Human Movement Science 66 (2019) 564–577

Contents lists available at ScienceDirect

Human Movement Science

journal homepage: www.elsevier.com/locate/humov

3D feedback and observation for motor learning: Application to the roundoff movement in gymnastics☆

Thibaut Le Naour*, Chiara Ré, Jean-Pierre Bresciani

Department of Neuroscience and Movement Science, University of Fribourg, Switzerland

ARTICLE INFO

Keywords:
Virtual reality
Self + expert modeling
Motor learning
3D feedback
Dynamic time warping

2010 MSC:
00–01
99–00

ABSTRACT

In this paper, we assessed the efficacy of different types of visual information for improving the execution of the roundoff movement in gymnastics. Specifically, two types of 3D feedback were compared to a 3D visualization only displaying the movement of the expert (observation) as well as to a more 'traditional' video observation. The improvement in movement execution was measured using different methods, namely subjective evaluations performed by official judges, and more 'quantitative appraisals based on time series analyses. Video demonstration providing information about the expert and 3D feedback (i.e., using 3D representation of the movement in monoscopic vision) combining information about the movement of the expert and the movement of the learner were the two types of feedback giving rise to the best improvement of movement execution, as subjectively evaluated by judges. Much less conclusive results were obtained when assessing movement execution using quantification methods based on time series analysis. Correlation analyses showed that the subjective evaluation performed by the judges can hardly be predicted/ explained by the 'more objective' results of time series analyses.

1. Introduction

We all regularly learn new motor skills throughout our life. Oftentimes, learning those skills is essential, as in rehabilitation therapies (e.g., Timmermans, Seelen, Willmann, & Kingma, 2009), in "gesture-based" professions such as surgeon (e.g., Porte, Xeroulis, Reznick, & Dubrowski, 2007), or in sports training (e.g., Schmidt, Lee, Winstein, Wulf, & Zelaznik, 2018). Providing feedback during this learning process has long been shown to play a central role to help the learner understand what he[1] has to do, or not to do. Many different types of feedback can be provided. Specifically, feedback can be communicated through different perceptual channels (e.g., auditive, visual, haptics, mixed) and via different technologies (e.g., virtual reality, video, haptics). Also, feedback can focus on the outcome of the movement, or rather on its pattern/form. Here we investigated how feedback affects motor learning with an observation-based paradigm. Previous studies have demonstrated that this paradigm can successfully be used to improve the learning of new motor skills and the learner's performance (Ste-Marie et al., 2012). We focused more particularly on feedback from self-observation (Dowrick, 1999), i.e., when the learner sees the execution of his/her own movements, and on expert observation, i.e., when the learner sees the 'ideal' movements performed by an expert. Both visual feedback about one's own movement and information about the expert's movement can be displayed during (concurrent) and/or after (delayed) movement execution. When the information to display is edited, for example by adding information (Williams, NetLibrary, Williams, & Hodges,

☆ Fully documented templates are available in the elsarticle package on CTAN

* Corresponding author.
E-mail address: thibaut.lenaour@unifr.ch (T. Le Naour).

[1] 'he' represents both males and females.

https://doi.org/10.1016/j.humov.2019.06.008
Received 15 March 2019; Received in revised form 12 June 2019; Accepted 12 June 2019
Available online 26 June 2019
0167-9457/

2003) or selecting the best movement to present, visual information about one's own movement is called 'self-modeling' and visual information about the expert is called 'expert-modeling'. Self-modeling allows the learner to build a better representation of his/her own body and movements (Scully & Newell, 1985), whereas expert-modeling, which is based on the concept of imitation (Gould & Roberts, 1981), tends to improve the reproduction of the reference movement or the acquisition of unfamiliar coordination patterns (Scully & Newell, 1985). The advantages provided by self-observation or self-modeling (i.e., the learner sees his own movement but has no information regarding the 'correct' execution) correspond to the limitations of expert-modeling, and *vice versa* (Ste-Marie et al., 2012). In line with this, several studies have shown that providing visual information combining expert and self-modeling (ES-M) usually gives rise to the best learning performance, i.e., better performance than visual information based on self-observation (Robertson, Germain, & Ste-Marie, 2017), self-modeling (Arbabi & Sarabandi, 2016) or expert-only modeling (Oñate et al., 2005; Arbabi & Sarabandi, 2016) alone.

In most cases, combined self-observation or self- and expert modeling feedback was provided via video (Arbabi & Sarabandi, 2016; Robertson et al., 2017; Oñate et al., 2005; Boyer, Miltenberger, Batsche, & Fogel, 2009; Barzouka, Sotiropoulos, & Kioumourtzoglou, 2015; Anderson & Campbell, 2015). Specifically, the method usually consisted in presenting to the learner the two models on a split screen in which two videos were played simultaneously (Boyer et al., 2009; Barzouka et al., 2015), or by viewing successively the video of the expert and the learner's own movements (e.g.,. Oñate et al., 2005). More specifically,Anderson and Campbell (Anderson & Campbell, 2015) provided participants with concurrent self-observation feedback using video in addition to the real-time demonstration of an expert. They created a concurrent video mechanism which overlaid the two video images (i.e., the expert recorded model with the concurrent self-observation). This type of visual information has been used to improve the learning of several different types of movements, e.g., volley pass (Barzouka et al., 2015), rowing (Anderson & Campbell, 2015), badminton serve (Arbabi & Sarabandi, 2016) or gymnastic movements (Baudry, Leroy, & Chollet, 2006). Unfortunately, in the majority of studies mentioned above, visual information was combined with verbal feedback (given by a teacher or a coach (Oñate et al., 2005; Boyer et al., 2009; Baudry et al., 2006; Barzouka et al., 2015)), which constituted a potential confounding variable. This point is supported by the results of Arbabi and Sarabandi (2016) who found a significant effect by adding a verbal feedback. To our knowledge, only few studies proposed video-based learning without providing any verbal feedback (Robertson et al., 2017; Arbabi & Sarabandi, 2016; Anderson & Campbell, 2015). In addition, the quality of movement execution was assessed using different methods in the different studies. In some studies, performance was 'subjectively' evaluated by experts who gave a score, mostly assessing the technical form of the movement (Boyer et al., 2009; Barzouka et al., 2015; Arbabi & Sarabandi, 2016; Robertson et al., 2017). Other studies focused on kinematic measures collected on a given portion of the movement, and assessing very specific features of movement execution, such as the alignment of body segments (Baudry et al., 2006), joints orientation (Oñate et al., 2005), or the length and rate of the rowing stroke (Anderson & Campbell, 2015). Such 'focused' and portion-specific methods of evaluation are not representative of the global form of the movement. Yet another type of evaluation consisted in focusing on the outcome of the movement, without any evaluation of its technical execution (e.g.,(Barzouka et al., 2015)). The disparity between the different methods of assessment of movement execution and learning makes the results obtained in the different studies quite hard to compare. With the exception of the study of Anderson and Campbell (2015), who superimposed the expert recorded video with the live learner video, two other limitations of the studies based on video feedback are directly related to the display method. Specifically, the two videos are presented next to one another, or one after the other. Therefore, the learner cannot simultaneously watch the two videos and directly compare the self- and expert-modeling. Moreover, although the two videos usually start at the same time, the temporal synchronization between them is not ensured (Boyer et al., 2009).

An alternative method to provide visual feedback about movement execution consists in using 3D representations in Virtual Reality (VR). Although harder to implement, the use of 3D applications grants the possibility to edit feedback either in 2D or 3D. Specifically, whereas video consists in capturing and then replaying a series of images, 3D is based on the synthesis of the captured movement (i.e., by implementing a reconstruction) in a virtual environment. Synthesis allows the author to edit the movement in order to add or remove information. For instance, feedback can be simplified by removing potential distractors, as for example facial information or morphological information, thereby allowing the learner to focus only on the most important and relevant information. Several studies have demonstrated that simplifying information does not affect the quality of learning. For instance, Poplu, Ripoll, Mavromatis, and Baratgin (2013) showed that in decision task tests, abstract representations do not degrade the consistency of the players' responses, and even improved response times. Breslin, Hodges, Williams, Curran, and Kremer (2005); Hayes, Hodges, Scott, Horn, and Williams, 2007 also observed no difference between video demonstration and point light display. An additional advantage of 3D feedback is that it provides the learner with a wider field of view (i.e., up to 360-degrees) via user control over the camera.

Different types of feedback were proposed in VR. All of them were displayed as concurrent feedback (i.e., as opposed to the previous mentioned studies, the feedback is presented at the same time as the execution of the movement). Furthermore, most of them relied on motion capture technologies. Several authors used the concept of 3D superimposition, which consists in displaying the avatar of the learner and the avatar of the expert simultaneously with a minimal SPATIAL offset (i.e., the learner's avatar overlays the expert's one) (Chua et al., 2003-Janua; Kimura, Kuroda, Manabe, & Chihara, 2007; Hoang, Reinoso, Vetere, & Tanin, 2016). Some authors chose to display one or several virtual experts next to the virtual participant (e.g., side by side or four teachers around the learner's avatar) (Chua et al., 2003-Janua; Kelly, Healy, Moran, & Connor, 2010; Chan, Leung, Tang, & Komura, 2011). Other authors proposed to highlight the 'errors', i.e., the body segments which are too far from the segments of the expert, with a red color (Chan et al., 2011). In another context, Eaves and colleagues (Eaves, Breslin, & Spears, 2011) combined video and motion capture techniques by blending the video displaying the expert with point lights displaying specific joints of the captured learner. Finally, Hoang et al. (2016) displayed a first person view of the two avatars through a Head Mounted Display (HMD).

