An Ontology for Reasoning on Body-based Gestures

Mehdi Ousmer, Jean Vanderdonckt
Université catholique de Louvain, LouRIM Institute
Louvain-la-Neuve, Belgium
{mehdi.ousmer,jean.vanderdonckt}@uclouvain.be

Sabin Buraga
Alexandru Ioan Cuza Univ. of Iasi, Fac. of Comp. Science
Iasi, Romania
busaco@info.uaic.ro

ABSTRACT

Body-based gestures, such as acquired by Kinect sensor, today benefit from efficient tools for their recognition and development, but less for automated reasoning. To facilitate this activity, an ontology for structuring body-based gestures, based on user, body and body parts, gestures, and environment, is designed and encoded in Ontology Web Language according to modelling triples ⟨subject, predicate, object⟩. As a proof-of-concept and to feed this ontology, a gesture elicitation study collected 24 participants × 19 referents for IoT tasks = 456 elicited body-based gestures, which were classified and expressed according to the ontology.

CCS CONCEPTS

• **Human-centered computing** → **Interaction devices**; **Gestural input**; *User studies*; • **Hardware** → *Sensors and actuators*;

KEYWORDS

Gesture interaction; Microsoft Kinect; Natural gestures; Ontology Web Language; Resource Description File.

ACM Reference Format:
Mehdi Ousmer, Jean Vanderdonckt and Sabin Buraga. 2019. An Ontology for Reasoning on Body-based Gestures. In *ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS '19), June 18–21, 2019, Valencia, Spain.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3319499.3328238

1 INTRODUCTION

Many sensors, like the Microsoft Kinect and LeapMotion, acquire full-body movements for a wide range of gestures (e.g., from hand-and-arm to full body) [12] to trigger actions in different contexts of use: for different users (e.g., civilian, military people, people with disabilities), in different environments (indoor vs outdoor, at home, in office, in-the-wild) for a wide range of applications (e.g., Human-Robot Interaction [5], games [21]). So many techniques [5, 11] have successfully demonstrated accurate recognition of Kinect-based gestures that their recognition no longer represents a significant challenge. And so does their incorporation into the design and the development of gesture user interfaces: many environments exist for design and analysis of Kinect gestures [17, 18]. Although this is an almost solved problem, the human management of all these gestures for other activities, such as automated reasoning and the appropriate mapping between a gesture and the action, still remains.

When gestures are predefined with their meaning and actions. For instance, armies use a vocabulary of about 130 hand-and-arm gestures to produce visual signs [13] for which the mapping gesture–meaning/action is straightforward and standardized. When gestures are progressively discovered. Gesture elicitation studies [3, 23] are conducted to ask participants to naturally map an action, that is materialized by a *referent* to a corresponding gesture.

Large sets of Kinect-based gestures are adequately handled by automata for recognition and development, but not for their management for human purposes, such as editing, clustering, distance computing, aggregation, composition, organization, searching strategies in an integrated way, which are considered in automated reasoning. This raises the need to structure such gestures according to a consistent representation that enables such a reasoning, with *intrinsic properties* (i.e., related to the gestures themselves) and *extrinsic properties* (i.e., related to the context of use in which they are produced). Whereas intrinsic properties should be independent of any such sensor, extrinsic properties are usually captured outside the range of sensing, which may require a human operator, unless machine learning comes into rescue.

To address these questions, this paper brings the following contributions:

- A sensor-independent ontology of body-based contextual gestures, with intrinsic and extrinsic properties.
- Its implementation in Ontology Web Language (OWL) using RDF triples ⟨subject, predicate, object⟩, each denoted by Uniform Resource Identifiers (URIs), to be processed by knowledge modeling software.
- A proof-of-concept of this ontology based on 456 body-based gestures issued from a gesture elicitation study.

EICS '19, June 18–21, 2019, Valencia, Spain

ACM ISBN 978-1-4503-6745-5/19/06.
https://doi.org/10.1145/3319499.3328238

2 RELATED WORK

This section is divided into two parts: a review of selected work on the Kinect-like gestures and an overview of existing elicitation studies performed on the human body.

