

Design and Evaluation of a Vision-Based UI for People with Large Cognitive-Motor Disabilities

Sergio Martínez¹, Antonio Peñalver² , Juan Manuel Sáez¹

¹Department of Computer Science and Artificial Intelligence, University of Alicante, Alicante, Spain; ²Center of Operations Research, Miguel Hernández University, Elche, Spain

Correspondence to: Antonio Peñalver, a.penalver@umh.es

Keywords: Brain Damage, Rehabilitation, Disabilities, Vision-Based User Interface

Received: December 12, 2020

Accepted: April 27, 2021

Published: April 30, 2021

Copyright © 2021 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

ABSTRACT

Recovering from multiple traumatic brain injury (TBI) is a very difficult task, depending on the severity of the lesions, the affected parts of the brain and the level of damage (locomotor, cognitive or sensory). Although there are some software platforms to help these patients to recover part of the lost capacity, the variety of existing lesions and the different degree to which they affect the patient, do not allow the generalization of the appropriate treatments and tools in each case. The aim of this work is to design and evaluate a machine vision-based UI (User Interface) allowing patients with a high level of injury to interact with a computer. This UI will be a tool for the therapy they follow and a way to communicate with their environment. The interface provides a set of specific activities, developed in collaboration with the multidisciplinary team that is currently evaluating each patient, to be used as a part of the therapy they receive. The system has been successfully tested with two patients whose degree of disability prevents them from using other types of platforms.

1. INTRODUCTION

The recovery of a patient suffering from traumatic brain injury (TBI) depends largely on the extent of the lesions, the parts of the brain affected, and the problems presented by the patient (at locomotor, cognitive and sensory levels). Damaged neural tissues are not recoverable, but there is a way to improve their effects thanks to a phenomenon known as plasticity [1], which is the nervous system's ability to minimize injuries by modifying its own structure and functional organization. To recover some of the functions affected by the TBI, the nervous system is capable of adapting neurons initially created for other functions to perform the tasks involved. Recovery through this phenomenon is slow but unlimited, so that the treatment is usually extended throughout the entire life of the patient [2]. Obviously, the treatment is much more complex when there exist multiple traumas causing different aftermaths. The apps developed in this

work are devoted to patients with this type of injury.

The aim of this work is to design a UI that allows patients that maintain at least mobility in one of their arms to interact with a computer so that it can be used as a tool for the therapy they are receiving and then ease the difficult task of communicating with their environment. The main idea is to capture the movement of a hand with a digital camera, which is used to control an avatar on the computer screen. The reason for using this type of interface is that the arm movement is very limited in most cases. Usually, most patients have a very controlled movement of the arm, but not of the hand, remaining always their fists closed. This prevents them to manipulate objects that require some skill using the fingers, such as a classical mouse. Furthermore, both the accuracy of the arm movement and the control of the force performed are really limited (in some cases by excess and others by lack of strength). This prevents the use of any kind of interface requiring any kind of pressure such as touch screens.

To overcome these limitations, we propose the use of a machine vision-based interface adapted to the abilities of the patients that allows interaction with the computer by means of a communication protocol. Through this interface, we develop a set of specific activities, in collaboration with the multi-disciplinary team that is currently evaluating the patients, so that such activities can be used as part of the therapy that they are carrying out. This opens an important range of possibilities allowing therapists to plan activities that involve a variety of multimedia content with which patients can interact.

The rest of the paper is organized as follows: in Section 2 we review some other projects related to the design of interfaces for people with brain injuries. Then, Section 3 is devoted to explaining artificial markers and how they can be used to detect the movement of the patient's hand. In Section 4 we describe the applications developed that use the marker designed to allow communication between the user and the computer. Then, in Section 5 we show the results of the evaluation of the tool after a period of test. Last, Section 6 summarizes conclusions and further work.

2. RELATED WORK

This section includes a review of some previous projects related to the recovery of a patient suffering from TBI. The continuous growing of mobile computing and the need of providing remote rehabilitation programs have led to the proliferation of several projects with similar objectives.

SITPLUS5 [3] is a computer vision-based tool using color markers among other peripherals like microphones. However, the purpose of the system is recreational and therapeutic while in our case, we rely mainly on patient communication. Besides, SITPLUS requires the configuration and calibration of several sensors, whereas our approach does not require any kind of calibration, so therapists can focus solely on the applications settings. Although SITPLUS can be used in a great number of disabilities, it is not suitable for people with serious locomotor and cognitive deficiencies as it happens in the case of the patients to whom our application is addressed, who only maintain mobility in one of their arms.

HABITAT [4] is a tool for creating and implementing interactive rehabilitation activities for persons suffering from acquired brain damage. It provides both a platform for rehabilitation of patients and tools for specialists to manage and monitor the outcome of patients. The interaction between tool and patients occurs through different hardware devices that allow both the undertaking of activities such and communication with specialists.

PREVIRNEC [5] is a software tool for remote rehabilitation providing a set of rehabilitation activities for patients, and utilities for the management and monitoring of the process for specialists. It uses virtual reality, building 2D and 3D scenarios where the affected persons learn everyday activities. The interaction in rehabilitation activities can be performed by any pointing device but requires a greater degree of mobility that many patients do not have.

RehaCom [6] is another rehabilitation tool developed in Germany by computer scientists and neuropsychologists. It has pre-set activities to be used as rehabilitation exercises, stimulating cognitive functions such as attention, concentration, logical reasoning, planning and problem solving, etc. The interaction is carried out through a special console or panel, comprising buttons and a small joystick.

NeuronUp [7] is a cloud-based tool for remote rehabilitation offering a wide variety of activities, simulations of real-world tasks and content management systems for the treatment of people affected by brain damage or positive aging (cognitive deficits caused by normal aging). It allows users to interact via a computer, a touch screen or even pen and paper. It allows to run on several portable devices such as PDA's or smartphones, with applications on various operating systems.

Indigo [8] is a game-based tool from Victoria University in Canada. Presenting the tool using the aspect of a game reduces anxiety that software rehabilitation tools cause on persons suffering from brain damage. The interface is simple, no windows, no toolbar and full screen. The pointing device is a mouse or a keyboard, so it is not suitable for people with severe motor impairments. Some tools are used for therapy and other to assess improvements in cognitive abilities of patients [9].

The tools described above are not suitable for people with a very limited level of mobility, as they use input devices that require a skill level much higher than they have. In other tools, the rehabilitation is performed using motion capture-based systems [10, 11]. Such systems are typically based on sensors that are attached to a body part and transmit the information via Wi-Fi or locally stored for further processing.

In addition to these general systems, there have been other rehabilitation systems designed specifically for disease specific disabilities, as shown in [12], for the rehabilitation of the upper extremities, or in [13] for the rehabilitation of knees. Some of these systems are based on mobile devices, as shown in [14, 15], allowing the use of mobile devices in the rehabilitation of patients suffering from a heart problem. Other tools, such as CloudRehab [16], provides a customizable platform for tele-rehabilitation, using a cloud-based technology, mobile devices, heartbeat bands and web technology, suitable for patients with brain damage.

3. ARTIFICIAL MARKERS AND HAND DETECTION

In computer vision, a synthetic marker is an object that is not part of the environment and it is introduced explicitly to be detected by a vision system. The features required by a marker depend on the application for which it is designed. We can say that a marker should be easily detectable in the environment (low ambiguity/high distinctiveness) where it will be used in order to achieve a fast and robust identification. There are two families of methods to detect markers: Statistical and Geometric.

