 

# A Grammar for Specifying Full-Body Gestures Elicited for Abstract Tasks

David Céspedes-Hernández [a,*], Juan Manuel González-Calleros [a], Josefina Guerrero-García [a], and Jean Vanderdonckt [b]
[a] *Facultad de Ciencias de la Computación, Benemérita Universidad Autónoma de Puebla, Mexico. E-mail: {dcespedesh,juan.gonzalez}@cs.buap.mx*
[b] *Université catholique de Louvain (UCLouvain), LouRIM. Place des Doyens, 1. B-1348 Louvain-la-Neuve, Belgium. E-mail: jean.vanderdonckt@uclouvain.be*

**Abstract.** A gesture elicitation study consists of a popular method for eliciting a sample of end end users to propose gestures for executing functions in a certain context of use, specified by its users and their functions, the device or the platform used, and the physical environment in which they are working. Gestures proposed in such a study needs to be classified and, perhaps, extended in order to feed a gesture recognizer. To support this process, we conducted a full-body gesture elicitation study for executing functions in a smart home environment by domestic end users in front of a camera. Instead of defining functions opportunistically, we define them based on a taxonomy of abstract tasks. From these elicited gestures, a XML-compliant grammar for specifying resulting gestures is defined, created, and implemented to graphically represent, label, characterize, and formally present such full-body gestures. The formal notation for specifying such gestures is also useful to generate variations of elicited gestures to be applied on-the-fly on gestures in order to allow one-shot learning.

Keywords:, Gesture elicitation study, gesture grammar, gesture recognition, gesture user interfaces, engineering interactive computing systems, one-shot learning

## 1. Introduction

*Natural Language Processing* (NLP) [1] is often defined as the science of making computers understand information in the same way as a person does. Human communication may be analyzed, using a basic taxonomy, according to whether messages are transmit- ted through verbal or non-verbal mechanisms. It is especially interesting that despite the fact multiple investigations mention that approximately 65 percent of the human communication is carried out through nonverbal mechanisms [2], it is possible to observe in the literature that most of the research is focused on analyzing speech, and that emphasis on *Gesture Recognition* (GR), considering facial expressions and body motion, has been given in the last two decades as consequence of the development of processing techniques like Machine Learning (ML) algorithms, data processing techniques, and of hardware devices.

GR for interaction purposes has been addressed during the last decades following different approaches and through its application to a number of domains such as education, sports training, rehabilitation, and robotics [3,4,5,6]. Current proposals report to achieve accuracy levels close to 100 percent using machine learning techniques and following traditional training procedures on which a large number of samples are gathered, processed and labeled. Due to its nature, the implementation of mechanisms for dynamically adapting gestural interfaces to a variety of contexts of use (defined as a user and related tasks, a platform, and an environment [26]), demands the consideration of an approach such that makes it possible for users to enrich the training of the recognition system providing a gesture only once, this is, training must be performed using a single sample.

We introduced a methodology [7] for GR relying on user-defined gestures sets considering a *One-Shot Learning* (OSL) approach [8]. However, even though the stages for achieving GR and for characterizing the

[*]Corresponding author. Tel. +52 12 345 678; Fax. +52 13 678 789; E-mail: dcespedesh@gmail.com
[*]Corresponding author. Tel. +52 12 345 678; Fax. +52 13 678 789; E-mail: dcespedesh@gmail.com

obtained gestures are described, there is a lack of a formal notation for describing such gestures, making it impossible to compute the gathered information and then, using it for enabling the training of a classifier applying OSL. Often, gesture recognizers work on gesture templates that are defined in their own physical format, such as points in space or vectors [20]

In this paper, we report on how a gesture elicitation study was conducted on full-body gestures for executing abstract tasks, from which a grammar was defined and represented through informal and formal means, making it possible to feed this information to a classifier and to carry out OSL. The purpose of using abstract tasks, observed through their instantiation on the possible commands over a streaming platform, is on that in this way, the obtained grammar will be useful for different interaction domains. We also provide description of rules to be followed for achieving training following the OSL approach.

The rest of the paper is divided into sections. Section 2 contains a brief description of the state of the art including work on full-body GR and an introduction to abstract interactive tasks. The setting up of the experiment and its performance are described in Section 3. Later, Section 4 is dedicated to present the gathered results, showing the obtained grammar and explaining the notation used. Section 5 discusses the insights and the contributions of the paper. Finally, in Section 6, conclusions and the future work to be addressed are presented.

## 2. State of the Art

This section provides a brief state of the art on full-body GR and including the definition and description of abstract interaction tasks.

### *2.1. Gesture Elicitation Study*

To illustrate the purpose of a Gesture Elicitation Study (GES) [23], let us consider a user interface designer willing to design a full-body gesture user interface to enable end users to control a smart home environment with commands such as turn lights on and off, dim lights, and make lights brighter. A sample $P$ of the target population is defined, typically at least 30 in number, i.e., $|P|\geq 30$, to participate in a GES. Each participant is presented with the effect of each function, the intensity of the lights is increased for the "make lights brighter" function, and the participant is elicited to propose a full-body gesture, for example captured by a camera, that would generate the desired effect. When the study is completed for all functions and participants, the designer has collected a resulting set of 120 elicited gestures = 4 (functions) $\times$ 30 (individual gestures). In order to decide which gesture to keep for each function, the designer browses the resulting set of gestures to understand whether there are any gestures in agreement for each function. This agreement could be acquired by computing various formula [25] or by frequency. If the consensus conditions are fulfilled, the designer can be confident that the elicited gestures are intuitive, and other end users would likely guess, learn, and perhaps prefer the same gestures than those elicited in the study.

