

Design of a Motion-based Evaluation Process in Any Unity 3D Simulation for Human Learning

Djadja Jean Delest Djadja, Ludovic Hamon and Sébastien George

LIUM - EA 4023, Le Mans Université, 72085 Le Mans, Cedex 9, France

Keywords: Virtual Learning Environment, Activity Evaluation, Motion Analysis, Unity.

Abstract: This paper discusses the usability of a generic method for the evaluation of the user activity in Virtual Learning Environments (VLE) and its implementation with Unity. In the context of motion-based tasks, the learning process relies on the observation and imitation of the task demonstrated by the teacher. The learner task is compared to the teacher one in terms of: (a) motions shape of the user and the manipulated artefacts and (b), the sequential order of 3D checkpoints that the user must collide with. The integration of the evaluation system into any existing VLE rises challenges regarding the system architecture and the Human Computer Interface to set up the evaluation process. A usability test related to the design of this process is conducted for a pool shooting, a dart throwing and a letter writing simulation. The preliminary results show that: (i) the integration of an existing VLE into the evaluation system is feasible despite issues related to the interaction assets and (ii), all participants are satisfied by their designed evaluation process for pool shooting and dart throwing, they were unable to set up a satisfying evaluation for letter writing due to scale issues.

1 INTRODUCTION

Virtual Reality (VR) technologies has been used to create effective Virtual Learning Environments (VLEs) in various domains such as sports training (Miles et al., 2012; Le Naour et al., 2019), education (Mikropoulos and Natsis, 2011), surgery (Bric et al., 2016; Roy et al., 2017), industry (Patle et al., 2019). From the motion-based traces generated by the final user, to the given advice by the system, thanks to relevant multimodal (*e.g.* visual, auditory, haptic), a complete processing chain can be set to evaluate the user activity.

Some VLEs evaluate the learner by detecting an actions sequence/pattern each action being composed of one or several gestures (Baudouin et al., 2007; Toussaint et al., 2015; Mahdi et al., 2019). However, their analysis are mainly based on the discrete states of the manipulated objects and not on the performed gestures. Another strategy consists in demonstrating the manual gestures to learn, by the expert (through an animated 3D avatar) that can be manually or automatically compared with those of the learner (Ng et al., 2007; Miles et al., 2012; Morel et al., 2016; Kora et al., 2015; Le Naour et al., 2019). An intuitive and understandable evaluation process can be set up even though this kind of system are mostly dedicated to the

learning of body motions. More accurate evaluation systems studies kinematic, dynamic or geometric key features (*e.g.* speed, acceleration, distance covered by joints) of the motion (Yamaoka et al., 2013; Aristidou et al., 2015; Senecal et al., 2018). The learner must build her/his own motion respecting those features, in an “ad-hoc” system requiring scientific knowledge to interpret them.

Most of the existing evaluation systems are merged within the VLE and do not consider the motions of all interactive artefacts (body and manipulated objects) for the evaluation. This paper proposes a method, its implementation and an usability test of a simulation-independent and motion-based evaluation system. The proposed system integrates an existing VLE to offer to the teacher intuitive tools to set up an evaluation process by: (i) choosing the virtual artefacts (body parts or scene objects) whose motions will be analysed (ii), defining task steps thanks to a set of ordered 3D virtual checkpoints and (iii) recording a demonstration of the task to learn that will be imitated by the learner. An automatic shape-based and key-features-based comparison is then realized between the teacher and the learner motions. The next section reviews the VLEs implying manual gestures and the evaluation process in terms of motion analysis, usability and task dependency. Section 3 presents a motion-

based and generic evaluation model of the user activity relying on the task demonstration, as well as the technical challenges related to its implementation in any existing Unity simulation. Section 4 presents an experiment where the user must set up an evolution process for three different tasks to learn: shooting of a pool ball, throwing a dart and writing a letter. The results of the experiment are discussed in section 5. Conclusion and perspectives ends this paper.

2 RELATED WORK

This section makes a non exhaustive review of motion-based VLEs according to : (i) their capabilities to create and evaluate new tasks (ii), the type of analysed trace and (iii), the type of pedagogical feedback. Some representative examples are detailed to highlight the advantages and drawbacks of three main and non-exclusive categories of VLE. Note: the similarity studies of the motion shape or key features using supervised learning algorithms are not in the scope of this paper. Indeed, they require a database made of a great amount of demonstrations for each considered task.

Procedures-based VLEs: A VLE is efficient if the observation needs of the final user are taken into account (Buche et al., 2004; Baudouin et al., 2007). Trace analysis models can consequently be designed to: (a) formalize and operationalize these needs and (b), monitor the evolution of the virtual artefact in the VLE. For instance, Toussaint et al. (Toussaint et al., 2015) analysed the learner activity, in terms of action, during a simulated surgical operation. Heterogeneous traces made of some discrete positions and orientations of a trocar and, gaze direction on monitoring devices, were recorded to check if some ordered sequences of parallel actions were correctly performed. However, some generic properties must be added to the trace analysis model to minimise a heavy re-engineering process in case of task changing or evolution of the observation needs (Buche et al., 2004). Baudouin et al. (Baudouin et al., 2007) analyzed the user activity for physics learning by using the MASCARET VLE. The monitoring function regarding the user activities is uncorrelated to the learning domain and the pedagogical strategy, thanks to two meta-models for the respective description of the environment and the activities. Thereby, a model to reify the traces coming from every object, action and event, can be easily built. The observation needs can also be the entry point of the design of the pedagogical scenario. Mahdi et al. (Mahdi et al., 2019) developed a VR editor allowing any teacher to set up

her/his learning scenario from existing VLEs. A scenario was made of actions defined by a set of predefined Virtual Behavioral interaction Primitives (VBP) (*i.e.* observation, navigation, manipulation, communication) that must be performed by the learner. All the previous cited work are based on the learning of procedures as a set of actions ordered or not. An action mainly consists in reaching some discreet states of the manipulated objects. Despite the integration of the application domain vocabulary through some multi-layers semantic models for trace analysis, the teacher cannot, alone, implement its own evaluation scenario (except in the case of Mahdi et al., 2019) or integrate new observation needs. In addition, the underlying motions of the user leading to the object manipulations are not evaluated leading to a loss of information for professional gesture learning.

