[2], AND CORNELIU TOADER[4]

[1]Department of Computers, Faculty of Automated Control and Computers, Politehnica University of Bucharest, 060042 Bucharest, Romania
[2]Faculty of Automatic Control and Computer Engineering, Gheorghe Asachi Technical University of Iași, 700050 Iași, Romania
[3]Department of Rehabilitation and Physical Medicine, Faculty of Medicine, Carol Davila University of Medicine and Pharmacy, 020022 Bucharest, Romania
[4]Neurosurgical Department, National Institute of Neurology and Neurovascular Diseases, 077160 Bucharest, Romania.

Corresponding author: Alin Moldoveanu (alin.moldoveanu@cs.pub.ro)

This work was supported by the TRAVEE Grant of the Romanian Executive Agency for Higher Education, Research, Development, and Innovation Funding—UEFISCDI, Joint Applied Research Projects Program, under Grant 1/2014(PN-II-PT-PCCA-2013-4-1580).

ABSTRACT As more and more people are left disabled by stroke each year, it is of vital importance to progress in the research of new ways to improve their condition and to ensure that they maintain their independence as much as possible in everyday life. A step in this direction of research was taken with TRAVEE, a system dedicated to neuromotor rehabilitation after stroke. To reach this goal, the TRAVEE has benefited from several innovative ideas and technologies—virtual reality, brain–computer interfaces, functional electrical stimulation, robotics, haptics, multimodal feedback, and a novel idea in information and communications technology systems for rehabilitation—visual augmentation as a form of feedback to the patient. Through visual augmentation, the TRAVEE immerses the patient in a virtual environment where his movements are rendered as being better than in the real world, and in this way diminishing his disability. We believe that this process—that is pending for patent—will greatly impact the recovery process after stroke, by providing more motivating sessions, while supporting the cortical reorganization process. This paper presents an overview of the TRAVEE system, the perspectives that supported it, details regarding its development, as well as the results of the clinical tests that were performed with the system.

INDEX TERMS Multimodal feedback, neuromotor rehabilitation, virtual reality, visual augmentation.

I. INTRODUCTION

According to the World Health Report [1], stroke affects 15 million annually. Out of them, a third die and a third are left with permanent disability.

According to the Heart Disease and Stroke Statistics 2018 [2] provided by the American Heart Association, stroke is a leading cause of disability in the United States. Approximately 90 million Americans are estimated to be living with a cardiovascular disease (CVD) or an aftereffect of stroke. Increasing the quality of life of those affected by stroke can therefore have a significant impact worldwide.

The TRAVEE system is the result of a national research project, undergone between 2014-2017. It is a neuromotor rehabilitation system for the upper limb, that took the first steps toward developing a low-cost solution that could be used on a large scale in the rehabilitation process.

The system combined multiple technologies (VR, BCI, FES, a robotic hand assistant device and haptic feedback), as well as complex ideas such as virtual therapist (VT), visual augmentation and multimodal feedback to develop a low cost, highly customizable rehabilitation solution. The resulting system has multiple functioning modes, a graphical user interface (GUI) dedicated to a non-technical healthcare practitioner and a database for storing the information regarding the patients.

One of the purposes of this project was to validate two ideas: the visual augmentation process in VR (transmitting to the patient an improved visual representation of his actual movements) and the eficacity of the virtual therapist, a virtual avatar that executes the movements that the patient must try to reproduce during his rehabilitation session.

TRAVEE was tested in iterations. The initial prototype was tested during two in-vivo testing sessions, in a medical settlement, in order to validate the technical solution. After refining the initial prototype into the final one, a clinical trial took place, in the same medical settlement, to qualitatively assess the final result of the project.

This paper presents the medical prerogatives that were used and supported the ideas of the TRAVEE project, the overview of the system functionality and its architecture as well as several technical implementation details. The article will also provide the results of the preliminary in-vivo tests and the clinical trial, along with their interpretations, conclusions and future development perspectives.

II. MEDICAL BACKGROUND AND PERSPECTIVES

A. STROKE AND REHABILITATION

Stroke is the main cause of adult disability; approximately 60% of survivors remain with dysfunctional sequelae, especially at the upper limb.

Rehabilitation therapy allows people with disabilities and activity limitations to gain and maintain optimal physical, intellectual, psychological and / or social functioning. It includes a broad and heterogeneous range of activities, therapeutic interventions and methodologies, in addition to standard medical care.

Over the past 15 years, significant scientific evidence has emerged that argue that intense and repeated training can influence the reorganization of the brain through the acquisition / revival of motor regimes. The learning of motor engram is done through internal processes associated with practice and experience, which leads to changes in the ability to move.

B. NEUROLOGIC PERSPECTIVES ON VISUAL AUGMENTATION

Neuroplasticity is the ability of the brain to undergo functional changes in the short term and also to undergo structural changes in the long-term to adapt to changes in the living environment, central or peripheral injuries, aging phenomena. Brain reorganization is the main mechanism for achieving neuroplasticity. The stimulation of brain reorganization is done by: enriching the environment, stimulating attention, social interactions, tactile stimulation, motor re-learning, direct brain stimulation.

The cortical reorganization for restoring the movement of the hand affected by stroke is done on three ways, which are not excluding one another:

1. Bilateral cortical activation, with significant recruitment of nerve networks in the unaffected hemisphere.

2. Increasing recruitment in secondary cortical areas in the affected hemisphere.

3. Recruiting nerve paths around the infarcted area.

A potential role in reorganization is the use of feedback (augmented or not) as a way to stimulate the reward mechanism underlying the learning process. The use of imagination or visual representations of movement is called motor imagery. According to an extensive study in the field of motor imagery in rehabilitation [3] there is at least theoretical and experimental proofs on healthy subjects for the support of this idea.

A study regarding the possibility of 'fooling' the brain into believing that the perceived improved feedback is the result of the motor action of the body was published in [4]. This paper presented the presence of techniques for fooling the brain in rehabilitation purposes starting from 1996 with Virtual Reality Box and Mirror Therapy [5], [6], both using mirrors to reflect the movements of the healthy hand in upper limb amputees to simulate the presence of the missing limb, in order to successfully alleviate or treat phantom pain, to Functional Electrical Stimulation consisting in application of electric currents on the missing limb also in the purpose of relieving phantom pain.

Other experiments also presented in this survey [4] include the use of Augmented Reality (AR) to amplify a small movement in order to trick the brain into thinking it was a wider, more ample one in the TheraMem system [7]. This system was tested on five patients and observed a high degree of motivation during the sessions with the system.

Another system that implements this idea is a Virtual Reality (VR) for "corrective learning" where small movements of the disabled arm generate full range movements in the VR to help the patient re-learn the given action by correcting the perceived feedback [8]. The system referred by [8] is called VirHab [9] and it augments movements by using image processing of video streams to replace the image of the disabilitated arm with a recording of a movement of the healthy one when an input device is actioned - a small ranged movement determines the visualization of a full range one. A similar system is presented in [10] and the presented study showed improvements in the involved patients on several disability scales that were maintained even after three months after the sessions with the system.

As there are previous researches that tested the feasibility of stimulating cortical plasticity by fooling the brain by providing virtual improved feedback, TRAVEE introduced the augmented feedback - tracking the body of the patient and displaying on the patient avatar in the virtual environment a slightly improved version of the detected movements, combined with multimodal feedback.

III. OVERVIEW OF PREVIOUS ICT REHABILITATION SOLUTIONS

Starting from a survey that was presented at the 8th International Conference on Speech Technology and Human-Computer Dialogue [11], we evaluated the existing literature regarding ICT systems for neuromotor rehabilitation. We observed two tendencies in the development of these kind of systems. Either the experiments used a unique technology, developed exclusively for the study, or they involved the use of commercially available solutions, in the aim of developing a more accessible system.

In the category of systems dedicated exclusively to rehabilitation, several experiments were presented, and will be mentioned in the following. The Rutgers Arm system [12] includes a forearm support that slides on a surface, to assist the patient in performing the movements necessary in a pick-and-place game that exercises the ability of following a given trajectory or a treasure-hunt game that tests arm endurance. The system also tested a game designed to exercise grasping gestures. The two subjects participating in the experiments with the system showed improvements in motor abilities and pinch and shoulder strength. The follower of the Rutgers Arm system is the Bright Arm [13] where the training table was completed with a rubber pear for monitoring the grasp strength in the palm, two infrared cameras placed above the head of the patient for movement tracking, a display and a computer connected to a remote medical server. Five games were available in this version of the system, and it was tested with 5 participants that – after the experiment – improved their shoulder strength, grasp strength, shoulder and elbow flexion and extension capabilities.

Another system that enhances the rehabilitation sessions using dedicated ICT is the ImAble [14], with its three configurations, all dedicated to rehabilitation using virtual games. The Able-B supports the disabled hand against gravity and moves it with the support of the healthy one. It uses a webcam to track the movements of the disabled hand by detecting a colored patch placed on it. The Able-M contains a sliding device to which the hand is strapped while sliding on a table and controlling a mouse for finger strength training. The Able-X consists of a lightweight handlebar that can be rotated in transversal and sagittal plane to control the movements of a pointer on screen. These systems are integrated with various games for static or dynamic target hitting. The three configurations (Able-B, Able-M and Able-X) were tested with five, three and 14 subjects respectively and in all cases improvements on the Fugl-Meyer scale were observed.

One of the systems that use commercially available solutions for rehabilitation is the Gertner Tele-Motion Rehabilitation System [15] that uses the Kinect to detect the movements of the patient. The patient performs certain rehabilitation movements that are translated to actions in specially designed video games. This system was tested on 18 subjects, 9 in the test group and 9 in the control group. Greater improvements were detected in the test group post-sessions, but a larger test is required for a definite result. The ioTracker is another system that uses Kinect to track the body movements of the patient as a form of input.

Other commercially available devices used in ICT systems for rehabilitation is the Wii. It was used for vestibular rehabilitation [16] in which over 50% of the 17 participants improved their balances indexes after the sessions.

Several studies [17], [18] used head mounted displays to immerse patients in virtual environments, with positive results on experiments with several patients, in the improvement of conditions such as memory and attention deficits: in [17] two patients were involved in ten sessions each with the system and in [18] the patients were evaluated using scales for attention deficit before and after using the system that immersed them in real-life scenarios, such as finding paths to certain destinations or memorizing information from the virtual world. The results showed improvements on both the Wechesler Memory Scale and on the Toulouse-Pieron scale.

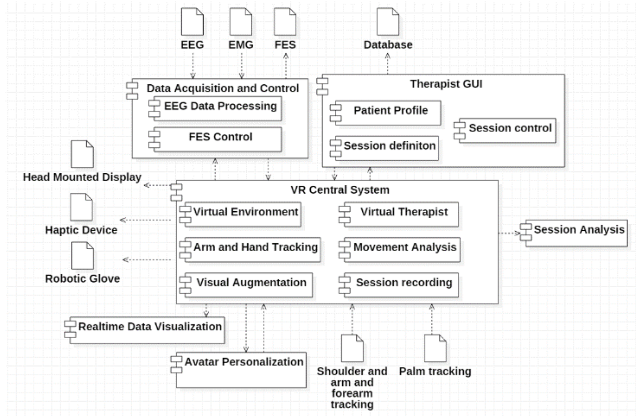


FIGURE 1. TRAVEE architecture overview.

The studied literature presents experiments in various fields of rehabilitation using ICT systems and most of them seem to have a positive influence on the rehabilitation procedures. TRAVEE is a complex system that combines several of the ideas that are already present in the existing literature with novel ideas, such as the visual augmentation, virtual therapist and multimodal feedback, using various technologies that are commercially available (Kinect, Leap Motion, Oculus Rift) or devices that are designed especially for the system (robotic glove, haptic device) as the system wants to evolve towards a low-cost solution. Several of the used technologies (for EEG, FES and EMG) are at the moment not low cost, but the desired evolution of the system is to replace them with accessible solutions at a satisfactory quality.

IV. TRAVEE NEUROMOTOR REHABILITATION SYSTEM - FUNCTIONALITY, ARCHITECTURE AND IMPLEMENTATION DETAILS

The system implements many original ideas, some original by themselves, others original in the context they were used. These are: the virtual therapist, that exemplifies the correct movement to the patient; the multimodal input, consisting of body tracking, EEG and EMG; multimodal feedback to the patient and visual augmentation of the patient's actions (an idea that is pending for patent).

The system integrates a variety of functioning modes in a modular architecture, presented in the image below.

The main components of the TRAVEE system are: the VR Central System, the Data Acquisition and Control component, the Therapist GUI, the Movement Analysis component, the Realtime Data Visualization component and the Avatar Personalization module. The rehabilitation sessions are recorded by the VR Central System. The resulted recordings are analysed using a standalone application, the Session Analysis component.

A. BIOPHYSICAL INPUTS

The system accepts input data from different devices, depending on the functioning mode: body tracking, body tracking + brain activity monitoring, detection of muscle activation in the limb. To stimulate the patient, the system generates many types of feedback: visual (true or augmented) through immersion in a virtual environment (VE), FES (Functional Electrical Stimulation), vibrations (haptic) and robotics.

B. BODY TRACKING

The tracking of the patient body is made using optical tracking devices: Kinect and Leap Motion. These devices are used to obtain information regarding the positions, rotations and scales of the main joints in the arm, forearm and palm of the user.

C. BCI

In the traditional therapy, the patients are asked to try to execute a certain movement with their impaired limb while they are imaging that movement. The goal is to perform a corresponding motor imagery (MI) task in order to produce a correct neural activation. The visual feedback of that action is obtained by using rope and pulley, if possible, a FES device to stimulate the corresponding muscles or a robotic device. In all cases the patient or the therapist are pulling the rope, trigger the FES or robotic device while the patients are imaging that movement. The problem is that for the patients it is very difficult to "see" that their impaired limb is moving because they are imaging so and not just because they or someone else is pulling the rope or pushing the button. This is the reason for which the causal loop cannot be closed and the recovery is blocked. On the other hand, the therapists don't have a real feedback from patients and they must rely on patients that they are really imagining that movement and carry on with the therapy. In reality, most of the patients, after a short time, lose concentration, they are getting bored, they start to think at something else like personal problems or even fall asleep. TRAVEE uses the BCI technology to determine if the patient is correctly performing the MI task. That can be used to trigger the FES, the robotic device or to update the patient avatar in the VE and to receive a corresponding feedback. Also, the therapist can have a feedback regarding the patient's mental activity and guide him in order to sustain and/or maximize this activity.

D. EMG

In case of the patients with residual motor potential or for those that start to have some minor muscle activity or to gain a small control over their limb due recovering therapy, electromyography (EMG) can be used as an alternative to detect the patient intention to make a movement. This is done by acquiring the EMG signal(s) and compare their amplitude(s) with a threshold. If it exceeds the threshold the patient intention is detected and can be used as a trigger signal for devices that guide/helps the patient to perform that movement and/or to update the patient avatar in VR.

FIGURE 2. Capture from the VE of the TRAVEE system.

E. FEEDBACK MODALITIES

1) VISUAL FEEDBACK (AUGMENTED OR DIRECT) THROUGH IMMERSION IN A VIRTUAL ENVIRONMENT (VE)

The visual feedback provided to the patient is obtained by immersing him or her in a VE where the patient sees the Virtual Therapist (VT) - an avatar that executes the current movement that the patient must try to perform, as well as an avatar of the patient (virtual representation of his or her body). The patient's avatar performs the movements of the patient either exactly as they are detected by the body tracking devices, either augmented - to be closer to the Virtual Therapist movements - before being applied to the patient's avatar.

2) VIBRATIONS (HAPTIC)

This feedback form consists in applying vibrations to certain key points on the hand or arm of the patient to inform him or her that the movement was sufficiently executed. The used haptic device was custom made for TRAVEE and it consists of vibrating motors attached to electrodes that are placed on the skin. The device is controlled by the system with commands that start and stop the application of vibrations.

3) LIGHTWEIGHT ROBOTICS

Robotics are an important feedback path and are represented by a glove actuated by five motors that support the extension of the fingers and hand. The device tested in the clinical setting was developed specifically for TRAVEE and includes five medium servo motors attached to a glove, controlled by an Arduino Mega 2560 development board.

4) FES

The functional electrical stimulation (FES) is a technique often used for recovering neuromotor functions in neuromuscular disabilities due to a central nervous system lesion. By artificially inducing a pulse train in muscle nerves, contractions of the respective muscles can be obtained in proportion to certain parameters of the stimulation signal. Thus, by modifying the stimulation signal parameters, intense muscle contractions can be induced to produce functional movements. The main requirement for electrical stimulation to produce the contraction of the target muscle is that both the muscle and the nerve that connects it with the spine must be intact. In the TRAVEE project FES is used to help the patient to perform the desired movement and/or to maintain the muscular tonus, reduce the spasticity, maintain the limb joints. A side effect of working with FES is that the electrical impulse is travelling back to the brain via the nerve. This is seen as a benefit because the brain is bombarded with information and it is forced to reorganize in order to process it.

F. THE VR CENTRAL SYSTEM

This is the central communication point and also the system server. The VR Central System is responsible with the VE (using the Oculus Rift device) in which the patient is immersed and the main logic of the application. Based on the session configuration and on the available input data it decides what kind of augmentation or feedback should be applied and controls the augmentation and feedback modalities. It directly controls the haptic device and the robotic glove. The VR Central System is also in charge with logging relevant information regarding the current session, such as the postures of the patient obtained from the body tracking devices, as well as data acquired from EEG and EMG devices.

1) VIRTUAL ENVIRONMENT

The Virtual Reality environment was implemented using the Unity game engine version 5.3.4. It contains an avatar for the therapist (VT) and an avatar for the patient. In the VE, the patient sees the representation of their own body, the virtual patient avatar, from a first-person point of view, in order to better identify with the movements of this avatar, in a sitting position, with the VT also in a seated position in front of the patient's avatar, as in the capture below.

The immersion is achieved through a Head Mounted Display, Oculus Rift.

2) VIRTUAL THERAPIST

The VT is an avatar placed in front of the patient avatar, that exemplifies the movements that the patient needs to try to reproduce in the real world. The VT avatar was made using the Adobe Fuse CC software that allows creating humanoid characters. The patient avatars were made using the open source Make Human software.

3) SESSION RECORDING

The session recording functionality is integrated with the VR Central System, and it consists of a mechanism that stores all the relevant information for each session in a .session file:: avatar poses obtained from the body tracking devices, data synthesized from the EMG and EEG devices, exercise codes.

Having this information is enough to be able to use the session analysis component and simulate the entire rehabilitation session, by performing the same analysis on the logged data as in real time during the session. The session recorder is started automatically when a new session is created.

4) MOVEMENT ANALYSIS

This component is coupled with the VR Central System. It analyses the patient posture using the data from the optical tracking devices. Each movement is evaluated based on several predefined parameters and classified by a score, representing the degree of correct execution of the current movement. Decisions regarding body tracking based augmentation and feedback are taken by the VR Central System according to this score.

5) BODY TRACKING

The tracking of the arm and hand were made using the Kinect and Leap Motion devices. Both of them were necessary, as the Leap Motion tracks the forearm, the joints of the palm, and the phalanges of each finger, while the Kinect device tracks - among others - the joint of the shoulder and the elbow. Each movement defined in the TRAVEE system is tracked by one of these devices.

The movements implemented by TRAVEE and their classification as either being tracked by Kinect or Leap Motion is presented below.

Movements tracked by Kinect: Forearm flexion-extension, Arm adduction-abduction, Arm anteduction-retroduction, Shoulder raise.

Movements tracked by Leap Motion: Palm flexion-extension, Finger flexion-extension, Thumb opposition, Forearm pronation-supination.

More details regarding the implementation of the hand tracking with the two devices are presented in [19].

6) VISUAL AUGMENTATION

During the rehabilitation session execution, the data from the input devices – tracking devices, BCI, EMG – is analysed by the VR Central System and, depending on the functioning mode, the movement is visually augmented.

The visual augmentation of a movement based on tracking data within the TRAVEE system is the process through which, during the execution of a certain movement in a rehabilitation session, the movement detected by the tracking devices is improved before being applied to the virtual avatar of the patient. This means that the patient tries to execute correctly the current movement in the session – exemplified by the therapist avatar – and the movement the patient observed on the patient avatar will be a slightly improved version of the real movement, as detected by the tracking devices.

The visual augmentation of the movement based on movement tracking data uses the score calculated for the movement and a previously set threshold.

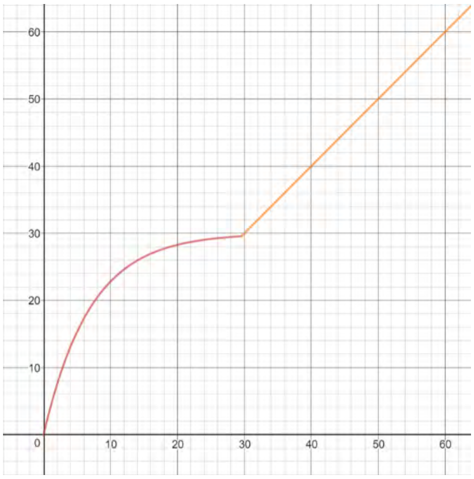


FIGURE 3. Graphical representation for augmentation function with threshold value 30.

The movement is augmented if the score is below the

$$\begin{cases} threshold \times (1 - e^{-\frac{x}{7}}), & x \leq threshold \times (1 - e^{-\frac{x}{7}}) \\ x, & x > threshold \times (1 - e^{-\frac{x}{7}}) \end{cases}$$

threshold with a factor. Several formulas of augmentation were tested, but the one we believe represents the envisioned visual augmentation of the TRAVEE system has the following form:

A graphical representation of this function, for a threshold value equal to 30 is presented in Fig. 3.

The augmentation algorithm is the following:

1) Evaluate the degree to which a movement was performed, based on the current body tracking data: for each movement we identified a joint or a set of joints that are most relevant and used them to calculate a degree of execution, referred to as score. The score is represented by a number, which is calculated differently for each movement, as it can represent a relevant angle or a distance between two bones or joints of the hand.
2) If the score is beneath the threshold for the currently executed movement, augment the relevant angles and distances of the movement according to the augmentation.
3) If the score is above the threshold, display on the patient avatar the pose obtained from the body tracking devices, without any alterations.

The process of visual augmentation based on body tracking is pending for a patent with the title: ''System, method and computer program for augmenting human movements''.

The joints used for the movements are presented in Fig. 4 and Fig. 5. As the system knows what the current exercise is, for each pose, it evaluates the current relevant angle or distance. This value is considered to be the score for the movement. Each type of movement has a predefined threshold.

The joints evaluated for each movement are presented in Figures 4 and 5.

FIGURE 4. The angles considered for evaluating the movement score for Forearm Flexion-Extension (top-left), Thumb Opposition(top-right), Arm Adduction-Abduction (bottom-left), Arm Anteduction-Retroduction (bottom-right).

FIGURE 5. The angles considered for evaluating the Palm flexion-extension (top-left), Forearm pronation-suppination (top-right), Finger flexion-extension (bottom-left) and Shoulder raise movements (bottom-right).

1) Forearm Flexion-Extension: the elbow joint was considered the most relevant. Therefore, the given score was the angle between the forearm and the arm. The maximum augmentation angle, up to which the movement was augmented, was set to 45 degrees.
2) Thumb Opposition: the angle between the direction of the first phalange of the thumb and the axis between

the base of the thumb and the base of the pinky fingers is considered the relevant angle. The maximum augmentation angle, up to which the movement was augmented, was set to 60 degrees.

3) Arm Adduction-Abduction: the angle between the direction of the arm and the direction of the spine is considered the relevant angle. The maximum augmentation angle, up to which the movement was augmented, was set to 60 degrees.
4) Arm Anteduction-Retroduction: the angle between the direction of the arm and the direction of the spine is considered the relevant angle. The maximum augmentation angle, up to which the movement was augmented, was set to 60 degrees.
5) Palm flexion-extension: the angle between the direction of the hand and the direction of the forearm is considered the relevant. The maximum augmentation angle, up to which the movement was augmented, was set to 45 degrees.
6) Forearm pronation-supination: the local Euler roll rotation angle of the forearm relative to the arm is considered the relevant. The maximum augmentation angle, up to which the movement was augmented, was set to 30 degrees.
7) Finger flexion-extension: all the angles between the phalanges of the fingers were analysed. The minimum angle between either two phalanges was chosen as the score for the movement. The maximum augmentation angle, up to which the movement was augmented, was set to 60 degrees.
8) Shoulder raise: this movement was more complicated to analyze as it did not have a relevant angle between two joints, so it was evaluated based on the distance between the position of the base of the neck and the position of the shoulder. The maximum distance up to which the augmentation was performed was defined at 0.6 units. The augmentation for this movement consisted in changing the position of the shoulder joint on the vertical axis with the calculated augmentation distance.

G. THE DATA ACQUISITION AND CONTROL COMPONENT

This component acquires several types of data from the patient, EEG and EMG and controls the FES.

1) EEG DATA PROCESSING

In the recent years, a series of scientific publications demonstrated that BCI (brain computer interface) and more precisely the ones based on motor imagery (MI) can stimulate the mirror neurons and induce neuroplasticity [20]–[22]. These evidences support the inclusion of BCI as an important tool for post-stroke recovery therapy to enhance the motor rehabilitation outcome. During the exercises where MI-based BCI is used, the patient is asked to imagine the movement of his hands in a random order. Motor imagery is a skill that must be learned by the patient during the so called "training phase". MI can be measured (real-time processing and classification of the EEG) and used to provide neurofeedback. The neurofeedback must be similar to the real motor activity that patient is asked to imagine [23]. The visual representation of the neurofeedback through the popular bar feedback (bFB) [24] or virtual reality (VR) [25] it is a very important component of the learning process because it actively involves the patient (meaning the patient's brain) in the task.

The MI based BCI assume that the exercises are performed with both hands. The method used to discriminate between the two imaginary tasks is Common Spatial Patterns (CSP). The method is based on the simultaneous diagonalization of two covariance matrices. Thus, the method allows to construct a new time series that maximizes the variance of the samples of a task, while minimizing the variance of samples of the other task. The matrices contains a set of spatial patterns, subject dependent, which provides information about the activity of a specific cortical area corresponding to imaging the movement of one of the hands. Given one projection matrix W, the decomposition of EEG signal for one trial X can be projected as:

$$Z = WX \tag{1}$$

where W^{-1} are sets of CSP models and are time-invariant EEG sources distributions [26]. After interpolation these CSP can be displayed as topographic maps [27].

Fig. 6 shows a set of CSP models for EEG recordings during MI for left and right hand which correspond to the firsts and respectively the lasts column of W^{-1}. The topographic distribution of these components correspond to expected contralateral activities of the sensorimotor rhythms induced by imagination of the movement. Another advantage of this method is that is not necessary the variances computation for the all n series. Müller-Gerking demonstrate that the optimal number of CSP models used to create a feature vectors is four, only first and last two rows of W [27]. The variance is calculated using a sliding window of T according to (2)

$$VAR_p = \sum_{t=1}^{T} (Z_{p(t)})^2 \tag{2}$$

where:

p – is the number of CSP filters ($p = 4$)

T – is the time window for which the variance is calculated ($T = 1.5$s)

To obtain the feature vectors the values are normalized and log (3)

$$f_p = log_{10}\left(\frac{VAR_p}{\sum_{p=1}^{4} VAR_p}\right) \tag{3}$$

In order to categorize a movement to be right or left hand a LDA (linear discriminant analysis) classifier is used based on the classification of the four feature vectors. The result of the LDA classifier is used as visual feedback for the patient, Fig.7.