Nowadays, artificial markers are widely used in different fields, like guiding and positioning autonomous robots or augmented reality. In [17] a technique for simultaneous localization and mapping (SLAM) using markers in black and white is described. In [18] color markers are used to improve the global localization. In the augmented reality context, markers allow to represent additional information on the real image as well as any type of virtual elements. In [19] four markers' systems for augmented reality are compared (ATK ARtoolkit, IGD, HOM and SCR). Another type of marker is based on the detection of infrared light. These have materials that are highly reflective when are illuminated with this type of light.

This method is the least visually intrusive but requires markers, cameras and special lighting. In [20] an infrared marker is used in an augmented reality system which facilitates maintenance of the industrial machinery where is implemented. To do this, the system shows the operator how to proceed with the work of replacement or repair the parts he is working with.

The traffic signals are being used as artificial markers more and more. Although they were not designed for this purpose, they meet the requirements of a marker. In [21] a traffic signals recognition system using the images captured by a smartphone is described. Moreover, many car manufacturers are including signs recognition systems in their vehicles.

3.1. Proposed Marker

For this project we use a marker based on [22] but we include some modifications in order to make it suitable for the aim of our project. The original synthetic marker is round shaped and based in color. **Figure 1** (left) shows the marker divided into three sectors with the colors red, green and blue.

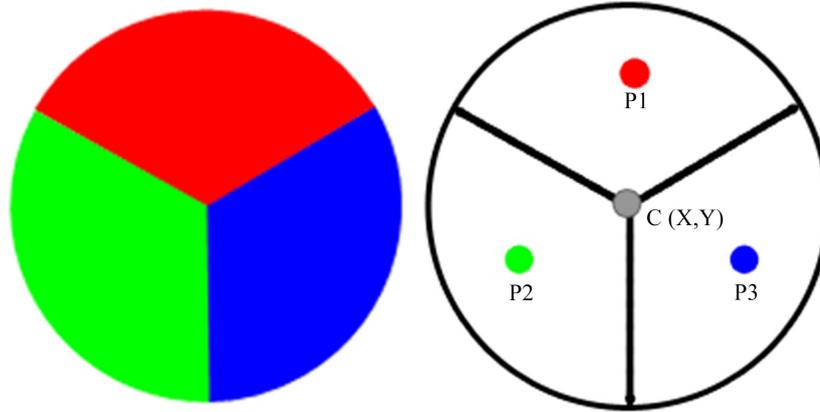


Figure 1. Original color-based marker (left) and interest points (right).

This technique does not assess each color separately but is based on the differences between them. In order to determine if a point $C (X, Y)$ matches the center of the marker, three points $P1$, $P2$, $P3$ at a distance of D pixels are evaluated. $P1$ is selected north, $P2$ is selected southwest and $P3$ is selected southeast. In **Figure 1** (right) we can see the points location.

The aim is to measure the area where probably each sector of the marker is placed. Therefore, $P1$ should match a red point, $P2$ should match a green point and $P3$ should match a blue point. To check if the above constraints are satisfied, a set of cascade filters are applied. For each filter, the difference between the R , G and B components of two points are computed. If at least five filters are fulfilled, then the pixel $C (X, Y)$ is considered as the center of the marker. The described marker has some drawbacks:

- The points $P1$, $P2$, $P3$ are selected with a fixed distance so the range of distances that can be detected by the camera is very limited. That is to say, if we leave too, the reference points would be outside the marker.
- The geometry does not allow the marker to be detected from any angle. Since each sector occupies 33% of the circle (120°), the maximum angle between the marker and the camera should be within the interval $[-60^\circ, 60^\circ]$.
- The result of the algorithm is a binary matrix where more than one pixel can be selected as the central point of the marker, so there is no single solution. This drawback is not significant within the context of this work, because it is more important to detect the presence of the marker than know its precise location.

Our marker should be suitable for people with large cognitive-motor deficits. The marker will be placed on the hand of the arm that has partial mobility so it should be detectable from any orientation. Because of the lesions of both patients, controlling movement of the arm is limited. In both cases, this implies that the position of the hand relative to the arm varies for each displacement. In addition, displacements are made not only in the horizontal plane but will also involve vertical movements. So, our marker should have the following features: 1) Independence between the viewing distance and the size of the marker, 2) Independence between the rotation angle of the camera and the marker and 3) Precise location of the center of the marker.

With these requirements in mind, we designed a new marker composed of three concentric discs that correspond with the pure colors of the RGB model. With this design, we got the marker meets the second feature required. As it is composed of three concentric circles, we can see the same shape for every rotation angle between camera and marker. The shape of the marker is not enough to fulfill the first and third features, so we use a combination of form, method of detection and localization, as discussed in the following sections.

The prototype of the system consists of a control PC, a screen and a webcam located over the hand of the user to capture his movements (**Figure 2**).



Figure 2. Structure of the prototype developed with the different parts: camera, user screen, and control PC related to the position of the user.

3.2. Marker Detection

In this section we propose a technique for determining the probability that a pixel in the image corresponds to the center of the marker. We have used the color difference technique proposed in [22], so the method is based on the RGB color model. We select regions of pixels taken on the concerned pixel row, and if necessary, we repeat the detection method also vertically (pixel column).

In **Figure 3** (left) we can see a picture of the marker captured by the camera and the line of pixels to consider (white). The graphs in **Figure 3** (right) represent the **R**, **G** and **B** components of the pixels of this line. The pixel position is introduced as the value for the X axis and the value of each component is introduced as the Y axis.

Analyzing the image, we can conclude the following: the values of the components do not reach the maximum theoretical value (256). Moreover, there is always a minimum value of each band, and color transitions are not ideal because soft slopes appear in areas of color change, which causes a small region of uncertainty. Last, the maximum value obtained for each component is different, therefore, the camera is not able to capture all colors equally.

The detection of the marker is performed by means of a matrix $P = \{p_{ij}\}$ where each p_{ij} is the probability that the pixel in position (i, j) represents the center of the marker. In order to avoid unnecessary calculations, we discard ($p_{ij} = 0$) pixels with values for the red component next to the green and blue values. Given a candidate pixel, we calculate p_{ij} from three regions of pixels on the left and right sides of the column j for the row i . Let N be the observed width of any of the disks in the row i , then the region Z_1 corresponds to the red zone of the marker and it is determined considering all the pixels in the interval $[j - N, j + N]$. Similarly, Region Z_2 (green) is determined by $[j - 2N, j - N] \cup [j + N, j + 2N]$ and region Z_3 (blue) is determined by $[j - 3N, j - 2N] \cup [j + 2N, j + 3N]$. Once the three regions $\{Z_1, Z_2, Z_3\}$ have been determined, we take the color values, called $\{P_1, P_2, P_3\}$ (**Figure 4** left). Instead of three isolated points, we consider three regions and their average, which is much more robust.

Then, we calculate p_{ij} by means of several cascade filters accumulating the positive differences between bands. **Figure 4** (right) shows the filters, where **R1**, **G1** and **B1** correspond to the components **RGB** of P_1 ; **R2** and **G2** corresponds to the components **R** and **G** of P_2 ; **R1**, **G1** and **B1** correspond to the components **RGB** of P_3 ; **X** represents the accumulated value of the differences (p_{ij}), and finally, **N** represents the size of the regions considered. As in [22] only some of the combinations of bands are checked because not all of them are significant.