Observation and Imitation-based VLEs: The contributions of IT environments in motion learning have been effective for several decades. For example, Le Naour et al. (Le Naour et al., 2019) made a review on several studies based on the observation of the expert demonstration through videos or VR environments for sports motions. According to Scully and Newell (1985), “the observation of a model provides the learner with essential information, notably regarding unfamiliar coordination patterns” (cited by Le Naour et al., 2019). In addition, VR systems can provide an accurate observation of the gestures thanks to the control of an anthropomorphic 3D avatar, the depth perception and the 3D navigation to observe it from different view points. Consequently, numerous VLEs were developed to capture and play the motions of the teacher and learner in various application domains such as violin (Ng et al., 2007), ball sports (Miles et al., 2012), disk throwing (Yamaoka et al., 2013), Japanese archery (Yoshinaga and Soga, 2015), golf (Kora et al., 2015), rugby throwing (Le Naour et al., 2019), etc. If one puts aside the complex processing chain of a motion capture session, any teacher can potentially use an observation and imitation-based system to record a motion to learn. However, the concurrent visualisation of the learner and teacher motions for comparison is confronted to the challenge of spatial and temporal synchronisation of the two motions. Morphology-independent representation (Kulpa et al., 2005; Sie et al., 2014; Morel et al., 2016) and some attempt to mix, “average” or choose an adapted expert model (Yoshinaga and Soga, 2015; Morel et al., 2016) exists to the detriment of keeping the bio-mechanical properties of the motion. Scaling the learner skeleton according to the height of the teacher and asking the former to temporally follow the expert demonstration remain an acceptable solu-

tion (Le Naour et al., 2019). The use cases mainly concern motions implying the body (or a sub-set of its members and joints) and not the manipulated objects. The assumption of “learning by observation”, is still a strong one without a feedback on their key features, as their absence can be a serious impediment to the learning.

Shape-based Systems: As an extension of observation and imitation-based VLEs, shapes-based VLEs offer, in addition, an evaluation based on spatial and temporal similarities of the motion shape (*i.e.* the shape of the trajectory made by the body members or joints in the 3D space). The motion shape is frequently considered for assessing the learner performances and progressions in sport through the Dynamic Time Warping algorithm (DTW) (Morel et al., 2016; Le Naour et al., 2019). DTW aims to compare the shape of two time series without considering the temporal aspect (*i.e.* the signal duration or frequency) by outputting: (a) a distance correlation matrix evaluating the distance between each point of the two signals and (b), a score being the sum of the minimum distances to pass from one signal to another (the lower the score is, the more similar the signals are). For example, Morel et al. (Morel et al., 2016) built an “average” motion from several demonstrations made by experts for tennis serve. They captured the motion variation by, firstly, using the DTW to temporally align the motions on the longest one. A “DTW”-like algorithm computes the distances (and their derivatives) between the joint positions to extract an average one by the accumulation of all the minimal distances. Two ad hoc formula were then applied to compute a spatial tolerance (*i.e.* based on the mean of the distances between the joints of every expert motion) and temporal one (based on the similarities between two correlation paths of two joints to the “average” corresponding one) regarding a new learner motion. Despite their interesting results for fault detection, the consistency of the bio-mechanical properties of the “average” motion can be questioned. For ball throwing in rugby, Le Naour et al. (Le Naour et al., 2019) worked on the benefits of the visualisation of the expert demonstration superimposed on the learner throw. The descriptors to assess the performances were the ball distance to a predefined target, the spatial distance between the expert movement and learner movement thanks to the DTW sum of the minimal distances, and the regularity around the mean of the spatial distances, to “express” the regularity of the motion over time (Le Naour et al., 2019). Their results showed the benefits of the superimposition modality to learn the movement pattern but not for the throw accuracy. Nonetheless, the ball motion were not analysed and the performance

descriptors were not given to the learner. In addition, the different phases of the throw were explained by the expert but not formalized and operationalized in the system. The effective but perfectible learning providing by observation and imitation-based systems encourages their use. In addition, they mainly requires some demonstrations intuitively performed by the teacher to be operational. Nonetheless, one can ask the relevancy of using such DTW-based descriptors as a pedagogical information given to the learner.

Key Features-based Systems: For a given learning situation, there is usually not a unique and a perfect motion to learn. The motion bio-mechanical features (*e.g.*, bring two feet together then push off upwards for a pin-point serve in tennis, maintain the triangle on the backswing in golf) are known and the learner has to build his own gesture respecting those features. The VLEs built on this principle can compute some kinematic, dynamic and geometric descriptors from the time series made of joint positions and orientations (Larboulette and Gibet, 2015). The main objective is to: (i) compare the values of the descriptors with reference ones and (ii), give advice according to those values. Therefore, each observation need must be translated into one or several computable descriptors. For example, Yamaoka et al. (Yamaoka et al., 2013) built a VLE to learn disk throwing by giving feedbacks on: (a) the take back amplitude (b), the height of the right hand (c), the height transition of the right hand (d), the angle of the right elbow and (e), the twisting of the waist. The descriptors considered by Ng et al. (Ng et al., 2007) for violin learning were: bowing trajectories, bowing and joint angles, velocity, acceleration and distance traveled by arms joints. For teaching dance, Aristidou et al. (Aristidou et al., 2015) considered 27 morphological independent evaluation metrics (*e.g.* such as distance between limbs, displacement, gait size, acceleration, velocity, jerk, volume, body part height) to characterize motions in terms of actions, emotions and intentions. Baldominos et al. (Baldominos et al., 2015) worked on medical rehabilitation through a virtual goal keeper simulation, where they analysed the trunk (elbow, shoulder and arms must be straight and perpendicular to the ground) and the arm (must be extended) to stop the ball. The evaluation strategy requires, for the final users, to have well-established knowledge in mathematics, physics and bio-mechanics to correctly interpret the descriptor values. To tackle this issue, some trace analysis models can give an understandable meaning to the descriptors according to the application domain vocabulary (Toussaint et al., 2015; Baudouin et al., 2007). For instance, the well-known Laban Movement Analysis (LMA) (Laban and Ull-

mann, 1988) was used by (Aristidou et al., 2015) to describe the motion according to its four dimensions (*i.e.* body, shape, effort and space) each characterized by a subset of their descriptors. Sometimes, even if the observation needs are taken into account, the evaluation system can fail. For example, Senecal et al. (Senecal et al., 2018) tried to automatically characterize the skills of SALSA dancers according to three levels of expertise and ten specific descriptors (*e.g.* velocity peaking for the feet, the main frequency at which the people move their feet, the linear correlation of leg motions). In a nutshell, two-thirds of the descriptors allowed assessing the level of the dancers. Even though the usability of key features-based systems can be enhanced by the integration of the expert vocabulary, some low-level descriptors (Larboulette and Gibet, 2015) must, *in fine* be chosen for each evaluation concept. This task often makes the system task dependent without a guarantee on its efficiency, even if the expert was in the design loop (Senecal et al., 2018). In addition, the teacher is used to his/her vocabulary, to correct the motion and this vocabulary is not necessarily linked to some bio-mechanical descriptors.

