Enabling Remote Deictic Communication with Mobile Devices: an Elicitation Study

Samuel Navas Medrano
Institute for Geoinformatics
University of Münster
Münster, Germany
s.navas@uni-muenster.de

Max Pfeiffer
Institute for Geoinformatics
University of Münster
Münster, Germany
max.pfeiffer@uni-muenster.de

Christian Kray
Institute for Geoinformatics
University of Münster
Münster, Germany
c.kray@uni-muenster.de

ABSTRACT

Mobile systems provide many means to relay information to a distant partner, but remote communication is still limited compared to face-to-face interaction. Deictic communication and pointing, in particular, are challenging when two parties communicate across distances. In this paper, we investigate how people envision remote pointing would work when using mobile devices. We report on an elicitation study where we asked participants to perform a series of remote pointing tasks. Our results provide initial insights into user behaviors and specific issues in this context. We discovered that most people follow one of two basic patterns, that their individual pointing behavior is very consistent and that the shape and location of the target object have little influence on the pointing gesture used. From our results, we derived a set of design guidelines for future user interfaces for remote pointing. Our contributions can benefit designers and researchers of such interfaces.

ACM Classification Keywords

H.5.2. User Interfaces – User-centered design

Author Keywords

Mobile interaction; gestures; participatory design.

INTRODUCTION

Deictic communication permeates everyday life and it is instrumental to successfully exchanging information. For example, it is used to establish joint attention [13], to express where people or things are, how they are placed and oriented, and which spatial relation they have to each other. Typical terms used to relay deictic information include: *this that, there, here, behind* or *over*. There is evidence that deictic expressions are amongst the first words children learn and use [9], and that they are present in almost every known language [12]. Since human communication is not limited to verbal exchanges, other modalities also play an important role. In

Figure 1. Mockup smartphone prototype used in the study

particular, facial expressions and gestures are often essential to understanding our communication partner – be it a human or a computer [51]. In face-to-face communication, we frequently use *deictic gestures* such as pointing to express or describe the location or shape of an object. For example, a person might utter "Please pass me that." to someone else while extending her arm and pointing her finger in the direction of the pen she would like to obtain. Without properly perceiving the pointing gesture, it will be difficult or impossible for the other person to correctly identify what the speaker means and to thus react appropriately.

Nowadays mobile technology enables us to also communicate remotely. While it facilitates bridging arbitrary distances, this type of communication is subject to some significant limitations in particular concerning deictic expressions. In the case of an audio-only connection, any gestural information is lost. There is also no joint perception of the environment each communication partner is located in: neither side knows which entities are located where with respect to the other party. Consequently, deictic expressions such as demonstratives (e.g. *there* or *here*) cannot be used well as they require a spatial context to make sense. Video connections have the potential to overcome some of these issues. While they enable each side to "see" the environment of the other party, the field of view afforded by conventional mobile devices is usually very limited. This can make it very difficult to bring both target and 'pointer' into the camera view at the same time, which in turn can severely affect deictic communication.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

MobileHCI '17 September 04-07, 2017, Vienna, Austria

© 2017 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-5075-4/17/09.

DOI: <http://dx.doi.org/10.1145/3098279.3098544>

The limited field of view might be one reason why there is little research on remote deictic communication. It is also not clear how such communication could be realized technologically and on the interface level. Finally, we currently do not know how people would intuitively/naturally perform pointing gestures for remote deictic communication. In this paper, we therefore explore how people use a mobile device to perform pointing gestures for remote deictic communication. Our goal is to develop a better understanding of key aspects in this context, to collect different types of gestures, and to gather qualitative feedback on the overall idea of remote deictic communication. For this purpose, we designed a lab-based elicitation study, in which we asked participants to point at different target objects during communication with a remote party.

Our main contributions are initial insights into natural user behavior for remote deictic communication and the identification of three categories of pointing gesture interaction: *free-hand pointing*, *see-through pointing* and *device pointing*. We also developed a tagging scheme for pointing gestures and initial guidelines for designing interfaces for remote pointing. Our contributions can benefit the development and investigation of future interfaces that support pointing across distances.