### *2.2. Full-body Gesture Recognition*

Gestures are (static or dynamic) expressive, meaningful body motions involving physical movements of the fingers, hands, arms, legs, head or face. Through gestures, it is possible to communicate in a nonverbal way and enrich verbal communication as well [9]. A systematic literature review including 60 works on GR was performed following a methodology [7]. As relevant insights, it was noticed that 27% of the works are focused on the study of gestures of the hands using leap motion sensors, depth cameras, wearable devices, and RGB cameras; 7% of the works are aimed at describing interaction by means of lower limbs, using wearable devices and depth cameras; and that only 10% of the analyzed works consider full-body interaction, most of them using depth cameras as data acquisition devices. A deeper revision to the considered works, reported that researchers on GR focus on the study of the most expressive body parts even though they use technology that enables the recognition of full-body gestures. This is understandable, as users are more likely using the most expressive body parts to interact with systems, i.e., hands and facial expressions, however, there are contexts on which users need to provide gestures using less expressive body parts, consider for instance, a rehabilitation system focusing on knee movement recovery. One of the works regarding full-body interaction is [10], in which users were asked to perform gestures for controlling a humanoid robot. In that study, Obaid et al. proposed navigational commands and analyzed gestures provided by users, reporting agreement scores, time performances, and graphical representations of the consensus set. Along with the elicitation of the gestures set, significant insights were discussed like the need of the definition of points of view when trying to recognize gestures for navigation, i.e., if users' motion is being tracked using cameras in front of them, it is possible to misunderstand the direction of the gestures. It is interesting for this research to point out that most of the proposals in the literature addressing elicitation studies have the purpose of providing the base for creating tailored solutions, as they are per-formed on specific tasks within determined applications. As the objectives of this work consider a generalization of the to be obtained gesture grammar, abstract tasks are considered. The following subsection is dedicated to the description of them.

### 2.3. Abstract Tasks

*Abstract tasks* are defined as standard actions [11]: it is possible to find a list of all the interactive tasks that users may perform over an application. Keeping awareness of such tasks, presented in Table 1, while designing user interfaces enables the definition of interaction mechanisms for their instantiations that are specific for the task and context of use.

Table 1
Standard tasks and use examples from Constantine [11].

| Standard action | Use example |
|---|---|
| Action/operation | Print symbol table, color selected shape |
| Start/go/to | Begin consistency check, confirm purchase |
| Stop/end/complete | Finish inspection session, interrupt test |
| Select | Group member picker, object selector |
| Create | New customer, blank slide |
| Delete, erase | Break connection line, clear form |
| Modify | Change shipping address, edit client details |
| Move | Put into address list, move up/up |
| Duplicate | Copy address, duplicate slide |
| Perform (& return) | Object formatting, set print layout |
| Toggle | Bold on/off, encrypted mode |
| View | Show file details, switch to summary |

The intention followed with the definition of abstract tasks was to provide a notation for designing graphic user interfaces. However, for the purposes of this proposal, they will be considered when preparing the experiment, making it possible to define a gesture grammar associated to abstract tasks enabling its use for different contexts. The following section is dedicated to the description of the experiment that was performed in order to obtain a gesture grammar associated to abstract inter-action tasks.

## 3. Setting up and Performance of an Experiment for Identifying a Gesture Grammar

In order to identify a gesture grammar for interaction, we decided to perform a gesture elicitation study as other authors in the literature have reported to do [12,13,14]. A first approach for this was on asking users to provide gestures for abstract tasks as such described in section 2.3. However, it was noticed that users needed to have specific examples of the correspondence of the tasks on determined applications for better understanding the requirement. This led us to decide using the Netflix streaming platform as base but keeping awareness of the relation of the analyzed tasks and abstract tasks. Developing an interaction mechanism for gesture-based applications is a challenging time-consuming activity. This, added to the objective of identifying gestures instead of recognizing them, made it evident that the implementation of a system for gesture recognition was out of the scope of this work, and that in its place, it would be necessary to apply a technique for gathering information of the interaction.

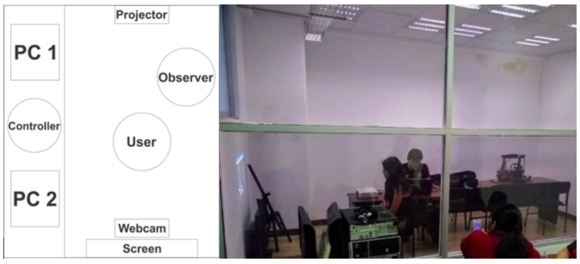


Fig. 1. Experiment setup following the Wizard of Oz technique [16] and its disposition in a Gesell chamber.

The Wizard of Oz technique has been used throughout the history of the development of interactive systems [15,21], and in particular, in the field of natural interfaces development as it is a way to collect data for mixed reality environments or movement commands for interaction in several application domains [16]. Following this technique, the designed experiment consisted on asking users to use their own body gestures to interact with the Netflix streaming platform using only specified body parts and giving them the feeling that those commands actually worked on the application, but actually providing the input to the system through keyboard and mouse commands.

The setup of the experiment, as shown in Fig. 1, took place in a Gesell room and included two computers (PC1 and PC2), the first one for retrieving users' movement information from a webcam and allowing to document the experiment; while the second one was connected to a projector for providing the user with feedback and simulating users' interaction with the system. An observer of the experiment was designated and located next to the user in order to provide support in case any doubt arises. The experiment controller managed the devices to document the experiment and to explain how each user performed the activity.

During the experiment, 90 users divided in 6 groups were asked to provide the gestures they would use to perform 23 commands within the Netflix streaming platform. The 23 tasks that were defined are presented in Table 2 along with their associated abstract task. When users carried out gestures for each of the requested commands, data was gathered including response time, a video record of the corresponding gesture, and a score given by the same user with respect to how natural they thought the gesture was for the task. As a total, 2070 records were acquired and then processed through manual labeling, manual comparison, to finally be assigned with an agreement score.

