Gestural interaction with 3D objects shown on public displays: an elicitation study

Guiying Du[1], Auriol Degbelo[1], Christian Kray[1], Marco Painho[2]

[1]Institute for Geoinformatics, University of Muenster, Muenster, Germany
[2]Nova Information Management School, Universidade Nova de Lisboa, Lisbon, Portugal

Abstract. Public displays have the potential to reach a broad group of stakeholders and stimulate learning, particularly when they are interactive. Therefore, we investigated how people interact with 3D objects shown on public displays in the context of an urban planning scenario. We report on an elicitation study, in which participants were asked to perform seven tasks in an urban planning scenario using spontaneously produced hand gestures (with their hands) and phone gestures (with a smartphone). Our contributions are as follows: (i) We identify two sets of user-defined gestures for how people interact with 3D objects shown on public displays; (ii) we assess their consistency and user acceptance; and (iii) we give insights into interface design for people interacting with 3D objects shown on public displays. These contributions can help interaction designers and developers create systems that facilitate public interaction with 3D objects shown on public displays (e.g. urban planning material).

Keywords: public displays, gestural interaction, elicitation, urban planning.

1 Introduction

An important measure to ensure the success of an urban planning project is to involve a broad range of citizens in the process; however, this is also a major challenge [1]. First, usually only a subset of the potentially affected citizens can be reached. Further, for those who do become involved, their level of participation is often very low. Key roadblocks to generating public involvement relate to how information about participation opportunities are distributed, what media are chosen to disseminate information, and other barriers that affect citizens' capability and willingness to actively take part. In this context, one way to potentially improve citizen participation is to use public displays [2, 3]. Public displays have become ubiquitous, and they are now found in many public or semi-public spaces such as shopping malls, transportation hubs, plazas, museums and various other urban settings. They can lower the barriers for citizens who want to take part in public decision-making processes such as urban planning in two ways: these public displays can make information available at locations that are frequently visited, and they can enable everyone to participate actively, regardless of their age, background, or experience.

In addition to providing easy access to urban planning materials and facilitating in-situ interaction, public displays also have the potential to stimulate learning. Interacting with public displays significantly increases recall [4], and Barth and Müller [5] also reported positive feedback from users regarding the use of public displays for learning. In another study, Giovannella et al. [6] implemented a mid-air gesture interface integrating smart learning and tourism by Kinect, which resulted in a very rapid learning curve. Nevertheless, there has been limited use of such displays (and displays that connect with mobile devices) for tracking and visualizing data in a

learning context [7]. Recent work [8] has also highlighted the need for further research on the promising combination between urban planning and interactive public displays: current research has indicated that public displays usually involve only low levels of public participation (displays focus on informing citizens rather than enabling them to voice opinions or make suggestions). The goal of this article is to address the following question: *How do users envision interacting with public displays - using mobile phones and gestures - to scrutinize urban planning material, i.e. 3D objects*? By answering this question, we aim to contribute insights into how public display designers can engage a broader range of citizens to more actively participate in urban planning projects.

In the following, we report on an elicitation study for determining the hand and phone gestures people make in the context of interacting with 3D objects shown on public displays. We asked participants to spontaneously perform gestures to accomplish tasks in the context of actively participating in urban planning. Our main contributions are as follows: (i) We identify two user-defined gesture sets that participants produced using their hands or using a mobile phone when performing several examination tasks with the 3D objects shown on a large public display; (ii) we assess the two gesture sets regarding their consistency and user acceptance; and (iii) we derive several implications for the design of interactive public displays in the context of interacting with 3D objects. Our contributions can help designers select suitable interaction modalities for citizen consultations via public displays (e.g. in the context of urban planning). In addition, our findings pave the way for the design of smart learning ecosystems [9] by connecting a network of citizens, urban planning materials, and public display technology in order to facilitate active citizen participation in urban planning processes.

In the following sections, we first review work related to using public displays for public participation and gesture interaction. We then describe the elicitation study we conducted, and then we report our key results and discuss their implications for system design and gestural interaction for public displays. Finally, we conclude our research by summarizing our main contributions and outlining future work.

2 Related Work

Since the aim of the current work is to gain insights into participants' perceptions about and needs for interacting with 3D objects shown on public displays, this section reviews previous research that has been done on two relevant topics: public displays for public participation and gestural interaction.

2.1 Public Displays for Public Participation

The International Association for Public Participation (IPA2)'s Public Participation Spectrum[1] defines five levels of realizing citizen participation: informing, consulting, involving, collaborating with, and empowering citizens. To engage such citizen participation, studies have explored using a variety of online technologies [10], and various means have also been explored for encouraging citizen participation specifically with public displays. For instance, some public displays have included voting applications [11, 12] regarding local issues. In another study, Hosio et al. [13] used applications on public displays to disseminate information about the construction

[1] https://www.iap2.org.au/About-Us/About-IAP2-Australasia-/Spectrum, (last accessed: Feb 20, 2018).

of a long-term renovation project and to enable citizens to provide in situ feedback to the institution responsible for the renovation project. Taking another view, Goncalves et al. [14] observed that public displays are instrumental in generating interest, but this interest comes at the price of noisy feedback. In another study, Behrens et al. [15] gave citizens the opportunity to express their feelings about local urban challenges on media façades through tangible artifacts. Despite these participation-invoking efforts, Du et al.'s survey [8] found that current research on public displays in urban settings still mainly targets low levels of participation. They observed that most research on public displays for public participation just address the *inform* level of IPA2's public participation spectrum. Further research efforts are thus needed to achieve higher levels of citizen participation, which may then result in "better governmental decisions that involve larger numbers of citizens and are, therefore, more acceptable and legitimate to the majority of people" [16].

The increasing use of large public displays in urban public life brings new challenges and opportunities for both designers and users, especially regarding how to provide suitable interaction modalities to retrieve information from or perform useful tasks for citizens in different scenarios. In addition, citizens are very diverse in their age, background, and experience with technology. Looking at current research on large interactive displays, there are four main interaction modalities for large public displays: touch, tangible objects, external devices, and body [17]. However, while the modalities of touch and tangible objects have been used more frequently for horizontal displays [18, 19], they may be unsuitable for large vertical displays, e.g. if displays are very large or unreachable. In this research, we want to make large displays accessible to a broader group of people. For this reason, in this article we focus on gestural interaction, i.e. hand gestures and phone gestures.