The EEG signals where acquired using a g.USBamp 16 channels biosignal amplifier device from g.tec medical

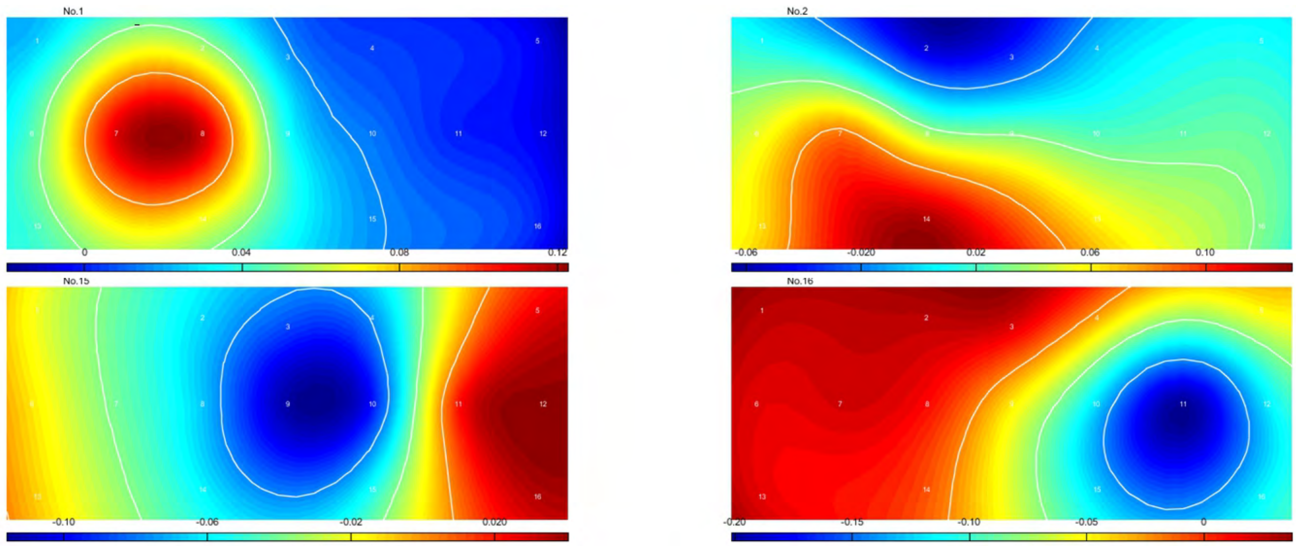


FIGURE 6. CSP over 16 channels for one of the patients during MI (left column – right hand, right column – left hand).

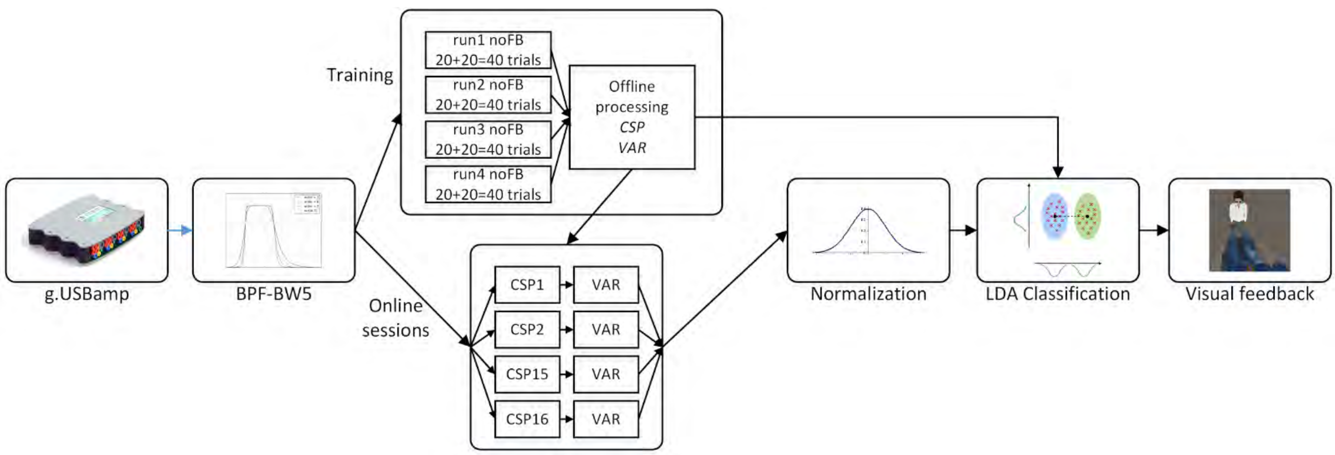


FIGURE 7. Workflow of BCI signal processing for visual feedback control.

engineering GmbH [28]. The electrodes are positioned on the EEG cap according to 10-20 International System in order to cover the sensorimotor areas of the brain, Fig. 8.

Before starting the recovery exercises there is a training session during which the patient must learn to imagine the movement. The session consists of 4 runs of 40 trials of hands movements, 20 for one hand and 20 for the other hand, in a randomized order without feedback. Each trial consists of 8 seconds of EEG recordings. At second 2, the patient hears a beep that informs him about the upcoming cue and at second 3 the cue (left or right) is presented, this representing the moment when the patient has to start imagining the movement. The feedback phase starts at second 4.25 and lasts till second 8. During the feedback phase, the patient has to imagine the movement of the hand dictated by the cue.

The training data recorded during the calibration phase is used to calculate the classifier that will be used for providing the feedback during the next phase.

After an online session, an error rate is calculated by comparing the cue presented to the patient with the classified movement at every sample time. For a number of N trials, the error rate is calculated as:

$$Err = \left(1 - \frac{Tcc}{N}\right) \cdot 100 \quad (4)$$

where Tcc is the number of correctly classified trials. The mean error rate and the minimal error are calculated during the feedback phase. Figure 9a presents the LDA classifier output for an online session. The dotted lines represent the output for each trial (blue for right and green for left) and the solid lines represent the averaged classification output for each class. Figure 9b presents as example the error rate for an online session, and the minimal error rate is marked with a red circle.

The system configuration using BCI to detect the patient intention to move is shown in fig. 10. The LDA output is

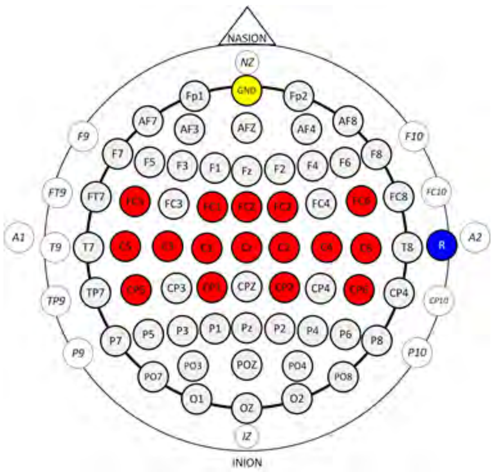


FIGURE 8. Position of the EEG electrodes according to 10-20 International System.

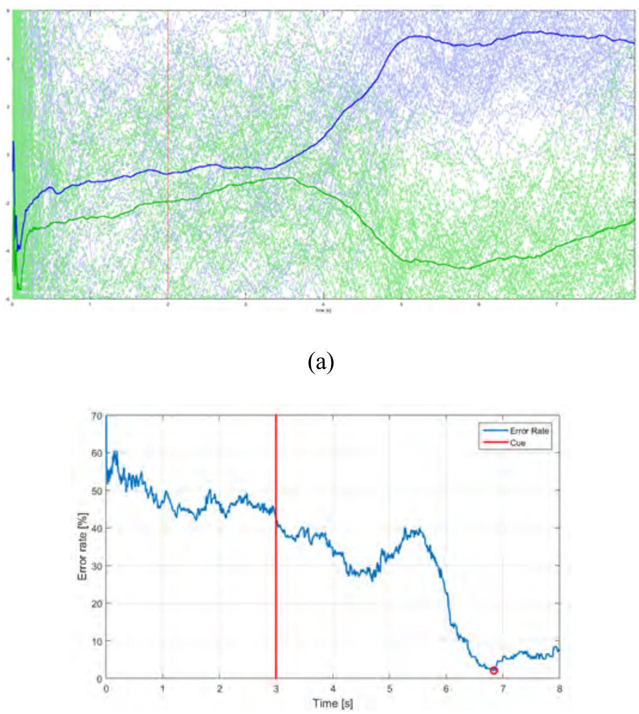


FIGURE 9. LDA classification output (using g.BSanalyze provided by g.tec medical engineering GmbH) and error rate.

used by the processing and control unit (PCU) to trigger the devices (robotic glove, robotic arm etc.) which helps the patients to perform the desired movement. At the same time it provides the patient with visual feedback he needs.

2) FES CONTROL

The FES is used in TRAVEE system to help the patient to perform the desired movement and as a technique to recover neuromotor functions by artificially inducing a pulse train in muscle nerves. Contractions of the respective muscles

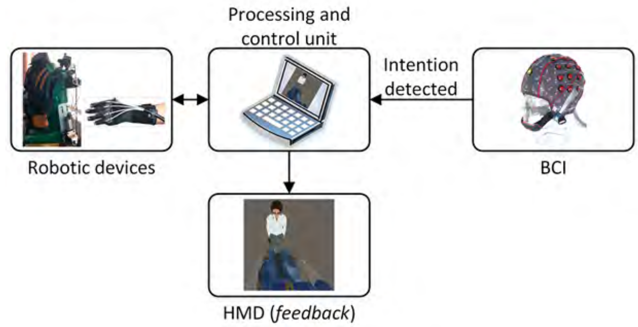


FIGURE 10. TRAVEE system configuration using BCI to detect the patient intention to move.

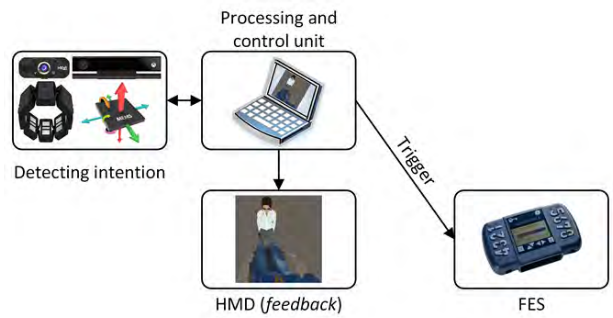


FIGURE 11. TRAVEE system configuration using BCI to detect the patient intention to move.

can be obtained in proportion to certain parameters of the stimulation signal. Thus, by modifying the stimulation signal parameters (timings for impulse rising, front and falling and current intensity), intense muscle contractions can be induced to produce functional movements. Because the muscle contraction is directly dependent by muscle tonus, skin resistance and electrode position, the FES parameters must be adjusted for every patient every time is used. The system configuration in which the FES device is used to help the patient to perform the desired movement is shown in fig. 11. This time the user intention is detected by using one of following devices: kinect, video + aruco markers, mio armband, IMU sensor etc.

The most used configuration is BCI - FES with additional robotic devices if needed (depends on the rehabilitation exercise). The patient must be able to seat without discomfort in a normal chair or wheelchair for 30 – 60 minutes, with his hands laid on the seat armrest. The exercise, for example flexion and extension of the hand fig. 11, is executed by the patient with his impaired hand but also with his healthy hand, one at a time. For this reason the FES electrodes are mounted on both hands over the finger extensors muscles (two channels).

The system configurations in which the BCI and FES devices are used are shown in fig 13.

In the first configuration, (Fig. 13 a), the BCI system component automatically triggers the FES component when it detects the patient intention to move and notifies the PCU.

FIGURE 12. Flexion and extension of the hand.

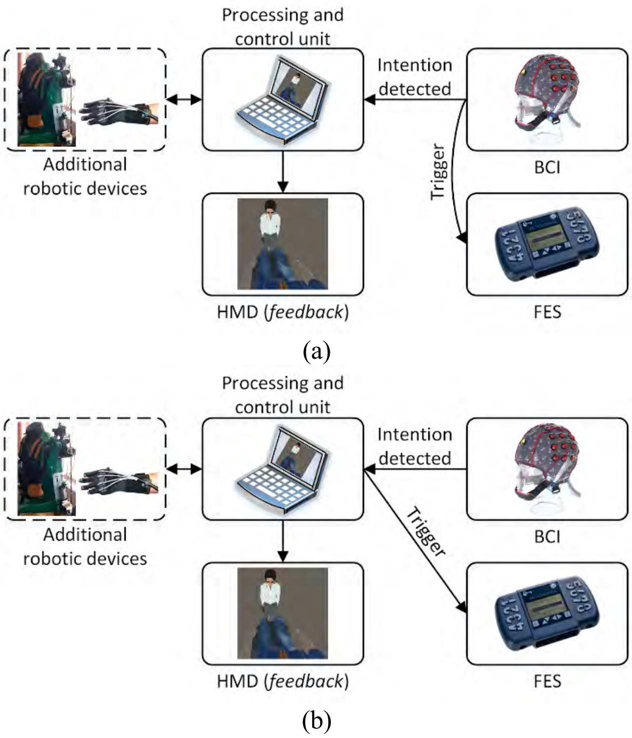


FIGURE 13. BCI - FES system configuration: a) BCI triggers FES & notifies PCU; b) BCI notifies PCU, PCU triggers FES.

In the second configuration, (Fig. 13 b), the BCI component only notifies the PCU about the patient intention and the PCU takes the decision to trigger the FES component. Fig. 14 shows a patient using the TRAVEE system configured as in Fig. 13. a.

H. THE THERAPIST GUI

This is the interface dedicated to the medical practitioner, which enables defining the patient profile (containing information regarding the patient, such as gender, age, weight, height, etc.), session configuring (exercises, durations, devices used) and analyzing statistics regarding the history of the sessions executions for the current patient. More details regarding the Therapist GUI are presented in past works [29]. This component is also integrated with a database that stores the patient and sessions information.

1) PATIENT PROFILE

The doctors are provided tools - in their dedicated user interface - to retain certain information regarding the patients that use the TRAVEE system for rehabilitation. The patient

FIGURE 14. Patient using the TRAVEE system.

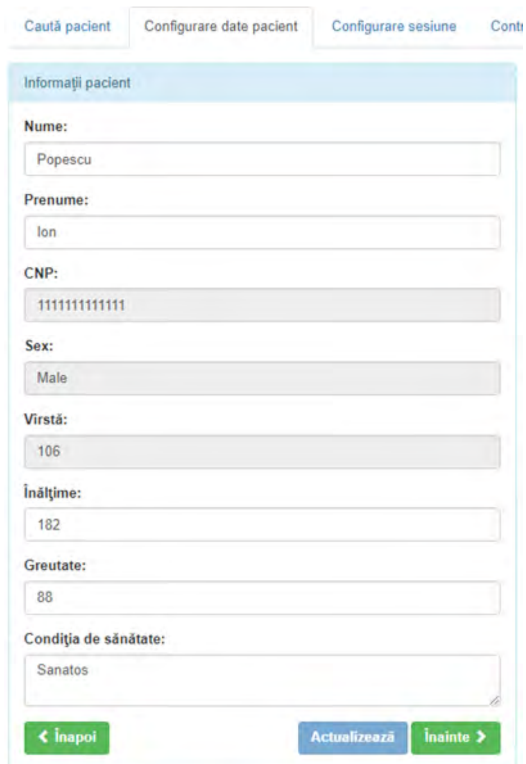


FIGURE 15. The patient profile configuration form.

profiles are defined by filling out a form with the following information: surname, name, personal identification number (PIN), gender, age, height, weight, health condition. This information is stored in a database and can be retrieved for further sessions.

2) SESSION DEFINITION

The doctor also has a view dedicated to the configuration of the rehabilitation session. In this view the doctor can select the exercises to be included in the rehabilitation session, their

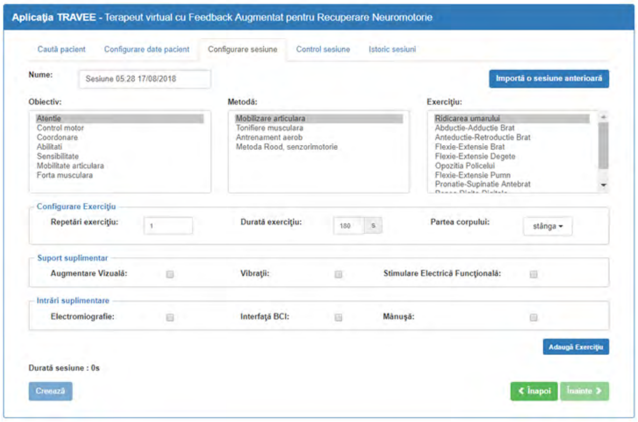


FIGURE 16. The session definition form.

FIGURE 17. The session control tools.

durations and the devices used for each one. The doctor also has the option to filter the available exercises based on their objective and the methods they are part of.

3) SESSION CONTROL

The session control view that is also a component in the interface dedicated to the doctors, allows the supervisor of the rehabilitation session to start, stop and pause the session. After the time chosen for an exercise has passed, the supervisor is asked to fill in a grade, evaluating the performances of the patient in the real world, based on the visual observations of the movement, as perceived by the supervisor.

4) SESSION ANALYSIS

The session analysis tool can analyze automatically many session recording files and extract synthetic data, so that the therapist can gather information without visually inspecting all the sessions.

This tool is a Unity application with a scene containing only the patient avatar, on which the recorded poses are played successively. As it was developed to automatically process many files without operator intervention, it allows the user to select a folder containing as many session recording files as necessary. It then automatically opens the session files one by one, and analyzes the poses in the file with the same algorithm described in the Visual augmentation of movements subchapter to determine the score for each pose. Using the variation of the scores and the other information in the files, the analysis tool calculates the following data:

TABLE 1. Augmentation and feedback pathways allowed for various input modalities.

	Visual augmentation	FES	Vibrations	Robotics
Body tracking	Augment	Yes	Yes	Yes
BCI	Fool	Yes	Yes	Yes
EMG	Fool	Yes	Yes	No

1) The execution times for each session
2) For each execution of an exercise in a session:

- The number of repetitions, as perceived by the system through the variations of the calculated scores for the tracked poses sequences. Each time the score changes the variation direction (was decreasing and is determined to be increasing, or if it was increasing and it is now considered to be decreasing), the algorithm records a change in the variation direction. Two successive changes in the variation direction is interpreted as a repetition.

- The average score for all the poses detected for the execution of a given exercise

The results of an analysis process is a file containing, for each recorded session: the total duration of the session and for each exercise in the session, the average score and the number of repetitions - as perceived by the Session Recording Analysis application.

5) THE REALTIME DATA VISUALIZATION COMPONENT

This component displays graphical representations of the EEG and EMG acquired data, to inform the doctor of their variations in time.

I. THE AVATAR PERSONALIZATION COMPONENT

This component allows the medical practitioner to change several characteristics of the virtual representation of the patient (gender, age, weight, hair and skin colour, clothes and hairstyle) in order to increase the immersion of the patient in the VE.

J. MODULAR AND INTEGRATIVE APPROACH

Using the dedicated graphical user interface, the doctor can define, for each rehabilitation session, a series of exercises and their durations, as well as the input devices to be used during the exercises and the feedback modalities.

Not all combinations of input modalities and feedback pathways implemented by TRAVEE make sense to be used together, therefore a set of allowed combinations was defined. These combinations are presented in Table 1 and are discussed below.

The Table 1 presents, for each input device, the available feedback modalities. The configurations that can be selected by the healthcare practitioner are limited by the conditions presented in this table but the doctor is not obliged to select all the available feedback devices for the session.

For the body tracking input as well as for the BCI one, any feedback can be implemented and used. For the EMG input, robotics would not be necessary, as the values of the electrical activations in the muscles will not be necessarily relevant if the muscles are actuated by the glove.

Depending on the selected inputs and feedbacks, the system behaviour changes. Therefore, several distinct functioning modes were defined. Out of these, the most significant ones are:

- Visual augmentation based on body tracking

During the session, the changes in the pose of the patient are analysed, and for each detected pose, the movement is slightly improved before being applied on the avatar in the VE, so that the patient perceives a better movement than he or she actually performed.

- Haptic feedback based on body tracking

For each pose detected by the optical tracking devices, the movement is analysed and when it is evaluated to be better than an established threshold, vibrations are applied on certain points on the hand of the patient. Therefore, the haptic feedback tells the patient when he or she has performed a good execution of the movement.

- BCI and FES

This functioning mode has a training phase in which the patient learns how to imagine the movement and the system computes an LDA classifier with a corresponding classification error. If the error rate is higher than 20% the training phase is repeated. If the error rate is lower than 20% the system can be switched to online mode where it can detect whether the patient is imagining the correct movement (with a certain degree of accuracy) or not and correspondingly activates FES for the respective hand. Therefore, the patient sees the feedback of what he is imaging.

- BCI and visual augmentation

This functioning mode is similar to the previous one, with the difference that instead of actually moving the patient hand through FES, the patient is immersed in the VE and the imagined movement is executed by the patient avatar in the VR.

- Robotic hand controlled mode based on body tracking

In this functioning mode, the position of the patient body as detected by the tracking devices is continuously evaluated. When the system detects that the patient cannot complete the movement, it activates the robotic glove for support, to help the patient perform the current exercise completely and correctly. Another function of this mode is that if no movement is detected in the patient hand, the robotic glove will start automatically to perform the whole movement, as the system will assume that the patient has no control of his hand muscles.

V. IN-VIVO TESTS AND CLINICAL TRIAL

The testing of the TRAVEE system took place in two stages. Initially, at the end of 2016, the initial prototype of TRAVEE was validated through two in-vivo testing sessions. Based on the observations made in these two tests the system was refined, to obtain the final prototype that was used during a clinical trial in May-June 2017.

For the tests to take place, permission was granted from the ethical council of the Neurological Recovery clinic of the National Institute of Recovery, Physical Medicine and Balneoclimatology (INRMFB) in Bucharest.

A. PRELIMINARY IN-VIVO TESTS

1) OBJECTIVES

The preliminary in-vivo tests were designed to test the initial prototype of the TRAVEE system, in order to determine whether it could successfully be applied to patients with neuromotor disabilities, what were the aspects that could make it easier to be used in a clinical settlement, and to test several functioning modes.

Two in-vivo testing sessions took place, in November and December 2016, respectively.

2) TECHNICAL DESCRIPTION

The first in-vivo tests evaluated the system for the Forearm flexion-extension, Arm anteduction-retroduction, Palm flexion-extension, Fingers flexion-extension. The second in-vivo test evaluated the system with the Forearm flexion-extension and Palm flexion-extension movements.

The hardware used in the first in-vivo testing session was: a computer running the TRAVEE VR Central System, Oculus Rift for immersion in the VE, BCI and FES. The second set of in-vivo tests used a computer running the VR Central System, Oculus Rift for immersion in the VE and the haptic feedback device.

The TRAVEE components that were tested during the preliminary in-vivo tests: the VR Central System, the Data Acquisition and Control and the Movement Analysis component.

The tested functioning modes: visual augmentation based on body tracking, visual augmentation based on BCI, FES controlled by BCI and haptic mode based on body tracking.

3) CLINICAL SETUP

Each in-vivo testing session took place in one day, at the Neurological Recovery clinic of the National Institute of Recovery, Physical Medicine and Balneoclimatology (INRMFB) in Bucharest.

Patients, as well as their families, have been informed about the device created in this research project. The Information Form was handed in, the questions and the unclear things were answered. Those who have accepted to participate in the test have signed the Informed Consent, in the presence of the medical team members and their families.

The patients selected by the doctors had various degrees of disability, ranging from patients with no motor control to patients who only had a slight tremor in their hand. All the patients had suffered a disability of their hand as a result of stroke. In the first in-vivo tests one patient tested the system with BCI and FES, and three patients tested only the VR Central System component with visual augmentation. In the second in-vivo testing session, three patients tested the VR Central System.

In the first in-vivo testing session, one patient executed a session containing the Palm flexion-extension movement with the BCI controlled FES augmentation, and three patients used the TRAVEE system in sessions with Palm flexion-extension and Forearm flexion-extension movements.

In the second in-vivo testing session, three patients tested the TRAVEE system for the Forearm flexion-extension and Palm flexion-extension movements with visual augmentation based on body tracking. One patient also tested the haptic feedback device.

4) INTERPRETATION OF THE RESULTS

The results of the in-vivo testing sessions were presented in previous works, for the first session [30] and the second session [31].

The results of the in-vivo tests were mainly technical conclusions regarding the usability of the TRAVEE system as well as possible improvements that could be brought upon the solution to prepare it for the clinical trial.

The participants to the in-vivo tests were asked to fill in questionnaires regarding their experience with TRAVEE, based upon which several conclusions were drawn.

1. During the test, what was the perceived level of tiredness?
2. During the tests did you feel dizziness?
3. During the tests did you feel nauseous?
4. During the tests did you feel any anxiety or fear?
5. The image perceived on the virtual glasses/monitor was clear?
6. Did you feel physical discomfort due to the system components?
7. Did you feel pain due to the FES/haptical stimulation?
8. How real did the avatar movements seem to you?
9. How well do you identify your movements to those of the avatar?
10. Did you feel that the movements of the avatar were different than yours (greater)?
11. Are the indications of the virtual therapist useful for the exercise execution?
12. How useful do you find such a rehabilitation system?

TABLE 2. Questionnaire responses of the patients for the two in-vivo testing sessions [30], [31].

Q	#1	#2	#3	#4	#5	#6	#7	Avg.
1	5	3	3	1	1	1	3	2.43
2	1	1	4	4	1	1	1	1.86
3	1	1	1	1	1	1	1	1.00
4	1	1	1	3	1	1	1	1.29
5	4	3	4	3	3	5	5	3.86
6	1	1	2	1	1	2	1	1.29
7	1				1	1	1	1.00
8	3	1	1	5	5	4	4	3.29
9	4	3	2	5	4	4	2	3.43
10		2	2	1	1	5	4	2.50
11		3	3	5	5	3	2	3.50
12		4	4	1	4	5	1	3.17

B. CLINICAL TRIAL

The effort necessary for the experiments associated with an extensive clinical trial are tremendous, therefore our goal was not to include in the tests a large number of patients, but to prove the validity of our system and the ideas that support it, and its use in a clinical environment. This decision was taken also because the system is still a prototype, not a final product, therefore we treated each patient participating in our trial as an individual test case, not necessarily aiming for statistical evidence as we believe it is still very early for such results.

1) OBJECTIVES AND APPROACH

The clinical trial took place between 28th April 2017 and 19th May 2017, at the National Institute for Rehabilitation, Physical Medicine and Balneoclimatology (INRMFB) in Bucharest. The tested configurations were chosen based on the degree of disability of each patient and included BCI, FES, VR and robotic glove.

From a clinical point of view, this study is an experimental acute one of a number of cases in which we followed, for each subject, the persistent therapeutic response in patients with stroke sequelae in the upper limb after post experiment and the possible occurrence of side effects.

The secondary goals were:

- Establish with maximum possible accuracy the clinical and functional profile of the patient after stroke that can benefit from a clinical and functional treatment with the TRAVEE system
- Determining the factors that restrict the application of the method
- Weaknesses of the device and corrective ways
- A qualitative assessment of the final prototype of the TRAVEE computerized system and to track the effects of TRAVEE during the development of the program.