Contrary to video-based experiments, the results obtained with VR feedback are contrasted. When the learner's and the expert's avatar were side by side, most studies reported significant improvements of movement execution (Smeddinck, 2014; Hoang et al., 2016; Kimura et al., 2007; Eaves et al., 2011), although significant effects failed to be obtained in other studies (Chua et al., 2003-Janua; Kimura et al., 2007). Some authors showed that ES-M in VR is significantly better than video feedback (Smeddinck, 2014; Hoang et al., 2016) and Kimura et al. (2007) found that superimposition of self- and expert-modeling allows participants to imitate the expert postures faster than with expert modeling. When visual information about the expert was displayed next to the feedback about the learner, only one study on dance training found significant training-evoked improvements (Chan et al., 2011). Motor enhancement was also shown to be significantly higher with ES-M than E-M. Importantly, in all studies mentioned above, improvement of movement execution was assessed using kinematic comparisons between the movement of the learner and the movement of the expert. Specifically, different features of single postures or time series of postures (i.e., the movements) were considered. These features concerned the whole body (Eaves et al., 2011) or a subset of joints (Chua et al., 2003-Janua; Chan et al., 2011; Eaves et al., 2011; Hoang et al., 2016), and they were expressed by rotations (Chan et al., 2011; Eaves et al., 2011; Smeddinck, 2014) or positions (Chua et al., 2003-Janua; Chan et al., 2011; Hoang et al., 2016). For that, several authors used the Dynamic Time Warping (DTW) algorithm (Berndt & Clifford, 1994) to synchronize movements before comparing the target features (Chan et al., 2011; Morel, Achard, Kulpa, & Dubuisson, 2018).

As highlighted above, video-based and 3D-based studies did not rely on the same methods to assess the efficacy of learning and the improvement of movement execution. Studies using video information mostly relied on subjective evaluations performed by experts, movement outcome, and specific pattern features, whereas VR-based studies were mostly based on geometrical comparisons performed on the whole movement. A possible explanation could be that the authors who used VR feedback preferred to rely on motion capture technologies. Specifically, motion capture provides accurate values about the spatial and temporal aspects of the movement, which grants the possibility to synchronize different movements. Also, most studies conducted using video information addressed research questions related to human movement science, whereas the studies conducted using VR feedback were mainly published in computer science journals. As few studies compared objective and subjective assessments (e.g., Burns, 2013), we believe that it would be interesting to know how the subjective evaluation of a movement performed by a human being, such as those performed in artistic sports, compare to an evaluation based on time series comparison methods, which might 'objectively' quantify performance.

This work had two objectives. First, we wanted to assess the efficacy of 3D[2] feedback and 3D demonstration for motor learning when the learner has to reproduce the movement of an expert. Participants had to learn the roundoff movement in gymnastics. The learning session was divided into several successive phases of movement execution and demonstration. Four different groups of learners were provided with different types of visual information during the learning phases, namely video-based expert modeling, 3D-based expert modeling, and 3D-based expert- + self-modeling feedback. These three types of visual information were presented with the sagittal and frontal views, whereas a fourth group received a 3D-based expert- + self-modeling feedback, but with free control over the camera displaying the virtual scene. This free control over the camera was motivated by previous findings on self-regulation (Ste-Marie et al., 2012). The second objective of this article was to compare the results obtained with different methods of assessment of movement execution. In particular, we wanted to know how the subjective evaluation performed by experts compares to quantitative evaluation methods based on time series analysis.

2. Materials and methods

2.1. Participants and design

Thirty-two participants (23 men, 9 women) aged from 20 to 26 (average 22.3) volunteered to participate in the study. All participants had a first experience in gymnastic but none were gymnasts. None of the participants had impaired motor skills and all were able to perform the task. Participants were informed about the nature of the study and gave written consent prior to their inclusion in the study. This was done in accordance with the ethical standards specified by the 1964 Declaration of Helsinki. Each participant was assigned to one of four groups, i.e., there were 8 participants per group. The assignment was controlled taking into account each participant's experience and level of 'expertise' (i.e., around five minutes before the beginning of the experiment, the participants had to perform a first roundoff evaluated by the expert gymnastics experimenter), so that the average performance was similar between groups. Each group was provided with a different type of visual feedback during the learning phase of the roundoff.

2.2. Apparatus and procedure

2.2.1. Movement

The roundoff is an important basic movement in gymnastics. It is used to gain speed before performing a sequence of flips and/or a salto. The proper execution of the roundoff requires good coordination of the body's limbs and of the torso. Learning this movement therefore relies on the ability to build a mental representation of one's own body both spatially and temporally. The roundoff is divided in three phases: (1) the run-up phase, which consists in bringing the hands to the floor one at a time while the body inverts. (2) The main phase, which consists of a wheel with a 1/4 turn of the hands; and (3) the last phase: the two feet grouped together

[2] 3D means that information were displayed with 3D representation using a monoscopic vision.

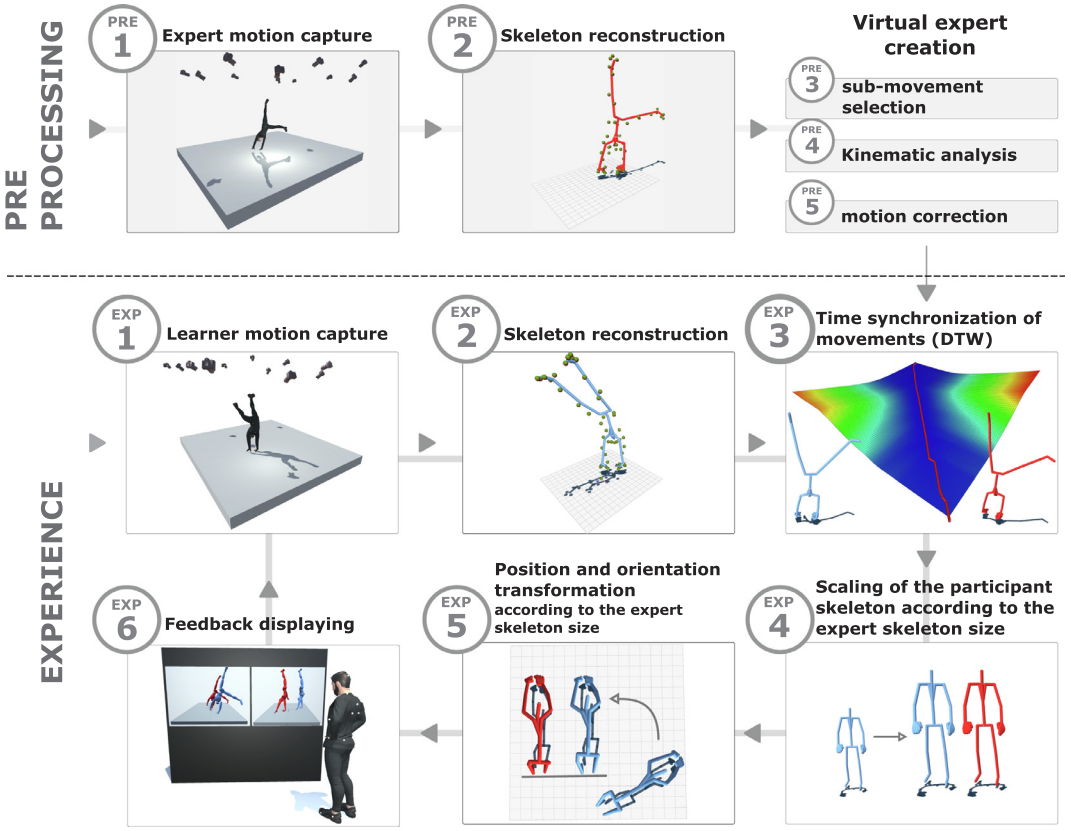

Fig. 1. Overview of our system for 3D visualization. The Step PRE 2 and EXP 2 were performed by the Motive software (Inc, 1996), whereas the other steps were computed via our custom-made application.

return to the floor at the same time, followed by a jump.

2.2.2. Capture

An Optitrack system (Inc, 1996) with 12 infrared cameras distributed in a 9 m × 6 m room was used to capture participants. The acquisition frequency was 120 frames per second. The cameras were specifically positioned and oriented to optimize capture in a ~38 m^3 volume centered on the participant (5 m × 2.5 m on the ground and 3 m high). Each participant wore a suit with 41 reflecting markers positioned on the whole body but the hands. Because a filtering system was necessary to resolve markers inversions and hands labeling problems (these problems occurred more particularly during hand contact with the ground), we decided to capture only the position and orientation of the wrists. Hands were always displayed as fully extended when providing visual feedback/ information during the learning phase. The reconstruction of the markers position and of the skeleton was performed using the Motive software (Inc, 1996).

