



**HAL**  
open science

## **Interaction by gesture recognition: a tool to virtually revive the automata of the Antiquity**

Jérôme Nicolle, Philippe Fleury, Olivier Lézoray

### ► **To cite this version:**

Jérôme Nicolle, Philippe Fleury, Olivier Lézoray. Interaction by gesture recognition: a tool to virtually revive the automata of the Antiquity. 22nd ConVRgence (VRIC) Laval Virtual 2020, Apr 2020, Laval, France. pp.110-117, <10.20870/IJVR.2020..3316>. <hal-02546867>

**HAL Id: hal-02546867**

**<https://hal.science/hal-02546867v1>**

Submitted on 18 Apr 2020

**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

# Interaction by gesture recognition: a tool to virtually revive the automata of the Antiquity<sup>1</sup>

Jérôme Nicolle<sup>1</sup>, Philippe Fleury<sup>2</sup>, Olivier Lezoray<sup>3</sup>

<sup>1</sup>Normandie Univ, UNICAEN, ERLIS, 14000 Caen, France

<sup>2</sup>Normandie Univ, UNICAEN, ERLIS, 14000 Caen, France

<sup>3</sup>Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC, 14000 Caen, France

**Corresponding author:** Jérôme Nicolle, [jerome.nicolle@unicaen.fr](mailto:jerome.nicolle@unicaen.fr)

**Keywords:** Automata – Antiquity – Virtual Reality – Gesture Recognition – Interaction

## Abstract

This article presents research works devoted to the virtual reconstruction of automata of the antiquity along with associated user interaction. This work is divided in two distinct parts: the virtual reconstruction of ancient automata and the development of an interaction method by gesture recognition. The final application links both the automata with gesture recognition to revive the astonishment feeling generated by the Greeks and Romans automata mechanisms. First the paper focuses on the analysis of ancient sources required to virtually restore some ancient automata (designed during a period ranging from the 3rd century BC to the 1st century AD). Virtual reconstructions make it possible to understand how a mechanism works and to study different hypotheses. In a second step, we introduce gesture recognition as a tool to virtually interact with automata. The construction of an interactive immersive application allows then the user to interact virtually with automata of the Antiquity.

## 1. Introduction

In the modern world, an automaton is defined as a machine "which, by means of mechanical, pneumatic, hydraulic, electrical or electronic devices, is capable of acts imitating those of animated bodies". This brief definition can also be applied to machines in Antiquity. In this paper, we present a virtual reconstruction of two automata of the Antiquity. This has been made possible after a study of the remaining ancient source materials that describe how they function. If the virtual reconstruction is interesting in itself, enabling a user to directly interact with an automaton can be beneficial in terms of user experience. In addition, it can bring to the user the astonishment that these automata were supposed to produce. This however supposes that the user can interact with the automata. This virtual interaction adds another dimension to the immersive application and the user becomes an actor for the manipulation of objects in real time. Indeed, the common everyday life gestures allow a natural interaction with automata in 3D environments (catching, releasing, pouring, etc.). This interaction also offers the possibility of checking the ergonomics of the object in order to know if the object can be easily manipulated. The union of both virtual reconstruction and virtual interaction makes it possible to understand how these automata worked and to reproduce the astonishment caused by these machines.

## 2. Automata in Antiquity

In Antiquity, many engineers worked on the study and design of automatic mechanisms (Chapuis, 1949) (Fayol, 1962). However, few real automata artefacts as well as descriptive texts have survived until today (Drachmann, 1948). Our study is based on automata conceived during a period that ranges between the 3rd century BC and the 1st century AD. In that time span, three surviving treatises on the subject can be considered. They were written by Philo of Byzantium (250 BC) and Hero of Alexandria (1st century AD).

---

<sup>1</sup> This work was funded by a thesis scholarship from the Normandy Region.

## 2.1. Available sources

Before any reconstruction can take place and in order to know both the exact dimensions and the shapes and materials of each automaton machine, it is essential to perform what is called an analysis of ancient source materials. This analysis usually takes place in three stages: i) the study of Latin and Greek literary sources, ii) the study of iconographic sources, and iii) the study of archaeological sources. Within the framework of ancient automata, only literary sources are available such as the treatise on *Pneumatics* by Philo and Hero (Hero of Alexandria, 1997) (Philo of Byzantium, 1974) and the one of the *Automata* by Hero (Hero of Alexandria, 1903a). Unfortunately, neither iconographic or archaeological sources are currently available.