Kinect Gestures

Many systems efficiently recognize Kinect-based gestures. Cicirelli et al. [5] developed a Neural Network classifier to recognize ten gestures performed by various people to drive an interface robot. Three Kinect devices are co-located indoor to ensure the accuracy of gesture recognition. Whereas Dynamic Time Warping (DTW) was used for recognizing Kinect dynamic gestures, a Bayesian classifier is preferred for the static gestures and postures [11].

Based on a taxonomy and a notation expressing 3D hand gestures captured with a Kinect, 43 user-defined gestures resulted from a GES for 6 TV actions[4]: Turn TV on/off, Go to previous/next channel, open/close blinds. Six intrinsic properties express fine-grained hand gestures: hand location, shape, movement, and orientation. An ontology for Kinect gestures has been already devised for categorizing such gestures [7], which also introduces a formal notation for reasoning on gestures. Other notations come from linguistics [10] or from multimodal interaction modelling [19].

Gesture Elicitation Studies

Understanding users' preferences and behavior with new interactive technology right from the early stage empowers engineers with valuable information to shape a product's characteristics for more effective and efficient use. This process is referred to as *Gesture Elicitation Studies* (GES) [23]. They have become popular to understand users' preferences for gesture input for a variety of contexts of use. For instance, Gheran et al. [9] studied ring-based gestures and improved the elicitation based on its classification. Such studies have been conducted along the three dimensions of the *context of use*: users and tasks, platforms and devices, and environments. Gestures are typically elicited in a particular *environment* in which devices are determined, such as a home or a surgery unit [15]. Gestures can be also constrained by type, such as hand gestures [2], whole-body gestures [20].

Kinect gestures were also elicited in a Wizard-of-Oz approach with six children for 22 tasks including object manipulation, navigation, and spatial interaction [6]. The authors manually analyzed these gestures to discuss five aspects, i.e., the influence of 2D touch screens, contextual cues, preferences, alternate approaches, and allocentric vs egocentric approaches for object manipulation. The manual analysis of these gestures by three researchers, who came to an inter-judge agreement, represents a significant amount of work.

Ten web browser actions initiated a GES comparing gestures, speech, and multimodal interaction [18], which was a replication of the Web-on-the-Wall experiment. Twelve TV actions were used in an elicitation study for TV control [8], where a Kinect captured hands gestures. Thirty-four actions for a Picture Archiving and Communication System (PACS) were used in a Kinect-based gesture elicitation [15]: surgeons were instructed to draw on a drawing board gestures they preferred for these actions. For each gesture group (gestures that share similar context), each surgeon expressed a common gesture for context (action selection) and a specific gesture for the modifier, thus enabling to reuse the modifier gestures that belonged to different contexts. This results into an elicitation of contextual gestures. A more recent GES [16] captured Kinect-based gestures, mainly the upper body, of a human person collaborating with an avatar. The *human gesture continuum* suggests that any human limb capable of some mobility can be the source of a GES, either in isolation (e.g., the legs), or combined with subsequent limbs (e.g., the legs with the feet), or considering the whole body.

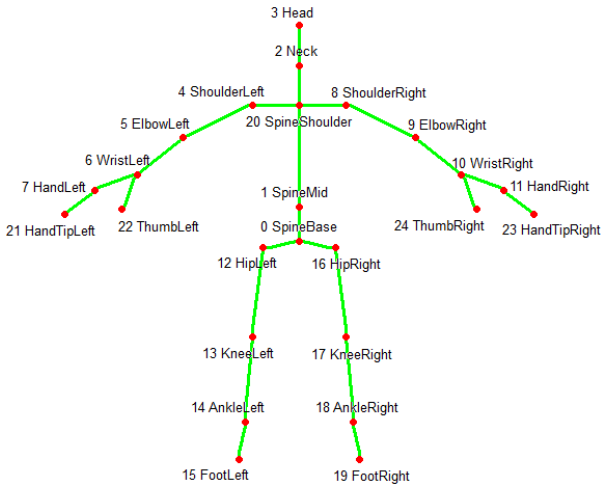

Figure 1: The 25 joints of a MS Kinect Skeleton.