2.2.3. Feedback projection

Visual information was projected on a 21-inch screen. Videos were displayed using a classic video player whereas 3D scenes were displayed using an application implemented by our research team (using C++/OpenGl with the SFML library (Gomila, 2007)). This application was dedicated to the feedback computation (Fig. 1) and the 3D display (120 frames per second).

2.2.4. Expert movement

For both video and 3D media, the reference expert's movement was that of an expert gymnast (16 years of experience, sports teacher and expert judge) whose performance was captured in our laboratory. The capture conditions were the same for the expert and for the learners. Two experts, one in gymnastics and the second one in motion capture, selected the reference movement among four repetitions of the movement performed by the expert. The movement was then cut from the beginning to the end of the roundoff (see Fig. 1). The reconstruction of the animation of the skeleton led to a fluid motion with no animation artifact.

2.2.5. Procedure

After a general warm up of 10 min, the participant put on the suit with 41 reflective markers. The suit was chosen among different available sizes (from s to xl) to fit the morphology of each participant. First, a live demonstration of the roundoff movement was

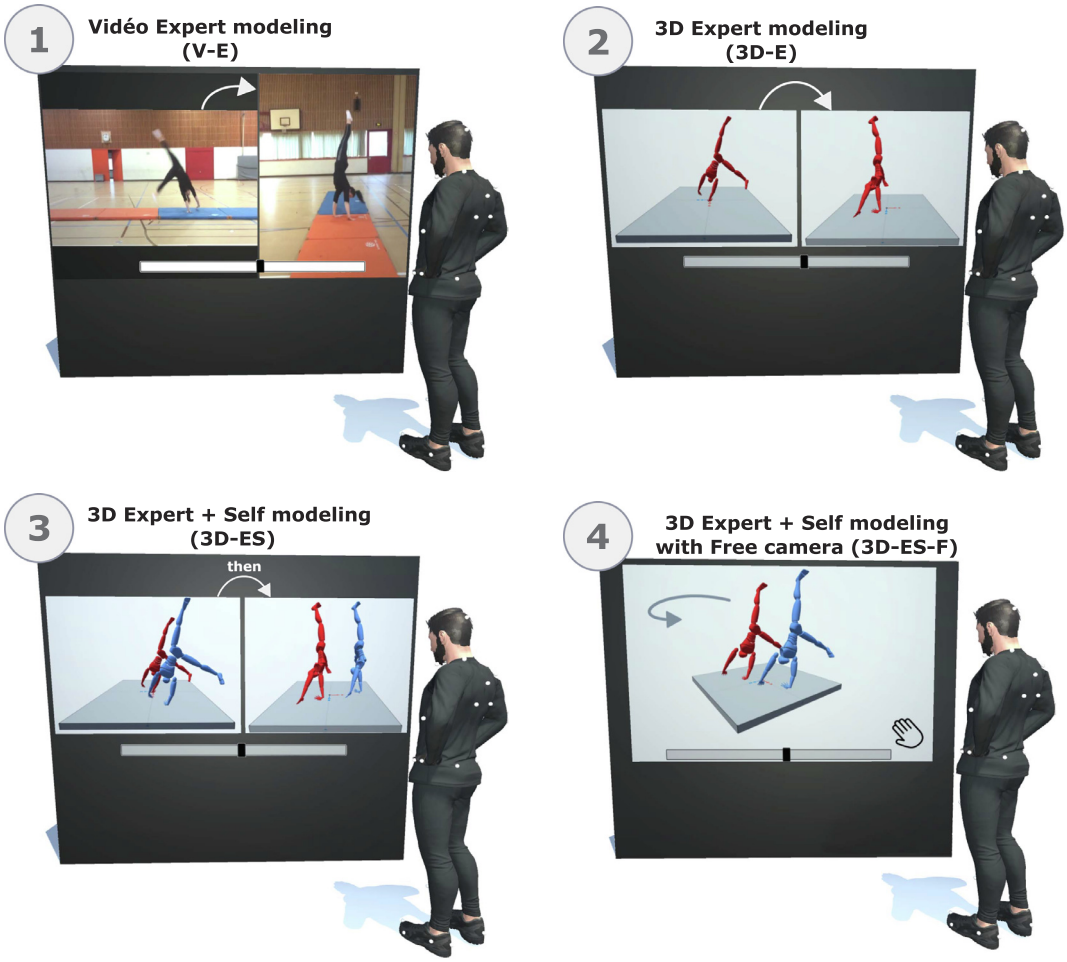

Fig. 2. Illustration of the four types of visual information provided to the learners.

performed by an expert. This demonstration was accompanied by a verbal description of the criteria of a good execution, namely: body straight with arms and legs fully extended during the handstand, feet pointed, legs together during landing. Participants then performed 2 trials to familiarize themselves with the movement, before being instructed about the experimental protocol. The experiment consisted of 3 phases:

- Pre-test: the participants were asked to perform 2 roundoff movements that were captured.
- Learning: the learning phase consisted in 9 sub-phases. For each sub-phase, the learner performed 2 roundoff movements before being provided with a specific observational learning about these two movements. The type of visual information provided during this observation phase varied between groups, as explained in the *observational learning* subsection.
- Post-test: the post-test phase was identical to the pre-test phase.

2.3. Observational learning

Four different types of visual information were provided to the four different groups, the number of observation phases and the total time of observation being the same in all groups:

- Expert-modeling using Video display (V-E): during the learning phase, the participants of this group watched two times a video of the expert performing the movement. Each video was shown with a sagittal and with a frontal view of the movement (Fig. 2.1). The sagittal view is commonly used by coaches and judges to observe the majority of errors performed by the participant. The frontal view is optimal to detect errors that occur when the hands or feet touch the ground, and for checking if the legs are correctly brought together.
- Expert-modeling using 3D display (3D-E): Here participants observed the movement of the virtual expert. As in the V-E condition, the expert's movement was shown two times, with two views each, namely a sagittal and a frontal view (Fig. 2.2).

- Expert + Self modeling using 3D display (3D-ES): This condition was similar to the 3D-E condition, but here the feedback showed the virtual learner him/herself as well as the expert (Fig. 2.3). Expert- and self-modeling feedback were side by side. As for the other observation phases, visual information was displayed two times with sagittal view and two times with frontal view.
- Expert + Self modeling using 3D display with free control (3D-ES-F): This feedback was identical to the 3D-ES feedback, but here participants were free to move the camera (translation, rotation and scale) to 'explore' the displayed feedback. It should be noted that for this intervention and the previous one, learners' movements were synchronised to the expert's one (this point is explained in the section *Dynamic Time Warping*).

For all conditions, participants were free to control the pace and the progression of the feedback (go forward, pause, go backward). However, observation time was limited to a total of two minutes, i.e., one minute by view for V-E, 3D-E, 3D-ES, and two minutes (free) for 3D-ES-F (these times were controlled by the experimenter).

2.3.1. Display

As illustrated in Fig. 2, the virtual characters representing the expert and the learner were displayed using simplified avatars (i.e., no mesh). This choice was motivated by previous works suggesting that using simpler feedback helps the learner to focus on the essential information (Poplu et al., 2013). The avatars of the expert and of the learner consisted of 21 joints, which were displayed in red and blue, respectively. Four additional steps were necessary for the 3D-ES and 3D-ES-F feedbacks. (1) As for the expert capture, an expert gymnast with good knowledge in motion capture segmented each roundoff movement (This step was performed manually to ensure that no error occurred). (2) A Dynamic Time Warping (DTW) procedure was applied to temporally map the movement of the learner with the movement of the expert (Fig. 1.3). This procedure is explained in details in the next Section 3 The avatar representing the learner was scaled to match the avatar of the expert (Fig. 1.4). (4) A geometrical transformation (translation + rotation) based on the pelvis of the avatars was automatically computed to 'align' (i.e., position and orientation) the avatar of the learner and the avatar of the expert (Fig. 1.5). Note that all of these steps took about one minute to be performed. Indeed, after the segmentation was done in the *Motive* software using a slider, the expert exported the movement to our application. The steps (2), (3) and (4), which were automatically computed, took less than 5 s in total.

2.4. Dynamic time warping

We used the Dynamic Time Warping algorithm (DTW) with two objectives: (1) synchronize the demonstration of the expert's movement with the feedback showing the movement of the learner, and (2) evaluate the distance between the movement of the learner and the movement of the expert. As introduced by Berndt and Clifford (1994), the DTW algorithm takes as input two time series and computes the best mapping between them. It also computes different measures described below.

2.4.1. Posture distance

This algorithm relies on a metric which expresses the distance between two time series. In our context, a movement is described by a series of postures (or skeleton) over time. A skeleton is represented by a set of hierarchically connected joints, with the hips being the root of the hierarchy. In our experiment, a skeleton included 21 joints representing the whole body (the hips, the two Spine joints, neck, head, left/right shoulder, right/left arm, right/left elbow, right/left wrist, right/left leg, right/left knees, right/left feet, right/left toes). Each joint j can be represented by a position $\boldsymbol{p}^j \in \mathbb{R}^3$ or by a rotation $\boldsymbol{q}^j \in \mathbb{SO}^3$. These two representations do not have the same meaning. Comparing two skeletons with positions is equivalent to comparing two point clouds. Specifically, a joint position is relative to the hips position and includes information about its position relative to its hierarchy. For example, the position of the left wrist includes the information of the left elbow, shoulder, spine and hips. On the other hand, rotations are expressed in the reference system of their parent joint. For instance, the left wrist rotation is only expressed relatively to the elbow joint. To summarize, positions express global information whereas rotations express local information. Here we used these two measures to evaluate the distance between the movement of the learner and the movement of the expert. The distance between two postures is finally given either by $d^p_{k,l} = \sum_j ||\boldsymbol{p}^j_k - \boldsymbol{p}^j_l||^2$ when using positions and by $d^q_{k,l} = \sum_j |||\log((\boldsymbol{q}^j_k)^{-1}\boldsymbol{q}^j_l)||^2$ when using rotations (represented by quaternion (Buss & Fillmore, 2001)). k and l correspond to the posture indexes of the expert's and learner's movement, respectively. $j \in [0, 21]$ is the index of the joint.

