IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 25 September 2023, accepted 31 October 2023, date of publication 6 November 2023, date of current version 10 November 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3330689

RESEARCH ARTICLE

Gamification and Virtual Reality for Tongue Rehabilitation

ANTONIO RODRÍGUEZ[1], **MIGUEL CHOVER**[2], **AND IMMA BOADA**[1]
[1]Graphics and Imaging Laboratory, Universitat de Girona, Girona, 17003 Catalonia, Spain
[2]Video Game Research Group, Universitat Jaume I, 12071 Castellón, Spain

Corresponding authors: Antonio Rodríguez (antonio.rodriguez@udg.edu) and Imma Boada (imma.boada@udg.edu)

This work was supported in part by Ministerio de Ciencia e Innovación (MICIN) (Ayudas a proyectos de I+D+i para la realización de pruebas de concepto 2021) under Grant PDC2021-120997-C32, and in part by MICIN under Grant PID2019-106426RB-C31.

ABSTRACT *Purpose:* Lingual exercises based on tongue movements are common in speech therapy. These exercises can be tedious for patients, but gamification and virtual reality (VR) with head-mounted displays (HMD) can serve as effective strategies to enhance their motivation and engagement. However, the use of these technologies can be challenging for therapists due to a lack of technological skills. *Material and Methods:* A new system to support HMD-based VR for gamified tongue rehabilitation exercises is proposed. The system offers a variety of games that challenge users to achieve their goals through tongue movements and sound interaction. These games support different interaction actions that can be set by therapists using easy-to-use editors. The system also provide functionalities for patient follow-up. The system has been implemented and tested considering different technologies such as mobile devices, personal computers, and HMD complemented with an external camera to properly capture the tongue movements. *Results:* The therapists found the system to be user-friendly, requiring no additional support for effective utilization. The system's versatility allows it to be used on mobile devices, as well as with augmented and virtual reality techniques, resulting in more engaging rehabilitation sessions. However, the sensibility of device movements to face detection strategies is a limiting factor of this configuration. In the case of using personal computers with HMD, better results are obtained and especially when virtual reality is considered. In this last case, it is better to consider illuminated scenarios to ensure the proper detection of facial movements. *Conclusion:* HMD-based VR for gamified tongue rehabilitation with ease to use editors to prepare sessions is a good strategy to improve patient engagement in tongue rehabilitation sessions.

INDEX TERMS Virtual reality, gamification, rehabilitation, speech therapy.

I. INTRODUCTION

The tongue plays a fundamental role in speech and requires it to be properly coordinated with other articulatory organs such as the lips, jaw, and palate, for the proper production of sounds and words. This muscular organ also must be controllable, flexible, strong, and with a wide range of movement, among other properties. Unfortunately, these features can be affected by different pathologies and medical conditions including: (i) speech disorders such as dysarthria, apraxia of speech, or articulation disorders [1]; (ii) swallowing disorders [2]; (iii) neurological conditions such as stroke, multiple sclerosis, etc. [3]; (iv) post-surgery or traumatic injuries [4]; (v) delays related to motor or speech development [5]; (vi) neuromuscular disorders [6]; (vii) voice disorders or (viii) age-related changes [7], among others. In these cases, the required rehabilitation sessions include tongue rehabilitation exercises as one of the key components. These exercises are designed to improve tongue strength, mobility, and coordination, and generally consist of a series of tongue movements such as extension and retraction, elevation along the palate, movements side to side, or pushing the tongue in different directions against something. Their simplicity makes them boring for the

The associate editor coordinating the review of this manuscript and approving it for publication was Tai-Hoon Kim.

patients and engagement strategies are required to encourage their practice. In this context, gamification and serious games are suitable approaches to be considered.

Gamification is defined as the use of game dynamics, mechanics, and aesthetics in non-playful environments to acquire, develop or improve an attitude or behavior [8]. One step further from gamification are serious games defined as games that besides entertainment aim to teach or transmit some knowledge or skill [9]. In recent years, both have been applied to the healthcare sector as strategies to improve patient involvement and motivation, promote healthy behaviors, or help patients manage their chronic conditions, among others. Different studies have shown the advantages of their application in this sector and particularly in the rehabilitation field [10], [11].

The idea of combining gamification and serious games with tongue rehabilitation is not new. Different authors proposed to integrate tongue exercises in a computer game environment where tongue movements are required to interact with the game to obtain rewards and overcome the game challenges. In these proposals, the main focus of interest has been centered on the devices used to control tongue movements. Basically, intrusive and non-intrusive approaches have been considered, depending if the contact of some physical component with the tongue is required or not.

Focusing on intrusive devices, the most popular one is the Tongue Drive System (TDS) which is composed of a small magnet secured to the tongue using tissue adhesive and an array of magnetic sensors mounted on headgear that is worn by the subject during the training. Kothari et al. proposed a tongue-controlled computer game for rehabilitation based on this device [12], [13]. Another device is the one proposed by Moreira et al. [14], the T-Station, an intra-oral joystick controlled by the tongue and able to provide variable resistance against tongue movements. This was used as an input device in computer games specifically designed for tongue training. In the study conducted by Schtern et al. [15], they utilized the Wave Speech Research System, which demonstrated remarkable precision in tracking the patient's tongue and other speech articulators within the oral cavity with submillimeter accuracy [16]. This system was employed in a game-based rehabilitation system designed to improve speech abilities. The same system was used by Haworth et al. [17]. Sasaki et al. [18] used electromyography signals acquired with surface electrodes attached at the underside of the jaw to control a tongue therapy computer game. The proposed system was adaptable to different games. More recently, Furlan et al. [14] proposed a new method for tongue rehabilitation with computer games using a joystick device properly adapted to control tongue movements. An evaluation of this device was carried out in [19]. For a survey on devices to track tongue movements in the context of human-computer-interaction strategies see [20].

Focusing on non-intrusive devices, camera-based solutions are the most popular. Miyauchi et al. proposed for tongue rehabilitation a shooter game and a catch-a-fish game controlled by the tongue movements captured using a Microsoft Kinect placed in front of the user's face. There are also RGB camera systems [21] able to track tongue positions but outside the mouth which limits its application scope and more in the context of game interaction. Moreover, there are studies not focused on rehabilitation but specifically aimed at communication that employ different strategies based on controlling tongue movements [22].