The existing VLEs dedicated to professional manual gestures suffer, at least, of four main lacks. First, the motion is not systematically considered as a time series made of geometrical data for the activity analysis. The whole interactive virtual artefacts (*i.e.* user's body and virtual artefacts) are not all considered as some potential objects of interest. There is a lack of editing possibilities regarding, the choice of the object of interest, the task requiring an evaluation and the motion key features that must be interpreted. Finally, all the evaluation systems are merged with the VLE. To counterbalance these issues, several technical and scientific challenges appear. Given an existing VLE and its pedagogical scenario:

- How to build a generic and motion-based evaluation model, applicable to a wide range of task?
- How to give an interaction paradigm allowing the teacher to intuitively create its own evaluation process and associated feedbacks?
- How to operationalize such a model and interaction paradigm in any existing VLE ?

In the next sections, we try to tackle those challenges. The next section presents a motion-based evaluation model, its correlated interaction paradigm and its integration in any existing Unity 3D simulation.

3 ACTIVITY EVALUATION

3.1 Main Principles

In our previous work, we proposed a method to evaluate the learner activity thanks to the analysis of the motion generated by her/his activity (Djadja et al., 2019). This method relies on a shape-based approach as it allows any teacher to build an evaluation process by making a demonstration of the task to learn (*cf.* section 2, shape-based VLEs).

Suppose a toy problem made of a glass with a ball inside, a deposit area and a container (*fig.* 2). The user has to learn to take the glass, put the ball into the container, turn over the glass and put it on the deposit area. Three steps are necessary to set up the evaluation process within the VLE. The teacher has to choose an Object of Interest (OI) whose motions will be analysed (*fig.* 1(b)). In this case, the OI can be the glass or the user's hand. Then, the teacher has to create and place 3D virtual CheckPoints (CP) (*fig.* 1(c)). CPs are 3D geometrical shapes (*i.e.* 3D rectangle or sphere) that can be placed, rotated and sized in the 3D space. The OI must collide with CPs for a dual reason: (1) to decompose the task in several steps if necessary (*e.g.* step 1: put the ball into the container, step 2: put the glass on the deposit area) and (2), automatically extract the Motion of the OI (MOI) as a set of positions and orientations through time. Consequently, three kinds of CP can be generated: the Starting CP (SCP), one or several Intermediate CP (ICP, optional) and Ending CP (ECP). When an OI collides with an SCP the system starts saving the motion. If the next CP, that the OI collides with, is the SCP, then the previous data are removed and the system starts recording a new motion. Finally, the system stops recording when the OI collides with the ECP. For the task example, the glass was chosen as an OI, a rectangular SCP was sized according to the height and placed close to the glass, a rectangular ICP was sized according to the height and placed close to the container and a rectangular ECP was sized according to the width and length and placed above the deposit area (*fig.* 2). Finally, once the CPs are placed, the teacher makes a demonstration of the task, visualises it, and repeat the demonstration until satisfaction (*fig.* 1(d)). The motion of the glass of the final demonstration is automatically stored.

The learner has to: visualise the teacher demonstration and reproduce it (*fig.* 1(ii)). To evaluate the quality of the performed motion, the DTW algorithm is applied to compare the OI motion shape manipulated by the learner and the teacher. The DTW score is given to the learner and must be under a

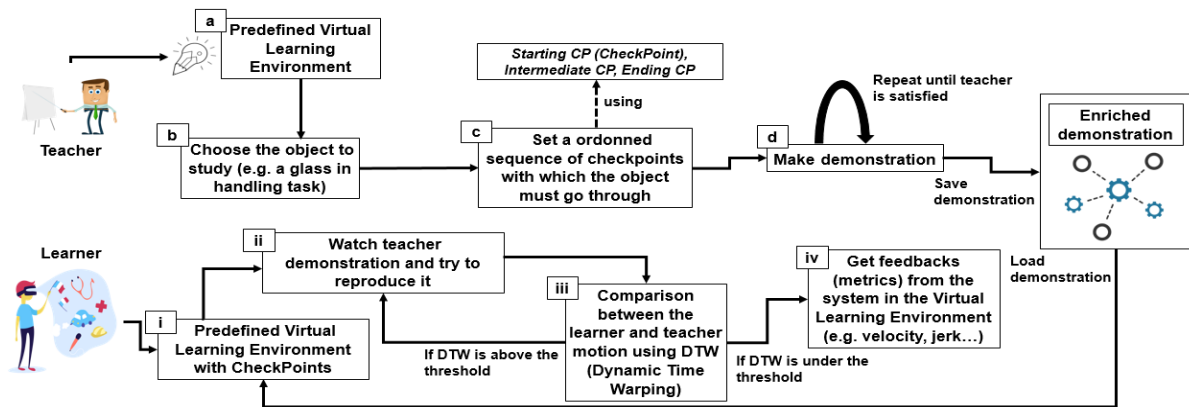


Figure 1: Motion-based evaluation method.

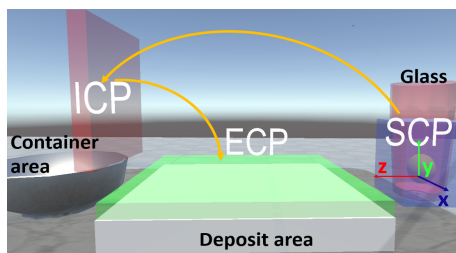


Figure 2: Manipulation task of glass.

threshold defined by the teacher (fig. 1(iii) and fig. 3(5)). If this is not the case, the learner has to reproduce the demonstration again (fig. 1(ii)). Once the threshold is reached, some motion key features are displayed to the learner such as kinematic descriptors (*e.g.* speed and jerk in this example (Larboulette and Gibet, 2015)) to have a more accurate analysis if needed (fig. 1(iv) and fig. 3(5)). In terms of feedback, the user can view within the VLE: the previous mentioned CPs (or not depending on the pedagogical strategy) and descriptors, the DTW score as well as the trajectory of the motion made by the expert and the learner (fig. 2 and 3(5)). Actually, if the trajectory is well known as a motion descriptor for navigation tasks (Terziman et al., 2011; Cirio et al., 2013), it can also be considered as a relevant understandable information for the evaluation of the current performance of the manual gesture. This last hypothesis will be studied in an experimental context.

