Capture, modeling and recognition of expert technical gestures in wheel-throwing art of pottery

Sotiris Manitsaris, Alina Glushkova, Frédéric Bevilacqua, Fabien Moutarde

▶ **To cite this version:**

Sotiris Manitsaris, Alina Glushkova, Frédéric Bevilacqua, Fabien Moutarde. Capture, modeling and recognition of expert technical gestures in wheel-throwing art of pottery. Journal on Computing and Cultural Heritage, 2014, 7 (2). hal-00975857v2

HAL Id: hal-00975857

https://minesparis-psl.hal.science/hal-00975857v2

Submitted on 29 Aug 2014

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.

Capture, modeling and recognition of expert technical gestures in wheel-throwing art of pottery.

S. MANITSARIS[1, 2, 3],
A. GLUSHKOVA[1, 2, 4],
F. BEVILACQUA[3],
AND
F. MOUTARDE[2]
University of Thessaly[1] – Greece, MINES ParisTech[2] - France,
IRCAM[3] - France, University of Macedonia[4] - Greece

This research has been conducted in the context of the ArtiMuse project that aims at the modeling and renewal of rare gestural knowledge and skills involved in the traditional craftsmanship and more precisely in the art of the wheel-throwing pottery. These knowledge and skills constitute the Intangible Cultural Heritage and refer to the fruit of diverse expertise founded and propagated over the centuries thanks to the ingeniousness of the gesture and the creativity of the human spirit. Nowadays, this expertise is very often threatened with disappearance because of the difficulty to resist to globalization and the fact that most of those "expertise holders" are not easily accessible due to geographical or other constraints. In this paper, a methodological framework for capturing and modeling gestural knowledge and skills in wheel-throwing pottery is proposed. It is based on capturing gestures using wireless inertial sensors and statistical modeling. In particular, we used a system that allows for online alignment of gestures using a modified Hidden Markov Model. This methodology is implemented into a Human-Computer Interface, which permits both the modeling and recognition of expert technical gestures. This system could be used to assist in the learning of these gestures by giving continuous feedback in real-time by measuring the difference between expert and learner gestures. The system has been tested and evaluated on different potters with a rare expertise, which is strongly related to their local identity.

General Terms: Technical gestures, Know-how, Modeling, Recognition, Wheel-throwing

Additional Key Words and Phrases: Perception, HCI, Inertial sensors, Machine-Learning

1. INTRODUCTION

Cultural expression is not limited in architecture, monuments, collection of objects, or music. It also includes fragile intangible live expressions, which involve knowledge and skills. Such characteristics can be human gestures, the color of our voice or facial expressions. They are controlled by the intelligence of the human creativeness depicted in music, dance, singing, theatre, human skills and handicraft. This kind of culture has been termed Intangible Cultural Heritage (ICH), which is manifested inter alia in oral traditions and expressions, performing arts (*music, dance, theatre etc.*) social practices, knowledge and practices concerning nature, universe and the traditional craftsmanship [Unesco 2003]. In traditional craftsmanship, the rare gestural knowledge primarily consists of hand and finger motions.

Authors' addresses: [1, 2, 3]S. Manitsaris E-mail: sotiris.manitsaris@uom.edu.gr. [1, 2, 4]A. Glushkova, E-mail: alina.glushkova@uom.edu.gr. [3]F. Bevilacqua, E-mail: frederic.bevilacqua@ircam.fr. [2]F. Moutarde, E-mail: fabien.moutarde@mines-paristech.fr. [1]Rural Space Lab, Department of Planning and Regional Development, University of Thessaly, Volos, Greece. [2]Robotics Lab, MINES ParisTech, Paris, France. [3]Sciences and

Technology for Music and Sound Lab, Ircam-CNRS-Université Pierre et Marie Curie,Paris, France. [4]Multimedia Technology and Computer Graphics Lab, Department of Applied Informatics, University of Macedonia, Thessaloniki, Greece.
Permission to make digital/hard copy of part of this work for personal or classroom use is granted without fee provided that the copies are not made or distributed for profit or commercial advantage, the copyright notice, the title of the publication, and its date of appear, and notice is given that copying is by permission of the ACM, Inc. To copy otherwise, to republish, to post on servers, or to redistribute to lists, requires prior specific permission and/or a fee. Permission may be requested from the Publications Dept., ACM, Inc., 2 Penn Plaza, New York, NY 11201-0701, USA, fax: +1 (212) 869-0481, permission@acm.org
© 2001 ACM 1530-0226/07/0900-ART9 $5.00 DOI 10.1145/1290002.1290003 http://doi.acm.org/10.1145/1290002.1290003

The project "ArtiMuse" aims at proposing a multidisciplinary research approach for the gesture recognition methodologies applied in musical and handicraft interactions. In order to preserve these gestural skills that require high expertise it is necessary to identify, record, analyze, model and recognize them. This paper presents the development of a methodology for the modeling of kinematic aspects of technical gestures using gesture recognition technologies based on wireless inertial sensors. This methodology has been implemented, applied and evaluated on the gestural skills of two expert potters for simple objects, such as bowls. The Human-Computer Interface called "ArtOrasis" has been developed for gesture capturing, modeling and recognition of expert gestures. It also proposes the time alignment of a gesture compared to a model gesture, which is a quite promising perspective for a performance comparison between expert and learner.

2. STATE OF THE ART

The study of human gestures has been of special interest in different research fields. In the ICH domain in particular, body and hand gestures are important means of communication, of expression and of creativity. Preservation and transmission of handicraft skills is often done by studying and analyzing recordings, verbal descriptions and documents. However, the important role of gestures has lead several researchers to model them using motion capture and gesture recognition technologies.

2.1 E-DOCUMENTATION AND DIGITAL ETHNOGRAPHY

Most existing methods for skill preservation are based on verbal descriptions of movements, often in conjunction with multimedia content such as graphic / photographic material [D. Chevallier, 1991]. Another widespread method is the video recording of skills and techniques of the gesture accompanied with verbal commenting. This method was applied to preserve knowledge of technical gestures in power stations in France during the manipulation of different technical tools. A video camera has been placed on the helmet of the workers to record their movements [Le Bellu 2012].