In order to make the marker suitable for detection in multiple scales (different camera distances), the process is repeated for different values of N , and only the maximum likelihood value obtained for each pixel is selected. Last, the process is repeated again for each pixel in the vertical direction (columns) and the maximum value obtained is selected.

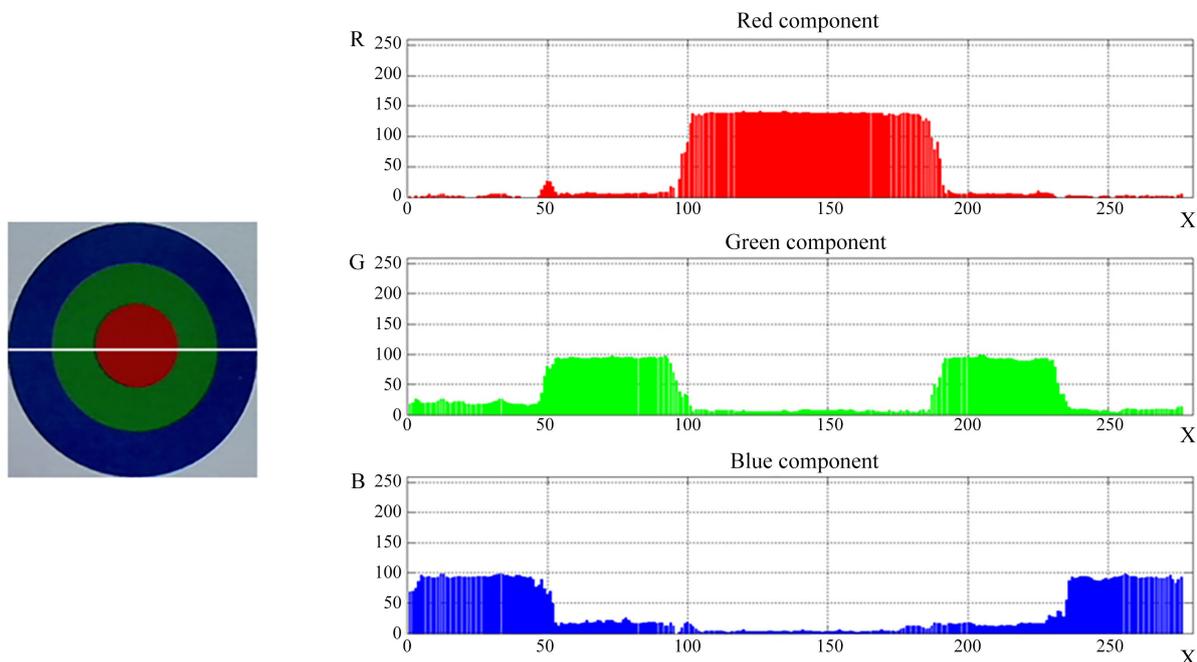


Figure 3. Marker image captured by the camera with the line of pixels to consider marked in white (left). Representation of the RGB values of the line (right).

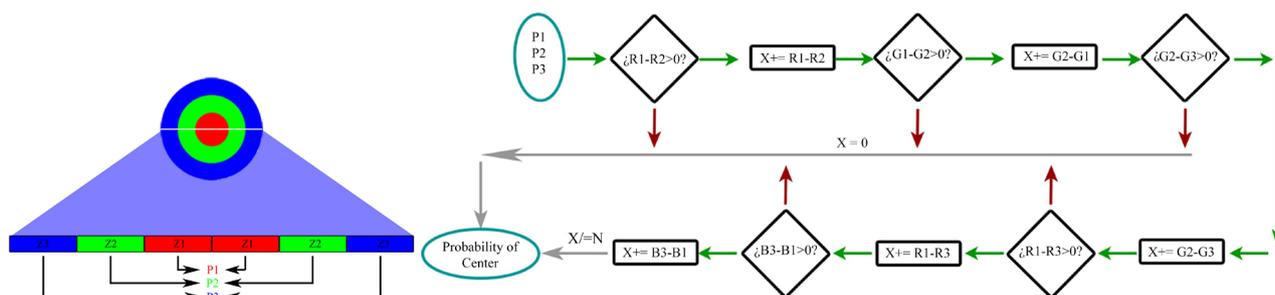


Figure 4. Synthetic marker (left-above). Central pixels row representation (left-below). Cascade filters to estimate probability of a pixel (right).

To determine the location of the marker, we need to figure out which pixel corresponds to the center of the marker from the probability map $P = \{p_{ij}\}$. In order to do that, we have developed an algorithm based on the known Mean Shift technique [23]. Mean shift is an iterative procedure for locating the maximum of a density function given discrete data sampled from that function. It is useful for detecting the modes of this density.

Let $V = \{v_{ij}\}$ be a voting matrix containing the probability that a pixel is located in the center of the marker with much less ambiguity than P . We calculate V from the centroid of a window of size ws , centered in (i, j) for each p_{ij} :

$$(i^*, j^*) = \frac{\sum_{u=i-ws}^{i+ws} \sum_{v=j-ws}^{j+ws} (u, v) P_{uv}}{\sum_{u=i-ws}^{i+ws} \sum_{v=j-ws}^{j+ws} P_{uv}}$$

Such a centroid accumulates the probability observed in p_{ij} :

$$V(i^*, j^*) \leftarrow V(i^*, j^*) + P_{ij}$$

This way, each point “votes” about the mean of its proximity. The main concern is the size of the window w_s , because this value can considerably affect the final result. As we know the size of the probability distributions (N_{\max}), we select this value as the size of the window. As we set $w_s = N_{\max}$, the entire width of the distribution is considered in the voting procedure and then all probabilities assigned to the same marker vote on the same point. This fact allows us to locate the marker very quickly because the algorithm is executed only one time.

In **Figure 5** (top) we can see the result of applying the algorithm to the image of **Figure 5** (bottom). The voting matrix V obtained from the detection matrix P is much less ambiguous, with a peak clearly distinguishable from the rest of points.

4. APPLICATIONS DESCRIPTION

In this section we describe the applications developed to ease the communication with the environment for people with large cognitive-motor disabilities. We developed two different applications: the first one is called IHM Creator and is used by the therapists and relatives to configure the activities. The second one is called IHM Runner; it uses the vision-based UI described in the previous section, and it is devoted to launching the different activities.