Obviously, the control of tongue movements is fundamental to gamify rehabilitation exercises. However, other strategies can be also considered to complement the proposals in order to engage patients. In this context, Virtual Reality (VR) can be considered a suitable strategy to improve patient engagement. VR is an advanced form of human-computer interface that allows the user to interact with and become immersed in a computer-generated environment in a naturalistic fashion. Through the use of technical devices such as head-mounted displays (HMD), desktop computers, video capture systems, tracking systems, and motion-sensing gloves, virtual experiences that closely resemble everyday environments and actions can be created and used for rehabilitation [23]. In recent years, the use of commercial and custom-made VR-based gaming systems for rehabilitation has significantly increased [24]. Proposed systems have been customized to target specific patients' disabilities including reduced motor function, mobility, postural control, or cognitive impairments, among others [25], [26]. In this paper, we will consider the application of VR for tongue rehabilitation using Head-Mounted-Displays (HMD). These devices equipped with multiple built-in cameras can track user movements allowing them to naturally interact with the objects of the virtual environment. Unfortunately, to be applied for tongue rehabilitation, they need to be complemented with an external camera that properly captures the tongue movements. Therefore HMD-VR environments complemented with external cameras will be considered.

The aim of the paper is to present a new system to support HMD-based VR for gamified tongue rehabilitation. To design the system the role of the therapist has been considered as important as the patient one. Particularly, our interest has been focused on non-technological skilled professionals providing an easy-to-use editor with functionalities that make the application of these technologies for the creation of rehabilitation exercises easy for them. The proposed system will be evaluated in a real scenario to demonstrate its well-performance.

Besides this introduction, the paper has been structured as follows. In Section II, the proposed tongue rehabilitation system is described. In Section III the Testing scenario is presented. In Section IV, Results and Discussion are given. Finally, Conclusions and Future Work are presented in Section V.

II. PROPOSED SYSTEM

A. SYSTEM REQUIREMENTS

The system requirements have been defined considering the needs of the therapists, responsible for rehabilitation sessions, and the patients who have to perform the exercises defined by the experts in the context of tongue rehabilitation sessions.

Since rehabilitation exercises are the key to the proposed system we will start presenting their requirements. Particularly, *rehabilitation exercises* have to:

- Take place in a virtual scenario that can be explored using VR with a consumer-grade headset;
- Run on mobile devices or tablets using their cameras or on a PC using a webcam when a VR headset is not available,;
- Be presented as games to motivate patients to practice with challenges to overcome and rewards to receive according to performance;
- Be interactable with tongue movements as the main interaction strategy to achieve exercise goals (presented as game challenges);
- Support sounds to complement tongue interaction to increase playability;
- Support different levels of difficulty to fit different patient profiles; and
- Provide different feedback strategies to fit patient needs.

Regarding rehabilitators, the proposed system has to allow the *rehabilitation team* to:

- Set exercise parameters regarding feedback, visualization mode, and the text information that will be provided to the patient;
- Configure the interaction required to achieve the exercise goals, i.e., tongue movements and sounds;
- Configure the visualization mode according to used technologies including virtual reality, augmented reality, mobile phone, personal computer, etc.
- Create rehabilitation sessions with different exercises;
- Configure the session length by giving the therapist the choice of how long (seconds/minutes) the play session will run for;
- Configure the target difficulty, such as the time between subsequent obstacles, or the magnitude of the required change in position;
- See the state of the rehabilitation session, interact with the objects on behalf of the patient, and change some of the configuration values during the session;
- Obtain information on patient progress and performance;
- Test exercises to see how the patient will see them.

Finally, focusing on *the patients* the system has to allow them to:

- Perform rehabilitation exercises in a game mode;
- Receive feedback on the actions that have to be carried out and also on their performance;
- Adapt the level of challenge to the patient's performance;
- Adapt exercises to patient technological resources since not all the patients will have VR devices;
- Consult their previous sessions to see the progress.

In addition to previous requirements, we consider necessary the support to multiple languages.

B. SYSTEM DESIGN

Taking into account the defined requirements, a system that supports therapists, in the preparation and execution of rehabilitation sessions, and patients, in the performance of rehabilitation sessions, is proposed. Exercises will take place in a game scenario created following the indications of the experts via system editors. Patients equipped with HMD will solve game challenges via tongue movements and sound actions which will be controlled via the device's camera and microphone, respectively. In case of HMD not being availabile computer, tablet and smartphone will be also supported. The system will provide automatic success/failure feedback. The main modules of the system with their interactions are presented in Figure 1 and described below.

1) THE TECHNOLOGY CONTROLLER

To allow the system to run on different technologies it is necessary to adapt the input and output actions to the used technology. To properly perform this module has three components named input, visualization, and output.

- The *Input Controller* obtains and interprets the input received from used technologies. For instance, in the case of using an HMD with an external camera to control face movements, the controller considers inputs from both devices.
- The *Communication Controller* sends data from the Input Controller to the system.
- The *Output Controller* is the game's visualization responsible and adapts the visualization to the technology. For instance, in Virtual Reality environments it controls the cameras and the parallax effects while in Augmented Reality one controls the position of the virtual objects in the real world.

2) THE THERAPIST MODULE

The therapist module has been designed to manage: (i) patients, (ii) rehabilitation exercises, and (iii) rehabilitation sessions. With this aim different functionalities grouped in components and accessed via ready-to-use interfaces that do not need any technical programming expertise to use are provided to the experts. These components are:

- The *Exercise Configuration* component connects with the exercises database allowing the therapist to select a rehabilitation exercise and configure:(i) *Feedback* which can be presented to the patient in a three-dimensional space, or in a two-dimensional one, and the text information that will be provided; (ii) *Patient interaction* referred to tongue movements and sound actions required to achieve the different game challenges of the exercise; and (iii) *Visualization mode* which can

