



EBAGG: Error-Based Assistance for Gesture Guidance in Virtual Environments

Florian Jeanne, Yann Soullard, Ali Oker, Indira Thouvenin

► To cite this version:

Florian Jeanne, Yann Soullard, Ali Oker, Indira Thouvenin. EBAGG: Error-Based Assistance for Gesture Guidance in Virtual Environments. 17th IEEE International Conference on Advanced Learning Technologies (ICALT 2017), Jul 2017, Timisoara, Romania. pp.472-476, <10.1109/ICALT.2017.32>. <hal-01517869>

HAL Id: hal-01517869

<https://hal.science/hal-01517869v1>

Submitted on 3 May 2017

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.



HAL Authorization

EBAGG: Error-Based Assistance for Gesture Guidance in Virtual Environments

Florian Jeanne, Yann Soullard, Ali Oker and Indira Thouvenin

Sorbonne universités , Université de Technologie de Compiègne, CNRS UMR 7253 Heudiasyc

57 Av. Landshut CS 60319, 60203 Compiègne cedex, France

Emails: {florian.jeanne - yann.soullard - ali.oker - indira.thouvenin} @hds.utc.fr

Abstract—Augmented feedback has been shown to improve interaction in virtual environments and to facilitate motor learning. Recent studies proposed this type of feedback to guide users, to highlight specific areas or to help them to perform a specific task. They can follow a path, pass through specific waypoints or even mimic an avatar. However these approaches do not show the gap between learners’ performance and the desired one. Our hypothesis is that by revealing this gap to the users, they will reduce it step by step and tend to the required performance. Thus, in this paper, we propose a new visual metaphor to guide trainees’ gestures by showing trajectory errors instead of showing the path to follow. In a first study we evaluated trainees’ improvement by measuring the mentioned gap. First results indicate that our approach allows an enhanced task performance.

Keywords—Virtual reality, visual feedback, guidance, gestures

I. INTRODUCTION

Within the last few years, the popularity of Virtual Reality (VR) technologies has substantially increased. It has attracted a wider public including gamers or new technology enthusiasts, and has been popularized among industrials. This immersive technology allows them to train their operators and workers in a safe, yet faithful to reality, environment.

Motor learning is defined as processes leading to relatively permanent changes in the capability for movement, acquired through practice or experience [1]. It involves physiological and cognitive mechanisms, making some complex tasks or gesture hard to perform. On the other hand, VR has shown some promising solutions in various fields such as sport [2], medical science [3] or technical gesture [4].

Augmented feedback, or extrinsic feedback, has been shown to improve interaction in virtual environments and to facilitate motor learning [5], [6]. It can show the gap between the learner’s performance and the desired one. It can be used to highlight certain aspects of the realised movement, or some part of the instructions that may have been forgotten. Thus, in this context of motor learning, augmented feedback fosters the perception of errors, allowing learners to correct on their own.

Therefore, in this paper, we propose a new type of visual guidance feedback for motor learning in virtual environments based on error orientation. It is called EBAGG which stands for Error-Based Assistance for Gesture Guidance. Our work is based on the theory of embodied cognition [7], which stipulates that people’s representations of their own environment depend on their prior perceptual and motor experiences. Thus by experiencing errors, users are supposed to perform better

their gestures. As the visual modality appears to be beneficial for motor learning, our paper is focused on the validation of visual guidance feedback. The paper is divided in several sections : section II is a review of related studies, section III is the presentation of our metaphor. The evaluation of its ability to improve gesture is described in section IV and finally, in section V we conclude and introduce future works.

II. RELATED WORKS

In this section, we reviewed several existing approaches for guidance in virtual environments, including techniques of trajectory visualization, assistance guiding users in real-time and error orientation.

A. Trajectory visualization prior to the gesture

During the last decades, one way to show a gesture was to display its trajectory.