These examples argue for the need of structuring body-based gestures in a way that is analyzable by some software. For example, to support the process of multimodal Kinect-based GES, KinectAnalysis [17] provides a "record-and-replay" metaphor where Kinect-based gestures are recorded, analyzed in real-time based on depth, audio and video streams, and visualized. It also offers KinectScript, a scripting language for querying recorded gestures and automating their analysis based on skeleton, distance, audio and gesture scripts. This paper is going towards the same direction in by attempting to systematically describe body-based gestures, with an extensible body-based gesture ontology that can be processed by automated reasoner. For example, KinectScript describes records of two Kinect-based skeletons with their right/left part, 4 positions, 2 distances, 20 joint terms (out of 25 of the Kinect, see Fig. 1), and two gesture names. Our ontology is aimed at expressing any body-based gesture, not just Kinect-based, for any set of skeletons, with an inheritance hierarchy of human limbs.

3 ONTOLOGY FOR BODY-BASED GESTURES

To represent body-based gestures in their context of use, an ontology was designed, including the user, the sensor, and the physical environment, and expressed in OWL, a W3C standard for expressing knowledge on the Web based on ⟨*subject*,*predicate*,*object*⟩ triples as defined in the Resource Description Framework (RDF) with 3 main classes (Fig. 2):

- *User*: general information regarding who the user is, plus observed anatomical body parts and joints: Body, BodyPart, Part, Limb, Arm, Leg, Bone, and Joint classes.
- *Sensor*: raw data provided by an acquisition device, such as a sensor like a Kinect or VicoVR[1].
- *Detected gestures and poses*: Gesture, GestureSegment, HandState, and Pose classes.

A *gesture* is made up of *strokes* that are binding *points*. A sensor captures a gesture as a suite of p_i=(x_i, y_i, z_i, w_i, t_i), where x_i, y_i, z_i, w_i are the 3/4D coordinates of each gesture point, t_i is the time stamp. To properly characterize gestures and perform associated tasks, the human body should be specified as formal components of the desired ontology. First, we define the joints as objects to be compared with respect to their relative positioning in a Cartesian space. From there, higher level relationships can be expressed based on Allen's interval algebra for temporal calculus [1]: `Left before Right`, `Left meets Right`, `Left overlaps Right`, `Left starts Right`, `Left finishes Right`, `Left during Right`, `Left and Right occur simultaneously`, `Right before Left`, `Right meets Left`, `Right overlaps Left`, `Right starts Left`, `Right during Left`, `Right finishes Left`. These thirteen relationships between two intervals can be consolidated them into the following relationships: `Above`, `Below`, `After`, `Before`, `ToTheLeftOf`, `ToTheRightOf`, `AtTheSameLength`, `AtTheSameHeight`, `AtTheSameDepth`, and `OverlapsWith`. Next, these joints are merged into segments to specify certain body poses, which results into a bone-like structure, establishing a human continuum from hands to feet. Because some poses may involve symmetric gestures, such as both arms raised above the head, mirroring segments can be tied together as one entity.

A *pose* represents the most fine-grained part of a gesture *G*, as it can be defined as small as just a simple comparison between two joints. Since manipulating and editing gestures is the primary goal of the ontology, the possibility of combining multiple atomic parts is important. For this purpose, the following logical constructs are provided: (1) merging two poses: "and" operation defined by the "&" symbol, (2) assuming the correctness of only one "or" denoted by the "|" symbol, and (3) matching the opposite of a pose: "not" logical operator specified by the "!" symbol. For example, a