2.4.2. DTW algorithm

For each metric previously introduced, we can compute a matrix of distances $\boldsymbol{M}$ between the movement of the expert $\mathbf{M_{exp}}$ and the movement of the learner $\mathbf{M_{lear}}$. The dimension of $\boldsymbol{M}$ is $m \times n$ where m and n are the number of postures into $\mathbf{M_{exp}}$ and $\mathbf{M_{lear}}$, respectively. As mentioned before, each element of $\boldsymbol{M}$ corresponds to the distance $d^p_{k,l}$ or $d^q_{k,l}$ between two postures $k \in [0, m]$ and $l \in [0, n]$. In a second step, an optimal warping path p_{dtw} is computed between $\mathbf{M_{exp}}$ and $\mathbf{M_{lear}}$, i.e., $p_{dtw}$is the path with the minimal total cost among all possible warping paths of $\boldsymbol{M}$. $p_{dtw} = \{p^i\}$ is represented by a vector of pairs of indexes where $p^i = (k, l)$ and $i \in L$ with L being the number of pairs in p_{dtw}. If two consecutive pairs have their indexes incremented (i.e., $p^i = (k, l)$ and $p^{i+1} = (k + 1, l + 1)$), that means that the two movements are naturally temporally synchronized between p^i and p^{i+1}. Otherwise, one of the movement needs to "wait" for the other so that the two movements are synchronized (i.e., $p^i = (k, l)$ and $p^{i+1} = (k, l + 1)$ or $p^{i+1} = (k + 1, l)$). In our context, this path, which corresponds to the best mapping between $\mathbf{M_{exp}}$ and $\mathbf{M_{lear}}$, is the mapping that we used to bind temporally the learner's movement with the movement of the expert. From this mapping, it is also possible to extract other information, such as (1) the global distance

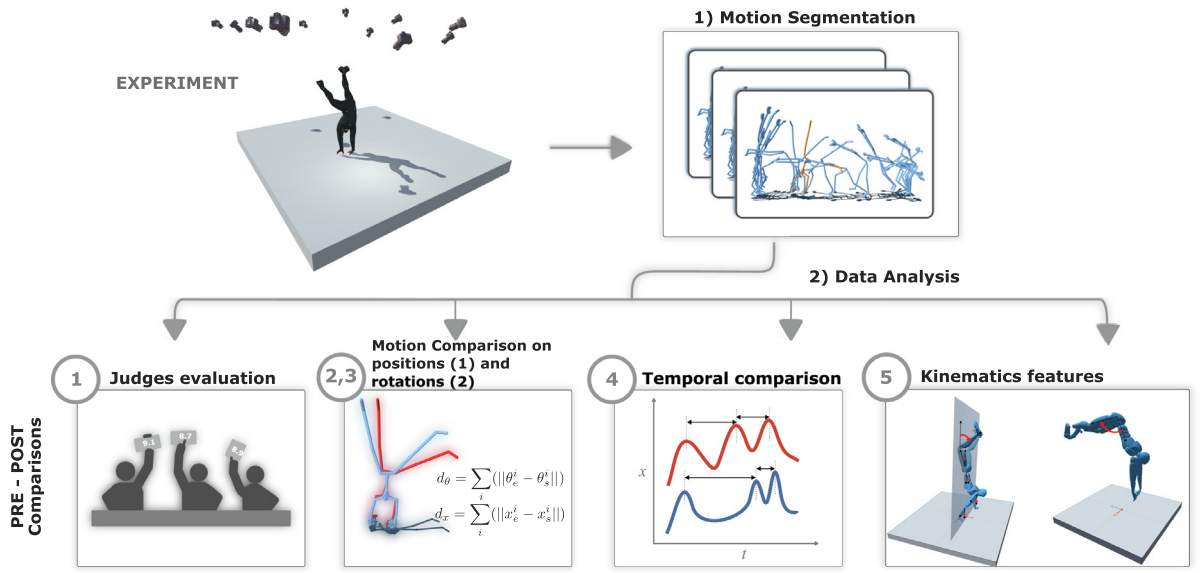

Fig. 3. Overview of the post processing stages.

$$d_{dtw} = \sum_{i \in L} d_i, \tag{1}$$

which expresses the combination of the spatial and temporal distances between the two movements. (2) The average distance, which normalizes d_{dtw} by the length of the path, describes better the spatial information:

$$d^s_{dtw} = \frac{d_{dtw}}{L}. \tag{2}$$

We can also compute the temporal difference between the two movements by summing up the number of occurrences for which the algorithm does not find a pair with 'corresponding' indexes, i.e., with indexes that are incremented simultaneously. This computation is explained by Algorithm 1.

2.5. Data analysis

2.5.1. Methods of evaluation

For each participant, we collected two roundoff movements in the pre and post phases. The movements were captured with the Optitrack motion capture system. All captured movements (4 * 32 = 128 sub-movements) were segmented by an expert gymnast with good knowledge of motion capture post-processing. Different analyses were then performed on the captured learners' movements. In particular, we quantified the spatial and temporal distance with the movement of the expert, we analyzed the kinematic features at given steps of the roundoff, and the movement was 'subjectively' evaluated by expert gymnastics judges that gave a score to each movement (see Fig. 3 for an illustration).

Algorithm 1 computation of error of time synchronization between the expert and learner movement based on DTW mapping. $l_{p_{dtw}}$ is the length of p_{dtw} and $p_{dtw}(i)$ is the distance computed for the i^{th} element of p_{dtw}

$d^t_{dtw} = 0$;
for all $key_{exp,lear}(i) \in p_{dtw}$ **do**
  if $(key_{exp}(i-1)$ equals $key_{exp}(i))$ || $(key_{lear}(i-1)$ equals $key_{lear}(i))$;
    d^t_{dtw} ++;
end

2.5.2. Spatial and temporal analysis

We used the DTW algorithm introduced above to extract the spatial and temporal distances. As for displaying the feedback in the ES condition, each movement of each learner was geometrically positioned and oriented to match the position and orientation of the hips of the expert. In addition, before the computation of the DTW, each skeleton of $\mathbf{M_{lear}}$ was scaled (but not retargeted) to match the size of the skeleton of the expert (as proposed by Chua and colleagues (Chua et al., 2003-Janua)).

2.5.3. Kinematic features

Inspired by the measures performed in previous works (Baudry et al., 2006; Oñate et al., 2005), we measured two specific aspects/criteria that are part of the evaluation guidelines for judges. Note that the experimenter focused on these evaluation criteria when giving the instructions at the beginning of the experience. The first criteria was the "vertical standing" of the participant at the handstand pose. This criterion corresponds to the instruction of holding straight with arms and legs fully extended. For each movement in the pre and post phases, we computed the sum of the angles between the sagittal plane of the participant and the orientation of specific body segments (see Fig. 3 for an illustration). The concerned segments were the legs, upper legs, torso, arms and forearms. The second measure computed the angle between the legs and the torso during the handspring phase. In this specific phase, the legs and torso of the learner are supposed to be at an 160–170 degrees angle (as what was observed for the movement of the expert). Here we took 165° as reference, which was the angle observed for the expert.

2.5.4. Evaluation by gymnastics judges

Two evaluation methods were used regarding the evaluation of the roundoff by judges. A first group of six official federal judges from the Swiss federation of gymnastics gave scores ranging from 1 to 10 for each roundoff. In total, each judge performed 128 evaluations, corresponding to the 32 participants × 4 movements per participant, i.e., two movements captured in the pre session and two movements captured in the post session. The order of presentation of the roundoff movements was fully randomized. A second group of five official federal judges evaluated the same four movements for each participant, as well as the reference movement of the expert. This time, the evaluation was performed using paired-comparisons. Specifically, for each evaluation trial, the judges were sequentially presented with two movements and asked to evaluate which of the two presented movements was the best (i.e., two-alternative forced choice). For each comparison, a score of 1 was attributed to the movement judged as being the best and a score of 0 was attributed to the other movement. Each judge was presented with ten comparison trials per participant, because there were five movements per participants (i.e., the four movements of the participant and the movement of the expert), and because each movement was compared to all other movements. In total, each judge performed a total of 320 comparison trials, corresponding to 10 comparisons × 32 participants. The order of presentation of the 320 trials was fully randomized, and so was the order of presentation of the movements for each paired comparison. After the comparisons, each movement had a score corresponding to the number of times it had been chosen as the best. For both methods of evaluation (i.e., scores and paired comparisons), the movements were presented to the judges using 3D animated media showing the movement in sagittal view, because this is the view currently used by judges in competition.