From a technical point of view, the results of the clinical trial were measured in the evolutions of the scores given by the system to each rehabilitation exercise, as processed from the recording file. As neither of the patients has taken part in more than six rehabilitation sessions with TRAVEE and many only participated to one or two sessions, the results were mostly specific to a qualitative clinical trial and not to a quantitative one. The system has been improved from the testing sessions, by refining the existing functionalities, as well as adding several new functioning modes as well as the recording function, described previously.

During the clinical trial, the TRAVEE system contained all the designed components: VR Central System, Data Acquisition and Control, Therapist GUI, 3D animations of the VT, tracking of the patient body movements, session recording, session analysis.

2) CLINICAL SETUP

30 patients with stroke were included in the study, 21 of them benefited from the complete experiment with the TRAVEE device.10 patients were tested for the response to BCI therapy, 2 patients were included in the mixed experiment, TRAVEE plus BCI, and one patient was included in the experiment with additional stimulation with FES and vibration stimulation.

For all the patients included in study the stroke was less than 12 months.

The general clinical profile of the patient included in the study was: conscious, temporal-space-oriented, cardio-respiratory balanced, no digestive or renal accusation, with central post-stroke motor neuron syndrome.

It is essential that passive mobilization applications that are made analytically and / or globally by the therapist to restore / revive the neural circuits defining the correct parameters of the movement: amplitude, direction, speed before TRAVEE training

The lot of patients had the following demographic characteristics:

- 15% women and 85% men
- ages between 43 and 79 years;
  The followed clinical parameters were:
- motor control
- spasticity, reflexes, other signs of hypertonia
- muscle strength
- vicious postures (joint, type, degree)
- synkinesis of the upper limb (type, description)
- coordination problems
- superficial and deep sensitivity
- articular mobility degree
- CRPS I complications, glenohumeral subluxation, thalamic pain

3) FUNCTIONAL EVALUATION

For functional evaluation assessed the degree of general dysfunctionality of an upper limb; to all patients this was in the range 2-5.

On the scale of functional independence regarding self-care and locomotion activity, the situation ranged from modify independence to 75% dependency (the Functional Independence Measurement scale).

Other scales used:

- Deficit scale: Manual Muscle Testing (MMT) for Muscle Strength Assessment, Ashworth Scale Assessment Scale, Mini Mental State Examination (MMSE) for cognitive status assessment, reflex score, fatigue scale
- Disability:
  - Action Research ARM Test (ARAT)
  - Box and Blocks Test
  - Motor Assessment Scale (MAS)
  - Rivermead Motor Assessment

4) INCLUSION CRITERIA

- Stable neurological status
- Conscious state
- Significant persistent neurologic motor deficit
- Functional disability at the level of at least two of the following: mobility, self-care capacity, communication, sphincterian control, swallowing
- Cognitive functions well preserved to allow learning
- Ability to communicate well enough to allow collaboration
- Physical exercise tolerance sufficient to perform the active program
- Achievable therapeutic goals

5) EXCLUSION CRITERIA

- Central motor neuron syndrome older than 6 months
- Spasticity > Ashworth Grade 2
- Instability of central neurological lesions; Progressive motor deficit
- Cardiac unstable or other co-morbidities requiring emergency medical care
- Intercurrent infections, other comorbidities that contraindicate inclusion in a medical recovery program
- Complete lack of proximal motor control at the level of the upper limb
- Uncontrolled psychiatric disorders
- Uncontrolled seizures
- Significant cognitive impairment with MMSE <18
- Bilateral marked deafness or hearing loss
- Amputations, ankyloses or severe limitations of joint mobility at the level of the upper limb, caused by diseases prior to neurological disease
- Multiple/ repeated central neurological lesions
- Co-existence of a peripheral neurological deficit at the level of the upper limb
- Absence of consent (informed consent) of the patient or family

6) RESULTS

a: TECHNICAL RESULTS. EVALUATIONS BASED ON THE AUTOMATED ANALYSIS MADE BY THE SYSTEM

The sessions that were performed with the VR Central System were recorded and then were analysed using the previously described Session Analysis application.

A total of 21 patients tested the TRAVEE system with visual augmentation. The recordings of the sessions were analysed and the most relevant ones are summarized below. The number of repetitions was determined automatically, based on the number of changes in the direction of variation of the calculated score for each movement.

Because each movement has different parameters used in its evaluation, the scores assigned to different movements cannot be compared. Also, the average scores for each patient are individual, based on his/her abilities in the exercised hand. A greater score indicates a larger amplitude of movement, therefore a possibly more complete execution.

Another important observation is that the performed clinical test evaluated more patients for a small number of sessions, to assess the usability of the system in various scenarios and various degrees of disability. For the results to be medically relevant, a more extensive clinical test would have been appropriate, with the same patients exercising for several sessions each day, for at least several weeks.

Out of the 21 patients that tested the TRAVEE system during the clinical trial, we selected for presentation in this paper those that had at least three rehabilitation sessions with the system.

i) Patient RV2

This patient had the most remarkable evolution with the TRAVEE system. Before the first session, the patient had a very strong tremor in the arm, that did not allow him to execute accurate and controlled movements. As soon as the Oculus and tracking devices were installed, the patient was immersed in the virtual environment, the session started and he was asked to repeat the movements shown by the virtual therapist; the tremor almost disappeared, being reduced greatly. The progress – as we were reported – was maintained outside of the virtual environment. Although we cannot determine exactly the reason for this improvement and we cannot necessarily connect it to the system, it is a coincidence that definitely requires further research.

Evolution of the Forearm Flexion-Extension movement

ii) Patient RV5

This patient came to the sessions regularly, was receptive to the idea of the system, had a positive attitude and a good evolution. For each session the patient had two repetitions of the Forearm Flexion-Extension, each of 180 or 200 seconds, during which, each time, performed approximately 20 repetitions, as evaluated by the system. The average scores did not vary significantly during the trial period, more sessions would have been required for statistical relevant information regarding the progress of the patient.

iii) Patient RV13

For this patient we observed an ascending trend for the average scores given by the system for the two movements executed for each of the three rehabilitation sessions in which the patient took part. At the same time, the number of repetitions detected by the system decreased. This observation could mean a more qualitative execution of the movements, at a slower pace, with better motion control.

The evolution of the Forearm Flexion-Extension movement is presented below.

TABLE 3. Exercises executed with TRAVEE by Patient RV2.

Date	Exercise	Secs	Reps	Avg
04.05	Forearm Flexion-Extension	200	17	65.85
04.05	Forearm Flexion-Extension	180	18	62.65
08.05	Forearm Flexion-Extension	200	26	59.91
08.05	Forearm Flexion-Extension	200	24	67.14
09.05	Forearm Flexion-Extension	200	26	59.92
10.05	Forearm Flexion-Extension	200	40	33.91
10.05	Forearm Flexion-Extension	200	17	24.4
11.05	Forearm Flexion-Extension	200	41	40.86
12.05	Forearm Flexion-Extension	200	26	65.97
12.05	Forearm Flexion-Extension	180	29	62.99

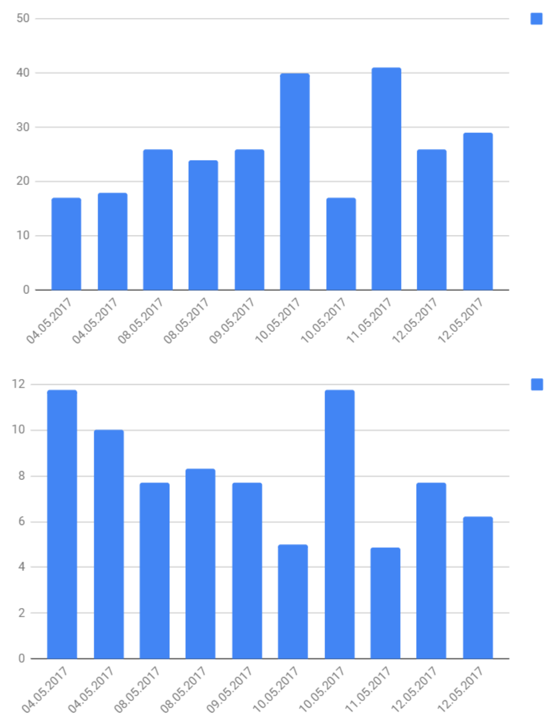


FIGURE 18. Average scores evolution (top) and average seconds/repetition evolution (bottom).

TABLE 4. Exercises executed with TRAVEE by Patient RV5.

Date	Exercise	Secs	Reps	Avg
09.05	Forearm Flexion-Extension	200	19	92.64
09.05	Forearm Flexion-Extension	200	23	89.28
11.05	Forearm Flexion-Extension	200	44	79.77
11.05	Forearm Flexion-Extension	200	26	74.1
12.05	Forearm Flexion-Extension	200	25	84.12
12.05	Forearm Flexion-Extension	180	23	83.74
19.05	Forearm Flexion-Extension	220	26	76.15
19.05	Forearm Flexion-Extension	200	27	87.83

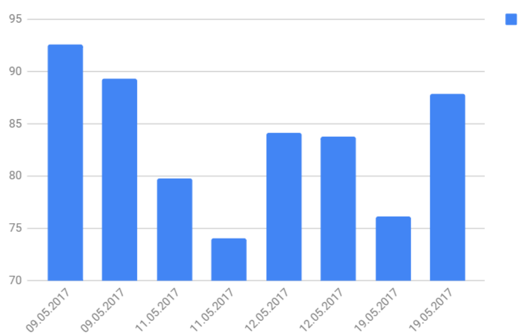


FIGURE 19. Average scores evolution (top) and average seconds/repetition evolution (bottom).

TABLE 5. Exercises executed with TRAVEE by Patient RV13.

Date	Exercise	Secs	Reps	Avg
09.05	Forearm Flexion-Extension	200	42	68.95
09.05	Palm Flexion-Extension	200	56	20.16
11.05	Forearm Flexion-Extension	200	30	75.38
11.05	Palm Flexion-Extension	180	47	27.57
19.05	Forearm Flexion-Extension	200	34	75.69
19.05	Palm Flexion-Extension	220	33	30.05

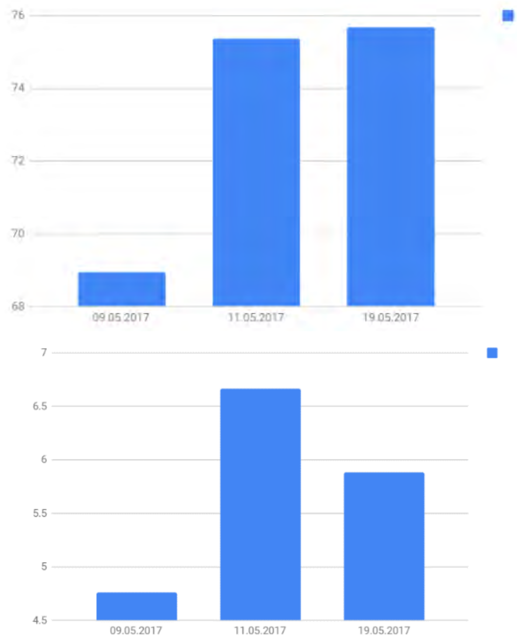


FIGURE 20. Forearm Flexion-Extension: Average scores evolution (top) and average seconds/repetition evolution (bottom).

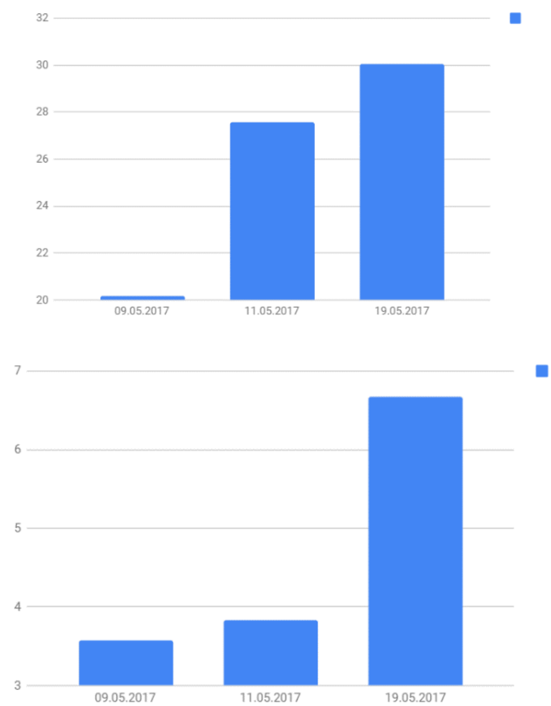


FIGURE 21. Palm Flexion-Extension: Average scores evolution (top) and average seconds/repetition evolution (bottom).

iv) Patient RV15

The patient also took part in several rehabilitation sessions with the system. Slight improvements were observed between the sessions regarding the number of repetitions detected by the system as well as the average scores.

v) Patient RV21

This patient took part in three rehabilitation sessions with the system. For all the three types of exercises there was a reduction in the average execution time, as perceived by the system.

TABLE 6. Exercises executed with TRAVEE by Patient RV15.

Date	Exercise	Secs	Reps	Avg
04.05	Fingers Flexion-Extension	200	32	17.86
04.05	Palm Flexion-Extension	180	10	31.82
04.05	Forearm Pronation-Supination	180	18	125.78
04.05	Forearm Flexion-Extension	180	35	71.5
05.05	Fingers Flexion-Extension	180	40	22.05
27.04	Fingers Flexion-Extension	310	80	18.29
27.04	Palm Flexion-Extension	300	37	35.42
28.04	Forearm Flexion-Extension	300	57	79.7

FIGURE 22. Fingers Flexion-Extension: Average scores evolution (top) and average seconds/repetition evolution (bottom).

b: CLINICAL RESULTS/SCORES

We underline that this clinical trial is an initial, acute-type experiment through its design team managed to adjust the TRAVEE program and bring it into its current form. This study will be followed by research to track the effectiveness and efficacy of TRAVEE in patients with stroke sequelae and to transpose the project into real life. The experiments aimed the adaption of the patients, their ability to learn, the ability to integrate TRAVEE into a complex, comprehensible medical poststroke recovery program.

TABLE 7. Exercises executed with TRAVEE by Patient RV21.

Date	Exercise	Secs	Reps	Avg
19.05	Forearm Flexion-Extension	200	29	101.75
19.05	Palm Flexion-Extension	200	43	26.97
19.05	Fingers Flexion-Extension	200	31	28.14
19.05	Fingers Flexion-Extension	120	31	29.13
22.05	Palm Flexion-Extension	200	22	21.28
22.05	Fingers Flexion-Extension	200	55	28.18
22.05	Palm Flexion-Extension	180	61	21.57
22.05	Forearm Flexion-Extension	200	36	96.17
25.05	Fingers Flexion-Extension	200	62	21.98
25.05	Palm Flexion-Extension	200	61	21.57
25.05	Forearm Flexion-Extension	200	36	96.17

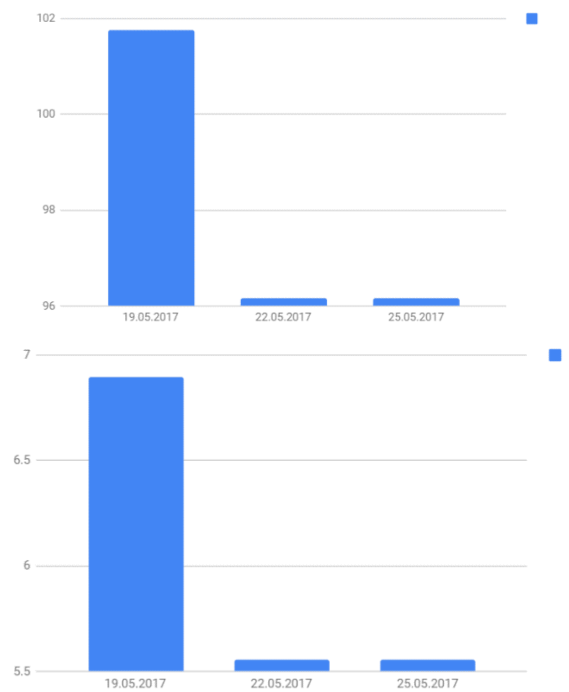


FIGURE 23. Forearm Flexion-Extension: Average scores evolution (top) and average seconds/repetition evolution (bottom).

From the point of view of the outcome of the acute experiment in each patient, this study led to increased motor control in the upper limb, especially proximal and intermediate, in 80% of patients. A statistically significant increase cannot be defined, but the evolution trend is positive. The lack of a positive response was seen in one of the patients with a low MMSE score (19, 20) and in 3 of the patients with

TABLE 8. Questionnaire answers.

Pat.	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12
1	2	2			3	4	1	2	3	2	2	3
2	2				3	3	2	3	3	3	3	3
3	3	2	1	1	3	4	1	2	3	2	4	3
4	3	2			3	3	1	2	2	4	2	2
5	4	2	1	1	3	3		2	3	4	2	2
6	4	2	1	2	3	4	1	1	2	4	2	2
7	2	2			3	4	1	2	2	2	2	3
8	3	2	1		3	3		2	3	2		3
8	4	3	2	3	3	4	3	2	3	4	3	3
10	2				3	4	1	2	2	4	2	3
11	2				4	3		3	3	2	4	3
12	4	3	2	3	4	4	3	2	2	3	2	2
13	2				4	3		3	3	2	3	3
14	4	1			3	3	2	3	3	1	3	2
15	4	2	1		2	3	3	3	3	2	3	2
16	2	2		1	4	3	1	3	3	2	3	2
17	4	2	1	1	2	3	1	2	3	2	3	3
18	2				3	3		2	3	3	2	2
19	4	1			3	3		3	2	2	2	2
20	2	2			3	1		3	3	2	2	3
21	4				3	4		3	3	2	3	3
Avg	3.00	2.00	1.25	1.71	3.10	3.29	1.62	2.38	2.71	2.57	2.60	2.57

MAS 2 measured on the MMT scale. The other patients with MAS 2 had a positive response after associating additional stimuli (BCI, Vibration, FES).

There were no serious adverse effects. As a common side effect present in all patients, we underline the fatigue that occurred more rapidly in those with higher cognitive impairment, with grade 2 spasticity and those with low muscular strength; the presence of abnormal movement patterns increased fatigue

Interpretation of results

Using the TRAVEE device for medical recovery of the upper limb function:

1. Allows improvement of motor control at the upper limb for patient after stroke, especially at the proximal and intermediate levels

2. This device is ideal to be use for patients with muscle strength 4 (MMT) patient, less than 2 Ashworth grade spasticity, with no abnormal movement patterns without severe cognitive impairment. Age and cardio-vascular associated pathology do not appear to negatively influence the patient's response to acute experimentation.

3. No serious adverse effects were seen. As a side effect we've identified fatigue. Patients also accused: dizziness, pain, feeling discomfort, but of low intensity, not interfering with the experiment. Just fatigue has the main cause of stopping the experiment.

4. Adding additional stimuli: functional electrical stimulation, vibrational stimulation, cerebral brain-computer brain stimulation seem to increase the positive effect on motor control in patients with lower muscular strength, even in plegical ones.

5. Validation of the method requires a prospective, double-blind, controlled clinical trial in batches of patients sufficiently large to have statistical power.

c: QUESTIONNAIRES

The patients that participated in the clinical trial received a questionnaire containing 12 questions. Each question had five answer options, on a scale from 1 to 5. The questions and the answers given by the patients are presented below.

Q1. During the training sessions, what was the perceived level of tiredness?

FIGURE 24. Fingers Flexion-Extension: Average scores evolution (top) and average seconds/repetition evolution (bottom).

Q2. During the training sessions, did you feel dizziness? If so, how intense?

Q3. During the training sessions, did you feel nauseous? If so, how intense?

Q4. During the training sessions, did you feel any anxiety or fear? If so, how intense?

Q5. During the training sessions, how clear was the image perceived on the virtual glasses/monitor?

Q6. During the training sessions, did you feel physical discomfort due to the system components? If so, how intense?

Q7. During the training sessions, did you feel pain? If so, how intense?

Q8. During the training sessions, how real did the avatar movements seem to you?

Q9. During the training sessions, how well did you identify your movements to those of the avatar?

Q10. During the training sessions, did you feel that the movements of the avatar were different than yours (greater)? If so, how much different?

Q11. During the training sessions, were the indications of the virtual therapist useful for the exercise execution? If so, how useful?

Q12. Do you consider that the training sessions with this system were useful for your rehabilitation? If so, how useful?

The responses received from the 21 patients are presented in the following table.

VI. CONCLUSIONS AND PERSPECTIVES

The current paper presents the vision implemented by the TRAVEE system, the medical background and perspectives upon which it was designed, as well as technical details regarding its implementation. TRAVEE is a system dedicated to medical neuromotor rehabilitation of the upper limbs that combines multiple technologies: VR, BCI, FES, robotics and haptics, with novel ideas, such as augmented feedback through natural movement augmentation and multimodal feedback. It was designed to support rehabilitation at several levels of disability - providing various degrees of support, from complete movement (through FES and robotics) to support for completing a movement either motor (robotic) or virtual (visual augmentation). The system was tested in a medical setting, during development in two in-vivo sessions, as well as after the final prototype was implemented, through a clinical trial. The paper presents the results of all the testing sessions, that correspond to those of a qualitative evaluation.The results we observed during the clinical trial show that the visual augmentation through VR has a great potential in rehabilitation, that must be further developed and researched.

The perspectives of future development of the system are vast and heterogenous. The main desired evolution for the system is the migration towards a low-cost solution. Providing an accessible system was one of the main targets of TRAVEE and - partially - it has succeeded. The areas in which we believe there is room for improvement are related to the EEG device which may be substituted by a low-cost solution (such as Emotiv Epoc [https://www.emotiv.com/epoc/]). This direction could assist TRAVEE to evolve into a commercially available product, with a wide applicability in the rehabilitation process. This commercial version could be based mainly on the VR component, arm and hand tracking and light robotics, with aspects of gamification. This solution could also be enhanced in clinical settings with the EMG and FES components.

Other possible paths of evolution for our research aim a better understanding of the effects that visual augmentation and multimodal feedback have upon the rehabilitation processes and on the cortical reorganization process. Another direction is to study whether the visual augmentation affects spasticity that appears in patients suffering after-effects of stroke, to test various environments and their influence on the sessions and study evolutions with various visual augmentation degrees and a proper comparison between classical rehabilitation sessions and the ones enhanced through visual augmentation and multimodal feedback.

REFERENCES

[1] World Health Report. *The World Health Report 2002—Reducing Risks, Promoting Healthy Life*. Accessed: Aug. 2018. [Online]. Available: http://www.who.int/whr/2002/en/

[2] *Heart Disease and Stroke Statistics 2018 At-a-Glance*. Accessed: Aug. 2018. [Online]. Available: https://www.heart.org/-/media/data-import/downloadables/heart-disease-and-stroke-statistics-2018—at-a-glance-ucm_498848.pdf

[3] T. Mulder, "Motor imagery and action observation: Cognitive tools for rehabilitation," *J. Neural Transmiss.*, vol. 114, no. 10, pp. 1265–1278, Oct. 2007.

[4] O. Ferche *et al.*, "From neuromotor command to feedback: A survey of techniques for rehabilitation through altered perception," in *Proc. E-Health Bioeng. Conf. (EHB)*, Nov. 2015, pp. 1–4.

[5] V. S. Ramachandran and D. Rogers-Ramachandran, "Synaesthesia in phantom limbs induced with mirrors," *Proc. Biol. Sci.*, vol. 263, pp. 377–386, Apr. 1996.

[6] B. L. Chan *et al.*, "Mirror therapy for phantom limb pain," *New England J. Med.*, vol. 357, no. 21, pp. 2206–2207, Nov. 2007.
[7] S. Hoermann, L. Hale, S. J. Winser, and H. Regenbrecht, "Augmented Reflection Technology for Stroke Rehabilitation—A clinical feasibility study," in *Proc. 9th Int. Conf. Disab., Virtual Reality Assoc. Technol. (ICDVRAT)*, Laval, France, 2012, pp. 1–9.
[8] J. Siegel-Itzkovich, "Hadassah tries to 'trick the brain' with virtual reality to treat physical ailments," *Jerusalem Post*, Dec. 2008. [Online]. Available: https://www.jpost.com
[9] U. Feintuch, M. Tuchner, A. Lorber-Haddad, Z. Meiner, and S. Shiri, "VirHab—A virtual reality system for treatment of chronic pain and disability," in *Proc. Int. Conf. Virtual Rehabil.*, Jun./Jul. 2009, pp. 83–86.
[10] S. Shiri *et al.*, "A novel virtual reality system integrating online self-face viewing and mirror visual feedback for stroke rehabilitation: Rationale and feasibility," *Topics Stroke Rehabil.*, vol. 19, no. 4, pp. 277–286, 2012.
[11] O. Ferche, A. Moldoveanu, and F. Moldoveanu, "An up to date survey of assistive technologies for neuro-motor rehabilitation using virtual reality," in *Proc. 8th Int. Conf. Speech Technol. Hum.-Comput. Dialogue*, Oct. 2015, pp. 211–220.
[12] R. Kizony *et al.*, "Evaluation of a tele-health system for upper extremity stroke rehabilitation," in *Proc. Int. Conf. Virtual Rehabil. (ICVR)*, Aug. 2013, pp. 80–86.
[13] K. Hata, K. Koyanagi, K. Nagata, T. Motoyoshi, H. Masuta, and T. Oshima, "Development and pretesting of reaching training software for post-stroke patient using SEMUL rehabilitation system for upper limbs," in *Proc. IEEE Int. Conf. Robot. Biomimetics*, Bali, Indonesia, Dec. 2014, pp. 1662–1667.
[14] S. Comani *et al.*, "Monitoring neuro-motor recovery from stroke with high-resolution EEG, robotics and virtual reality: A proof of concept," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 23, no. 6, pp. 1106–1116, Nov. 2015, 10.1109/TNSRE.2015.2425474.
[15] G. J. Barton, M. B. Hawken, R. J. Foster, G. Holmes, and P. B. Butler, "Playing the Goblin Post Office game improves movement control of the core: A case study," in *Proc. Int. Conf. Virtual Rehabil.*, Zurich, Switzerland, Jun. 2011, pp. 1–5.
[16] T. Pintaric and H. Kaufmann, "Affordable infrared-optical pose tracking for virtual and augmented reality," in *Proc. Trends Issues Tracking Virtual Environ. Workshop*, 2007, pp. 44–51.
[17] E. Tunik, S. Saleh, H. Bagce, A. Merians, and S. V. Adamovich, "Mirror feedback in virtual reality elicits ipsilesional motor cortex activation in chronic stroke patients," in *Proc. Int. Conf. Virtual Rehabil.*, Zurich, Switzerland, Jun. 2011, pp. 1–7.
[18] V. Crocher, P. Hur, and N. J. Seo, "Low-cost virtual rehabilitation games: House of Quality to meet patient expectations," in *Proc. Int. Conf. Virtual Rehabil.*, Philadelphia, PA, USA, Aug. 2013, pp. 94–100.
[19] A. Voinea, A. Moldoveanu, F. Moldoveanu, and O. Ferche, "Motion detection and rendering for upper limb post-stroke rehabilitation: Multimedia applications for medical and healthcare education and learning; Rehabilitative and assistive technologies," in *Proc. 5th Int. Conf. E-Health Bioeng. (EHB)*, Nov. 2015, pp. 1–4.
[20] A. A. Frolov *et al.*, "Post-stroke rehabilitation training with a motor-imagery-based brain-computer interface (BCI)-controlled hand exoskeleton: A randomized controlled multicenter trial," *Frontiers Neurosci.*, vol. 11, p. 400, Jul. 2017.
[21] R. Ortner, D. C. Irimia, J. Scharinger, and C. Guger, "A motor imagery based brain-computer interface for stroke rehabilitation," *Studies Health Technol. Inform.*, vol. 181, pp. 319–323, Sep. 2012.
[22] S. R. Soekadar, N. Birbaumer, M. W. Slutzky, and L. G. Cohen, "Brain-machine interfaces in neurorehabilitation of stroke," *Neurobiol. Disease*, vol. 83, pp. 172–179, Nov. 2015.
[23] C. Neuper and G. Pfurtscheller, "Neurofeedback training for BCI control," in *Brain-Computer Interfaces* (The Frontiers Collection). Berlin, Germany: Springer, 2010, pp. 65–78.
[24] A. Lechner, R. Ortner, and C. Guger, "Feedback strategies for BCI based stroke rehabilitation: Evaluation of different approaches," *Replace, Repair, Restore, Relieve—Bridging Clinical and Engineering Solutions in Neurorehabilitation* (Biosystems & Biorobotics), vol. 7, W. Jensen, O. Andersen, and M. Akay, Eds. Cham, Switzerland: Springer, 2014.
[25] R. G. Lupu, D. C. Irimia, F. Ungureanu, M. S. Poboroniuc, and A. Moldoveanu, "BCI and FES Based therapy for stroke rehabilitation using VR facilities," *Wireless Commun. Mobile Comput.*, vol. 2018, Apr. 2018, Art. no. 4798359.
[26] A. Espinosa, R. Ortner, D. Irimia, and C. Guger, "Brain computer interfaces as stroke rehabilitation tools: Optimization of current strategies," in *Proc. 5th Int. Conf. Adv. Cognit. Technol. Appl. Cognit.*, Valencia, Spain, May/Jun. 2013, pp. 1–25.
[27] J. Müller-Gerking, G. Pfurtscheller, and H. Flyvbjerg, "Designing optimal spatial filters for single-trial EEG classification in a movement task," *Clin. Neurophysiol.*, vol. 110, no. 5, pp. 787–798, 1999.
[28] *Home—G.Tec—Guger Technologies*. Accessed: Aug. 2018. [Online]. Available: http://www.gtec.at/
[29] R. G. Lupu *et al.*, "Virtual reality system for stroke recovery for upper limbs using ArUco markers," in *Proc. 21st Int. Conf. Syst. Theory, Control Comput. (ICSTCC)*, vol. 1, Sinaia, Romania, Oct. 2017, pp. 548–552.
[30] O.-M. Ferche, A. Moldoveanu, and F. Moldoveanu, "The TRAVEE system for neuromotor recovery: Architecture and implementation," in *Proc. 6th IEEE Int. Conf. E-Health Bioeng. (EHB)*, Sinaia, Romania, Jun. 2017, pp. 575–578.
[31] O.-M. Ferche *et al.*, "The TRAVEE neuromotor rehabilitation system: *In-vivo* testing," in *Proc. Zooming Innov. Consum. Electron. Int. Conf. (ZINC)*, May/Jun. 2017, pp. 30–33.