In the field of ICH, for many years, ethnologists have studied the characteristics (*arts and techniques, oral traditions and living expressions*) of groups and communities in their surroundings. [D. Chevallier, 1991]. Enora Gandon worked on wheel-throwing gestures in a cultural context. She studied the impact of the cultural background of the human on the development of motor skills in Wheel-throwing art [E. Gandon, 2011]. Kuo-An Wang studied the case of weaving Chinese traditional items with Bamboo [Kuo-An Wang et. al, 2011]. He created a digital archive with approximately1200 objects accompanied by images and videos presenting the gestures involved in the creation of the

objects. Through meetings and interviews conducted with the craftsman, Kuo-An Wang has identified a set of 20 basic gestural patterns of weaving. Then, he connected each of the digitized objects file with a combination of those patterns.

However some significant limitations can be identified in these methods. While describing a gesture on a piece of paper using photos or figures, the gesture is limited in two dimensions and it does not represent any realistic information about how the gesture has been performed. In the case of video recording, the gesture is also represented in two dimensions but still limits the information that can be extracted.

2.2 MOTION CAPTURE AND GESTURE RECOGNITION

The use of innovative technologies for motion capture permit to overcome some of the limitations mentioned above, to achieve a faithful record of the gesture and model it stochastically. A significant number of studies have been based on various techniques for modeling and recognition of gestures based on motion capture. These technologies can be subdivided in 3 categories: a) marker-based, b) marker-less and c) inertial motion sensors.

Marker-based approaches use optical-markers and active computer vision, which require expansive commercial systems, such as Vicon Peak or Optitrack. This type of sensor has been used for the modeling of music performances of a violin player [Rasamimanana N. et al. 2009, Demoucron M. et al., 1994]. One of the most important limitations of the marker-based gesture recognition systems is that they are not robust to occlusions.

Marker-less technologies do not require subjects to wear special equipment for tracking and are usually based on passive computer vision approaches. For example, Microsoft Kinect is a low-cost depth camera that provides good results for the recognition of global body postures, such as dance gestures [Raptis, 2011]. It provides Cartesian representation of the human motion and it is usually used for tracking joints of the body. Nevertheless, it is less precise for hand gestures. To solve this problem there is an ongoing research aiming at the creation of a hand skeletal model for finger detection. A hand skeletal model for depth images, provided by the PMD CamBoard Nano time of flight camera, has been applied to capture music-like finger gestures [Dapogny et al., 2013] based on Shotton's algorithm and pixel wise classification through Random Decision Forest. This model is currently being developed and adapted for pottery-like finger gestures, as a training database is required. An elaborated algorithm is also necessary for scene and object segmentation in the case of technical gesture recognition in wheel-throwing art of pottery. Moreover, the depth cameras are self and scene occlusion-dependent and the development of hand/finger skeletal model for the capturing of finger gestures in pottery is a challenging approach in a medium term vision.

Inertial Motion Sensors [R. Aylward et al., 2006, T. Coduys et al., 2004, T. Todoroff, 2011, E. Fléty et al. 2011] or commercial interfaces, such as the Wii joystick [D. Grunberg, 2008], permit to track gesture features continuously and in real-time. These sensors have been tested and used in dance and music performances [Bevilacqua et al., 2007, 2010]. For example, inertial sensors have been used for motion capture aiming at the archaeological reconstruction, understanding and interpretation of different possibilities of use of an ancient Iron Age roundhouse [S. Dunn et al, 2011].

2.3 PRESSURE MEASUREMENT AND MECHANICAL STRESS

In case of wheel throwing pottery, a study has been conducted to evaluate potters' skills by taking into consideration the mechanical characteristics of the objects created by the potters. The Von Mises stress index [J. Lemaitre et al., 2004] has been used to measure the mechanical stress operating in the object and it has been related to the throwing difficulty, proposing thus an idea for potters' skills assessment [E. Gandon et al., 2011]. In this object-oriented approach, the pressure is measured by analyzing the mechanical characteristics of the created object after completion, and not in real-time during the creation of the vessel. Therefore, despite its interest, this mechanical stress estimation cannot easily be used to build an interactive pedagogical tool, because interactivity requires real-time analysis of the gesture performed. Since designing this type of tool is one of our final goals, our approach for the preservation and modeling of gestural skills is based on the analysis of the kinematic aspects of the gesture required for the creation of the object, which can be easily captured and analyzed in real-time.

3. METHODOLOGY

In the methodology proposed below, the goal is not a simple video recording with verbal descriptions, nor just the digitalization of information. The objective is to study and to model rare gestural know-how involved in handicrafts, to gather data about different biomechanical, kinematic aspects of a technical gesture (*distances between the hands, angles of the vertebral axis, rotations of the joints, gesture's trajectory etc.*), as well as the body postures and to create information about its various parameters. The modeling of the gestural know-how and the effective recognition of gestures have been done in different methodological phases as described below and represented in the figure 1.

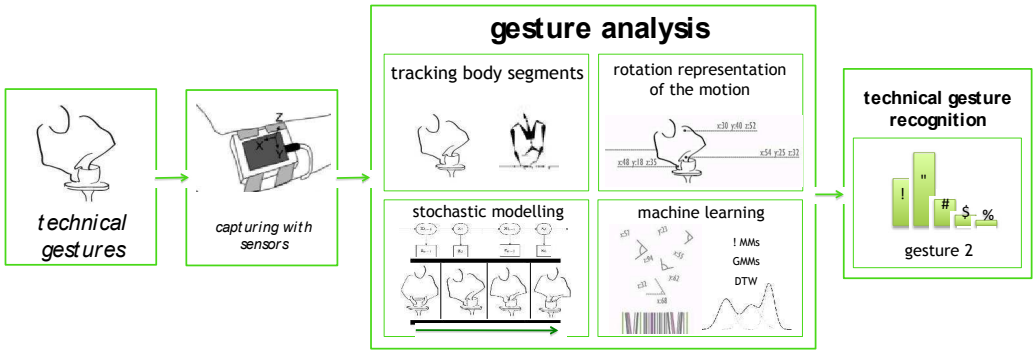


Fig. 1. Gesture recognition pipeline based on the Animazoo suit of wireless motion sensors for the upper-part of the potter's body.