2.5.5. Data analysis and statistics

We intended to compare the effect of the video vs different types of 3D feedback for the acquisition of the roundoff. To this end, we used different measures inspired by previous studies in human movement sciences and data sciences. The measures that we used as dependent variables are summarized below and illustrated in Fig. 3. Specifically: (1) Two dependent variables related to the subjective evaluation performed by the judges, namely the scores and the results of the paired-comparisons. (2) Two dependent variables concerned the comparison between the joint positions of the participants and those of the expert. (3) Two dependent variables concerned the difference between the joint rotations of the participants and those of the expert. It should be noted that joint positions are relative to the pelvis position, whereas rotations provide a more "local" information, as they are expressed relatively to the parent joint. Note that for differences in position and rotation, one dependent variable was the global distance (i.e., global distances on positions (2.a) and rotations (3.a), see Eq. (1)), and the other dependent variable was the average distance between the movements of the participants and the movement of the expert (i.e., average distance on positions (2.b) and rotations (3.b), see Eq. (2)). (4) As distance-related measures mainly 'evaluated' the spatial characteristics of the movement, we also used a metric which quantifies the temporal difference between the movement of the participants and the movement of the expert(see Algorithm 1). (5) A last type of analysis focused on specific kinematic features that are expected if the roundoff movement is executed properly: (5.a) on the vertical standing of the body and (5.b) on the right angle between the legs and the torso at the handspring pose.

First, for each dependent variable, the average performance (average value of the 2 roundoff repetitions) measured in the pre-phase was compared between the four groups. This was done using either a one-way ANOVA (when data was parametric) or a Kruskal-Wallis test (when data was non-parametric). This test allowed us to make sure that the initial performance was similar between groups. Then, for each group and each method of evaluation, we compared the average performance in the pre-phase with the average performance in the post-phase. This was done using either a Student's t-test for repeated measures (when data was parametric) or a Wilcoxon signed-rank test (when data was non-parametric). In order to compare the efficacy of the four types of visual information, we computed the 'motor improvement' for each group and each participant. Motor improvement corresponded to the difference between the post and the pre performance. The average motor improvement was compared between the 4 groups using either a one-way ANOVA (when data was parametric) or a Kruskal-Wallis test (when data was non-parametric). For all tests, alpha was set at 0.05, and it was Bonferroni-corrected when relevant. For each test, normality was tested by the Shapiro-Wilks Normality test.

Finally, to try to identify which spatial or temporal characteristics of the movement could affect the subjective evaluation of the judges, we performed different correlations between the scores given by the judges and the other dependent variables except the kinematics measures because they did not characterize the whole movement. The correlations were performed using Pearson correlation coefficient (because data was parametric for all performed correlation tests). We also performed correlations between the results of the paired-comparisons (subjective evaluation of the judges) and the dependent variables related to pace and positions. Because the movements were always compared 'within' learner, the values entered into the correlation analysis were always

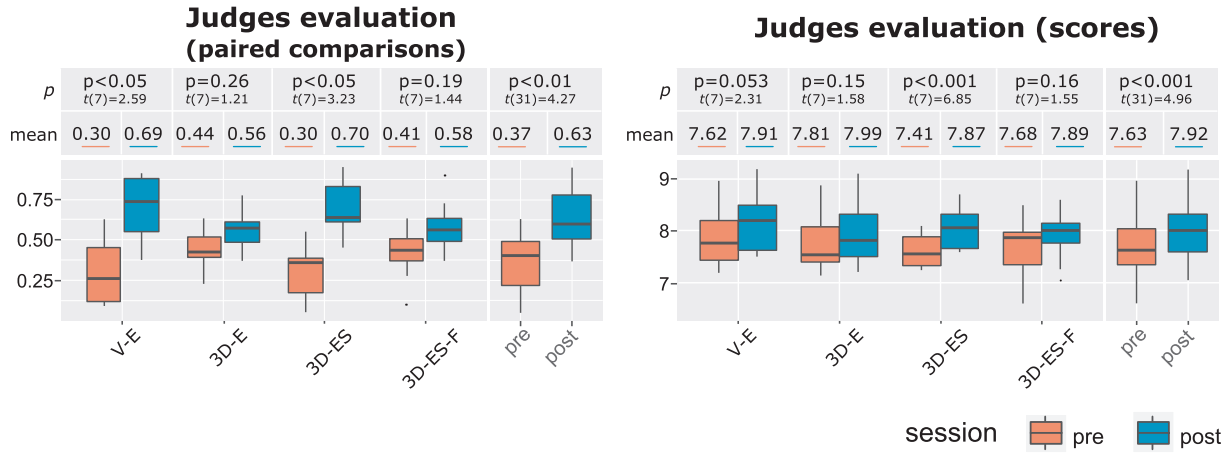

Fig. 4. Subjective evaluation performed by the official judges using paired-comparisons (left panel) and 'traditional' scores (right panel).

difference values, that is, for each learner, the difference between the value measured in the post session and the value measured in the pre session. Furthermore, before computing the correlations (still using Pearson correlation coefficient), we removed the influential points considered as outliers using the Cook's Distance (Cook, 1977) with a cutoff of $4/n$ (2 outliers with temporal data and 3 outliers with spatial data).

3. Results

When comparing initial performance (pre-phase) between groups, no significant difference was observed for any of the dependent variables. Concerning performance improvement as evaluated by the judges, Fig. 4 shows the box-plots and p-values relatives to the difference between the performance in the post and pre phases. One can see that the results obtained with the two methods of evaluation are similar. In particular, one can see that the two types of visual information giving rise to the largest improvement were the video demonstration of the expert (significant improvement for the comparisons and p-value barely failing to reach significance for the scores) and the 3D feedback displaying the participant and the expert (expert + self, $p < 0.05$ for both methods of evaluation). Concerning all other dependent variables measuring performance improvement, the results are summarized in Fig. 5. Note that for all these dependent variables, a value closer to zero means that the participant comes closer to the ideal/expert movement, i.e., it indicates a better performance.

Concerning the global distance on positions (2.a), none of the four types of visual information led to a significant improvement, though there was a tendency for the 3D demonstration displaying the movement of the expert. Note that a significant improvement was observed when taking all participants together (i.e., pooling all four groups together). The same general pattern was observed for the average distance (2.b). Remember that this variable expresses the average spatial distance over time between participants' postures and the expert's postures. Here we observed significant improvements with the 3D feedback displaying the expert, as well as when pooling all groups together. To get a concrete idea, the average joint-to-joint distance between the participants and the expert was around 13.9 cm in the pre phase (i.e., 292 cm divided by 21 joints) and 13.2 cm in the post phase (i.e., 277 cm/ 21 joints). Please note that we conducted the same analysis using the Derivative Dynamic Time Warping (i.e., DDTW) algorithm on positions (Keogh & Pazzani, 2001), and that the results were similar to those obtained with the DTW.

The pattern of results is slightly different with the comparisons based on rotations (3.a and 3.b). Here, there was no significant improvement whatsoever, and the only significant difference between the post and the pre phase actually indicated a worsening of the performance. This worsening was observed with the video demonstration when quantifying global distance (3.a). Note that performance also worsened for the 3D-ES-F, though this was not significant.

Regarding temporal differences in movement execution between participants and the expert (4), we observed a significant improvement for two groups, namely the group that was provided with video demonstration and the group that was provided with 3D feedback of self + expert with free exploration of the feedback. We also observed a significant improvement when pooling all participants together (i.e., when collapsing groups).

Finally, concerning kinematic features, the only significant improvement was observed for the angle between the torso and the legs during the handspring phase. This improvement was observed for the group that received video demonstration.

We then compared performance improvement (difference post - pre) between groups. In other words, we assessed whether some types of feedback gave rise to significantly larger improvements than other types of visual information. No significant difference between groups was observed for any of the dependent variables. The only dependent variable for which the main effect of the type of visual information barely failed to reach significance (Kruskall-Wallis $X^2_{(3)} = 7.22$, $p = 0.065$) was the global distance for rotations comparisons. It should be noted that to check if these results might have been affected by fatigue, we conducted the same tests after half of the training session had been completed, and the results were similar.

