



**HAL**  
open science

# What is wrong with your gesture? An error-based assistance for gesture training in virtual environments

Florian Jeanne, Yann Soullard, Indira Thouvenin

## ► To cite this version:

Florian Jeanne, Yann Soullard, Indira Thouvenin. What is wrong with your gesture? An error-based assistance for gesture training in virtual environments. 11th Symposium on 3D User Interfaces (3DUI 2016), Mar 2016, Greenville, SC, United States. IEEE, 2016, 2016 IEEE Symposium on 3D User Interfaces (3DUI). <hal-01298592>

**HAL Id: hal-01298592**

**<https://hal.science/hal-01298592v1>**

Submitted on 9 May 2016

**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

# What is wrong with your gesture ? An error-based assistance for gesture training in virtual environments

Florian Jeanne

\*

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

Yann Soullard

†

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

Indira Thouvenin

‡

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

## ABSTRACT

Virtual environments are commonly used to train people, especially for gesture training. A major challenge consists in error perception. In this paper, we propose to guide the user gesture with an error-based feedback in a 3D virtual environment. In a preliminary experiment, trainees reproduce an expert gesture from an aluminium foundry. Their gesture performance is evaluated according to two critical error criteria to improve the interaction for training.

**Keywords:** 3D interaction, Immersion, Training, Gesture

## 1 INTRODUCTION

In an industrial context, new technicians are generally trained by an expert, who first presents the tasks and gestures of the technical process and then assists them in the first stages of training. Such training is performed on a real workstation in the factory and may lead to safety problems due to trainees lack of expertise. Besides, they may cause some damage on apparatus, leading to important costs for the company.

On the last few years, gesture training systems using virtual environments have been proposed for specific applications such as medical science [6, 7], non-verbal communications [8] or technical gestures [3]. However, most of systems do not provide any kind of feedback regarding errors made by a trainee.

To improve industrial gesture training, we propose a training system based on learning through error with a virtual environment. The gesture error is measured using a machine learning method. Then, sensory feedback are generated in a 3D virtual environment. The machine learning method is a Dynamic Time Warping algorithm (DTW) [1] which enables to evaluate the matching between a trainee gesture and an expert one on error criteria. The DTW computes a distance for each error criterion (e.g. speed, orientation): the more important the distance is, the worst the error is. Each error criterion refers to a particular 3D visual feedback.

We illustrate our approach with a specific technical gesture from an assembly line in an aluminium foundry. Trainees have to reproduce an expert gesture for mold cleaning with a compressed-air blower. We observe two types of errors: gestures speeds and blower inclination errors. Gestures are tracked using infrared cameras detecting marker positions and orientations.

## 2 TRAINING SYSTEM IN VIRTUAL ENVIRONMENT

We identify the most important error to be corrected and then provide an appropriate guidance to improve trainee gestures.

\*e-mail: florian.jeanne@hds.utc.fr

†e-mail: yann.soullard@hds.utc.fr

‡e-mail: indira.thouvenin@hds.utc.fr

## 2.1 Gesture analysis based on error recognition

The motion tracking system we use extracts marker positions and orientations. However, these features may not be adequate to recognize different types of errors. Thus, we compute higher level features referring to spatio-temporal descriptors and allowing a better recognition of the different error classes.

Working with technical gestures leads to various constraints limiting the choice of a model. Indeed, one generally has to deal with only a few number of gesture examples since they must be manually labeled by an expert, which is costly. Besides, one wants to get interpretable outputs in order to provide relevant feedback to trainees. Based on these comments, a DTW algorithm [1] seems to be an appropriate choice for such a task. Indeed, a DTW may be efficient on small training samples, in particular with one gesture template. In addition, it provides interpretable results in the form of a distance between two time series. Besides, unlike many learning methods, a DTW directly works with labeled examples and it does not model the data distribution (which creates a bias).