2.2 Gestural Interaction

Since gestures are considered to be an intuitive method of interaction, it is not surprising that much research has been done in this area for a broad range of applications. For instance, Medrano et al. [20] looked into remote pointing when using mobile devices, and they identified three categories of pointing gesture interactions, namely free-hand pointing, see-through pointing and device pointing. Rovelo et al. [21] examined gestures for interacting with 360° panoramic recordings, both for an individual and collocated usage. Further, Kray et al. [22] studied how people use gestures on mobile phones to interact with other types of devices (i.e. another phone, a tabletop, and public display). They reported that the concept of phone gestures was very easy to understand and to put into practice; their participants indicated that phone gestures would work well for interacting with public displays. Wobbrock et al. [23] stressed the importance of involving users in coming up with gestures for a given task, reporting that "three experts cannot generate the scope of gestures that 20 participants can". Further outcomes from previous studies include a gesture set for 3D manipulations of distant objects [24], a gesture set for the exploration of large datasets through active tokens [25], and insights from sign language interpreters about hand gestures that are most comfortable when performed repeatedly [26].

Example work specifically directed towards gestural interaction with public displays include the following studies [27–29]: Fikkert et al. [27] identified a set of gestures through which commands (e.g. panning and zooming) can be issued to a large interactive display 'with ease'. Panning and zooming were also the focus of a study by Nancel et al. [28], though they looked closely into the performances of the different types of gestures used for these two tasks. They found that two-handed gestures were faster than one-handed ones and that linear gestures are generally faster

than circular ones [28]. Walter et al. [29] investigated the usability of a system that allows people to vote on a given topic, and they concluded that people (if provided no hint) use pointing and dwelling gestures to successfully select items. As the studies mentioned above illustrate, gestural interaction is a vibrant area of research and has the promise of immediate usability (when implemented appropriately). The study presented in the next sections aims at exploring gestures that are helpful (and natural) to people when it comes to interacting with 3D objects.

3 User Study

User-centered design is an approach that puts the user in the center to elicit input from them when interacting with technologies and allowing them to define intuitive and easy interactions [30]. In the spirit of participatory design, one aim of our elicitation study is to explore user-defined hand gestures and phone gestures to interact with 3D objects shown on large public displays. The motivation for the study, as mentioned in Section 1, is that public displays hold great potential for providing a broad set of citizens with access to urban planning material and that higher interactivity with displays can lead to higher information recall [4]. We focus only on *enabling users to examine the 3D objects shown on public displays*, and in particular on tasks such as *showing the back/right/left side, repositioning, resizing (bigger/smaller) and selecting a building.*

3.1 Overview and Rationale

Immersive technologies have been employed in urban planning processes for decades, either for experts or for different groups of stakeholders. In our study, we represented urban planning material as 3D objects integrated into panoramic video footage, which was projected on three large screens in a room. This setting is also known as an Immersive Video Environment (IVE). IVE is a type of audiovisual simulation that provides a feeling of immersion, where users are immersed in panoramic video footage to provide them with a strong sense of being at the real-world site depicted in the video. This immersion can promote user engagement [31]. 3D objects are increasingly used when presenting urban planning projects to citizens and can be combined with the IVE. While 3D objects provide a realistic and intuitively understandable view of what is planned, they also, however, introduce new challenges regarding how they can be examined more closely. These are two reasons why we had participants interact with 3D objects. We chose to investigate the use of both hand gestures and phone gestures, because they are two representative interaction modalities [17]. In addition, hand gestures can lead to a more immersive user experience [32] because they do not require any external device. Furthermore, many people are very familiar with smartphones, and using these devices helps to solve some privacy problems: people can input personal data without worrying about it being visible for third parties.

As one goal of our study was to elicit user-defined hand and phone gestures, we did not want the participants' behaviors to be influenced by technical issues such as gesture recognition issues or smartphone sensor technologies. No feedback from the system, i.e. IVE was provided during their performance. We also provided the participants with a transparent mockup prototype phone (as shown in Fig. 1) instead of a real phone. All participants were encouraged to disregard any gesture recognition or sensor technologies issues, and we asked them think of the mockup prototype as a futuristic smartphone. They were told that the mockup prototype could have any

features they wished for and that it would be capable of understanding and recognizing all the gestures they would perform. In this way, we followed the same principles as followed by previous research [20, 23, 33]. In addition, we wanted to remove the *gulf of execution* [34] from the user-device dialogue to make sure our observation of users' unrevised behavior was not influenced by the gesture recognition issue or the sensor technologies.

All the tasks that participants performed were played back to them as audio messages generated via a free online text-to-speech service[2]. Two additional questions for evaluating the ease and appropriateness of each gesture were also played by audio message after each task. The rationale behind this was to avoid users' misunderstanding of the tasks because of English pronunciation problems and to ensure consistent delivery of the instructions. All participants were video-recorded during the whole study session. Our study followed a within-group design. We used two panoramic videos spanning all three screens, which were captured from two different sites of our city. As said above, we focused on how participants examined 3D objects.

3.2 Selection of Tasks

According to IPA2, two key goals for citizen consultation are to keep the public informed and to obtain public feedback. To choose the suitable interaction activities for realizing these goals, we determined which tasks to include in our study by first classifying tasks into two categories: examining urban planning material, i.e. 3D objects, and giving feedback on it. Since existing research has explored providing feedback via public displays various ways, such as through entering text or voting [35, 36], we focused mainly on the first category: examining 3D objects. In doing so, we also followed the typology of general interactivity [37] for geographic visualization. The selected tasks are representative of typical tasks that can be used in the scenario of examining 3D objects shown on large public displays. In total, we asked each participant 14 questions, seven for each gesture type, using the following two templates: *Which hand gesture would you use to do ACTION*? and: *How would you use your smartphone to do ACTION*?, where ACTION stands for:

- Show the back side of the building?
- Show the right side of the building?
- Show the left side of the building?
- Move the building from its location to another location?
- Make the building smaller?
- Make the building bigger?
- Select the building?

3.3 Apparatus and Materials

We conducted the study in a lab environment. The two panoramic videos overlaid with 3D buildings were displayed in the IVE consisting of three big screens connected to a single PC running Windows 7. A MacBook Pro was used to play all the audio message questions during the whole study. The 3D building objects used in our study

[2] http://www.fromtexttospeech.com/, (last accessed: Nov 20, 2017).

were downloaded from the 3D Warehouse[3]. One 3D model was a model of supermarket[4] and another one was a skateboard shop[5]. We used Sketchup Pro 2017[6] to export the pictures of each model to PNG format with transparent backgrounds, which were then overlaid over the videos by Final Cut Pro X. We used two cameras in our study. One Canon EOS 550D[7] camera was put beside participants on a tripod to view them from the side . Another GoPro Hero4 camera was situated on the top of the front section of the IVE to view them from the front. There was a moderator sitting close to the Canon EOS 550D camera throughout the study session. With this setup, we attempted to capture all the details when participants were performing surface gestures on the phone. All the participants were guided to stand in the same location of the room in front of the IVE. The location was marked by a white paper with two footprints. Fig. 1 depicts the study settings and also shows the transparent mockup prototype used in the phone-gesture condition.