Table 2

Defined tasks for the gesture elicitation study and their association to abstract interactive tasks from [11].

| Id | Task | Associated abstract task |
|---|---|---|
| 1 | Log in | Select, Start/go/to |
| 2 | Search movie or series | Select, Action/operation |
| 3 | Adjust volume | Modify |
| 4 | Play video | Start/go/to |
| 5 | Change chapter | Perform (& return), Select |
| 6 | Activate subtitles | Select, Toggle |
| 7 | Change language | Perform (& return), Select |
| 8 | Go forward 10 seconds | Action/operation |
| 9 | Go backward 10 seconds | Action/operation |
| 10 | Add to playlist | Create |
| 11 | Log out | Stop/end/complete |
| 12 | Pause video | Stop/end/complete |
| 13 | Deactivate subtitles | Select, Toggle |
| 14 | Skip intro | Stop/end/complete |
| 15 | Mute | Toggle |
| 16 | Exit video | Stop/end/complete |
| 17 | Move up | Move |
| 18 | Move down | Move |
| 19 | Move left | Move |
| 20 | Move right | Move |
| 21 | Select show | Select, Start/go/to |
| 22 | Change profile | Select, Modify |
| 23 | View full screen | View |

Those gestures having the highest agreement scores for each of the defined tasks were considered as part of the grammar, graphically represented for quick identification, textually described and characterized in terms of an anthropomorphic notation [7], for later using a formal notation in order to allow their processing and inclusion as part of a classifier. The results of the performed experiment and the proposed gesture grammar are described in the next section.

## 4. A Full-body Gesture Grammar for Interaction considering Abstract Tasks

Once the experiment was prepared and the 90 participants were recruited, they were divided into 6 groups of 15 users each, considering balance in terms of gender, age and whether they were left-handed or right-handed. Each of the groups was meant to fol-low the same protocol for the experiment as introduced in the previous section, with the particularity of having to interact with the system using different parts of their bodies. The first group was instructed to use only facial expressions, the second group was able to use only movement from neck and shoulders, the third group could only provide gestures using their arms, group number 4 was only allowed to use their hands, group 5 used their legs for interacting, and group six had to use only their feet. The results of the performance of the experiment are addressed in the following subsections.

### *4.1. Obtained Grammar for Interaction with a Streaming Platform and its Correspondence with Abstract Tasks*

As a total, 124 different gestures were identified during the elicitation study. Such gestures and combinations of them, were used by the participants for expressing all of the commands within the streaming platform. Regarding facial expressions(Table 3), 16 gestures were identified, grouped according to the tasks in which they were used, labeled, described and characterized.

When studying gestures made using only neck and shoulders, we selected 21 of the samples that users provided in Table 4. As it was done with the facial expressions, this set of gestures was also labeled, grouped, described and characterized.

After analyzing gestures that subjects provided using their arms, considering only motion from shoulders, elbows and wrists, we selected and decomposed them in 22 different gestures as presented in Table 5.

Table 3
Facial expressions obtained in the gesture elicitation study.

| **Gesture** | **Tasks** | **Description** | **Characterization** |
|---|---|---|---|
| E001 | 1, 2, 3, 6, 7, 8, 11, 13, 17, 21, 22 | Move tongue up | Move tongue up |
| E002 | 1, 2, 3, 6, 7, 11, 13, 18, 21, 22 | Move tongue down | Move tongue down |
| E003 | 1, 2, 6, 7, 11, 13, 19, 21, 22 | Move tongue left | Move tongue left |
| E004 | 1, 2, 6, 7, 11, 13, 20, 21, 22 | Move tongue right | Move tongue right |
| E005 | 1, 2, 6, 7, 11, 13, 21, 22 | Keep tongue out | Keep tongue straight |
| E006 | 2 | Blink with right eye | Minimize right eyelid distance |
| E007 | 4, 12 | Double blink with right eye | Minimize right eyelid distance twice |
| E008 | 5, 11 | Inflate right cheek | Right cheek inflation |
| E009 | 6, 7, 13 | Open mouth | Increase lip distance |
| E010 | 8 | Inflate right cheek once | Right cheek inflation, then right cheek contraction |
| E011 | 9 | Inflate left cheek once | Left cheek inflation, then left cheek contraction |
| E012 | 10, 14 | Smile | - |
| E013 | 11 | Inflate left cheek | Left cheek inflation |
| E014 | 15 | Close mouth hiding lips | Minimize lip distance |
| E015 | 16 | Frown | - |
| E016 | 23 | Blow a kiss | - |

Table 4
Gestures from neck and shoulders obtained in the gesture elicitation study.

| **Gesture** | **Tasks** | **Description** | **Characterization** |
|---|---|---|---|
| N001 | 1, 6 | Raise right arm to the front for selection | Right shoulder flexion |
| N002 | 2, 5 | Raise right shoulder twice | Move right shoulder up twice |
| N003 | 2 | Raise left shoulder twice | Move left shoulder up twice |
| N004 | 4 | Nod twice | Neck flexion then neck extension twice |
| N005 | 7 | Rotate neck to the right | Neck right lateral rotation |
| N006 | 8 | Raise right shoulder | Move right shoulder up |
| N007 | 8 | Lower left shoulder | Move left shoulder down |
| N008 | 9 | Raise left shoulder | Move left shoulder up |
| N009 | 9 | Lower right shoulder | Move right shoulder down |
| N010 | 10, 21 | Tilt head back twice | Neck extension twice |
| N011 | 11, 16, 23 | Take right shoulder back | Move right shoulder back |
| N012 | 11, 16, 23 | Take left shoulder back | Move left shoulder back |
| N013 | 12 | Nod once | Neck flexion then neck extension |
| N014 | 13 | Raise right arm to the front twice | Right shoulder flexion twice |
| N015 | 14 | Tilt head to the right twice | Neck right lateral flexion twice |
| N016 | 15 | Move head from left to right twice | Neck left lateral flexion, then neck right lateral flexion twice |
| N017 | 1, 2, 3, 6, 7, 17, 21, 22 | Move head back | Neck extension |
| N018 | 1, 2, 3, 6, 7, 18, 21, 22 | Move head forward | Neck flexion |
| N019 | 1, 2, 6, 7, 19, 21, 22 | Move head left | Neck left lateral flexion |
| N020 | 1, 2, 6, 7, 20, 21, 22 | Move head right | Neck right lateral flexion |
| N021 | 22 | Move head from left to right | Neck left lateral flexion then neck right lateral flexion |