Capturing, modeling, online recognition and time alignment, are fundamental stages and extensions towards the creation of an interactive pedagogical tool for the transmission of gestural skills based on sensorimotor learning. This comparison could be done in real time in order to provide a sonic or optical feedback to the learner, and thus drive him/her to correct his/her gestural errors.

3.1 HANDICRAFT AND EXPERT SELECTION

The first step is to select the type of handicraft and the craftsman to be studied. The handicraft selected should respect some criteria and be compatible with the goal of the research and with its' technical conditions and constraints. Since we are using gesture capture technologies, one of the important criteria for technical feasibility and application of the methodology is to avoid use of specific tools during gesture execution.

3.2 IDENTIFICATION OF THE EFFECTIVE GESTURES

The second phase concerns the identification of effective gestures. Effective gestures have a direct impact on the material and it is important to identify them and to distinguish them from other auxiliary gestures. This goal can be achieved in interaction and collaboration with the expert, by conducting interviews and observing him while working. He should show and describe a complete sequence of gestures, effective and auxiliary, in order to define a dictionary $GD = \{G_i\}_{i\in\mathbb{N}}$ of his/her effective gestures.

3.3 GESTURE CAPTURING

In the context of expert technical gestures, a possible simplification of the complexity of the human body can be based on rotation of segments, which is the case of the data stream (*observation vector*) provided by the sensors. Therefore, the expert's body segments are tracked and the sensor captures kinematic properties about angle rotations in 3D space and record them as a sequence of observation vectors $Y_{0:k} = \{y_1 \dots y_k\}_{k\in\mathbb{N}}$, where $y_{j_{j\in\mathbb{N}}}$ is one observation vector and k the total number of observations. The vector $y_j = [y_1^j \dots y_n^j]_{j,n\in\mathbb{N}}$ represents the n kinematic descriptors for a given time stamp j.

3.4 GESTURE MODELING AND MACHINE-LEARNING

3.4.1 GESTURE REPRESENTATION

Once the data acquisition is completed, then $Y_{0:k}$ should be normalized. Euler angles, Quaternions are possible rotation representations of the motion. Euler rotation is a rotation about a single Cartesian axis. According to the Euler's Rotation Theorem, every orientation can be described as a rotation from some other reference orientation as a sequence of three elemental rotations (*precession, nutation, and intrinsic rotation*). Quaternions are representing orientations and rotations of objects in 4D, where there is one real axis and three imaginary axes (*i, j, and k*). Another way to model the rotations of the motion of the human body is the Direction Cosine Matrix (DCM), which is based on a triad of unit vectors. The rotation is described by specifying the coordinates of the triad of unit vectors in its current position, based on a non-rotated coordinate axes that is used as a reference.

3.4.2 GESTURE MODELING AND ONLINE CHARACTERISATION

One of the difficulties in gesture recognition is that the same gesture can be performed in a variety of ways, in particular the change in speed of execution. For this reason, one of the authors has developed a system called "Gesture Follower" that can be seen as a

hybrid approach between Dynamic Time Warping (DTW) and Hidden Markov Models (HMM) [Bevilacqua 2007, 2010]. This system is a template-based method, which allows the use of a single gesture to define a gesture class. This requirement is necessary due to the limited access of gesture data in our application. Nevertheless, we use a HMM formalism to compute in real-time computation measures between the template and the incoming data flow.

As described by Rabiner (1989), Hidden Markov Model can be used to model recorded time series (*training procedure*) and to compute the likelihoods (*one per HMM*) that the hidden state sequence $X_{0:k} = \{x_1 \dots x_k\}_{k\in\mathbb{N}}$, generated the new observation sequence $Y_{0:k} = \{y_1 \dots y_k\}_{k\in\mathbb{N}}$. In our case, we are interested in the following information:

- The likelihood $\psi(x, y)$ that the observation sequence was produced by the different models. The sequence with the maximum likelihood ψ_{max} to generate $Y_{0:k}$ indicates the gesture G_i from the gesture dictionary $GD = \{G_i\}_{i\in\mathbb{N}}$.
- The likeliest state sequence $X_{0:k}$ of the observation sequence $Y_{0:k}$. This state-sequence allows for obtaining a time-warping between the model and the observed sequence.

In order to greatly simplify the learning procedure and to guarantee a high temporal precision in the gesture modeling, we associate each template to a state sequence. We define one state for each sample data of the template (*applying a constant sampling rate*). We compute the different likelihoods in real time using the well-known for-ward procedure that return the results incrementally, and that can be implemented efficiently even in the case of models with a large number of states [Bevilacqua 2010].

4. CASE STUDY

4.1 SELECTION OF THE WHEEL-THROWING POTTERY AND OF THE POTTERS

For the implementation of this methodology the wheel-throwing pottery has been chosen mainly for two reasons. First of all, because of the high social and cultural value of this traditional profession for the local communities of the Macedonia Region in northern Greece (*community of the potter A*) and of the French Riviera (*Côte d'Azur*) in Southern France, (*community of the potter B*) which is also famous for the large number of ceramists working there. An important number of associations and independent experts are actively promoting this handicraft in these regions. A list of candidate craftsmen has been established. Two experts from the regions above have been selected. They detain a very high level of expertise in wheel-throwing pottery art and also important pedagogical experience. The potter A is teaching this handicraft in a centre for therapy and social reintegration of people with substance dependencies and the potter B is a senior craftsman with more than twenty years of experience in practicing wheel-throwing pottery. The second criterion is more technical and linked to our system's technical characteristics. The wheel-throwing pottery is based on gesture control of the material. There should be no interference of the hands of the potter and his material with specific tools or other intermediate mechanisms.

4.2 BASIC GESTURES FOR THE CREATION OF A BOWL

The two selected expert potters presented to us a complete sequence of gestures that are used to create a bowl with a simple shape (*a bowl*) with different quantities of clay.

It has been asked to the potter A to create 5 bowls of 18-20 cm of diameter, 10 cm of height, with approximately 1.3 kg of clay. Additionally, it has been asked from the potter B to create bigger bowls of the same shape of those created by potter A, with 20-23 cm of diameter and 13 cm of height with 1.75 kg of clay.

After meticulous observation of video recordings of the gestures of the two potters and after the interviews conducted with them, we have concluded on the following 4 basic gestural phases of creating a simple bowl (Figure 3).