3.4 Participants

Twenty-eight participants, twenty-one males and seven females, between the ages of 22-39 (mean=28, SD=4.9) were recruited for our study. They had different professional backgrounds. Most of them had lived in Germany for the last two years, but some participants had primarily lived in other European countries, American countries or Asian countries over last two years. There were no special requirements about participants' age or prior experience regarding participation in urban planning processes. All participants had a moderate level of English. Recruitment was done through emails, flyers, Facebook, and word of mouth. Each participant received 10 EUR as a reward at the end of the study.

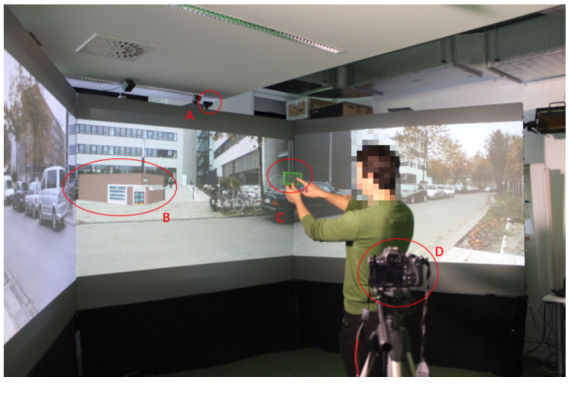

Fig. 1. Study setup: The participant with the transparent mockup prototype (C) stood on the footprint mark in front of the IVE showing the 3D objects (B), while the GoPro Hero4 camera (A) and the Canon camera (D) recorded the participant's behaviors.

[3] https://3dwarehouse.sketchup.com/, (last accessed: Nov 20, 2017).
[4] https://3dwarehouse.sketchup.com/model/fec488ae8cbf0c8035d6a087b4694131/Meijer-Supermarket, (last accessed: Nov 20, 2017).
[5] https://3dwarehouse.sketchup.com/model/76e9d3b5554893e272396bc529d8c9c/Small-Time-Skates, (last accessed: Nov 20, 2017).
[6] https://www.sketchup.com/download/all, (last accessed: Nov 20, 2017).
[7] https://www.canon.de/for_home/product_finder/cameras/digital_slr/eos_550d/, (last accessed: Nov 20, 2017).

3.5 Procedure

At the beginning of the study, each participant was given a brief explanation of the objective and the procedure. After they read and understood the study, they were asked to sign the consent form. Then the moderator guided each participant to stand at a marked position in front of the large display. Before starting the main part of the study, the moderator spent several minutes introducing the IVE and the tasks in the study. The moderator encouraged participants to ask any question regarding the study. The moderator also told participants that they could think aloud when performing the tasks. The main part of the study started when participants were clear about what was going to happen and what they should do.

After the setup, the moderator gave an urban planning story to each stimulus, i.e. the panoramic video with the 3D objects overlaid. Then each participant was given about one minute to become familiar with the IVE and the stimuli. The audio message describing the task was played next, and participants began to perform the task. After each task, the participants were asked how easy/how appropriate it was to come up with an action for the particular task. The order of exposure to each of the interaction modalities, i.e. the hand and the smartphone, was counterbalanced. For each condition, the order of tasks and videos were randomized across conditions and participants.

Once the two scenarios were finished, each participant was given two questionnaires. The first one asked for participants' background information, and the second aimed to get general feedback and attitudes about the hand-gesture and smartphone-gesture interactions. Finally, the moderator wrapped up the session and handed out their reward. The duration of each study was about 45 minutes.

4 Results

In the following section, the results of our study will be presented. The section starts with a brief introduction of the taxonomies for hand gestures and phone gestures used during the analysis, and it goes on to describe the hand- and phone-gesture sets obtained in the study, the agreement scores between participants, and their subjective ratings.

4.1 Taxonomy used in the analysis

Arnheim and McNeil [38] described that gestures consist of four stages: preparation, stroke, hold, retraction. The stroke phase describes the step of performing the gesture, so we firstly extracted this phase of all proposed gestures. The gestures were further analyzed by the taxonomies. Inspired by relevant work about classification of gestures [23, 30, 33, 39], we then derived our taxonomies for further analyzing users' hand gestures and phone gestures. In order to analyze the gestures in as much detail as possible, we made changes to previous taxonomies. The taxonomy for user-defined hand gestures was modified and extended from Obaid et al. [30, 39] and Ruiz et al. [33]. The taxonomy for user-defined phone gestures was modified and extended from Obaid et al. [30, 39], Ruiz et al. [33], and Wobbrock et al. [23]. We analysed the hand gestures according to five dimensions: form, nature, body parts, temporal, and complexity dimensions. Each dimension consists of multiple categories, as shown in Table 1. The phone gestures were analyzed along seven dimensions (Table 2): form, nature, touch-fingers, temporal, complexity, spatial, and type of gestures dimensions.

The form dimension in the air-hand gesture taxonomy was adopted from Obaid et al. [30, 39] without changes. In the phone gesture taxonomy, we combined the form of surface gestures [23] and the form of motion gestures [33] into the form dimension. We modified the categories of the body-parts dimension from Obaid et al. [30, 39], because we were eliciting air-hand gestures that only involved hands but no other body parts. In the taxonomy of the phone gestures, we replaced this dimension by the touch-fingers dimension, which describes the number of fingers involved in performing gestures.

We extended the nature dimension originally from Obaid et al. [30] according to the research presented by Wobbrock et al [23]. The categories of our nature dimension include deictic, iconic, metaphorical, abstract and symbolic gestures. The temporal dimension is also adopted from Wobbrock et al. [23] to show whether the ongoing recognition of gestures is needed or not. The complexity dimension adopted from Ruiz et al. [33] aims to capture as how complicated a gesture is perceived to be.

Table 1. Taxonomy for user-defined hand gestures (modified and extended from Obaid [30, 39] and Ruiz [33]).