Moreau *et al.* [8] also used the technique of waypoints. In a virtual environment for skill transfer and acquisition of obstetric gestures, learners had to reproduce a technical gesture represented by concentric “guide spheres”, located at specific points of the reference path. Trainees had to pass the forceps blades the closest to the center of those spheres to perform as good as the reference. Covaci *et al.* [2] displayed the ideal trajectory with ellipses representing the position of the ball in a basketball free-throw simulator, in like manner as Moreau’s guide spheres. However, these types of feedback are not reactive and not adaptive to the users’ gestures.

B. Concurrent visual feedback for guidance

Due to its potential to enhance motor learning, concurrent feedback is more and more visually displayed.

A solution is to display an avatar performing gestures to reproduce. *YouMove* [9] was a system for motor learning using a skeleton displayed on an augmented “mirror”. Learners performed in front of a screen where they could see themselves overlaid by a skeleton reproducing the requested gesture. However, like the solutions presented in the last section, these avatars do not show gesture error.

For hand movement guidance, Sodhi *et al.* [10] projected AR guidance hints on users’ body. A red dot was showing the relative position of the user’s hand while a blue segment was showing the path to follow around this red dot. As it is a 3D visual hint, the segment projected a shadow on the user’s body, helping him to identify the direction to take. This last solution is close to what we propose, however it does not provide any feedback on the errors committed, letting users deduce whether their gesture is correct.

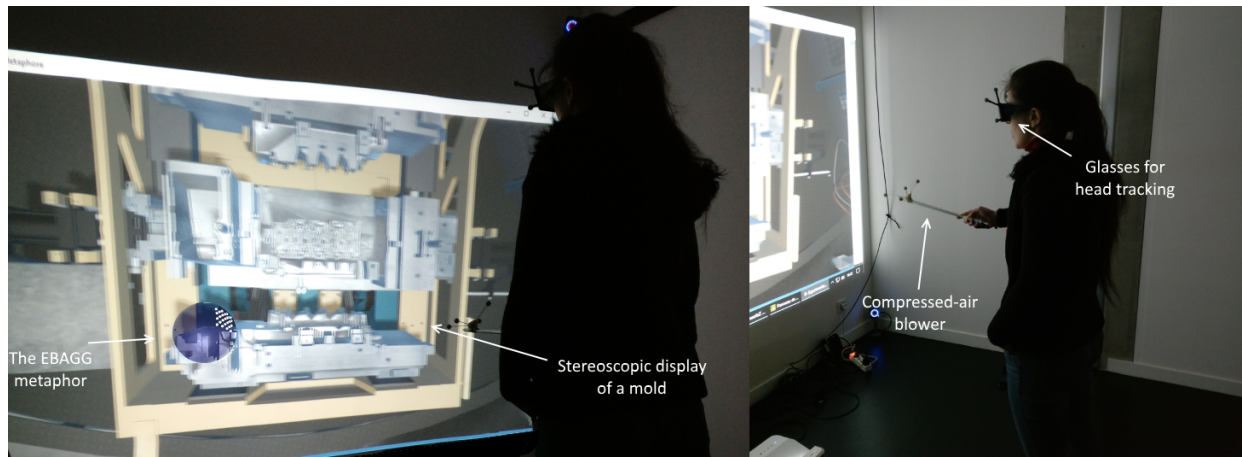


Figure 1. The EBAGG metaphor and the experimental setup

C. Error orientation feedback

For orientation feedback, you can either choose to display the ideal path to follow or indicate trajectory errors. In this paper, we decided to focus on error feedback.

Marchal-Crespo *et al.* conducted an experiment in [11] where participants had to drive in a driving simulator, with forces applied by steering wheel to their hands when errors were made. The aim of these forces were to correct the steering wheel motion and to put the users back on the right way. Also, the assistance was only provided when needed. If users had performed well and had had only a few mistakes, the guidance level was decreased and forces would be applied only if large errors were made.

Vibrotactile feedback is also used to highlight error. Lieberman *et al.* proposed in [12] a wearable suit including several vibrotactile actuators to guide users' arm motion. These actuators would vibrate to indicate motion errors. For instance, if users had to bend their wrist in the same way the teacher did, one actuator placed near by would vibrate if they bent too far. Also, as the error increased, vibrations would become stronger. This type of feedback informs users the direction to take to correct their gesture. As those examples which are error based, in our study we propose another type of feedback employing the visual modality.