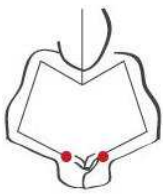
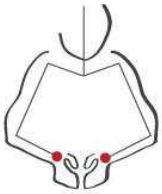
						
4 basic gestural phases	**P_1 Centering and bottom opening**		**P_2 The raise**	**P_3 The first configuration**	**P_4 The final configuration and removing**	
Potter A 4 gestures	G_1^A Centering and bottom opening		G_2^A The raise	G_3^A The first configuration	G_4^A The final configuration and removing	
Potter B 6 gestures	G_1^B Centering the clay	G_2^B Opening the bottom	G_3^B The raise	G_4^B The first configuration	G_5^B The final configuration	G_6^B Removing the object

Fig 3. Basic phases and gestures per potter for the creation of a bowl

Since the dimensions of the bowls created by the two experts are different, the required gestures inside the gestural phases are also different. Obviously, there is a common track between the gestures of the two experts, even if the dimensions of their objects are slightly different. The bigger an object is, the more gestural work it requires. Consequently, since the object of the potter A is smaller, the 4 basic gestures that we have identified correspond exactly to the 4 basic gestural phases presented above. For the bigger object, the potter B has more clay to manage and he is paying more attention to shape refining.

They are also considered as representative of the high-level wheel-throwing pottery skills, since the creation of the object takes in case of the potter A only 60-75 sec and for potter B 120-140 sec. The fluidity and the speed of potters' gestures as well as the hand coordination are elements that constitute the basis of the rare know-how and gestural skills.

All gestures have duration of 15-25 seconds. The centering and bottom opening P_1 consists of fixing of the clay on the wheel, hands are pressing steadily on the material aiming at the opening of the bottom. Then, the potter's hands are picking up the clay, defining the height of the bowl through the second gesture, P_2 the raise of the clay.
Then the body posture is changing, slightly turning on the right or on the left side for the first configuration of the shape, for P_3. Precise finger gestures are specifying the basic form of the object. The fingers of the one hand are fixing the clay and of the other are forming the object. His hands are too close to each other, touching the inner and outer sides of the clay respectively. After this stage, the potter is making the final configuration of the shape P_4. His fingers are controlling and equalizing the bowl thickness and at the end the potter passes a very fine wire between the bowl and the wheel in order to take the bowl.

4.3 EXPERIMENTATIONS: CAPTURING THE POTTER'S GESTURES

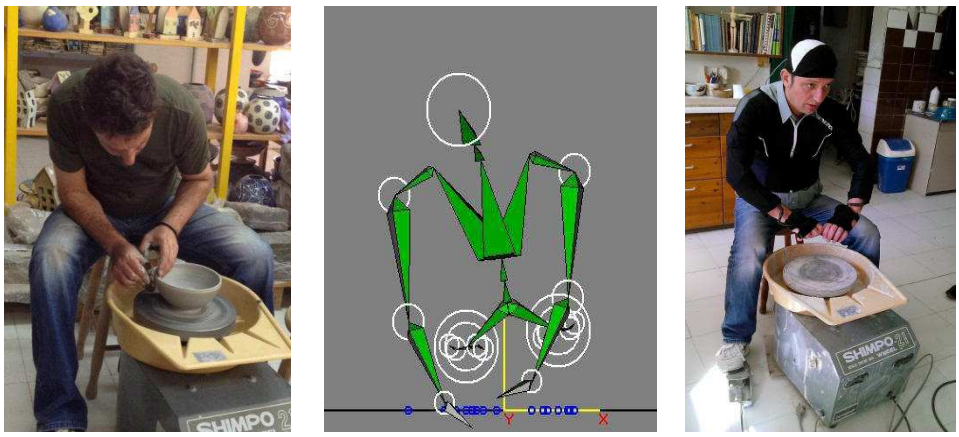
After the definition of the above effective gesture the potters are asked to put the inertial motion capture suit that can easily provide real-time access to motion information and permits the data acquisition (Figure 4).

This suit contains 11 inertial sensors (*gyroscopes and magnetometers*) and it is covering for the upper part of potter's body, his wrists, his neck and his head. It provides an automatic filtering for the correction of magnetic disruption. It has been selected for the gesture capturing and the implementation of the methodology described below. It is occlusion-independent and it provides a high precision rotational representation of body segments.

The 11 sensors are integrated in the suit and after the calibration they provide and capture information related to the XYZ axis rotations with the use of integrated gyroscopes, accelerometers and magnetometers. The posture of the upper-part of the potter's body can be derived by the data obtained from the suit but it is not the case with his position in 3D space. These data are recorded following a hierarchical structure and more precisely the Bounding Volume Hierarchy (BVH).

Since magnetometers are used among other sensors in the suit with the inertial sensors, the quality of data captured can be influenced by magnetic disturbances. During the first day of data acquisition with the potter B these disturbances were very strong since he was using an old model of wheel, containing many metallic devices. Despite the fact that data are online corrected by the system if weak magnetic disturbances are identified, the data acquired at the first day were of a very bad quality. For this reason another data acquisition session has been realized with the use of a more modern wheel with less magnetic disturbances.

The following table I, lists the different parts of potter's body, which motions have been captured for the preservation of his gestural know-how. Some of them may play a more important role in the technical gesture depending on the type of handicraft, but all the following body articulations are involved in the performance of the gesture of the craftsman. We are also aware about the important role of fingers in wheel-throwing process. Finger tracking constitutes an important step and we are currently working on the creation of a skeletal model trained on the potter's finger gestures.



(a) (b) (c)

Fig 4. (a) The potter A, Theodoros Galigalidis, pottery teacher at the "Therapy Center for Depending Individuals" in Thessaloniki, is performing the G_3 (*first configuration*). (b) Skeleton reconstruction. (c) The potter is wearing the suit with the inertial sensors for preliminary analysis.

Table I. Body segments and gesture descriptors for rotations in 3D space

Body segments	Gesture descriptors (y_n^k)
- Spine - Right Shoulder - Right Arm - Right Forearm - Right Palm - Left Shoulder - Left Arm - Left Forearm - Left Palm - Hips - Head	- Direction Cosine Matrix, y_{99}^k - Euler angles, y_{33}^k - Quaternions, y_{44}^k

4.4 POTTER'S GESTURE MODELLING WITH THE ARTORASIS SYSTEM

After the gesture capturing using the inertial sensors, the data is normalized in [-1, 1] using Euler angles, Quaternions and Directors Cosine Matrix as described in the methodology.