As illustrated in Fig. 6, no linear correlation was observed between the spatial and temporal characteristics of the movement and

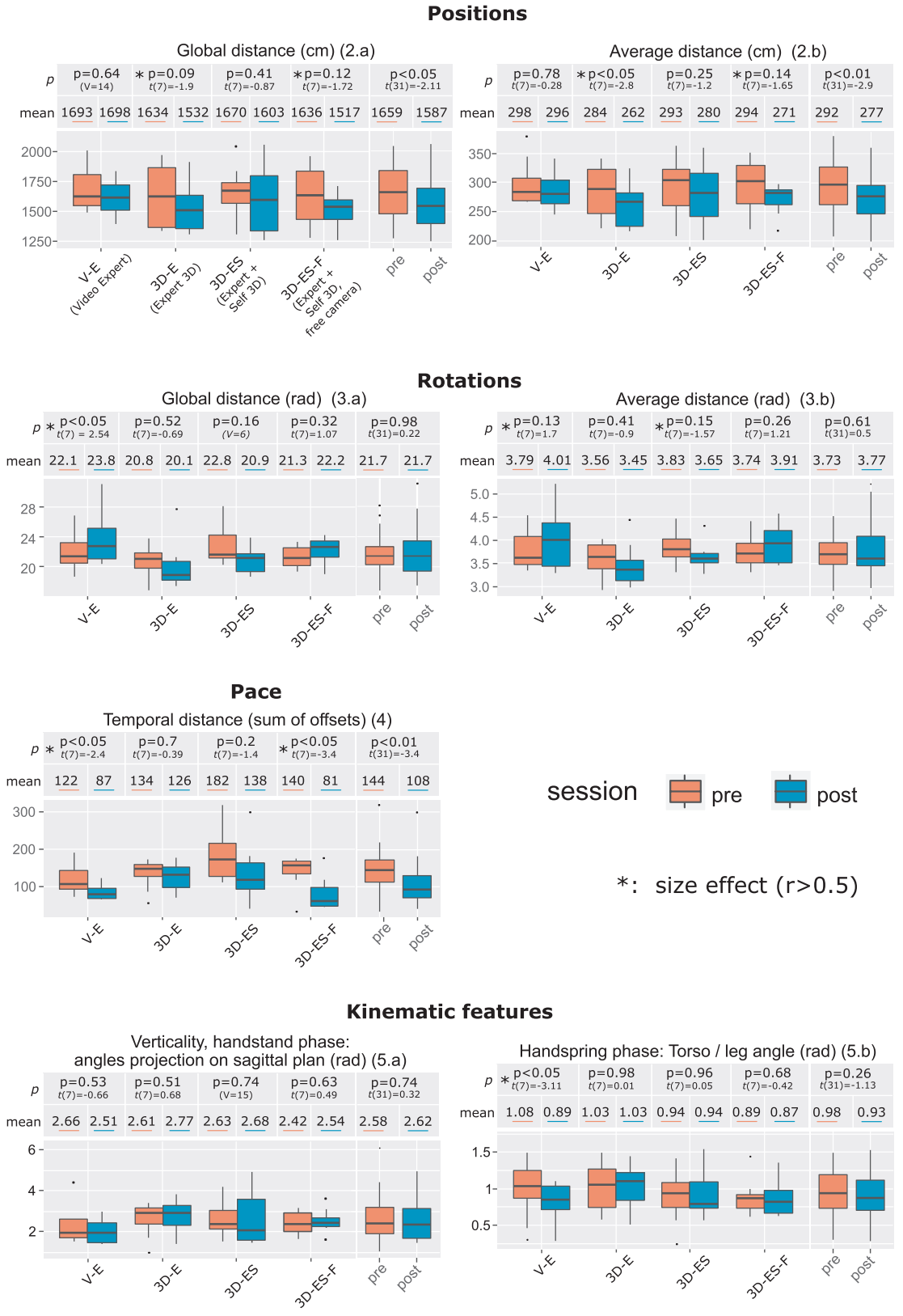

Fig. 5. Time series and kinematic features comparisons.

Linear correlations (Pearson test)

	(2.a)	(2.b)	(3.a)	(3.b)	(4)
(1)	0.15 p = 0.09	0.12 p = 0.17	-0.08 p = 0.37	0.13 p = 0.15	-0.15 p = 0.08
(4)	0.15 p = 0.08	0.02 p = 0.83	0.13 p = 0.15	-0.15 p = 0.08	
(3.b)	0.54 p <0.01	0.57 p < 0.01	0.79 p < 0.01		
(3.a)	0.55 p <0.01	0.47 p <0.01			
(2.b)	0.93 p <0.01				

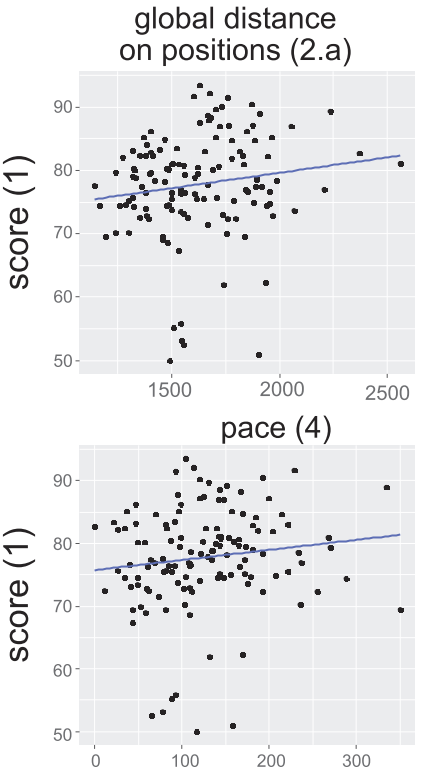

Fig. 6. Correlation between the different measures 'evaluating' the whole movement (i.e., the kinematics measures are not considered here). (1) Represents the subjective evaluation performed by the judges, (2.a) and (2.b) give the global and average distances between the joint positions of the participants and those of the expert. (3.a) and (3.b) represent the global and average distances between the joint rotations of the participants and those of the expert. (4) Corresponds to the temporal difference between the movement of the participants and the movement of the expert.

the score given by the judges. Multiple linear regressions also failed to provide any result (the highest observed R was $r^2(125) = 0.07$). Regarding the other correlations presented in the figure, as expected, strong positive linear relationships were found between positions measures ($r(126) = 0.93$), as well as between rotations measures ($r(126) = 0.79$). We also observed moderate positive relationships between the positions and rotations for average measures ($r(126) = 0.47$) as well as for global measures ($r(126) = 0.57$).

Regarding the correlations with paired-comparisons results, we observed a moderate positive linear correlation between the evaluation of the judges and temporal values ($r(28) = 0.58$, $p < 0.001$), whereas a moderate negative linear correlation was observed between the evaluation of the judges and spatial values ($r(27) = -0.42$, $p < 0.03$). These correlations are illustrated in Fig. 7.

4. Discussion

The main objective of this study was to compare learning outcomes of a round-off as a result of different types of visual information for motor learning. 3D expert-modeling, 3D expert- + self-modeling and 3D expert- + self-modeling with free control of the camera were provided to three different groups of participants. In addition, a fourth group was provided with a video demonstration showing the movement of the expert (i.e., expert-modeling). The learners performed ten sessions of 2 roundoff movements, and visual information was provided after each session but the last one. The effectiveness of each type of visual information on motor learning was assessed by two qualitative measures from official gymnastics judges, five measures based on the DTW algorithm, and two measures dedicated to specific kinematics features of the movement.

Regarding movement evaluation by official judges, two different methods were used. One method consisted for the judges to give a score to each evaluated movement, whereas the other method consisted in paired-comparisons. It is important to mention first that the two evaluation methods provided consistent results. In regard to the two qualitative measures, we observed a tendency towards an improvement of movement execution for all groups, and this improvement was significant when pooling all learners together. When considering the four types of visual information separately, our main finding is that providing 3D feedback about the learner and the expert with sagittal and frontal view significantly improved the execution of the roundoff movement. This improvement was observed with both evaluation methods. This finding is in line with previous studies using video feedback displaying self- + expert-modeling (Baudry et al., 2006; Boyer et al., 2009) (or self-observation + expert-modeling (Robertson et al., 2017; Barzouka et al., 2015)). However, to our knowledge, this is the first time that such improvement is observed with 3D feedback.

Providing learners with a video demonstration showing the movement of the expert also significantly improved the execution of the roundoff movement. This result confirms the results reported in previous studies using video demonstration and subjective

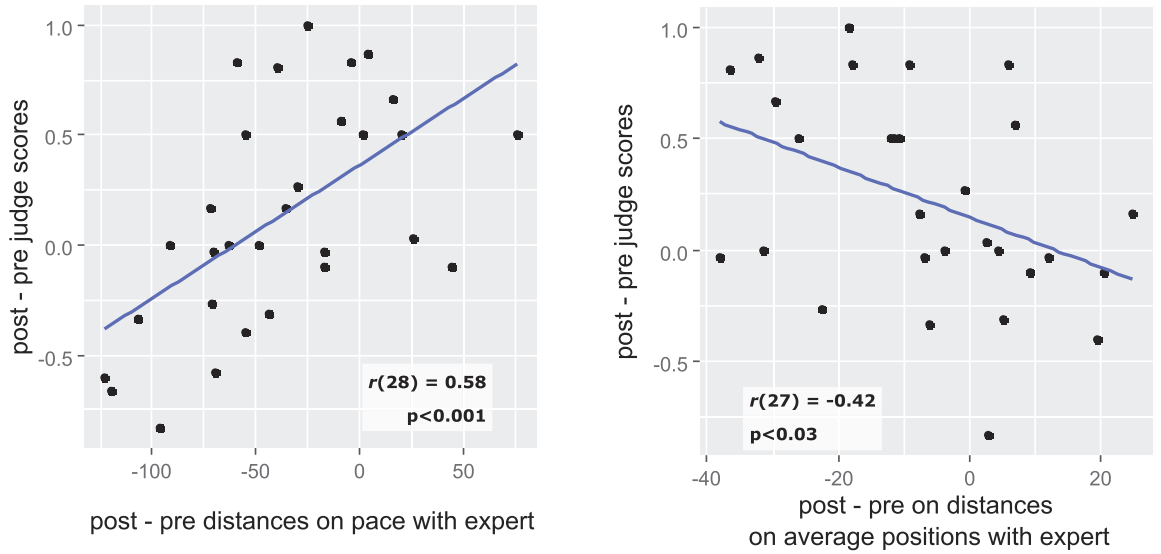

Fig. 7. Correlation between the subjective evaluation performed by the judges and a. pace quantification (left panel) and b. distance on average positions (right panel).