III. ERROR-BASED ASSISTANCE FOR GESTURE GUIDANCE

The metaphor we propose is a visual guidance feedback for gesture learning. The goal here is to set parameters depending on the error value.

A. Concept

Our visual feedback can take different shapes. It can be primitives such as a sphere, a cylinder, a cube or even a patch, but it can also have no particular shape visible. Indeed, no matter which one is chosen, there will be particles on its surface. So, for instance, it can be designed as an invisible or visible sphere with particles moving around its surface. Each shape can be specific to a certain error criterion: the sphere

may be better to represent a three-dimensional space, whereas a cuboid would fit better for representing the orientation of a tool.

The EBAGG metaphor determines trajectory errors committed by users, it means that it shows the direction on which they have deviated. The particles fulfill this role by appearing on the surface while vibrating. In fact, when there is no error, they are invisible to users. They are only visible when an error is made or is no longer acceptable. In addition, the number of visible particles depends on the importance of the error. The bigger the measured value is, the higher the number of visible particles is. To determine the error value, we use a dynamic time warping algorithm (DTW), presented in the next section. The idea here is to indicate errors of direction to users instead of showing the direction to take.

B. Error calculation

To determine the performance of users, we compare their gesture with a reference gesture recorded beforehand. We calculate a distance between two gestures, and according to the gap we obtain, we indicate the error on the metaphor with the particles. To get the distance measure, we use a dynamic time warping algorithm [13][14] which provides interpretable results as a gap between two time series. Unlike learning methods (e.g. Hidden Markov Models), a DTW algorithm directly deals with the labeled examples (and does not model them) so the use of a DTW algorithm allows us to not have a gap between the data modeling and the true data distribution. In addition, gestures are generally multidimensional sequences, as they are represented by several time series (e.g. signals provided by different motion capture) so we use a natural extension of the standard DTW algorithm to multidimensional sequences, called a 1D-DTW on various dimensions in [15]. Besides, we need our feedback to be a real-time guidance. So, inspired by [16] we define an early recognition scheme for this algorithm, i.e. to deal with incomplete gestures. Finally, we parallelize calculations on different process to reduce even more the global calculation. This version of the DTW

algorithm allows us to use its results early and to be close to realtime with an offset of approximately 100ms.

The goal of a DTW algorithm is to find an optimal alignment between two time series, and a distance is calculated between these two series based on the optimal alignment. If $\mathbf{x} = (x_1, \dots, x_N)$ and $\mathbf{q} = (q_1, \dots, q_M)$ are two time series, the DTW algorithm computes a cumulative cost $C(n, m)$. This cost is defined as:

$$C(n, 1) = \sum_{i=1}^n d(x_i, q_1) \quad (1)$$

$$C(1, m) = \sum_{i=1}^m d(x_1, q_i) \quad (2)$$

For $1 < n \leq N$ and $1 < m \leq M$ the cost is computed using the following equation:

$$C(n, m) = \min\{C(n-1, m), C(n-1, m-1), C(n, m-1)\} + d(x_n, q_m) \quad (3)$$

Here, $d(x_n, q_m)$ is the distance between two points x_n and q_m (the Euclidean distance). Initial and final conditions are defined in such manner that the process begins and ends with the alignment of the first and last elements of the two series.

C. Implementation of EBAGG

Concerning the first implementation of our visual metaphor, we made several choices to make it relevant for gesture guidance with position as the only error criterion. First of all, the shape of EBAGG should represent a three dimensional space. A sphere or cylinder would both fit to show directional errors, however we don't need to pay attention to orientation since position is our only error criterion. A sphere can represent the three directions equally because of its central symmetry, so every point of its surface is at equal distance of the center. Thus, we assume that a sphere, paired with the particles, is the simplest and best shape to differentiate every direction of motion so that spatial information can be easily understood by users.