## 2.2. The automated models

The study of both these treatises makes it possible to draw up a list of five automata that can be virtually reconstructed. Each automaton has a specific mechanical system and all five automata provide a representative catalog of the different ancient techniques used (pneumatic, hydraulic, weight system, counterweight, etc.). Thus, we have modeled: 1) The miniature game of Heracles and Ladon (Hero of Alexandria), 2) The automatic maidservant distributing water and wine (Philo of Byzantium), 3) The intermittent fountain of the owl and songbirds (Hero of Alexandria) (Carrer, 2016), 4) The automatic doors of a miniature temple (Hero of Alexandria), 5) The mobile-based automatic theater (Hero of Alexandria) (Prou, 1881). In the next sections, due to space constraints, we present only two virtual reconstructions of these ancient automata: the automatic maidservant and the miniature game of Heracles and Ladon.

## 3. Virtual reconstruction

### 3.1. The maidservant of Philo of Byzantium

The automaton of the automatic maidservant distributing water and wine is described in paragraph 30 of the *Pneumatics* by Philo of Byzantium. However, no Greek texts are available today for this treaty, only Arabic versions are available. Before one can understand the functioning of the mechanism, it is useful to specify that it is hidden inside the maidservant's body to be invisible to the public. The text tells us that the automaton is entirely made of copper or silver, and that the maidservant is represented standing, with a jug in the right hand (see Figure 1). It also tells us that the bowl is weighed down by a ballast at its base and that the distributed mixture is two thirds of wine and one third of water. Indeed, during Antiquity, wine was stored in a very concentrated state and it had to be diluted with water to be drinkable. This explains the presence of the two tanks in the maidservant since the mixture is not prepared in advance in the tank but made by the automaton. When the bowl is placed on the servant's left hand, the left arm is lowered. This arm is connected to a butt by two trunnions. In mechanics, a trunnion is a cylindrical part around which a part receives a rotational movement. One of the trunnions (A) is fixed and serves as an axis of rotation for the butt. The other (B) connects the arm to the butt (C) with a rod (D). This butt is weighed down at its base, thus serving as a counterweight to the mechanism to hold the arm in place. When the arm is lowered, it pulls on the end of the butt, the trunnion which is not fixed acts as a lever and the mechanism starts. Two pipes (E, F) come out of the tank, one for wine and the other for water. Two solid rods (H, G) form the end of the butt and are nested in these two pipes. When the mechanism is activated, the rods slide in the pipes using leather seals. These allow silent operation of the machine, so guests using the maidservant do not suspect the presence of a mechanism inside. A slot is present on each solid rod. The rod that comes into contact with the wine hose (G) is the longer. The slot at the end is twice as long as the one on the other rod (H). When the slit is at the level of the pipe, it allows air to enter it for a certain period of time. When the slit has completed its passage through the pipe, the solid rod blocks the entry and the passage of air. The tank, dimensioned from head to chest, is divided into two equal parts serving as a reserve for wine and water. The head is a tight cover and offers the possibility of filling the tanks. The air enters the tank through the two pipes, replaces the vacuum and drives out the liquid (wine and water) that flows to the ewer by two pipes hidden in the forearm of the servant. The two slots are of unequal dimensions so as to respect the ratio of two thirds of wine and one third of water. When the flow of liquids is