A prerequisite for the creation and the application of our methodology was to design the ArtOrasis system and interface (Figure 5). This gesture recognition system is entirely implemented in MaxMSP, an interactive programming environment that uses the Jitter toolbox and it aims at the recognition of technical gestures. ArtOrasis can also be used for capturing, modeling, and recognition. It also provides functionalities for the visualization of the skeleton of the craftsman.

The machine learning engine of the ArtOrasis is based on a hybrid Hidden Markov Model and Dynamic Time Warping approach, which is implemented into the Gesture Follower (GF) [Bevilacqua et Al. 2007] patch for MaxMSP (developed by the IMTR research team of IRCAM).

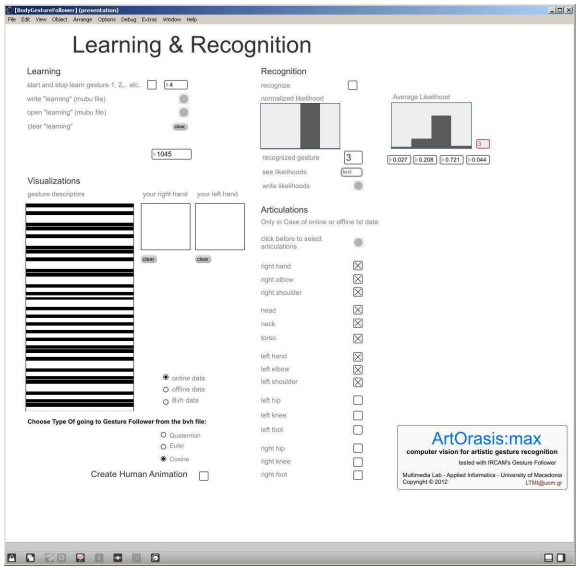


Fig 5. Screenshot of the interface of the ArtOrasis system presenting the learning and recognition phases

In case of wheel throwing pottery 11 segments of the human skeleton listed in the table below have been selected and used for the training of ArtOrasis system. Concerning the different gestures separation, the training has been based on the 4 effective gestures identified during the second stage.

According to the model defined previously the user (*researcher, potter, learner*) of ArtOrasis can define and choose which are the most important parts of the body that participate in the execution of a technical gesture and train the system based on ones. This stage corresponds to the machine learning phase of the methodology. The training of the gesture recognition system is also based on the effective gesture separation defined in second methodological step. After the training of the system the last step is the gesture recognition that is evaluated below.

5. EVALUATION

One of the final goals of our research is the design of a real-time pedagogical tool that can help transmission, and thus preservation of gestural know-how. To attain this goal, we need to compare gesture realization by an apprentice with the recorded and modeled gestures by experts. A pre-requisite before estimating such similarity, is the automated recognition by the system of what particular step the apprentice is trying to perform. For this reason we are convinced that online technical gesture recognition is essential for the comparison of handicraft skills between apprentices and expert. Furthermore, the segmentation of the data captured into a set of specific gestures, and the training of models, provides the data with a semantic dimension.

In order to validate our approach, and evaluate the recognition accuracy of the system for all the P_i, it has been asked to each of the expert potters to create five bowls. All the gestures G_i^A from the potter A and G_i^B from the potter B that are involved in all the four

phases P_i have been recorded in real conditions (*co-articulated gestures and without rest*). It has to be mentioned that very often, expert craftsmen are not available to create many copies of exactly the same object since this procedure is considered as a creative art process or because of ageing. Nevertheless, in case of the potter A the repeatability of his gestures can be considered as being of a high level, since he was very concentrated and careful in the way he performed the gestures. In case of the potter B the repeatability is of a medium level since he is easily disturbed in his everyday work by external elements (neighbors visiting his atelier, etc.)

The gesture recognition rates have been evaluated based on the « jackknife » method [Abdi, Williams 2010]. In our case, jackknifing means estimation of the recognition accuracies for manually segmented gestures (*isolated gestures*) by using subsets of the available gestural data. The basic idea behind the jackknife variance estimator lies in systematically recomputing the statistic estimate leaving out one or more observations at a time from the sample set.

Practically, a dataset contains observations of all the G_i^A and G_i^B. In total, five observations for each gesture have been recorded and distinct databases for learning and test have been used in five iterations. For each iteration, one dataset is left out to be used as the learning database and train one model M_i per gesture G_i until all the data sets are used once and the four remaining datasets are used as a database for testing. Two metrics have been used to evaluate the system:

$$Precision = \frac{\#\ of\ True_Recognitions}{\#\ of\ True_Recognitions\ +\#\ of\ False_Recognitions} \quad (1)$$

$$Recall = \frac{\#\ of\ True_Recognitions}{\#\ of\ True_Recognitions\ +\#\ of\ Missed_Recognitions} \quad (2)$$

So, for the potter A, the first evaluation phase has been done using Euler angles after normalization for training the HMMs M_i^A. For each of the eleven body segments, one Euler angle per axis has been computed. The table II shows the results of the five iterations of the jackknifing as well as the Precision and Recall per gesture of the potter A. Twelve queries for recognition per G_i^A have been asked to the M_i^A. Both Precision and Recall were at 100%.

Table II. Precision and Recall per gesture from the potter A based on five iterations of jackknifing using Euler angles.

		Maximum likelihoods (ψ_{max})				
		M_1^A	M_2^A	M_3^A	M_4^A	**Recall**
Observations ($Y_{0:k}^A$)	G_1^A	**20**	-	-	-	100%
	G_2^A	-	**20**	-	-	100%
	G_3^A	-	-	**20**	-	100%
	G_4^A	-	-	-	**20**	100%
Precision		100%	100%	100%	100%	

The Quaternions and the Direction Cosine Matrix have also been used for the training of the M_i models. For both Quaternions and Direction Cosine Matrix, all the observations $Y_{0:k}$ from G_2^A and G_3^A that have been given as query to ArtOrasis gave true recognized (Recall) but there are some cases where M_2^A and M_3^A gave maximum likelihood for false recognitions (Precision). In table III we can see that all three ways of the motion representation give excellent results for the recognition of all the effective gestures.