[1]See http://www.vicovr.com, http://www.vicovr.io

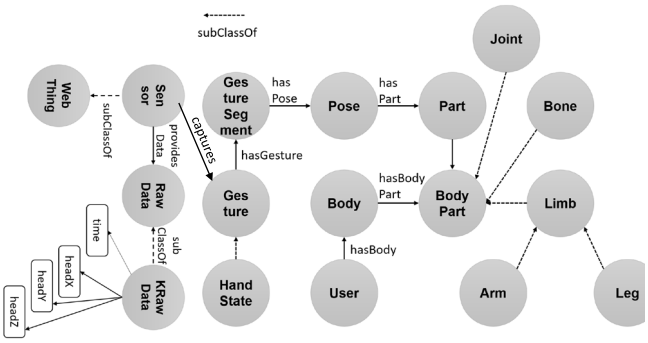

Figure 2: The ontology for body-based gestures.

given pose p could be considered as a composition of three sub-poses: $p = p_1\&(p_2|!p_3)$. To determine a joint's position relative to another one, it is necessary to determine the length between them in space. Since the distance between the wrist and the hand can be smaller than between hand and elbow, the difference in height and angle between the two should be proportional, resulting in the same confidence rate when comparing both. For maintaining this proportion, we needed to calculate the shortest distance between any two given joints and add the positions of each intermediary joint to the final ratio. To investigate the angle between two interconnected segments, the concept of *limb* is defined as the peripheral parts of the human body, such as arms and legs. We considered as limbs the connecting bones between hip and ankle (as feet) and the ones between shoulder and wrist (as arms). This allowed to analyze and combine the rotation and angle for hands and feet. A model of the pose is being created and then compared to the actual position of the body from a Kinect frame, resulting in a degree of confidence.

The `Sensor` class represents the generic sensor which is used to identify the user. The property `providesData` specifies the relation between a sensor and the provided raw data denoted by a generic `RawData` class. This raw data could be expressed by various RDF constructs, such as in our case a CSV file for each gesture and then transformed in RDF triples, which is expressed by `KRawData`, a subclass of `RawData`. An instance of `KRawData` stores numerical data about the position of observed body parts – e.g., head, neck, spine, shoulder (left and right), elbow (left and right), hand (left and right), and so on for every Cartesian axis. An excerpt is:

```
:kinect rdf:type Sensor ;
        providesData (:data1, :data2,..., :dataN).
:data1 rdf:type KRawData ;
        time  0.0110013^^xsd:double ;
        headX 0.0021632^^xsd:double ;
        headY 0.8330318^^xsd:double ;
        headZ 2.2899450^^xsd:double ;
        ...
        headLeftX -0.2059579^^xsd:double ;
```

In the next steps of processing, the detected gestures and poses will be specified by high-level ontological constructs described below. The object property `detectsUser` defines the relation between a `Sensor` instance and a detected user.

The `User` class is populated by a sub-set of properties coming from the `Person` class[2] of Schema.org, a collaborative initiative to structure common data on the Web. Additionally, the `Sensor` class could refer to the *Web Thing* concept defined by the Web of Things (WoT) Architecture[3]. In our case, could be considered as the digital representation of a physical IoT device, which also directly provides the interface for natural interaction. A *Web Thing* representation could be specified by RDF model, including metadata of interest.

The `Body` class is identified using the `trackingId` functional property, having as a value a RDF literal. It also includes multiple instances of the `BodyPart` class, specified with the property `hasBodyPart`, expressed in Turtle[4]:

```
Body rdf:type owl:Class .
trackingId rdf:type owl:DatatypeProperty ,
                    owl:FunctionalProperty ;
  rdfs:domain Body ; rdfs:range  rdf:Literal .
hasBodyPart rdf:type owl:ObjectProperty ;
  rdfs:domain Body ; rdfs:range  BodyPart .
```