The remainder of this paper is structured as follows. We first review relevant related work and then describe in detail the elicitation study we conducted. We also report on the analysis we applied to the collected data before highlighting the key results we obtained. Based on these results, we then derive a series of guidelines for designers. A short discussion points out implications and limitations of our work. The final section briefly summarizes our main contributions and provides an outlook on future research.

RELATED WORK

In the following, we first discuss deictic communication in general to highlight its importance and its different facets. We then present previous work on remote collaboration, which defines the broader technological context of our research. In addition, we review work on gestures and gestural communication in general including work on pointing gestures. The section concludes by discussing elicitation studies as a means to derive intuitive gestures.

Deictic Communication

Deictic communication is an active research field across multiple disciplines such as linguistics and cognitive psychology. Also due to this aspect, there is no universally agreed-upon definition that covers all possible facets or uses. The term is derived from the Greek word 'deixis', which refers to pointing or indicating. This can include the use of gestures, demonstratives, pronouns or other means. Tomasello [55], for example, compared the information shared between speech and gestures in languages, while Dixon [14] did the same with demonstratives. Diessel [12] argued that the primary words used for spatial deictic communication across languages are demonstratives, i.e. *this*, *that*, *here* and *there*.

There is a substantial body of work highlighting that deictic communication plays a significant role in communication and

the evolution of languages. According to the British National Corpus word frequency count¹, demonstratives appear with the highest frequency terms in a language. Bühler [7] argued that demonstratives play a major role in the evolution of grammar. Furthermore, Clark and Eve [10] observed that demonstratives are among the first words infants acquire, and Iverson [27] found that gestures can predict the acquisition of language. Diessel [13] reported findings indicating that demonstratives are more related to gestures than other words.

People are more likely to use gestures when they are communicating spatial information than when their message does not contain spatial information [2][35]. It is more challenging to express and understand this type of information without the use of gestures [47]. Occasions, where spatial deictic information plays a major role, include describing the distribution of a room or building [53], an irregular shape [25], or motion in space [31]. Giving directions [3] is another example for a person trying to transmit spatial information. This is particularly relevant in a mobile remote communication setting with one partner navigating in situ, while a person at another location remotely provides directions.

Gestural Communication

As mentioned in the previous section, communication is not limited to (spoken or written) language alone. Non-verbal aspects in general and gestures in particular have been the subject of extensive research for quite some time now. Efron [16] proposed a set of detailed categories of what he called emblems as communicative body movements. Ekman & Friesen [17] and Graham and Argyle [24] later modified his classification and suggested extensions such as Freedman's [19] representational gestures, Wiener et al.'s [59] pantomimic gestures or McNeill's [39] iconic, metaphoric and deictic gestures. McNeill [39] also proposed to distinguish between the pre-stroke, stroke and post-stroke phases of gestures, which facilitates the differentiation between static and dynamic gestures. Poggi [46] introduced a detailed classification of gestures based on a different set of variables ranging from the semantic content, goal, and awareness to cognitive construction.

Since gestures are an important aspect of communication, gestural interfaces have become an established field in Human-Computer Interaction. Several approaches for gesture recognition have been suggested, using a broad range of technologies and sensors (see [40, 48] for a survey). Devices such as the Wii Remote² or the Kinect camera³ have popularized gesture interaction. The broad array of sensors on modern smartphones and mobile devices has enabled mobile gestural interaction as well.

Work on pointing gestures in Human-Computer Interaction can be traced back to the "Put that there" system [6], which used polhemus sensors to track arm movements for interacting with big displays. More recently, approaches such as

¹<http://www.natcorp.ox.ac.uk/using/index.xml?ID=freq>, accessed 25 May 2017

²<http://www.nintendo.com/wiiu/accessories>, accessed 9 February 2017

³<http://www.xbox.com/en-US/xbox-one/accessories/kinect>, accessed 9 February 2017

Mayer's et. al. [37] proposed pointing interaction using an allocentric perspective and investigated hand pointing as well as gaze pointing [1]. Such approaches frequently require an instrumented environment to work properly, thereby restricting mobility to that augmented area. Other approaches have looked at enabling pointing gestures in a more general mobile context. Magnusson [36] used a separate movement sensor for pointing that was connected to a smartphone via Bluetooth. The ShowMe system [4] facilitates pointing (amongst other gestures) in a mobile setting by recording arm movements and transmitting them to a communication partner to support remote collaboration.