Field	Value	Description
Form	static	A static body posture is held after a registration phase
	dynamic	The gesture contains the movement of one or more body parts during the stroke phase
Nature	deictic	The gesture is indicating a position or direction
	iconic	The gesture visually depicts an icon and directly represents a real-world property
	metaphoric	The gesture visually depicts an icon and describes a real-world property in an abstract way
	abstract	Gesture mapping is arbitrary
	symbolic	The gesture is an artificial symbol that does not represent a real-world property but represents a meaning that needs to be learned and is often culture specific
Body parts	right hand only	The gesture is performed with the right hand only
	left hand only	The gesture is performed with the left hand only
	two hands	The gesture is performed with two hands
Temporal	discrete	Action occurs after completion of gesture
	continuous	Action occurs during gesture
Complexity	simple	Gesture consists of a single gesture
	compound	Gesture can be decomposed into simple gestures

Table 2. Taxonomy for user-defined phone gestures (modified and extended from Obaid [30, 39], Ruiz [33], and Wobbrock [23]).

Field	Value	Description
Types of gestures	surface gesture	Deliberate movements of the device by end-users to invoke commands
	motion gesture	Two-dimensional gestures using the touchscreen of the smartphone as a mobile surface computer
	mixed gesture	Combine the surface and motion gesture
Form	static pose	Hand pose is held in one position
	dynamic pose	Hand pose changes in one position
	static pose and path	Hand pose is held as hand moves
	dynamic pose and path	Hand pose changes as hand moves
	Single-Axis motion	Phone-Motion occurs around a single

		axis
	Tri-Axis motion	Phone-Motion involves either translational or rotational motion, not both
	Six-Axis	Motion occurs around both rotational and translational axis
Temporal	discrete	Action occurs after completion of gesture
	continuous	Action occurs during gesture
Spatial	ST	Perform the gesture while looking through the transparent screen of the device
	DP	Device was used an extension of their arm or remote control
	SSU	Device was used as current smartphone and held on one of the hands
Touch-fingers	One finger	The gesture is performed with one finger only
	two fingers	…two fingers
	multi-fingers	… more than two fingers
Complexity	simple	Gesture consists of a single gesture
	compound	Gesture can be decomposed into simple gestures
Nature	deictic	The gesture is indicating a position or direction
	iconic	The gesture visually depicts an icon and directly represents a real-world property
	metaphoric	The gesture visually depicts an icon and describes a real-world property in an abstract way
	abstract	Gesture-referent mapping is arbitrary
	symbolic	Gesture visually depicts a symbol

Regarding the spatial dimension, we got inspiration from [20]. Some participants preferred to perform gestures by looking through the transparent screen of the mockup device. These gestures were classified as 'see-through' (ST) gestures. Some participants also used the mockup device as an extension of their arms or remote control. These gestures were labelled as 'device-pointing' (DP) gestures. We also found that some participants designed gestures that mimicked actions occurring during normal use of smartphones. These gestures were categorized as 'standard smartphone-use' (SSU).

A total of 196 hand gestures were collected. As shown in the overall taxonomy distribution of the hand gestures (see Fig. 2), these gestures tended to be simple dynamic gestures which were performed involving the right hand and which required continuous recognition and real-time feedback. The overall taxonomy distribution of the phone gestures illustrates the breakdown of the 196 phone gestures observed in our study. As shown in Fig. 3, more surface gestures were found than motion gestures, and more than 90% of gestures were performed by one or two fingers. Similar to the hand gestures, most of phone gestures were simple gestures and required continuous recognition and real-time feedback.

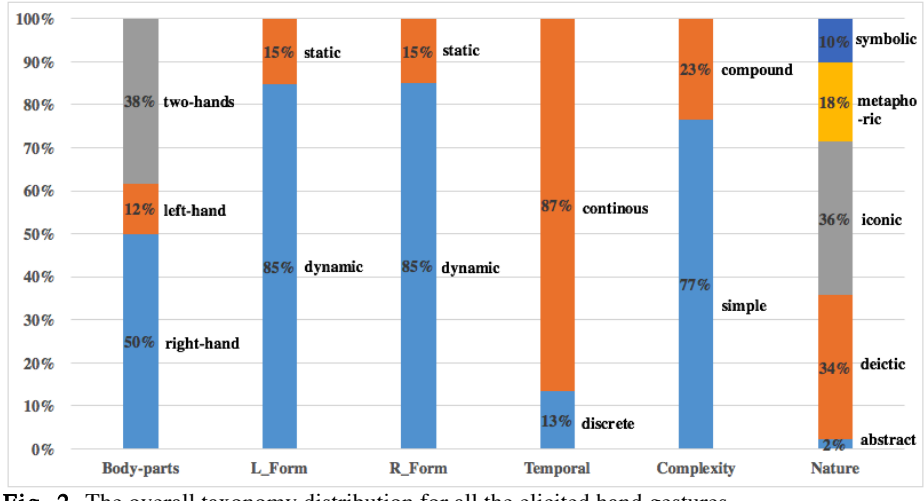

Fig. 2. The overall taxonomy distribution for all the elicited hand gestures.

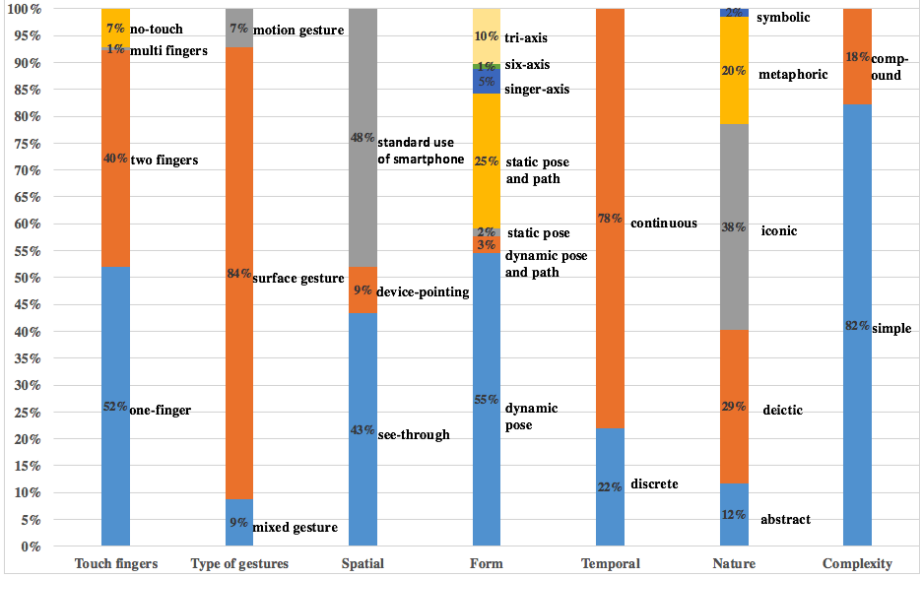

Fig. 3. The overall taxonomy distribution for all the elicited phone gestures.