The `BodyPart` class represents any elements of the human body and is a super-class of the `Limb`, `Bone`, and `Joint` sub-classes. `Limb` is itself a super-class of `Arm` and `Leg` subclasses and has at least one `Bone` instance is defined through the property `hasBone`. The `Arm` class has only two disjoint individuals belonging to the `RightArm` and `LeftArm` classes. Similar constructions are made for the `Leg` class. The `Bone` class is a sub-class of `BodyPart` and is formed of exactly two `Joint` instances via `hasJoint` of cardinality 2 (each bone has only two joints). The `Joint` specifies the human body joints and is a sub-class of `BodyPart`. The Kinect device promotes 25 individuals of this class (Fig. 2), of which KinectScript [17] exploits 22. Another device may introduce more or less instances leaving this class definition untouched in our ontology. The `Part` class denotes the tie between two instances of the `BodyParts` and an action specific to a certain context of use. This action is populated by the `Action` class defined by Schema.org model. The class is also enriched by properties defined in relation to itself: `isAbove`, `isBelow`, `isAtTheSameHeightOf`, `isAtTheSameLengthOf`, `isAtThe - SameDepthOf`, `isToTheRightOf`, `isToTheLeftOf`, and `over lapsWith`. These properties are useful to perform automatic reasoning on the ontology in order to detect complex or compound gestures and/or poses, depicting the relationships between different `BodyPart` instances. For example, the triple `RightHand isAbove Head` specifies that the right hand of the tracked user is placed above her head.

The `Pose` class represents a union of instances of the `Part` classes, defined through the object property `hasPart`, and specifies the relationship between different parts of the user's body. For example, a specific pose is denoted in RDF:

```
aPose rdf:type Pose ;
      hasPart RightHandToTheLeftOfRightElbow ;
      hasPart RightHandAboveRightElbow .
```

The class `GestureSegment` embodies a series of poses defined with the `hasPose` property – an example:

```
aGestureSegment rdf:type GestureSegment ;
  LeftHandBelowAtTheSameLengthWithLeftElbow .
```

The `Gesture` class defines a gesture with at least one `GestureSegment` instance by relying on the object property `hasGestureSegment`.

4 FEEDING THE ONTOLOGY

We conducted a GES following the original methodology [22, 23] to collect users' preferences for body-based gestures. We selected 19 referents common to IoT from current literature [9, 23] divided into three groups: a single *unary* action (i.e., Start player), 10 *binary* actions (i.e., turn on/off a device like a TV, the air conditioning, the light, the heating system, the alarm system, answer/end a phone call), and 4 *linear* actions (i.e., increase/decrease volume, go to next/previous item in a list, brighten/dim lights).

Participants

Twenty-four voluntary participants (11 Females, 13 Males; aged from 23 to 53 years, $M = 34.54$, $Md = 31$, $SD = 9.87$) were recruited for the study via a contact list in different organizations. The participants' occupations included employees, executives, and independents in domains such as administration, economics, science, and transportation. Usage frequencies were captured: computer, smartphone, tablet, game console, and Kinect devices. Participants reported frequent use of computers, smartphones, and some of Kinect.

Apparatus

The experiment took place in a usability laboratory to control the experiment. Sheets of paper were used for showing the referents to participants, each paper contained a referent depicted by a descriptive action text and two abstract images (one before the action, one after) without any reference to any particular system. All the gestures were recorded by a camera and a Kinect camera placed in front of the participant to capture their body based gestures. In order to keep the study focused, we asked the participants to minimise their lower-body movement, without any instrumentation.

[2] https://schema.org/Person
[3] https://w3c.github.io/wot-architecture/
[4] https://www.w3.org/TR/turtle/

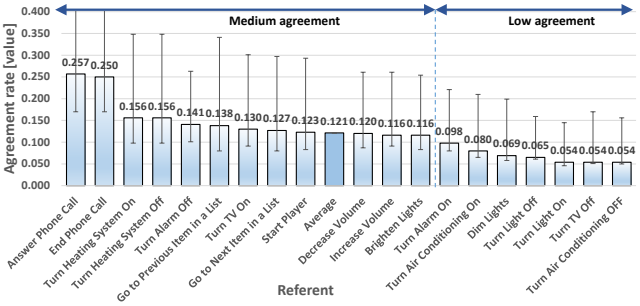

Figure 3: Referents in decreasing order of their agreement rate. Error bars show a confidence interval for $\alpha = .05$.