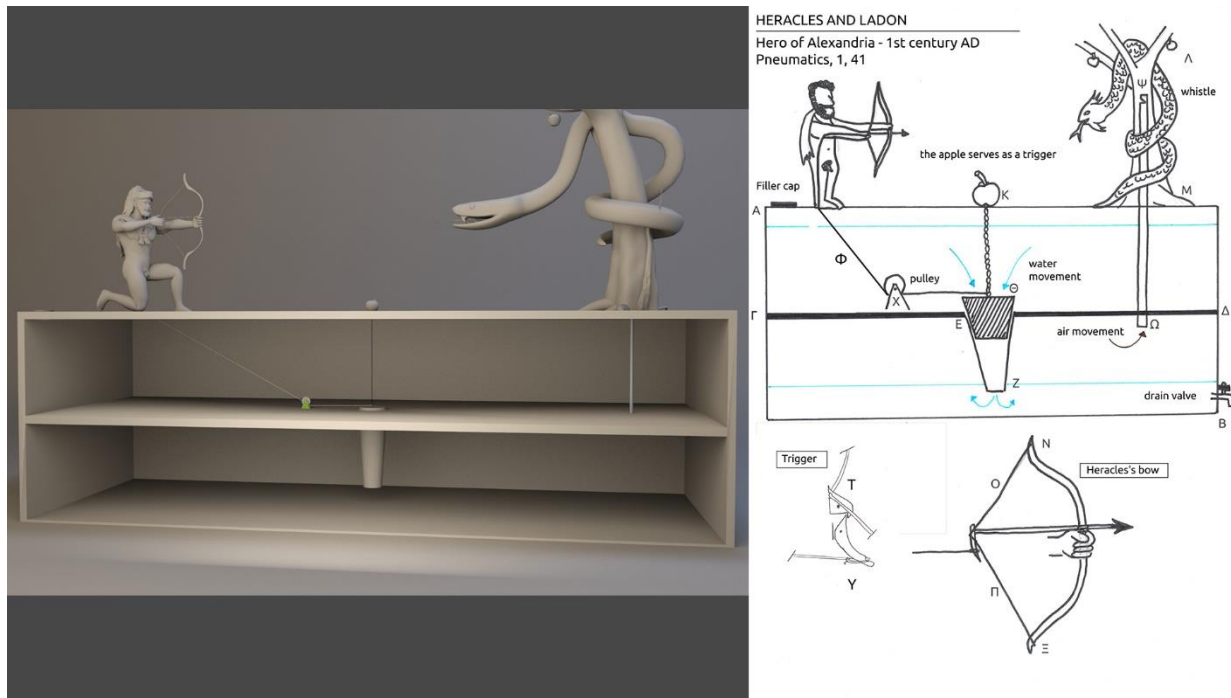


Figure 2 : Internal mechanism (right) and virtual reconstruction (left) of the miniature game of Heracles and Ladon.

#### 4. Interaction by gesture recognition

After modeling the virtual reconstruction of the two automata, we will now focus on how to interact with them. We consider an interaction where the user performs actions with his hands. This therefore requires the ability to recognize hand gestures in real time in order to trigger an item of the mechanism.

##### 4.1. Development of a new method

Recent approaches for gesture recognition are based on machine learning techniques, and in particular on deep learning techniques. We have considered the state-of-the-art approach of Devineau (Devineau, 2018). This method is today one of the most effective in the classification of gestures on the DHG 14/28 Gesture Dataset database (Smedt, 2017). We describe this method in the sequel. The hand is represented by a set of 22 hand joints, each represented by its 3D coordinates within a time period that captures the gesture to be recognized. The approach of Devineau considers the 66 coordinates (X, Y, Z for the 22 hand joints) separately within a time period of 100 frames. Each coordinate (denoted as  $C_i$ ) is processed using three different branches that have the structure of a convolutional neural network (CNN), made of a cascade of three 1D convolutions and pooling layers. This enables to extract three sets of features for each coordinate, that are concatenated to represent each joint independently. Then, all the features of the 66 joints are concatenated into a single vector that represents the gesture to be recognized. It is the input of a fully connected layer that performs the gesture recognition into 14 or 28 gestures. Figure 3 presents an overview of this architecture (when one considers only the  $C_i$  coordinates). Devineau's method does not take into account the physical relationships that exists between the hand's joints (the fingers) at all and that play an important role in the description of a gesture. Our improvement over Devineau's approach will consist in taking into account the movements of the fingers. Indeed, during the performance of a gesture, each finger has a precise position and orientation as well as specific coordination of its joints. To that aim, we add a fourth branch of processing that will compute dedicated features for each finger  $D_i$ , enabling to capture higher-level cues. Each finger is composed by the following joints, as shown in Figure 3a: finger 1 (inch joints 3-4-5-6), finger 2 (index, joints 7-8-9-10), finger 3 (major, joints 11-12-13-14), finger 4 (ring finger, joints 15-16-17-18), finger 5 (little finger, joints 19-20-21-22). Joints 1 and 2, that represent the wrist and the palm of the hand, are not used to isolate the fingers. In the new branch we have designed, each finger (denoted as  $D_i$ ) is processed by three 2D convolutions and pooling layers. We

consider 2D convolution layers since we have 4 joints for each finger on 100 consecutive frames. This is similar to the processing performed by the architecture of Devineau for each 1D temporal coordinate signal.