3.2 Enhancements and HCI Design

Three toy problems (a manipulation task, a throw task, and a navigation task) were implemented, by using the Unity engine, to illustrate some examples of evaluation configurations (*i.e.* chosen OI, placement of CPs, trajectory display) (Djadja et al., 2019). Despite this method was designed to allow teachers

to easily build and assess learner activities in any VLE, the current evaluation system was merged with the VR environments. In addition, the system took into account only one ICP leading to the consideration of two-step tasks at most. Finally, no interaction paradigm was designed to create, size, rotate and place the ICP. An IT designer has to configure every CPs through the unity developer interface.

Therefore, our evaluation method was enhanced by considering several ICPs and, thus, tasks with more than two steps. Each ICP are numbered to make an ordered sequence of CPs with which the OI must go through (fig. 1, left, (c)). An interaction paradigm was also designed to: (a) choose the OI (b), create and configure the CPs and (c), perform, play and record the demonstration. The designed interface was made in the context of “scale 1” VR environments where the user is immersed with a VR helmet, two hand controllers and a motion capture system. VLEs can benefit of advanced interaction paradigms and especially those relying on gestures to offer more natural interface with affordance properties (Emma-Ogbangwo et al., 2014). Consequently, grasping a CP is realized by touching it with the hand virtual avatar and holding a controller button. The CP placement, translation and rotation are achieved by the placement, translation and rotation of the user hand holding the VR controller. The OI is chosen thanks to an usual process mixing a virtual ray for selection and a controller button for validation, leading to a color change of the selected object. Every virtual artefact with a rigidbody can be chosen as an OI as well as each body member of the user (*cf.* section 3.3 for more details).

The figure 3 (3, 4 and 5) presents the teacher virtual interface as a 2D menu regrouping the functionalities linked to: (1) the application control (*i.e.* OI choice mode, CPs creation mode, demonstration mode) and set of useful functionalities to make the demonstration (*i.e.* resetting the virtual scene, play/-

pause/slow down the current demonstration, etc.). This menu is attached in position and rotation to the user's right hand, while the combination of a virtual ray and a controller button of the left hand is used to interact with the menu buttons. It can be noted that the CP size can be set thanks to this 2D interface (fig. 3(4)) according to the 3D Cartesian axes, as the preliminary test for sizing the CP through hand gestures were not conclusive. Beforehand of the resizing process, the CP must be selected with a similar interaction procedure to the OI selection procedure. Furthermore, the choice of a body member is also made thanks to the menu listing all the available body members. Indeed, the set of virtual body members depends of the motion capture interface and must be known in advance.

With the presented method and interaction paradigm, we hope to give the users the means to intuitively select the OI, place and configure the CPs, make their demonstrations/try to reproduce the activity, and receive relevant feedbacks by a combination of a shape-based approach (*i.e.* observation and comparison of the demonstration and the motion trajectory) and a key features approach (*i.e.* DTW score and kinematic features display). Despite the choice of the kinematic features must be discussed, the overall HCI and evaluation principles must be first tested (*cf.* section 4). Furthermore, if one wants to implement the evaluation method as a module adaptable to any existing VLE, technical challenges rises regarding the architecture of the used VR engine. The next section focuses on those challenges and proposed an implementation in the context of the Unity engine.

3.3 Implementation Details

Our goal is to implement an motion-based activity evaluation module which can be adapted to any kind of VLE. Our evaluation module was made using Unity version 2019. Two main strategies can be considered to build this module. The module can be built as a Unity plugin, that could be added to any kind of Unity VLE. The plugin will contain the functions related to: the demonstration analysis, the user interactions and as well as the data storage of the evaluation configuration. The VLE could be then natively modified. However using a plugin requires to have minimum Unity skills that goes against the targeted final user of this study *i.e.* non IT teachers. Therefore, an architecture of the evaluation tool that imports an existing VLE is proposed in figure 4.

An existing VLE made with Unity is mainly composed of 3d objects (meshes) and user interaction scripts. Those 3d objects usually have a *collider* and

can have a *rigidbody* attached to them. The *collider* allows the 3d object to detect collision with other objects, while the *rigidbody* allows Unity to control the motion of this object using the physics engine. The architecture main concept relies on the export the VLE with its components in a file that can be imported and operationalized by the evaluation tool. Following this principle, the Unity developer must export its VLE (fig. 4, right side) to allow the teacher to import it with our evaluation tool (fig. 4, left side). The exportation *via* the unity "save project" function cannot be used as it needs the unity developer interface. In addition, the "build executable" function can also not be used as the executable file cannot be decomposed in the previous mentioned components during the importation.

The Unity engine proposes a method, called "assetbundle" to save an existing VLE in a single file. Actually, an assetbundle is "an archive file that contains platform-specific non-code Assets (such as Models, Textures, Prefabs, Audio clips, and even entire scenes) that Unity can load at runtime"¹. This method allows exporting a VLE and operationalized it by importation in our evaluation system. However, three important issues appears: (a) an assetbundle does not contain the user interaction scripts (unless they are pre-compiled), (b), the VLE, load at runtime does not save any modification applied in the 3D scene structure (*e.g.* change in the position of 3D objects, creation of new objects, etc.) and (c) conflicts can occurs if the VLE interactions used the same inputs as the evaluation tool interface.

For (a) and (c), the user interactions scripts were manually imported and the interactions conflicts manually resolved if existing. The section 5 proposes some perspectives to automatically process them. Regarding (b), one must known that all the created objects during a simulation at runtime in Unity are destroyed after the simulation stops. So the CPs creation and configuration will destroyed after each simulation for example. The OI choice, the correlated CPs and their configuration as well as the demonstration must be saved for a given VLE. Consequently, a SQL database is built in which the OI choice (through a unique id), each CPs configuration (shape, position and orientation) and the name of file containing the OI motion are saved. Indeed, a set of positions and orientations each separated by a fixed time-step, representing the OI motion is save in an unique file. Note that, each object, with a rigidbody, that collides the OI has its motion also replayed.