4.2 User-Defined Gesture Sets and Agreement Scores

The core of our study aim is to generate user-defined gesture sets. This process was structured as follows: firstly, for one task t, all gestures produced were grouped into a set P(t); then we classified all gestures in P(t) into subsets, which contain identical

gestures $P_i(t)$, with $i \in 1, 2 \ldots n$, and n is the value of the total number of identified subsets for task t. The subset $P_i(t)$ with the largest size was then chosen as the representative gesture for the task t for our user-defined set. We also checked if there was more than one gesture candidate for a task. Second or third gesture candidates were only chosen when they were accounted for at least half of the first gesture candidate. However, if the representative gestures for different tasks were the same, then a conflict occurred. That is because one gesture cannot result in two or more outcomes. To resolve the conflict, we assigned the gesture to the task that was associated with that gesture the most often. Also if the first gesture candidate of one task was the second or the third gesture candidate of another task, we removed this gesture as the alternative gesture for the other task. Table 3, Table 4, and Table 5 show all the gesture candidates for each of the seven tasks in the two conditions, how often the participants performed each task with the gesture candidate, and all the gesture candidates' taxonomies. Our process of generating user-defined gestures can be traced back to previous work [23, 30, 33, 39]. In our study, the first candidate gesture of *Show back* conflicts with that of *Show right* and *Show left* in both conditions. Compared with the other two actions, the action *show back* did not have the largest group, so we moved the second candidate gesture as the first candidate gesture for that action.

To evaluate the degree of consensus among participants with the proposed gesture, we use the formula as used by Wobbrock et al. [23]. They calculate an agreement score AS(t) for each task t, where:

$$AS(t) = \sum_{i=1}^{n} \left(\frac{P_i(t)}{P(t)}\right)^2 \quad (1)$$

The range of AS(t) is $[P_i(t)', 1]$, and $P_i(t)'$ corresponds to all the participants choosing different gestures for task t, while 1 means all the participants performed the same gesture for task t. So we can say if there is a high agreement score for task t, then all the participants have a similar understanding of how to perform the task by gesture. But when there is a low agreement score for task t, participants found it difficult to think of a similar appropriate gesture for task t. Fig. 4 shows agreement levels for hand gestures and phone gestures.

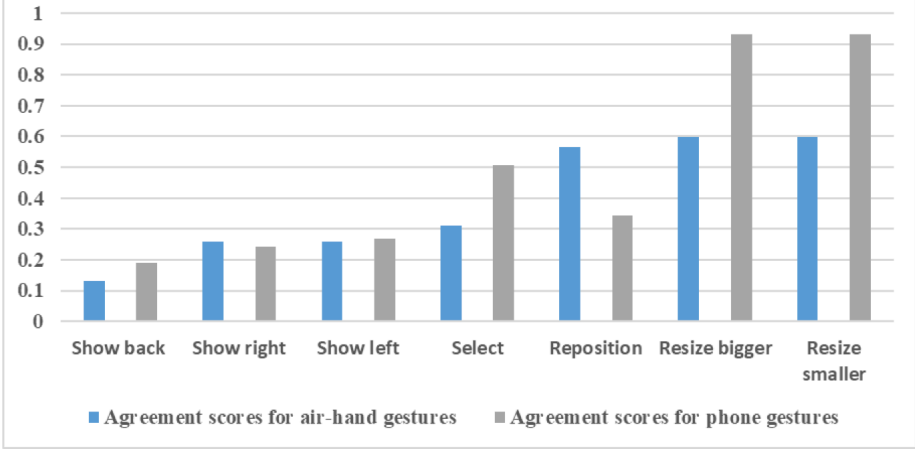

Fig. 4. Agreement scores for user-defined gestures.

Table 3. Gesture candidates for performing the seven tasks by hand.

Action	Gesture candidates	Occur-rences	Form	Gesture Nature	Body Parts	Temporal	Comp-lexity
Show right	Swipe left	35%	Dynamic	Deictic	One hand	Continuous	Simple
Show left	Swipe right	35%	Dynamic	Deictic	One hand	Continuous	Simple
Select	One hand air-point	52%	Static	Deictic	One hand	Continuous	Simple
Resize smaller	Move two hands linearly closer	75%	Dynamic	Iconic	Two hands	Continuous	Simple
Resize bigger	Move two hands linearly spread	75%	Dynamic	Iconic	Two hands	Continuous	Simple
Reposition	Hand-point to the building, then move the hand to another location, then loosen hand	68%	Dynamic	Deictic	One hand	Continuous	Comp-ound
Show back	Two hands perform clockwise motion along Y-axis	18%	Dynamic	Metaph-oric	Two hands	Continuous	Simple

Table 4. Gesture candidates for performing the seven tasks by smartphone (First part).

Action	Gesture Candidates	Occurrences	Type of gestures	Form
Show right	Swipe right/left	46%	Surface	Static pose and path
Show left	Swipe right	43%	Surface	Static pose and path
Select	Tap	68%	Surface	Dynamic pose
Resize smaller	Pinch to zoom out	96%	Surface	Dynamic pose
Resize bigger	Pinch to zoom in	96%	Surface	Dynamic pose
Reposition	Keep on pressing the building on the smartphone, move the smartphone to another location and then release the finger;	43%	Mixed	Tri-Axis motion
	Drag the building on the smartphone	39%	Surface	Static pose and path
Show back	Move fingers around each other on the building	18%	Surface	Dynamic pose

Table 5. Gesture candidates for performing the seven tasks by smartphone (Second part).

Action	Touch fingers	Temporal	Complexity	Gesture Nature	Spatial
Show right	One finger 77% Two fingers 23%	Continuous	Simple	Deictic	SSU 46% ST 54%
Show left	One finger 75% Two fingers 25%	Continuous	Simple	Deictic	SSU 42% ST 58%
Select	One finger 95% Two fingers 5%	Discrete	Simple	Abstract	SSU 47% ST 53%
Resize smaller	Two fingers	Continuous	Simple	Iconic	SSU 48% ST 52%
Resize bigger	Two fingers	Continuous	Simple	Iconic	SSU 48% ST 52%
Reposition	One finger 92% Two fingers 8%	Continuous	Compound	Metaphoric	DP 25% ST 58% SSU 17%
	One finger 91% Two fingers 9%	Continuous	Simple	Iconic	ST 36% SSU 64%
Show back	Two fingers 80% Multi-fingers 20%	Continuous	Simple	Metaphoric	SSU 80% ST 20%

4.3 Subjective Ratings

After each action, the participants answered the two following questions using a 5-point Likert scale:

- How easy was it for you to produce this gesture? Answers were given on a scale from 1 = "quite hard" to 5 = "quite easy".
- How would you rate the appropriateness of your gesture/action to the task? Answers were given on a scale from 1 = "quite inappropriate" to 5 = "quite appropriate".