Remote Communication and Collaboration

Modern technology enables remote human interaction and eliminates the requirement to be co-present in order to interact [23, 29, 26]. Unsurprisingly, remote communication has had a profound impact on how people live and work and has changed how people understand time and space [8, 5]. Along these lines, some argue that mobile phones liberate people from place [58, 22] and that the 'sense of belonging to a place' is superseded by 'a sense of belonging to a communicative network' [18]. Based on Palen et. al.'s [44] definition of conversational space, Rettie [49] argues that people can feel as if they were in two places at the same time during remote communication. Apart from the communicative, social and individual considerations pertaining to remote collaboration and communication, there is also a technical component to consider.

The concepts of remote collaboration, mobile communication and presence are well-established fields in Human-Computer Interaction and have been investigated in the context of many different domains. Gestures, in general, and pointing gestures in particular have also received some attention in the context of remote collaboration. For example, Fussell [20] and Kirk [30] explicitly investigated tools and approaches what allowed gesture interaction for supporting remote collaboration in video systems. Gauglitz [21] proposed a framework to enable mobile remote collaboration on tasks that involve the physical environment. The ShowMe system [4] mentioned in the previous subsection is an example of a system that facilitates remote gestural communication.

Elicitation Studies for Gestural Interfaces

In order to design user interfaces for gestural/remote communication, different approaches are possible. Rather than researchers or designers creating such interfaces by themselves, participatory design [52] advocates involving users into the design process. Those users then adopt an active role at different stages in the design process. Particularly in the early stages, elicitation studies can play an important role in this context. In such studies, users are asked to perform specific tasks spontaneously or "instinctively" without being restricted by predefined interaction mechanisms or technology. This approach has the potential to not only empower users but also to produce interactions that are more likely to be user friendly than the ones designed by experts [43]

Figure 2. Study setup: participants stood on the footprint mark in front of the IVE (a), while a front and a side cameras (b,c), a top webcam (d) and a Kinect (e) were recording their movements

User-centred methodologies and elicitation studies have also been used to generate gestural interactions. Wobbrock et al. [60] used this approach in the context of interaction with a tabletop computer. Morris [42] adopted a user-centered approach to elicit speech and gesture interactions for various functions of a web browser. They also described the max-consensus and consensus-distinct ratio metrics for analysing elicitation study results. Vavatu and Zait [56] used a similar approach to elicit gestures in the scenario of TV interaction. Döring et al. [15] also used the same approach to elicit a gesture set for multitouch interaction on a steering wheel.

Elicitation studies have also found a place in the context of mobile interaction. Kray et al. [33] carried out such a study to see what actions or gestures people produce in the context of activities involving two different devices. Jokela et al. [28] looked at a more specific case focusing on wearable devices and two everyday tasks. Another example is work by Shimon et al. [54] who carried out an elicitation study to collect gestures to interact with the back of a (mobile) device.

STUDY

From previous work, we can thus conclude that deictic communication is an important component of interpersonal interaction and that there is a substantial potential for it in a mobile context, e.g. for supporting people during navigation. While previous work has covered various aspects related to deictic communication, gestural interaction and remote collaboration, we can also observe a gap in research on remote deictic communication and in particular, pointing in a mobile setting. It is not clear how such communication could be realized technologically and on the interface level. Specifically, we currently do not know how people would intuitively/naturally perform pointing gestures for remote deictic communication.

Figure 3. Experimenter setup: the laptop was used to trigger audio messages to participants (e.g. task descriptions) and to observe their behavior via the installed cameras (IVE, Kinect and webcam video stream)

Aim

In our study, we thus investigate how people use a mobile device to perform pointing gestures for remote deictic communication. The overall aim is to gain an initial understanding of relevant aspects in this context, to collect different types of gestures and to gather qualitative feedback on the general idea (of remote pointing). In the spirit of participatory design, we carried out an elicitation study to collect gestures that people spontaneously and naturally produce when they perform deictic communication while using a mobile device for remote communication.