¹<https://docs.unity3d.com/Manual/AssetBundlesIntro.html>

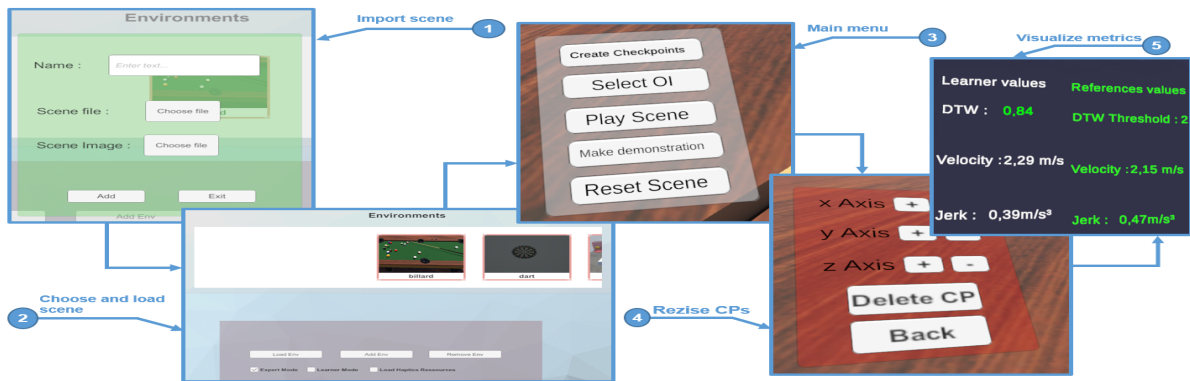


Figure 3: 2D interface menu for teacher.

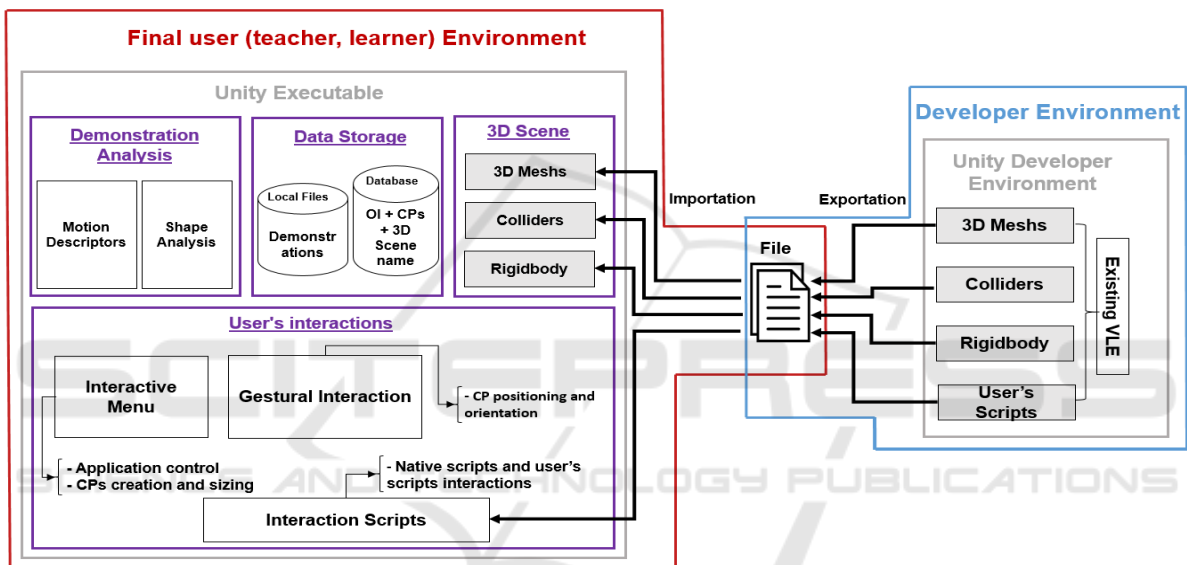


Figure 4: Architecture of the evaluation tool.

4 USABILITY STUDY

The evaluation method must be tested from the teacher and the learner perspectives given several specific professional learning situations. The objective is to have a positive effect on the learning process in terms of: (i) motor-skills acquisition in accordance with the given feedback and (ii), satisfaction degree of the author of the evaluation. Beforehand, a first study must be conducted to have a qualitative feedback, from a teacher point of view, on the evaluation method. This section studies the usability and relevancy of the implemented interactions and evaluation method with simple well-known tasks.

4.1 Protocol

First, each candidate received an oral explanation of all concepts used in section 3 as well as the principles of the HCI, the feedbacks and their role as a teacher. In terms of feedback, a short explanation was provided regarding the meaning of the DTW score (i.e. the lower the score is, the more similar the motions are) and the jerk descriptor (i.e. representing the smooth/tremor of the motion).

Three videos were shown to the candidates. These videos present a user working as a teacher on three tasks: the manipulation task described in section 3 (fig. 1 right), a navigation task and a last one showing a ball throw in a bin (Djadja et al., 2019). In each video, the teacher in the VLE chose an OI, placed the CPs, performed a demonstration of the task until

satisfaction². After watching the three videos, the participant were invited to put on the HTC Vive helmet, take the controllers and set up the evaluation process for the manipulation task. The participant is provided with all necessary assistance from technical view points (i.e. how to perform the interaction), must create and configure the CPs and make the demonstration according to what they watched on the corresponding video. They were not advised to consider the DTW score and the key features for building their evaluation as they were not asked to play the role of the learner for this first experiment.

Only 3D rectangular shapes and spherical ones were available for CPs. No qualitative or quantitative data are recorded at this stage as the goal is to become familiar with the system by practicing in the manipulation scene. This first step is necessary, as some people were not used to VR. In addition, the non or semi gestural based interactions (e.g. select mode or functions with the virtual menu) must be apprehended. Then, the participant was informed that he/she will play the role of a teacher on three different tasks from those observed for practise *i.e.* throwing dart on a target, pool shooting and letter writing. For each participant and each task, the recorded data are: the time from the beginning of the simulation to the creation of a satisfactory demonstration, the evaluation configuration (*i.e.* chosen OI, CPs placement, orientation and size) and the performed demonstration. The duration of this experiment was not timely limited and the user is free to stop when he/she wants. The participant was provided with all necessary assistance from technical view points, but was free to choose the OI, configure the CPs and make the demonstration according to their will. The main goals of each task are explained below.

Dart Throwing: the objective is to test our method in a context of “throwing task” learning situation. The dart is a sports activity that requires strength and accuracy. The purpose of the simulation is to throw a dart in the target center from a fixed distance (fig. 5, left). The candidate OI can be the virtual dart or the user’s hand.

Pool Shooting: the objective is to test our method in a “manipulation task” learning situation. The purpose of the simulation is to shoot the white ball with the cue, the white ball must collide with the orange ball placed near a hole (fig. 5, middle). The orange ball must fall into the hole. The position of the white and the orange ball are the same for every participant. The candidate OI can be the virtual white ball, orange ball, cue or one of the virtual user’s hand.