ALIN MOLDOVEANU is currently a Full Professor with the Computer Science and Engineering Department, Faculty of Automatic Control and Computers, Politehnica University of Bucharest, where he teaches software engineering and virtual reality.

He is actively researching in areas, such as virtual and augmented reality (exploring and applying immersion, sensory substitution, and distorted reality), eHealth (assistive and rehabilitative solutions and prevention of hospital acquired infections), and eLearning (3-D MMO mixed-reality campuses). He is the Director or is responsible for national or European research projects in areas, such as Sound of Vision, TRAVEE, HAI-OPS. A detailed log of projects and publications is available at: https://cs.pub.ro/index.php/people/userprofile/alin_moldoveanu.

OANA-MARIA FERCHE professional experience include being a Teaching Assistant with the Politehnica University of Bucharest, from 2014 to 2018, teaching software engineering and computer graphics. She is a Computer Science Engineer and is currently a Software Programmer with InfoWorld company, which is active in the medical IT&C market.

She is currently pursuing the Ph.D. degree in virtual reality. Her thesis is on a system for neuromotor rehabilitation using virtual reality. As part of her Ph.D. studies, she researched in the virtual reality in medical rehabilitation field, in order to participate in the TRAVEE national project, that developed and tested a system for neuromotor recovery after stroke. A selection of her published work can be found here: https://www.researchgate.net/profile/Oana_Maria_Ferche/contributions

FLORICA MOLDOVEANU is currently a Professor with the Department of Computer Science and Engineering, Politehnica University of Bucharest. Her research and teaching activity is focused on computer graphics, computer vision, software engineering, and e-health. She coordinates the master program in computer graphics, multimedia and virtual reality with the Faculty of Automatic Control and Computers. She is also the President of Health Level 7 Romania Association.

ROBERT GABRIEL LUPU is currently an Associate Professor with the Computer Department, Faculty of Automatic Control and Computer Engineering, Gheorghe Asachi Technical University of Iași, where he teaches computer programming, object oriented programming, and the design of human–computer interaction systems.

He is actively researching in areas such as assistive technology (eye tracking-based communication systems for patients with neuro locomotor disabilities and sensory substitution), e-Health (patient monitoring and VR- and BCI-based rehabilitative solutions). He is responsible for or is a member of national or European research projects in areas, such as TRAVEE, SoV, and SIACT. A detailed list of publications is available at: https://scholar.google.com/citations?user=66PuewoAAAAJ&hl=en

DELIA CINTEZĂ was the Director in Research and Development with the National Institute for Rehabilitation, Physical Medicine and Balneoclimatology, from 2009 to 2010, where she is currently the Medical Director. She was the President of the Romanian Association of Balneology, from 2010 to 2014.

She is currently a member of several medical associations, including the Rehabilitation and Physical Medicine association, the Romanian Association for the Study of Pain, and the Romanian Osteoporosys Association.

She is also an Assistant Professor with the Carol Davila University of Medicine and Pharmacy, Bucharest, teaching courses and activities in the specialty of physical medicine, rehabilitation, and balneoclimatology.

Her research activity includes over 120 papers published in national and international journals, and participation in over 20 research projects and medical studies.

DANUT CONSTANTIN IRIMIA is currently a Lecturer within the Faculty of Electrical Engineering, Gheorghe Asachi Technical University of Iași, where he teaches industrial robotics, system theory, and signal acquisition and processing. His current research interests involve human motion analysis and synthesis, neuroprostheses, biomechanics, and brain–computer interface.

CORNELIU TOADER has been a Neurosurgeon with the Neurosurgical Department, National Institute of Neurology and Neurovascular Diseases, Bucharest, Romania, since 1992. He is also an Assistant Professor with the Faculty of Medicine, Carol Davila University of Medicine and Pharmacy, Bucharest.

During his career, he has operated on more than 2000 patients. His surgical activity consists mostly in cerebro-vascular pathology, mainly intracranial aneurysms (over 1000 patients), but he also operates arterio-venous malformations, brain tumors, cranial base tumors, and spinal pathology.

In 2009, he was named the Chief of the National Institute of Neurology and Neurovascular Diseases.

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 4798359, 8 pages
https://doi.org/10.1155/2018/4798359

Research Article

BCI and FES Based Therapy for Stroke Rehabilitation Using VR Facilities

Robert Gabriel Lupu ,[1] Danut Constantin Irimia,[2] Florina Ungureanu,[1] Marian Silviu Poboroniuc,[2] and Alin Moldoveanu[3]

[1]*Computer Engineering Department, "Gheorghe Asachi" Technical University of Iasi, Iasi, Romania*
[2]*EUEDIA Department, "Gheorghe Asachi" Technical University of Iasi, Iasi, Romania*
[3]*Computer Engineering Department, Politehnica University of Bucharest, Bucharest, Romania*

Correspondence should be addressed to Robert Gabriel Lupu; lupu.robert@gmail.com

Received 22 September 2017; Accepted 14 February 2018; Published 5 April 2018

Academic Editor: Evdokimos I. Konstantinidis

In recent years, the assistive technologies and stroke rehabilitation methods have been empowered by the use of virtual reality environments and the facilities offered by brain computer interface systems and functional electrical stimulators. In this paper, a therapy system for stroke rehabilitation based on these revolutionary techniques is presented. Using a virtual reality Oculus Rift device, the proposed system ushers the patient in a virtual scenario where a virtual therapist coordinates the exercises aimed at restoring brain function. The electrical stimulator helps the patient to perform rehabilitation exercises and the brain computer interface system and an electrooculography device are used to determine if the exercises are executed properly. Laboratory tests on healthy people led to system validation from technical point of view. The clinical tests are in progress, but the preliminary results of the clinical tests have highlighted the good satisfaction degree of patients, the quick accommodation with the proposed therapy, and rapid progress for each user rehabilitation.

1. Introduction

The worldwide statistics reported by World Health Organization highlight that stroke is the third leading cause of death and about 15 million people suffer stroke worldwide each year [1]. Of these, 5 million are permanently disabled needing long time assistance and only 5 million are considered socially integrated after recovering. Recovering from a stroke is a difficult and long process that requires patience, commitment, and access to various assistive technologies and special devices. Rehabilitation is an important part of recovering and helps the patient to keep abilities or gain back lost abilities in order to become more independent. Taking into account the depression installed after stroke, it is very important for a patient to benefit from an efficient and fast rehabilitation program followed by a quick return to community living [2]. In the last decade, many research groups are focused on motor, cognitive, or speech recovery after stroke like Stroke Centers from Johns Hopkins Institute [3], ENIGMA-Stroke Recovery [4], or StrokeBack Consortium funded by European Union's Seventh Framework Programme [5]. Important ICT companies bring a major contribution to the development of technologies and equipment that can be integrated into rehabilitation systems. For example, *Stroke Recovery with Kinect* is a research project to build an interactive and home-rehabilitation system for motor recovery after a stroke based on Microsoft Kinect technology [6].

In the last years, the virtual reality (VR) applications received a boost in development due to VR headset prices that dropped below $1000, allowing them to become a mass-market product [7]. The VR was and still is especially used for military training or video games to provide some sense of realism and interaction with the virtual environment to its users [8]. Now it attracts more and more the interest of physicians and therapist which are exploring the potential of VR headset and augmented reality (AR) to improve the neuroplasticity of the brain, to be used in neurorehabilitation and treatment of motor/mental disorders [9]. However, considering the diversity of interventions and methods used, there is no evidence that VR therapy alone can be efficacious compared with other traditional therapies for a particular type of impairment [10]. This does not mean that the potential of VR was overestimated and the results are not the ones that were expected. The VR therapy must be complemented with other forms of rehabilitation technologies like robotic therapy, brain computer interface (BCI) and functional electrical stimulation (FES) therapy, and nevertheless traditional therapy to provide a more targeted approach [11].

*Saebo*VR is a virtual rehabilitation system exclusively focusing on activities of daily living and uses a virtual assistant that appears on the screen to educate and facilitate performance by providing real-time feedback [12]. The neurotechnology company MindMaze has introduced MindMotion PRO, a 3D virtual environment therapy for upper limb neurorehabilitation incorporating virtual reality-based physical and cognitive exercise games into stroke rehabilitation programs [13]. At New York Dynamic Neuromuscular Rehabilitation, the CAREN (Computer Assisted Rehabilitation Environment) based on VR is currently used to treat patients poststroke and postbrains injuries [14]. EVREST Multicentre has achieved remarkable results regarding the use of VR exercises in stroke rehabilitation [15].

Motor imagery (MI) is a technique used in poststroke rehabilitation for a long time ago. One of its major problems was that there was not an objective method to determine whether the user is performing the expected movement imagination. MI-based BCIs can quantify the motor imagery and output signals that can be used for controlling an external device such as a wheelchair, neuroprosthesis, or computer. The FES therapy combined with MI-based BCI became a promising technique for stroke rehabilitation. Instead of providing communication, in this case, MI is used to induce closed-loop feedback within conventional poststroke rehabilitation therapy. This approach is called paired stimulation (PS) due to the fact that it pairs each user's motor imagery with stimulation and feedback, such as activation of a functional electrical stimulator (FES), avatar movement, and/or auditory feedback [16]. Recent research from many groups showed that MI can be recorded in the clinical environment from patients and used to control real-time feedback and at the same time, they support the hypothesis that PS could improve the rehabilitation therapy outcome [17–21].

In a recent study, Irimia et al. [22] have proved the efficacy of combining motor imagery, bar feedback, and real hand movements by testing a system combining a MI-based BCI and a neurostimulator on three stroke patients. In every session, the patients had to imagine 120 left-hand and 120 right-hand movements. The visual feedback was provided in form of an extending bar on the screen. During the trials where the correct imagination was classified, the FES was activated in order to induce the opening of the corresponding hand. All patients achieved high control accuracies and exhibited improvements in motor function. In a later study, Cho et al. [23] present the results of two patients who performed the BCI training with first-person avatar feedback. After the study, both patients reported improvements in motor functions and both have improved their scores on Upper Extremity Fugl-Meyer Assessment scale. Even if the number of patients presented in these two studies is low, they support the idea that this kind of systems may bring additional benefits to the rehabilitation process outcome in stroke patients.

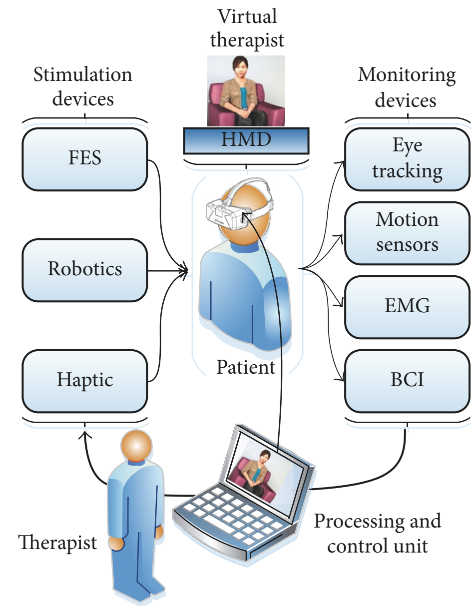


Figure 1: TRAVEE system architecture.

2. General System Architecture

The BCI-FES technique presented in this paper is part of a much more complex system designed for stroke rehabilitation called TRAVEE [24], presented in Figure 1. The stimulation devices, the monitoring devices, the VR headset, and a computer running the software are the main modules of the TRAVEE system. The stimulation devices help the patient to perform the exercises and the monitoring devices are used to determine if the exercises are executed properly, according to the proposed scenarios. Actually, the TRAVEE system must be seen as a software kernel that allows defining a series of rehabilitation exercises using a series of USB connectable devices. This approach is very useful because it offers the patient the options to buy, borrow, or rent the abovementioned devices according to his needs and after connection, the therapist may choose the suitable set of exercises.

The TRAVEE system is based on a new and promising rehabilitation concept which implies the augmented/magnified feedback of the movement of the impaired limb and can be successfully applied especially in the early stages of the rehabilitation therapy in order to close the loop that may trigger the mirror neurons [25]. These mirror neurons intermediate learning, indirectly controlling the brain plasticity and the technique is known as mirror therapy for stroke rehabilitation [26]. Despite the advantages of mirror therapy in comparison with other standard techniques, some disadvantages are obvious: it is difficult to explain to a patient how the mirror helps him: monotony, the patient's condition

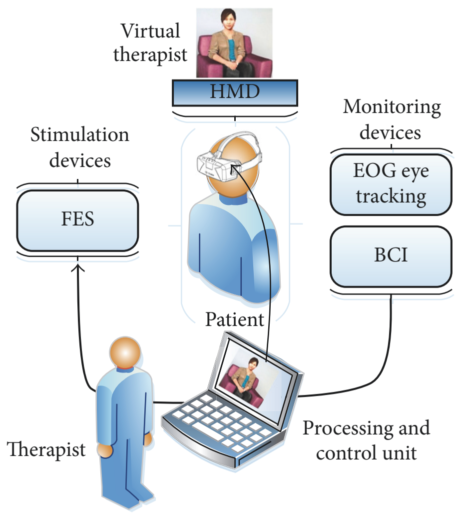


Figure 2: The BCI-FES TRAVEE subsystem.

Figure 3: The hand rehabilitation exercise.

Figure 4: Patient executing a rehabilitation exercise.

and position, the lack of challenging task, and so on. [27]. By replacing the physical mirror with a VR headset the patient has the same visual feedback that is needed to close the loop that triggers the mirror neurons but without disadvantages of the mirror therapy mentioned above. Once the patient is immersed in the virtual world he is no longer a disabled person and this has a good impact on patient's self-esteem. Within the TRAVEE project, encouraging results were obtained for the development of a virtual reality system for poststroke recovery using an inertial movement unit, a glove with sensors, a Myo Armband with electromyography sensors, and an Oculus Rift headset [28]. An alternate implemented system contains a Leap Motion device for patient's limbs movements monitoring, a VR headset, and a haptic module attached to patient's arm also offering better results than standard therapy methods [29].

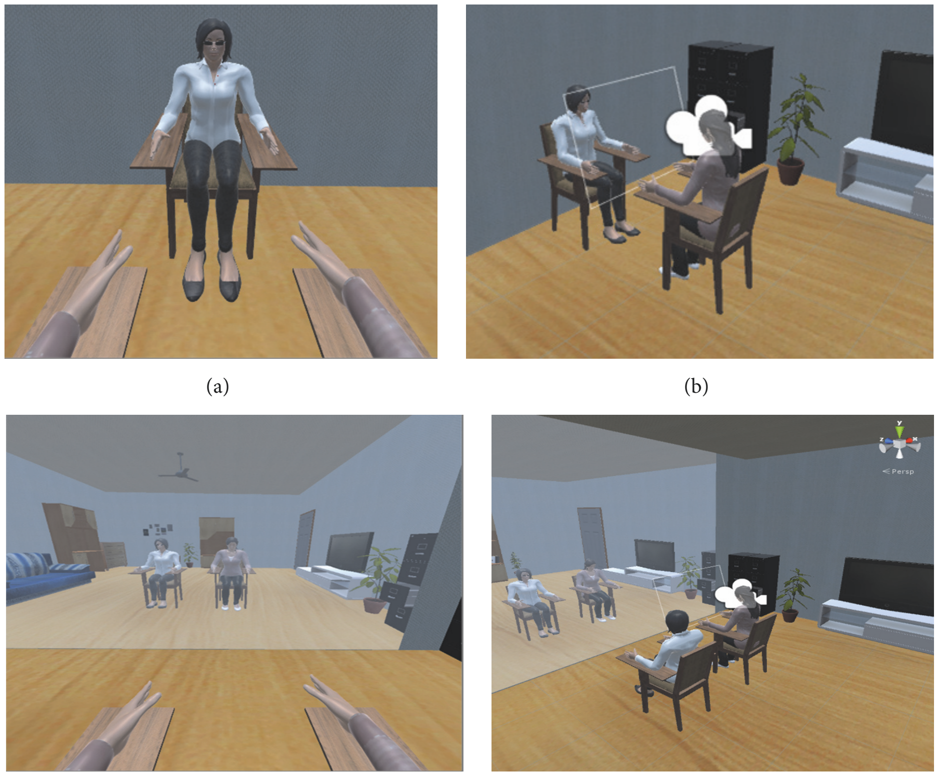
3. Materials and Methods

For the current study, the BCI-FES TRAVEE subsystem is composed of FES as stimulation device, BCI and an electrooculography (EOG) system as monitoring devices, Oculus Rift as VR headset, and a laptop, Figure 2.

The rehabilitation exercise was focused on flexion and extension of hand and fingers (Figure 3). The patient is seated in a wheelchair or normal chair. The FES electrodes are mounted on extensors muscles of both hands as shown in Figure 3 and the FES software module is started in order to determine the FES parameters (intensity and timings of the current impulse: rising, front, and falling). Then, the EOG electrodes and EEG helmet are mounted and the correct acquisition of the signals is verified. Before attaching the VR headset, the therapist sits in front of the patient explaining what he will see by showing him the following: the virtual therapist will raise the hand like in Figure 3 (the left hand of the therapist is the right hand of the patient); a big arrow will appear on the upper left or right of the screen depending on virtual therapist indications and the patient will also hear sounds from the left or the right. After explanations, the VR headset is mounted on (Figure 4), EOG system is calibrated, and the recovery exercise may begin, but not before the real therapist tells the patient that he has the possibility of choosing between two views: front view (the virtual therapist is located in front of the patient) or mirror view (the virtual therapist is located on the left side and a mirror is in front of them, like in a dance room) presented in Figure 5.

For the EOG calibration, a red spot appears for 2.5 seconds on a white background displayed on the VR system in different places, in the following order: center, upper right, center, upper left, center, lower left, center, lower right, and center. The user has to gaze at the spot in each location. The calibration is very important for an accurate calculation of the gaze points (eye tracking) during the tests.

In order to provide VR and FES feedback according to the patient's imagined movement, a set of spatial filters and classifier have to be created [22]. First, we are recording 4 runs of training data. Each run consists of 20 right- and 20 left-MI trials, in a random order. We use the trial time course and signal processing algorithms presented in [22]. Each trial lasts 8 seconds. At second 2 a beep informs the user about the upcoming cue. At second 3, the cue is presented and marks the moment when the user has to start imagining the movement shown by the virtual therapist until the end of the trial. While recording the test data, starting with second 4.25, the user sees the virtual hand indicated by the cue moving, and at the same time, the neurostimulator induces the patient's corresponding hand opening. After the spatial filters and classifier are created, we are recording 2 more runs, where the VR and FES feedback are provided to the patient between seconds 4.25 and 8 of each trial only if the classification result is correct. By comparing every sample of the classification result with the presented cue for each



(a) (b) (c) (d)

Figure 5: The VR environment in which the patient is immersed: (a) and (c) patient views; (b) and (d) world views; (a) the therapist in front of the patient; (c) the therapist on the left side of the patient with mirror in the front.

trial during the last 2 runs, we are calculating a control error rate course for that session. Except the first session, while recording the 4 train data runs, we are using the set of spatial filters and classifier calculated in the previous session of that patient only if the control error rate for that session was smaller than 20%.

4. EEG and EOG Recording

The BCI-FES subsystem consists of a 16-channel biosignal amplifier (g.USBamp, g.tec medical engineering GmbH) and an 8-channel neurostimulator (MOTIONSTIM8, KRAUTH+ TIMMERMANN GmbH). The EEG signals are collected from 12 positions over the sensorimotor areas according to the 10–20 International System, as seen in Figure 6(a). The last four channels are used in differential mode to record the vertical and horizontal EOG. Figure 6(b) presents the EOG electrodes position of the subject's head. The EEG and EOG data are sampled at 256 Hz and notch-filtered for excluding the 50 Hz noise. The EEG data are bandpass filtered between 8 and 30 Hz and then fed to the processing algorithm that performs spatial filtering with the Common Spatial Patterns (CSP) method [30, 31] and Linear Discriminant Analysis (LDA) classification [22, 32]. The EOG data are filtered with a moving average filter in order to calculate the average of the last 128 samples.

To acquire EOG signals the same EEG device was used but from all the EEG electrodes of the gTec–g.USBamp, 4 of them were used for EOG signals. The eye tracking is necessary because patient needs constant motivation and attention during training/recovering session from a therapist. In fact, after a while, the patient does not pay attention any more, is falling asleep, or is looking at/thinking of something else. By using the electrooculography (EOG) based eye tracking, the system is able to determine if the patient is concentrated and warns the patient if he is not. Figure 7 presents the output of the implemented algorithm for detecting the gaze point of the subject on the image in front of him. Figure 7(a) shows the processed HEOG and VEOG while Figure 7(b) displays the movement of the gaze point based on HEOG and VEOG.

5. Technical and Clinical Testing

The online signal processing and classification of the EEG signals were done by using the Common Spatial Patterns 2 class BCI Simulink model provided by g.tec medical engineering GmbH and the offline analysis of the data was done using g.BSanalyze software provided by the same company. For the EOG processing we developed a Simulink block containing an algorithm that processes the EOG signals and outputs the x-y gaze normalized coordinates with respect to the center point of the image displayed on the VR system. The whole

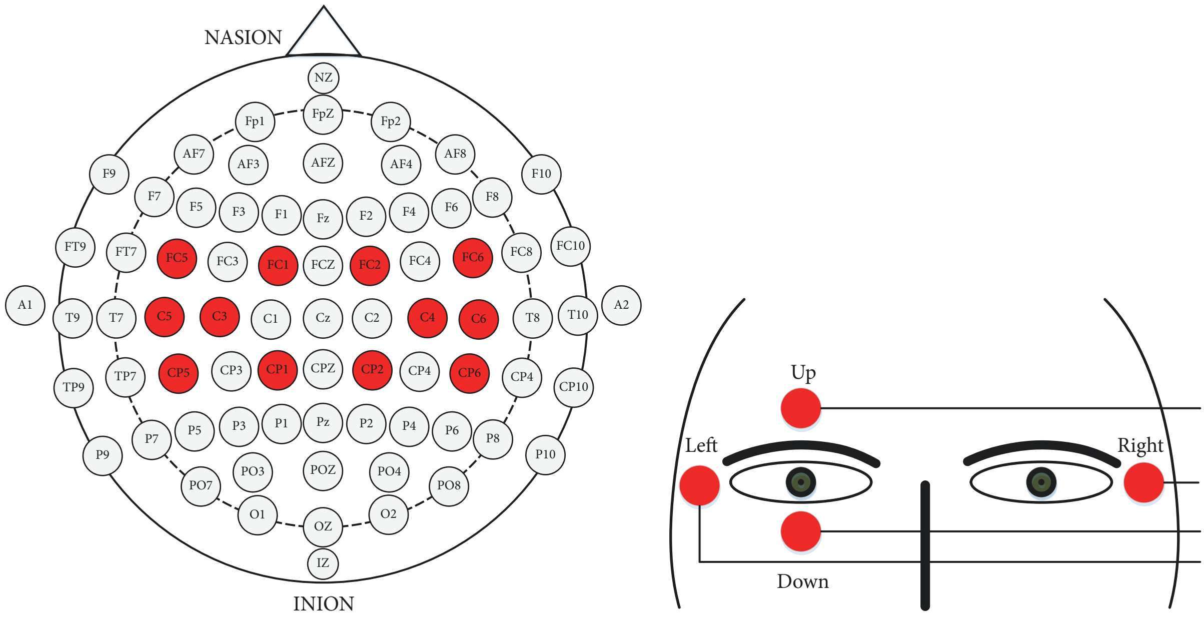


Figure 6: (a) EEG electrodes positions according to the 10–20 International System; (b) EOG electrodes displacement.