A DTW is defined to find an optimal alignment between two time series. It computes a distance between the two series based on the optimal alignment. Let  $\mathbf{x} = (x_1, \dots, x_N)$  and  $\mathbf{q} = (q_1, \dots, q_M)$  be two time series. The DTW algorithm consists in computing a cumulative cost  $C(n, m)$  such that  $C(n, 1) = \sum_{i=1}^n d(x_i, q_1)$ ,  $C(1, m) = \sum_{i=1}^m d(x_1, q_i)$  and for  $1 < n \leq N$  and  $1 < m \leq M$  the cost is computed using the following equation:

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

where  $d(x_n, q_m)$  measures the distance between the  $n^{\text{th}}$  value of  $\mathbf{x}$  and the  $m^{\text{th}}$  value of  $\mathbf{q}$ . Initial and final conditions are defined to ensure that the process begins and ends with the alignment of the first and last elements of  $\mathbf{q}$  and  $\mathbf{x}$ .

For each error criterion, we have a unidimensional time series referring to the feature values computed from the marker positions and orientations at each frame. We apply one DTW per error criterion (speed and inclination): each of them evaluates the matching between a trainee gesture and an expert gesture and provides a distance which informs on the gap between the two gestures for the given type of error. These measures are then provided to the assistance module in order to select an appropriate guidance.

## 2.2 Error-based assistance

The guidance can be multimodal and be related to different nature of feedback such as visual [6], auditory [2] or haptic [5]. Besides, Galligan et al. [4] define different types of feedback: there are intrinsic feedback from the kinaesthetic senses, extrinsic feedback such as performance or results, positive and negative feedback to inform users on the performance level, terminal feedback which are provided before or after the performance, and finally concurrent feedback which are given during the performance. The authors point out that the type of feedback may depend on trainees themselves or on gestures they have to learn, e.g. some gestures require real-time feedback whereas others require an a posteriori feedback.

At the end of a trainee performance, the higher distance provided by the error recognition module refers to the most critical error. Thus, the guidance referring to this most critical error is given to the user. We provide extrinsic feedback enabling users to visualize their performance. With such feedback, we expect trainees to improve their gestures by observing their performance. When the speed is the most critical error, we display the gesture trajectory using spheres whereas we use capsules for an inclination error (Figure 1). Capsules are orientated in the same direction as the tracked marker during the gesture. A color coding informs trainees on the differences between their gesture and the reference one. This is determined according to the optimal alignment provided by the DTW: sky blue and dark blue refer respectively to a slightly and much smaller value than the expert gesture value, green is for a close value whereas orange and red refer respectively to a slightly and much higher value.

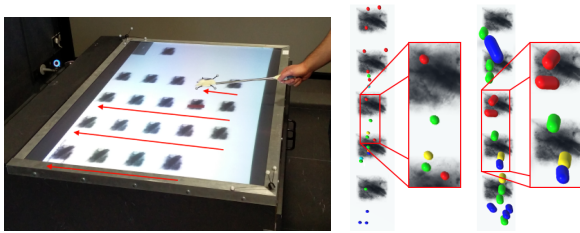


Figure 1: On the left: Illustration of a gesture being performed (red arrows show the directions). On the right: two types of feedback provided a posteriori (the speed criteria represented by balls and the inclination one by capsules)

### 3 PRELIMINARY EXPERIMENT

We applied our approach to the context of aluminium foundry. Within this context, technicians perform various technical gestures. Also, as a first step, we selected one of these gesture which consist of cleaning a mold using a compressed-air blower.

#### 3.1 Experimental design

The task consists in cleaning 4 lines of 5 dirt, as illustrated in Figure 1). An expert gesture is beforehand recorded. It consists in cleaning the mold using the blower from the front to the back and from the left to the right with a constant speed when the blower moves from the front to the back of the mold. It has to be performed at a proper speed, otherwise technicians would waste time if they are too slow or the mold would not be correctly cleaned if they are too quick. Finally, the blower has to be lifted up when the technician reaches the back of the mold.