²<https://drive.google.com/drive/folders/1u0DzKAKkK-WsDiohsdGdeTlugQHqFH71?usp=sharing>

Letter Writing: the objective is to test our method to learn a “manipulation task” with small and more accurate object that the previous simulation. The purpose of the simulation is to write the “a” letter on paper put on a desk (fig. 5, right). Some horizontal lines are drawn on the paper. The used interface for writing is the Geomagic Touch haptic arm, its stylus controlling the virtual pen in the VLE. The HTC Vive hand controllers are still used for setting the evaluation configuration and controlling the application. The candidate OI can be the virtual stylus or the user’s hand.

After the experiment, the candidate fills an anonymous questionnaire³ regarding the perception of the user on (non-exhaustive list): the use difficulty, the time spent to achieve the tasks, the virtual menus and its interactions, the functionalities, the information given by the system, the appreciation of the performed demonstration in each task, etc.

4.2 Results & Analysis

18 people, aged between 20 and 30 years old, from different backgrounds (*i.e.* mainly students from IT, commercial, biology schools, with a VR experience or not) participated in this experiment. The training sessions were correctly performed by all participants. The dart throwing and the pool shooting were also completed by all participants, except for the demonstration step of the dart simulation (see below). However, none of them succeeded in building an evaluation process in the writing letter simulation (cf. section 5). Some evaluation configurations made by the participants are available by following this link².

4.2.1 Evaluation Configuration

Dart Throwing: 17 participants chose the dart as the Object of Interest (OI) and one the left hand. The way the CheckPoints (CPs) were placed varies from one participant to another. The majority of participants (15) give small shapes to the CPs. It seems that their goals was to cover only the surface through which the OI will pass. 3 participants chose large shapes. A possible explanation lies in the covering of the learner’s body in order to consider the morphology of all the learners. The Starting CP (SCP) was placed above the line where the participants should throw the dart (fig. 6). 5 participants used one Intermediate CP (ICPs) and 4 used two ICPs. The other does not used ICPs. They mainly used the cube shape (16) for the SCP. However, they mainly used sphere for the Ending CP

³<https://docs.google.com/forms/d/1KffjDL8vTFRiKFy0aYlxYOyn0yKQaVUOf4MTEonZdJM/edit?usp=sharing>



Figure 5: (left) Arrow and target in the dart simulation, (middle) White ball, orange ball, and cue in the pool simulation, (right) Virtual stylus, paper and desk in the writing simulation.

(ECP) (10). The ECP was placed just in front of the target. 7 participants were accurate with the ECPs and tried to cover exactly the surface of the target while 11 overflow it, apparently to provide feedbacks to those who missed the target.

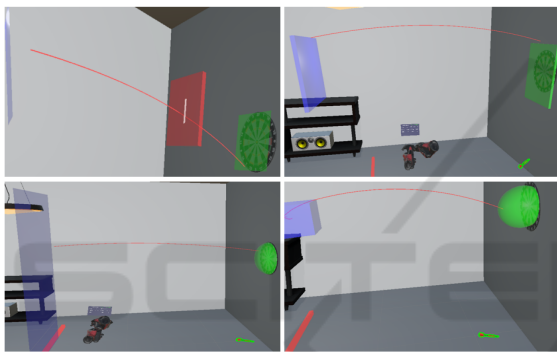


Figure 6: Dart throwing: examples of the CPs positioning and the OI trajectory.

Half of the participants failed to reach the dart target during the demonstration stage. Figure 6 shows some examples of CPs positioning and the trajectory generated by the dart.

Pool Shooting: 7 participants chose the orange ball as the OI and 9 chose the white ball. One of them chose the cue and another one chose the right hand as OI. Several approaches were adopted.

For the one who chose the cue as OI, it seems that he/she wanted to show to the learner where the head of the cue should start and end in order to have an effect when the white ball is smashed regardless the force applied. He/she then placed the SCP close to the border of the table and at the end of the cue when the learner set his/her stance. The ECP was placed just before and near the white ball, while the ICP was placed between the SCP and ECP. All the CPs had small shapes, probably to make the task more accurate.

For those who chose the white ball as OI, apparently they tried to give to the learner a very accurate path to follow. The SCP was placed just after

the white ball with a small shape. They also toggled between the cube and the sphere for the CPs shapes. 6 participants used ICPs. For those participants, the ICP was placed after the SCP and the ECP was near or merged with the orange ball. One of the participants gave a specific shape to the ICP *i.e.* a rectangular shape going from the SCP to the ECP. Another one merged the ICP with the orange ball, then placed the ECP after the orange ball but not in the hole (fig. 7, at the top left).

For those who chose the orange ball, the SCP was placed just after this ball. One ICP was used by only one participant. The ECP were placed inside the hole in which the orange ball was supposed to fall. They used alternatively the cube and the sphere for the shape of all the CPs and they mainly gave a small size to the CPs to apparently target the task accuracy.

All participants succeeded in realising a demonstration. Figure 7 shows some examples of CPs positioning and the trajectory generated by the chosen OI.

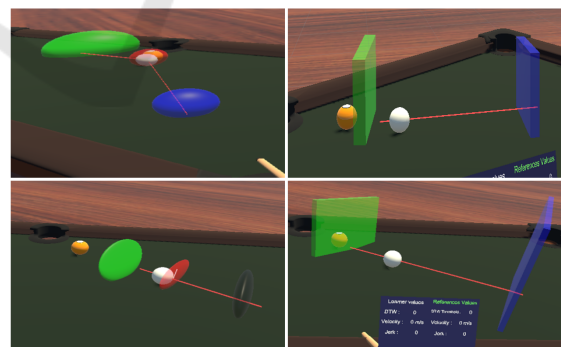


Figure 7: Pool shooting: examples of the CPs positioning and the OI trajectory.

4.2.2 Time to Perform a Satisfactory Configuration

Dart Throwing: The minimum time was 4,53 minutes and the maximum time was 16,56 minutes. The average time per user is 7,7 minutes while the standard deviation is 3,75.

Pool Shooting: The minimum time was 4,36 minutes and the maximum time 29,41. The average time per user is 8,22 minutes while the standard deviation is 5,74.

The time to perform these two evaluation configurations is reasonable, as only few minutes are required to set up a proper configuration for non professionals of the two considered application domains. However, these results are strongly dependant of the evaluation HCI presented in this study and the VLE interactions.

4.2.3 Questionnaire

This questionnaire³ is made of 7 questions based on a 1-5 Likert scale adapted from the method described by Lewis, 1995. Each answer for each question proposes an appreciation of the evaluation method or the HCI with textual qualifying adjectives. The answers are shown in figure 8.