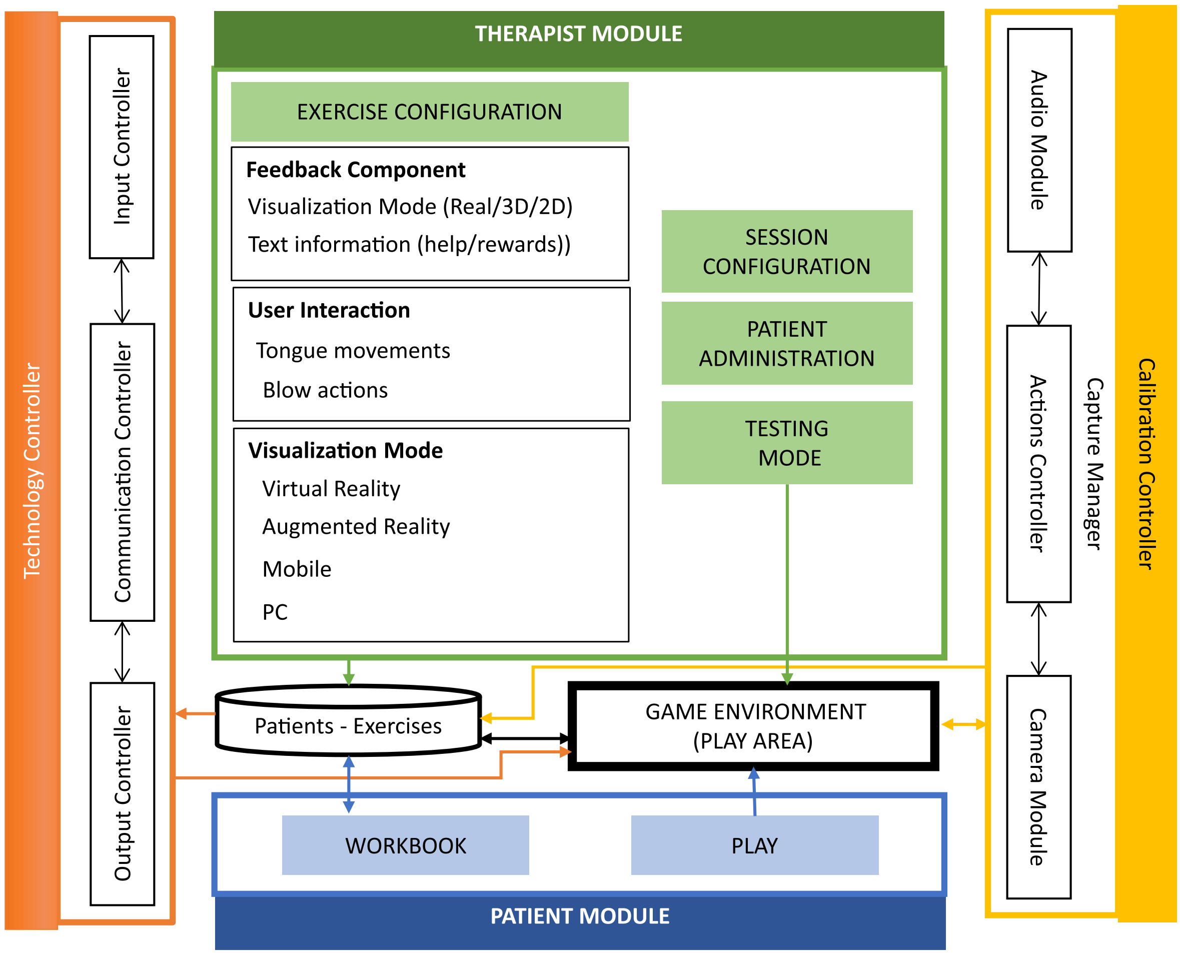

FIGURE 1. **The main modules of the system and the supported profiles.**

be in a Virtual Reality scenario, an Augmented Reality scenario, or via Computer, Tablet or Smartphone.

- The *Session Configuration* component provides functionalities to assign exercises to a patient in the context of a session. The sessions assigned to the patient will define the patient workbook.
- The *Patient Administration* component is used to register the information of the patient and also to follow up on the performance of previous sessions. All this information is registered in the patients' database.
- The *Testing* component allows the therapist to see the exercise as the patient will see in order to check if the proposed configuration is the correct one or not. Note that this module requires a connection with the *Calibration Controller* and *Play Area* components described below.

3) THE PATIENT MODULE
This module is composed of two components designed to provide functionalities to the patient:

- The *Workbook component* with functionalities to access personal data, the sessions (with all the exercises that have to be done and the ones that have been assigned and solved), statistics of the number of errors, time, etc.
- The *Play component* acquires patient actions (tongue movements and sounds) using the camera and microphone and converts them into control signals that are sent to the game scenario to perform the rehabilitation exercise. To properly perform it requires a previous calibration process carried out by the *Calibration Controller*.

4) PLAY AREA
The *Play Area* module presents the rehabilitation exercise to the patient or to the expert as a pre-visualization previous to exercise acceptance accessed via testing mode. The presented exercise is presented by applying the configuration parameters related to actions, feedback, etc. determined by the expert. This module is connected to the *Calibration Controller* module that processes user interaction in the context of the gameplay moment to properly proceed, i.e., feedback, reward update, etc. The Play Area also contains a report zone with information on the current state of the game.

5) THE CALIBRATION CONTROLLER
This module has been designed to control user interactions and transform them into game actions. Since only tongue movements and some sounds are supported it is necessary

a calibration process to configure interaction parameters to better fit patient features. To properly perform this module has three components named audio, camera, and action.

- The *Camera component* is responsible for controlling tongue movements. To properly perform it requires a pre-processing step, carried out with the expert, to calibrate tongue movements. This process, using the device's camera or if required via an external camera connected to the system, acquires images of the patient's face in a normal position, sticking out the tongue to the left, right, up, and down, and also pressing on the palate. Then, using the MediaPipe Face Detection [27] a 3D facial face mesh in each position is obtained. For each mesh, two reference points on the lips, d_t and d_b, (corresponding to the upper lip and bottom one, and used to detect if the mouth is opened or closed), the distance between eyes, d_e, (used to normalize), and the Histogram of Oriented Gradients (HOG) are computed. The HOG counts occurrences of gradient orientation in a localized portion of an image which in our case corresponds to the region from the nose to the chin. To improve performance 160×100 pixels images are considered. All this information is stored in the patient data folder and taken as reference information to process patient actions (see Figure 2).
  During the execution of the rehabilitation exercise, the camera is continuously acquiring information from the patient's face. To control if the patient is in the correct position, the MediaPipe Face Detection [27] obtains the 3D facial face mesh and compares it with the reference images previously acquired. In case of no correct position a message indicating the patient to correct position appears. To identify the patient movement the similarity between reference images and acquired ones is computed by weighting the difference between HOGs. The most similar reference image is selected and the corresponding action is applied.
- The *Audio component* controls the sound intensity. To properly perform this component requires a calibration process to identify minimum and maximum intensity level thresholds for each user. To carry out this process the system needs the patient to first remain in silence and then produce sounds. As in the previous case, under the therapist's control, the system saves the sound intensity when the patient is doing sound action and in silence. Then, during game playing the intensity level is extracted, saved into a buffer, and compared with registered patient thresholds.
- The *Action Controller* obtains information from the audio and camera component and compares it with registered information to determine if it is an acceptable action or not. For instance, a blow action is accepted only when the tongue is in an up position and a blowing sound is also detected, the horse sound is accepted only when the tongue goes from an up position to the down one accompanied by a last hard sound.