assessments (e.g., Zetou, Tzetzis, Vernadakis, & Kioumourtzoglou, 2002; Kalapoda, Michalopoulou, Aggelousis, & Taxildaris, 2003; Rodrigues, Ferracioli, & Denardi, 2010). Surprisingly, movement execution did not improve significantly (though there was a trend) for the learners who were provided with a 3D demonstration showing the movement of the expert only. This result is somehow surprising because some studies have shown that displaying biological motion information (e.g., using point light demonstrations) is comparable to video demonstration (e.g., Rodrigues et al., 2010). And as mentioned above, when learners were provided with a video demonstration showing the movement of the expert, movement execution improved significantly. A possible explanation is that the learning time was not sufficient to become familiar with the type of graphical representation underlying our 3D visual information. For the learners provided with 3D-ES feedback, this issue might have been compensated by the additional information about the learner's own movement. Specifically, providing both expert- and self-modeling allows the learner to directly compare his/her own movement with the movement of the expert, which has been shown to favor motor learning (Lejeune, Decker, & Sanchez, 1994; Famose, Hébrard, & Simonet, 1979; Fery & Morizot, 2000).

Regarding the group that was provided with both expert- and self-modeling feedback with free control over the camera, the lack of significant feedback-evoked improvement might result from some difficulty experienced by the learners to optimally use the camera. Specifically, in the other conditions, the viewpoint/viewing perspective on the movement was optimal for the whole viewing duration. This was not the case for the group that had free control over the camera, because learners in this group 'played around' with the viewing angle.

When directly comparing improvement between groups, to our surprise, none of the two feedback methods combining expert- and self-modeling gave rise to significantly larger improvement as compared to the other two types of visual information. There was actually no significant difference in improvement between the four groups. This finding is not consistent with previous studies assessing the effects of video-(Oñate et al., 2005; Boyer et al., 2009; Barzouka et al., 2015; Arbabi & Sarabandi, 2016; Robertson et al., 2017) and Virtual-Reality-based visual information (Chan et al., 2011; Hoang et al., 2016). The lack of significant difference between groups in our experiment could result from several factors. In particular, participants sometimes require several sessions (until five distributed over several days) before benefiting and taking full advantage of visual information (Shea, Lai, Black, & Park, 2000; Augusto et al., 2012). This can notably be exacerbated by the complexity of the movement to be learned. This was likely the case with the roundoff movement, because this movement requires good coordination, proprioception and balance. An additional factor that likely affected the 'slow' learning rate of participants in our study is the complexity of the provided visual information. This was particularly the case for the feedback methods combining expert- and self-modeling, because the displayed feedback then consisted of two animated humanoids with 21 joints each. In some cases, this visual feedback might have been difficult to use optimally.

Regarding the dependent variables related to body kinematics, pace, joints position and joints orientation, significant improvements were only observed for the measures quantifying the temporal features of the movement, as well as for the absolute positions. On the other hand, none of the four types of visual information contributed to improve 'local' spatial aspects of the movements (i.e., local rotations) or other specific features such as vertical standing during the handstand or legs orientation during the handspring phase. In addition, considering all groups and all measures, it is difficult to extract any consistent pattern. For instance, the significant improvement observed on average positions with 3D expert-modeling wasn't observed on any other dependent variable. Overall, the most interesting results were observed for the temporal aspects of the learned movement, i.e., the difference between the pace of the learners' movement and the pace of the expert's movement. Specifically, in addition to the global improvement, two groups significantly improved their pace, namely the group that was provided with a video demonstration showing the movement of the expert

and the group that received 3D feedback about the expert and oneself with full control over the viewing angle. This finding is surprising for three reasons. First, the instructions given to the learners did not mention the temporal aspects of the roundoff movement. Therefore, the observed improvement must be a 'byproduct' of other characteristics of the movement. In addition, for all four types of visual information, the learners had full control over the time of the displayed movements. Specifically, the learners could freely accelerate or decelerate the displayed visual information with the mouse. In this context, the temporal aspects of the movement are likely more difficult to grasp, although it is still possible to get an overview of the pace. Finally, in the 3D-ES-F condition (as in the 3D-ES condition), the movement of the learner was synchronized with the movement of the expert. This means that the learner did not have any information regarding the pace of his/her own movement.

Concerning the measures based on the Dynamic Time Warping algorithm, only few studies assessed the quality of the movement with this method, and all of them used virtual reality (Kelly et al., 2010; Chan et al., 2011; Burns, 2013; Morel et al., 2018). Only Chan and colleagues (Chan et al., 2011) used this method in the context of motor learning (dance movements) and applied it on positions and rotations. In contrast to what we observed, they found that providing feedback combining expert- and self-modeling significantly improves the global distances on rotations and positions. Our results are however consistent with those of Chua et al. (2003-Janua) who used a method very similar to the DTW and did not observe any improvement when providing combined self- and expert-modeling feedback in 3D.

In addition to comparing the efficacy of different types of visual information for motor learning, this work also aimed at comparing the subjective evaluation performed by judges and other types of quantitative measures related to kinematic and dynamic aspects of the movement. Incidentally, and even though there are of course exceptions, those two types of evaluation are traditionally used in different fields, namely in human movement science and in computer animation. From a global point of view, it is difficult to draw any conclusion based on our results (see boxplots presented in Figs. 4 and 5). When aggregating all groups, some measures gave rise to consistent tendencies, such as the results observed for 'global and average positions', 'pace', and subjective evaluations by the judges. However, when considering the groups separately, the effects observed with a given measure were seldom correlated to those observed with other measures. In particular, when assessing the relations between the subjective evaluations performed by the judges and the other quantitative measures, none of the tests showed any correlation. On the other hand, when examining the improvement values (i.e., post-pre differences), we observed moderate negative correlations between the evaluation performed by the judges and variables such as the pace and the spatial features of the movement. In other word, a slight correlation could be observed when focusing on the improvement.

A possible explanation for this lack of consistency between the two types of analysis could be that the measures based on the DTW were too approximate when considering the global values. This could be explained, for example, by morphological differences between the learners and the expert (Burns, 2013; Morel et al., 2018). However, the pattern was not different for the pace measure which did not take into consideration any spatial information.

To conclude, the global values provided by the different measures do not allow us to draw conclusions about the relations between the subjective evaluation performed by judges and quantitative measures used in time series analysis. Based on the results of this study, it seems inappropriate to use the DTW method to evaluate movements usually assessed by experts such as gymnastics or dance gestures. Our results also suggest that analyses based on improvement values might be more appropriate. Taking our results into account, the results of some studies in which no learning-evoked improvement has been observed using virtual reality-based feedback (Chua et al., 2003-Janua) could actually be interpreted in a different way and give rise to different conclusions. With regard to a possible substitution of the evaluation of gymnasts via a method based on artificial intelligence, we believe that the DTW method as set up here is not a viable option. However, other types of analysis such as those based on features selection (Nanopoulos, Alcock, & Manolopoulos, 2001; Tang, Alelyani, & Liu, 2014) would probably help to better 'explain' or predict the subjective evaluation performed by judges. In this context of automatic evaluation, it would also be interesting to try learning-based methods such as neural networks (Schmidhuber, 2015).

References

Timmermans, A. A., Seelen, H. A., Willmann, R. D., & Kingma, H. (2009). Technology-assisted training of arm-hand skills in stroke: concepts on reacquisition of motor control and therapist guidelines for rehabilitation technology design. *Journal of Neuroengineering and Rehabilitation, 6*(1), 1.

Porte, M. C., Xeroulis, G., Reznick, R. K., & Dubrowski, A. (2007). Verbal feedback from an expert is more effective than self-accessed feedback about motion efficiency in learning new surgical skills. *The American Journal of Surgery, 193*(1), 105–110. https://doi.org/10.1016/j.amjsurg.2006.03.016 URLhttp://www.sciencedirect.com/science/article/pii/S0002961006006398.

Schmidt, R. A., Lee, T., Winstein, C., Wulf, G., & Zelaznik, H. (2018). Motor Control and Learning, 6E. *Human Kinetics*.

Ste-Marie, D. M., Law, B., Rymal, A. M., Jenny, O., Hall, C., & McCullagh, P. (2012). Observation interventions for motor skill learning and performance: An applied model for the use of observation. *International Review of Sport and Exercise Psychology, 5*(2), 145–176. https://doi.org/10.1080/1750984X.2012.665076.

Dowrick, P. W. (1999). A review of self modeling and related interventions. *Applied and Preventive Psychology, 8*(1), 23–39. https://doi.org/10.1016/S0962-1849(99)80009-2.

Williams, A. M. (A. Mark), NetLibrary, I., Williams, M. E., & Hodges, N. E. (2003). *Skill acquisition in sport: Research, theory and practice.* London; New York: Routledge.

