



**HAL**  
open science

# A study on improving performance in gesture training through visual guidance based on learners' errors

Florian Jeanne, Indira Thouvenin, Alban Lenglet

► **To cite this version:**

Florian Jeanne, Indira Thouvenin, Alban Lenglet. A study on improving performance in gesture training through visual guidance based on learners' errors. 23rd ACM Symposium on Virtual Reality Software and Technology (VRST'17), Nov 2017, Göteborg, Sweden. pp.1-10, 10.1145/3139131.3139144 . hal-01585864

**HAL Id: hal-01585864**

**<https://hal.science/hal-01585864>**

Submitted on 20 Feb 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A study on improving performance in gesture training through visual guidance based on learners' errors

Florian Jeanne  
Sorbonne universités,  
Université de Technologie de  
Compiègne,  
CNRS UMR 7253 Heudiasyc  
57 Av. Landshut CS 60319  
60203 Compiègne cedex  
florian.jeanne@hds.utc.fr

Indira Thouvenin  
Sorbonne universités,  
Université de Technologie de  
Compiègne,  
CNRS UMR 7253 Heudiasyc  
57 Av. Landshut CS 60319  
60203 Compiègne cedex  
indira.thouvenin@hds.utc.fr

Alban Lenglet  
Sorbonne universités,  
Université de Technologie de  
Compiègne,  
CNRS UMR 7253 Heudiasyc  
57 Av. Landshut CS 60319  
60203 Compiègne cedex  
alban.lenglet@hds.utc.fr

## ABSTRACT

Gesture training, especially for technical gestures, requires supervisors to point out errors made by trainees. Virtual reality (VR) makes it possible to reduce reliance on supervisors (fewer interventions and of shorter duration) and to reduce the length of training, using extrinsic feedback that provides training or learning assistance using different modalities (visual, auditory, and haptic). Visual feedback has received much attention in recent decades. Users can be guided by a *metaphor* in a virtual environment. This metaphor may be a 3D trace of canonical movements, a visual cue pointing in the right direction, or gestures by an avatar that the trainee must mimic. However, with many kinds of feedback, trainees are not aware of their errors while performing gestures. Our hypothesis is that guiding users with a dynamic metaphor based on the visualization of errors will reduce these errors and improve performance. To this end, in a previous work we designed and implemented a new 3D metaphor called *EBAGG* to guide users in real time.

In the present paper we evaluate *EBAGG* in relation to two other visual cues: first, a feedforward technique that displays the trace of a reference movement, and, second, a concurrent orientation feedback. The results of the user study show that *EBAGG* outperformed the others in improving users' performances over a training session. Moreover, the information assimilated during training with this dynamic feedback had a persistent effect when the metaphor was no longer displayed.

## CCS CONCEPTS

• **Human-centered computing** → **Interaction techniques**; Virtual reality; • **Software and its engineering** → *Virtual worlds training simulations*;

## KEYWORDS

Virtual reality, visual feedback, performance, guidance, gestures, user study

### ACM Reference format:

Florian Jeanne, Indira Thouvenin, and Alban Lenglet. 2017. A study on improving performance in gesture training through visual guidance based on

learners' errors. In *Proceedings of VRST '17, Gothenburg, Sweden, November 8–10, 2017*, 10 pages.

<https://doi.org/10.1145/3139131.3139144>

## 1 INTRODUCTION

Motor learning involves physiological and cognitive mechanisms acquired through practice or experience, which makes it a complex task. Feedback from an instructor is crucial to correct gesture or position errors that are made by the learner during the learning process. Once the instructor has departed, learners no longer receive feedback, and learning the gestures becomes considerably more challenging.

Immersive environments are known to be valuable tools for training in a variety of contexts including sporting performance [Covaci et al. 2015], the health sector [Moreau et al. 2007] and even assembly and maintenance scenarios [Louison et al. 2017]. Learners can train in safer and longer training sessions than in real environments, which has particular advantages in scenarios such as industrial maintenance where conditions are difficult. Training can take place away from the machines that are devoted to production, meaning that that these machines can neither be damaged by nor cause injury to learners. These immersive environments may include multisensory feedback that can act as a virtual coach.

In the absence of an instructor during gesture training, learners need augmented feedback (also known as extrinsic feedback). This is defined as information that cannot be given without the help of a trainer or a display [Sigrist et al. 2013], and is used to help learners visualize gestures, to guide them, and to correct or to report errors that are made. However, according to the guidance hypothesis, continuous concurrent feedback during the acquisition phase of learning leads to a dependency on the feedback used [Schmidt 1991; Sigrist et al. 2013]. When acquiring a motor skill, learners need to memorize it through their own intrinsic feedback, proprioception, that is to say the sensation of relative positions of their own body parts and the strength of the effort required to perform a movement. The guidance hypothesis forces them to ignore those feelings and to focus mostly on the feedback [Sigrist et al. 2013]. Learners may thus have good results during the initial practice, but then poor results at the transfer phase and when feedback is withdrawn.

In a previous work [Jeanne et al. 2017], we proposed a new metaphor for movement guidance based on error visualization. This feedback, *EBAGG* (*Error-based assistance for gesture guidance*),

*VRST '17, November 8–10, 2017, Gothenburg, Sweden*

© 2017 Association for Computing Machinery.

This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in *Proceedings of VRST '17, November 8–10, 2017*, <https://doi.org/10.1145/3139131.3139144>.

warns learners in real-time when they make trajectory errors. *EBAGG* does not explain how to correct the errors, but shows the direction in which the errors are occurring.

In this paper, we explore several approaches to visual guidance for gesture training in an industrial context. We compare *EBAGG* with two visual cues and analyze the impact of the different types of feedback on learners' performances. Our paper is organized as follows. Section 2 reviews related works on different visual cues for movement guidance, including feedforward and feedback cues. In section 3 we describe the design of our experiment. Section 4 presents the quantitative and qualitative results of the user study, and these are discussed in Section 5.

## 2 RELATED WORKS

Different types of feedback from different paradigms can be used in motor learning to guide users. In this section we look at previous work on guidance techniques: first we focus on feedforward techniques, then we study feedback used for real-time orientation, and finally error-based feedback.

### 2.1 Feedforward for guidance

In the context of maritime or fluvial navigation, the ideal trajectory can be suggested by displaying multiple traces showing the limits of the path that is to be followed. Benton and Walker [Benton and Walker 2004] used augmented reality (AR), and Fricoteaux *et al.* [Fricoteaux *et al.* 2014] used VR, to display this suggested path as in a head-up display. The path may also be represented in the form of waypoints or as a continuous line [Benton and Walker 2004; Fricoteaux *et al.* 2014]. Although this type of feedforward was originally designed for navigation, displaying traces is may also be used in gesture training.

In [Moreau *et al.* 2007], Moreau *et al.* focused on gesture analysis to compare learners' gestures with those of an expert, using the childbirth simulator BirthSIM. This was a virtual environment designed for skill transfer and acquisition of obstetric gestures, which allowed novices to acquire experience without any risk. The goal was for learners to understand thoroughly the gesture prior to training in the delivery ward. In this virtual environment concentric spheres ("guide spheres") with different radii (one, two and three centimeters), were placed along the reference path. Trainees were required to position the forceps blades as close as possible to the center of these spheres to reproduce the reference gesture.