C. REHABILITATION EXERCISES

For the therapists to prepare rehabilitation sessions, the following games have been created:

- The *horse race game*. Different horses must compete to reach the finish line first. To advance the patient should hit the tongue against the palate. While advancing they bump into stones along the way that will be thrown at the patient's face and they must be avoided by moving the tongue according to the direction of the stone (see Figure 3(a)).
- The *feed the horses game*. Three horses locked in their stable need to be fed. By sticking the tongue down the food elements that appear in the scenario can be selected. Once caught, food must be given to the horse whose stable color has changed by moving the tongue to the left, right, or up depending on which of the three horses needs food. If the user takes too long, the horse becomes nervous and tries to escape from its stable. To calm it down it will be necessary to hit the tongue against the palate (see Figure 3(b)).
- The *snake charming game*. A snake coiled into a basket has to move following the sound produced by the user by blowing with the tongue folded so that the tip touches the roof of the mouth. When this sound is produced the snake starts to dance with the objective to touch the bell with the same rope colour as the one of the snake (see Figure 3(c)).
- The *way to the stable game*. A map with the way to the stable appears on the screen and the user has to guide the horse to it. The horse will automatically move forward on the map following tongue movements to turn left or right. In case of obstacles along the way, the horse can jump them if the patient hit the tongue against the palate (see Figure 3(d)).

D. SYSTEM INTERFACES

The system provides therapist interfaces to assign exercises to a patient and also an interface to calibrate the camera and sound according to patient features. This calibration process is required once and is done under therapist supervision. The actions that the patient has to perform are defined by default and can be modified using the exercise configuration interface. The main parts of this interface are illustrated in Figure 4. From top to down, and for the Horse race game, the therapist can set the tongue movements (up, down, left, right, against the palate, blow, reproduce horse trot sound and none). The none option allows the game to perform actions automatically and is used in case of patients with minimal interaction degree to familiarize themselves with the game. Regarding feedback, the therapist can select between realistic, illustrative images, or no images. There is also the possibility to add help text.

E. THE SYSTEM IMPLEMENTATION

The proposed system has been implemented with Unity3D [28] requiring to run a device with an internet connection

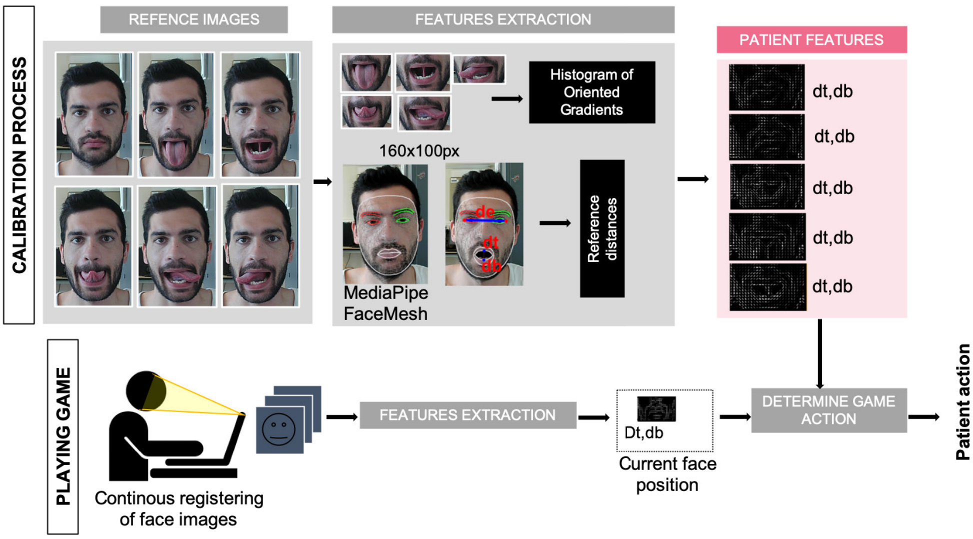

FIGURE 2. **The calibration process acquires face images in the normal position and with the tongue in different positions. Then, extracts reference measures from these images that will be used to control the interaction to be applied during game playing.**

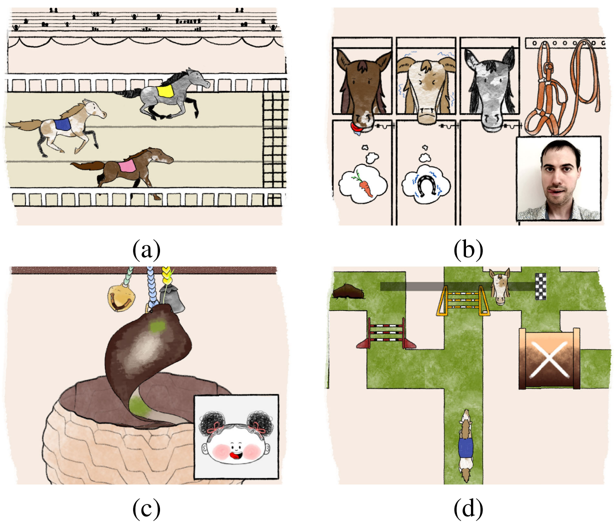

FIGURE 3. **(a) The horse race; (b) Feed the horses; (c) Snake charming and (d) The way to the stable games.**