Scully, M., & Newell, D. K. M. (1985). Observational learning and the acquisition of motor skills: Toward a visual perception perspective. *Journal of Human Movement Studies, 11*, 169–186.

Gould, D. R., & Roberts, G. C. (1981). Modeling and motor skill acquisition. *Quest, 33*(2), 214–230. https://doi.org/10.1080/00336297.1981.10483755.

Robertson, R., Germain, L. St., & Ste-Marie, D. M. (2017). The effects of self-observation when combined with a skilled model on the learning of gymnastics skills. *Journal of Motor Learning and Development,* 1–30.

Arbabi, A., & Sarabandi, M. (2016). Effect of performance feedback with three different video modeling methods on acquisition and retention of badminton long service. *Sport Science, 9*, 41–45.

Oñate, J. A., Guskiewicz, K. M., Marshall, S. W., Giuliani, C., Yu, B., & Garrett, W. E. (2005). Instruction of jump-landing technique using videotape feedback: Altering lower extremity motion patterns. *American Journal of Sports Medicine, 33*(6), 831–842. https://doi.org/10.1177/0363546504271499.

Boyer, E., Miltenberger, R. G., Batsche, C., & Fogel, V. (2009). Video modeling by experts with video feedback to enhance gymnastics skills. *Journal of Applied Behavior Analysis, 42*(4), 855–860. https://doi.org/10.1901/jaba.2009.42-855.

Barzouka, K., Sotiropoulos, K., & Kioumourtzoglou, E. (2015). The effect of feedback through an expert model observation on performance and learning the pass skill in volleyball and motivation. *Journal of Physical Education and Sport, 15*(3), 407–416. https://doi.org/10.7752/jpes.2015.03061.

Anderson, R., & Campbell, M. J. (2015). Accelerating skill acquisition in rowing using self-based observational learning and expert modelling during performance. *International Journal of Sports Science and Coaching, 10*(2–3), 425–437. https://doi.org/10.1260/1747-9541.10.2-3.425.

Baudry, L., Leroy, D., & Chollet, D. (2006). The effect of combined self- and expert-modelling on the performance of the double leg circle on the pommel horse. *Journal of Sports Sciences, 24*(10), 1055–1063. https://doi.org/10.1080/02640410500432243.

Poplu, G., Ripoll, H., Mavromatis, S., & Baratgin, J. (2013). How do expert soccer players encode visual information to how do expert soccer players encode visual information. *Research Quarterly for Exercise and Sport, 1367*(November), 37–41.

Breslin, G., Hodges, N. J., Williams, A. M., Curran, W., & Kremer, J. (2005). Modelling relative motion to facilitate intra-limb coordination. *Human Movement Science, 24*(3), 446–463. https://doi.org/10.1016/j.humov.2005.06.009.

Hayes, S. J., Hodges, N. J., Scott, M. A., Horn, R. R., & Williams, A. M. (2007). The efficacy of demonstrations in teaching children an unfamiliar movement skill: The effects of object-orientated actions and point-light demonstrations. *Journal of Sports Sciences, 25*(5), 559–575. https://doi.org/10.1080/02640410600947074 pMID: 1736554, arXiv: https://doi.org/10.1080/02640410600947074.

Chua, P. T., Crivella, R., Daly, B., Hu, N., Schaaf, R., Ventura, D., Camill, T., Hodgins, J., & Pausch, R. (2003). Training for physical tasks in virtual environments: Tai Chi. *Proceedings – IEEE Virtual Reality, 2003*, 87–94. https://doi.org/10.1109/VR.2003.1191125.

Kimura, A., Kuroda, T., Manabe, Y., & Chihara, K. (2007). A study of display of visualization of motion instruction supporting. *Educational Technology Research, 30*(1–2), 45–51. https://doi.org/10.15077/etr.KJ00004963315.

Hoang, T. N., Reinoso, M., Vetere, F., Tanin, E. (2016). Onebody: remote posture guidance system using first person view in virtual environment. In: Proceedings of the 9th Nordic Conference on Human-Computer Interaction 25:1–25:10.https://doi.org/10.1145/2971485.2971521.

Kelly, P., Healy, A., Moran, K., & Connor, N. E. O. (2010). A virtual coaching environment for improving golf swing technique. *Science,* 51–56. https://doi.org/10.1145/1878083.1878098.

Chan, J. C. P., Leung, H., Tang, J. K. T., & Komura, T. (2011). A virtual reality dance training system using motion capture technology. *IEEE Transactions on Learning Technologies, 4*(2), 187–195. https://doi.org/10.1109/TLT.2010.27.

Eaves, D. L., Breslin, G., & Spears, I. R. (2011). The short-term effects of real- time virtual reality feedback on motor learning in dance. *Presence: Teleoperators and Virtual Environments, 20*(1), 62–77. https://doi.org/10.1162/pres_a_00035.

Smeddinck, J. D. (2014). Comparing modalities for kinesiatric exercise instruction. *CHI'14 Extended Abstracts on Human Factors Computing Systems,* 2377–2382.

Berndt, D. J., Clifford, J. (1994). Using dynamic time warping to find patterns in time series. In: KDD workshop, Seattle, WA, pp. 359–370.

Morel, M., Achard, C., Kulpa, R., & Dubuisson, S. (2017). Automatic evaluation of sports motion: A generic computation of spatial and temporal errors. *Image and Vision Computing, 64*(2018), 67–78. https://doi.org/10.1016/j.imavis.2017.05.008.

Burns, A.-M. (2013). *On the Relevance of Using Virtual Humans for Motor Skills Teaching: a case study on Karate gestures.* Université Rennes 2 (Jan. 2013) URLhttps://tel.archives-ouvertes.fr/tel-00813337.

Inc, N. (1996). Optitrack and motive from naturalpoint (1996). URLhttps://optitrack.com/.

Gomila, L. (2007). Simple and fast multimedia library. URLhttps://www.sfml-dev.org/.

Buss, S. R., & Fillmore, J. P. (2001). Spherical averages and applications to spherical splines and interpolation. *ACM Transactions on Graphics, 20*(2), 95–126. https://doi.org/10.1145/502122.502124.

Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics, 19*(1), 15–18.

Keogh, E. J., Pazzani, M. J. (2001). Derivative Dynamic Time Warping. pp. 1–11.https://doi.org/10.1137/1.9781611972719.1. arXiv:https://epubs.siam.org/doi/pdf/10.1137/1.9781611972719.1, URLhttps://epubs.siam.org/doi/abs/10.1137/1.9781611972719.1.

Zetou, E., Tzetzis, G., Vernadakis, N., & Kioumourtzoglou, E. (2002). Modeling in learning two volleyball skills. *Perceptual and Motor Skills, 94*, 1131–1142. https://doi.org/10.2466/pms.2002.94.3c.1131 (3_suppl).

Kalapoda, E., Michalopoulou, M., Aggelousis, N., & Taxildaris, K. (2003). Discovery learning and modelling when learning skills in tennis. *Journal of Human Movement Studies, 45*(5), 433–448.

Rodrigues, S. T., Ferracioli, M.de C., & Denardi, R. A. (2010). Learning a complex motor skill from video and point-light demonstrations. *Perceptual and Motor Skills, 111*(2), 307–323. https://doi.org/10.2466/05.11.23.24.25.PMS.111.5.307-323 URLhttp://journals.sagepub.com/doi/10.2466/05.11.23.24.25.PMS.111.5.307-323.

Lejeune, M., Decker, C., & Sanchez, X. (1994). Mental rehearsal in table tennis performance. *Perceptual and Motor Skills, 79*(1), 627–641.

Famose, J. -P., Hébrard, A., Simonet, P. (1979). Contribution de l' aménagement matériel du milieu à la pédagogie des gestes sportifs individuels, STAPS.

Fery, Y.-A., & Morizot, P. (2000). Kinesthetic and visual image in modeling closed motor skills: The example of the tennis serve. *Perceptual and Motor Skills, 90*(3), 707–722. https://doi.org/10.2466/pms.2000.90.3.707 pMID: 1088374, arXiv:https://doi.org/10.2466/pms.2000.90.3.707.

Shea, C. H., Lai, Q., Black, C., & Park, J. H. (2000). Spacing practice sessions across days benefits the learning of motor skills. *Human Movement Science, 19*(5), 737–760. https://doi.org/10.1016/S0167-9457(00)00021-X.

Augusto, F., Eduardo, J., Modenesi, A., Guedes, K., Paula, T. D., Peterson, A., Elisa, M., & Piemonte, P. (2012). Motor learning, retention and transfer after virtual-reality-based training in parkinson's disease – effect of motor and cognitive demands of games: A longitudinal, controlled clinical study. *Physiotherapy, 98*(3), 217–223. https://doi.org/10.1016/j.physio.2012.06.001.

Nanopoulos, A., Alcock, R., & Manolopoulos, Y. (2001). Feature-based classification of time-series data. *International Journal of Computer Research, 10*(3), 49–61.

Tang, J., Alelyani, S., Liu, H. (2014). Feature selection for classification: a review. Data classification: Algorithms and applications, 37.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks, 61*, 85–117. https://doi.org/10.1016/j.neunet.2014.09.003 URLhttp://www.sciencedirect.com/science/article/pii/S0893608014002135.