Fig. 5 and Fig. 6 show the mean values for users' ratings of easiness and appropriateness of the hand gestures and phone gestures, respectively. We applied a two-way repeated ANOVA and found that neither the means of the ratings for ease nor appropriateness differed significantly with the interaction modalities, i.e. hand and smartphone. However, they did differ significantly with the tasks, $F(6) = 2.9225$, $P<0.05$ but not as a function of both tasks and interaction modalities. A one-way repeated measures ANOVA revealed that the means of rated appropriateness differed significantly between the actions, with $F(6) = 3.231$, $P<0.01$ for all the phone gestures. The task *show back* received significantly lower ratings of the appropriateness than the other actions, while the tasks *resize bigger, resize smaller, select* had significantly higher ratings for appropriateness than other actions. No significant difference of the means of the rated easiness of gestures were found for hand gestures or phone gestures. The means of the rated appropriateness also did not differ with tasks for phone gestures.

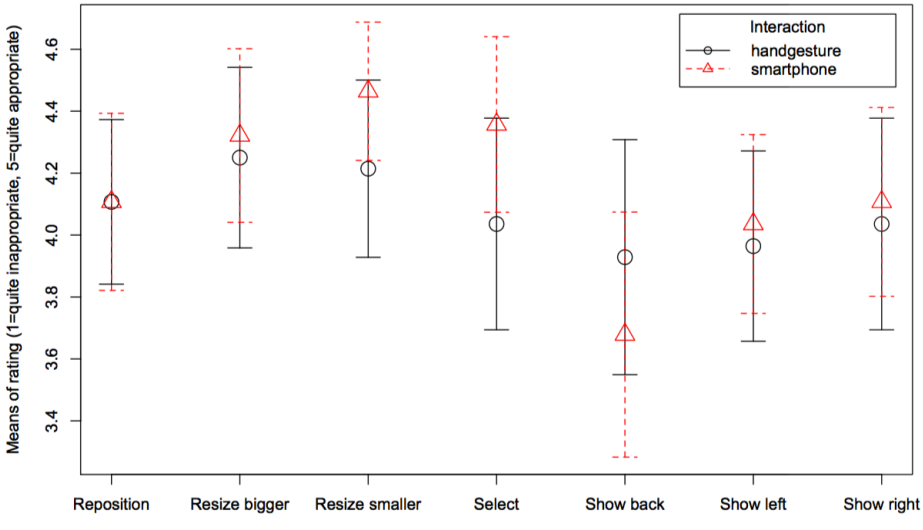

Fig. 5. User ratings for the appropriateness of the proposed gestures for the seven tasks.

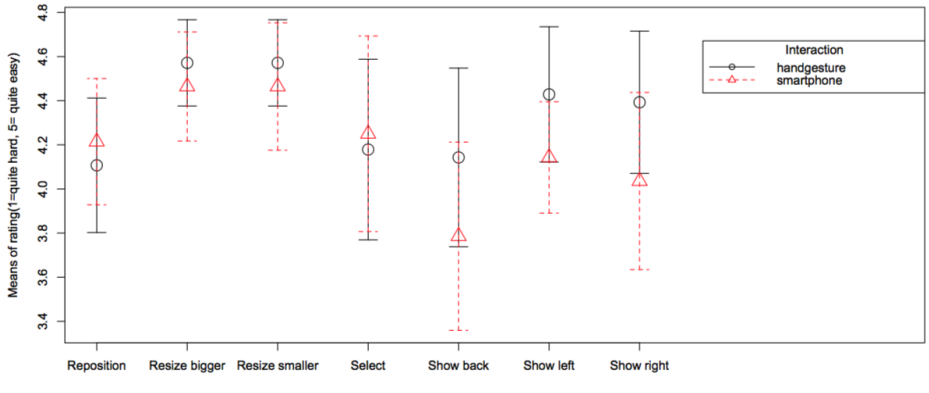

Fig. 6. User ratings for the easiness of the proposed gestures for the seven tasks.

5 Discussion

In this section, we discuss the user-defined gestures we observed in our study as well as the implications of our results for system design and interface design.

5.1 User-Defined Gestures

The distributions of hand gestures and phone gestures reveal some common characteristics. Both types of user-defined gestures tend to be simple and continuous. There is a similar distribution for user-defined hand gestures (iconic and deictic; 70%) and user-defined phone gestures (iconic and deictic; 67%). This suggests that users

expect gestural interaction to provide immediate responses and continuous control regarding the urban planning information by means of simple gestures. It also indicates that users prefer to have their actions directly and visually depicted on the 3D objects when examining them in the IVE. People found the task *show back* hard to understand as indicated by the low scores for agreement and appropriateness. This indicates no common concept exists among people for this task. Consequently, it may make sense to remove this task and to instead rely on performing *show left/right* twice.

Dynamic right-hand gestures were most preferred according to the taxonomy distribution of the elicited hand gestures, both for left-hand gestures and right-hand gestures (dynamic; 85%). Even though recognizing dynamic gestures is technically more challenging than static postures, this finding implies future gestural interfaces for 3D objects might require the former.

According to the taxonomy distribution of phone gestures, one finger (52%) and two fingers (40%) touch were most popular among the participants. It may make sense to not distinguish between one-point touch and two-point touch for supporting interactions with 3D objects by a smartphone. This observation is similar to what we observed for hand gestures. During our study, we found people usually did not consider the number of fingers when performing the tasks by hands. For example, when performing the gesture *hand point*, some participants pointed to the display while bending all the fingers resulting in a fist, while some others pointed to it by using one finger or more fingers. We also observed that more surface gestures (84%) were generated than motion gestures (9%), although participants were well aware that they could imagine that the mockup phone supports any type of technology or function they wished. This may be the result of our participants having extensive smartphone experience, since more than 90% of them frequently used smartphones. Medrano et al. [20] showed in a previous study that user preferences were influenced by the technology experience of the participants.

5.2 Implications for System Design

Regarding the system design for the two user-defined gesture sets, we can point out several challenges. Except the gesture candidate for the task *select*, the other gesture candidates in the user-defined hand gesture set are dynamic and continuous. Developers should consider that users will not all be the same height, may stand at different locations in front of a large public display, and may expect immediate responses during the gestures. Hence, it is important to find a suitable recognition system that provides a wider tracking range and facilitates synchronous responses. Most of the gesture candidates in the user-defined set are of the deictic type, but there are also some iconic and metaphoric types. This suggests that the recognition system needs to be able to recognize both types. A suitable recognition system should also meet the requirement of recognizing one-handed or two-handed gestures while ignoring the number of fingers.