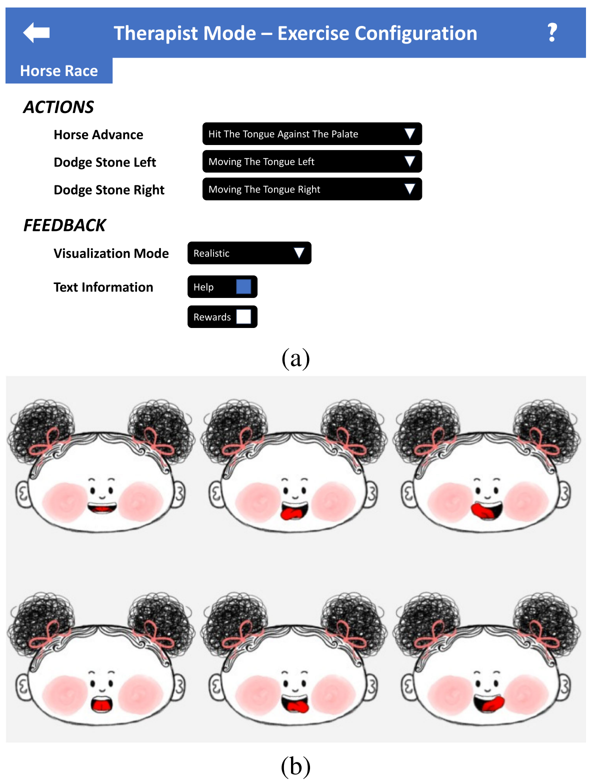

FIGURE 4. **(a) the interface for the therapist to configure interaction mode, feedback, and help messages. (b) Feedback images to guide the tongue movements.**

(to access the repositories of the system to generate the scene and also to store the patient performance); a camera and a micro (to capture patient interaction), and, depending on the configuration, external devices such as the HMD. The implementation follows a client-server architecture (see Figure 5) where the client, using a Unity3D application, controls all the game events and sends, via JSON [29], the information to the server. This analyzes the information, using PHP [30], and administrates the database using MySQL [31].

III. TESTING

To test the system our interest has focused on technological aspects and also on the therapist's role as the creator of the exercises. Patient testing has been postponed to a second phase and described as future work. Taking into account these considerations, three scenarios have been considered to evaluate: (i) System functionalities with respect to therapist needs; (ii) System performance considering different device configurations; and (iii) Face detection and illumination conditions for system support in virtual reality.

A. SYSTEM FUNCTIONALITIES

To evaluate the system's usability three experimented therapists have been considered. These have been asked to prepare

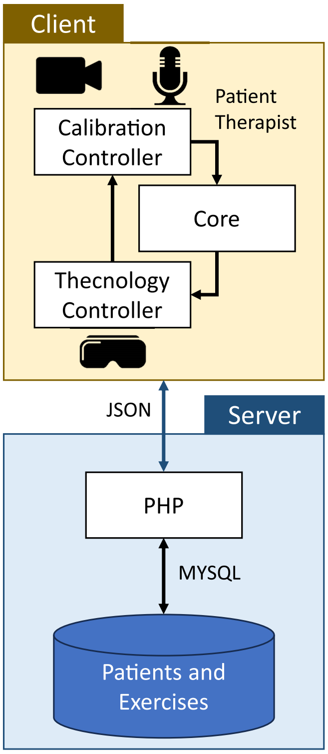

FIGURE 5. The main components of the client-server architecture of the system.

a rehabilitation session using the four available games. Once prepared they have been asked to play and evaluate if they cover the patient's needs. No information was given to them but only access to the system with a brief introduction of each interface (see Figure 4) and the help description that can also be accessed from interfaces. We want to assess if the information provided in the configuration session and rehabilitation exercises interfaces is enough.

B. SUPPORTED TECHNOLOGIES

The Unity implementation of the system allows this project to compile in different platforms and hence the system to run in different technologies. To evaluate the system performance different configurations have been considered in a laboratory session with technology experts as users. These configurations, illustrated in Figure 6, consider the system running on: (i) a mobile device, which can be a tablet or a smartphone, and using its frontal camera and microphones to acquire user interaction; (ii) a mobile device with a back camera to allow game elements to be combined with the real elements of the room where the user performs rehabilitation; (iii) a mobile device with a HMD such that the game is executed on the mobile device and visualized in the HMD; (iv) a desktop computer with a WebCam and a microphone to acquire user interaction; (v) a desktop computer connected to a HMD; and (vi) a desktop computer connected to a HMD with illumination on user's face to better acquire information. The three last scenarios are used to evaluate the effect of illumination conditions on user performance.

Since the user is equipped with a HMD, the illumination conditions are fundamental for the system to properly detect the reference points such as eyes, nose, and mouth required to interpret the tongue movements. To evaluate the effect of illumination, the three last scenarios (iv, v, and vi) have been considered playing each game, in sessions of 60 seconds. In these sessions, the tongue movements have been registered and the movements that have been detected incorrectly have been counted.

To evaluate these scenarios three parameters will be considered: (i) the *Cost* which measures the additional development cost caused by the use of technology. Possible values are low, medium, and high depending on extra implementation hours. (ii) *Robustness* that measures the capabilities of the system to detect user interaction. Possible values are very low, low, medium, and high depending on the detected interactions. (iii) *Novelty* that measures the novelty of used technology with low, medium, high, and very high as possible values.

IV. RESULTS

The results will be presented considering the testing scenarios designed to evaluate the system functionalities and the performance of different technologies with respect to tongue movement acquisition.

A. SYSTEM FUNCTIONALITIES

The three therapists were able to create a rehabilitation session with the four exercises and with no need for help other than the one provided by the system interface. The time to create the sessions ranged from 80 seconds (when default options were considered) to 240 seconds (when all the options were personalized). They were also able to calibrate the system with no need for help considering the calibration with the patients an easy process.

Created sessions were tested by the therapist and all of them agree that sessions are motivating for the patients and that exercises can be ease adapted to the needs of the patients. All of them agree that the system is a good complement to rehabilitation sessions.