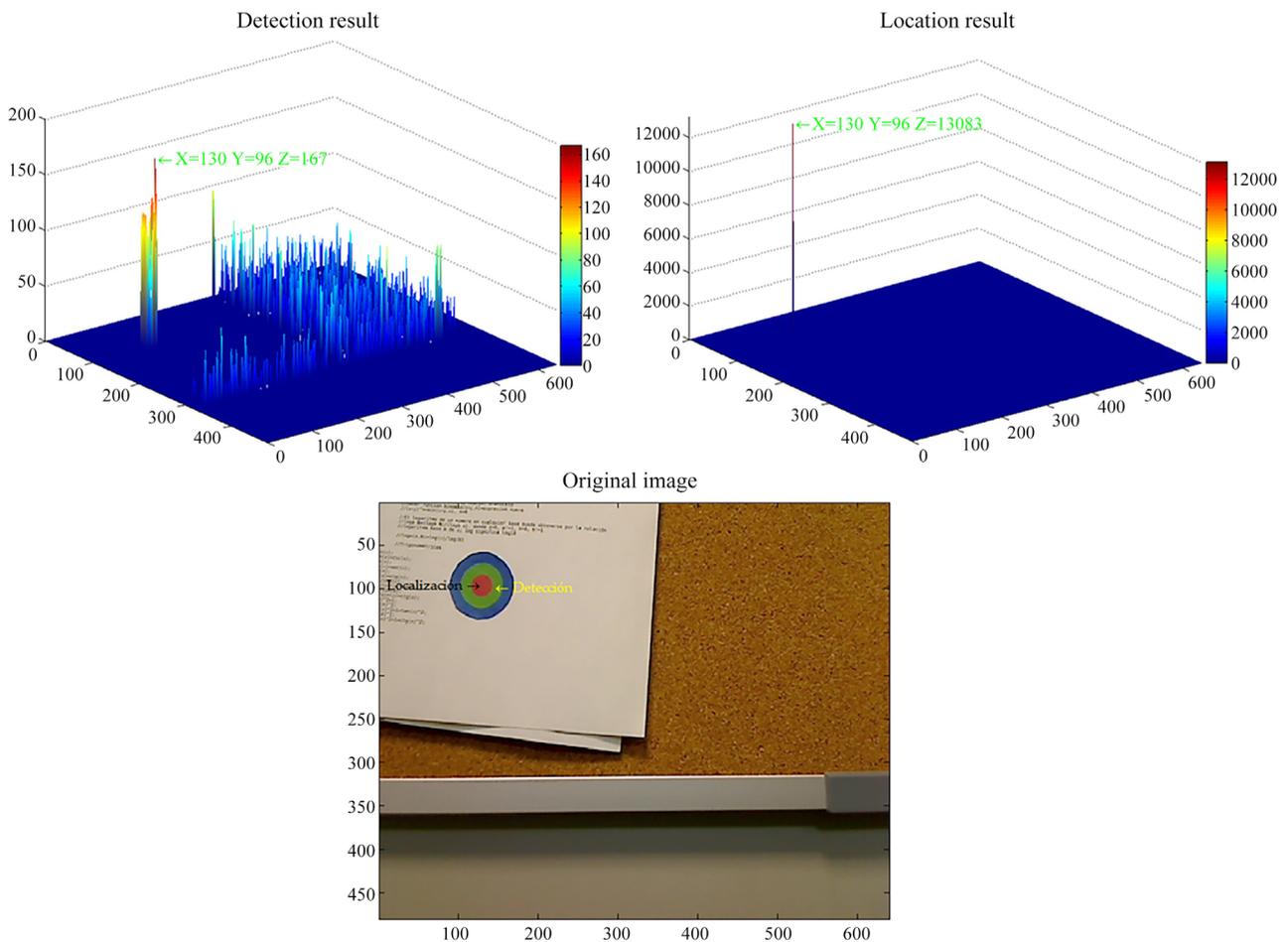


Figure 5. CascDetection matrix P (top left) and voting matrix V (top right) obtained in the location process. Original image (bottom).

After multiple meetings with the team of therapists, we decided the system should have two basic types of activities: the first one shows the patient a list of videos on the screen and then he can select one of them to be played. According to the therapists, the activity shows two or four different videos each time (depending on his/her cognitive abilities). The interface shows options to select, play and stop a video. The second activity allows patients to interact with their environment. The application allows to select simple commands like “Call mom”, “I am hungry”, etc. using a text-to-speech interface.

4.1. IHM Creator

IHM Creator allows both therapists and relatives to configure the activities that will be executed later by the patients. Taking into account the feedback received from the therapists, we decided the features that should be customizable:

- Avatar selection. This option allows to specify the “Avatar” (cursor) and its size. We have found that each patient responds better to a certain type of avatar.
- Activity “Video” configuration. With this option, the user can configure the behavior of the “Video” activity. The video configuration allows therapists to select the audiovisual material that will be used in each session. The screen is divided into four zones that can be activated or deactivated so that the patient can select one of the videos and play it. This way, the number of videos to display is specified depending on the cognitive level of the patient.
- Activity “Text to Speech” configuration. This option allows to associate different images to the texts that the IHM Runner application can play, like “calling mom”, “calling dad” and so on. The UI speaks the text using the same synthetic voice used in IHM Runner. It also allows to specify how many times the text will be reproduced.

The application settings are stored in text files, so that different configurations can be selected in different sessions. **Figure 6** (left-above) shows the cursor type and size selection screens and the UI for the video activity configuration (left-below). **Figure 6** (right) shows the Text-to-Speech configuration interface: Initially the list of images is empty (above left). Then the list of available images is displayed (above right) and finally we can see the appearance of the screen when two of the images have been selected.

4.2. IHM Runner

The activities launcher consists of three different screens. The first one allows to select one of the previously stored configuration files (with IHM Creator), configure some parameters of the webcam and exit the application. If the therapist selects a configuration file for the video activity, a new screen is showed with a list of available videos depending on how it was configured (ranging between one and four). Then, a “Play” blue symbol is displayed over the active sectors. If the camera detects the marker in the image, the cursor is displayed on the corresponding position. In order to start the video playback, the cursor must be placed on the activation zone. When the video is started, it is played in full screen. When it stops, the screen with the list of videos is showed again.

To achieve this behavior, a state machine is designed. The states are as follows: Rest (the cursor is not on any trigger point), Activation (the avatar is located on a trigger point) and Playback (the video is playing full screen).

If the user locates the cursor on any of the activity zones, then the cursor is displayed in red. **Figure 7** (above right) shows the cursor in red color over the first video. If the avatar stays over the same position for long enough, then the state “Playback” is reached, and the video starts playing. During the playback state, the cursor is not displayed on the screen. When it ends, the cursor is displayed on the previous zone, but the activation icon is not showed to avoid that the user can play the video again without moving the avatar.

The design of the Text-to-Speech activity is similar. Therapists have found that many users are able to read and understand written sentences, so we designed an activity capable of read a text associated with an image. This opens a wide range of possibilities for communication concerns.

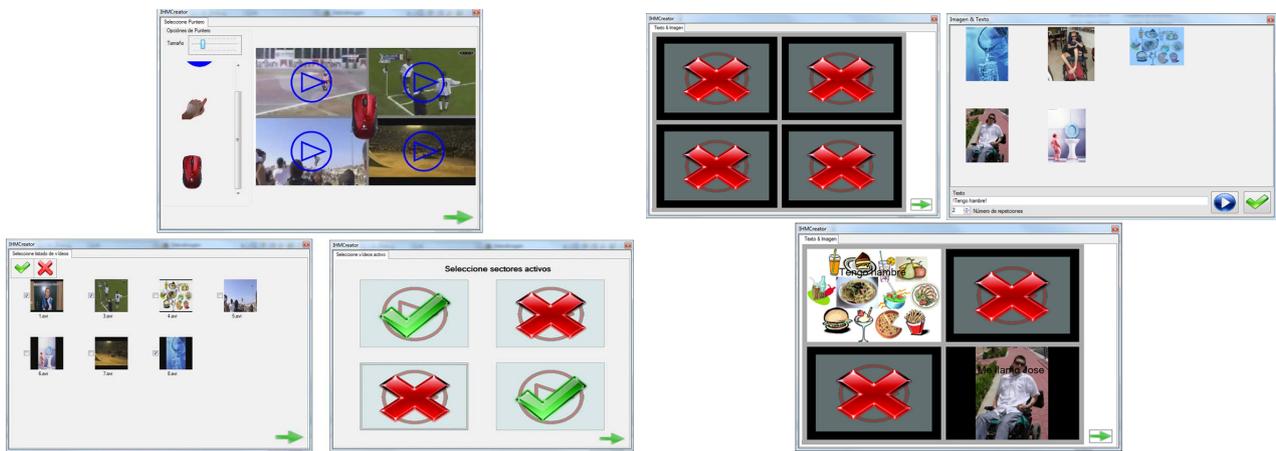


Figure 6. Avatar type and size screen (left-above). Video list selection screen (below left) and active sections screen (left-below). Text-to-Speech configuration interface showing two different commands that one of the patients can select: “I am hungry” and “My name is José”.