In the case of the user-defined phone gesture set, we observed a strong need for surface recognition technology. There are iconic, deictic, metaphoric, and abstract types of gestures with different forms. This means the sensor should be sensitive enough to recognize diverse patterns and forms of surface gestures. It was very common for the same gesture candidate to be performed by users but with different numbers of fingers. Developers could provide a choice for users to decide the number of fingers involved or decide to ignore the number of fingers while focusing on the shape, location and dynamics of the gesture. Another challenge is the mixture of motion gesture and surface gesture for one task. To support this type of gesture, different sensor technologies have to be integrated during the system design. In

addition, a major trend that emerged was that people preferred to hold the smartphone with the screen facing the 3D objects with one hand, while performing surface gestures with the other hand. They imagined that there should be a synchronous response to their gestures from the urban planning materials (i.e. 3D objects) both on the smartphone and the large public display. This case may necessitate technologies like a rear-mounted camera and low-latency connectivity to support communicating the data and interaction events between the smartphone and the large display. A related challenge is to quickly and reliably compute the geometric mapping between the smartphone image and the display, which is known as display registration [40].

5.3 Implications for User Interface Design

The high agreement scores regarding the *resizing* tasks imply that user interfaces for interacting with 3D objects shown on public displays may readily implement the gesture set identified in this work. The lower agreement between participants for other tasks suggests, on the contrary, that user interface designs may have to accommodate the variety of options expressed by the participants. In addition, with the exception of the task *reposition*, the degree of the consensus among the participants regarding the elicited phone gestures was higher than the degree of consensus for the hand gestures. People rated phone gestures as more appropriate than hand gestures, although they commented in general that phone gestures were less intuitive to recall than hand gestures. Participants found it particularly hard to come up with an appropriate phone gesture for the task *show back*. The higher degree of consensus for phones gestures than for hand gestures may be explained through participants' general familiarity with phone usage.

Qualitative feedback collected from our participants suggests some implications for the design of interfaces based on hand gestures. There are four frequent negative aspects mentioned by participants: frustration resulting from the inability of the system to detect gestures properly, a lack of confidence based on previous bad experiences, social embarrassment, and lack of privacy when performing gestures in front of the public.

Regarding the design of interfaces based on phone gestures, four frequent negative aspects were also given: the need to have a smartphone and an app to perform the interaction, the high effort resulting from switching between two screens, the lack of motivation to connect their personal device with the public display due to privacy risks and security issues, and the limited screen size of the smartphone. The latter may be related to the 3D objects not being well suited for exploration on a small screen, e.g. larger maps showing a planned project.

5.4 Limitations of the Study

Most participants were under 40 years old, highly educated, and experienced smartphone users. It is quite possible that the gestures obtained were influenced by the participants' technology experiences and backgrounds. Repeating the study with participants from other groups (e.g. children, older people, or people with little technology experience) would lead to a more complete picture of users' perceptions and needs regarding interacting with 3D objects shown on public displays.

6 Conclusion

In this paper, we presented an elicitation study exploring two interaction modalities for facilitating interaction with 3D objects shown on public displays. We recruited 28 participants for our study, during which we asked them to watch panoramic videos overlaid with 3D objects in an immersive video environment. They then performed seven tasks of examining 3D objects using their hands only (first condition) and then using a mockup of a futuristic smartphone (second condition). In total, we elicited 196 hand gestures and 196 phone gestures, which we analyzed to derive two gesture sets that can inform research and practice on interaction with 3D objects on public displays. In addition, we also collected qualitative feedback about the easiness and appropriateness of the elicited gestures. Participants mostly agreed on gestures for *resizing* 3D objects, while gestures involving the manipulation of buildings (e.g. *select, show back, show right, show left*) led to much lower agreement scores.
An immediate step for future work is to implement the identified gesture sets in a system and to evaluate their usability. Further studies regarding the functions that resulted in low agreement scores are also highly desirable, as are studies that replicate our setup with different user groups (e.g. different age range, technological background or culture).

Acknowledgements: The authors gratefully acknowledge funding from the European Commission through the GEO-C project (H2020-MSCA-ITN-2014, Grant Agreement No. 642332, http://www.geo-c.eu/), and comments from the reviewers that helped improve the article.

References

1. Münster S., Georgi C., Heijne K., Klamert K., Noennig J.R., Pump M., Stelzle B., van der Meer H.: How to involve inhabitants in urban design planning by using digital tools? An overview on a state of the art, key challenges and promising approaches Procedia Comput. Sci., 112, pp. 2391–2405 (2017)
2. Brignull H., Rogers Y., Brig H.: Enticing People to Interact with Large Public Displays in Public Spaces in Rauterberg, M., Menozzi, M., and Wesson, J. (eds.) Proceedings of the 7th International Conference on Human-Computer Interaction (INTERACT'03). pp. 17–24. IOS Press (2003)
3. Memarovic N., Langheinrich M., Alt F., Elhart I., Hosio S., Rubegni E.: Using Public Displays to Stimulate Passive Engagement, Active Engagement, and Discovery in Public Spaces Proc. 4th Media Archit. Bienn. Conf. Particip. (MAB '12), pp. 55–64 (2012)
4. Alt F., Schneegass S., Girgis M., Schmidt A.: Cognitive effects of interactive public display applications Proceedings of the 2nd ACM International Symposium on Pervasive Displays - PerDis '13. p. 13. ACM Press, Mountain View, California (2013)
5. Barth K., Muller W.: Interacting with public displays for informal learning: design issues and first Experiences 2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT). pp. 92–94. IEEE (2017)
6. Giovannella C., Iosue A., Moggio F., Rinaldi E., Schiattarella M.: User experience of kinect based applications for smart city scenarios integrating tourism and learning Proceedings - 2013 IEEE 13th International Conference on Advanced Learning Technologies, ICALT 2013. pp. 459–460 (2013)
7. Verbert K., Govaerts S., Duval E., Santos J.L., Van Assche F., Parra G., Klerkx J.: Learning dashboards: an overview and future research opportunities Pers. Ubiquitous Comput., (2013)
8. Du G., Degbelo A., Kray C.: Public displays for public participation in urban settings: survey Proc. 6th ACM Int. Symp. Pervasive Displays - PerDis '17, pp. 1–9 (2017)
9. Giovannella C.: Smart Learning Eco-System: " FASHION " OR " BEEF "? J. e-learning