Feedforward guidance is also used in sports training. In [Covaci *et al.* 2015], Covaci *et al.* proposed a basketball free-throw simulator and analyzed the transfer of motor skills from training in virtual environments to real practice. The authors compared several training paradigms, looking in particular at the effectiveness of visual guidance. Learners were helped in a way similar to the childbirth simulator mentioned above, via the display of a series of ellipses representing successive positions of the ball along the ideal trajectory.

This kind of visual cue enables learners to visualize the ideal movements to be emulated, but lacks dynamism in that it neither reacts nor adapts to users' gestures.

### 2.2 Orientation feedback

Tönns and Klinker review in [Tönns and Klinker 2009], 3D arrows as guidance feedback using AR in automotive environments. They distinguish three types of arrow, of which the first two are designed to reduce drivers' inattention without distracting them from the driving task. The first is an arrow displayed in the direction of the nearest danger, while the second is a bird's eye sketch of a car with a 2D arrow, also pointing in the direction of the nearest danger. Both are updated in real-time, allowing drivers to correct their trajectory. The third type of arrow discussed are navigational arrows displayed on the road at intersections that implicitly indicate the road to take. However, occlusion with other vehicles is an issue, especially if the driver is stuck in dense traffic. In this situation the assistance becomes less visible, possibly even hard to identify and to understand. These arrow solutions allow a direct mapping of the path but do not provide any feedback on errors. George *et al.* proposed a similar kind of feedback in [George *et al.* 2012]. The authors use the metaphor of weather vane to indicate a danger, its dangerousness and its criticality to the driver of a vehicle.

Likewise, Sodhi *et al.* proposed AR visual cues for hand movement guidance [Sodhi *et al.* 2012]. Four different types of feedback can be used and are projected directly onto the user's hand. The first one is a 2D black arrow (called *Follow Spot*) in a white disk that indicates if the hand needs to move up or down by changing its direction and its size. Another 2D arrow (*Hue Cue*) is presented in the next section. Then there is the *3D Arrow* pointing in the direction to take and the *3D Pathlet* metaphor showing a segment of the path. The latter includes a red dot representing user's relative position and a blue segment indicating the path to follow. However, when the hand is not visible or oriented at an extreme angle these features are no longer visible.

Moreover, for movement guidance, Henderson and Feiner proposed an HMD application using AR to assist users in the training of a maintenance assembly task [Henderson and Feiner 2011]. Real-time feedback such as dynamic 3D arrows and labels are provided to trainees to inform them of the next movements to be performed. The color and size of each arrow change in response to a user's activity.

Rovelo *et al.* [Rovelo *et al.* 2015] propose a mid-air gesture guidance system called *Gestu-Wan*. The authors divided gestures in different parts, gathered to form a tree. *Gestu-Wan* shows the most recent gesture to have been performed correctly, together with a set of different possible gestures from the tree that follow on from this previous gesture. Each of these gestures is sketched with a black silhouette, whereas the user's current gesture overlays them with a blue silhouette for comparison. This technique can be useful for communicating the next gesture to perform during simple mid-air gesture learning, but relatively difficult to use with complex or rapid series of gestures.

*YouMove* [Anderson *et al.* 2013] is a system for motor learning using a skeleton displayed on an augmented "mirror". Learners perform in front of a screen where they can see themselves overlaid by a skeleton reproducing the requested gesture. Yang *et al.* [Yang and Kim 2002] used a "ghost" metaphor to display the reference gesture during hand movement training. A gesture is recorded *a priori* and replayed by a semi-transparent avatar. Trainees use a

head-mounted display (HMD) to visualize the virtual environment and see themselves with an avatar. The two avatars are superposed and trainees attempt to reproduce the gesture performed by the ghost. These techniques are good for guidance, but limited to specific gestures. They do not show errors committed by learners who must deduce themselves where and when corrections are needed.

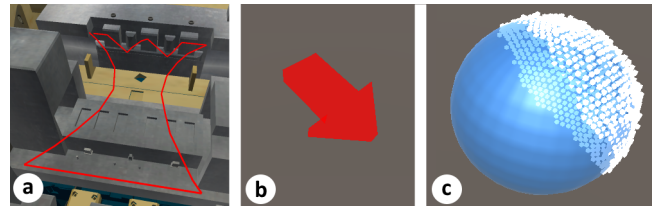
*MotionMA* is a system that automatically models a gesture performed by a user (generally an expert) and compare it with gestures performed by other users (generally trainees) in order to give them real-time feedback on where they have to improve [Velloso et al. 2013]. Users are represented by a skeleton avatar with each bone colored according to its score. The system also warns trainees when they are not moving at the right speed.

### 2.3 Error-based assistance

Wei *et al.* developed in [Wei et al. 2005] a system where users had to reproduce specific movements with a robotic arm to reach a target. They performed these movements with a visual distortion and some movements had error augmentation. This experiment showed that increasing the perceived error can accelerate the rate of learning during neurorehabilitation. The authors propose that error augmentation may trigger the recovery process. Rehabilitation of stroke patients have also been studied with VR systems. Kapur *et al.* proposed in [Kapur et al. 2010] a sleeve that can recognize upper limb movement and give tactile feedback. Ideomotor limb apraxia patients often have difficulty imitating movements that are demonstrated visually, so the use of vibrotactile feedback can be advantageous.

Vibrotactile feedback is not only used for rehabilitation, but also to highlight errors. To improve human motor learning, Lieberman and Breazeal proposed the system that they call *TIKL: tactile interaction for kinesthetic learning* [Lieberman and Breazeal 2007]. *TIKL* is a robotic wearable suit incorporating several vibrotactile actuators that analyzes learners' arm movements and gives real-time feedback on any errors they make. For instance, if learners have to bend their wrists in the same way the supervisor did, the actuator placed on the movement side will vibrate if the learner's wrist is bent too far inwards. Moreover, the greater the error, the stronger the vibration. Schönauer *et al.* also studied vibrotactile feedback for motion guidance in [Schönauer et al. 2012]. The system includes vibrotactile actuators placed on the user's arm, a pneumatic feedback vest which provides stronger tactile sensations, and two types of visual feedback, namely a ghost avatar and directional arrows.

Haptic guidance is defined by Feygin *et al.* as physical assistance provided by a haptic interface to guide through the ideal motion [Feygin et al. 2002]. Haptic augmented feedback extends this definition and refers to any kind of haptic perception that guides towards and not necessarily through, the desired movement [Sigrist et al. 2013]. In [Marchal-Crespo et al. 2010], Marchal-Crespo *et al.* studied the effect of haptic guidance in a driving simulator. Learners had to drive in the virtual environment and forces were applied by the steering wheel to the learners' hands to bring the vehicle back on the right path. The level of guidance was reduced when learners were steering without large errors, and increased when errors were made. This study showed that haptic guidance is beneficial for long-term retention of a driving skill.