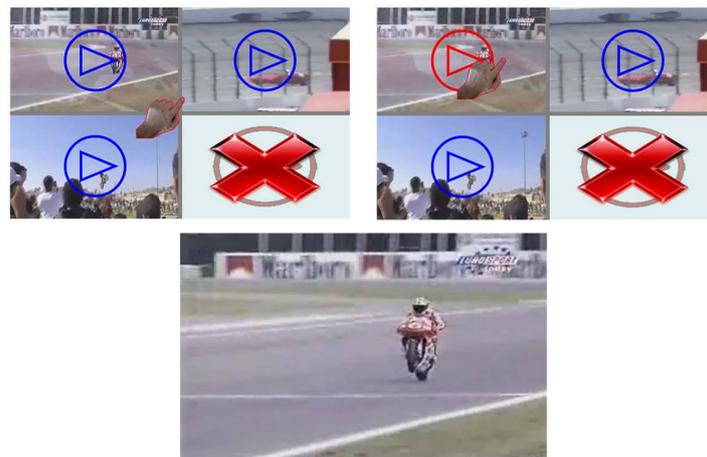


Figure 7. The video activity. Rest state (above left). Activation state (above right) and Playback state (below).

The screen is divided again into four different sections or zones that can be activated or deactivated, depending on the configuration file. This zone shows the image associated to the text that will be “spoken” by the computer. In this case, the state machine has the same two first states than the video activity, but the third one is different, so we call it “Speaking”. **Figure 8** (above left) shows one of the configuration schemes designed, that allows to express basic needs like “I am hungry”, “I am thirsty” or “I am wet”. The text is read by means of a synthetic voice instead of a recording in order to improve portability to other patients and environments. In **Figure 8** (above right) we can see the “Activation” state with the activation symbol changing to red and the “Speaking” state (below) when the application reads the text associated to the selected image (“I am thirsty”, in this case). We can also see that the selected avatar is the marker itself, because some patients find very easy to make the association between the movement of the marker stuck in his hand and the one shown on the screen.

5. EVALUATION

This section is devoted to describing the experiments performed to verify the correct operation of the system. We conducted two kinds of experiments: the first one was performed in the laboratory, with the

aim of testing the correct operation of the marker designed and the second one was conducted to evaluate the user experience using the activities described in the previous sections. This study has been carried out with the express consent of the families of the patients, to whom we thank them for their total collaboration in carrying out the tests.

The first patient, Carlos (**Figure 9** left), aged 26, suffered multiple serious head injury which caused him a subdural hematoma, massive brain edema with effacement and hypointensity of the brainstem cisterns, with a broken left wall and left temporal contusion focus. This produces him a tetraparesis master flexor (muscle stiffness that keeps the body flexing) and severe cognitive and sensory problems. These lesions were caused by a traffic accident in 2007. Since then, he is receiving a multidisciplinary treatment (Occupational Therapy, Physiotherapy and Speech Therapy). Thanks to this treatment he has recovered mobility of the left side, part of head control and he has increased cognitive abilities. Nowadays it has been found that he has certain capabilities such as visually identification of people, objects, colors, shapes, sounds and some words. He can also perform some simple mathematical operations such as addition and subtraction, understand simple commands and read some single words and short phrases.

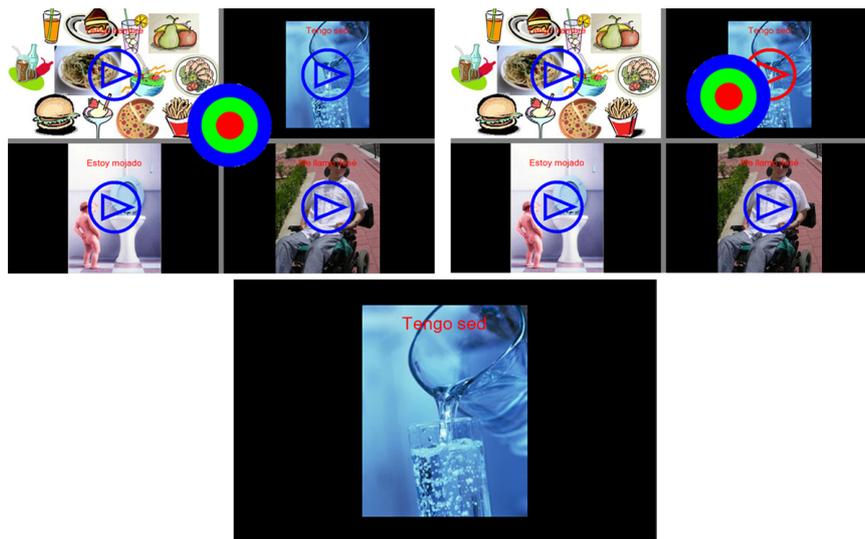


Figure 8. Text to speech activity in rest state (above left), activation state (above right) and speaking state (below) for the text “I am thirsty”.



Figure 9. Carlos (left) and José (Right).

The second one, José (**Figure 9** right) aged 24, suffered a TBI that caused him severe multiple injury in the back right dorsolateral portion of the midbrain. This injury causes mainly spastic tetraparesis with an extensor pattern (muscle stiffness keeps the body in continuous extension) and severe cognitive and sensory problems. As in the previous case, the lesions were caused by an accident. Furthermore, after the accident he was implanted a baclofen pump by surgical (a muscle relaxant acting on the spinal cord to relax the muscles). Through this treatment the spasticity of José has been reduced. The muscle relaxant acts both in the spastic muscles and the others, causing that José (unlike Carlos), keeps the torso and upper limbs completely relaxed, making him difficult to perform voluntary movements. The multidisciplinary treatment that he has been receiving since the accident allowed him to recover part of the mobility of the higher limbs, so his cognitive abilities were increased. Currently, it has been found that he has certain capabilities to visually identify people and objects, distinguish colors, shapes, sounds, words and understand orders of medium/high complexity level. These skills have been discovered by indirect ways, because in both cases they cannot verbalize. Because of this and their cognitive problems, communication with them is a really difficult task.

Training and simply assessing the capabilities of both patients are very complex tasks, requiring specific procedures. It is important to consider that the conditions of these patients do not affect equally all the abilities. For instance, José can answer simple questions with his eye (he looks up for “yes” and down to say “no”) but he still confuses the arm he wants to move (laterality problem). Carlos recognizes his family and can even perform addition and subtraction, but his communication skill is not as good as José’s one. Therefore, we cannot forget that, despite having lost many skills, others remain intact. They are like a puzzle with missing pieces. When brain damage occurs in childhood (e.g. due to a hypoxia) it impacts the development and learning process, so the effects tend to be global (abilities are not completely acquired). However, these patients were adults when they suffered their TBI’s. This is why they retained many abilities, sometimes hidden under their impaired sensory, motor or cognitive.