Study Design

For our study, we tried to balance realism and openness to ensure eliciting useful gestures. Initially, we considered placing the experimenter in a separate room and performing the experiment remotely via a video-call. We reckoned that this setup would closely resemble the envisioned scenario and thus result in a realistic setting for the participants. However, we decided against this approach as it might have biased the participants towards using the camera to perform deictic communication. Instead, the participants and the experimenter were placed in the same room but were separated by a heavy black curtain. While this avoided interaction between the experimenter and the participant and thus better simulated the remote communication scenario, it also meant that the experimenter was nearby in case of any problems.

In terms of the mobile device to be used, we first considered using a turned-off smartphone during the experiment but then decided against it for a similar reason: we wanted to elicit interactions that were not limited by today's technologies or a specific phone model. Eventually, we opted for crafting a transparent mock-up prototype [45, 57] as a futuristic mobile device (see Figure 1). We reasoned that the prototypical nature of the mock-up and the absence of clearly identifiable controls such as buttons would free participants from the limitations of current devices.

In order to facilitate the elicitation of 'natural' gestures, we wanted to expose participants to a realistic environment but also needed to maintain a high degree of control. We therefore used panoramic video material of urban scenarios and played them back in an immersive visualization environment. We selected videos from an existing pool that contained target objects with different spatial properties (e.g. small/large, wide/narrow, individual objects/groups of objects). In order to trigger pointing gestures, we designed two sets of questions

that we played back as audio instructions (generated via a text-to-speech system⁴). The rationale behind this decision was to simulate a remote communication scenario, where the person in situ would hear but possibly not see the remote communication partner. The first set of questions (see Table 1) aimed at eliciting gestures for specific objects. Consequently, they all used the same structure: "Could you please show me the X?", where X was replaced by an object shown in the panoramic video. With the second set of questions our goal was to give participants more freedom with respect to object choices and gestures (see Table 2). In total, there were ten questions for each of the three videos we used (see Figure 4): eight targeting specific objects and two open questions. We randomized the order of the questions to counter any order effects that could result from a fixed sequence.

Participants

We recruited participants through flyers and online channels such as international students groups from the local university community. Twenty people with an average age of 25 years (SD: 5.17, Range: 20-34 years) and approximately equal balance of gender (nine males and 11 females) participated in the experiment. All participants were right-handed, owned smartphones, and had an university-level education. They all stated that they were fluent English speakers and were residents of the city where the panoramic videos were recorded. In addition, they also reported being familiar with all the recorded spots. For the study, a monetary compensation of 10 EUR was offered to the participants.

Apparatus and Material

To ensure that all participants experience the same conditions, the experiment was performed in a lab environment. We used an Immersive Video Environment (IVE) [41] to display panoramic video footage (see Figure 2). With this setting we were able to realistically simulate environments that participants would encounter in a field study while maintaining full control and ensuring consistent experiences across all participants. The IVE consisted of three big screens connected to a single PC running Windows 7, which played back panoramic videos spanning all three screens. Stereo speakers were hidden behind the center screen and were used to play back audio instructions during the experiment.

Two Canon EOS 550D⁵ cameras were used to capture a front view and a side view of the participants. A Logitech HD Pro Webcam C920⁶ situated on the top of the front section of the IVE was used as a backup recording device and for allowing the experimenter to have a live video stream of the participants during the experiment. In addition, a Kinect 2 camera was placed in front of the participants (see Figure 3 for full setup). The Kinect, the webcam, and the speakers were connected to the laptop of experimenter (see Figure 2) with Kinect Studio 2.0v. The laptop was also running a second instance of the IVE

⁴<https://play.google.com/store/apps/details?id=com.google.android.tts>, accessed 25 May 2017