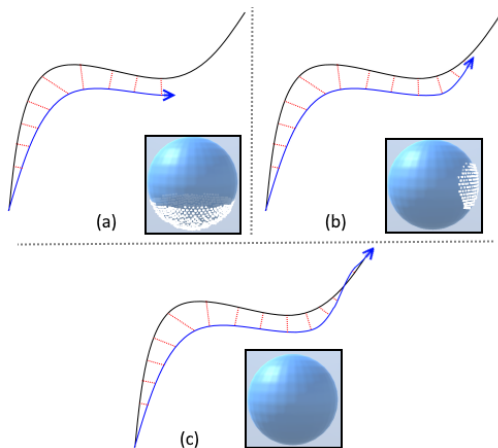
**Figure 1: An overview of the 3D cues used to guide users during the experiment. (a) is the path of the reference gesture displayed by a succession of red segments, (b) is a red 3D arrow showing the direction to take to get back on the correct path and (c) is the metaphor *EBAGG* displaying trajectory errors.**

As mentioned previously, Sodhi *et al.* also proposed a metaphor, *Hue Cue*, based on a display of trajectory errors. *Hue Cue* is an arrow (*Follow Spot*) surrounded by positive and negative spatial coloring representing a two-dimensional ( $x$  and  $y$ ) space. To perform the right gesture, users have to move towards the positive coloring and away from the negative coloring. Here, the arrow *Follow Spot* is used to correct depth errors ( $z$ ). This type of visual cues have proven to guide accurately for movement of a single body part in space at a controlled speed.

## 3 USER STUDY

*EBAGG* was introduced in a previous paper [Jeanne et al. 2017], and as mentioned in section 1, it is a metaphor based on error visualization. *EBAGG* takes the form of a sphere, which makes mapping of the 3D space easier (Figure 1-(c)). The sphere includes particles on its surface which appear when errors are made and in the direction of those errors; for example, if users deviate too much to the right of the ideal trajectory, they appear on the right side of the sphere. The number of visible particles depends on the value of errors: the greater the error, the larger the number of particles. Moreover, with this 3D cue, we don't show the path to follow, but instead shows subjects' deviations from the ideal path in order that they may perform the requisite adjustments and rejoin the correct path. In the manner of a sports coach, this feedback suggests corrections in the trajectory to perfect the gesture. The basic idea behind this metaphor was inspired by the work of Varela *et al.* on *enaction*, defined as the adaptation of people's behavior in reaction to changes in the environment in which they are located [Varela et al. 1992]. Learners modify their gestures according to the behavior of the metaphor, which changes in response to their trajectory, as we can see on Figure 2.

In order to evaluate the performances and usability of *EBAGG*, we conducted a user study where we compared it to standard metaphors using different paradigms. The goal was to determine whether *EBAGG* metaphor could attain the same level of performance as the standards metaphors for a technical gesture training task in a virtual foundry. *EBAGG* clearly belongs to the error-based assistance category, thus we first need to pick several cues from other paradigms: feedforward and orientation feedback.



**Figure 2: The behaviour of EBAGG.** The reference trajectory is represented in black, the user’s trajectory in blue and the distance between the two sequences is in red. (1) The gap was important so EBAGG displays many particles on the side of the error. (2) The gap has been reduced so there are less particles. (3) The user’s trajectory is similar to the reference one, hence there is no particle.

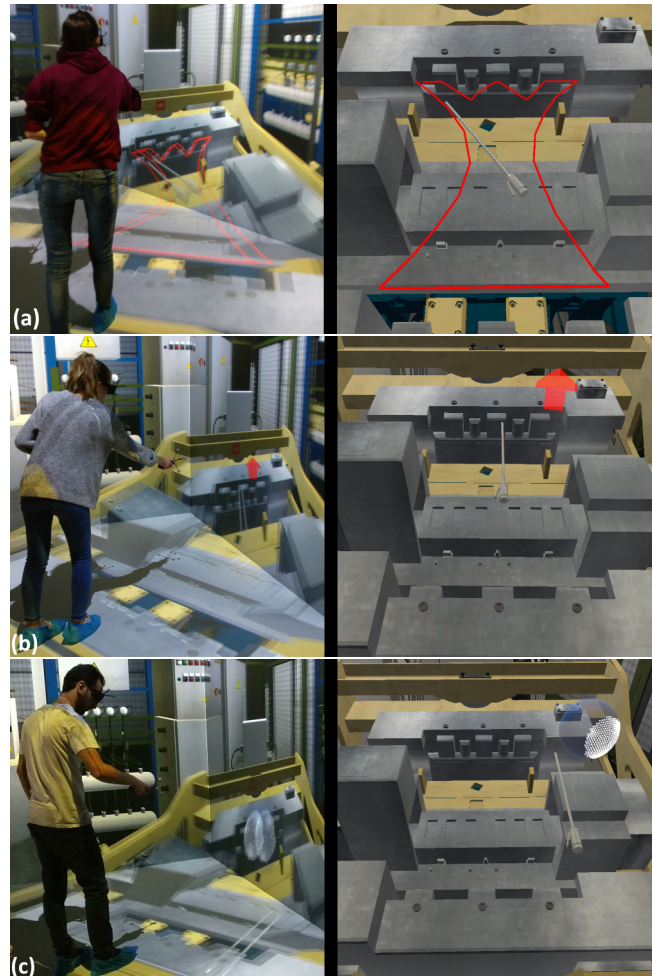
### 3.1 Guidance hints

In the light of the related works described in Section 2, we selected two types of feedback against which to compare EBAGG’s performance.

Feedforward techniques are very useful for guiding subjects through the ideal motion, since they provide information about the gesture to perform before subjects have started to perform it. Thus we decided to design a 3D path of the reference gesture as shown in Figure 1-(a). The path was composed of a number of red segments positioned along the ideal motion in the same way as in [Covaci et al. 2015; Moreau et al. 2007]. Subjects were required to follow this path to reproduce the reference gesture.

We also chose to compare EBAGG to a technique of orientation feedback. We decided against including in our study a third paradigm based on an avatar, because this would have involved tracking the whole of the subject’s body and using a third person view, whereas feedforward and orientation feedback use a first person view. The orientation feedback solution that we adopted was a 3D arrow, as shown in Figure 1-(b). This red arrow points in real time to the direction of the ideal motion so that subjects can rejoin the correct path after deviating from it. Like EBAGG’s particles, this arrow appears when subjects are making errors and disappears when they are not. However, it does not directly show the error that has been made, but rather the direction in which the correction is to be made. Subjects have to deduce from the presence of this visual cue that they are making errors.

EBAGG and the 3D arrow provide feedback in real time. We used a specific guidance algorithm based on the dynamic time warping algorithm (DTW) presented in a previous paper. The DTW algorithm provides interpretable results in the form of a discrepancy between two time series, for example between the reference gesture



**Figure 3: Each figure includes a view on a user using a specific hint and a view of this visual cue in the virtual environment.**

and the gesture actually performed [Berndt and Clifford 1994]. We use the computed distance to determine whether users are sufficiently far from the ideal motion to warrant the provision of feedback. To perform real-time calculation, we have adapted the DTW algorithm as presented in the previous paper.