This first experimental case shows that, at least for the creation of simple objects in wheel throwing-art, online gesture recognition based on machine-learning can be successfully applied, and can therefore be used as a first step for "capturing of gestural skills" related to pottery.

Table III. Comparative table for Precision and Recall of Euler, Quaternions and Direction Cosine Matrix (DCM) representations

	Euler		**Quaternions**		**DCM**	
	Precision	Recall	Precision	Recall	Precision	Recall
G_1^A	100%	100%	80%	94%	100%	85%
G_2^A	100%	100%	100%	90%	87%	100%
G_3^A	100%	100%	100%	87%	95%	100%
G_4^A	100%	100%	85%	100%	100%	95%
Total	**100%**	**100%**	**91%**	**92%**	**96%**	**95%**

More precisely, the Direction Cosine Matrix needs the most computational power since a 3x3 matrix has to be calculated and sent as an input to the HMMs, which may be a very important constraint for real-time applications. Taking into consideration the fact that only the upper part of the potter's body contributes in a direct way to the creation of the object while he/she is seating on a chair in front of the wheel, we can conclude that the degrees of freedom of his/her body are really reduced. Additionally, DCM are widely used on animation but not for analysis, recognition or modeling of rotations. Quaternions

impose the non-Euclidean space, which cannot be easily interpreted by humans and consequently they cannot be meaningful for learning purposes.

With regards to the potter B, we have also evaluated the recognition accuracy based on jackknife method. In the table IV, the precision and recall for his G_i^B are presented. During this test we use Euler angles since they have been previously identified as the most relevant descriptor. Like in the previous example 20 queries for recognition per G_i have been asked to the M_i^B. The precision and recall are perfect for G_1^B to G_4^B. For G_5^B, there is one false recognition since G_4^B and G_5^B are very similar.

The difference between G_4^B and G_5^B is that, in G_5^B the potter B defines the shape with a tool and a sponge but in G_4^B he defines the shape without any tool, just with his hands. Gesture G_6 has the lowest recognition rate because the potter was very disturbed. The repeatability of this gesture is low and it has a direct impact on its recognition rate. Even if the number of G_i^B is increased compared to G_i^A, the Precision and Recall are still excellent.

Table IV. Precision and Recall per gesture based on 5 iterations of jackknifing using Euler angles-Potter B.

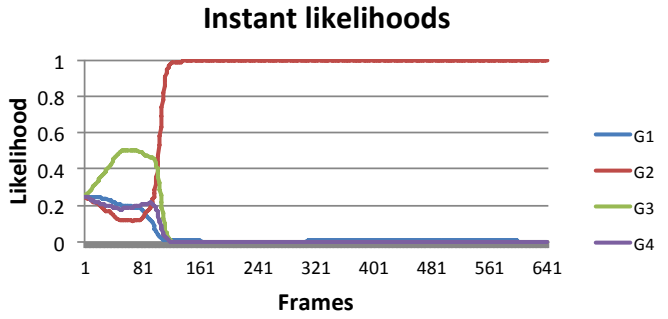
		Maximum likelihoods (ψ_{max})						
		M_1^B	M_2^B	M_3^B	M_4^B	M_5^B	M_6^B	**Recall**
Observations ($Y_{0:k}^B$)	G_1^B	**20**	-	-	-	-	-	100%
	G_2^B	-	**20**	-	-	-	-	100%
	G_3^B	-	-	**20**	-	-	-	100%
	G_4^B	-	-	-	**20**	-	-	100%
	G_5^B	-	-	-	1	**19**	-	95%
	G_6^B	2	-	-	-	-	**18**	90%
Precision		91%	100%	100%	100%	95%	90%	

The possibility to do cross-potters evaluation has been rejected since it is not meaningful to compare discrete gestural skills of a rare expertise, which are strongly related with the local identity of each expert.

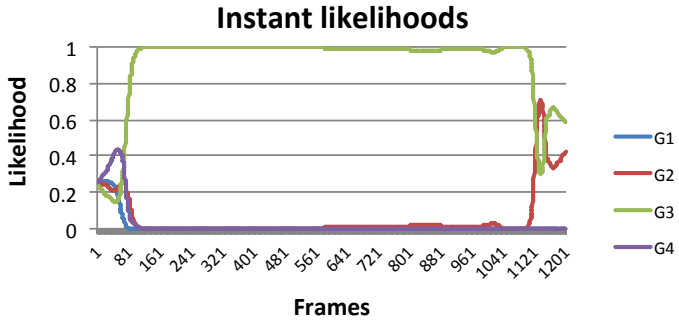
The recognition rate is very high but is effective after a latency of 80-120 frames (*1.5 to 2 seconds*). This latency corresponds to the time that is needed to separate the different HMM models based on their associated maximum likelihood. Co-articulation has an impact to the computation of the instant likelihood and this effect usually becomes more important during the transitions between gestures. It lasts short time periods and the gestures are not clearly recognized during these periods. More precisely, co-articulation can be defined as the fusion of distinct actions into larger and holistically perceived chunks [Hardcastle et al., 1999].

On the left of figure 6 for example, G_2^A is given to ArtOrasis. At the beginning, there is a maximum likelihood alternation effect between the M_i^A. After the first 1.5 seconds,

M_2^A has the maximum likelihood until the end of the observations. As we can see, the likelihood of M_2^A is very high comparing to other models and this can be justified by the fact that the expert repeats the gestures in a very precise way.



(a)



(b)

Fig 6. Instant likelihoods per frame using Euler angles. (a) On the left G_2^A and (b) on the right G_3^A are given as test inputs to the models.

On the first figure a) of the figure 6, there is still a latency of about 1.5 second before M_3^A becomes the maximum likelihood. Then, this effect is normalized and almost instantly the M_3 gives the maximum likelihood for about 1040 frames. Just after the frame 1121, the maximum likelihood starts to alternate between M_2^A and M_3^A due to co-articulation. In G_2^A (*the raise*), the potter needs to clean his hand before to take a small tool and starts the first configuration of the shape. During the co-articulation phase, the right hand goes under the left and vice-versa, which is a common phase with the gesture G_2^A also.