In this work we have exploited the fact that both patients were regular computer users before the TBI. Therefore, they were used to the mechanics of using a mouse. Objectively, a mouse is very difficult to use since it controls an avatar on the screen through an object which is not directly observed, and it is also in a perpendicular plane (table) to the plane where it operates (screen).

5.1. Laboratory

To check the performance of the detection and localization techniques and the behavior of the developed programs, we conducted several experiments. The first experiment was to test the detection and location of the marker in a complex image with strong color changes. In this context, we consider that an image is complex if there are areas, in addition to the marker, with strong transitions between RGB components.

In **Figure 10** (right) we show an image where the user’s clothes include strong transitions between red and blue. This is an adverse scenario for the detection and localization since these transitions generate high gradients, like the ones generated by the marker. **Figure 10** (left) shows two local maxima for the probability distribution, that hinder the location of the marker. However, after applying the location algorithm we obtain the matrix shown in **Figure 10** (center), where we can see there is just one peak with the global maximum of the function. Therefore, we can say that our marker behaves robustly even under adverse color conditions.

In the second experiment, we checked how the marker behaved to natural lighting changes, gradually moving from daylight to artificial light and how the tool behaves during a large use. When we say “large use” we mean up to four hours, which is the maximum time that the patients can be in their wheelchairs. We kept the system running for six hours in the time slot between 13:00 h and 19:00 h (dark) focusing a fixed marker. During that time, between 13:00 h and 16:00 h the capture was made only with natural light. Between 16:00 h and 18:00 h approximately (sunset), the capture was made with both natural and artificial light. From 18:00 h the capture was performed only with artificial light so in this period lighting conditions did not change.

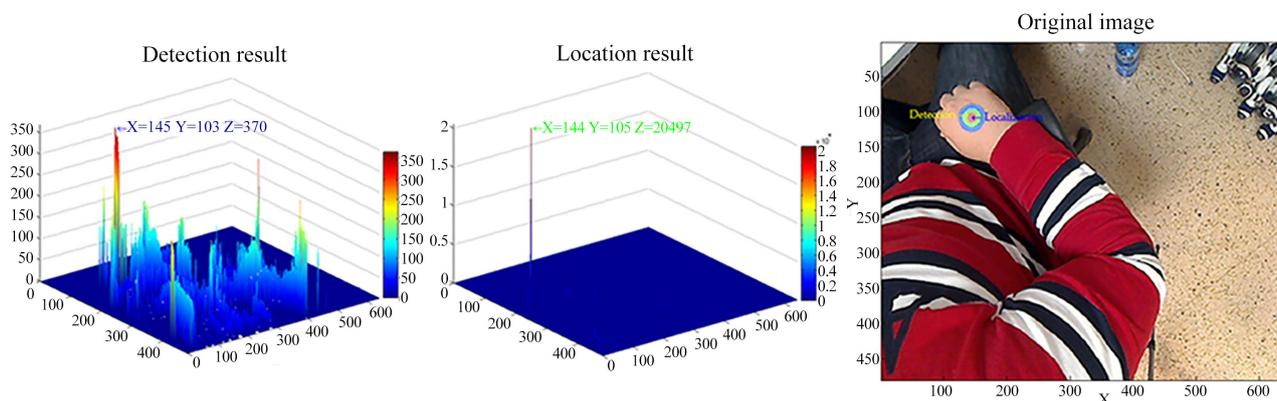


Figure 10. High gradient transitions experiment. Resulting probability matrix (left). Location matrix (center). Original image (right).

We tried to simulate an ordinary situation, in which the camera is facing down, artificial light comes from above (behind the camera) and the natural light is lateral (through the balcony). **Figure 11** shows the probability and location matrices for three representative captures made during the experiment: at 13:00 h with natural light, at 16:00 h with both, natural and artificial light, and at 19:00 h with artificial light only. As we can see, the location matrix is the same, but the contrast with only natural light is high, and therefore, the probability scale is much higher, and some probabilities of low magnitude appear.

We also conducted an experiment in order to test the detection and localization process using multiple markers at different scales. **Figure 12** (left) shows the original image including three markers with three different sizes. The picture in the center shows the detection matrix and the picture on the right side shows the location matrix. As we can see, there are three groups of peaks in the detection matrix corresponding to each marker. In the location matrix each group of peaks became a single peak, so we can conclude that the method is suitable for different scales.

5.2. Users Experience

In this section we describe some experiments conducted at the homes of José and Carlos. Both of them had a working prototype of the system, including all necessary hardware and software. Note that these tests were not performed in a single session, but five consecutive months have been necessary, performing four to five monthly sessions of two to three hours each.

In the first session with Carlos, he was taught the known activity “Simon Says”. Initially he showed some interest, staring at the screen where the application was reproduced. After a certain time (about 30 minutes), Carlos completely lost his interest in the application, although during that time family and therapist tried to motivate him. This test was repeated in two subsequent sessions with different schedules, as cognitive capacity and attention of Carlos varies greatly depending on the day and time, getting the same results. Finally, therapists concluded that the activity was too abstract and childish for him, so it did not offer any kind of motivation to him.

In the second test with Carlos, he was shown the video selection activity. At first, as with the previous activity, he showed a lot of interest, with the difference that this application did not decrease their attention over time, as videos changed continuously, which he found very inspiring.

After many sessions with this activity (it is difficult to manage Carlos working more than an hour and a half because he got tired and lost concentration), the results were slightly encouraging. The greatest progress was that Carlos let someone else guide his hand to activate some video. This indicates that Carlos understood that moving his hand produced a reaction on the screen, because he allowed another person to lead his movement (it’s usually not like that), keeping a great muscular rigidity. However, the activity did not provide enough motivation to get that Carlos moved the hand himself.