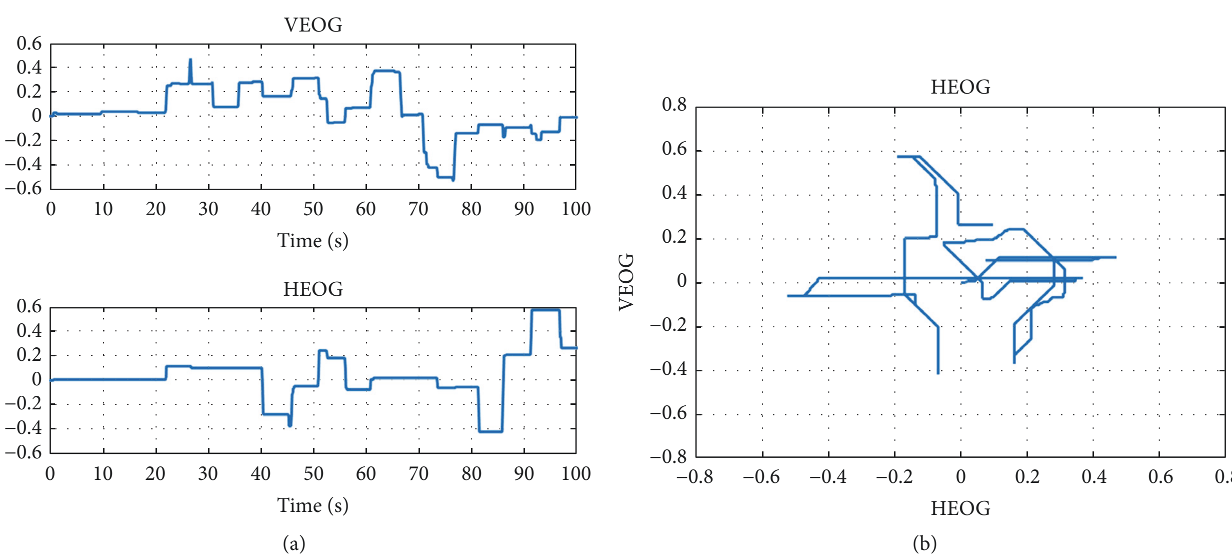


Figure 7: (a) HEOG and VEOG recorded for 100 seconds; (b) the gaze position on the image during 100 seconds of recording.

system was first tested on 3 healthy people and then some fine tunings were done based on their suggestions in order to get high accuracy and a good repeatability coefficient. All three-healthy people achieved low control error rates, comparable to the ones presented by Ortner and colleagues in [33].

Before starting the tests on patients within clinical environment, this study was approved by the institutional review board of the National Institute of Rehabilitation, Physical Medicine and Balneoclimatology from Bucharest, Romania, and each patient signed informed consent and an authorization for videos and photographs release before starting the study. The general clinical profile of the patients included in the study was afebrile, aware, temporospatial oriented, and cardiorespiratory balanced, without digestive or reno-urinary complains, with poststroke central neuromotor syndrome. From the whole patients, one-third was women and two-thirds were men, with ages between 52 and 79 years old. The inclusion criteria was stable neurological status; stable consciousness state; significant and persistent neuromotor deficit; disability for at least two of the following: mobility, self-help capacity, communication, sphincter control, deglutition; sufficient cognitive functions to allow learning; communication ability; sufficient physical exercise tolerance.

The clinical tests are in progress and until this moment the proposed system was tested on 7 patients. Each of them performed three training sessions, and all of them were able

Table 1: Mean and minimal control error rate values for seven patients.

Subject	Session	Mean error [%]	Minimal error [%]
S1	1	20.62	5.48
	2	20.62	7.11
	3	26.48	19.70
S2	1	23.96	11.97
	2	24.60	14.10
	3	28.83	21.00
S3	1	33.56	22.78
	2	37.00	21.35
	3	35.58	29.51
S4	1	32.58	24.77
	2	31.54	24.61
	3	37.21	26.22
S5	1	18.50	7.36
	2	19.72	10.72
	3	20.80	9.45
S6	1	19.20	6.37
	2	19.25	7.68
	3	19.58	1.95
S7	1	28.19	15.00
	2	25.53	13.56
	3	21.91	5.13
Mean values		*25.96*	*14.56*

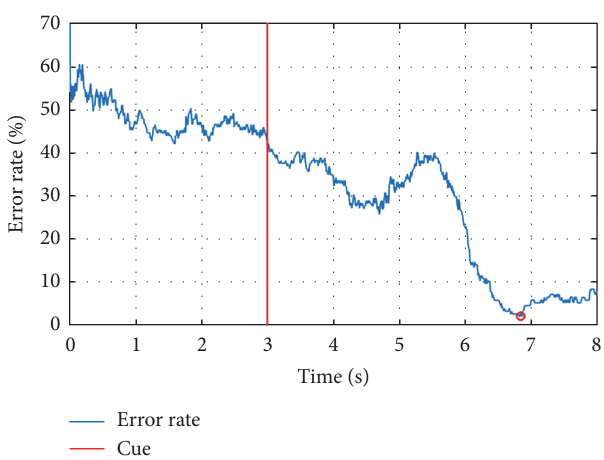


Figure 8: The error rate in time for subject S6, session 3.

to achieve a low control error rate over the whole system. Table 1 presents the mean and minimal control error rate achieved by each patient. The mean error rate is calculated as the mean of the errors for each time point between seconds 4.25 and 8 of the last 2 runs. Figure 8 presents the error rate in time for subject S6, session 8, when he achieved the lowest control error rate, indicated by the red circle at second 6.8.

Except for subjects S3 and S4, all patients exhibited control error rates lower than 20% in at least one session. At this time of the study, it is premature to make evaluations of the rehabilitation outcome of the patients, but, based on their feedback after each session, the VR system makes them remain focused on the task that they have to perform, and they see everything like an interactive game. The fact that they are cognitively involved in this task, unlike having a passive or bored attitude, obviously brings additional benefits to rehabilitation process outcome.

At the beginning, it was difficult for the patients to understand how to concentrate on imagining the movement of their impaired limb as part of the rehabilitation exercise. For those with a low-level education, it was unclear how such a concentration effort regarding their limb movement will help them. This was observed especially when the system was used only with BCI module without VR. The indications on what they had to do were very poor in information (just a simple sound and an arrow to indicate left or right). Also, the activity around the patient disturbed him very easily from imagining the movement. The patients needed around 5 training sessions in order to learn how to imagine the movements and to obtain a good neurofeedback. By adding VR, the number of training sessions was decreased to one or (very rarely) two.

Analyzing the questionnaires, it was concluded that the average user satisfaction was around 3, the answers being highly influenced by the patients' understanding of the rehabilitation therapy because most of them expected to recover themselves based on the therapist's activity and not to be consciously involved in the rehabilitation process. That depends also on the education degree. However, the overall patients' impression was that they felt and saw an encouraging improvement in recovering after using the proposed system.

For the next months, we plan to organize two groups of patients: a test group and a control group. The test group will perform up to 25 sessions of training with the system, while the control group will perform only classical rehabilitation therapy. When finishing the study, the results will be compared between groups and a statistical analysis will be performed on the results to see if the test group function improvements are statistically and significantly higher than the ones of the control group.

6. Conclusions

In this paper, a BCI-FES system for stroke rehabilitation is presented. Besides stimulation device, the BCI and EOG systems supervise how exercises are performed and the patient's commitment and Oculus Rift headset facilitates the patient's immersion in VR. By using this system, the patient is not distracted by the real environment or by events around him. He is just immersed in VR where the virtual therapist tells and shows him how to perform every exercise and a red big arrow is shown every time. The patient is focused most of the time, but if he loses his concentration the eye tracking system detects this and gives a warning.

The technical performances were validated by testing the system on healthy persons with good knowledge in assistive technologies. The healthy people achieved low control error rates, comparable to the ones reported in the literature.

The clinical tests are in progress, but the preliminary ones are very encouraging regarding fast accommodation and satisfaction of each patient. This approach of combining VR and BCI and FES facilities can effectively speed up the rehabilitation period and increase the users' optimism and the desire to exercise and recover lost skills. By involving the brain via BCI and VR the system proved to be more effective than the standard techniques.

The clinical tests last for several months for a significant number of subjects but once these will be completed the Likert questionnaires and technical files of all subjects will be analyzed.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the Romanian National Authority for Scientific Research (UEFISCDI), Project 1/2014 Virtual Therapist with Augmented Feedback for Neuromotor Recovery (TRAVEE).

References

[1] Stroke Statistics, The Internet Stroke Center, http://www.strokecenter.org/patients/about-stroke/stroke-statistics/, last visit September 2017.

[2] Recovering After a Stroke: A Patient and Family Guide, http://www.strokecenter.org/wp-content/uploads/2011/08/Recovering-Aftera-Stroke.pdf, last visit September 2017.

[3] Johns Hopkins Institute - Strock Centers, http://www.hopkinsmedicine.org/neurology_neurosurgery/centers_clinics/cerebrovascular/stroke/.

[4] ENIGMA-Stroke Recovery, http://enigma.ini.usc.edu/ongoing/enigma-stroke-recovery/, last visit September 2017.

[5] StrokeBack Project, http://www.strokeback.eu/project.html, last visit September 2017.

[6] Stroke recovery gets a boost from Kinect last visit September, http://www.microsoft.com/en-us/research/blog/stroke-recovery-gets-a-boost-from-kinect/.

[7] T. Bradshaw, "Virtual Reality gets its mass-market headset on, Financial Times," https://www.ft.com/content/f8087e6e-8c66-11e6-8aa5-f79f5696c731 last visit.

[8] A. Lele, "Virtual reality and its military utility," *Journal of Ambient Intelligence and Humanized Computing*, vol. 4, no. 1, pp. 17–26, 2013.

[9] K. Laver, S. George, S. Thomas, J. E. Deutsch, and M. Crotty, "Virtual reality for stroke rehabilitation: an abridged version of a Cochrane review," *European Journal of Physical and Rehabilitation Medicine*, vol. 51, no. 4, pp. 497–506, 2015.

[10] B. H. Dobkin and A. Dorsch, "New evidence for therapies in stroke rehabilitation.," *Current Atherosclerosis Reports*, vol. 15, no. 6, p. 331, 2013.

[11] W.-P. Teo, M. Muthalib, S. Yamin et al., "Does a combination of virtual reality, neuromodulation and neuroimaging provide a comprehensive platform for neurorehabilitation?—A narrative review of the literature," *Frontiers in Human Neuroscience*, vol. 10, article no. 284, 2016.

[12] Benefits of Virtual Reality for Stroke Rehabilitation, http://www.saebo.com/benefits-virtual-reality-stroke-rehabilitation/, last visit September 2017.

[13] L. Panjwani, Virtual Reality Therapy Designed to Help Stroke Patients Recover last visit September, http://www.rdmag.com/article/2017/08/virtual-reality-therapy-designed-help-stroke-patients-recover.

[14] Virtual Reality in Stroke Rehabilitation at NYDNR, https://nydnrehab.com/treatment-methods/neurorehab/virtual-reality-in-stroke-rehabilitation/.

[15] Stroke Outcomes Research Canada, sorcan.ca/current-projects/,.

[16] N. Sabathiel, D. C. Irimia, B. Z. Allison, C. Guger, and G. Edlinger, "Paired associative stimulation with brain-computer interfaces: A new paradigm for stroke rehabilitation," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Preface*, vol. 9743, pp. 261–272, 2016.

[17] K. K. Ang, C. Guan, K. S. G. Chua et al., "A large clinical study on the ability of stroke patients to use an EEG-based motor imagery brain-computer interface," *Clinical EEG and Neuroscience*, vol. 42, no. 4, pp. 253–258, 2011.

[18] F. Pichiorri, F. De Vico Fallani, F. Cincotti et al., "Sensorimotor rhythm-based brain-computer interface training: The impact on motor cortical responsiveness," *Journal of Neural Engineering*, vol. 8, no. 2, Article ID 025020, 2011.

[19] R. Ortner, D. C. Irimia, J. Scharinger, and C. Guger, "A motor imagery based brain-computer interface for stroke rehabilitation," *Stud Health Technol Inform*, vol. 181, pp. 319–323, 2012.

[20] S. R. Soekadar, N. Birbaumer, M. W. Slutzky, and L. G. Cohen, “Brain-machine interfaces in neurorehabilitation of stroke,” *Neurobiology of Disease*, vol. 83, pp. 172–179, 2015.

[21] A. Remsik, B. Young, R. Vermilyea et al., “A review of the progression and future implications of brain-computer interface therapies for restoration of distal upper extremity motor function after stroke,” *Expert Review of Medical Devices*, vol. 13, no. 5, pp. 445–454, 2016.

[22] D. C. Irimia, M. S. Poboroniuc, R. Ortner, B. Z. Allison, and C. Guger, “Preliminary results of testing a BCI-controlled FES system for post-stroke rehabilitation,” in *Proceedings of the 7th Graz Brain-Computer Interface Conference 2017, September 18th – 22nd*, Graz, Austria, 2017.

[23] W. Cho, A. Heilinger, R. Xu et al., “Hemiparetic Stroke Rehabilitation Using Avatar and Electrical Stimulation Based on Non-invasive Brain Computer Interface,” *International Journal of Physical Medicine & Rehabilitation*, vol. 05, no. 04, 2017.

[24] TRAVEE, Virtual Therapist with Augmented Feedback for Neuromotor Recovery, http://travee.upb.ro/, last visit september 2017.

[25] D. Cinteza, “Modern Concepts of Recovery and Rehabilitation - CNS Affections (Mirror System),” *Balneo-Research Journal*, vol. 3, 2012.

[26] M. E. Michielsen, R. W. Selles, J. N. van der Geest et al., “Motor recovery and cortical reorganization after mirror therapy in chronic stroke patients: a phase II randomized controlled trial,” *Neurorehabilitation and Neural Repair*, vol. 25, no. 3, pp. 223–233, 2011.

[27] T. Muzaffar, R. K. Wadhwa, B. Diganta, N. Laisram, and SY. Kothari, “Evaluation of Mirror Therapy for Upper Limb Rehabilitation in Stroke,” in *Vol 24(3): 63-9*, vol. 24, p. 63, September, IJPMR, 2013.

[28] R. G. Lupu, F. Ungureanu, and A. Stan, “A virtual reality system for post stroke recovery,” in *Proceedings of the 20th International Conference on System Theory, Control and Computing, ICSTCC 2016*, pp. 300–305, Romania, October 2016.

[29] R. G. Lupu, N. Botezatu, F. Ungureanu, D. Ignat, and A. Moldoveanu, “Virtual reality based stroke recovery for upper limbs using leap motion,” in *Proceedings of the 20th International Conference on System Theory, Control and Computing, ICSTCC 2016*, pp. 295–299, Romania, October 2016.

[30] J. Müller-Gerking, G. Pfurtscheller, and H. Flyvbjerg, “Designing optimal spatial filters for single-trial EEG classification in a movement task,” *Clinical Neurophysiology*, vol. 110, no. 5, pp. 787–798, 1999.

[31] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Müller, “Optimizing spatial filters for robust EEG single-trial analysis,” *IEEE Signal Processing Magazine*, vol. 25, no. 1, pp. 41–56, 2008.

[32] S. Lemm, B. Blankertz, T. Dickhaus, and K.-R. Müller, “Introduction to machine learning for brain imaging,” *NeuroImage*, vol. 56, no. 2, pp. 387–399, 2011.

[33] R. Ortner, J. Scharinger, A. Lechner, and C. Guger, “How many people can control a motor imagery based BCI using common spatial patterns?” in *Proceedings of the 7th International IEEE/EMBS Conference on Neural Engineering (NER ’15)*, pp. 202–205, April 2015.


Journal of Engineering
The Scientific World Journal
International Journal of Rotating Machinery
Journal of Sensors
Advances in Multimedia
Advances in Civil Engineering
Hindawi
Submit your manuscripts at
www.hindawi.com
Journal of Control Science and Engineering
Journal of Robotics
Journal of Electrical and Computer Engineering
Advances in OptoElectronics
VLSI Design
International Journal of Navigation and Observation
Modelling & Simulation in Engineering
International Journal of Aerospace Engineering
International Journal of Chemical Engineering
International Journal of Antennas and Propagation
Active and Passive Electronic Components
Shock and Vibration
Advances in Acoustics and Vibration

Virtual Reality Based Stroke Recovery for Upper Limbs Using Leap Motion

Robert Gabriel Lupu, Nicolae Botezatu, Florina Ungureanu, Daniel Ignat
Computer Department
"Gheorghe Asachi" Technical University of Iasi
Iasi, Romania
robert@cs.tuiasi.ro

Alin Moldoveanu
Computer Department
University POLITEHNICA of Bucharest
Bucharest, Romania
alin.moldoveanu@cs.pub.ro

***Abstract*— In this paper a virtual reality based stroke recovery system for upper limbs is described. The patient is immersed in the virtual environment through the use of an Oculus Rift device and interacts with the system by using Leap Motion as input device. The patient experience is enriched by providing haptic feedback when interacting with objects in the virtual environment. The recovery therapy relies on TRAVEE system's state-of-the-art paradigm of using augmented feedback during early recovery stages to create new recovery possibilities.**

Keywords—stroke; recovery; leapmotion; virtual reality, augmented feedback

I. Introduction

According to the American Stroke Association, only in 2010 the prevalence of stroke was around 33 million people worldwide with nearly half of them heaving the first stroke. In 2013 it become the second leading cause of death [1]. From 85% of stroke survivors, only 10% recover completely and 25% with minor impairments. The rest remains with severe impairments (40%) and they require special care or worse, special care in nursing homes (10%) [2]. Recovery involves relearning motor skills and is possible due to brain neuroplasticity (brain ability to reorganize itself) [3]. A successful stroke recovery depends firstly on the amount of damage on the brain and secondly by the skills and experience of an interdisciplinary rehabilitation team [4]. In order to obtain the maximum results, the recovery exercises rely on visual feedback. The patient needs to see the movement of the impaired limb and because in the early stages of recovery, this is not possible, a mirror is used [5]. The patient thinks and moves his healthy limb and what he sees in the mirror trick his brain in believing that his impaired limb moves as well. In time, however, the patient loses motivation; gets tired because of bad position and no challenging tasks are available [6]. These problems can be overcome by the use of virtual reality (VR) and motion trackers technologies [7]. Despite the differences between VR and non VR scenarios in terms of moving [8], it has big potential for stroke recovery by improving efficacy and patient motivation.

This paper presents one of many approach of a bioinformatic integrated system (TRAVEE) that helps the patients to recover after stroke.

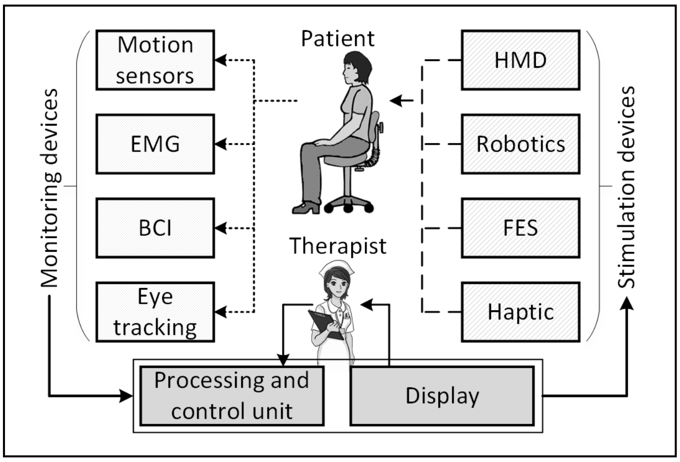


Fig. 1. TRAVEE system

The system relies on a hardware and software architecture and rehabilitation concept with contributions beyond the state of the art [9]. The ideas that underline this new rehab concept are: the use of augmented and magnified feedback; the use of virtual therapist as guidance for exercises [9]. This approach brings the recovery therapy for upper limbs to a new level, with promising results as healthcare professionals considered. Another idea promoted by the TRAVEE rehab concept is the customization of the system according to patient recovery needs. In this way, the patient can borrow, rent or buy the devices (i.e. off-the-shelf commercial devices) that he needs for his recovery exercises and when the session ends, he can return them back or even sell them if the recovery is complete for that motor function according to specialist. He can continue with other devices in the same manner if the recovery therapy requires.

There are two categories of devices used by the TRAVEE system: devices for patient monitoring and stimulation and devices for data processing. The devices from the first category are plug-and-play and optional, the system can use any combination of them. The devices from the second category are mandatory because it guides the recovery exercises.

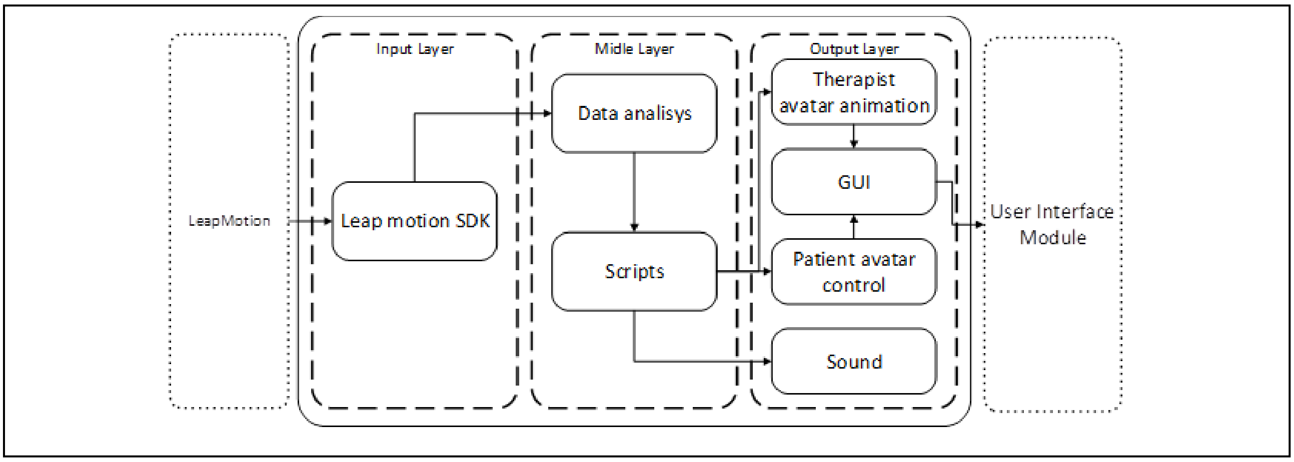


Fig. 2. Processing and control unit architecture

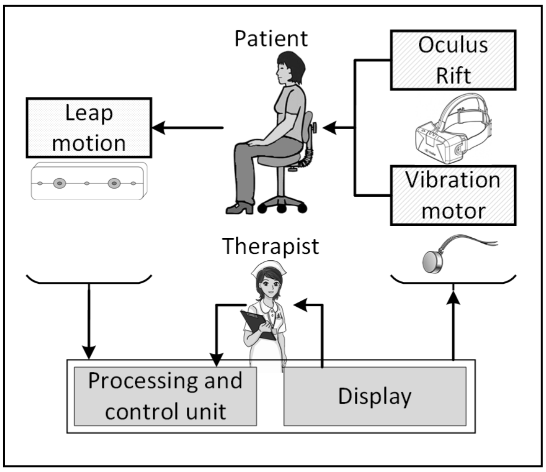


Fig. 3. Custom TRAVEE system

From hardware design perspective the system must ensure in the same time continuous monitoring of the patient movements and patient stimulation. This must be done in accordance with the patient needs of recovery at that moment in order to keep the system cost as low as possible.

From a software design perspective, the system is a computer program that communicates with the corresponding software modules of the connected devices. In this way the therapist can load and define the recovery exercises adding or removing components as needed.

II. Custom System Architecture

Starting from already demonstrated effects of the mirror therapy [10], a custom system was configured in order to replace the mirror with a VR headset. The system contains a Leap Motion device for patient's upper limbs movements monitoring, a VR headset (Oculus Rift) to immerse the patient in a virtual reality environment and a haptic modules attached to patient's arm (Fig. 3). The virtual environment is built using Unity and is designed to resemble a familiar environment for the patient (e.g. the actual therapy room used by the patient).

Once a recovery exercise is configured and started, the *Processing and control unit* acquires real-time data, processes and stores it and updates the output stimulation equipment (i.e. VR headset and haptic subsystem). The therapist can monitor the exercise (i.e. view the Virtual Environment as the patient sees it through the VR headset) on the display of the *Processing and control unit*.

The VR scene contains both patient and therapist avatars seated at a table with their hands resting on it. In the real world the patient also sits at the table with his hands on the table, with the VR headset and leap motion mounted on his head (Fig. 4). The presence of the therapist avatar is needed in order to guide the recovery exercises and to motivate the patient.

The feedback for the patient is further augmented by using a haptic device (i.e. vibration motors) that stimulates the tactile sense on the arm when certain conditions occur.

A. Processing and control unit architecture

Figure 2 presents the structure of the Processing and control unit. It is based on a three tier paradigm which helps decouple the input, processing and output stages of the system. The input layer is represented by the services provided by the Leap Motion SDK. The acquired data are sent to the middle layer which takes care of the processing. It includes the Data analysis block which analyzes the input data (i.e. arm and finger parameters - position, rotation, length, thickness) and computes the amplitude of the movements. This layer also features a configuration block that dictates the behavior of the output based on scripted exercises. This is the main part that adapts the system and triggers the actions of the avatars based on the exercises described in section three. The output layer contains the animation modules for the therapist and for the patient: the Therapist avatar animation module contains predefined animations that match the exercises for the intended setup; the Patient avatar control module generates animation based on a wireframe arm model. The sound generation module is based on the Microsoft Speech Platform SDK. The scene is rendered using the Unity3D engine.

B. I/O components

Leap Motion is a commercial device built to track forearm, hand and finger movement. It contains two cameras and three IR LEDs. Based on these elements, the embedded processing unit generates and transmits a grayscale image to the PC using an USB connection. These images are further processed by through the use of the associated SDK.

In order to animate the patient's avatar, the data provided by the Leap Motion device is read and processed through the use of the *HandController* and *Frame* classes. The *HandController* class instantiates the communication with the device and the *Frame* class is used to access the information related to forearm, hand and finger position. Each frame contains a still image of the scene recorded by the device; the base class provides a series of functions for data access.

Currently, most games have limited visual output capabilities on a computer, tablet or phone display. Projects similar to the Oculus Rift one have the capability to change the way we interact with technology leading to a revolution in this field. The Oculus Rift device is composed of a headset equipped with a 7 inch display with a resolution of 1280x800 pixels (i.e. 640x800 pixels for each eye). The person equipped with the device can move in the 3D environment generated by the game application through the use of an orientation sensor. The control loop of the device reads the sensor, tracks the head movement and updates the VR perspective rendered on the display. This approach makes use of natural body movements and discards the use of auxiliary input devices (e.g. joystick, mouse).

The haptic module is a custom one based on our previous work [11] and consists of an aceMote v1.0 embedded platform that controls five ERM vibration motors. The motors are interfaced by the DRV2605L haptic drivers that integrate a licensed version of the TouchSense 2200 software from Immersion. This library contains haptic effects carefully designed and implemented that are ready to use. For this haptic driver, there are available 100 licensed haptic effects grouped into 7 software selectable libraries.

III. Patient setup for upper limbs recovery

On the first use, the patient is always assisted by the therapist, who describes the contents of the recovery session and configures the system accordingly. The patient sits on a chair with the arms resting on a table. The therapist sits at the same table, facing the patient and describes the exercises that the patient must follow (Fig.5 and Fig.6). Next, the therapist puts on the VR headset and starts the system which runs the same exercises as for the patient. In this stage the patient is requested the view the demo exercises, as the therapist sees them through the VR headset, on the system's monitor. In the next stage the headset is mounted on the patient and he is requested to move the head and look around in order to get used to the virtual environment. Once the patient feels comfortable enough the therapist starts the exercise. From this point on, the patient is requested to follow only the instructions of the VR avatar that resembles the therapist (Fig. 7 and Fig.8).