Table 5

Gestures from arms obtained in the gesture elicitation study.

| Gesture | Tasks | Description | Characterization |
|---|---|---|---|
| A001 | 1, 4, 12, 21 | Flex right elbow | Right elbow flexion |
| A002 | 1, 2, 3, 4, 7, 8, 9, 10, 12, 14, 15, 16, 17, 18, 19, 20, 21, 23 | Extend right arm to the front | Right shoulder flexion and right elbow extension |
| A003 | 2, 16 | Using right arm, draw a circle in the air to the right | Shoulder external rotation |
| A004 | 3 | Move right arm up | Right shoulder flexion |
| A005 | 3, 15 | Move right arm down | Right shoulder extension |
| A006 | 5 | Extend right arm to the front flexing the elbow | Right shoulder flexion and right elbow flexion |
| A007 | 5 | Rotate right arm to the left | Right shoulder internal rotation |
| A008 | 13, 22 | Extend right arm to the right flexing the elbow | Right shoulder abduction and right elbow flexion |
| A009 | 6, 13 | Rotate right arm to the front | Right shoulder internal rotation |
| A010 | 7 | Rotate the right wrist to the front twice | Right wrist flexion twice |
| A011 | 8, 14 | Rotate the right wrist to the right twice | Right wrist rotation to the right twice |
| A012 | 9 | Rotate the right wrist to the left twice | Right wrist rotation to the left twice |
| A013 | 10 | Flex and extend the right elbow twice | Right elbow flexion, then right elbow extension twice |
| A014 | 11 | Put both arms together creating an 'X' | Right shoulder flexion, left shoulder flexion, right elbow flexion and left elbow flexion |
| A015 | 15, 18, 21 | Flex right wrist down | Right wrist flexion |
| A016 | 17 | Flex right wrist up | Right wrist extension |
| A017 | 19 | Flex right wrist to the left | Right wrist ulnar deviation |
| A018 | 20 | Flex right wrist to the right | Right wrist radial deviation |
| A019 | 22 | Rotate left arm to the front | Left shoulder internal rotation |
| A020 | 23 | Extend left arm to the front | Left shoulder flexion and left elbow extension |
| A021 | 23 | Rotate right arm to the right. | Right shoulder external rotation |
| A022 | 23 | Rotate left arm to the left. | Left shoulder external rotation |

In the same way, the 23 gestures from the hands included in Table 6 were gathered during the experiment.

After completion of the analysis of the upper body, the experiment continued with the analysis of lower limbs. Table 7 presents the 27 gestures from legs, considering hips, knees and ankles, on which use for interacting with the streaming platform subjects agreed.

Finally, in order to complete the study, gestures made using the feet were reviewed. Table 8 summarizes the information gathered for the 15 selected gestures for interacting using feet.

A gesture grammar allows not only the identification of gestures for interaction, but it also enables user training through the use of graphical representations. Fig. 2 reproduces an excerpt of the complement of the non-formal representation of the gesture grammar corresponding to arms and legs gestures. If the informal representation of the gesture grammar, which was obtained considering the data gathered in the elicitation study, is useful for proposing an interaction mechanism and for user interfaces engineering, the objective of our proposal is aimed at automation of the gesture recognition and classification training processes focusing on OSL, a formal representation for storing and processing the incoming data is required. The next subsection describes the proposed notation for this purpose.

### *4.2. Formal Description of the Gesture Grammar*

Enriching the proposal in [7] with a notation for formally describing gestures and hence enabling the storage and processing of incoming data for further feeding to a classification or recognition stage, in this paper we define an XML file with a structure such that complies with anthropomorphic metrics and contains the possible motion and body parts according to the capability model [17]. The XML structure considers both eyes' point of fixation, vertical rotation, horizontal rotation, and eyelid distance, which is particularly useful for expressing head and shoulders gestures [24]; the brow head position, brow arch position, and the brow tail position of both eye-brows; nasal tip position, nasal