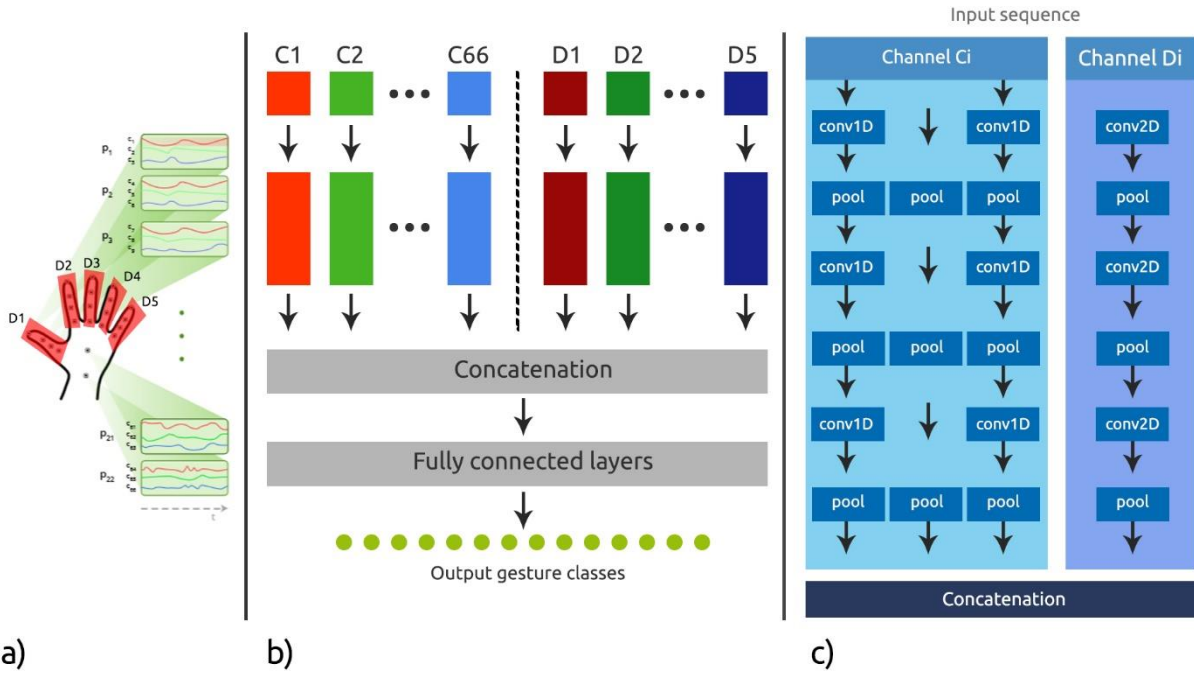


Figure 3: CNN architecture of our gesture recognition method. a) Illustration of the different input parameters of the model: the channels  $C_i$  for each joint's coordinates and the channels  $D_i$  representing the fingers. b) General view of the neural network. c) Illustration of the parallel branches allowing the features to be extracted.

Theoretically the management of each finger should allow a gain of precision in the recognition of gestures. This can indeed capture overall movement features of each finger that represent statistics taken at a higher semantic level. The features extracted for each finger are concatenated together with the output of the Devineau's approach and transmitted to the fully connected layer that performs the classification. This last layer classifies the different gestures into 14 categories when used with the DHG 14/28 Gesture Dataset. This new method that we have developed obtains a classification rate of 89.97% on the DHG 14/28 gestures database. This is slightly behind the approach of Devineau that has a classification rate of 91.28%. The standard non-deep approach of De Smedt obtains a score of 88.24% (De Smedt, 2016). However, in the context of the interactive immersive application, we do not need to recognize all the gestures that are present within this database. In particular, after looking closely at the results, we have noticed that for the "catch" gesture, our method is more effective than that of Devineau with a recognition rate of 95.2% versus 94.8%. This gesture being of paramount importance for the virtual interaction with automata, shows the benefit of our approach that captures the higher-level cues of the finger. In addition, this motivates the building of a specific dataset that contains only the gesture we are interested in.