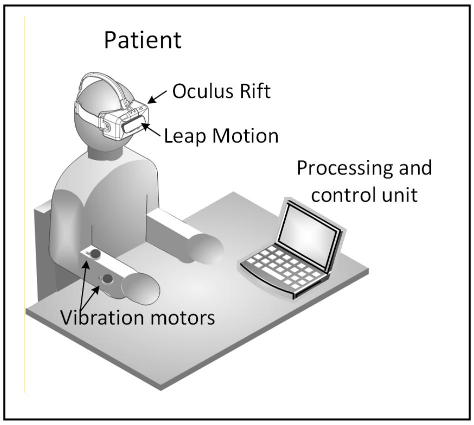


Fig. 4. Patient setup for right hand recovery exercise

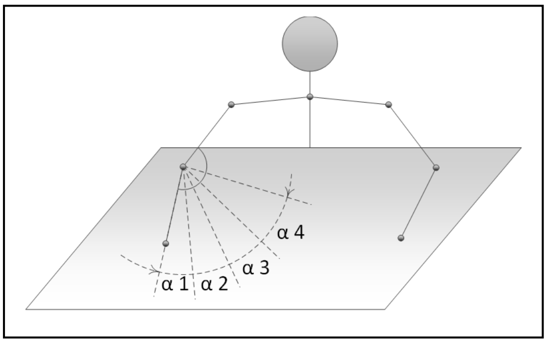


Fig. 5. Upper limbs recovery exercise without gravity

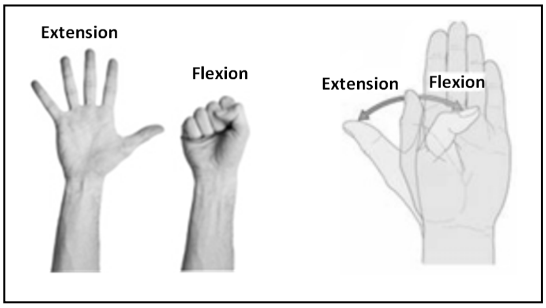


Fig. 6. Two hand recovery exercises

There are three exercises implemented, one for the arm (Fig. 5) and two for the hand (Fig. 6). The arm exercise starts with the forearm resting on the table, perpendicular to the body. The patient is required to slide the forearm on the table, back and forth, bringing it close to the body and moving it back to the start position (i.e. actively using the elbow joint).

In order to recover the hand mobility, two exercises were implemented: flexion and extension of the palm, and flexion and extension of the thumb (Fig. 6).

IV. Recovery Sessions

The neuromotor rehabilitation is divided in three session types, with mirror, augmented and real feedback. Each session type is repeated until the devised session objective is achieved.

Fig. 7. Patient and therapist avatars during exercises – general perspective

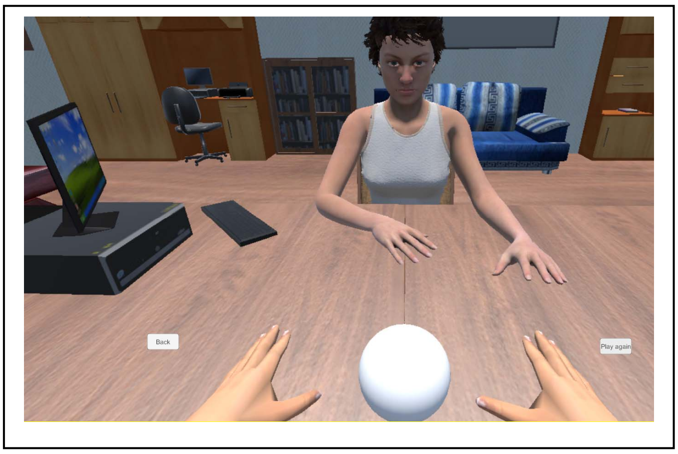


Fig. 8. Therapist avatar during exercises – patient perspective

The first type of session is employed immediately after the stroke, when the patient cannot move the paralysed arm. It uses mirror therapy by immersing the patient in the virtual environment: the patient can see both his hands moving, thus making him believe he has full control of the affected arm. During the exercise, the system is configured to track the movement of the healthy arm and to update the position of both arms of the avatar in the virtual environment. In this way the patient is under the impression that he moves both his arms. The visual feedback activates the mirror neurons in the premotor cortex. This is particularly important because all mirror neurons present a congruency between visual actions and motor responses.

After the patient has at least a weak control over his arm (i.e. one can visually observe the intention to move the arm), the second type of session is started. The sensitivity of the Leap Motion device is gradually set in order to detect the slightest intention of arm or finger movement. During the exercise, all movements executed by the patient are augmented by a factor that is directly related to the amplitude of the movement – for low amplitude movements the amplification factor will be at the maximum value and it will linearly decrease to an amplification of 1 for maximum arm or finger movement (i.e. the maximum amplitude movement that can be achieved by a healthy individual). The effect on the motor cortex is important because the patient realizes that he moves the hand at his will. The augmented feedback is extremely necessary because it transforms the visual information in knowledge. For all three types of exercises the maximum amplitude interval is divided in four, each subinterval having an amplification factor of 2.5, 2, 1.5 and 1 respectively. For example, for the arm movement exercise (Fig. 5) the maximum amplitude is 90 degrees. The subintervals of 22.5 degrees each, labeled $\alpha 1$ to $\alpha 4$, are associated to the above mentioned amplification factors.

The third session type is employed when the patient has regained partial control over his arm. The movements are no longer augmented and the patient is further motivated by assigning different tasks to him. The purpose of the VR is to motivate the patient through the proposed tasks and scenarios. In this stage the feedback is supplemented by the use of the haptic module.

In the setup for arm exercises, the patient pushes or hits a ball as hard as he can in order to move it and send it as far as possible. Of the five vibration motors, one is mounted on the side of the hand (i.e. between the thumb and the back of the hand) and the other four are mounted on the forearm at equal distance between the wrist and the elbow. The motors are selectively activated when the patient hand in the VR touches the ball. The intensity of the vibration is proportional to the velocity of the movement.

For the hand exercises a single vibration motor is placed in the palm. Based on the exercise type, it is activated when the fist is closed or when the thumb is in flexion.

During the last two recovery sessions, the therapist can set different values for the feedback movement amplitude. In this way the difficulty of the exercise is altered based on the patient's status. More, due to the fact that the movement can be tracked during the exercises, the therapist can observe the evolution and can determine the recovery degree of the patient.

V. Conclusions

This paper presented the design of an easily adaptable system for post-stroke rehabilitation. Based on the TRAVEE paradigm, it used virtual reality and haptics for augmented feedback during exercise sessions. Besides the actual system, three exercise types were defined (forearm, hand and thumb), each divided in three successive session types (mirror, augmented and real feedback).

The technical performances of the proposed system were extensively tested in terms of hardware and software performance. The test techniques and updated versions of the application have contributed to ensure the stability, accuracy, repeatability of the acquired data and data loss have not been reported. The software fulfilled the load, volume and stress testing. The analyses of the data generated by the application revealed a very good concordance between the real position of the limb/hand/palm and the calculated position.

Presently, pilot case studies are conducting to evaluate ease of use and efficacy from the patient's side and rehabilitation process. Preliminary tests suggest user acceptance of the technology. The project is in progress and in the next work package the assessment of its efficiency and usability in medical clinics or at home will be conducted by the Romanian National Institute of Rehabilitation, Physical Medicine and Balneoclimatology. The individuals after stroke with intact cognition and sitting balance are selected for the study and they do not receive any other intensive rehabilitation. The subjects will follow a recovery scheme used in similar studies [12]. Each patient should perform gradually the sets of exercises for a maximum of 30 minutes, 3 or 4 times a day, for at least 10 weeks. After 5 weeks and at the end of the period, well-known outcome measures will be performed: Motor Activity Log, Wolf Motor Function Test, Nine-Hole Peg Test and Chedoke Arm and Hand Activity Inventory [13]. A software module collects for each user the frequency and duration of use and indexes regarding patient's performance in comparison with avatar. Finally, the individual files will be merged and statistically analyzed by the use of R language.

Further work includes defining new exercise types and a long term usability study performed with stroke impaired subjects.

Acknowledgment

This work is supported by the Romanian National Authority for Scientific Research (UEFISCDI), Project 1/2014 Virtual Therapist with Augmented Feedback for Neuromotor Recovery (TRAVEE).

References

[1] American Stroke Association, Heart Disease, Stroke and Research Statistics At-a-Glance, https://www.heart.org/idc/groups/ahamah-public/@wcm/@sop/@smd/documents/downloadable/ucm_480086.pdf, last visit june 2016

[2] National Institute of Neurological Disorders and Stroke, Stroke Rehabilitation Information, march 29, 2016, http://www.ninds.nih.gov/disorders/stroke/stroke_rehabilitation.htm, last visit june 2016

[3] S. M. Cameron, K. L. Campbell, J.D. R. Colin, and L. A. Boyd, "Promoting neuroplasticity for motor rehabilitation after stroke: considering the effects of aerobic exercise and genetic variation on brain-derived neurotrophic factor", Phys Ther. 2013 December; 93(12): 1707–1716. Published online 2013 August 1. doi: 10.2522/ptj.20130053 PMCID: PMC3870490

[4] S. Knecht, S. Hesse, and P. Oster, "Rehabilitation after stroke", Dtsch Arztebl Int. 2011 September; 108(36): 600–606. Published online 2011 September 9. doi: 10.3238/arztebl.2011.0600 PMCID: PMC3183303

[5] G. Yavuzer, R. Selles, Sezer N., S. Sutbeyaz, J.B. Bussmann, F. Koseoglu, M.B. Atay, and H.J. Stam, "Mirror therapy improves function in subacute stroke: a randomized controlled trial", Arch Phys Med Rehabil 2008, 89:393-8

[6] T. Muzaffar, R.K. Wadhwa, D. Borah, N. Laisram, and S.Y. Kothari, "Evaluation of mirror therapy for upper limb rehabilitation in stroke", International Rehabilitation Forum Interscience Summit IRF2014, Suzhou International Expo Center, 16-10 oct. 2014, China

[7] A. Baldominos, Y. Saez, and C. García del Pozo, "An approach to physical rehabilitation using state-of-the-art virtual reality and motion tracking technologies", Conference on Health and Social Care Information Systems and Technologies, HCist 2015 October 7-9, 2015

[8] M. A. Just, P. J. Stapley, M. Ros, F. Naghdy, and D. Stirling, "A comparison of upper limb movement profiles when reaching to virtual and real targets using the Oculus Rift: implications for virtual-reality enhanced stroke rehabilitation", Proc. 10th Intl Conf. Disability, Virtual Reality & Associated Technologies Gothenburg, Sweden, 2–4 Sept. 2014

[9] S. Caraiman, A. Stan, N. Botezatu, P.C. Herghelegiu, R.G. Lupu, and A. Moldoveanu, "Architectural design of a real-time augmented feedback system for neuromotor rehabilitation", Proceedings of 2015 20th International Conference on Control Systems and Computer Science, CSCS 2015, University Politehnica of Bucharest; Romania; 27-29 May 2015

[10] Mirror Therapy, Heart & Stroke Foundation - Canadian Partnership for Stroke Recovery, http://www.strokengine.ca/intervention/mirror-therapy/, last visit june 2016

[11] A. Stan, N. Botezatu and R. G. Lupu, "The design of a scalable haptic system used for impaired people assistance", 2015 20th International Conference on Control Systems and Computer Science, Bucharest, 2015, pp. 863-866.

[12] C. E. Lang, M. D. Bland, R.R.Bailey, S. Y. Schaefer and R. L. Birkenmeier, "Assessment of upper extremity impairment, function, and activity after stroke: Foundations for clinic al decision making", Physical Therapy Faculty Publications, Paper 19, http://digitalcommons.wustl.edu/pt_facpubs/19,, 2013

[13] P.J. Standen, K. Threapleton, L. Connell, A. Richardson, D. J. Brown, S. Battersby, C. J. Sutton, and F. Platts, "Patients' use of a home-based virtual reality system to provide rehabilitation of the upper limb following stroke", Physical Therapy, vol. 95, pp. 350 – 359, March 2015

The 6th IEEE International Conference on E-Health and Bioengineering - EHB 2017
Grigore T. Popa University of Medicine and Pharmacy, Sinaia, Romania, June 22-24, 2017

Usability Assessment of Assistive Technology for Blind and Visually Impaired

Maria Dascalu[1], Alin Moldoveanu[1], Oana Balan[1], Robert Gabriel Lupu[2], Florina Ungureanu[2], Simona Caraiman[2]
[1] Department of Computer Science, University Politehnica of Bucharest,
Bucharest, Romania, maria.dascalu@cs.pub.ro, alin.moldoveanu@cs.pub.ro, oana.balan@cs.pub.ro
[2] Department of Computer Engineering, "Gheorghe Asachi" Technical University of Iasi
Iasi, Romania, robert@tuiasi.ro, fungurea@tuiasi.ro, sarustei@tuiasi.ro

***Abstract*— In this paper we describe the usability assessment of a system designed to help blind and visually impaired people to navigate and perceive the environment. The proposed system is based on sensory substitution, remapping the vision stimuli into audio and haptic ones. The goal of this study is to aid the development of the sensory substitution device (SSD) by understanding how the different choices in encoding and rendering the environmental information affects the user's perception and experience while using the system. The preliminary results are presented to show the usability and usefulness of the proposed system.**

Keywords*— *usability; visual impaired; assistive technology; virtual environment; sensory substitution.

I. Introduction

In 2010, the World Health Organization [1] reported 285 million of visually impaired, 246 million with low vision and 39 million of blinds worldwide.

For a person, the possibility to orient and move by himself relies on information gathered from the indoor/outdoor surroundings about position, direction and stationary/moving objects: walls, stairs, bumps and holes respectively, moving vehicles, people [2] etc. To gather this information, a healthy person relies mostly on the visual sense and barely on the hearing or haptic ones. But for a blind or visually impaired, independent orientation is a real challenge due to poor environmental information received through audio and haptic channels.

Due to so called "human auditory system", any healthy person can recognize and find the source of a sound. This is possible because the auditory system uses two classes of cues: monaural and binaural. The first-class cues let the system find/detect the sound source in median plane and whether the sound is coming from back or front. The second-class cues give information about sound source in horizontal plane [3].

The blind people and visually impaired (VIP) build their own spatial map based on surrounding or their own sounds. By tracking these sound sources, they are able to orient themselves. The second information channel, the haptic channel, might be used as alternative information channel for those with hearing impairment, or just as an empowering channel to gather more information, helping the user to build an improved surrounding map for better orientation [4].

SSDs can be used for building spatial surrounding maps in a much faster way, and with more information about static and moving objects. This kind of device is using special video cameras and converts/encodes the detected information in such a way that can be transmitted on the user available channel, audio and/or haptic. That implies the remapping of the visual information into audio/haptic information based on predefined algorithms.

II. State of the Art on Sensory Substitution Devices

Sensory substitution stands for remapping of stimuli of one sensory into stimuli for another sensory. Usually this is done to bypass a defective sense (vision, hearing). A blind or deaf person doesn't lose the ability to see or hear, but only the possibility to convey stimuli to the brain [5]. There are invasive and non-invasive methods for sensory substitution, the last one being mostly used by the SSD due to advance on human computer interaction technology. For blinds, we can speak about Auditory Visual substitution and Tactile Visual substitution [5].

Kai Wun et al. presented in [6] a SSD that assists the visual impaired to avoid obstacles. It is a head-mounted device with stereo camera mounted on eyeglasses to compute depth. In addition, the system is capable of live video streaming through 3G network based on witch a normal person gives indications.

The EyeMusic [7] tool developed by University of Jerusalem provides visual information through bone conductance headphones and based on 25x40 images supplied by a camera mounted on eyeglasses. It uses various timbres, pitches and notes to create a mental image of the visual scene in front of the user.

Michael Bujacz developed and algorithm for sonification of 3D scenes by mixing image processing methods with audio representation. It uses depth algorithms for object segmentation and builds stereoscopic images. For every detected object the sonification algorithm generates the corresponding sound [8].

978-1-5386-0358-1/17/$31.00 ©2017 IEEE

The VIBE project developed at Gipsa-Lab and LPNC from Grenoble proposed an audio guided system by generating sounds as summation of sinusoidal sounds produced by virtual sources (a set of pixels grouped in receptive fields). The receptive field state is the mean of gray levels in its area [9].

Stiles and Shimojo presented the vOICe device which is an auditory SSD that translates vertical position to frequency, left-right position to scan time (encoded in stereo), and brightness to sound loudness [10]. In fact, this device assists the blind by encoding a video stream into a sound pattern.

In the same research area, our project, the Sound of Vision (SOV) project, is aiming to design, implement and validate a wearable system to assist VIP by creating and conveying an auditory representation of the surroundings, continuously updated and delivered in real time to the user. In addition, haptics is used to enhance some relevant information. The SOV project, approaches and devices are widely presented on the project site [11] and in [12].

III. USABITLITY ASESSEMENT OF THE SOV DEVICE

The usability tests of the first prototype of the SOV device were performed with two purposes: (1) to test the prototype functionality and (2) to accommodate test-takers with the prototype. In order to do that without any physical risks and fear of doing mistakes, the system was tested in a virtual environment (VE). The VE simulates highly abstracted scenes from real world, e.g. using boxes of different dimensions. The first tests in VE were the ego-static tests (VE1) where the test-taker had analyzed a scene presented to him/her and to identify the objects properties based on the received audio and haptic information. The second category of VE tests were the ego-dynamic tests (VE2) where the user had the possibility to move in the virtual scene with the help of the computer keyboard. The goal for the test-taker was to apply the knowledge from VE1 to navigate and avoid stationary objects. Both test categories (VE1 and VE2) were carried out by test-takers using firstly the audio headset and then repeat most of the tests with the haptic vest as stated below:

- for audio, 8 testing tasks were available (7 in VE1 and 1 in VE2): in VE1, the user's abilities to identify the direction, the elevation, the width, the distance and the quantity of a generic object, the type of an object and also one's ability to manage in a complex scene (in which several objects of different dimensions were present) were tested; in VE2, the user's ability to hit/find one box was evaluated;
- for haptics, 8 tasks were available (7 in VT1 and 1 in VT2): in VE1, the user's ability to identify the direction, the width, the height, the distance and the quantity of a generic object, the type of an object and also one's ability to manage in a complex scene were evaluated; in VE2, the user's ability to hit/find one box was tested.

During test sessions, the tester and test-taker stood next to each other as it is presented in Fig. 1. For each task, the users had: (1) a baseline testing session, in which they tried to solve the tasks, without any training; (2) a training period, in which feedback was offered to them and (3) three sessions of self-testing. The complex task from VE1 and the task related to finding a box in VE2 did not contain the training part.

The tests took place in three different locations: University Politehnica of Bucharest (UPB), Technical University of Iasi (TUI), and the High School for Visually Impaired Persons from Targu Frumos, Romania.

Fifteen sighted (but blindfolded) and blind test-takers participated in both audio and haptics usability tests, having ages between 19 and 67, with an average of 33 years old. All the subjects were informed about the SOV project and the testing methodology and signed a consent form before the tests: in total, there were 8 blind and 7 sighted, 11 men and 4 women. Before starting the tests, audiometric check was performed, to evaluate if any hearing impairment exists; all the test-takers passed the check.

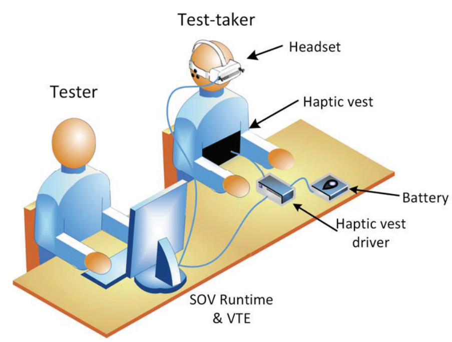


Figure 1. Environment settings tests

IV. ASSESSMENT RESULTS AND ANALYSIS

The results (minutes spent on a task and score) were registered for both audio and haptic tasks, as follows: baseline testing (B) - minutes and score [0%-100%], training (Tr) - minutes, first session of self-testing (ST1) - minutes and score [0%-100%], second session of self-testing (ST2), if necessary - minutes and score [0%-100%], third session of self-testing (ST3), if necessary - minutes and score [0%-100%], final testing (T) - minutes and score [0%-100%]. A task is considered passed if the user obtains over 75% in the final score in two out of three self-testing sessions.

The average results obtained by the 15 test-takers were calculated, for both blind and sighted persons. If one analyzes the final score obtained in the testing part and the minutes spent on a specific task, there is no straight-forward correlation: e.g. for audio-elevation task, the test-takers spent in average 30 minutes and obtained a 53 % score, while for audio-type task they spent only 7 minutes and obtained much better results, 96%. Based on the average final scores, the test-takers failed the following 8 tasks: audio elevation, audio distance, audio complex scene, audio box, haptics distance, haptics quantity, haptics complex scene and haptics box.

As a general observation, the average time spent on audio was 17 minutes, while the average time spent on haptics was 12 minutes. The average score obtained at final testing in audio was 70%, while the one obtained at haptics was 71%. Consequently, both models have similar precision. The blind persons did slightly better in testing, but they failed the same tests, except for the audio distance one.

To identify the learning progress of the users, a set of learnograms, for both audio and haptics, blinded and sighted persons were built. We were particularly interested in finding out the learning curves of the users. A learning curve shows the increase of learning (score) on vertical axis with experience (minutes) on horizontal axis. The learning curves were analyzed especially for those users who failed the final tests.

TABLE 1. SUMMATIVE RESULTS FOR LEARNING PROGRESS FOR AUDIO TASKS (P-PASSED; L- ASCENDING LEARNING CURVE)

	Direction		Elevation		Type		Width		Distance		Quantity		Complex Scene		1 Box	
	P	L	P	L	P	L	P	L	P	L	P	L	P	L	P	L
Number of test-takers (out of 15)	12	14	2	7	15	15	13	14	7	7	15	15	6	9	5	6

TABLE 2. SUMMATIVE RESULTS FOR LEARNING PROGRESS FOR HAPTIC TASKS (P-PASSED; L- ASCENDING LEARNING CURVE)

	Direction		Type		Width		Height		Distance		Quantity		Complex Scene		1 Box	
	P	L	P	L	P	L	P	L	P	L	P	L	P	L	P	L
Number of test-takers (out of 15)	15	15	14	14	11	13	9	11	5	6	8	8	1	3	4	7

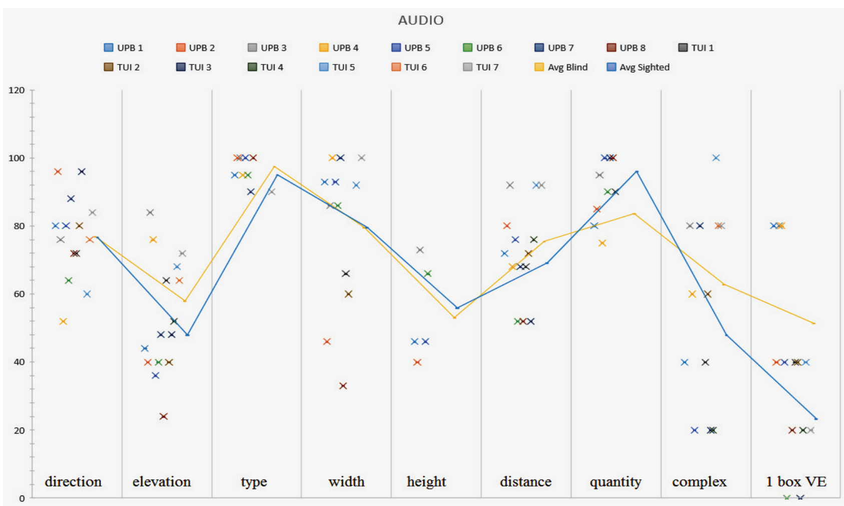


Figure 2. Final scores of all users for audio tasks; the yellow and blue lines show the average scores for blind, respectively sighted people

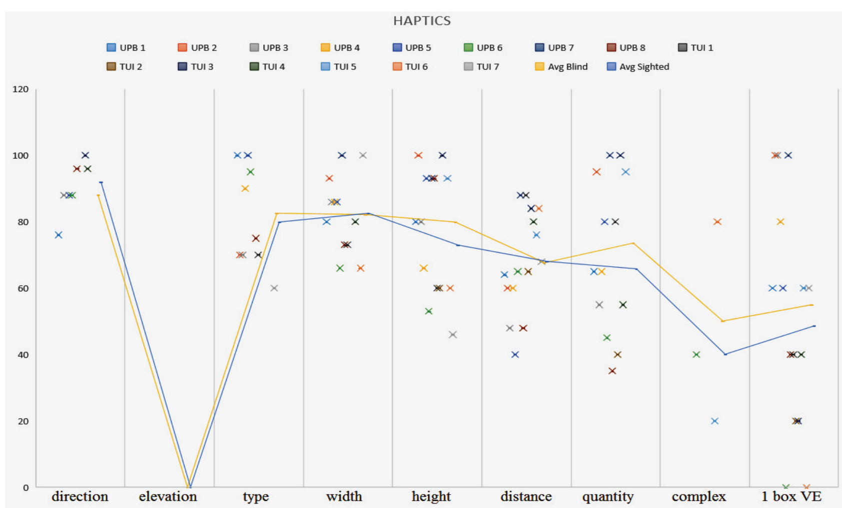


Figure 3. Final scores of all users for haptic tasks; the yellow and blue lines show the average scores for blind, respectively sighted people

The summative Table 1 shows whether the test-takers passed the audio tasks or failed the tasks and also whether they have ascending learning curve or not (based on the available learnograms). If the learning curve is ascending, even for the users who failed the final test, we consider that their results can be improved, but they need a longer period of training or the model is not intuitive enough (thus hard to be learnt).

The summative Table 2 shows whether the test-takers passed the haptic tasks (over 75% in the final score) or failed the tasks and also whether they have ascending learning curve or not (based on the available learnograms). If the learning curve is ascending, even for the users who failed the final test, we consider that their results can be improved, but they need a longer period of training or the model is not intuitive enough (thus hard to be learnt).

We noticed that there are tasks in the audio model which are very hard to be solved, e.g.: elevation, distance, complex scene, box task. Although at distance task, the users almost passed, we noticed that the learning curve is constant. For the other 3 tasks, the learning curve for most users is slightly ascending, thus, with a lot more training, those users might pass them. At the complex scene and box scene, the results were fluctuating. There were no significant differences registered between blind and sighted persons. Also, the older test-takers obtained similar results with the other users. Looking at the dispersion of results per each audio task (see Figure 2), the most intuitive models are type, width and quantity (most users obtained high scores). Still, at the tasks which were found difficult by some users, others were able to obtain good scores, so, with more practice, the tasks are doable.

With haptics, the distance, quantity, complex scene and box task were hard to be solved. For all those tasks, for most of the users who failed the test, the learning curve was fluctuant. So, more training does not guarantee better results. We considered that all the users who passed the tests have learning curves, but this is not always the case. There were no significant differences between the results of the older users and the others. Looking at the dispersion of results per each audio task (see Figure 3), the most intuitive models are direction, type and width (most users obtained high scores). At the complex scene, the haptic model was very hard to learn, while the results obtained at the box and quantity tasks were highly disparate and no useful conclusion can be drawn.