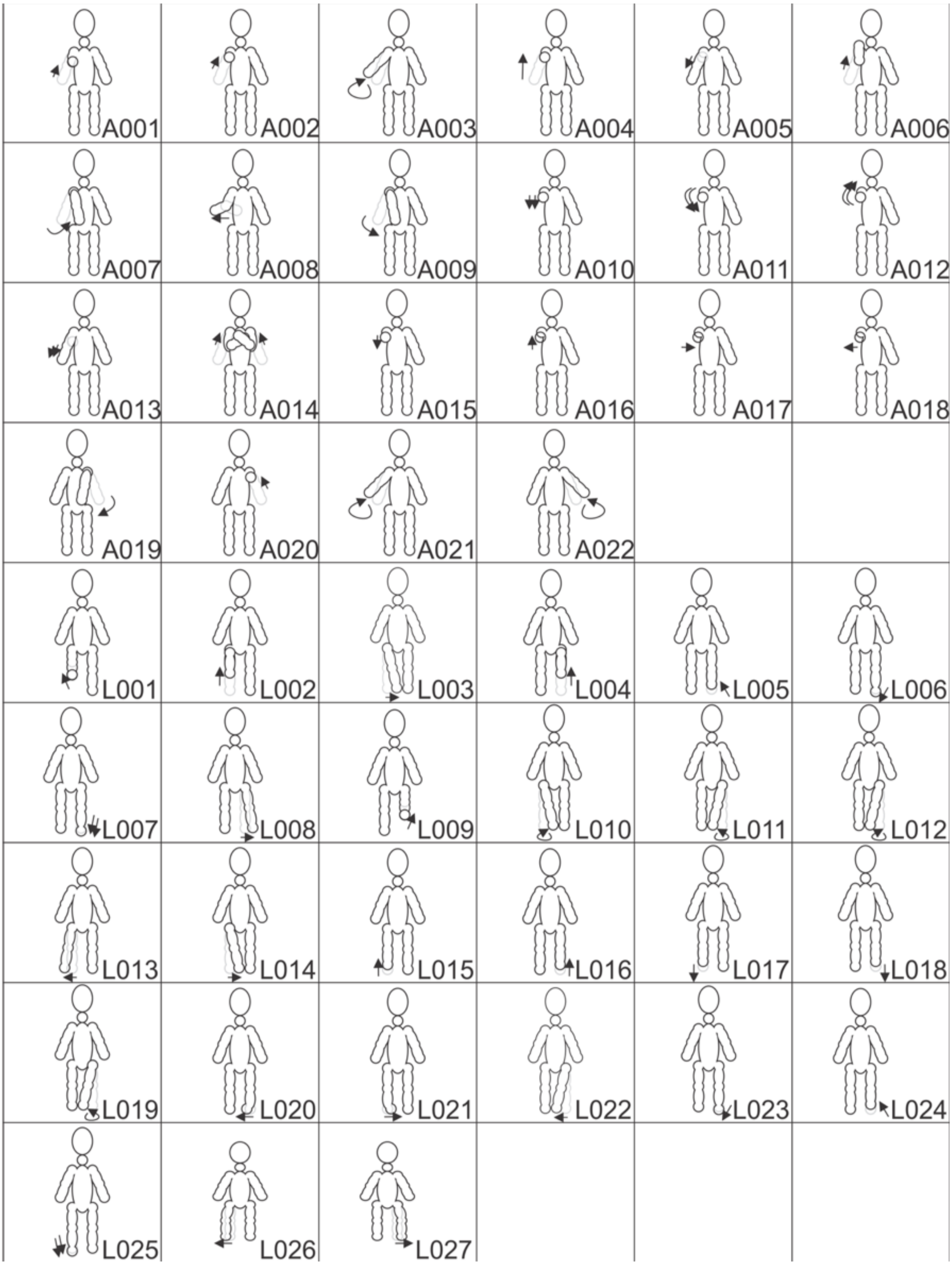


Fig. 2. Excerpt of the non-formal graphical representation of the identified gestures from arms and legs.

Table 6
Gestures from hands obtained in the elicitation study.

| Gesture | Tasks | Description | Characterization |
|---|---|---|---|
| H001 | 1, 11, 12, 13, 15, 22 | Open right hand with the palm towards the system | Right wrist extension and all fingers' joints extended |
| H002 | 1 | Put all finger tips together | Flexion of the A joints of all fingers in the right hand |
| H003 | 2 | Make 'Ok' sign to the system with the right hand | Extension of all joints of fingers 1, 2 and 3, and flexion of all joints of fingers 4 and 5 making their tips touch |
| H004 | 3, 17 | Open right hand with the palm up | Extension of all joints in the fingers of the right hand, supination of the right forearm |
| H005 | 3, 18 | Open right hand with the palm down | Extension of all joints in the fingers of the right hand, pronation of the right forearm |
| H006 | 3, 17 | Move the hand up | Move hand up |
| H007 | 3, 18 | Move the hand down | Move hand down |
| H008 | 4 | Close left fist leaving index and middle finger together extended | Flexion of all joints in fingers 1, 2 and 5 of left hand, and extension of all joints in fingers 3 and 4 of left hand |
| H009 | 5, 19 | With the right palm open, move hand to the left | Extension of all joints in the fingers of the right hand, supination/pronation of the right forearm and move hand to the left |
| H010 | 6 | Using right hand, extend index finger to the front | Flexion of all joints in fingers 1, 2, 3 and 5 of right hand, extension of all joints in finger 4 |
| H011 | 6 | Using left hand, extend index finger to the front | Flexion of all joints in fingers 1, 2,3 and 5 of left hand, extension of all joints in finger 4 |
| H012 | 7 | With right hand, make an 'L' sign | Flexion of all joints in fingers 1, 2 and 3 of the right hand, and extension of all the joints in fingers 4 and 5 of the right hand |
| H013 | 8 | Close right fist leaving index and thumb extended, touch twice thumb tip using index tip | Flexion of all joints in fingers 1, 2 and 3 of the right hand, and extension of all the joints in fingers 4 and 5 of the right hand, then flexion of all joints in fingers 4 and 5 making their tips touch twice |
| H014 | 9, 21 | Close right fist leaving index and middle finger together extended | Flexion of all joints in fingers 1, 2 and 5 of the right hand, extension of all joints in fingers 3 and 4 of the right hand |
| H015 | 10 | Using right hand, make 'thumbs up' sign | Flexion of all joints in fingers 1, 2, 3 and 4 of the right hand, extension of all the joints in finger 5 of the right hand |
| H016 | 11, 13, 15 | Close right fist | Flexion of all of the joints in all the fingers of right hand |
| H017 | 14, 20 | With the right palm open, move hand to the right | Extension of all joints in the fingers of the right hand, supination/pronation of the right forearm and move hand to the right |
| H018 | 16 | Close right fist with the palm towards the system | Flexion of all joints in all fingers of the right hand and extension of the right wrist |
| H019 | 16 | Move right hand down | Move hand down |
| H020 | 22 | Lower finger tips | Flexion of A and B joints in all fingers of the right hand |
| H021 | 23 | Open left hand with the palm towards the system | Left wrist extension and all fingers' joints extended |
| H022 | 23 | Move right hand to the right | Move right hand to the right |
| H023 | 23 | Move right hand to the left | Move left hand to the left |