B. EVALUATION OF SUPPORTED TECHNOLOGIES

The evaluation of considered technologies with respect to *Development Cost*, *Robustness* and *Technology novelty* is presented in Table 1

Although the system correctly runs on a mobile device, the errors in face detection due to device movements are the main limitation of this option. Moreover, when augmented or virtual reality is considered the user motivation increases but also the sensitivity of the system to device movements. In addition, the development cost with virtual reality increases due to communication requirements between mobile devices and virtual reality glasses. In this last case, the detection of face movements is still more difficult. Therefore, to use the system in a mobile device the

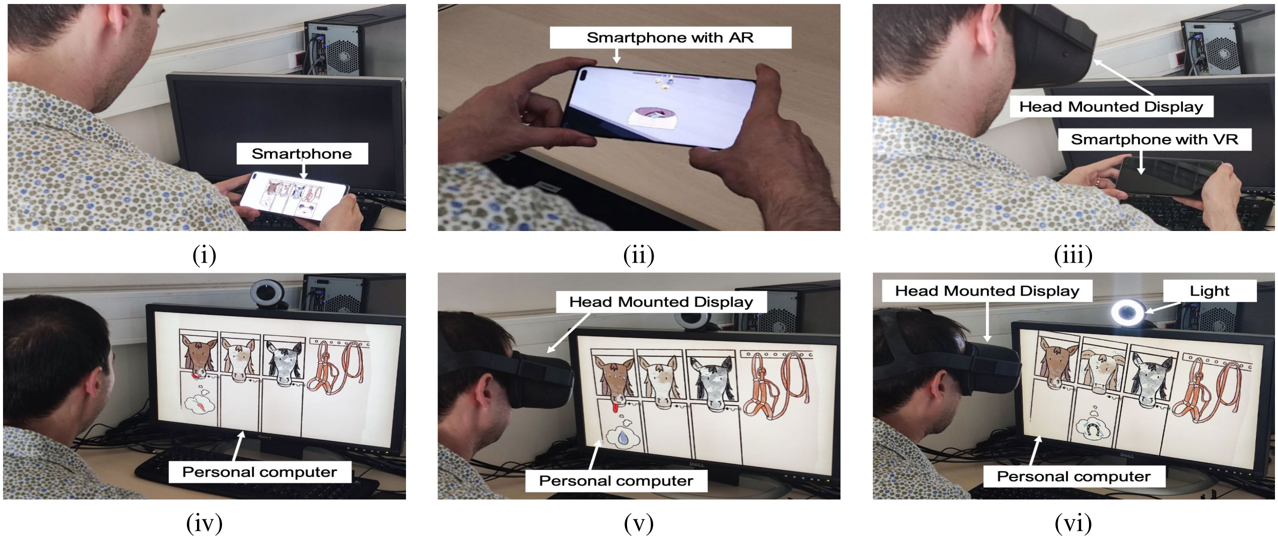

FIGURE 6. The different configurations that have been considered to test the system.

TABLE 1. Tested scenarios evaluated in laboratory and considering development cost, system robustness and technology novelty as assessment measures.

Technology	Cost	Robustness	Novelty
Mobile device	low	medium	medium
Mobile device with Augmented Reality	medium	low	very high
Mobile device with Virtual Reality	high	very low	very high
Desktop Computer	low	high	low
Desktop Computer with Virtual Reality and no illumination	medium	low	high
Desktop Computer with Virtual Reality and illumination	medium	medium	high

device movements have to be minimal to ensure the proper performance.

Using the desktop computer system, the extra mobility and motivation that virtual and augmented reality in a mobile device can give are lost. However, this becomes an advantage for the system since the calibration of cameras and microphones becomes more robust as well as the face detection process. Note that the extra cost of a WebCam and a microphone is required. Adding virtual reality to this configuration increases user motivation with no additional cost. Again the main limitation is on the face detection process to locate the tongue since virtual glasses create shadows on the face. To tackle this problem a light can be added to ensure the proper detection of reference points such as eyes, nose, and mouth. From these experiments, the desktop configuration with virtual reality and proper illumination is considered the best one.

C. IMPACT OF VR TECHNOLOGIES ON USER INTERACTION DETECTION

The last experiment considers different scenarios including a personal computer with and without HMD and with and without external illumination, a mobile device with and without HMD, and with and without external illumination. and evaluate their capabilities to detect user interaction. For each scenario, a rehabilitation session with the user playing each game in sessions of 60 seconds has been considered. An observer counted the number of actions that were correctly, incorrectly, and not detected (when the user performs the action and after 3 seconds the system has not been able to detect it) by the system. The average values for each scenario are collected in Figure 7 where blue, orange, and grey bars correspond to correct, incorrect, and no detected actions. Illum indicates a scenario with extra illumination to better detect the face.

Note that the best results are obtained with the personal computer with the HMD and using an illumination device to better detect the reference face positions, i.e., *PC-HMD Illum* case. For Mobile device configurations, the best one is with a device without HMD and with no extra illumination. This case only requires the device to be in a fixed position with stable and sufficient illumination to detect the face.

Comparing the different scenarios using a personal computer or a mobile device, it can be seen that in the case of using HMD the personal computer is the best option. However, in case of not using HMD the best one is achieved with the mobile device. This only requires the device to be in a fixed position. This is an advantage for the proposed system since the conditions required to properly perform are minimal.

Regarding worst results, these are achieved with the personal computer and the mobile device with HMD and no illumination conditions. The HMD make the detection of face reference points more difficult for the system.

D. FINAL REMARKS

Although a more detailed assessment is required, to end this section, some remarks regarding the cost-effectiveness

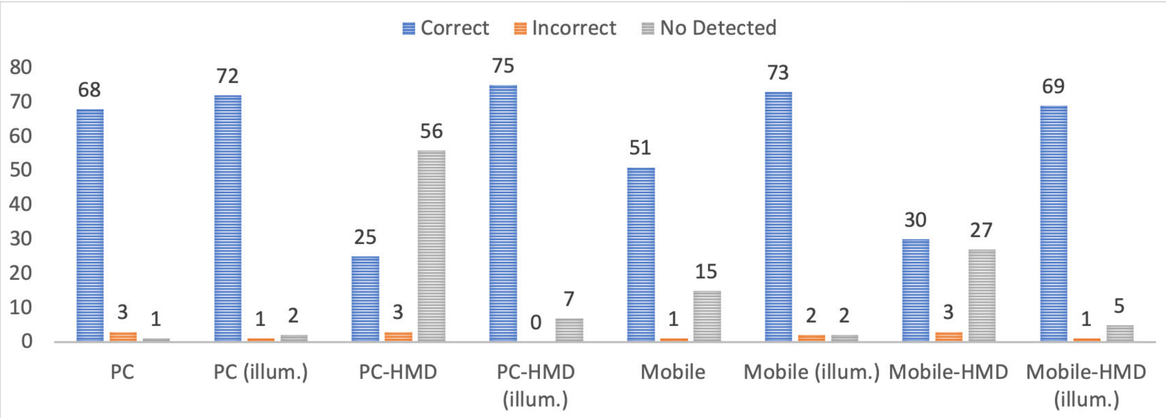

FIGURE 7. The different configurations that have been considered to test the system where PC represents personal computer and *Illum.* represents a scenario with external illumination to better detect face reference points.