Procedure

Pre-Test phase. The participants were welcomed to the setting by the researchers and were first asked to sign an informed consent form compatible with GDPR regulation. Then, they were given information about the study and the general process of the experiment. They were also asked to fill a sociodemographic questionnaire. The researchers collected the sociodemographic data about each participant in order to use some of these parameters in the study. The questionnaire gives general information (e.g., age, gender, domain of activity) and questions about use of technologies.

Test phase. During this phase, the experimenter explained to participants what body interaction is all about, the following tasks that they had to perform, and the allowed types of gestures. Participants operated with the belief that no technological constraint was imposed in order to preserve the natural and intuitive character of the elicitation. Participants were presented with the 19 referents, i.e., actions to control objects in an IoT environment, for which they elicited a gesture that fits the referent well, is easy to produce and remember. Participants were instructed to remain as natural as possible. The referents were randomly ordered per participant based on a random number generator (www.random.org). The THINKING TIME between the first showing of the referent and the moment when the participant knew which gesture she would perform was measured in seconds with a stopwatch. After eliciting each gesture, the participant rated its GOODNESS-OF-FIT from 1 to 10 to express to what extent she thought her gesture was appropriate to the presented referent. Each session took approximately 45 min.

Post-test phase. At the end of each session, the participants filled in the IBM PSSUQ (Post-Study System Usability Questionnaire) [14], which enables them to express their level of satisfaction with the usability of the setup and the testing process. This questionnaire benefits from a $\alpha = 0.89$ reliability coefficient. Each question is measured using a 7-point Likert scale (1=strongly disagree to 7=strongly agree) and four measures are computed: *system usefulness* (1-8), *quality of the information* (9-15), *quality of the interaction* (16-18), and *system quality* (19).

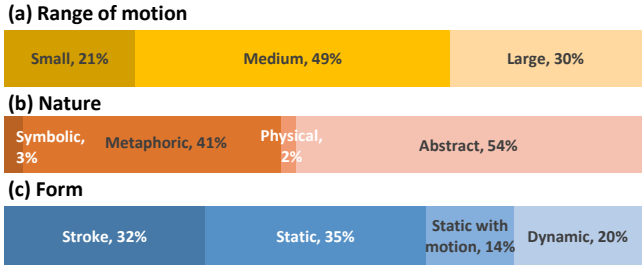

Figure 4: Distributions. See https://youtu.be/vngyHlVl7dM for examples.

Results and discussion

Agreement Rate. Fig. 3 depicts the referents in decreasing order of their agreement rate computed by AGaTE [22]. Overall, agreement scores and rates are medium in average magnitude, in particular for rates (which are the most demanding ones) between .257 and .116 for the global sampling ($M = .121$, $SD = .066$). Most referents belong to the *medium* range according to Vatavu and Wobbrock's method [22] to interpret the magnitudes of agreement rates. These results are very similar to the other rates reported in the GES literature [22]. Hence, our results fall inside medium consensus ($< .3$) category with their average in the same interval (highlighted bars in Fig. 3). To decide the consensus gesture, four criteria were considered: the agreement rate, the individual frequency of occurrence for each Referent, the associative frequency when two referents are symmetric (e.g., "Go to next channel" and "Go to previous channel") to take into account the consensus by pair, and the unicity.

Classification. All individual elicited gestures were classified according to several criteria [9, 21]: range of motion (Fig. 4a), nature (Fig. 4b), form (Fig. 4c), and hand distribution (Fig. 5). All participants were left handed and mainly used their dominant hand to issue each gesture (72%). When they switched to bimanual mode, they slightly preferred asymmetry (17%) over symmetry (11%). Based on these criteria, a classification of 53 individual (unique) gestures falling into 23 categories of gestures, each category potentially having sub-categories has been reproduced. Regarding the average THINKING TIME, there is no particular correlation between the thinking time for pairs of related referents. The GOODNESS OF FIT is distributed into six regions depending on its respective value and current interpretation. Overall, the value for most gestures (Max=8.89, Min=6.26) belongs to the "excellent" region ($v > 7$, 12/22=55%) or the "good" region ($v \in [5.5, 7]$, 9/22=41%) between 3.36 and 8.14 for the global sampling ($M = 6.78$, $SD = 1.63$). These results are above the average values: participants were particularly happy with the gestures they chose, which reinforces the acceptability of the elicited gestures. The main gesture categories are: swipe, rotate a control knob, tap, raise, remote control, point with finger, shape a phone, and press.