### 3.2 Expected outcomes and hypotheses

Participants were divided in four groups: the first three groups were allotted respectively to the 3D path, the 3D arrow, and EBAGG, while the fourth group was a control group in which subjects received no feedback during training. The goal of the experiment was to compare the EBAGG metaphor to standard metaphors, namely a visualization of the ideal motion and a 3D arrow pointing in the direction of the ideal motion. We studied the levels of performance reached by participants for each 3D visual cue.

Before conducting the user study we made the following hypotheses:

- **H1:** Subjects in the three groups receiving feedback during the training phase would perform better than subjects from the control group.
- **H2:** Subjects in the *EBAGG* group would perform better than those in the other groups.
- **H3:** Subjects in the *EBAGG* group would continue to perform as well once the metaphor was no longer displayed.

### 3.3 Participants

The study was performed with forty-eight healthy participants (twelve females and thirty-six males) aged from 18 to 60 years old (mean: 27; std dev: 9.36). As mentioned previously, they were divided in four groups of twelve subjects. Only six participants were left-handed but they had the same experimental conditions since the dynamic visual cues (the arrow and *EBAGG*) were following the wand. They had variable experience with VR devices, especially *CAVEs*, and were drawn from among people working in the laboratory and students at the university. Participants did not receive any remuneration for their participation.

### 3.4 Apparatus

The experiment was performed in a *CAVE* (cave automatic virtual environment). This is a cubic immersive VR system comprising several screens where our stereoscopic 3D projectors display images on the paired screen. Each screen that constitutes a wall of the *CAVE* is a 3.4m x 2.5m glass screen, while the floor is a 7m x 3.4m reflective painted surface (Figure 3). This *CAVE* has a movable wall that can change it from a closed U-configuration to an open L-configuration. For the present experiment the L-configuration was chosen since it allows us more space on either side of the user for the display of the virtual machine.

The *CAVE* has a motion tracking system including eight infrared OptiTrack Prime 13W cameras and two OptiTrack Prime 13 cameras<sup>1</sup>, which provide a resolution of 1280x1024 pixels with an adjustable frame rate between 30 and 240 fps (frames per second). This system tracked the several reflective markers placed on the stereoscopic glasses worn by subjects, and hence their point of view. The images distorted by the corners of the screens could thus be readjusted to fit users' vision. Some markers were also placed on a controller whose buttons let users start or stop recording their gestures.

The application ran on a 3D rendering computer with two CPU Intel Xeon E5-2667 at 2.90 GHz (max 3.50 GHz) and two GPU Nvidia M6000. The virtual environment was designed using the Unity game engine<sup>2</sup> and represented a virtual metal foundry. Moreover, a virtual compressed-air blower was co-located with the controller.

### 3.5 Procedure

Participants were asked to perform a technical gesture in front of a virtual mold with or without the help of a visual cue according to their group. First of all, they were immersed in a basic training environment where they could become familiar with the VR equipment. During this familiarization phase the experiment supervisor explained the task they were going to perform. In a previous

experiment we had asked the participants to reproduce the gesture thirteen times, but that had proved insufficient. This time they were asked to reproduce a technical gesture (corresponding to the removal from the mold of residual particles from a previous casting) twenty times. Using the virtual air blower, they had to move through specific locations while moving in the three dimensions of space, as shown on Figure 1-(a). The gesture was shown twice to the participants, who were not allowed to practice it before training so that the experimental conditions were identical each time. For the three groups who were to have a visual aid the supervisor explained the concept and the behavior of the relevant cue. Once these explications had been completed, training could begin.

The experiment included three training phases for the three feedback groups: a *pre-training* phase of five iterations where subjects did not receive any feedback on their performances; a *training* phase of ten iterations during which participants were assisted by the 3D path, the 3D arrow or the *EBAGG* metaphor according to their group; a *post-training* phase of five iterations where feedback was withdrawn and participants had to perform gestures without any help. The 3D path was a static visual cue located above the mold along the reference gesture, while the 3D arrow and *EBAGG* were dynamic cues following the controller and consequently the subject's hand. The control group simply had to reproduce the gesture twenty times without any help. After performing the twenty gestures, participants were given a questionnaire about the experiment and where applicable about the 3D cue.

## 4 RESULTS

This section describes the results obtained from the experiment described in Section 3 above. The results are analyzed for each group; we performed a linear regression on the first fifteen iterations to see whether the no-feedback training or the feedback training had an impact on performance. We obtained straight lines corresponding to the equation 1, the values are gathered in table 1. Then we analysed the results of post-training for each group.

$$y = \alpha * x + \beta \quad (1)$$

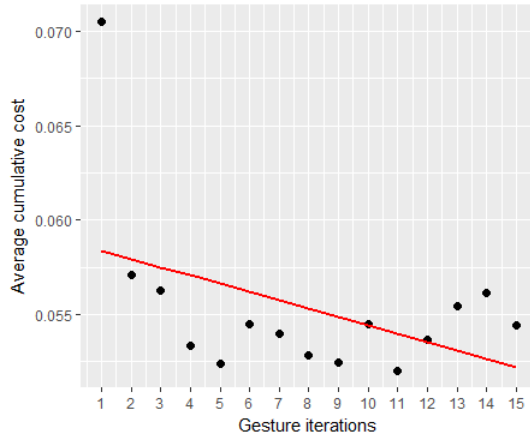
**Table 1: Values of the slope and y-intercept of the equation 1 for each group**

	Group			
	Control	3D Path	3D Arrow	EBAGG
$\alpha$ values	-0.0004408	-0.002269	-0.000056	-0.0008376
$\beta$ values	0.0588256	0.059406	0.04546	0.0480623

Data from a subject's gesture was acquired from the moment that the subject began the gesture by clicking on a specific button of the controller, up to the moment when the subject ended the gesture by clicking again on the same button. During data acquisition the position of the controller was saved and then sent to the DTW algorithm at a frequency of five messages per second. The DTW algorithm compared users' gestures to a reference gesture recorded prior to the experiment. It returned a matching between the last position of the controller and a specific position of the reference

<sup>1</sup><http://optitrack.com>

<sup>2</sup><https://unity3d.com>



**Figure 4: Linear regression on the average cumulative cost of the first fifteen iterations for the control group**

time sequence, and the Euclidean distance between the two. A cumulative cost on the partial gesture that takes account on this distance was also provided by the DTW. We use this cumulative cost as a performance scores since a higher cumulative cost indicates a greater deviation by the subject from the reference gesture.

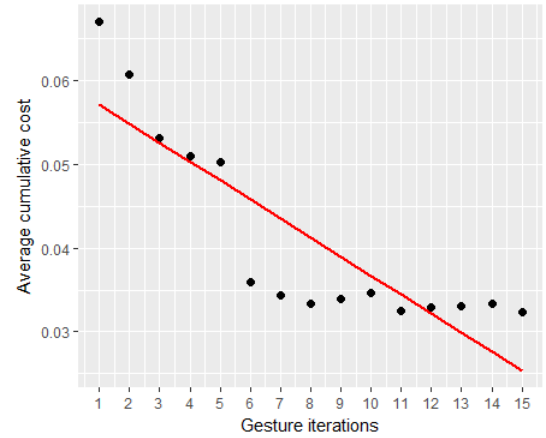
## 4.1 Training analysis

**4.1.1 Control group.** Figure 4 shows the linear regression for the first fifteen iterations in the group that had no feedback during the training phase.