We evaluate users capability to reproduce the technical gesture performed by the expert. Thus, we use a multidimensional DTW which measures the matching between two multidimensional time series. Here gestures are 2-dimensional sequential data where the two dimensions refer to the two error criteria (the speed and the inclination of the blower).

#### 3.2 Experimental procedure

The expert begins the process by describing the gesture to the trainee. Then users have to reproduce it 10 times. Between each training iteration, visual feedback (one of the two illustrated in Figure 1) is provided to the trainee according to the error criterion having the higher distance with the expert gesture (i.e. speed or inclination).

#### 3.3 Participants and platform

15 users contributed to our experiment, 12 men and 3 women from 18 to 53. All the participants were novices with our training system.

We used 6 infrared cameras: 2 OptiTrack Prime 13W and 4 OptiTrack Prime 13 (optitrack.com), both with a frequency of 240 Hz ensuring a low latency. A stereoscopic display of a mold with dirt were done on a horizontally laid screen. A compressed-air blower enables to clean the mold remotely. Markers were fixed at the end of the prop and on the head and they were tracked using the system.

## 4 RESULTS AND DISCUSSION

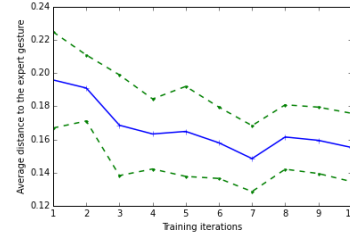


Figure 2: Average gap (distance computed by the DTW) of user performances with the expert one during training iterations.

User performance generally improves during the training process. A Wilcoxon test between the first and the last training iteration shows a significant improvement (the p-value is equal to 0.038 which is significant to a standard 0.05 threshold). In addition, the standard deviation slightly decreases. This illustrates that the gap between user performance decreases during the training. Finally, we point out that user performances seem to remain constant on the last iterations. This may be due to a loss of concentration after some training iterations.

To conclude, those first results seem promising and show that trainee gestures get closer to the expert gesture as expected. In our future works, we will experiment our system on more complex gestures and compared it to other automatic strategy of feedback. Besides, we will focus our research on adaptive dynamic guidances to provide customized assistances.

### ACKNOWLEDGEMENTS

This work is funded by the European Union and the French region Picardie. Europe acts in Picardie with the European Regional Development Fund (ERDF).

### REFERENCES

- [1] D. J. Berndt and J. Clifford. Using dynamic time warping to find patterns in time series. In *KDD Workshop*, pages 359–370, 1994.
- [2] E. Boyer, B. Caramiaux, S. Hanneton, A. Roby-Brami, O. Houix, P. Susini, and F. Bevilacqua. Legos project - state of the art. Technical report, 2014.
- [3] L. Da Dalto, F. Benus, and O. Balet. The use and benefits of virtual reality tools for the welding training. In *63rd Annual Assembly & International Conference of the International Institute of Welding, Istanbul, Turkey*, 2010.
- [4] F. Galligan. *Advanced PE for Edexcel*. Heinemann, 2000.
- [5] R. W. Lindeman, Y. Yanagida, K. Hosaka, and S. Abe. The tactapack: A wireless sensor/actuator package for physical therapy applications. In *HAPTICS*, page 51. IEEE Computer Society, 2006.
- [6] 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*, 41(6):991–1000, 2008.
- [7] C. Papadopoulos, H. Choi, J. Sinha, K. Yun, A. E. Kaufman, D. Samaras, and B. Laha. Practical chirocentric 3d platform for immersive environments. In *2015 IEEE Symposium on 3D User Interfaces, 3DUI 2015, Arles, France, March 23-24, 2015*, pages 31–34, 2015.
- [8] B. Ravenet, M. Ochs, and C. Pelachaud. From a user-created corpus of virtual agent non-verbal behavior to a computational model of interpersonal attitudes. In *Intelligent Virtual Agents*, pages 263–274, 2013.