In parallel, we applied a Levene's test in order to detect the equality of variances between the Euler angles for the potter A. According to this test, the variances of the four gestures in three axis are not equal. By applying the One-Way-ANOVA test, we observe that the mean values for all the four gestures are not equal on the axis X by comparing pairs of gestures. Also, equalities are extracted between G_1^A - G_2^A on the axis Z and G_2^A - G_3^A on the axis Y. The conclusion of the statistical analysis is that the four gestures are

different in terms of means and this fact contributes to the very high recognition rates. The results of the One-Way-ANOVA for the potter B are very similar to those of the potter A.

As it has been discussed before, the recognition is useful for the system in order to distinguish the gestures between them. But when the gesture is correctly recognized then it is important to have information about its temporal evolution (*time index*) compared to its model. This can be used not only to measure the different performances of the same potter, but also as a way to measure the distance between the performances between expert and learner in real-time. To do this, a temporal rescaling or time warping of the gesture can be directly obtained from our hybrid DTW-HMM approach. Time alignment experiments have been done for the right palms of the two potters. In figure 7, the time alignment of G_1^B for the potter B is shown.

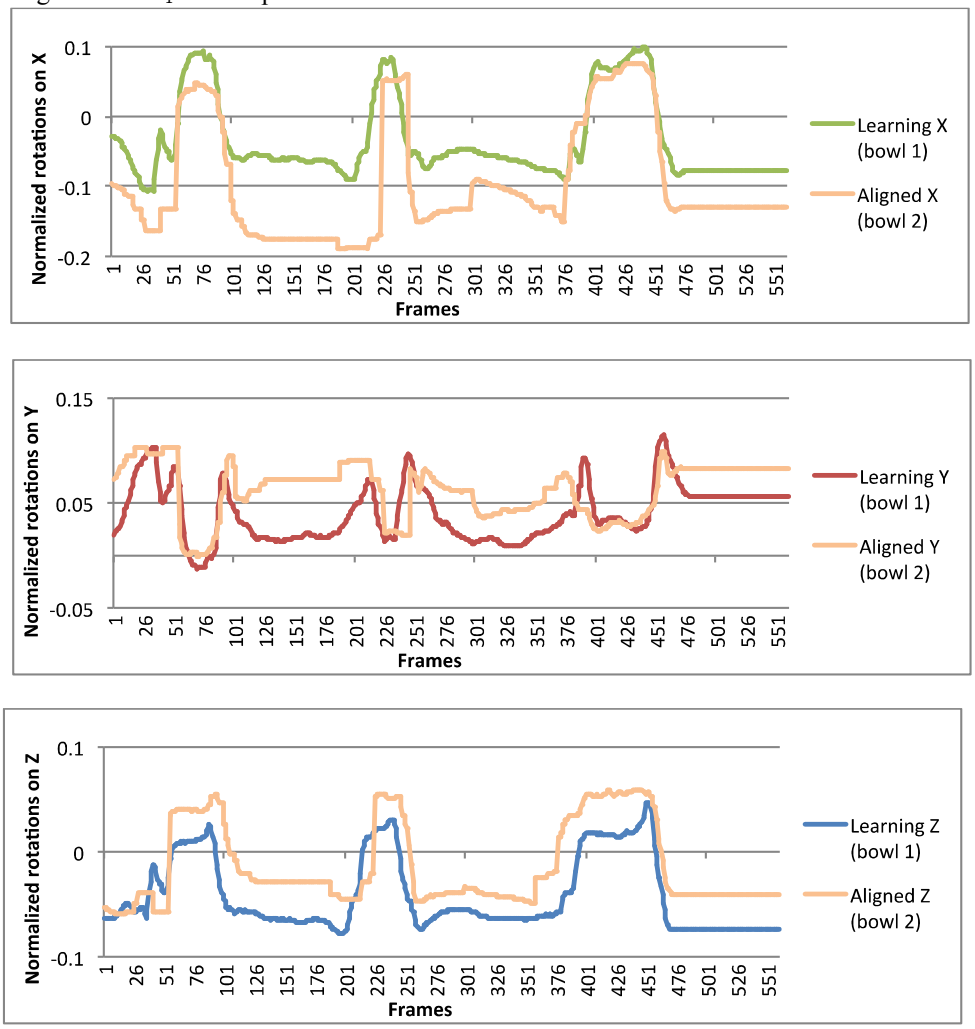


Fig.7 Example of aligned data from the right palm of the potter B for G_1^B (*Opening the Bottom*)

Normalized rotations of the first bowl for the right palm have been given to ArtOrasis as a learning sequence. Then, the normalized rotations of the second bowl have been given both as recognition and time alignment sequence. The results presented above show that for the G_1^B the aligned sequence fits well with the one that has been used for the learning of the HMMs for all three axis. Finally, from this alignment, it is possible to examine precisely where the main differences occur between the different sequences. . As previously mentioned, such difference could be displayed in real-time and offered thus feedback to the potter.

CONCLUSION

Conscious of the need for transmission of the ICH and for its preservation, we proposed a methodology for wheel-throwing gestural know-how preservation through capturing, modeling, online recognition and time alignment of different performances for the most fundamental gestures. In order to validate this methodology, the technological prototype ArtOrasis has been developed. It is able to capture rotations of body segments using inertial sensors and recognize the expert postures and gestures based on machine learning techniques. Both methodology and system have been evaluated on basic expert gestures for the creation of a bowl with the help of different gesture descriptors. Through experiments conducted on two different case studies of rare expertise, we show that gesture recognition with machine-learning can be successfully applied to the creation of simple objects in wheel-throwing pottery. This illustrates that technical gesture recognition can be used for modeling of gestural skills, which is a first required step for "capturing of ICH" in a form facilitating its transmission by interactive pedagogical tools. The high level of recall and precision of gestures recognition is justified by the fact that there are no equalities of variance but also by the expertise of the potters, who repeated the gestures in a very precise way.