Figure 2 illustrates this implementation. (a) The user is too far from the reference trajectory so the distance between the two performances is important. Therefore, the EBAGG metaphor displays many particles on the side of the error. (b) The user's trajectory is now close to the reference one, hence there is less particles on the sphere. (c) Finally, the user's trajectory is similar to the reference, so the feedback does not need to display any particle.

IV. EVALUATION OF ERROR REDUCTION

A. Participants

Twenty-two healthy participants, including five women and seventeen men, aged from 21 to 42 years old (mean: 25.3; std dev: 4.6) took part in this experiment. Three of them were left-handed. They were people working in the laboratory or students at the university. None of them had ever used the EBAGG metaphor, nor any other error-based feedback modules before this experiment. Participants had normal or corrected-to-normal vision.

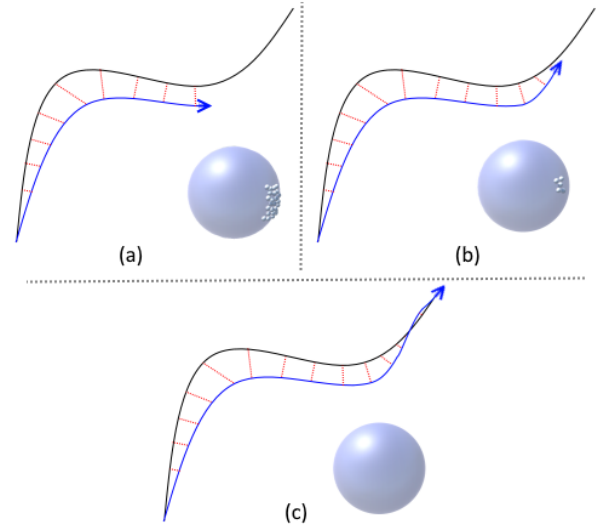


Figure 2. EBAGG metaphor's behaviour according to the distance between a user's performance and the reference performance. The reference trajectory is represented in black, the user's trajectory in blue and the distance between the two sequences is in red.

B. Setup and Stimuli

The experiment took place in a room equipped with motion capture cameras. There were five OptiTrack Prime 13W cameras, with an adjustable frequency between 30 fps (frames per second) and 240 fps. For this experiment, they were sampling at 120Hz, for a latency of 8ms. These cameras have had ten LEDs emitting infrared beams reflecting on mocap makers placed on the tracked tool. Thus, they were tracked by the cameras that were sending their positions to the virtual environment through Motive, the OptiTrack software. As we can see on figure 1, the stereoscopic display of a factory containing a mold was done on a wall by using the ultra short throw projector Optoma EH319USTi, paired with active 3D glasses Volfoni EDGE RF. Subjects were standing still equipped with the 3D glasses approximately 150 cm. in front of the display. The stimuli used to modulate participants gesture in this experiment was the sphere aforementioned as the EBAGG metaphor.

C. Experimental design and Task

Participants had to perform two technical gestures with a compressed-air blower in front of a virtual mold displayed on the wall. They had to pass through a specific trajectory and not to be too close or too far from the mold to have an efficient blowing.

a) *Gesture 1*: The first gesture was a straight line in three dimensions. The difficulty here was to go from one point to another by moving on a single dimension in the three dimensional space.

b) *Gesture 2*: The second gesture was a bit more complex. It was square-shaped with an arc instead of a straight line for the right side. This time, participants had to move on two different directions and had to stay in the same plane all

along the gesture. In both cases, the main difficulty was to position the air blower properly at the right depth.

The experiment began with a phase of familiarization of the virtual environment, mostly an adaptation to the stereoscopy, and of the EBAGG metaphor. The concept was detailed to the participants and then, they had the possibility to test the metaphor's reaction to their movements. Finally the two exercises could begin once the gesture to reproduce was shown twice and explained by the supervisor.