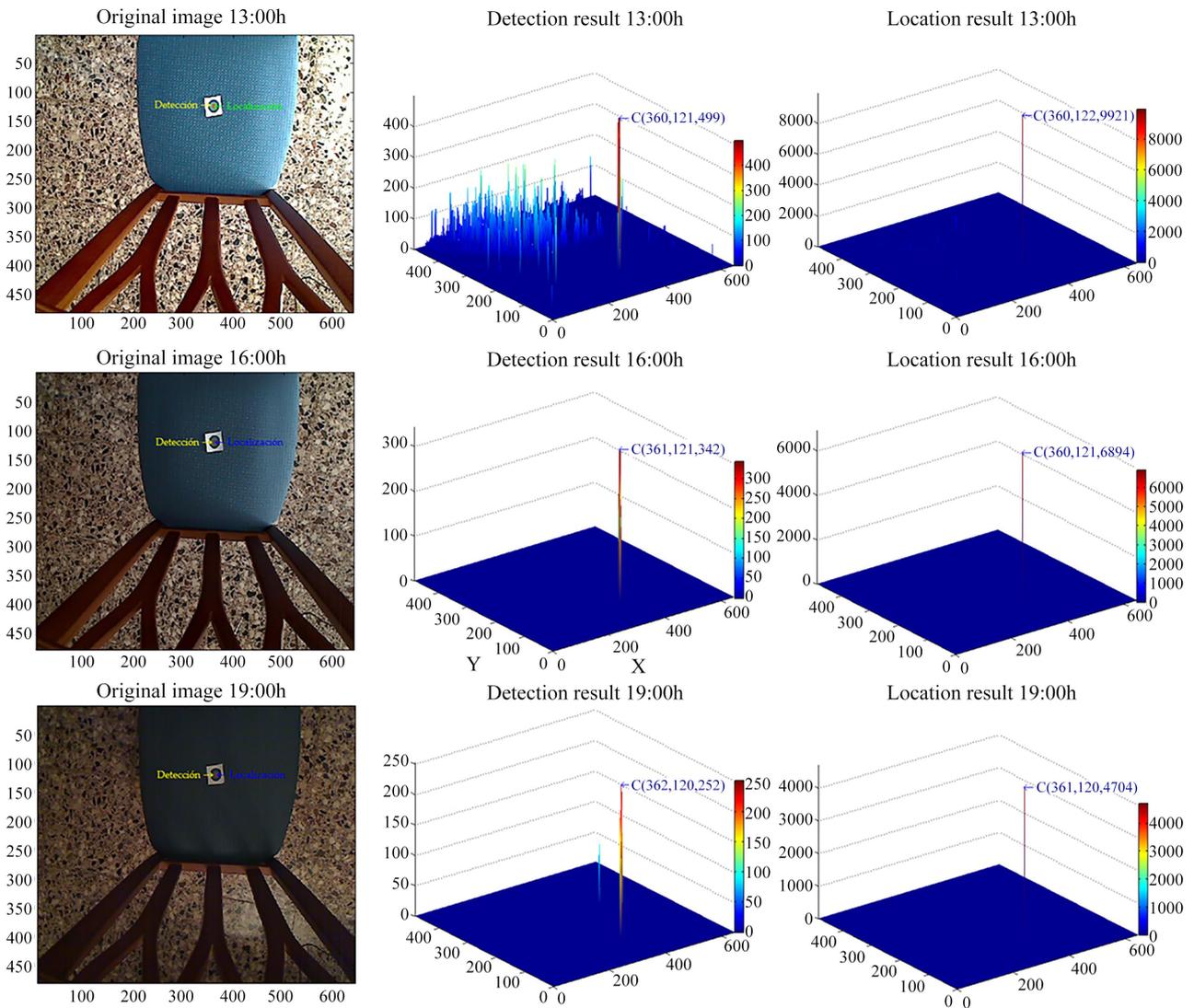


Figure 11. Capture with different illumination conditions (first column), Probability matrix (second column) and Location matrix (third column) for captures performed at 13:00 h (first row), 16:00 h (second row) and 19:00 h (third row).

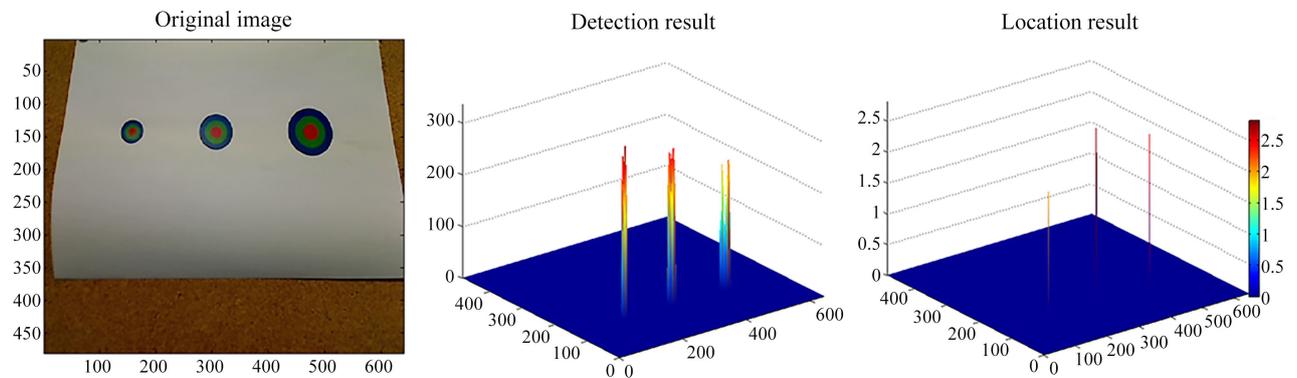


Figure 12. Experiment for multiple markers with different scales. Original image (left), detection matrix (center), and location matrix (right).

In the third test with Carlos, we used the Text to Speech activity. First, the test is performed by replacing the image of his parents by a black screen with the text “Call Mom” and “Call dad”. In this way, we intended to force Carlos to read the contents of the screen, trying to get him more motivated. The results obtained were more successful than those obtained with the Videos activity, but again we did not fulfill the original goal, so we cannot conclude that this movement was completely voluntary. However, Carlos had a greater motivation than with the previous activity, probably because he was able to recognize the usefulness for him.

None of the three tests with Carlos managed to reach the initial objectives of the applications, but little progress had been made. It is difficult to determine a clear reason, but we think it was a motivation problem. Possibly, other applications based on different motivations (the type of Text to Speech) could greatly improve the results.

Some months later, due to a recent change in his treatment, he was still in the process of associating the movement of his hand to the avatar. The new medication kept him conscious but also more altered, making it difficult to pay attention during the use of the system. During the experiments, we focus on strengthening the hand-avatar association. To do this, at times he remained heedful and responsive, we stirred his hand to activate any video or text. Typically, after several demonstrations, we managed that he was able to change some text or video many times during the session.

Figure 13 (first row) shows a picture of Carlos located in the usual position where the system is used, performing an activity of playing a video. The camera is placed over him, focusing on his hand. On the left we see the control PC. In front of him there is a screen in which the activity is displayed. Initially, the state of the activity is rest. In the central part of the figure, we see the same scene, but in this case, we can see that Carlos’ hand is displaced and the avatar is on a trigger point, so the Play symbol appears in red. Finally, the third image of the figure shows how the video is full screen played after the activation time is elapsed.

The fact that Carlos could see the marker attached to his hand, and that it matched with the avatar he was watching on the screen, helped him to understand that the horizontal movements he was performing with his hand, became vertical movements of the avatar on the screen.



Figure 13. Carlos (first row) and José (second row) performing the Video and Text to Speech activities respectively. Each column shows a different state of the activity: Rest (first column), Activation (second column) and Playback/Speaking (third column).

Like Carlos, the motivation of José changes depending on the day and time. Due to other therapeutic activities that both perform throughout the day, in the morning they are usually less tired and therefore more receptive. After communicating directly with José and asking him some questions about the “Simon Says” activity (he is able to answer simple questions with the eyes), we concluded that the application was childish for him, and therefore did not motivate him enough.

The higher cognitive level of José allows him to fully understand the hand-avatar association. As we said before, José finds very difficult to make movements with his hand, due to the lack of strength. However, once past the effort to move his hand at the beginning of the session, he collaborated actively selecting different sectors of each activity by himself.

After two sessions with the Video selection activity, José was able to select any of the four videos voluntarily. To verify this, he was prompted to select each particular video himself, and he got it.

José’s right hand is the one with greater mobility. It is the hand he uses since 2007 to perform most of the work that therapists did with him, because he had not moved voluntarily the other hand hitherto. However, after a few sessions with the system, José began to show more skill with his left hand than with his right one. Later, the sessions were performed with either hand. Note that this evolution of the left hand could be due to some other factor, but therapists attributed it to the motivation provided by the system.

Finally, we describe the results of the activity Text to Speech. In this case, it was not necessary to change the images of the family with texts as in the case of Carlos, because at the time of this test, José had some skill in the use of the application to play videos. The display showed repeated images of José’s mother and sister. Their presence was necessary to motivate the activity. The relative should come when José called them. Every time José called his sister, and she came, he was very pleased.