We used a Shapiro-Wilk test on the data that showed that data was normally distributed. From the analysis of variance we obtained that there is no significant effect ( $F(1, 13) = 3.1549, p = 0.09909$ ) and that the slope (Equation 1 and Table 1) has a low value, which means that the linear regression is almost constant and that subjects did not significantly improve. The linear regression therefore does not fit well the data ( $R^2 = 0.1953$ ), and in the absence of assistance subjects' performances failed to improve, suggesting that the repetition of gestures is not sufficient as a training strategy.

**4.1.2 3D path group.** Figure 5 shows the linear regression for the first fifteen iterations in the group that had the 3D path cue during the training phase.

We used a Shapiro-Wilk test on the data that showed that data was normally distributed. This time, the analysis of variance showed a significant effect for this visual cue ( $F(1, 13) = 36.88, p = 0.00004$ ). Moreover, since  $\alpha = -0.002269$ , the regression line falls steeply (Equation 1 and Table 1), which suggests an effect of training over time, and the R-squared value is  $R^2 = 0.7393$ , which indicates that this model fits well the data relatively well for the first fifteen iterations. The 3D path cue is seen to have improved subjects' gesture training performances under the experimental conditions described. Displaying the reference path improved overall performance insofar as subjects only needed to follow this path in order to obtain good results. As we can see, the average value of the cumulative cost decreased considerably from  $\bar{y}_{pre} = 0.0564$  during pre-training to



**Figure 5: Linear regression on the average cumulative cost of the first fifteen iterations for the 3D path group**

$\bar{y}_{tbegin} = 0.0344$  and  $\bar{y}_{tend} = 0.0329$  during respectively the first five iterations and the last five iterations during training.

**4.1.3 3D arrow group.** Figure 6 shows the linear regression for the first fifteen iterations in the group that had the 3D arrow cue during the training phase.

We used a Shapiro-Wilk test on the data that showed that data was normally distributed. Again, the analysis of variance did not show any significant effect for the 3D arrow cue ( $F(1, 13) = 0.0312, p = 0.8625$ ). The regression line is almost horizontal, with a slope value of  $\alpha \sim 10^{-5}$ , which indicates that the 3D arrow did not improve subjects' performances. In addition, the the R-squared value is  $R^2 = 0.0024$  which indicates that the linear regression does not fit well the data. The mean cumulative cost for the pre-training phase is  $\bar{y}_{pre} = 0.0465$ ,  $\bar{y}_{tbegin} = 0.0423$  for the first five iterations of the training phase and  $\bar{y}_{tend} = 0.0463$  for the end of training phase. The improvement is minimal, and we can also remark a deterioration in performance during the training phase.

**4.1.4 EBAGG group.** Figure 7 shows the linear regression for the first fifteen iterations in the group that had the EBAGG metaphor during the training phase.

We used a Shapiro-Wilk test on the data that showed that data was normally distributed. Here, the analysis of variance revealed a significant effect for the EBAGG metaphor over the first fifteen iterations ( $F(1, 13) = 28.74, p = 0.00013$ ). The value of the slope is low ( $\alpha \sim 10^{-4}$ ), hence the regression line sharply decreases and the R-squared value is  $R^2 = 0.6885$ . So this model fits the data fairly well, and we may deduce that the EBAGG metaphor improved subjects' gesture training performances. In contrast to the results for the 3D path group, performance does not improve quickly at the beginning of training phase, and there is no obvious gap between the fifth and the sixth iterations. Nevertheless, the average cumulative value also decreases from  $\bar{y}_{pre} = 0.0464$  for the pre-training phase to  $\bar{y}_{tbegin} = 0.0395$  and  $\bar{y}_{tend} = 0.0383$  for the first five and the last five iterations in the training phase.

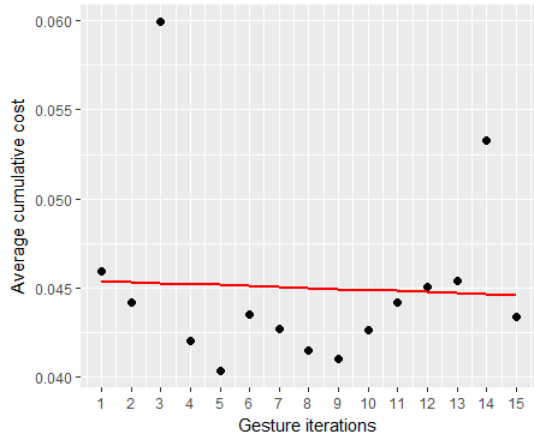


Figure 6: Linear regression on the average cumulative cost of the first fifteen iterations for the 3D arrow group

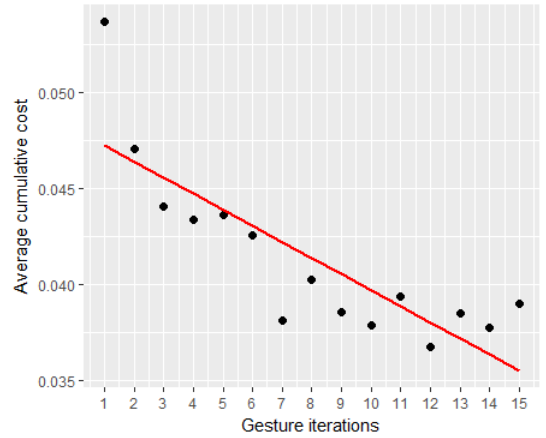


Figure 7: Linear regression on the average cumulative cost of the first fifteen iterations for the EBAGG group

## 4.2 Post-training analysis

Visual cues were withdrawn for the feedback group and the purpose of this phase was to determine whether these visual cues had any impact on performance in the very short-term.

As we can see from Figure 8, different groups had different levels of performance as regards the pre-training phase; for the control group and the 3D path group the average value of the cumulative cost is  $\bar{y}_{pre} \approx 0.057$ , whereas for the 3D arrow group and EBAGG group the value is  $\bar{y}_{pre} \approx 0.046$ . The post-training results between the different groups are therefore not comparable, since the discrepancy between them might be explained by something other than training. The impact of outliers could have been reduced with more participants. We therefore chose to perform an intragroup, as opposed to an intergroup, comparison. We analyzed the difference in performances between pre-training and post-training, and also between the end of training and the post-training. We used the Shapiro-Wilk test to verify whether data was normally distributed, and when it was verified, we conducted a two-sample Student's  $t$ -test.

For the control group, participants' performance during the post-training is in the same order of magnitude as the previous phases:  $\bar{y}_{pre} = 0.0579$ ,  $\bar{y}_{tbegin} = 0.0537$ ,  $\bar{y}_{tend} = 0.0543$  and  $\bar{y}_{post} = 0.0538$ . These results from the post-training confirm and match our linear regression: the level of performances remains constant.