V. CONCLUSIONS

This study was performed to support the development of an SSD, as an alternative method for the blind people to acquire information about the surrounding space. As shown in other studies [13], by encoding the visual information into tactile or auditory stimuli, such devices have a great potential of helping the blind and visually impaired. However, SSDs require a careful design process, based on user feedback, as well as extensive training and testing to obtain accepted functionality.

The training and testing in virtual environment have a key role for VIPs' accommodation with audio and haptic encoding. For a blind person, it is very important to be confident with the SSD, to have courage to use it in real world as well, and to perceive the training and testing sessions like serious games.

Based on the feedback of the conducted tests, the hardware modules were redesigned, and audio and haptic encoding models were improved. An important decision which derived from the tests was the simultaneous use of audio and haptic devices for empowering the SOV system. Thus, one of the stimuli can overcome the lower performance of the other. For example, in case of elevation, the audio encoding leads to better results but for the height, haptics offered a better perception. An important future work is the cognitive load assessment of VIPs using the SOV system, based on analysis of electroencephalography and electrodermal activity records.

ACKNOWLEDGMENT

The authors wish to thank the volunteers for involvement and assistance. This work is supported by the European Union's Horizon 2020 research and innovation program under grant agreement No 643636 "Sound of Vision".

REFERENCES

[1] World Health Organization, Global Data on Visual Impairements, 2010, http://www.who.int/blindness/GLOBALDATAFINALforweb.pdf, last visit march 2017.

[2] Merja Saarela, "Solving way-finding challenges of a visually impaired person in a shopping mall by strengthening landmarks recognisability with iBeacons", Häme University of Applied Science (HAMK) Finland, http://www.hamk.fi/english/collaboration-and-research/smart-services/matec/Documents/solving-way.pdf, last visit march 2017.

[3] J. C. Middlebrooks, D. M. Green, "Sound Localization by Human Listeners", Annual Review of Psychology, **42**, 1991, pp.135-159.

[4] D. I. Ahlmark, "Haptic Navigation Aids for the Visually Impaired", PhD. Thesis, Lulea University of Technology 2016, www.diva-portal.org/smash/get/diva2:990226/FULLTEXT01.pdf, last visit march 2017.

[5] V. J. D. Gomez, "A computer-vision based sensory substitution device for the visually impaired (See ColOr)", phd thesis, Univ. Genève, 2014, no. Sc. 4642, file:///C:/Users/robert/Downloads /unige_34568 _attachment01.pdf, last visit march 2017.

[6] Kai Wun Lin, Tak Kit Lau, Chi Ming Cheuk, Yunhui Liu, "A wearable stereo vision system for visually impaired", International Conference on Mechatronics and Automation (ICMA), Chengdu - China, 2012.

[7] EyeMusic, Amir Amedi's Lab, Hebrew University of Jerusalem, http://brain.huji.ac.il/site/em.html, last visit march 2017.

[8] M. Bujacz, "Representing 3D scenes through spatial audio in an electronic travel aid for the blind", Ph.D. Thesis, Technical University of Lodz, http://www.bujaczm.os1.kappa.pl/Bujacz%20Michal%20-%20PhD%20Thesis.pdf, last visit march 2017.

[9] Barthelemy Durette, Nicolas Louveton, David Alleysson, Jeanny Herault, "Visuo-auditory sensory substitution for mobility assistance: testing TheVIBE", Workshop on Computer Vision Applications for the Visually Impaired, Paris, 2008.

[10] N. R. B. Stiles, S. Shimojo, "Auditory Sensory Substitution is Intuitive and Automatic with Texture Stimuli", Sci. Rep. **5**, , 2015, pp. 1-14.

[11] Sound of Vision, Natural sense of vision through acoustics and haptics, http://www.soundofvision.net/, last visit aplril 2017.

[12] F. Ungureanu, R. G. Lupu, S. Caraiman, A. Stan, "A Framework to Assess Cortical Activity of Visually Impaired Persons During Training with a Sensory Substitution Device", CSCS21-2017, Bucharest, in press.

[13] Ó. I. Jóhannesson, O. Balan, R. Unnthorsson, A. Moldoveanu, Á. Kristjánsson, "The Sound of Vision Project: On the Feasibility of an Audio-Haptic Representation of the Environment for the Visually Impaired", Brain Sci. 6, 2016, pp. 20.

Architectural design of a real-time augmented feedback system for neuromotor rehabilitation

Simona Caraiman, Andrei Stan, Nicolae Botezatu, Paul Herghelegiu, Robert Gabriel Lupu
Computer Department
"Gheorghe Asachi" Technical University of Iasi
Iasi, Romania
robert@tuiasi.ro

Alin Moldoveanu
Computer Department
University POLITEHNICA of Bucharest
Bucharest, Romania
alin.moldoveanu@cs.pub.ro

***Abstract*— This work presents the design aspects of TRAVEE, a neuromotor rehabilitation system. The TRAVEE system relies on innovative concepts for improving the rehabilitation process and increasing the patient recovery rate. One such concept is to present the patient an augmented feedback as part of a learning process based on neuroplasticity. Most of the rehabilitation exercises are based on visual feedback aimed at restoring the brain function for upper limbs control. This feedback is provided in a virtual reality setting by presenting the patient with a virtual model of his/her body. The movements executed by the patient are augmented in the virtual reality. Assisting the patient by a virtual therapist when executing the recovery exercises is another original feature of the proposed system. TRAVEE is a complex system, that integrates virtual reality, robotics, electrical stimulation, electromyography and brain-computer interfaces to boost the rehabilitation process. The challenges posed to the architectural design of TRAVEE reside in the complexity of its functioning in a setup that integrates a variety of devices, with real-time operation constraints and requirements of keeping the system at accessible costs, easy to install and use. In this paper, we present the proposed hardware and software architecture for TRAVEE. We analyze and discuss the advantages of our approach and the mechanisms that address the various constraints in TRAVEE.**

***Keywords*— stroke; visual feedback; rehabilitation; virtual/augmented feedback;realtime software**

I. INTRODUCTION

According to World Heart Federation [1], 15 million people suffer a stroke each year. Only four million successfully recovers, six million die and five million are left permanently disabled. That makes stroke the major cause of long term disability, with huge economical and social impact. For stroke survivors, rehabilitation is very important in order to relearn skills that were lost with part of the brain. That is possible due to brain neuroplasticity [2], which represents the brain ability to reorganize itself by creating new connections between neurons. This is possible with repetitive training [3], exercises functional relevance and the intensity of practice [4]. Most of the rehabilitation exercises are based on visual feedback in restoring brain function for limbs control. The visual feedback can be achieved simple by using a mirror during exercises, in which the patient can see the movements of his healthy arm and have the impression that is his impaired arm. Studies have shown that mirror therapy has significant recovery results on grip strength, accuracy of arm movements [5], speed and dexterity [6]. However there are some disadvantages like limited choice of activities, repeated motivation and good mental function, patient position and condition, limited perspective, lack of increasingly challenging tasks [7].

The disadvantages mentioned above can be overcome by the use of virtual reality (VR) and interactive video gaming. Taking into consideration the computer graphics development, VR approach in stroke rehabilitation is relatively recent [8]. The patient interacts with virtually simulated environment customized for his/her condition. The system provides therapist with valuable data about patient's progress based on which the therapist can adjust the game's difficult level. The possibility to change the game (environment, scenario), to increase the challenges along with the patient progress, to view the improvements in physical ability, engages and motivates the patient, increases the self-esteem [8]. Liang et al. observe that traditional therapy combined with virtual kitchen for hand training is more effective for motor function and ability for patients in convalescent phase [9]. This study demonstrates that VR is not used only for repetitive task to recover the motor function, but also to retrain the patient's activities of daily living (ADL). Programs like driving a car or a scooter, shopping in a supermarket have been tested and evaluation studies reported good effects of VR [10]. However there are some constrains regarding visual perception of spatial relationships between VR objects when guiding movements [11]. System interaction body parts driven by patients and system virtual objects have independent coordinates references [12] and that may affect the effectiveness of rehabilitation systems. Other studies reported that older people are reticent [13] on using VR therapy despite the demonstrated user friendly environments [14].

978-1-4799-1780-8/15 $31.00 © 2015 IEEE
DOI 10.1109/CSCS.2015.106

Augmented reality (AR) is a composite view of a computer generated image and real environment [8]. For stroke patients, an augmented reality based rehabilitation systems allows to control the information from the environment and how the patient interacts with it [15]. That can be used as stimuli for the therapy. The problem of patient movement consistency from VR where the patient have to map his movement with object movement from VR scene, no longer exists in AR. Luo et al. developed a system with AR and mechanical devices for repetitive practice like grasp-release tasks or moving virtual objects[16]. The patient is able to see his hand in the augmented reality scene and the information provided can be controlled in this environment. An example is the following: if, in his environment, a patient is unable to walk, is unable to move by himself his lower or upper limbs, a combination of AR and a robotic device, provides feedback to the patient and that is the visual sensation of walking, a very important component in motor learning [17]. However, the recovery results are around 80% for lower limb (the patient can learn to walk again) [17], [18] and a maximum of 20% for upper limb [17], [19]. Some causes for such low recovery percentage for the upper limb are: discontinuity of rehabilitation exercises once the patient is released from hospital [20], [21]; complexity of rehabilitation exercises [22], [23]; overusing of healthy arm leads to learned non-use for disabled one [20], [24]. Solutions to overcome these causes are: trained personnel for 30 hours / week [17] with high costs; tools for repetitive exercises that can be used even at home; the possibility to adjust these tools accordingly with patient needs.

Another approach in assisted stroke rehabilitation is the use of robotic devices, in scenarios that continues the traditional therapy or for evaluation [25], [26]. These devices can be used alone or with VR [27] to increase patient motivation and involvement [28]. The robotic assisted approach results do not exceed to much the traditional approach, but the costs involved (6-axis robots cost around US$60,000 without taking into consideration the development and tools) are too higher to justify the use of it. However the potential of such a system is not to be ignored but rather improved by using devices like Brain Computer Interface (BCI), Electromyography (EMG), Functional Electric Stimulation (FES), haptics. In this way new concepts and rehabilitation scenario can be developed in order to significantly improve the stroke recovery results. Designing the robotic devices for specific rehabilitation exercises can lower the cost, making it available for large scale home use. Yet, the design must take into consideration the possibility to integrate that particular robotic device into another system. In this way, the therapist can have the possibility to imagine and build a modular rehabilitation system that best fits the patient needs. Recent experiments of combining robotics and/or VR with BCI [29], [30] reveals the fact that motor imagery development has the same importance in recovery as development of motor action itself [31], [32]. The patient is trained to think a movement of his impaired limb, the BCI detects an EEG pattern of that specific thought that can be used as a trigger for robotics and/or FES. In the same way patterns in EMG signals can be used as triggers for FES devices, to detect muscle fatigue and to adjust speed and force during exercises.

II. TRAVEE SYSTEM

The TRAVEE system is an integrated bio-informatic system useful in recovery process of patients with neurolocomotor disabilities caused by stroke, accidents and brain surgery. It includes a rich hardware-software architecture and rehab concept with many contributions beyond state of the art. Out of these, should be noted some very important aspects: the brand new idea of using augmented and magnified feedback; the fact that tasks and guidance are provided by a virtual therapist - concept new in the field of rehab and considered extremely promising by the healthcare professionals; prospects of large-scale usable, with immense socio-economic impact; the modular design that allows the therapist to add/remove monitoring and stimulation/assisting devices in correspondence with rehabilitation exercises; the user interface of every module gives the possibility to modify working parameters for every patient or to follow the patient progress. This modularity can lower the costs for the patient by using only the needed modules. These hardware modules can be rented or purchased or even sold if no longer needed.

From design perspective, the TRAVEE system is intended to be a software kernel that interacts with different pluginable components, allowing the therapist to build/configure, to log, to visualize, to analyze rehabilitation exercises.

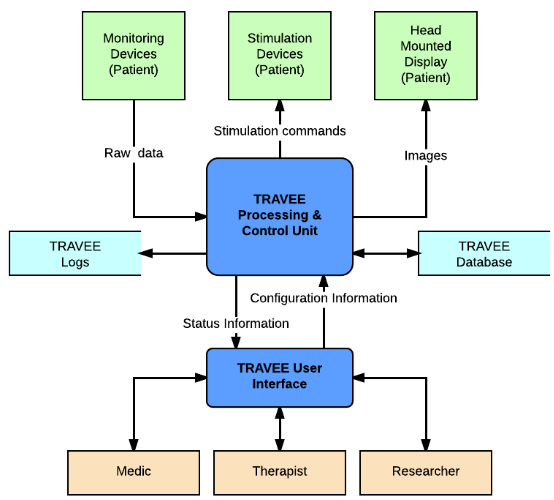


Fig. 1. TRAVEE system context diagram

The TRAVEE system will contain two main equipment categories: equipment for patient monitoring and training and equipment for data analysis and processing as presented in Fig. 1. Devices from the first category are optional pluginable components for the system: the system can use them all, but they are not mandatory. Results of processed raw data from monitoring devices will be used to trigger the stimulation devices with respect to rehabilitation exercise. The patient can view his body postures and movements in an augmented reality scene rendered on screen or head-mounted display (HMD).

Taking into account the necessary processes of augmented recovery, the TRAVEE system architecture reflects the components categories involved (hardware, robotics equipment, software) and how information is exchanged between them. All these interconnected components form a complete information loop closed by the patient. The basic idea is a more natural interaction of the patient with the system, the goal being the artificial closing of the causal loop that controls neuroplasticity / learning / recovery. In this way the system architecture is scalable both in terms of variable volume of information to process and terms of variable number of sensors controlled by the system and from which it receives information. This variability is given by the complexity of recovery exercise.

A key element taken into account in the "TRAVEE" system is interoperable synchronization of the main subsystems that perform both patient activities monitoring and physical, biological and psychological stimulation. From this perspective, the main challenge consists in relevant and real-time management of information provided by different components of the system. The solution of multilayer configuration and management of data stream can relieve the main processing module from the computation pressure and decision with the cost of losing, by filtering, the information considered irrelevant at a lower level. Because some data provided by hardware modules may be affected by errors (ex. motion inertial sensors are not so precise) multiple data synchronization between accelerometer, gyroscope, magnetic sensor and optic sensors (leap motion, kinect, video camera) is required.

Analyzing the system working scenarios, a series of constrains has been taken into consideration during system design. All constrains were classified in nine main categories: environment (mobility, easy to use, the therapist position), patient status (the system addresses only to patients with cognitive capabilities, the system components must not prevent patient from exercises, the system must draw patient attention whenever he loses concentration, button for discomfort signaling), resource availability (dynamic configuration, easy to add/remove components), interoperability (modular architecture, scalability), data storage (patient profile, patient history, local and remote data access), security (security access, patient confidentiality), real time system reaction, module testing, network communication.

III. HARDWARE ACHITECTURE

The used hardware devices must ensure the system functionalities of permanent patient monitoring during the exercises and the movement and stimulation of the upper limb that needs to be recovered. The monitoring function determines the correctness of the exercise performed by the patient. The monitoring is also used to provide real-time information to update the virtual environment. The system will aggregate and synchronize information gathered from multiple monitoring equipments to interpret the status and actions of the patient.

The stimulation function aims on one hand to restore and to maintain the muscular tone and / or to assist the patient when performing the recovery exercises. On the other hand, it aims to attract and maintain the focus and attention of the patient. In this respect, the stimuli generation for the patient will be realized on two levels: physic at the level of the affected upper limb and visual through augmenting the initial movement during the recovery exercise.

The processing and control functions aim to synchronize all events and decisions to allow the system to act as a whole. Both hardware equipments and software components will be selected to fulfill the system constraints regarding the performance and operational safety.

The hardware required to implement the system is shown in the Fig. 2. These equipments presented in the figure are chosen to meet the monitoring, stimulus generation and data presentation requirements.

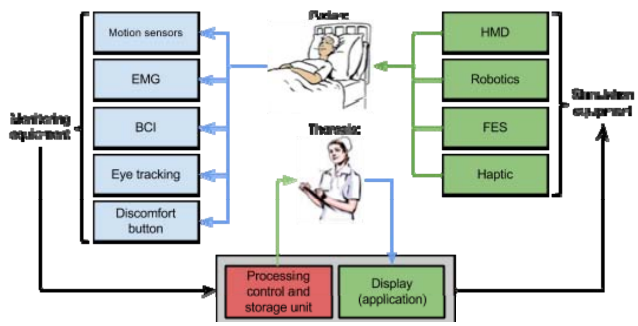


Fig. 2. TRAVEE hardware system architecture

Multiple equipments are used to meet the user requirements. In this way, the proposed architecture is maximal and allows flexibility in choosing concrete system configurations in order to fulfill the system objectives. Throughout the development of the project optimal configurations will be determined as tangible results will be produced by implementing different architectural versions.

A. *Monitoring components*

The monitoring equipments are designed to retrieve data on the position and movements of the patient's affected limb. Data about muscle and brain activity are also monitored. This information is send to the processing unit to be subject to specific processes that output appropriate commands for the stimulus generation equipment. Also, information obtained from the monitoring equipment are stored and presented to other users (e.g. physician, therapist, researcher) to further develop custom recovery scenarios. The monitoring equipments use different technologies (inertial sensors, optical sensors, etc.) and are described in the followings:

- Inertial sensors [33] are specific devices that allow movement tracking mainly by measuring the acceleration (accelerometers). In practice IMU (inertial measurement unit) units are used to reduce the acceleration conversion errors. These units contain an accelerometer, a gyroscope and a magnetic sensor.

- Optical sensors are devices used to track and monitor movements by analyzing image flows. Among such devices we can mention Leap Motion [34] [35] [36], [37], Kinect [38], video cameras and markers. Leap Motion can be used to track minor movements by using stereo vision IR sensors that determine the position of the limb.

Kinect sensor can be used to easily identify human silhouettes and track their movement. The standard Kinect system detects 25 joints, finger movements (thumb and at least another finger). It also estimates what muscles are involved in certain movements, can detect facial expressions and can estimate the heart rate by analyzing color changes of the human face. As disadvantages, the Kinect sensor requires the upper body detection to estimate the human skeleton. Also, it does not accurately compute the depths of the edges of an object and present cumulative errors that make slight movements hard to track.

Video camera [39] and markers represent a classical solution to detect object's movement. It requires the use of cameras and markers attached to the region of interest. The images acquired by the cameras are processed to obtain data about the marker movements.

- EMG [40] devices record the electrical activity. This is done by analyzing the electrical impulses of the peripheral nervous system in the relevant areas.

- BCI [41], [42] device allows investigating the possibilities of determining the intention of making a movement.

- ET device (Eye Tracking) [43] allows estimating / determining the gaze direction by using a video camera and specific image processing tools.

B. Stimuli generating components

- HMD Device [44], [45] represents a video system which has one or two screens through which images of the patient's virtual model can be projected. This virtual model will mimic the real life position of the patient and will reproduce its movements.

- FES devices [40], [46], [47] allow the application of electric currents to activate nerves in the extremities affected by paralysis as a way of augmenting the feedback and to stimulate the affected limb muscles accordingly.

- Vibrating devices [48] allow generation of vibrations with configurable / adjustable properties (e.g. intensity). Applying these devices on the patient allows the generation of a haptic feedback that might be associated with the properties of the performed movements.

C. Processing and control component

The system control and the data processing are performed by a computing component that allows the efficient implementation of all functionalities. The computing component must have the needed computing power and the amount of memory to process real-time data of different types (e.g. images, data acquired from the position sensors). The computing system must also allow efficient and secure data storage in a local database.

The computing system must be portable to allow easy use in different locations by different users (physician, therapist). Multi-users access to the system resources must be secure. The installed operating system has to provide support for all the devices that are referred to in the hardware system architecture.

IV. SOFTWARE ARCHITECTURE

The main constrains for the TRAVEE system with implications over the software architecture are: operation environment, patient condition, interoperability, data storage and access, security, network communication, real time data processing, integration of various devices for monitoring and stimulation. Thus, to deal with these constrains the software component has an event driven architecture (EDA) which is an architectural model based on event generation, detection, consumption and feedback. Another motivation for EDA is that all monitoring sensors work in an event-driven mode: event detection or threshold exceeding. Moreover, event hierarchies allow the conversion of sensor output from higher to lower granularity. This facilitates processing of complex events. The input/output model is straightforward: every system component signals its status and reacts to signals from the other components. The advantage of EDA is that it simplifies the design, development and testing by minimizing the connections between the system software components, making them simple to use and highly independent.

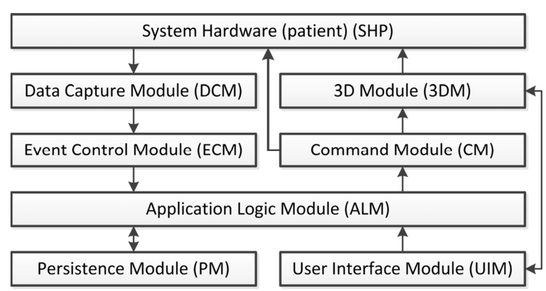


Fig. 3. System modules

The system processes are administrated in a *sense-analyze-respond* manner. First (*sense*) implies real time processing of data acquired from the patient monitoring devices. That generates a stream of events associated with the patient status and actions. By analyzing these events the patient status/actions can be estimated. The analysis component continuously evaluates the *when* clause of the *when-then* rules to determine if the *then* clause is to be executed. *When-then* rules are defined within predefined scenarios associated to the recovery session. The system response consists in execution of the *then* clauses, generating commands for the stimulation equipments. For every generated event, minimal information will be added: type, unique id, time stamp and component id (sensor, device, user interface). Depending on the sources, the events are classified in user, monitoring and error events. The events can be simple and are associated to monitoring functions like “button pressed”, “discomfort”, “movement detection”; or can be complex and are generated based on other simple events like muscular activity, neurologic activity, limb movements.

The modules included in the TRAVEE software architecture are presented in Fig. 3. Patient status and monitoring function are provided by Data Capture Module (DCM) and Event Control Module (ECM). The Application Logic Module (ALM) analyzes the received events in order to call the Command Module (CM) that generates individual commands for the stimulation equipments. There is a separate module (3DM) for the visual stimulation. It is designed to generate the 3D augmented reality scene and to update the patient avatar according to his/her actions. The patient avatar is the patient’s virtual body model in the 3D reality scene. The motion sensors detect the movements of the impaired limb and generate events to update the virtual body model. Raw data, events and actions are stored by Persistence Module (PM) for further analysis. The therapist interacts with the system through User Interface Module (UIM).

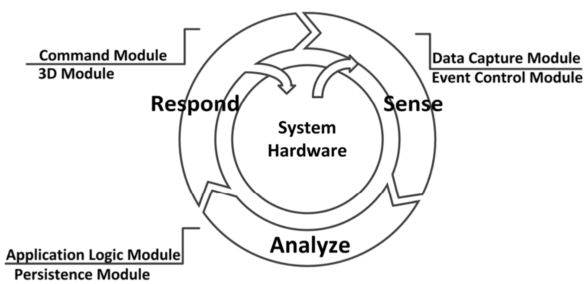


Fig. 4. *Sense - analyze – respond* mapping functions

The three functions of an EDA system, i.e. sense, analyze and respond, are mapped on the TRAVEE software components as described in Fig. 4. The proposed functional decomposition allows handling simple events generated by individual monitoring devices and also the aggregation of such data thus forming complex events. Thus, the system is able to detect a patient’s action more efficiently by correlating data coming from several sources. This correlation is provided by the ALM given its event analysis role.

V. Conclusions

In this paper we have presented the design aspects of TRAVEE, a neuromotor rehabilitation system. The key aspect of TRAVEE is to provide the user with an augmented feedback in a virtual environment, in order to speed up and improve the rehabilitation of an impaired upper limb. TRAVEE also aims to improve the user rehabilitation process by introducing a virtual therapist that will be a part of the virtual environment. The proposed hardware and software design ensures the system scalability and enables the connection of different devices for monitoring and stimulation. That gives the therapist the possibility to create various rehabilitations scenarios. The software architecture of the TRAVEE system takes into account the requirements for real-time processing of data and integration of various devices for monitoring and stimulation. To this end, the architecture is designed using an event-driven approach.

Last but not least the TRAVEE system has to be seen more as a new concept, the concept of digital therapist in stroke rehabilitation by promoting modularity, flexibility, customization, providing patients with professional specialized care, increasing self esteem and having the feeling of independence.

Acknowledgment

This work is supported by the Romanian National Authority for Scientific Research (UEFISCDI), Project 1/2014 Virtual Therapist with Augmented Feedback for Neuromotor Recovery (TRAVEE).

References

[1] World Heart Federation, Stroke, http://www.world-heart-federation.org/cardiovascular-health/stroke/, last visit feb. 2015

[2] Azari, Nina PView Profile; Seitz, Rudiger J., Brain plasticity and recovery from stroke, American Scientist88.5 (Sep/Oct 2000): 426-431.

[3] Van Peppen R., Kwakkel G., Wood-Dauphinee S., Hendriks H., Van der Wees P., Dekker J., The impact of physical therapy on functional outcomes after stroke: What’s the evidence? Clinical Rehabilitation, vol.18, no.8, pp. 833-862, ISSN 0269-2155, 2004

[4] French B., Thomas L., Leathley M., Sutton C., McAdam J., Forster A., Langhorne P., Repetitive task training for improving functional ability after stroke, vol.40, no.4, pp. 98-99, ISSN 00392499, 2009

[5] Yavuzer G., Selles R., Sezer N., Sutbeyaz S., Bussmann J.B., Koseoglu F., Atay M.B., Stam H.J., Mirror therapy improves function in subacute stroke: a randomized controlled trial, Arch Phys Med Rehabil 2008, 89:393-8

[6] Garry M.I., Loftus A., Summers J.J., Mirror, Mirror on the wall viewing a mirror reflection of unilateral hand movements facilitates Ipsilateral M1 excitability, Exp Brain Res 2005; 163:118-112

[7] Muzaffar Tufail, Wadhwa R.K., Borah Diganta, Laisram Nonica, Kothari S.Y., Evaluation of Mirror Therapy for Upper Limb Rehabilitation in Stroke, International Rehabilitation Forum Interscience Summit IRF2014, Suzhou International Expo Center, 16-10 oct. 2014, China

[8] Kimberlee Jordan, Marcus King Augmented Reality Assisted Upper Limb Rehabilitation Following Stroke, Augmented Reality - Some Emerging Application Areas, 2011, ISBN 978-953-307-422-1

[9] Liang M, Dou ZL, Wang QH, Xiong W, Zhen YD, Application of virtual reality technique in rehabilitation of hemiplegic upper extremities function of stroke patients, Chinese Journal of Rehabilitation Medicine 2013, pp: 114-118

[10] Kate Laver, Stacey George, Julie Ratcliffe and Maria Crotty, Virtual reality stroke rehabilitation – hype or hope, Australian Occupational Therapy Journal (2011) 58, 215–219, doi: 10.1111/j.1440-1630.2010.00897.x

[11] Hossein Mousavi Hondori, Maryam Khademi, Alison McKenzie, Lucy Dodakian, Cristina V. Lopes, Steven C Cramer, Utility of Augmented Reality in Relation to Virtual Reality in Stroke Rehabilitation, Stroke Volume 45,1 February 2014

[12] Borko Furht, HandBook of augmented reality, Springer 2011, ISBN 978-1-14614-0063-9, e-ISBN 978-1-4614-6, DOI 10.1007/9/978-1-4614-0064-6

[13] Lee J., Ku J., Cho W., Hahn W., Kim I., Lee S., A virtual reality system for the assessment and rehabilitation of the activities of daily living, Cyberpsychology and Behavior, 2003

[14] Rand D., Katz N., Weiss P., Intervention using the VMall for improving motor and functional ability of the upper extremity in post stroke patients, European Journal of Physical and Rehabilitation Medicine, 2009

[15] Levin M., Can virtual reality offer enriched environments for rehabilitation, Expert Review of Neurotherapeutics, Vol.11, No.2, February 2011, pp.153-5, ISSN: 1473-7175

[16] Xun Luo, T. Kline, H.C. Fischer, K.A. Stubblefield, R.V. Kenyon, D.G. Kamper, "Integration of Augmented Reality and Assistive Devices for Post-Stroke Hand Opening Rehabilitation", 27th Annual International Conference of the Engineering in Medicine and Biology Society IEEE-EMBS 2005., vol., no., pp.6855–6858

[17] M. King and J. Kimberlee, Augumented Reality Assisted Upper Limb Rehabilitation Following Stroke, Augmented Reality – Some Emerging Application Areas, 2011, ISBN 978-953-307-422-1

[18] C. E. Skilbeck, D. T. Wade, R. Hewer Langton and V. A. Wood, Recovery after stroke, Journal Neurology, Neurosurgery and Psychiatry, 1983, vol. 46

[19] G. Kwakkel , B. Kollen , J. van der Grond and A. Prevo , Probability of regaining dexterity in the flaccid upper limb, Impact of Severity of Paresis and Time Since Onset in Acute Stroke, Stroke, vol. 34, 2003.