Figure 13 (second row) shows José performing the activity Text to Speech with the three states: Rest (left), Activation (center) and Speaking (right).

6. CONCLUSIONS AND FURTHER WORK

In this work, a vision-based human-computer interface for people with high cognitive-motor deficits is proposed. Specifically, we focused on patients with severe TBI, whose symptoms include not only motor problems but also cognitive and sensory ones. The developed prototype consists of a camera that reads the horizontal movements of the patient’s hand and played them vertically on the screen, through a system of synthetic markers. The system can be easily configured to the needs of each particular patient. Thus, users should not have a special strength or skill to use the interface.

First, a synthetic color and shape-based marker minimizing the ambiguity with respect to other elements of the scene is designed. This marker is composed of three concentric rings, colored with the three basic components of the RGB model.

The marker detection system provides a matrix of probabilities presenting some local maxima groups around the marker, hindering its location. In order to determine the exact center point of the marker, a MeanShift-based algorithm is developed. The detection and localization processes are very robust and fast (near 30 fps). Furthermore, the method does not require calibration and it is also invariant to changes in illumination.

A module to configure activities like Videos or Text to Speech has been developed. It allows to configure features such as the avatar to use, the list of videos, the screen active actors, the texts associated with the images, etc. This allows the therapist to adapt activities to each session and family planning videos on leisure sessions. Furthermore, we have designed a software able to load and run these configurations using the marker detection and location systems.

The joint use of both applications increases usability of the system in a larger number of situations in everyday life of users. Despite the enormous limitations of this type of patients, the results obtained are encouraging and the designed system is usable and suitable as a complement to the usual therapy that they receive.

As future work, we intend to conduct a clinical trial with several people to evaluate the system with

other patients. On the other hand, it is desirable to extend the number of available activities while maintaining the possibility of setting them.

CONFLICTS OF INTEREST

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

REFERENCES

1. Castao, J. (2002) Plasticidad neuronal y las bases científicas de la neurohabilitación. *Revista de Neurología*, **34**, 130-135. <https://doi.org/10.33588/rn.34S1.2002033>
2. Christensen, A.L. and Uzell, B.P. (1999) *International Handbook of Neuropsychological Rehabilitation*. Springer Science + Business Media, New York.
3. Sitplus Project. <http://sitplus.crea-si.com>
4. Navarro, F.J., Navarro, E. and Montero, F. (2012) HABITAT: Una herramienta para el soporte de actividades interactivas tiles en el tratamiento del daño cerebral. *Proceedings of XIII Interacción 2012*, Elche, 3-5 October 2012, 139-146. .
5. Instituto Guttmann (2011) Universidad Autónoma de Barcelona. PREVIRNEC.
6. Rehacom Project. <http://www.hasomed.de/en/products/rehacom-cognitive-therapy.html>
7. Neuron. Up 2012. <https://www.neuronup.com>
8. Computer Science Canada University of Victoria (2001). <https://webhome.cs.uvic.ca/~mserra/indigo/index.html>
9. Bonneville, E., Muzio, J.C. and Serra, M. (2000) Usability Issues in Software to Assist People with Brain Injuries. *6th ERCIM Workshop "User Interfaces for All"*, Florence, 25-26 October 2000. http://ui4all.ics.forth.gr/UI4ALL-2000/files/Long_papers/Bonneville.pdf
10. Zhou, H. and Hu, H. (2007) Human Motion Tracking for Rehabilitation: A Survey. *Biomedical Signal Processing and Control*, **3**, 1-18. <https://doi.org/10.1016/j.bspc.2007.09.001>
11. Zhou, H., Hu, H. and Harris, N. (2005) Application of Wearable Inertial Sensors in Stroke Rehabilitation. *Engineering in Medicine and Biology Society, EMBS, 27th Annual International Conference*, Shanghai, 1-4 September 2005, 6825-6828.
12. Zhou, H., Hu, H., Stone, T. and Harris, N. (2007) Use of Multiple Wearable Inertial Sensors in Upper Limb Motion Tracking. *Medical Engineering & Physics*, **30**, 123-133. <https://doi.org/10.1016/j.medengphy.2006.11.010>
13. Ayoade, M., Morton, L. and Baillie, L. (2011) Investigating the Feasibility of a Wireless Motion Capture System to Aid in the Rehabilitation of Total Knee Replacement Patients. *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 5th International Conference*, Dublin, 23-26 May 2011, 404-407. <https://doi.org/10.4108/icst.pervasivehealth.2011.246132>
14. Gay, V., Leijdekkers, P. and Barin, E. (2009) A Mobile Rehabilitation Application for the Remote Monitoring of Cardiac Patients after a Heart Attack or a Coronary Bypass Surgery. *Proceedings of the 2nd International Conference on Pervasive Technologies Related to Assistive Environments*, Corfu, 9-13 June 2009, 2-4. <https://doi.org/10.1145/1579114.1579135>
15. Salminen, J., Koskinen, E., Kirkeby, O. and Korhonen, I. (2009) A Home-Based Care Model for Outpatient Cardiac Rehabilitation Based on Mobile Technologies. *3rd International Conference on Pervasive Computing Technologies for Healthcare, PervasiveHealth 2009*, London, 1-3 April 2009, 1-8.
16. Ruiz-Zafra, A., Noguera, M., Benghazi, K., Garrido, J.L. and Cuberos, G. (2013) CloudRehab: Plataforma para la TeleRehabilitación de Pacientes con Daño Cerebral. *Proceedings of XIV Interacción 2013*, Madrid, 17-20 Sep-

tember 2013. 35-38.

17. Okuyama, K., Kawasaki, T. and Kroumov, V.I. (2011) Localization and Position Correction for Mobile Robot Using Artificial Visual Landmarks. *Advanced Mechatronic Systems (ICAMechS)*, 2011 *International Conference*, Zhengzhou, 11-13 August 2011, 455-461.
18. Guo, Y. and Xu, X. (2006) Color Landmark Design for Mobile Robot Localization. *IMACS Multiconference on Computational Engineering in Systems Applications*, Beijing, 4-6 October 2006, 1868-1874. <https://doi.org/10.1109/CESA.2006.313618>
19. Zhang, X., Fronz, S. and Navab, N. (2002) Visual Marker Detection and Decoding in AR Systems: A Comparative Study. *International Symposium on Mixed and Augmented Reality*, Darmstadt, 30 September-1 October 2002, 97-106.
20. Wang, T., Liu, Y. and Wang, Y. (2008) Infrared Marker Based Augmented Reality System for Equipment Maintenance. 2008 *International Conference on Computer Science and Software Engineering*, Volume 5, 816-819. <https://doi.org/10.1109/CSSE.2008.8>
21. Lai, C.H. and Yu, C.C. (2010) An Efficient Real-Time Traffic Sign Recognition System for Intelligent Vehicles with Smart Phones. 2010 *International Conference on Technologies and Applications of Artificial Intelligence (TAAI)*, Hsinchu, 18-20 November 2010, 195-202.
22. Coughlan, J., Manduchi, R. and Shen, H. (2006) Cell Phone-Based Wayfinding for the Visually Impaired. *1st International Workshop on Mobile Vision*, Volume 1. <https://users.soe.ucsc.edu/~manduchi/papers/WayfindingFinal.pdf>
23. Comaniciu, D. and Meer, P. (2002) Mean Shift: A Robust Approach toward Feature Space Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **24**, 603-619. <https://doi.org/10.1109/34.1000236>