The experiment was divided in three phases for both gestures: firstly a *pre-formation phase* of three iterations with no feedback, then a *feedback formation phase* of seven iterations where the EBAGG metaphor was indicating trajectory errors to trainees, and finally a *post-formation phase* of three iterations with no feedback. Both exercises requested participants to reproduce the gesture thirteen times.

Here are our principal and secondary hypothesis:

- **Principal Hypothesis:** There will be a significant improvement of performances between the pre-formation phase and the post-formation phase;
- **Secondary Hypothesis:** There will be a significant difference between the first iterations and the last iterations of the feedback phase according to the complexity of the task.

D. Results and discussion

To evaluate participants' performances, their gestures were compared by the DTW algorithm to a reference gesture (one for each gesture to reproduce) recorded prior to the experiment. The DTW returned two pieces of information: on one hand the matching of the last position of the tool with the reference one and the distance computed between the two, and on the other hand the cumulative cost computed on the partial gesture. As mentioned before, the matching was used as an input for our metaphor to represent trajectory errors according to the vector between the two positions and the distance value. The cumulative cost acted like a score used to evaluate the global performance of users for each iteration. The more this cost were close to zero, the more users were close to the reference performance.

As the raw data obtained revealed to be not normally distributed (tested by the Kolmogorov-Smirnov test for normality), we conducted a Wilcoxon Test for our evaluation. This is not surprising given the fact that DTW provides a cumulative cost. The Wilcoxon test is a nonparametric test designed to evaluate the difference between two conditions as before-after treatments for the data which is not normally distributed. Thus, we conducted two tailed Wilcoxon test between the average values of pre-formation phase, post-formation phase and the first and last three iterations of feedback formation phase.

To complete this quantitative evaluation, we have submitted a questionnaire to the participants, which included 12 questions adapted from the Witmer presence questionnaire [17]. The rating score ranged from 0 (worst rating) to 10 (best rating). It provided us a subjective evaluation of this experiment and might explain certain results. The questions were about

users' adaptation to virtual reality and the environment itself, the use of the metaphor and about the training.

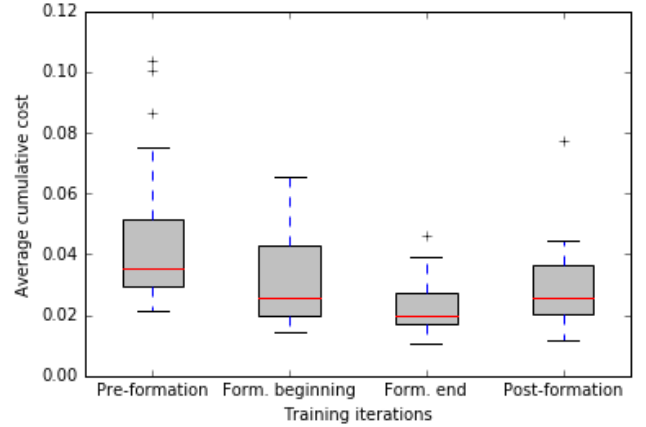


Figure 3. Average values of cumulative cost per user over the training iterations for the first gesture across 3 phases ("form." stands for formation)

a) *Results of First Gesture:* Figure 3 shows the box plots we obtain for the pre-formation phase, the feedback formation phase and the post-formation phase for the first gesture. When we compare the pre and post phase's iterations, the statistical analysis reveals that the average value improves from pre-formation (0.046) to post-formation (0.026) ($p < 0.05$). As this is an overall improvement of users' performances before and after the feedback, we deduce that **Principal Hypothesis** is confirmed for the first gesture.

Regarding our secondary hypothesis, the Wilcoxon analysis revealed that there is a significant difference between the first three iterations (0.032) and the last three iterations (0.023) of the feedback formation phase ($p < 0.005$). This means that feedback metaphor was rapidly sufficient to correct participants gesture, shown by better performance for the last iterations. As proposed by the second hypothesis, this result can be explained by the fact that the first gesture is not a complex one. Thus, only few iterations were enough to increase the performance.

b) *Results of Second Gesture:* Figure 4 shows the box plots we obtain for the pre-formation, the feedback formation and the post-formation phase for the the second gesture. Regarding our **Principal Hypothesis** for the second gesture, the Wilcoxon analysis revealed a significant effect between pre-formation phase and post-formation phase ($p < 0.05$). According to these results, subjects increased their performance in the post-formation phase (0.101) compared to the pre-formation phase (0.114).