ridge wrinkling, left nostril flaring, and right nostril flaring of the nose; the lip distance and position of the mouth; the tip position, curling, turning and folding of the tongue; and the inflation and contraction of both cheeks as elements related to facial expression. For the neck gestures, the structure allows to record information from flexion, extension, right lateral rotation, left lateral rotation, right lateral flexion, and left lateral flexion. In a similar way, for the upper limbs, the proposed structure contemplates the input of information regarding flexion, extension, abduction, adduction, external rotation, and internal rotation of both shoulders; flexion, extension, and hyperextension of both elbows; supination and pronation of both forearms; flexion, extension, radial deviation, and ulnar deviation of both wrists; supination and pronation of both hands; and flexion, extension, adduction, abduction, and hyperextension of finger joints. Finally, regarding the lower limbs, the proposed structure allows entering information of flexion, extension, adduction, abduction, external rotation, and

Table 7
Gestures from legs obtained in the gesture elicitation study.

| Gesture | Tasks | Description | Characterization |
|---|---|---|---|
| L001 | 1, 6 | Flex right knee backwards | Right hip extension and right knee flexion |
| L002 | 2, 10 | Flex right knee to the front | Right hip flexion and right knee flexion |
| L003 | 2 | Rotate right knee to the left | Right hip internal rotation |
| L004 | 3, 13, 14 | Flex left knee to the front | Left hip flexion and left knee flexion |
| L005 | 3 | Raise left knee | Left hip flexion |
| L006 | 3 | Lower left knee | Left hip extension |
| L007 | 4 | Step twice using left foot | Left hip flexion, left knee extension and left ankle plantar flexion twice |
| L008 | 5, 16, 19, 22 | Move left leg to the left | Left hip abduction and left knee extension |
| L009 | 7 | Flex left knee backwards | Left hip extension and left knee flexion |
| L010 | 8 | Draw a circle to the right on the floor using right leg extended | Right hip external rotation |
| L011 | 9 | Draw a circle to the left on the floor using left leg extended | Left hip external rotation |
| L012 | 10 | Rotate left knee to the right | Left hip internal rotation |
| L013 | 11, 20 | Move right leg to the right | Right hip abduction and right knee extension |
| L014 | 11 | Move right leg to the left | Right hip adduction and right knee extension |
| L015 | 12 | Raise right tiptoe | Right ankle dorsiflexion |
| L016 | 12 | Raise left tiptoe | Left ankle dorsiflexion |
| L017 | 12 | Lower right tiptoe | Right ankle plantar flexion |
| L018 | 12 | Lower left tiptoe | Left ankle plantar flexion |
| L019 | 13 | Rotate left knee to the left | Left hip external rotation |
| L020 | 15 | Move left tiptoe to the right | Left ankle inversion |
| L021 | 15 | Move right tiptoe to the left | Right ankle inversion |
| L022 | 16, 22 | Extend left leg to the right | Left hip adduction and left knee extension |
| L023 | 17 | Move left leg to the front | Left hip flexion and left knee extension |
| L024 | 18 | Move left leg backward | Left hip extension and left knee extension |
| L025 | 21 | Step twice using right foot | Right hip flexion, right knee extension and right ankle plantar flexion twice |
| L026 | 23 | Sitting down, move right leg to the right | Sitting down (right hip flexion, right knee flexion, left hip flexion and left knee flexion), right hip abduction |
| L027 | 23 | Sitting down, move left leg to the left | Sitting down (right hip flexion, right knee flexion, left hip flexion and left knee flexion), left hip abduction |

internal rotation of the hips; flexion and extension of both knees; dorsiflexion, plantar flexion, eversion, inversion, pronation, supination, lateral rotation, and medial rotation of both ankles; and flexion and extension of each of the five toes of both feet. By having a structure for formal representation of gestures, it is possible to manually – or automatically – record and process data coming from a source for different purposes, such as automated analysis of guidelines [22]. Fig. 3 instantiates the A006 gesture in the proposed XML structure. The information encoded in the XML file may express degrees, centimeters, or relative metrics. A discussion on the use of the proposal and regarding additional results that were gathered during the realization of the experiment is presented in the following section.

## 5. Discussion

The grammar presented in the previous section may be used for the achieving the engineering of gesture-based user interfaces. The association between the abstract tasks defined in [11] and specific commands within the analyzed domain, as well as the identification of gestures which users relate to those commands, allows generalization of the obtained gesture grammar. However, it is important to point out that the nature of gestural user interfaces makes it necessary to consider cognitive, cultural and linguistic aspects for customization purposes [18]. The population which participated in the experiment consist-ed of 47 female and 43 male subjects, all of them Mexicans aged between 16 and 57 years ($M$=38.19, $SD$=11.88).

Table 8
Gestures from legs obtained in the elicitation study.

| Gesture | Tasks | Description | Characterization |
|---|---|---|---|
| F001 | 1, 2, 6, 7, 17, 21, 22 | Move right foot to the front | Move right foot to the front |
| F002 | 1, 2, 6, 7, 18, 21, 22 | Move right foot backwards | Move right foot back |
| F003 | 1, 2, 6, 7, 9, 19, 21, 22 | Move right foot to the left | Move right foot to the left |
| F004 | 1, 2, 6, 7, 8, 14, 20, 21, 22, 23 | Move right foot to the right | Move right foot to the right |
| F005 | 23 | Move left foot to the left | Move left foot to the left |
| F006 | 1, 2, 6, 7, 21, 22 | Step twice | Flexion of right foot toe 5 twice |
| F007 | 3, 14 | Move right tiptoe up | Dorsiflexion of right ankle |
| F008 | 3 | Move right tiptoe down | Plantar flexion of right ankle |
| F009 | 4, 12, 16 | Move right tiptoe up twice | Dorsiflexion of right ankle twice |
| F010 | 5, 13 | Tilt right foot to the right | Supination of the right ankle |
| F011 | 6, 7 | Tilt right foot to the left | Pronation of the right foot |
| F012 | 10 | Move right toe up | Extension of right foot toe 5 |
| F013 | 11 | Move right tiptoe to the left | Inversion of right ankle |
| F014 | 11 | Move left tiptoe to the right | Inversion of left ankle |
| F015 | 15 | Flex all toes from right foot | Flexion of both joints in all foot toes of right foot |