## 4.2. Interaction Gesture Dataset

For the purpose of our application, a new gesture database was created in February 2019. This database, called *Interaction Gesture Dataset (IGD)*, brings together two useful gestures for interacting with ancient automata. They have been obtained with the "Perception Neuron" motion capture suit:

1. *Catch*: used to detect when an object has been taken. The combination with the movement of the arm allows to know in which direction the user is carrying the object. The gesture begins with an open hand and ends with a closed hand.
2. *Pour*: is combined with the first gesture. It tells if the user is pouring something, especially when filling a tank. The entire gesture is carried out with a closed hand.

The *Interaction Gesture Dataset* was built by a panel of 4 people (3 are right-handed and 1 is left-handed). Each person performed the gesture 30 times, which gives  $4 \times 30 = 120$  executions per gesture. The dataset provides a total of 240 gestures. We plan to distribute this new database on a dedicated website this year. However, the Perception Neuron suit used to build this new dataset is different from the Intel Real Sensor. It provides a hand skeleton with only 10 joints unlike the Intel Real Sense sensor that is made of 22 joints (see Figure 4 a and b). The model of Devineau, as well as our new approach presented in the previous section, is designed to take as input a hand skeleton of 22 joints. Therefore, to train these models and further classify them, we have to transform the 10 joint skeletons into 22 joint skeletons. To do this, we use a specific linear interpolation with physical constraints to take into account the shape of the hand and the position of the joints (see Figure 4 c and d). To interpolate the coordinates of a missing joint, we consider its two closest neighbors and compute a weighted average. On this IGD database a classification rate of 84.26% was obtained with our new approach. The state-of-the-art approach of Devineau obtains a lower classification rate of 83,57% on these same data. This shows the advantage of having built this second specific database because it specifically contains the gestures that we need to interact with ancient automata and enables us to obtain better results than the state-of-the-art approach of Devineau.

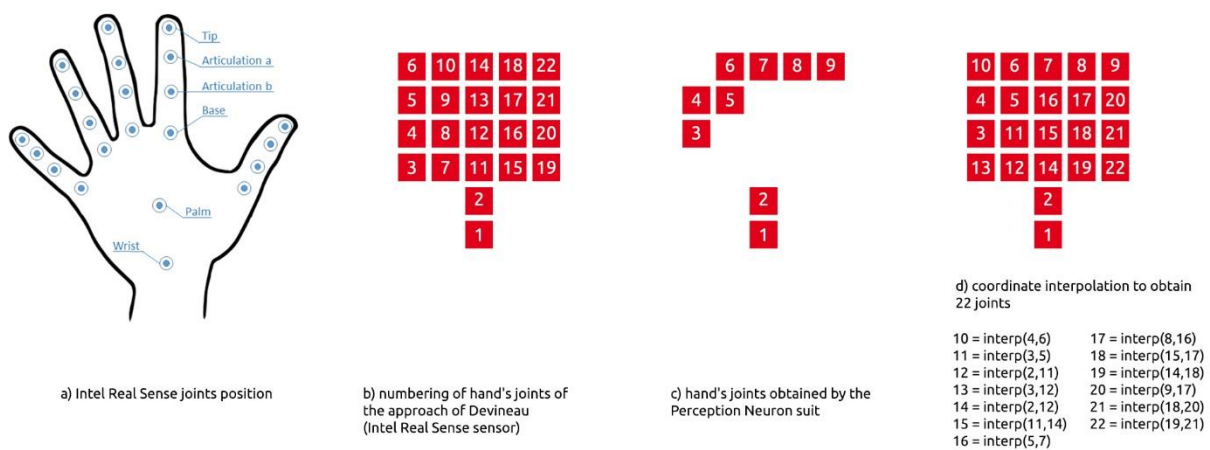


Figure 4: Principle of interpolation of coordinates in order to restore a skeleton of the hand with 22 joints, similar to the Intel Real Sense sensor.