For the 3D path group, however, the improvement in performance obtained during the first fifteen iterations decreased during the training phase, as Figure 8 shows. The average value of the cumulative cost increases from  $\bar{y}_{tend} = 0.0329$  to  $\bar{y}_{post} = 0.0441$  (two-sample  $t = -33.7, p < 0.001$ ) but is still lower than  $\bar{y}_{pre} = 0.0564$  (two-sample  $t = 3.81, p < 0.05$ ). This suggests that a lot of the information assimilated during the training phase with the 3D path cue was lost when this visual cue was withdrawn. Training might nevertheless still have been of some benefit to subjects in improving performance.

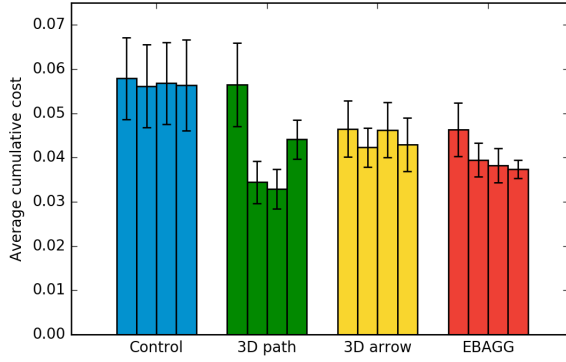
For the 3D arrow group there is no significant improvement between the pre-training and the post-training phase (two-sample  $t = 1.02, p = 0.37$ ) as suggested by the linear regression.

For the EBAGG group, if we compare the data from the pre-training phase ( $\bar{y}_{pre} = 0.0464$ ) to the data from the post-training phase ( $\bar{y}_{post} = 0.0374$ ) we can notice a significant improvement (two-sample  $t = 4.57, p = 0.009$ ). We can also remark from Figure 8 that subjects' levels of performance during post-training were fairly close to their levels of performance at the end of training, even slightly better. They decrease from  $\bar{y}_{tend} = 0.0383$  to  $\bar{y}_{post} = 0.0374$ . However, this difference is not significant (two-sample  $t = 1.64, p = 0.15$ ), which means that when subjects no longer had feedback, they reused a large amount of the information they acquired from the metaphor during the training phase. The slight improvement could partly be explained in terms of an over-correction by the subjects, who may have been more focused on the metaphor than on the gesture being performed. Moreover, we notice in Figure 8 that the 95% confidence interval is very short for the post-training of the EBAGG group. This means that the standard deviation is also low, and hence that there were few differences between subjects' levels of performance during the post-training phase.

Finally, we carried out a sharper analysis of the effect on training for the 3D path cue and the EBAGG metaphor. We calculated the mean squared error (Equation 2) and hence compared the last five values obtained from the experiment to the values predicted by our linear models. These values are reported in table 2 for the 3D path and in table 3 for EBAGG. We did not study the values of the control group, nor those of the 3D arrow group, given that neither the lack of assistance nor the 3D arrow cue had any significant effect, and neither were of any benefit to subjects over the first fifteen iterations.

$$MSE = \frac{1}{n} \sum_{i=j}^k (y_i - \hat{y}_i)^2 \quad (2)$$





**Figure 8: Average cumulative cost of the twenty iterations per group over the three training phases. On this graph and for each group, the first bar corresponds to the pre-training phase, the second and third bars correspond to the training phase and the fourth bar corresponds to the post-training phase. Each bar represents five iterations of the gesture and the error bars represent 95% confidence intervals.**

**Table 2: Average cumulative costs estimated by the model and empirical values obtained from the experiment for the 3D path group (*Est.*=estimated, *Emp.*=empirical)**

	Iterations				
	16	17	18	19	20
Est. $y$	0.02310	0.02083	0.01856	0.01630	0.01403
Emp. $y$	0.04381	0.04322	0.04434	0.04453	0.04471

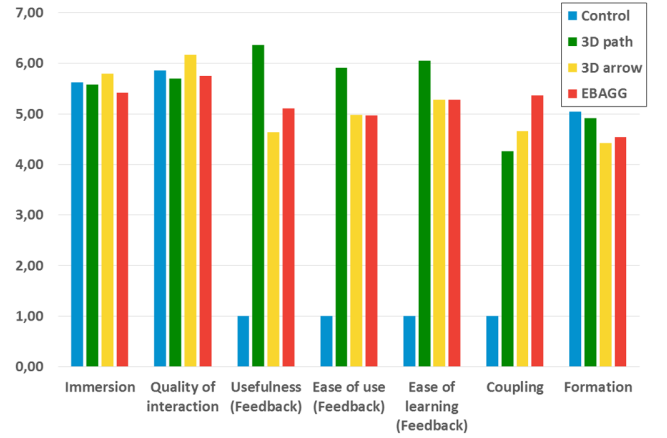
**Table 3: Average cumulative costs estimated by the model and empirical values obtained from the experiment for the EBAGG group (*Est.*=estimated, *Emp.*=empirical)**

	Iterations				
	16	17	18	19	20
Est. $y$	0.03466	0.03382	0.03299	0.03215	0.03131
Emp. $y$	0.03806	0.03744	0.03638	0.03735	0.03768

Here in the equation 2,  $n$  is the number of iterations ( $n = 5$ ),  $j$  is the first iteration studied ( $j = 16$ ), and  $k$  the last one ( $k = 20$ ).

From Table 2 we obtained a value of  $MSE_t = 6.66 \times 10^{-4}$  and from Table 3 we obtained  $MSE_e = 2.08 \times 10^{-5}$ .

In both cases we remark a deterioration with respect to the estimated values of the linear models, that is to say empirical values are higher than estimated values. For the 3D path, the empirical values had a mean value of  $\bar{y}_{emp} = 0.0441$  versus  $\bar{y}_{est} = 0.0186$  for the estimated values, while for EBAGG the empirical values had a mean value of  $\bar{y}_{emp} = 0.0374$  versus  $\bar{y}_{est} = 0.0330$  for the estimated values. This might be explained by a lack of feedback during the post-training phase; participants did not perform as well as during the training phase. However, the discrepancy is much smaller for



**Figure 9: Results from the questionnaire using a seven-point Likert scale.**

EBAGG, as can be seen in Figure 8. For EBAGG post-training performances were slightly better, but not significantly, than at the end of training, while for the 3D path they were significantly lower.

### 4.3 Subjective evaluation

After performing the gesture twenty times the subjects were asked to fill in a questionnaire. This questionnaire included twenty-six questions and was adapted from the Witmer Presence questionnaire [Witmer and Singer 1998] and from the USE Questionnaire [Lund 2001]. The goal was to obtain a subjective assessment by the participants regarding the experiment and the different types of feedback with respect to seven criteria: immersion, quality of interaction, usefulness of the visual hint, ease of use of the visual cue, ease of learning of the visual cue, quality of coupling and training. Results are reported in Figure 9.