The long-term goal of this research is the development of the appropriate methodology and technology for collecting, recording, classifying and modeling of hand gestures that constitute a rare know-how in various types of handicrafts. Since the role of finger touching on the material is very important, a future goal of this research is to extend the current methodology by combining it with computer vision in order to capture finger movements as well. To propose a completed methodology it would be also interesting to combine kinematic data (upper part of the body and fingers) with kinetic information about the pressure brought on the clay. But also the object detection and scene segmentation would give precious information about the evolution of the creation process and also about the progression of the gesture. The results of this study will not only contribute to the preservation of this gestural know-how but also to the development of a system aiming at the transmission. These results could be also used for renewal of ICH by proposing gestural metaphors for the creation of augmented musical performances based on handicraft gestures.

ACKNOWLEDGEMENTS

The research project is implemented within the framework of the Action «Supporting Postdoctoral Researchers» of the Operational Program "Education and Lifelong Learning" (Action's Beneficiary: General Secretariat for Research and Technology), and is co-financed by the European Social Fund (ESF) and the Greek State.

REFERENCES

AYLWARD, R., DANIEL LOVELL S., AND JOSEPH A. PARADISO, 2006, A Compact, Wireless, Wearable Sensor Network for Interactive Dance Ensembles, In Proceedings of *BSN 2006, The IEEE International Workshop on Wearable and Implantable Body Sensor Networks*, Cambridge, Massachusetts, April 3-5, pp. 65-70

ABDI, H., WILLIAMS, L.J., 2010, « Jackknife », In Neil Salkind (Ed.), Encyclopedia of Research Design. Thousand Oaks, CA: Sage.

BAKIS, R. 1976. Continuous speech recognition via centisecond acoustic states. In Proc. 91st Meeting of the Acoustic Society in America .

BEVILACQUA, F., GUÉDY, F., SCHNELL, N., FLÉTY E. AND LEROY N., 2007, 'Wireless sensor interface and gesture-follower for music pedagogy'. In Proceedings of *the International Conference of New Interfaces for Musical Expression*, New York, USA, pp 124-129.

BEVILACQUA, F., ZAMBORLIN, B., SYPNIEWSKI, A., SCHNELL, N., GUÉDY, F. AND RASAMIMANANA, N., 2010 'Continuous realtime gesture following and recognition', *LNAI 5934*, pp.73–84.

CHEVALLIER D., 1991, « *Savoir faire et pouvoir transmettre: Transmission et apprentissage des savoir faire et des techniques* », Edition MSH France

CODUYS T., HENRY C. AND CONT A., 2004, TOASTER and KROONDE: High-Resolution and High-Speed Real-time Sensor Interfaces, In Proceedings of the *International Conference on New Interfaces for Musical Expression*, Hamamatsu, Japan.

DAPOGNY A., DE CHARETTE R., MANITSARIS S., MOUTARDE F., GLUSHKOVA A., 2013, Towards a Hand Skeletal Model for Depth Images Applied to Capture Music-like Finger Gestures, *10th Int. Symposium on Computer Music Multidisciplinary Research (CMMR'2013),* Marseille France

DEMOUCRON M., ASKENFELT A. AND CAUSSÉ R., 1994, Observations on bow changes in violin performance. In Proceedings of *Acoustics, Journal of the Acoustical Society of America*, volume 123, page 3123.

DUNN S., WOOLFORD K., BARKER L. ET AL, 2011, Motion in Place: a Case Study of Archaeological Reconstruction Using Motion Capture, CAA2011 - Revive the Past: In Proceedings of the *39th Conference in Computer , Applications and Quantitative Methods in Archaeology*, Beijing, China. 12-16 April

DYMARSKI P., 2011, *Hidden Markov Models, Theory and applications*, Published by InTech Croatia

FLÉTY E., MAESTRACCI C., 2011, Latency improvement in sensor wireless transmission using ieee 802.15.4, Proceedings of the International Conference on New Interfaces for Musical Expression, pages 409–412, Oslo, Norway

GANDON E., 2011, *Influence of cultural constraints in the organization of the human movement: proposition of a theoretical framework and empirical support through the example of pottery-throwing* (France / India Prajapati / India Multani Khumar), PhD thesis, Marseille University, France

GANDON, E., CASANOVA, R., SAINTON, P, COYLE, T., ROUX, V., BRIL, B., & BOOTSMA, R.J., 2011, A proxy of potters' throwing skill: ceramic vessels considered in terms of mechanical stress., *Journal of Archaeological Science, 38, 1080-1089.*

Grunberg D., 2008 Gesture Recognition for Conducting Computer Music. Retrieved January 10, from: http://schubert.ece.drexel.edu/research/gestureRecognition

HARDCASTLE, W., AND HEWLETT, N., 1999 *Coarticulation: theory, data and techniques*. Cambridge: Cambridge University Press.

LE BELLU S., LE BLANC B., 2010, How to Characterize Professional Gestures to Operate Tacit Know-How Transfer? *The Electronic Journal of Knowledge Management* Volume 10 Issue 2 (pp142-153)

LEMAITRE, J., CHABOCHE, J.L., 2004, Mécanique des Matériaux Solides. Dunod, Paris.

RAPTIS M., KIROVSKI D., HOPPE H., 2011, Real-Time Classification of Dance Gestures from Skeleton Animation. Eurographics/ In Proceedings of *ACM SIGGRAPH Symposium on Computer Animation.*

RABINER, L. R., 1989, A tutorial on hidden markov models and selected applications in speech recognition. In *Proceedings of the IEEE*, pages 257–286.

RASAMIMANANA N. AND BEVILACQUA F., 2009, Effort-based analysis of bowing movements : evidence of anticipation effects. *The Journal of New Music Research*, 37(4): 339 – 351

TODOROFF T., 2011, Wireless digital/analog sensors for music and dance performances. In Proc. the *International Conference on New Interfaces for Musical Expression*, pages 515–518, Oslo, Norway

WANG K.A., LIAO Y.C. , CHU W.W. , YI-WU CHIANG J., CHEN Y.F. AND CHAN P.C., 2011, Digitization and Value-Add Application of Bamboo Weaving Artifacts, C. Xing, F. Crestani, and A. Rauber (Eds.): ICADL 2011, *LNCS 7008*, pp. 16–25, 2011. Springer-Verlag Berlin Heidelberg

UNESCO 2003. 1998. *Segmentation and Categorization of Phonemes in Continuous Speech*. Technical Report TR-CST25JUL98, Center for Sensor Technology, University of Pennsylvania.