[20] Feys HM, De Weerdt WJ, Selz BE, Cox Steck GA, Spichiger R, Vereeck LE, Putman KD, Van Hoydonck GA, Effect of a Therapeutic Intervention for the Hemiplegic Upper Limb in the Acute Phase After Stroke : A Single-Blind, Randomized, Controlled Multicenter Trial, Stroke. Journal of the American Heart Association, vol. 29, 1998.

[21] Carter K, Anderson C, Hacket M, Feigin V, Barber PA, Broad JB, Bonita R; Auckland Regional Community Stroke (ARCOS) Study Group, Trends in ethnic disparities in stroke incidence in Auckland, New Zealand, during 1981 to 2003, Stroke, vol. 37, no. 1, 2006.

[22] Nudo RJ, Milliken GW, Jenkins WM, Merzenich MM, Use-Dependent Alterations of Movement Representations in Primary Motor Cortex of Adult Squirrel Monkeys, The Journal of neuroscience : the official journal of the Society for Neuroscience, vol. 16, no. 2, 1996.

[23] Plautz EJ, Milliken GW, Nudo RJ, Effects of repetitive motor training on movement representations in adult squirrel monkeys: role of use versus learning, Neurobiology Learning and Memory, vol. 74, 2000.

[24] Sterr A, Elbert T, Berthold I, Kölbel S, Rockstroh B, Taub E, Longer Versus Shorter Daily Constraint-Induced Movement Therapy of Chronic Hemiparesis: An Exploratory Study, Archives of physical medicine and rehabilitation, vol. 83, 2002.

[25] Gert Kwakkel, Boudewijn J. Kollen, and Hermano I. Krebs, Effects of robot-assisted therapy on upper limb recovery after stroke: a systematic review, Neurorehabilitation and Neural Repair, vol. 22, 2008

[26] Susan E. Fasoli, Hermano I. Krebs, Joel Stein, Walter R. Frontera, Richard Hughes, Neville Hogan, Robotic therapy for chronic motor impairments after stroke: follow-up results, Archives of Physical Medicine and Rehabilitation, vol. 84, 2003

[27] Prange GB, Jannink MJ, Groothuis-Oudshoorn CG, Hermens HJ, Ijzerman MJ, Systematic review of the effect of robot-aided therapy on recovery of the hemiparetic arm after stroke, Journal of rehabilitation research and development, vol. 43, no. 2, 2006

[28] Housman SJ, Scott KM, Reinkensmeyer DJ, A randomized controlled trial of gravity-supported, computer-enhanced arm exercise for individuals with severe hemiparesis, Neurorehabilitation and neural repair, vol. 23, 2009

[29] Ang KK, Guan C, Chua KS, Ang BT, Kuah C, Wang C, Phua KS, Chin ZY, Zhang H, A clinical study of motor imagery-based brain-computer interface for upper limb robotic rehabilitation, Proceedings of IEEE Engineering in Medicine and Biology Society Annual Conference, 2009.

[30] S. Bermudez i Badia, A. Garcia Morgade, H. Samaha and P.F.M.J. Verschure, Using a Hybrid Brain Computer Interface and Virtual Reality System to Monitor and Promote Cortical Reorganization through Motor Activity and Motor Imagery Training, IEEE Transaction on Neural Systems and Rehabilitation Engineering, vol. 21, no. 2, march 2013.

[31] Giacomo Rizzolatti, Corrado Sinigaglia, Mirrors in the brain-How our minds share actions and emotions, Oxford University Press, Oxford, 2008.

[32] Steven M. Platek, Kyrre Wathne, Niall G. Tierney, Jaime W. Thomson, Neural correlates of self-face recognition: an effect-location meta-analysis, Brain Research, vol. 1232, 2008

[33] LPMS-B: 9-Axis IMU / AHRS / Motion Sensor with Bluetooth (wireless) Connectivity, http://www.lp-research.com/9-axis-imu-with-bluetooth-wireless-connectivity/, last visit april 2015

[34] Leap Motion, https://www.leapmotion.com/, last visit april 2015

[35] Leap Motion for Virtual Reality, https://www.leapmotion.com/product/vr

[36] Frank Weichert et al., Analysis of the Accuracy and Robustness of the Leap Motion Controller, 2013, Sensors 2013, doi:10.3390/s130506380

[37] Maryam Khademi et al., Free-Hand Interaction with Leap MotionController for Stroke Rehabilitation, 2014, CHI '14 ExtendedAbstracts on HumanFactors în ComputingSystems.

[38] Kinect for Windows v2, https://www.microsoft.com/en-us/kinectforwindows/, last visit april 2015

[39] SmartCam Full HD 1080p WiFi Camera, http://www.samsung.com/us/security-and-monitoring/ip-cameras/SNH-E6411BN, last visit april 2015

[40] NeuroTrac MyoPlus4, http://veritymedical.co.uk/Products/13/NeuroTrac%C2%AE+MyoPlus4

[41] Emotiv Epoc, http://emotiv.com/product-specs/Emotiv%20EPOC%20Specifications%202014.pdf

[42] g-HIamp Multichannel amplifier, http://www.gtec.at/Products/Hardware-and-Accessories/g.HIamp-Specs-Features, last visit april 2015

[43] Robert Gabriel Lupu, Florina Ungureanu, Valentin Siriteanu, Eye tracking mouse for human computer interaction, E-Health and Bioengineering Conference (EHB), 2013

[44] Oculus Development kit 2, https://www.oculus.com/dk2/

[45] Sony Wearable HDTV, http://store.sony.com/wearable-hdtv-2d-3d-virtual-7.1-surround-sound-zid27-HMZT3W/cat-27-catid-3D-Personal-Viewer

[46] Functional Electric Stimulation - Motionstim 8, http://www.kundt-hamburg.de/medel/doku/sciencemode/index.html, last visit april 2015

[47] R. Rupp, A. Kreilinger, M. Rohm, V. Kaiser, G.R. Müller-Putz, Development of a non-invasive, multifunctional grasp neuroprosthesis and its evaluation in an individual with a high spinal cord injury, 34th Annual International Conference of the IEEE EMBS San Diego, USA, 2013

[48] DRV2603, Haptic Drive with Auto-Resonance Detection for Linear Resonance Actuators, SLOS754A –JUNE 2012–REVISED JANUARY 2014, http://www.ti.com/lit/ds/symlink/drv2603.pdf, last visit april 2015

The 4th IEEE International Conference on E-Health and Bioengineering - EHB 2013
Grigore T. Popa University of Medicine and Pharmacy, Iaşi, Romania, November 21-23, 2013

Eye Tracking Mouse for Human Computer Interaction

Robert Gabriel Lupu, Florina Ungureanu, Valentin Siriteanu
Department of Computer Engineering
"Gheorghe Asachi" Technical University of Iasi
Iasi, Romania
robert@cs.tuiasi.ro, fungurea@cs.tuiasi.ro, valentin.siriteanu@gmail.com

***Abstract*— In many cases, persons with neuro-locomotor disabilities have a good level of understanding and should use their eyes for communication. In this paper a reliable, mobile and low-cost system based on eye tracking mouse is presented. The eye movement is detected by a head mounted device and consequently the mouse cursor is moved on the screen. A click event denoting a pictogram selection is performed if the patient gazes a certain time the corresponding image on the screen.**

***Keywords*— assistive technology, image processing, pupil detection, video glasses.**

I. INTRODUCTION

Assistive technology (AT) promotes greater independence for people with disabilities by enabling them to perform tasks that they were formerly unable to accomplish. However, the communication with patients having neuro-locomotor disabilities is a great challenge even today [1]. Usually, the communication with these patients requires continuous presence of a caregiver who should guess patient's basic needs. There is a category of people with severe speech and motor impairment or with neuro-locomotor disabilities who cannot speak and cannot use sign language. If these patients have a good level of understanding and perception, they should use their eyes for Human-Computer Interaction (HCI). Eye tracking (ET) techniques measure the person's eye movements so that the gaze point at any time and the eyes shifting are established accurately. Different invasive or noninvasive methods for eye movement measurement were investigated. Today, some vendors (e.g. Tobii or MyGaze) provide commercial remote camera-based eye-tracker systems for which the light source and camera are permanently affixed to a monitor. These systems require the patient's presence in front of the monitor and calibration procedure for any new dialog session and do not fit with the aims of AT initiative. Furthermore, these commercial systems are expensive, exceeding 10 000 USD.

As an alternative, some mobile and low cost devices for HCI were developed by different research groups [2], [3]. ETRA Conferences join together companies and researchers involved in eye tracking technologies and highlight new hardware and software solutions.

In one of our previous research project, a communication system for people with disabilities, named *ASISTSYS,* was designed and implemented in concordance with the international guidelines and rules regarding assistive technology (AT). *ASISTSYS*, presented in detail in [3], is based on a mobile device for patient's gaze measurement and also an optimized algorithm for video eye tracking implemented on embedded system. The eye tracking system was composed from a webcam mounted on a glasses frame, a mobile device with BeagleBoard xM for image acquisition and processing, a monitor for displaying words correlated with patient's needs and software application written in C++ and Qt. The prototype of the proposed system has been tested in a neurologic recovery clinic and was rated by patients with 210 points from 225 maximum possible. The medical staff evaluation revealed an overall score was of 18 from a maximum of 25. Despite favorable general assessment, a few drawbacks were revealed: the quality of the acquired images, the use of a monitor for displaying the user graphic interface, the sensitivity of ET algorithm to light intensity and the selection of an image or word by looking at it and blinking.

Our recent research was focused on new hardware and software solutions to improve the reliability, mobility and usability of the communication system. The proposed eye tracking method was oriented towards the possibility to be used by patients for email, messenger and social sites. In this paper we propose an eye tracking mouse (ETM) system using video glasses and a new robust eye tracking algorithm based on the adaptive binary segmentation threshold of the acquired images. The proposed system allows the patient to communicate his needs, to browse a graphical user interface and to select an image or a word, using only his eyes.

978-1-4799-2373-1/13/$31.00 ©2013 IEEE

II. EYE TRACKING MOUSE ARCHITECTURE

The proposed ETM system consists of two hardware devices, webcam and video glasses and the software application running the eye tracking algorithm. The webcam, mounted on a video glasses frame with the help of an aluminum bar, has a modified system lens in order to be used at a short suitable distance (less than ten centimeters) from user's eyes. It captures images only in infrared light by using an infrared filter on top of the lens. Six infrared LEDs provide constant illumination of the eye so that the natural light has an insignificant influence on pupil detection.

The video glasses display copies of the computer screen for both eyes so that the patient sees a 16:9 widescreen 1.9 m display, as seen from 3 m [4]. The software application detects the pupil and maps its webcam position on computer screen in concordance with patient's gaze direction. Therefore, the mouse cursor is moved in the point of screen coordinates. By gazing at that point for one to two seconds, the software generates left click event. In this way the patient can point and click.

Unlike the previous approach, video glasses were used instead of computer monitor so that the head position of the patient does not affect the eye tracking algorithm after calibration. The software application was written in C++ and C# using Visual Studio 2010 and OpenCV library for image processing. The software application is organized on two layers, as it is presented in Fig. 2.

The input layer is written in C++ and consists in three modules: *Feeder, pInitializer* and *pTracker*. The *Feeder* module provides for *pInitializer* continually acquired and pre-processed images until ROI (Region Of Interest), binary segmentation threshold and mapping coefficients are obtained. After these values are validated, the *pTracker* module detects eye pupil and *mTracker* determines the mouse coordinates.

The output layer written in C# defines how information provided by the input layer are processed. So, the *Point mapper* calculates the new cursor coordinates based on webcam pupil coordinates. The mapping coefficients can be loaded from a local file or can be also updated when the *mTracker* is not running.

The output layer written in C# defines how information provided by the input layer are processed. So, the *Point mapper* calculates the new cursor coordinates based on webcam pupil coordinates. The mapping coefficients can be loaded from a local file or can be also updated when the *mTracker* is not running. The *Calibration* component of *mInitializer* module displays nine points on screen, one at a time. The patient has to look straight to each of them for one or two seconds and the corresponding positions of the pupil are recorded. Then, using the Sheena and Borah [5] equations, the mapping coefficients are determined.

The *User Interface (UI)* module moves the cursor in the position provided by *Point mapper*. The click event is generated if the cursor stays in a certain position for one second.

Fig. 1. Head mounted Eye Tracking System.

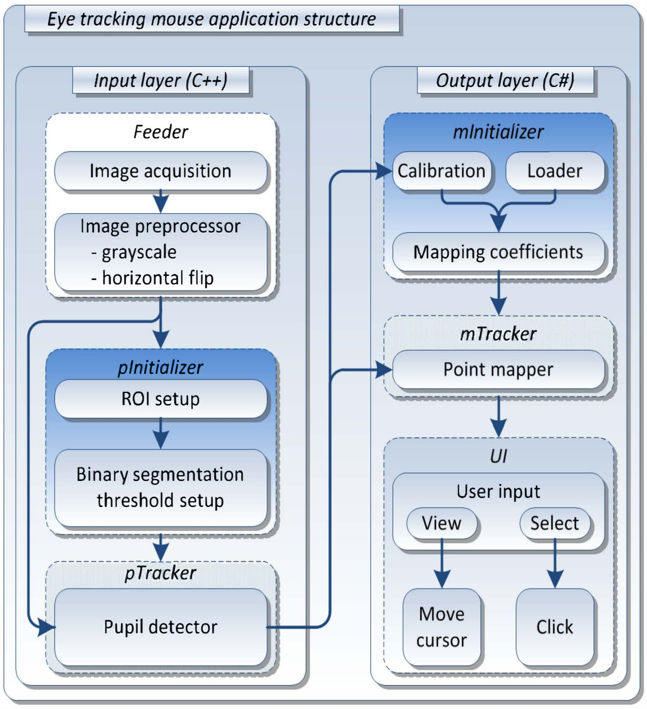


Fig. 2. Eye tracking mouse software application.

III. EYE TRACKING MOUSE METHOD

The key point of eye tracking mouse application is the pupil detection algorithm. Firstly, Starburst algorithm [5] was implemented. This is a hybrid algorithm because it uses two main approaches: model-based and feature-based. The algorithm starts by detection and removal of corneal reflection. Next step is to find candidate feature points (located on pupil contour) and then RANSAC (RANdom SAmple Consensus) algorithm is applied to find feature point consensus set. Those points are used to find best fitting ellipse for pupil contour. The weak point of Starburst algorithm is the parameters of ellipse instability during the ellipse fitting for every frame. Therefore, the center of the detected pupil varies as it is shown in Fig. 3, and consequently the cursor position on screen is changing even if the patient stares at a fixed point. Due to the instability of the cursor position on screen, the Starburst algorithm could not be used for pupil detection in order to point on screen and to generate a left click event.

Another feature-based approach presumes that pupil correspondent pixels are the blackest ones from the image. In order to obtain a binary image of the eye, an inverse binary segmentation with a given threshold was used.

The new image contains white pixels in the same area where pupil correspondent pixels from originally image were located. The center of mass for this image is the center of the pupil. The main advantage of this approach is that the position of the pupil center has insignificant variation and therefor the cursor position on screen is stable on both axes as it is presented in Fig. 4. The challenge of this approach is to determine automatically the binary segmentation threshold. The proposed algorithm for eye tracking mouse, hereinafter referred as ETAST (Eye Tracking with Adapted Segmentation Threshold) is based on binary segmentation of the image and its diagram is presented in Fig. 5.

The proposed algorithm is performed in three stages. In *Preprocessing* stage images with 640 x 480 resolution are acquired. Then, each image is converted in grey scale and flipped horizontally. If the tracking algorithm was initialized previously the *Tracking* task is launched, otherwise *Initializing* task starts.

The ROI coordinates, segmentation threshold and mapping coefficients are established in the *Initializing* stage. The ROI coordinates specifying where the eye ball is located on the image are determined using the *HaarCascadeFilter* from OpenCV library for several frames. The function returns the coordinates of the upper corner of a rectangle that square the eye and the length and high of the rectangle. The obtained values are averaged in order to obtain the final rectangle coordinates. After ROI determination, a mask image with the same dimensions is generated. The mask presented in Fig. 6 is applied in order to eliminate noise pixels.

Binary segmentation threshold is computed using one frame. It starts by calling the Starburst algorithm in order to find the ellipse that fits the pupil contour. Then, the binary segmentation threshold is incremented from default value until white pixels outside the ellipse are obtained in segmented image. The *Initializing* stage ends by calling calibration process to determine new mapping coefficients or by loading the old ones. According to Parkhurst [5], the calibration method which has the lowest error degree is based on biquadratic function. This nonlinear mapping function needs nine calibration points for determining coefficients values. The points with known coordinates are displayed on the monitor in a 3x3 grid and divide the screen in four quadrants. The mapping functions are widely described in [6]. *Tracking* task detects the pupil coordinates by calculating the center of mass for segmented image. Then, the mapping equations are applied and the new cursor coordinates are obtained corresponding to patient gaze point on video glasses screen.

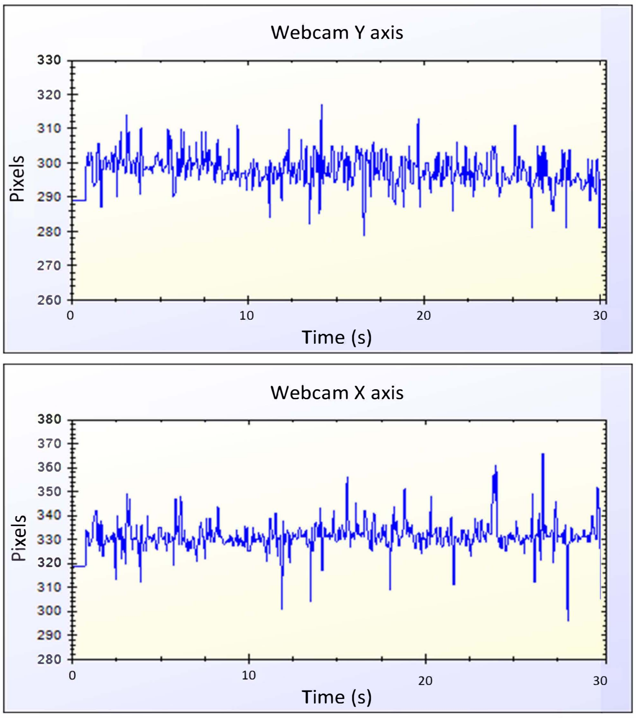


Fig. 3. Variation of pupil center position using Starburst algorithm.

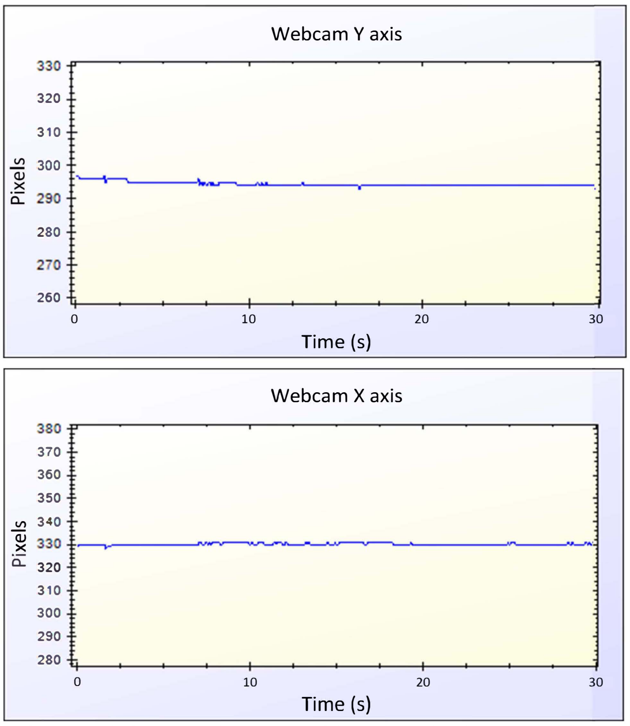


Fig. 4. Variation of pupil center position using binarization algorithm.

In order to assess the performance of ETAST method the rows gaze data provided by the ETM were recorded. Some widely accepted metrics are briefly presented in [7] and refer to fixation, saccadic eye movement, smooth pursuit, scanpath, etc. The analysis of experimental data reveals that ETAST algorithm is not sensitive to the noise generated by involuntary blinking or inherent pupil movement and provides good results for scanpath metric.

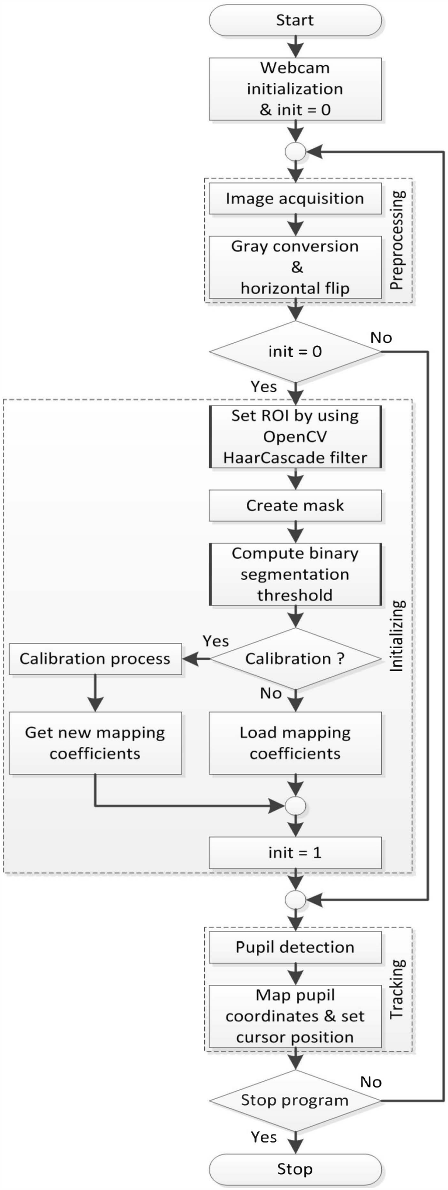


Fig. 5. ETAR algorithm.

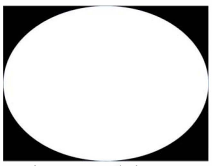


Fig. 6. Mask image.

IV. RESULTS AND CONCLUSIONS

We performed an experimental procedure in order to evaluate the usability, accuracy and reliability of the ETM system. Twenty participants, ranging from 19 to 27 were involved in the experiment. All of them had normal or corrected to normal vision, they had no prior experience with eye tracking and needed a minimal training for system using. The volunteers were asked to fill a Linkert questionnaire with 8 questions concerning ergonomics, learning and adaptation time for system using, ease of use, level of fatigue, usefulness for healthy people or ETM shortcomings. The system was rated with 603 points from 800 maximum possible. The poor rating came from two makeup ladies and this is easily justified. Because the pupil detection is based on the darkness points from the acquired image, the mascara from the eyelashes affects the algorithm performance. This is not a drawback for the people with severe neouro-locomotor disabilities but must be solved for others future ETM applications. All participants agreed that the ETM was easy to use without any discomfort.

The obtained positive results proved that ETM method is a reliable and low-cost solution for HCI and fits with the assistive technology goals to provide efficient solutions for patient's communication and independent activities.

The social impact of the proposed ETM system may be significant allowing the social reinsertion of the disabled persons and increasing their self-respect. For many disabled people, such a communication system could help them to continue their intellectual and social life or to pass easier the difficult period of medical recuperation. In addition, taking into account that many people with disabilities do not afford a suitable communication system, this low-cost system could successfully replace the more expensive ones [8].

The proposed mobile device should be also useful for people with limited hand functions or should be integrated in different virtual and augmented reality systems for recovering and rehabilitation process targeting persons suffering from neuromotor paralysis in the spirit of the new paradigm of Cyber-Physical Systems [9].

ACKNOWLEDGMENT

The authors thank Mihai Bulea for his kindly support and encouragement.

REFERENCES

[1] C. Rotariu, H. Costin, V. Cehan and O. Morancea, "A communication system with severe neuro-locomotor handicapped persons", Proc. Int. Conf. Biomedical Electronics and Biomedical Informatics, Rodos, Greece, pp. 145-149, Aug. 2008.

[2] J.S. Agustin, E. Mollenbach, M. Barret, M. Tall and D.W. Hansen, "Evaluation of a low-cost open-source gaze tracker", Proceedings of ETRA 2010, Austin, USA, pp. 77-80, March 2010.

[3] R.G. Lupu, F. Ungureanu and R.G. Bozomitu, "Mobile embedded system for human computer communication in assitive technology", Proceedings IEEE ICCP 2012, Cluj-Napoca, Romania, pp. 209-212, August 2012

[4] Vuzix wrap 1200, video eyewear http://www.vuzix.com/consumer/, last visit august 2013.

[5] D. Li and D. J. Parkhurst, "Starburst: A robust algorithm for video-based eye tracking", Elsevier Science, 2005.

[6] D. Sheena and B. Borah, "Compensation for second-order effects to improve eye position measurements, pp. 257-268, in "Eye movements: Cognition and visual perception", University of Illinois Press, 1981.

[7] S.V. Wass, T.J. Smithand M.H. Johnson, "Parsing eye-tracking data of variable quality to provide accurate fixation duration estimates in infants and adults", Behav. Res. Methods, Vol. 4, pp 229-250, 2013.

[8] O. Geman, H. Costin, "Parkinson's Disease Prediction based on Multistate Markov Models", Int. J. of Computers, Communication and Control, ISSN 1841-9836, 8(4): 525-537, August, 2013.

[9] A. Moldoveanu, F. Moldoveanu, V. Asavei, A. Egner and A. Morar, "From HTML to 3DMMO - a roadmap full of Challenges", Adv. In Intelligent Control Systems & Computer Science, AISC 187, I. Dumitrache Ed., Springer-Verlag Berlin Heidelberg, 2013, pp. 379-392.