Knowl. Soc., 10, pp. 15–23 (2014)
10. Naranjo Zolotov M., Oliveira T., Casteleyn S.: E-participation adoption models research in the last 17 years: A weight and meta-analytical review, (2018)
11. Claes S., Slegers K., Vande Moere A.: The Bicycle Barometer: Design and Evaluation of Cyclist-Specific Interaction for a Public Display Proc. 2016 CHI Conf. Hum. Factors Comput. Syst., pp. 5824–5835 (2016)
12. Schiavo G., Milano M., Saldivar J., Nasir T., Zancanaro M., Convertino G.: Agora2.0: enhancing civic participation through a public display C&T 2013, pp. 46–54 (2013)
13. Hosio S., Goncalves J., Kostakos V., Riekki J.: Exploring civic engagement on public displays in Saeed, S. (ed.) User-Centric Technology Design for Nonprofit and Civic Engagements. pp. 91–111. Springer International Publishing (2014)
14. Goncalves J., Hosio S., Liu Y., Kostakos V.: Eliciting situated feedback: A comparison of paper, web forms and public displays Displays, 35, pp. 27–37 (2014)
15. Behrens M., Valkanova N., Fatah gen. Schieck A., Brumby D.P.: Smart Citizen Sentiment Dashboard : A Case Study Into Media Architectural Interfaces Proc. Int. Symp. Pervasive Displays, pp. 19–24 (2014)
16. Milakovich M.E.: The Internet and Increased Citizen Participation in Government eJournal Democr., (2010)
17. Ardito C., Buono P., Costabile M.F., Desolda G.: Interaction with Large Displays ACM Comput. Surv., 47, pp. 1–38 (2015)
18. Rogers Y., Lim Y.K., Hazlewood W.R.: Extending tabletops to support flexible collaborative interactions Proceedings of the First IEEE International Workshop on Horizontal Interactive Human-Computer Systems, TABLETOP'06. vol. 2006. pp. 71–78 (2006)
19. Dohse K.C., Dohse T., Still J.D., Parkhurst D.J.: Enhancing multi-user interaction with multi-touch tabletop displays using hand tracking Proceedings of the 1st International Conference on Advances in Computer-Human Interaction, ACHI 2008. pp. 297–302 (2008)
20. Medrano S.N., Pfeiffer M., Kray C.: Enabling remote deictic communication with mobile devices: An elicitation study Proc. 19th Int. Conf. Human-Computer Interact. with Mob. Devices Serv. MobileHCI 2017, (2017)
21. Rovelo G., Vanacken D., Luyten K., Abad F., Camahort E., Val D., N C.D.V.S.: Multi-Viewer Gesture-Based Interaction for Omni-Directional Video pp. 4077–4086 (2014)
22. Kray C., Nesbitt D., Dawson J., Rohs M.: User-defined gestures for connecting mobile phones, public displays, and tabletops Proc. 12th Int. Conf. Hum. Comput. Interact. with Mob. devices Serv. - MobileHCI '10, pp. 239 (2010)
23. Wobbrock J.O., Morris M.R., Wilson A.D.: User-defined gestures for surface computing Proceedings of the 27th international conference on Human factors in computing systems - CHI 09. p. 1083 (2009)
24. Liang H.-N., Williams C., Semegen M., Stuerzlinger W., Irani P.: User-defined surface+motion gestures for 3d manipulation of objects at a distance through a mobile device Proc. 10th asia pacific Conf. Comput. Hum. Interact. - APCHI '12, pp. 299–308 (2012)
25. Valdes C., Eastman D., Grote C., Thatte S., Shaer O., Mazalek A., Ullmer B., Konkel M.K.: Exploring the design space of gestural interaction with active tokens through user-defined gestures Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14. pp. 4107–4116. ACM Press, Toronto, Canada (2014)
26. Rempel D., Camilleri M.J., Lee D.L.: The design of hand gestures for human–computer interaction: Lessons from sign language interpreters Int. J. Hum. Comput. Stud., 72, pp. 728–735 (2014)
27. Fikkert W., Van Der Vet P., Van Der Veer G., Nijholt A.: Gestures for large display control Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), 5934 LNAI, pp. 245–256 (2009)
28. Nancel M., Wagner J., Pietriga E., Chapuis O., Mackay W., Nancel M., Wagner J., Pietriga E., Chapuis O., Mackay W.: Mid-air Pan-and-Zoom on Wall-sized Displays pp. 177–186 (2011)
29. Walter R., Bailly G., Valkanova N., Müller J.: Cuenesics: using mid-air gestures to select items on interactive public displays Proc. 16th Int. Conf. Human-computer Interact. with Mob. devices Serv. - MobileHCI '14, pp. 299–308 (2014)
30. Obaid M., Kistler F., Kasparavičiūtė G., Yantaç A.E., Fjeld M.: How would you gesture

navigate a drone?: a user-centered approach to control a drone Proc. 20th Int. Acad. Mindtrek Conf. - Acad. '16, pp. 113–121 (2016)
31. Drettakis G., Roussou M., Reche A., Tsingos N.: Design and Evaluation of a Real-World Virtual Environment for Architecture and Urban Planning Presence Teleoperators Virtual Environ., 16, pp. 318–332 (2007)
32. Ren G., Li C., O'Neill E., Willis P.: 3D freehand gestural navigation for interactive public displays IEEE Comput. Graph. Appl., 33, pp. 47–55 (2013)
33. Ruiz J., Li Y., Lank E.: User-defined motion gestures for mobile interaction Proc. 2011 Annu. Conf. Hum. factors Comput. Syst. - CHI '11, pp. 197 (2011)
34. Hutchins E., Hollan J., Norman D.: Direct Manipulation Interfaces Human-Computer Interact., 1, pp. 311–338 (1985)
35. Steinberger F., Foth M., Alt F.: Vote with your feet: Local community polling on urban screens Proc. 3th Int. Symp. Pervasive Displays (PerDis '14), pp. 44 (2014)
36. Cheverst K., Fitton D., Dix A.: Exploring the Evolution of Office Door Displays BT - Public and Situated Displays: Social and Interactional Aspects of Shared Display Technologies Presented at the (2003)
37. Crampton J.W.: Interactivity Types in Geographic Visualization Cartogr. Geogr. Inf. Sci., 29, pp. 85–98 (2002)
38. Arnheim R., McNeill D.: Hand and Mind: What Gestures Reveal about Thought Leonardo, 27, pp. 358 (1994)
39. Obaid M., Kistler F., Häring M., Bühling R., André E.: A Framework for User-Defined Body Gestures to Control a Humanoid Robot Int. J. Soc. Robot., 6, pp. 383–396 (2014)
40. Pears N., Jackson D.G., Olivier P.: Smart phone interaction with registered displays IEEE Pervasive Comput., 8, pp. 14–21 (2009)