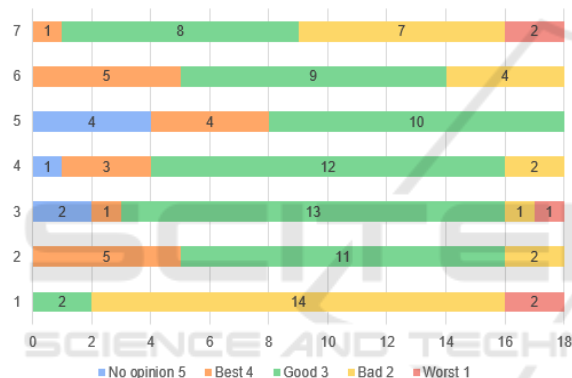


Figure 8: Questionnaire results - worst (1) to best (4) appreciations, (5) without opinion.

For the clarity of the figure, the best appreciation was encoded with 4, while the worst appreciation was encoded with 1. “No opinion” is encoded with 5. 14 participants thought that they carried out their demonstration slowly (question 1). The difficulty experienced for system control (question 2), the overall use of the system (question 3) and on the appreciation of the user interface (question 4 for the menu & 5 for the information presentation) were at least satisfactory by 14 users. Nevertheless, a minority (6) suggested some improvements, particularly in terms of the user interface (cf. the end of this section) in the commentary part of the questionnaire. 14 of the participants were at least satisfied with their demonstration on the pool simulation (question 6) and 9 participants on the dart throwing task (question 7).

5 DISCUSSIONS

The aim of this experiment was to evaluate the usability and the user experience with a motion-based evaluation tool of the learner activity in VE. The current study was made according to the perspective of the teacher building the evaluation process *i.e.* choosing an OI whose motions will be analysed, positioning an ordered sequence of 3D virtual CPs, demonstrating the task by manipulating the OI through the CPs. A menu and gesture-based HCI was designed to set up the evaluation process that was tested through three different examples: a dart, a pool and a writing VLE simulation. The participants were unable to complete the writing simulation as the issue winning unanimous support was the inability to accurately size, position and orientate very small CPs with the HTC Vive controller and the current interactions.

Regarding the evaluation method, we found that users chose a variety of OIs from one simulation to another. While most of the users chose the dart in the first simulation, all dynamic objects were chosen for the pool simulation with a majority dedicated to the balls. Therefore, the object relevancy varies from one participant to another, especially for simulations with more than 2 dynamic objects (only the hand and the dart can be a relevant choice in the first simulation). The CP configuration strategy also strongly varied even for the same OI. One participant seems to be interested in the task accuracy by accurately positioning CPs while others apparently gave a strong assistance by multiplying the ICPs. For the CP resizing, some users gave small shapes to the CPs while others made shapes that almost encompass the user’s body. With an acceptable amount of satisfied participants by their demonstration on the two first tasks (14/18 for pool and 9/18 for dart) performed in a few minutes, this tends to show the: (i) the usability of the HCI for building an motion-based evaluation process based on 3D CPs and demonstrations (ii), the effective need to give to the teacher the means to set up his/her own evaluation process, as their learning objectives seems to be very different for each participant. These results are first insights that clearly encourages future work and development regarding the presented method. Nevertheless, with regard to the slowness felt by the participants (question 1), an interview should have been conducted to understand the problem origin. The bad appreciation of the dart demonstrations (question 7) by half of the participants could be related to their performances (*i.e.* they fails to reach the target) and to the usability of the dart simulation.

One can not forget that the learner point of view was not considered as well as the expertise of profes-

sionals of the three considered tasks. For reminder, section 3.3 discussed two issues regarding (a) the non exportation of the VLE interaction scripts and (c) the conflicts between these interactions and the evaluation interface. For the issue (a), an alternative could be a file that will contains an assetbundle and the user interactions scripts, thanks to a custom exportation and importation module. Imported scripts files could respect a specific style in the file naming by considering at least the class name and the game object name. Indeed, a process must be designed to keep the links between the interaction scripts and the game objects for Unity. For the issue (c), a synthesis interaction module can be integrated to inventory the set of the used controls and search conflicts in this set. A possible solution can be to automatically propose a first menu to switch between the native interaction inputs and the evaluation interactions if there is at least one conflict.

An additional 4-Likert scale question was asked to the participant regarding the number of functionalities. 11 participants thought that there are enough functionalities, while 7 want to have more functionalities. The participants suggest several improvement ideas in the commentary part of the questionnaire, such as a guiding system for the learner by adding contextual information (*e.g.* texts, animations, pre-defined audio feedbacks) or a CP copy past system. Finally, the recording time of the demonstration can be extended to go beyond the collision of the ECP. Indeed, up to now, once the OI touches the ECP, the recording of the OI motion stops. Therefore, the demonstration scene can not be fully recorded and replayed. For instance, in the pool case, if the ECP is closed to the orange ball, the demonstration replay will never show the orange ball falling in the hole. To counterbalance this issue, two CP categories can be implemented: “scene CPs” will delimit the recording of all dynamic objects while, “OI CP” will keep the same role as the current CPs presented in this paper.

6 CONCLUSION

In this article, a motion-based evaluation method and its HCI were presented for the evaluation of the user activity in Virtual Learning Environments (VLE). The main principles of this method relies on the choice of a dynamic Object of Interest (OI) in the simulation whose motions will be: (a) observed by the learner following a learning based on the observation and imitation of the teacher demonstration (b), analysed in terms of spatial similarities thanks to the DTW algorithm and some kinematic key features available as

feedbacks in the VLE. To set up the evaluation process, a gestural and menu-based HCI was designed to allow the teacher to chose the OI among the available VLE dynamic objects. Then, the teacher has to create an ordered sequence of 3D virtual CheckPoints (CP), with which the OI must collide with. An architecture was designed to operationalize this method through a module that can be associated to any VLE. An implementation was studied in the context of the Unity engine. An experiment was carried out to assess the usability and the user experience of the evaluation module from the teacher perspective. Regarding the architecture implementation, the creation of an environment allowing to import an existing VLE in order to associate the evaluation module, seems to be the best way, for a non IT user, despite some issues in terms of the integration of the VLE user interactions. The experiment showed several interesting results justifying the relevancy of the proposed evaluation method. In a nutshell, the diversity of the OI choice and the CP configuration, combined with the satisfaction of the majority of the participant tends to show the relevancy of the proposed method for its adaptation capabilities to different learning objectives, modeled by the triplet “OI, CP, demonstration”. However, several limitations appear and among them, one can note the non adaptation of the HCI to accurately configure small CPs. In addition, some participants expressed the will to consider several OIs at once, as they felt that using only one OI will not fully fill their learning objectives. According to the experiment analysis and the participant feedbacks (*cf.* section 5), the future work will focus on an multi-OI evaluation system. A synthesis module allowing to better integrate the VLE interaction while resolving conflicts with the evaluation interactions will be developed. We also aims to design an automatic tool for the creation, placement and orientation of CPs once the OIs are chosen, to assist the teacher in the evaluation configuration. An experiment will be conducted with biology professors to validate our preliminary results in a real teaching situation.