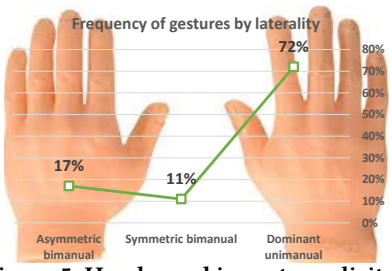

Figure 5: Hands used in gesture elicited.

5 CONCLUSION AND FUTURE WORK

An ontology for structuring body-based gestures, based on user, body and body parts, gestures, and environment, has been encoded in Ontology Web Language according to modelling triples ⟨subject, predicate, object⟩. As a proof-of-concept and to feed this ontology, a gesture elicitation study was conducted with twenty-four participants who elicited a set of 456 body-based gestures for 19 referents associated to frequent IoT tasks. These initially elicited gestures were then structured according to criteria to come up with a classification of 53 individual (unique) gestures falling into 23 categories of gestures, each category potentially having subcategories. We release this gesture set to the community of practice (CoP). This is a first step towards a more comprehensive understanding of how body-based gestures can be formally expressed to facilitate several reasoning tasks: beyond querying based on the ontology vocabulary (Fig. 2) and its formal syntax, a preliminary investigation of this formalization has been tested for conducting usability testing experiments. The ontology provides the following potential benefits: (1) it defines a semantic model of gestures combined with knowledge associated to the context (contextual gestures), as opposed to a syntactical model; (2) it introduces a shareable and reusable gesture representation useful for reasoning based on these relationships (e.g., the `subClass` relationship is part of query), but also for experiments and database management; (3) it is extensible; (4) it provides a coherent navigation from one concept of the to another, which is key for data integration, mining, and analytics.

REFERENCES

[1] James F. Allen. 1983. Maintaining Knowledge About Temporal Intervals. *Commun. ACM* 26, 11 (Nov. 1983), 832–843.

[2] Idil Bostan, Oğuz Turan Buruk, Mert Canat, Mustafa Ozan Tezcan, Celalettin Yurdakul, Tilbe Göksun, and Oğuzhan Özcan. 2017. Hands As a Controller: User Preferences for Hand Specific On-Skin Gestures. In *Proc. of DIS '17*. 1123–1134.

[3] Zhen Chen, Xiaochi Ma, Zeya Peng, Ying Zhou, Mengge Yao, Zheng Ma, Ci Wang, Zaifeng Gao, and Mowei Shen. 2018. User-Defined Gestures for Gestural Interaction: Extending from Hands to Other Body Parts. *Int. J. of Human-Comp. Interaction* 34, 3 (2018), 238–250.

[4] Eunjung Choi, Heejin Kim, and Min K. Chung. 2014. A taxonomy and notation method for three-dimensional hand gestures. *Int. Journal of Industrial Ergonomics* 44, 1 (2014), 171–188.

[5] Grazia Cicirelli, Carmela Attolico, Cataldo Guaragnella, and Tiziana D'Orazio. 2015. A Kinect-Based Gesture Recognition Approach for a Natural Human Robot Interface. *International Journal of Advanced Robotic Systems* 12, 3 (2015), 22.

[6] Sabrina Connell, Pei-Yi Kuo, Liu Liu, and Anne Marie Piper. 2013. A Wizard-of-Oz Elicitation Study Examining Child-defined Gestures with a Whole-body Interface. In *Proc. of IDC '13*. 277–280.

[7] Natalia Díaz Rodríguez, Robin Wikström, Johan Lilius, Manuel Pegalajar Cuéllar, and Miguel Delgado Calvo Flores. 2013. Understanding Movement and Interaction: An Ontology for Kinect-Based 3D Depth Sensors. In *Ubiquitous Computing and Ambient Intelligence. Context-Awareness and Context-Driven Interaction*, Gabriel Urzaiz, Sergio F. Ochoa, José Bravo, Liming Luke Chen, and Jonice Oliveira (Eds.). Springer International Publishing, Cham, 254–261.