Again, we conducted another Wilcoxon analysis between the first iterations and the last iterations of the feedback formation phase. Although there is an improvement for the descriptive statistics between the first iterations (0.119) and the second iterations (0.116) of the feedback phase, the difference is not significant ($p > 0.05$). According to our secondary hypothesis, these results show that more complex gestures necessitate more iterations. This is why, post-formation phase reveals only

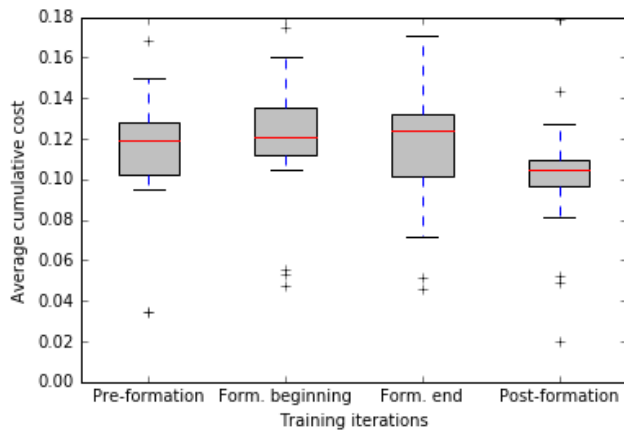


Figure 4. Average values of cumulative cost per user over the training iterations for the second gesture across 3 phases ("form." stands for formation)

an increase of performance (0.101) after the integration of the whole formation iterations by the participants. Compared to the first gesture, our results confirm the idea that the number of formation iterations are strongly dependent on the complexity of the gesture.

c) *Qualitative Results:* Overall, participants felt that their interaction with the environment were natural and that they were in control of their movements (with an average rating of 7.2 out of 10, std. dev: 1.8). However, they needed some time to fully understand the metaphor's behaviour (with an average rating of 5.5 out of 10, std. dev: 2.2). Lastly, they managed to concentrate on assigned task rather than on the visual feedback (6.3 out of 10, std. dev: 2.2).

V. CONCLUSION

In this paper, we presented a new visual feedback for gesture guidance based on error display. To do so, we compared user's gesture in real-time to a reference and measured the gap between the two performances. Depending on the value of this gap, we used a particle system located on a sphere to represent the errors. If this value exceeded a certain threshold, particles became visible and vibrated to attract users' attention. The greater the value was, the bigger the number of visible particles was.

An experiment was conducted to evaluate whether our feedback is able to improve users' performance when performing technical gestures. Results showed that in the case of the two gestures to reproduce, participants performed better after a few iterations with our metaphor. Moreover, this study revealed that the use of this feedback presented a cognitive load which could modify the number of iterations necessary for the training, depending on gesture complexity.

In our future works, we will conduct an experiment to compare our error-based feedback with other types, such as trajectory visualization, to determine whether EBAGG has better performances. Moreover, the guidance hypothesis [18], [19] stipulates that providing too much augmented feedback during acquisition may lead to a dependency of the feedback. We will

hence study the effect of our feedback on this dependency in comparison to the other studied types of feedback.

ACKNOWLEDGMENT

This work is funded by the European Union and the French region Picardie. Europe acts in Picardie with the European Regional Development Fund (ERDF). We thank Montupet Group and Reviattech for helping us improve EBAGG.