of the system will be considered. The benefits of virtual environments and their positive cost-effectiveness in relation to traditional rehabilitation sessions have been pointed out by several authors [32], [33]. Our system also provides different advantages that directly impact the rehabilitation cost-effectiveness. Particularly, it provides functionalities to ease the creation of personalized rehabilitation exercises that adapt to individual progress while requiring fewer human resources. Moreover, it allows a single rehabilitator to control different patients simultaneously reducing in this way labor costs. The accessibility of virtual rehabilitation sessions from home minimizes the expenses associated with commuting and facility upkeep. Focusing on patients, the gamification of rehabilitation exercises is expected to engage them, and as a consequence motivation and adherence to the treatments will increase leading to potentially faster recovery times. Regarding hardware needs, the system supports different configurations to accommodate different budgets. From Figure 6 it can be seen that the system can run on a very economical configuration with a smartphone (with a cost near two hundred euros) to a more advanced one with a computer with a camera and a loudspeaker and virtual reality glasses (with a cost over one thousand euros). In all cases, the first inversion on hardware will be offset by long-term savings. Therefore, the proposed system is a cost-effective solution for modern rehabilitation needs.

V. CONCLUSION AND FUTURE WORK

Patient engagement is critical to a successful recovery in tongue rehabilitation sessions. However, the simplicity of the exercises makes them boring for the patients and they usually lose interest. Gamification and virtual reality strategies are alternatives for creating more engaging sessions for patients. Moreover, the reduction in technology costs makes the implementation of these solutions more affordable. Unfortunately, the lack of technological skills of many therapists makes the application of these solutions challenging, and easy-to-use frameworks are required. In this paper, a new system that integrates gamification, virtual reality, and easy-to-use content editors for the therapist to prepare gamified tongue rehabilitation exercises has been proposed. The system has been tested in a real scenario considering different technologies and user profiles obtaining very satisfactory results.

Our future work will be centered on the extension of the system to provide more exercises and support to other facial movements. In addition, deeper testing will be carried out considering different profiles of patients and therapists. The evolution of patient improvements will be also evaluated as well as the cost-effectiveness of the system. Finally, more advanced solutions based on artificial intelligence strategies will be explored to automate the calibration process that is currently required by the system. It is also planned to compare the system with other approaches.

DECLARATIONS

A. CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

B. OPEN ACCESS

DATA AVAILABILITY

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

REFERENCES

[1] S. Hajesmaeel-Gohari, S. Goharinejad, E. Shafiei, and K. Bahaadinbeigy, "Digital games for rehabilitation of speech disorders: A scoping review," *Health Sci. Rep.*, vol. 6, no. 6, p. e1308, Jun. 2023.

[2] W. Zhao, *Rehabilitation Therapy of Neurological Training of Swallowing Dysfunction*. Singapore: Springer, 2019, pp. 467–480.

[3] A. S. Mefferd and M. S. Dietrich, "Tongue- and jaw-specific articulatory changes and their acoustic consequences in talkers with dysarthria due to amyotrophic lateral sclerosis: Effects of loud, clear, and slow speech," *J. Speech, Lang., Hearing Res.*, vol. 63, no. 8, pp. 2625–2636, Aug. 2020.

[4] T. Nakanishi, Y. Yamamoto, K. Tanioka, Y. Shintani, I. Tojyo, and S. Fujita, "Effect of duration from lingual nerve injury to undergoing microneurosurgery on improving sensory and taste functions: Retrospective study," *Maxillofacial Plastic Reconstructive Surgery*, vol. 41, no. 1, p. 61, Dec. 2019.

[5] N. L. Potter, Y. Nievergelt, and M. VanDam, "Tongue strength in children with and without speech sound disorders," *Amer. J. Speech-Lang. Pathol.*, vol. 28, no. 2, pp. 612–622, May 2019.

[6] G. Umemoto, "Tongue dysfunction in neurological and neuromuscular disorders: A narrative literature review," *World J. Otorhinolaryngol.*, vol. 5, no. 2, pp. 58–64, 2015.

[7] K. Iyota, S. Mizutani, S. Oku, M. Asao, T. Futatsuki, R. Inoue, Y. Imai, and H. Kashiwazaki, "A cross-sectional study of age-related changes in oral function in healthy Japanese individuals," *Int. J. Environ. Res. Public Health*, vol. 17, no. 4, p. 1376, Feb. 2020.

[8] S. Deterding, D. Dixon, and R. Khaled, "Gamification: Toward a definition," in *Proc. ACM CHI Conf. Human Factors Comput. Syst.* Vancouver, BC, Canada: ACM, 2011, pp. 12–15.

[9] D. Michael and S. Chen, *Serious Games: Games That Educate, Train, and Inform*. Muska & Lipman/Premier-Trade, 2006.

[10] D. Avola, L. Cinque, G. L. Foresti, and M. R. Marini, "An interactive and low-cost full body rehabilitation framework based on 3D immersive serious games," *J. Biomed. Informat.*, vol. 89, pp. 81–100, Jan. 2019.

[11] C. S. González-González, P. A. Toledo-Delgado, V. Muñoz-Cruz, and P. V. Torres-Carrion, "Serious games for rehabilitation: Gestural interaction in personalized gamified exercises through a recommender system," *J. Biomed. Informat.*, vol. 97, Sep. 2019, Art. no. 103266.

[12] M. Kothari, P. Svensson, X. Huo, M. Ghovanloo, and L. Baad-Hansen, "Force and complexity of tongue task training influences behavioral measures of motor learning," *Eur. J. Oral Sci.*, vol. 120, no. 1, pp. 46–53, Feb. 2012.