During the experiment, it was possible to confirm that there are many-to-one mappings from concepts to gestures and vice versa as stated in the literature [9]. To observe this, notice that for the 23 studied tasks, users provided a total of 2070 gestures (345 gestures for each of the six groups). From those gestures, 124 gestures were selected according to their agreement scores and after disambiguation and division of complex gestures into simple, atomic ones. Besides, as it may be seen in Tables 3-8, some of the gestures are used for accomplishing several tasks while others are specifically related. Just after providing each gesture, users were asked to give a score between 1 and 10, according how difficult, memorable and descriptive they thought the given gesture was for representing the required command. We expected to see relationships existing between the response times and users' grades for each gesture and the agreement scores or final assessment, however, in a similar way as reported in [7] there does not seem to be any.

As part of the followed experimentation protocol, users were asked to fill in the IBM CSUQ [19] in order to get data regarding their appreciation towards the system in terms of usability. The IBM CSUQ instrument is composed of 19 questions, with a 7-point Likert scales concerning system use (Q1-Q8), information quality (Q9-Q15), UI quality (Q16-Q18), and general feedback to the system (Q19).

Results of the applied IBM CSUQ reported that interaction using lower limbs was less satisfying for users as they assessed it with an average of 5.13 out of 7 points ($SD$=1.17) while after simulating interaction using upper limbs, subjects gave an average score of 5.8 out of 7 ($SD$=.79). This difference may be caused due to lack of expressiveness of the lower limbs, but also by the fact that subjects are used to use their hands for interacting with applications rather than their legs or feet. Along with the quantitative information which was gathered, i.e., response time, gesture score, agreement score, and usability evaluation, it was possible to get qualitative insights regarding the experiment. During the experiment, participants be-longing to the neck and shoulders, legs, and feet groups complaint about how difficult was for them to use only such parts of their bodies to interact with an application and for achieving complex tasks. It was necessary to introduce them to a scenario on which interacting with limited parts of the body was not only desired but critical, e.g., due to disability, for rehabilitation purposes, and due to characteristics of the context.

Regarding the resultant proposal in terms of the gesture grammar and its non-formal and formal notation, it is possible to use it already for user training purposes, for gesture-based applications design, and for starting the training of an automatic gesture classifier. Nevertheless, some clarifications need to be done. The XML file structure was created in compliance with the user capability model from [17], but its extension with supplementary interesting features is possible through the addition of tags into its body. Simplification of the structure is also possible through tags removal. This reduction may not only apply to complete blocks of in-


Fig. 3. Formal representation of the A006 gesture along with its graphical representation.

formation (researchers may remove facial expressions from the file if not interested on tracking such type of interaction), but also to specific features. Another interesting consideration is on that the proposed structure for formal representation of gestures includes opposite information, e.g., flexion-extension, inflation-contraction, abduction-adduction. These opposite tags may not be present in a final instantiation as they have repetitive information and their use depends on the approach followed.

The representation of static gestures may be achieved through the use of one single XML file, while for dynamic gestures it is necessary to group a series of files according to different times on which significant changes occur. Additional tags should be added to the basic structure in order to support the inclusion of the order or appearance and time or speediness.

In order to achieve the general objective of supporting OSL as mentioned in the introduction, the formal notation enables a gesture to be understandable by computers, however, so far it would not be possible to complete a classifier training process with one single sample. Thus, we identified the need for defining a mechanism for using data gathered for one gesture for training an automatic gesture classifier. To this purpose, we propose to apply slight variations on the formally represented information in order to stimulate multiple samples for carrying out the training process.

The definition of such transformations implies the knowledge of motion restrictions, e.g. maximum rotation degrees for a joint, maximum distance between joints; as well as of tolerance levels in which one gesture may be varied without generating ambiguity. In the following section, conclusions and the future work to be addressed are introduced.

## 6. Conclusion and Future Work

In this paper, we report the results of a full-body gesture elicitation study performed analyzing tasks within the Netflix streaming platform and aligned to abstract interactive tasks. Such results were presented through a non-formal representation using a textual description and a motion characterization, and then using a formal notation by means of a flexible XML structure containing tags for recording information describing the characteristics of the gestures depending on the followed approach. For the use of this formal notation, it is necessary to assume the existence of a base position from which all of the recorded movements may be relative to.

This work includes the presentation of quantitative and qualitative insights as well, in terms of usability, number of gestures gathered, agreement scores, grades provided for each gesture, and subjects' considerations and limitations. To the best of our knowledge, there are no other proposals considering abstract tasks during elicitation studies nor defining transformations on single samples for addressing OSL. To compare the obtained gestures with other proposals, it will be necessary to analyze such related works considering the tasks that were studied and associating them to abstract tasks. This, along with the definition of a complete set of trans-formations and variations for addressing OSL are identified as future work.

## Acknowledgments

Jean Vanderdonckt acknowledges the support of the project EDBS Emerald Casting Assistant funded under Convention No. 7901 of MecaTech Competitivity Pole, Walloon Region, Belgium.

## References


[1] Weng, F. & Zhang, Q. (2011). U.S. Patent No. 7,930,168. Washington, DC: U.S. Patent and Trademark Office.

[2] Anders, S. What are you really saying? Physician leadership journal, 2(2), 82.