REFERENCES

- [1] R. A. Schmidt and T. Lee, *Motor control and learning*. Human kinetics, 1988.
- [2] A. Covaci, A.-H. Olivier, and F. Multon, "Visual perspective and feedback guidance for vr free-throw training," *Computer Graphics and Applications, IEEE*, vol. 35, no. 5, pp. 55–65, 2015.
- [3] R. Moreau, V. Ochoa, M. T. Pham, P. Boulanger, T. Redarce, and O. Dupuis, "A method to evaluate skill transfer and acquisition of obstetric gestures based on the curvatures analysis of the position and the orientation," *Journal of biomedical informatics*, vol. 41, no. 6, pp. 991–1000, 2008.
- [4] M. Le Chénéchal, T. Duval, V. Gouranton, J. Royan, and B. Arnaldi, "Vishnu: Virtual immersive support for helping users-an interaction paradigm for collaborative remote guiding in mixed reality," in *3DCVE 2016: International Workshop on Collaborative Virtual Environments*. IEEE, 2016, pp. 1–5.
- [5] M. K. Holden, "Virtual environments for motor rehabilitation: review," *Cyberpsychology & behavior*, vol. 8, no. 3, pp. 187–211, 2005.
- [6] R. Sigrist, G. Rauter, R. Riener, and P. Wolf, "Augmented visual, auditory, haptic, and multimodal feedback in motor learning: a review," *Psychonomic bulletin & review*, vol. 20, no. 1, pp. 21–53, 2013.
- [7] F. J. Varela, E. Thompson, and E. Rosch, *The embodied mind: Cognitive science and human experience*. MIT press, 1992.
- [8] R. Moreau, M. T. Pham, T. Redarce, and O. Dupuis, "A new learning method for obstetric gestures using the birthsim simulator," in *Robotics and Automation, 2007 IEEE International Conference on*. IEEE, 2007, pp. 2279–2284.
- [9] F. Anderson, T. Grossman, J. Matejka, and G. Fitzmaurice, "Youmove: enhancing movement training with an augmented reality mirror," in *Proceedings of the 26th annual ACM symposium on User interface software and technology*. ACM, 2013, pp. 311–320.
- [10] R. Sodhi, H. Benko, and A. Wilson, "Lightguide: projected visualizations for hand movement guidance," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2012, pp. 179–188.
- [11] L. Marchal-Crespo, S. McHughen, S. C. Cramer, and D. J. Reinkensmeyer, "The effect of haptic guidance, aging, and initial skill level on motor learning of a steering task," *Experimental brain research*, vol. 201, no. 2, pp. 209–220, 2010.
- [12] J. Lieberman and C. Breazeal, "Tiki: Development of a wearable vibrotactile feedback suit for improved human motor learning," *IEEE Transactions on Robotics*, vol. 23, no. 5, pp. 919–926, 2007.
- [13] D. J. Berndt and J. Clifford, "Using dynamic time warping to find patterns in time series," in *KDD workshop*, vol. 10, no. 16. Seattle, WA, 1994, pp. 359–370.
- [14] E. J. Keogh and M. J. Pazzani, "Derivative dynamic time warping," in *Sdm*, vol. 1. SIAM, 2001, pp. 5–7.
- [15] G. Ten Holt, M. Reinders, and E. Hendriks, "Multi-dimensional dynamic time warping for gesture recognition," 2007.
- [16] R. Muscillo, M. Schmid, S. Conforto, and T. D'Alessio, "Early recognition of upper limb motor tasks through accelerometers: real-time implementation of a dtw-based algorithm," *Computers in biology and medicine*, vol. 41, no. 3, pp. 164–172, 2011.
- [17] B. G. Witmer and M. J. Singer, "Measuring presence in virtual environments: A presence questionnaire," *Presence: Teleoperators and virtual environments*, vol. 7, no. 3, pp. 225–240, 1998.
- [18] A. W. Salmoni, R. A. Schmidt, and C. B. Walter, "Knowledge of results and motor learning: a review and critical reappraisal," *Psychological bulletin*, vol. 95, no. 3, p. 355, 1984.
- [19] R. A. Schmidt, "Frequent augmented feedback can degrade learning: Evidence and interpretations," in *Tutorials in motor neuroscience*. Springer, 1991, pp. 59–75.