REFERENCES

- Aristidou, A., Stavrakis, E., Charalambous, P., Chrysanthou, Y., and Himona, S. L. (2015). Folk dance evaluation using laban movement analysis. *J. Comput. Cult. Herit.*, 8(4):20:1–20:19.
- Baldominos, A., Saez, Y., and del Pozo, C. G. (2015). An approach to physical rehabilitation using state-of-the-art virtual reality and motion tracking technologies. *Procedia Computer Science*, 64:10–16.
- Baudouin, C., Beney, M., Chevaller, P., and LEPALLEC,

- A. (2007). Recueil de traces pour le suivi de l'activité d'apprenants en travaux pratiques dans un environnement de réalité virtuelle. *Revue STICEF*, 14.
- Bric, J. D., Lombard, D. C., Frelich, M. J., and Gould, J. C. (2016). Current state of virtual reality simulation in robotic surgery training: a review. *Surgical Endoscopy*, 30(6):2169–2178.
- Buche, C., Querrec, R., De Loor, P., and Chevaillier, P. (2004). Mascaret: Pedagogical multi-agents system for virtual environment for training. *Journal of Distance Education Technologies*, 2(4):41–61.
- Cirio, G., Marchal, M., Olivier, A.-H., and Pettre, J. (2013). Kinematic evaluation of virtual walking trajectories. *IEEE Transactions on Visualization and Computer Graphics*, 19:671–80.
- Djadja, D. J. D., Hamon, L., and George, S. (2019). Modeling and Evaluating of Human 3d+t Activities in Virtual Environment. In *The 14th European Conference on Technology Enhanced Learning*, pages 696–700, Delft, Netherlands. Springer.
- Emma-Ogbangwo, C., Cope, N., Behringer, R., and Fabri, M. (2014). Enhancing user immersion and virtual presence in interactive multiuser virtual environments through the development and integration of a gesture-centric natural user interface developed from existing virtual reality technologies. *HCI International*.
- Kora, T., Soga, M., and Taki, H. (2015). Golf learning environment enabling overlaid display of expert's model motion and learner's motion using kinect. In *19th Annual Conference of Knowledge-Based and Intelligent Information & Engineering Systems, KES-2015, Singapore, September 2015 Proceedings*.
- Kulpa, R., Multon, F., and Arnaldi, B. (2005). Morphology-independent representation of motions for interactive human-like animation. *Comput. Graph. Forum*, pages 343–352.
- Laban, R. V. and Ullmann, L. (1988). *Mastery of Movement*. Princeton Book Co Pub.
- Larboulette, C. and Gibet, S. (2015). A review of computable expressive descriptors of human motion. In *2nd International Workshop on Movement and Computing*, pages 21–28.
- Le Naour, T., Hamon, L., and Bresciani, J.-P. (2019). Superimposing 3d virtual self + expert modeling for motor learning: Application to the throw in american football. *Frontiers in ICT*, 6:16.
- Lewis, J. R. (1995). IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction*, 7(1):57–78.
- Mahdi, O., Oubahssi, L., Piau-Toffolon, C., and Iksal, S. (2019). Towards an editor for VR-oriented educational scenarios. In *The 14th European Conference on Technology Enhanced Learning*, Delft, Netherlands.
- Mikropoulos, T. A. and Natsis, A. (2011). Educational virtual environments: A ten-year review of empirical research (1999–2009). *Computers & Education*, pages 769 – 780.
- Miles, H. C., Pop, S. R., Watt, S. J., Lawrence, G. P., and John, N. W. (2012). A review of virtual environments for training in ball sports. *Computers & Graphics*, pages 714 – 726.
- Morel, M., Kulpa, R., Sorel, A., Achard, C., and Dubuisson, S. (2016). Automatic and Generic Evaluation of Spatial and Temporal Errors in Sport Motions. In *11th International Conference on Computer Vision Theory and Applications*, pages 542–551.
- Ng, K., Weyde, T., Larkin, O., Neubarth, K., Koerselman, T., and Ong, B. (2007). 3d augmented mirror: A multimodal interface for string instrument learning and teaching with gesture support. In *ICMI*, pages 339–345.
- Patle, D. S., Manca, D., Nazir, S., and Sharma, S. (2019). Operator training simulators in virtual reality environment for process operators: a review. *Virtual Reality*, pages 293–311.
- Roy, E., Bakr, M. M., and George, R. (2017). The need for virtual reality simulators in dental education: A review. *The Saudi Dental Journal*, 29(2):41 – 47.
- Senecal, S., Nijdam, N. A., and Thalmann, N. M. (2018). Motion analysis and classification of salsa dance using music-related motion features. In *Proceedings of the 11th Annual International Conference on Motion, Interaction, and Games*, pages 11:1–11:10, New York. ACM.
- Sie, M., Cheng, Y., and Chiang, C. (2014). Key motion spotting in continuous motion sequences using motion sensing devices. In *2014 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, pages 326–331.
- Terziman, L., Marchal, M., Multon, F., Arnaldi, B., and Lécuyer, A. (2011). Short paper: Comparing virtual trajectories made in slalom using walking-in-place and joystick techniques. In *Proceedings in JVRC11: Joint Virtual Reality Conference of EGVE-EuroVR*, pages 55–58.
- Toussaint, B.-M., Luengo, V., Jambon, F., and Tonetti, J. (2015). From heterogeneous multisource traces to perceptual-gestural sequences: the petra treatment approach. In Conati, C., Heffernan, N., Mitrovic, A., and Verdejo, M. F., editors, *Artificial Intelligence in Education*, pages 480–491, Cham.
- Yamaoka, K., Uehara, M., Shima, T., and Tamura, Y. (2013). Feedback of flying disc throw with kinect and its evaluation. *Procedia Computer Science*, 22:912 – 920.
- Yoshinaga, T. and Soga, M. (2015). Development of a motion learning support system arranging and showing several coaches' motion data. *Procedia Computer Science*, 60:1497 – 1505.