The questionnaire was designed with statements grouped into those seven categories, and responses were entered using a seven-point Likert scale. Participants were asked to rate their level of agreement with each statement (from 1=*strongly disagree* to 7=*strongly agree*). The first statements concerned the quality of immersion, whether subjects felt immersed in the virtual environment and whether they were comfortable with the equipment. Overall, the participants were satisfied with the quality of immersion (mean: 5.79, standard deviation (std): 1.13). Only eight of the forty-two participants experienced discomfort with the stereoscopic glasses, mostly because they were too large. Regarding the quality of interaction, they felt that their interactions were natural (mean: 5.87, std: 1.18).

There followed (except for subjects in the control group) a series of statements from the USE questionnaire about the feedback received during the training phase. This section included the following three categories: usefulness of the visual cue, ease of use of the visual cue, ease of learning of the visual cue. Here, the 3D path had the best results. In the opinion of the participants the 3D path was the most useful with a mean score of 6.36 (std: 0.8), the easiest to use with a mean score of 5.92 (std: 1.26) and the easiest to learn how to use with a score of 6.06 (std: 1.07). This positive assessment

can almost certainly be explained by the simplicity of this type of feedback in terms of understanding it and using it. The easiest, quickest way to communicate a path is to display its trace, whereas the other two types of feedback are focused on error correction.

The next statements were about the coupling, which in the present context is defined as a mutual and continuous influence between a human user and a visual cue. They regulate each other and co-evolve [Varela et al. 1992]. Dynamic cues adapt their behavior according to users' actions, and users change the way they act depending on the feedback they get from the visual cues. This time, *EBAGG* had the best results (mean: 5.36, std: 1.42) and the 3D path the worst results (mean: 4.26, std: 2.48). This suggests that according to this questionnaire *EBAGG* has the greatest impact on training. Among the three types of feedback the 3D path had the worst results, whereas as we saw in the previous paragraph it had the best results regarding usefulness, ease of use and ease of learning. A possible explanation is that the 3D path is static, while the two others are dynamic. Users may be more likely to appropriate dynamic types of feedback as a part of themselves, an extension of their arm, since it was following the tool they were holding, than a static visual cue which seemed to belong to the virtual mold.

Finally, there were two statements about training and whether subjects felt they had improved over the twenty iterations of the gesture and whether they felt they needed a guidance cue during this experiment. Results are similar for the four groups, but the control group had the highest results with a mean score of 5.04 (std: 1.73), which is unsurprising given that they had no feedback during the training phase.

## 5 DISCUSSION

The results from the control group indicate that the repetition of gestures has little or no significant effect over training. Levels of performance remained constant. Participants reported in the questionnaire that they needed some feedback during training so they could be aware of their errors and correct them. Moreover, subjects from all three other groups with feedback had better performances, which suggests that our first hypothesis **H1** is validated.

The *EBAGG* metaphor showed significant results regarding training. First, there was a performance gap between the pre-training phase and the post-training phase. Participants were efficiently trained to reproduce the reference gesture, they improved continuously over the twenty iterations. The *EBAGG* group performed better than the other groups during post-training phase. The questionnaire showed that the coupling between the human user and the visual cue was better with the *EBAGG* metaphor. This would suggest that our second hypothesis **H2** is validated. However, we should like to conduct the experiment with a fresh group of participants to get more data and to compare the quantitative results of post-training. We note some similarities in the levels of performance for the last iterations of the training phase and the post-training phase, and even a slight improvement in performances, although this improvement was not significant. The consistency in levels of performance suggests that the information assimilated by subjects during training remained in their memory; they remembered the errors displayed by *EBAGG* and the corrections they needed to

make to adjust their gesture. They could still access this information and use it even when the metaphor was no longer displayed, in order to perform the gesture properly. This suggests that our third hypothesis **H3** is validated.

The performances of subjects whose visual guidance cue was the 3D path improved sharply during the training phase where the feedback was displayed. This may be explained by the effectiveness and ease of use of the 3D path, as suggested by the results of the questionnaire. Displaying the trace of the reference path seemed more natural to the participants, whose task only involved following the path. However, we remark a significant deterioration in performances in post-training. This might be linked to the guidance hypothesis mentioned in the introduction to this paper. Subjects may have been narrowly focused on properly following the 3D path rather than on reproducing the reference gesture. Thus, when the 3D path was no longer displayed, they had difficulty reproducing the gesture with the same level of performance as before, which may account for this significant deterioration. Moreover, the comparison of mean squared errors (MSE) shows that the gap between estimated and empirical values was much higher for the 3D path than for *EBAGG*, which might indicate a dependency generated by the 3D path. *EBAGG*, on the other hand, seemed to generate very little dependency. One possible explanation could be that subjects using the 3D path mostly relied on feedback of a visual kind, while those using *EBAGG* focused on their proprioception and on the feelings they had during the previous iterations. However, this experiment was an analysis of performances during a training session. We would like to conduct the same experiment as part of a study of long-term retention and to check whether the results are the same.

The results of the 3D arrow group indicate that this visual cue did not have any benefits at all in this kind of training. We even remark a deterioration in performances during the training phase. The responses to the questionnaire may shed some light on these results. Although subjects quickly understood the behavior of the 3D arrow during the familiarization phase, they struggled to understand what the 3D arrow was indicating. With no framework of reference, subjects were unable to understand precisely in which direction the arrow was pointing, and they reported that it was more of a hindrance than a help.

Finally, some of the participants reported some limitations of the *EBAGG* metaphor. When the controller was close to their body, the sphere became gigantic because it was too close to the virtual cameras, or no longer visible because it was behind the virtual cameras. Subjects had to step back in order to distance the controller from their body and hence to obtain *EBAGG* with a manageable size. Another limitation is the rate of computation of the DTW algorithm. This frequency is crucial for the behaviour of *EBAGG*. The higher the rate of computation, the more often the position of the particles on the surface of the metaphor are updated. The rate of computation is also linked to the number of successive positions of the reference gesture. Our DTW algorithm actually explores the entire time series at every computation to find the best matching. It also recomputes the previous matching to update the previous matchings. We were therefore limited to the number of points that the reference gesture could accept, so that our DTW algorithm could perform the requisite computations in real time. At specific

points in the gesture, such as the two curves (Figure 1-(a)), the gap between two positions was too large. So when participants made trajectory errors, EBAGG indicated erroneously that in addition to these errors the participants were also too far ahead in their gesture, because the last matching had been made using a previous point. Consequently, errors were recorded that had not necessarily been made. A possible improvement would be to perform GPU (graphics processing unit) computing. This would allow a more precise reference gesture with more points. In this scenario, the question arises as to whether EBAGG would still outperform the 3D path. Since the 3D path is a static visual cue, the addition of points would not change its shape, but EBAGG would be more precise and updated more often. It therefore appears clear that EBAGG would still outperform the 3D path.

## 6 CONCLUSION

In this paper we presented a user study designed to compare levels of performance obtained using three visual cues as part of gesture training in an industrial context. One of the visual cues under consideration was EBAGG, a metaphor that we introduced in [Jeanne et al. 2017]. This visual feedback is based on an error display to guide users in real time while performing a gesture. We compared EBAGG to two other visual cues, namely a feedforward technique that displays the trace of a reference motion (3D path), and a concurrent orientation feedback (3D arrow). We evaluated subjects' performances quantitatively using results from a DTW algorithm, and qualitatively with a questionnaire.