[8] Haiwei Dong, Nadia Figueroa, and Abdulmotaleb El Saddik. 2015. An Elicitation Study on Gesture Attitudes and Preferences Towards an Interactive Hand-Gesture Vocabulary. In *Proc. of MM '15*. 999–1002.

[9] Bogdan-Florin Gheran, Jean Vanderdonckt, and Radu-Daniel Vatavu. 2018. Gestures for Smart Rings: Empirical Results, Insights, and Design Implications. In *Proc. of DIS '18*. 623–635.

[10] Dafydd Gibbon. 2009. Gesture Theory is Linguistics: On Modelling Multimodality as Prosody. In *Proc. of the 23rd Pacific Asia Conf. on Language, Information and Comp.* City Univ. of Hong Kong, 9–18.

[11] M. Giuroiu and T. Marita. 2015. Gesture recognition toolkit using a Kinect sensor. In *IEEE Int. Conf. on Intelligent Computer Communication and Processing (ICCP)*. 317–324.

[12] J. Han, L. Shao, D. Xu, and J. Shotton. 2013. Enhanced Computer Vision With Microsoft Kinect Sensor: A Review. *IEEE Transactions on Cybernetics* 43, 5 (2013), 1318–1334.

[13] Army Publishing Headquarters, Department of the Army. 2017. Visual Signals, TC 3-21.60. http://www.apd.army.mil. (March 2017).

[14] James R. Lewis. 1995. IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction* 7, 1 (1995), 57–78.

[15] Naveen Madapana, Glebys Gonzalez, Richard Rodgers, Lingsong Zhang, and Juan P. Wachs. 2018. Gestures for Picture Archiving and Communication Systems (PACS) operation in the operating room: Is there any standard? *PLOS ONE* 13 (06 2018).

[16] Pradyumna Narayana, Nikhil Krishnaswamy, Isaac Wang, Rahul Bangar, Dhruva Patil, Gururaj Mulay, Kyeongmin Rim, Ross Beveridge, Jaime Ruiz, James Pustejovsky, and Bruce Draper. 2019. Cooperating with Avatars Through Gesture, Language and Action. In *Intelligent Systems and Applications*. Springer, Cham, 272–293.

[17] Michael Nebeling, David Ott, and Moira C. Norrie. 2015. Kinect Analysis: A System for Recording, Analysing and Sharing Multimodal Interaction Elicitation Studies. In *Proc. of EICS '15*. ACM, 142–151.

[18] Michael Nebeling, Elena Teunissen, Maria Husmann, and Moira C. Norrie. 2014. XDKinect: development framework for cross-device interaction using Kinect. In *Proc. of EICS'14*. 2014, 65–74.

[19] George V. Popescu, Helmuth Trefftz, and Grigore Greg Burdea. 2014. Multimodal Interaction Modeling. In *Handbook of Virtual Environments, 2nd ed.* CRC Press, 411–434.

[20] Jaime Ruiz and Daniel Vogel. 2015. Soft-Constraints to Reduce Legacy and Performance Bias to Elicit Whole-body Gestures with Low Arm Fatigue. In *Proc. of CHI '15*. 3347–3350.

[21] Chaklam Silpasuwanchai and Xiangshi Ren. 2015. Designing concurrent full-body gestures for intense gameplay. *International Journal of Human-Computer Studies* 80 (2015), 1 – 13.

[22] Radu-Daniel Vatavu and Jacob O. Wobbrock. 2015. Formalizing Agreement Analysis for Elicitation Studies: New Measures, Significance Test, and Toolkit. In *Proc. of CHI '15*. 1325–1334.

[23] Jacob O. Wobbrock, Meredith Ringel Morris, and Andrew D. Wilson. 2009. User-defined Gestures for Surface Computing. In *Proc. of CHI '09*. 1083–1092.