[13] M. Kothari, P. Svensson, J. Jensen, T. D. Holm, M. S. Nielsen, T. Mosegaard, J. F. Nielsen, M. Ghovanloo, and L. Baad-Hansen, "Tongue-controlled computer game: A new approach for rehabilitation of tongue motor function," *Arch. Phys. Med. Rehabil.*, vol. 95, no. 3, pp. 524–530, Mar. 2014.

[14] R. M. M. M. Furlan, G. A. Santana, W. F. Bischof, A. R. Motta, and E. B. de Las Casas, "A new method for tongue rehabilitation with computer games: Pilot study," *J. Oral Rehabil.*, vol. 46, no. 6, pp. 518–525, Jun. 2019.

[15] M. Shtern, M. B. Haworth, Y. Yunusova, M. Baljko, and P. Faloutsos, "A game system for speech rehabilitation," in *Proc. Motion Games, 5th Int. Conf.* Rennes, France: Springer, Nov. 2012, pp. 43–54.

[16] J. J. Berry, "Accuracy of the NDI wave speech research system," *J. Speech, Lang., Hearing Res.*, vol. 54, no. 5, pp. 301–1295, 2011.

[17] H. Brandon, H. Elaine, K. Elaine, K. Melanie, B. M. A. B. Show, and Y. Yunusova, "Electromagnetic articulography in the development of 'serious games' for speech rehabilitation," in *Proc. Int. Workshop Biomechanical Parametric Model. Human Anatomy (PMHA)*, 2014, pp. 1–8.

[18] M. Sasaki, K. Onishi, T. Arakawa, A. Nakayama, D. Stefanov, and M. Yamaguchi, "Real-time estimation of tongue movement based on suprahyoid muscle activity," in *Proc. 35th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2013, pp. 4605–4608.

[19] M. Amaral, R. Furlan, G. Santana, and C. Figueiredo, "Feasibility of using a tongue-activated joystick by healthy pre-teen children," *Int. J. Orofacial Myol.*, vol. 45, no. 1, pp. 13–30, Nov. 2019.

[20] Y. Cao, H. Chen, F. Li, and Y. Wang, "CanalScan: Tongue-jaw movement recognition via ear canal deformation sensing," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, May 2021, pp. 1–10.

[21] S. Niu, L. Liu, and D. S. McCrickard, "Tongue-able interfaces: Prototyping and evaluating camera based tongue gesture input system," *Smart Health*, vol. 11, pp. 16–28, Jan. 2019.

[22] T. Yu, "Application of B+M-mode ultrasonography in assessing deglutitive tongue movements in healthy adults," *Med. Sci. Monitor*, vol. 21, pp. 1648–1655, Jun. 2015.

[23] C. M. Hayre, D. Müller, and M. Scherer, *Virtual Reality in Health and Rehabilitation*. Boca Raton, FL, USA: CRC Press, 2021.

[24] M. C. Howard, "A meta-analysis and systematic literature review of virtual reality rehabilitation programs," *Comput. Hum. Behav.*, vol. 70, pp. 317–327, May 2017.

[25] K. Tornbom and A. Danielsson, "Experiences of treadmill walking with non-immersive virtual reality after stroke or acquired brain injury—A qualitative study," *PLoS ONE*, vol. 13, no. 12, 2018, Art. no. e0209214.

[26] K. Lohse, C. Hilderman, and K. Cheung, "Virtual reality therapy for adults post-stroke: A systematic review and meta-analysis exploring virtual environments and commercial games in therapy," *PLoS ONE*, vol. 13, no. 12, 2014, Art. no. e93318.

[27] MediaPipe. *MediaPipe Face Mesh*. Accessed: Jan. 31, 2023. [Online]. Available: https://google.github.io/mediapipe/solutions/face_mesh.html.

[28] Unity Technologies. *Unity Real-Time Development Platform*. Accessed: Nov. 29, 2022. [Online]. Available: https://unity.com/

[29] F. Pezoa, J. L. Reutter, F. Suarez, M. Ugarte, and D. Vrgoč, "Foundations of JSON schema," in *Proc. 25th Int. Conf. World Wide Web*, Apr. 2016, pp. 263–273.

[30] S. S. Bakken, Z. Suraski, and E. Schmid, *Php Manual*, vol. 1. iUniverse, Incorporated, 2000.

[31] Oracle. *Mysql*. Accessed: Jan. 31, 2023. [Online]. Available: https://www.mysql.com/

[32] F. Fatoye, T. Gebrye, C. E. Mbada, C. T. Fatoye, M. O. Makinde, S. Ayomide, and B. Ige, "Cost effectiveness of virtual reality game compared to clinic based McKenzie extension therapy for chronic non-specific low back pain," *Brit. J. Pain*, vol. 16, no. 6, pp. 601–609, Dec. 2022.

[33] R. Lloréns, E. Noé, C. Colomer, and M. Alcañiz, "Effectiveness, usability, and cost-benefit of a virtual reality-based telerehabilitation program for balance recovery after stroke: A randomized controlled trial," *Arch. Phys. Med. Rehabil.*, vol. 96, no. 3, pp. 418–425, 2015.

ANTONIO RODRÍGUEZ received the M.Sc. and Ph.D. degrees in computer science from Universitat de Girona, in 2016 and 2020, respectively. He is currently an Associate Professor with the Computer Science and Applied Mathematics Department, Universitat de Girona, and a Researcher with the Graphics and Imaging Laboratory. His research interests include serious games, e-learning, and medical applications, among others.

MIGUEL CHOVER received the Ph.D. degree in computer science from the Polytechnic University of Valencia, in 1996. He is currently a Professor with the Department of Computer Languages and Systems, Jaume I University, Castellón. He is also the Director of the Video Game Research Group and a member of the Institute of New Imaging Technologies (INIT), Jaume I University. His research interests include geometric modeling, interactive visualization, and video game technology. He is an active member of the Spanish Association of Computer Graphics (EUROGRAPHICS S.E) and a member of the Executive Committee of the Spanish Society for Video Game Sciences (SECiVi).

IMMA BOADA received the Ph.D. degree in computer science from the Polytechnic University of Barcelona, in 2001. She is currently a Professor with the Computer Science and Applied Mathematics Department, Universitat de Girona, Catalonia. She is also the Director of the Institut Informàtica i Aplicacions. She is a member of the Graphics and Imaging Laboratory, Universitat de Girona, where she carries out research and development on visualization, image processing, serious games, and e-learning.