EBAGG had better results than the other groups. Subjects' performances improved continuously over the iterations during training, even though EBAGG was less easy to learn and use than the 3D path. Furthermore, the 3D arrow disturbed the participants and the 3D path generated dependency, reducing subjects' performances. In the case of EBAGG subjects' performances remained constant during the post-training phase, at the same level as during the training phase with feedback. EBAGG also had the best coupling with human users, which meant that it had the greatest impact on training. Our results show that EBAGG generated less dependency than the other types of feedback.

An important direction for future works is to analyze long-term retention. In this paper we have focused on EBAGG's ability to improve users' performances during gesture training, but we have not yet investigated its impact on learning over a long period. It would also be interesting to improve computation performance by assigning computation to a separate dedicated computer so as not to diminish performances of rendering.

## ACKNOWLEDGMENT

The authors would like to thank the Hauts-de-France Region and the European Regional Development Fund (ERDF) 2014/2020 for the funding of this work. We thank Montupet Group and Reviattech.

## REFERENCES

Fraser Anderson, Tovi Grossman, Justin Matejka, and George Fitzmaurice. 2013. YouMove: enhancing movement training with an augmented reality mirror. In *Proceedings of the 26th annual ACM symposium on User interface software and technology*. ACM, 311–320.

C Benton and R Walker. 2004. Augmented reality for maritime navigation: The next generation of electronic navigational aids. In *Proc. 7th Marine Transportation System Research & Technology Coordination Conf., Washington, DC*.

Donald J Berndt and James Clifford. 1994. Using Dynamic Time Warping to Find Patterns in Time Series. In *KDD workshop*, Vol. 10. Seattle, WA, 359–370.

Alexandra Covaci, Anne-Helene Olivier, and Franck Multon. 2015. Visual Perspective and Feedback Guidance for VR Free-Throw Training. *Computer Graphics and Applications, IEEE* 35, 5 (2015), 55–65.

David Feygin, Madeleine Keehner, and R Tendick. 2002. Haptic guidance: Experimental evaluation of a haptic training method for a perceptual motor skill. In *Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2002. HAPTICS 2002. Proceedings. 10th Symposium on*. IEEE, 40–47.

Loïc Fricoteaux, Indira Thouvenin, and Daniel Mestre. 2014. GULLIVER: A decision-making system based on user observation for an adaptive training in informed virtual environments. *Engineering Applications of Artificial Intelligence* 33 (2014), 47–57.

Paul George, Indira Thouvenin, Vincent Fremont, and Véronique Cherfaoui. 2012. DAARIA: Driver assistance by augmented reality for intelligent automobile. In *Intelligent Vehicles Symposium (IV), 2012 IEEE*. IEEE, 1043–1048.

Steven J Henderson and Steven K Feiner. 2011. Augmented reality in the psychomotor phase of a procedural task. In *Mixed and Augmented Reality (ISMAR), 2011 10th IEEE International Symposium on*. IEEE, 191–200.

Florian Jeanne, Yann Soullard, Ali Oker, and Indira Thouvenin. 2017. EBAGG: Error-Based Assistance for Gesture Guidance in Virtual Environments. In *IEEE ICALT 2017*. Timisoara, Romania. <https://hal.archives-ouvertes.fr/hal-01517869>

Pulkit Kapur, Mallory Jensen, Laurel J Buxbaum, Steven A Jax, and Katherine J Kuchenbecker. 2010. Spatially distributed tactile feedback for kinesthetic motion guidance. In *2010 IEEE Haptics Symposium*. IEEE, 519–526.

Jeff Lieberman and Cynthia Breazeal. 2007. TIKL: Development of a wearable vibrotactile feedback suit for improved human motor learning. *IEEE Transactions on Robotics* 23, 5 (2007), 919–926.

Céphise Louison, Fabien Ferlay, Delphine Keller, and Daniel R Mestre. 2017. Operators' accessibility studies for assembly and maintenance scenarios using virtual reality. *Fusion Engineering and Design* (2017).

Arnold M Lund. 2001. Measuring Usability with the USE Questionnaire. *Usability interface* 8, 2 (2001), 3–6.

Laura Marchal-Crespo, Stephanie McHughen, Steven C Cramer, and David J Reinkensmeyer. 2010. The effect of haptic guidance, aging, and initial skill level on motor learning of a steering task. *Experimental brain research* 201, 2 (2010), 209–220.

Richard Moreau, Minh Tu Pham, Tanneguy Redarce, and Olivier Dupuis. 2007. A new learning method for obstetric gestures using the BirthSIM simulator. In *Robotics and Automation, 2007 IEEE International Conference on*. IEEE, 2279–2284.

Gustavo Rovelto, Donald Degraen, Davy Vanacken, Kris Luyten, and Karin Coninx. 2015. Gestu-Wan-An Intelligible Mid-Air Gesture Guidance System for Walk-up-and-Use Displays. In *Human-Computer Interaction*. Springer, 368–386.

Richard A Schmidt. 1991. Frequent augmented feedback can degrade learning: Evidence and interpretations. In *Tutorials in motor neuroscience*. Springer, 59–75.

Christian Schönauer, Kenichiro Fukushi, Alex Olwal, Hannes Kaufmann, and Ramesh Raskar. 2012. Multimodal motion guidance: techniques for adaptive and dynamic feedback. In *Proceedings of the 14th ACM international conference on Multimodal interaction*. ACM, 133–140.

Roland Sigrist, Georg Rauter, Robert Riener, and Peter Wolf. 2013. Augmented visual, auditory, haptic, and multimodal feedback in motor learning: a review. *Psychonomic bulletin & review* 20, 1 (2013), 21–53.

Rajinder Sodhi, Hrvoje Benko, and Andrew Wilson. 2012. LightGuide: projected visualizations for hand movement guidance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 179–188.

Marcus TÖnnis and Gudrun Klinker. 2009. Augmented 3D arrows reach their limits in automotive environments. In *Mixed Reality In Architecture, Design And Construction*. Springer, 185–202.

Francisco J Varela, Evan Thompson, and Eleanor Rosch. 1992. *The embodied mind: Cognitive science and human experience*. MIT press.

Eduardo Velloso, Andreas Bulling, and Hans Gellersen. 2013. MotionMA: motion modelling and analysis by demonstration. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1309–1318.

Yejun Wei, James Patton, Preeti Bajaj, and Robert Scheidt. 2005. A real-time haptic/graphic demonstration of how error augmentation can enhance learning. In *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*. IEEE, 4406–4411.

Bob G Witmer and Michael J Singer. 1998. Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and virtual environments* 7, 3 (1998), 225–240.

Ungyeon Yang and Gerard Jounghyun Kim. 2002. Implementation and evaluation of "just follow me": an immersive, VR-based, motion-training system. *Presence: Teleoperators and Virtual Environments* 11, 3 (2002), 304–323.