### 4.3. Integration of the method in an interactive immersive application

The use of the designed interaction method within the interactive immersive application makes it possible to obtain an effective and efficient level of detection of user gestures. When the Trigger is activated by the user's hand, the algorithm performs gesture detection every 20 frames in order to limit the risk of errors, by aggregating successive temporal gesture recognitions. In the worst case (approximately 1 in 12 times), we will have to do the gesture twice so that it is well recognized. A demonstration of the interactive application was led during a public show on April 4, 2018 (see Figure 4), in front of an audience of 264 people of very varied age<sup>2</sup>. Its purpose was to virtually interact with Philo of Byzantium's automatic maidservant to demonstrate it and let the audience understand how its internal mechanism works. This presentation was really well received by the audience. Each automaton is integrated in its context (see Figure 6) and the interaction with the user

<sup>2</sup> <https://www.youtube.com/watch?v=bkrmy6Nlpuc&t=2140s>

makes it possible to judge the degree of reliability of its reconstruction. The method has obtained good results in terms of real-time gesture recognition (a classification rate of 90.4% was recorded afterwards on the fifty gestures performed during the demonstration).



*Figure 5 : Virtual interaction of the automatic maidservant of Philo of Byzantium.*



*Figure 6: Contextualization of automata.*

## **5. Conclusion**

In this paper, we have considered the virtual reconstruction of ancient automata. The approach has required background knowledge both in studying ancient sources and learning how machines work. The study ancient sources has enabled us to propose a virtual reconstruction of two automata. Machine learning has enabled to provide gesture recognition as a tool for interacting with the designed virtual automata. Linking 3D models of machines with virtual interaction adds an extra dimension to the user's experience and enables to appreciate the effectiveness of virtual restitution. This virtual immersion revives the thaumaturgy effect of using automata

on users. The whole immersive application has been realized under Unity 3D, where the user navigates and interacts in a natural way with the ancient automata.

## 6. References

- Carrer, J. (2016). *Étude de la valorisation d'un objet antique : l'exemple de l'automate des oiseaux chanteurs et de la chouette d'Héron d'Alexandrie* [Mémoire de Master sous la dir. de M.T. Cam].
- Chapuis, A., & Droz, E. (1949). *Les automates, figures artificielles d'hommes et d'animaux : histoire et technique*. Éditions du Griffon,
- Chapuis, A., Gélis, É., & Haraucourt, E. (1984). *Le Monde des automates : étude historique et technique*. Slatkine.
- Devineau, G., Moutarde, F., Xi, W., & Yang, J. (2018). *Deep Learning for Hand Gesture Recognition on Skeletal Data*. 106-113.
- Devineau, G., Xi, W., Moutarde, F., & Yang, J. (2018). *Convolutional Neural Networks for Multivariate Time Series Classification using both Inter- & Intra- Channel Parallel Convolutions*.
- Drachmann, A. G. (1948). *Ktesibios, Philon and Heron: a study in ancient pneumatics*. E. Munksgaard.
- Fayol, A. (1962). *Précurseurs oubliés : de Ctesibios à Judson*. Éditions et Imprimeries du Sud-est.
- Hero of Alexandria. (1903a). *Automates* (W. Schmidt (éd. et trad.)). Teubner.
- Hero of Alexandria. (1903b). *Pneumatiques* (W. Schmidt (éd. et trad.)). Teubner.
- Hero of Alexandria. (1997). *Pneumatiques* (G. Argoud & J.-Y. Guillaumin (éd. et trad.)). Publications de l'Université de Saint-Étienne.
- Marcinkowski, A., & Wilgaux, J. (2004). Automates et créatures artificielles d'Héphaïstos : entre science et fiction. *Techniques & Culture.*, 43-44. <http://journals.openedition.org/tc/1164>
- Philo of Byzantium. (1903). *Pneumatiques* (B. Carra de Vaux (éd. et trad.)). Imprimerie nationale.
- Philo of Byzantium. (1974). *Pneumatiques* (F. D. Prager (trad.)). L. Reichert.
- Prou, V. (1881). *Les théâtres d'automates en Grèce au IIe siècle avant l'ère chrétienne : d'après les Αυτοματοποιικα d'Héron d'Alexandrie*. Imprimerie Nationale.
- Smedt, Q. D., Wannous, H., & Vandeborre, J. (2016). *Skeleton-Based Dynamic Hand Gesture Recognition*. 1206-1214.