[3] Schlömer, T., Poppinga, B., Henze, N., & Boll, S. (2008, February). Gesture recognition with a Wii controller. In Proceedings of the 2nd international conference on Tangible and embedded interaction (pp. 11-14). ACM.

[4] Liang, R. H., & Ouhyoung, M. (1998, April). A real-time continuous gesture recognition system for sign language. In Proceedings third IEEE international conference on automatic face and gesture recognition (pp. 558-567). IEEE.

[5] Daniloff, J. K., Noll, J. D., Fristoe, M., & Lloyd, L. L. (1982). Gesture recognition in patients with aphasia. Journal of Speech and Hearing Disorders, 47(1), 43-49.

[6] Jia, P., Hu, H. H., Lu, T., & Yuan, K. (2007). Head gesture recognition for hands-free control of an intelligent wheel-chair. Industrial Robot: An International Journal, 34(1), 60-68.

[7] Céspedes-Hernández, D., & González-Calleros, J. M. (2019). A methodology for gestural interaction relying on user-defined gestures sets following a one-shot learning approach. Journal of Intelligent & Fuzzy Systems, 36(5), 5001-5010.

[8] Wang, P., Liu, L., Shen, C., Huang, Z., van den Hengel, A., & Tao Shen, H. (2017). Multi-attention network for one shot learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2721-2729).

[9] Mitra, S., & Acharya, T. (2007). Gesture recognition: A survey. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 37(3), 311-324.

[10] Obaid, M., Häring, M., Kistler, F., Bühling, R., & André, E. (2012, October). User-defined body gestures for navigational control of a humanoid robot. Proceedings of International Conference on Social Robotics (pp. 367-377). Springer, Berlin, Heidelberg.

[11] Constantine, L. L. (2003, June). Canonical abstract prototypes for abstract visual and interaction design. Proceedings of International Workshop on Design, Specification, and Verification of Interactive Systems DSV-IS '03 (pp. 1-15). Springer, Berlin, Heidelberg.

[12] Vatavu, R. D., & Zaiti, I. A. (2014, June). Leap gestures for TV: insights from an elicitation study. Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video (pp. 131-138). ACM.

[13] Connell, S., Kuo, P. Y., Liu, L., & Piper, A. M. (2013, June). A Wizard-of-Oz elicitation study examining child-defined gestures with a whole-body interface. Proceedings of the 12th International Conference on Interaction Design and Children (pp. 277-280). ACM.

[14] Ruiz, J., Li, Y., & Lank, E. (2011, May). User-defined motion gestures for mobile interaction. Proceedings of the ACM International Conference on Human Factors in Computing Systems CHI' 11 (pp. 197-206). ACM Press, New York.

[15] Dow, S., Lee, J., Oezbek, C., MacIntyre, B., Bolter, J. D., & Gandy, M. (2005, April). Wizard of Oz interfaces for mixed reality applications. Proceedings of CHI'05 Extended Abstracts on Human Factors in Computing Systems (pp. 1339-1342). ACM Press, New York.

[16] Höysniemi, J., Hämäläinen, P., & Turkki, L. (2004, June). Wizard of Oz prototyping of computer vision-based action games for children. Proceedings of the 2004 conference on Interaction design and children: building a community (IDC '04) ACM Press, New York, 27–34.

[17] Kaklanis, N., Moschonas, P., Moustakas, K., & Tzovaras, D. (2013). Virtual user models for the elderly and disabled for automatic simulated accessibility and ergonomy evaluation of designs. Universal access in the information society, 12(4), 403-425.

[18] Chandler, J., & Schwarz, N. (2009). How extending your middle finger affects your perception of others: Learned movements influence concept accessibility. Journal of Experimental Social Psychology, 45(1), 123-128.

[19] Lewis, J.R. (1995). IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use. International Journal of Human-Computer Interaction, Vol. 7, No. 1, 1995, pp. 57-78. https://doi.org/10.1080/10447319509526110.

[20] Vanderdonckt, Jean and Roselli, Paolo and Pérez-Medina, Jorge Luis. !FTL, an Articulation-Invariant Stroke Gesture Recognizer with Controllable Position, Scale, and Rotation Invariances. Proceedings of the 20th ACM International Conference on Multimodal Interaction (ICMI '18). ACM Press, 2018, 125–134.

[21] Schipor, Ovidiu-Andrei and Vatavu, Radu-Daniel and Vanderdonckt, Jean. Euphoria: A Scalable, event-driven architecture for designing interactions across heterogeneous devices in smart environments Information and Software Technology, 109, 43–59.

[22] Vanderdonckt, Jean and Beirekdar, Abdo Automated Web Evaluation by Guideline Review Journal of Web Engineering, 4(2), 2005, 102–117.

[23] Wobbrock, Jacob O. and Morris, Meredith Ringel and Wilson, Andrew D. User-defined Gestures for Surface Computing Proceedings of ACM International Conference on Human Factors in Computing Systems (CHI '09) ACM Press, New York, 2009, 1083–1092.

[24] Vanderdonckt, Jean and Magrofuoco, Nathan and Kieffer, Suzanne and Pérez, Jorge and Rase, Ysabelle and Roselli, Paolo and Villarreal, Santiago Head and Shoulders Gestures: Exploring User-Defined Gestures with Upper Body In Design, User Experience, and Usability. User Experience in Advanced Technological Environments Springer International Publishing, Cham, 2019, 192–213.

[25] Vatavu, Radu-Daniel and Wobbrock, Jacob O. Formalizing Agreement Analysis for Elicitation Studies: New Measures, Significance Test, and Toolkit Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15) ACM Press, New York, 2015, 1325–1334.

[26] Calvary, Gaëlle and Coutaz, Joëlle and Thevenin, David and Limbourg, Quentin and Bouillon, Laurent and Vanderdonckt, Jean A Unifying Reference Framework for multi-target user interfaces Interacting with Computers, 15(3), 2003, pp. 289-308.