⁵http://www.canon.de/for_home/product_finder/cameras/digital_slr/eos_550d/, accessed 9 February 2017

⁶<http://www.logitech.com/en-roeu/product/hd-pro-webcam-c920>, accessed 9 February 2017

Could you please show me the...		
#	Target	Asked in...
1	... car?	V1 & V3
2	... street?	V1, V2 & V3
3	... trash bin?	V1, V2 & V3
4	... bikes?	V1, V2 & V3
5	... white building?	V1
6	... post?	V1
7	... traffic sign?	V1 & V3
8	... bus stop information sign?	V1
9	... pharmacy?	V2
10	... tree?	V2
11	... plants?	V2
12	... entrance of the bank?	V2
13	... street lamp?	V2 & V3
14	... building under construction?	V3
15	... cathedral?	V3

Table 1. Object-specific questions used to elicit pointing gestures while seeing one of three panoramic videos (V1 to V3); order was randomized in the study.

If you would like to...	
#	Target
1	"... open a business here where would you place it?"
2	"... rest where would you take a sit?"

Table 2. Open questions used to elicit pointing gestures; each one was asked for all videos (V1, V2 & V3); order was randomized in the study.

to play the IVE background audio at the same time as playing the audio instructions. Data from the Kinect device and the Logitech webcam were also recorded during the experiment.

All participants were placed in the same spot of the IVE (Figure 2.a), which was marked by two footprints. The mock-up smartphone we built (see Figure 1) was handed to the participants prior to the experiment. We told them this is a phone from the future, which has all features and sensors they could imagine.

Procedure

After welcoming the subjects, they received printed documents informing them about the experiment, a participation consent form, and an initial questionnaire with a number of demographic questions (age, gender, dominant hand and expertise of using smart-phones). Any questions participants had while reading through the documents were answered by the experimenter. After participants had signed the consent and filled in the questionnaire, the experimenter placed them on a marked position (Figure 2.a) in front of the center screen. He then provided information regarding the IVE and detailed instructions of the tasks to perform. Participants were asked to behave in the IVE as they would in real life. We also instructed them to pretend that the mock-up phone was a futuristic smartphone capable of doing whatever they could imagine, equipped with any sensor and/or running any kind of app. Furthermore, we asked them to imagine being in a remote communication scenario such as a phone conversation. We also informed the participants that a remote voice would ask them to perform

Hand	Dominant	
	Non-Dominant	
	Both	
Target	Single	
	Multiple	
Stroke	Dynamic	
	Static	
Pointing	Free-Hand Pointing	F1, F2, F3, F4, F5
	See-Through Pointing	P, L
	Device Pointing	

Table 3. Tagging scheme used to classify deictic pointing gestures

ten different tasks in three different scenarios (30 tasks in total) and that they should use the phone for fulfilling those task. These instructions were repeated up to three times during the introduction to ensure participants were clear about what they needed to do and what they had to imagine. This was important as there was no further direct interaction between the experimenter and participants during the study. Finally, the experimenter provided participants with the opportunity to ask any question regarding the procedure of the experiment.

After the setup was completed, the main part of the study began. For each task, the audio message of the task was played, and participants then tried to perform a corresponding pointing gesture. They were also allowed to request repeating the audio message, or to indicate that they wanted to skip to the next task. (The last situation could occur if participants were not able to complete the task or did not know how to do it). The audio message was played back in stereo by speakers placed directly in front of the participant. The order of the tasks was randomised across scenarios and subjects. The experimenter was behind a curtain, not visible and did not talk to the participant during this part of the study.

Once the three scenarios were completed, subjects were given a final questionnaire, in which they were asked different questions related to their habits of usage of mobile devices. This included questions such as how much they use their smartphone, their general attitude about using a mobile device in the situation that the experiment introduces, what sensors and features they imagined the mock-up device had and some further questions. After completing the questionnaire, the experimenter debriefed them and they received a small payment to compensate them for their time.

Collected data

We collected different types of data during the study. We asked participants to fill in qualitative questionnaires before and after the main study. The initial questionnaire collected demographic information about the participants. In the final questionnaire, we asked them if they would use a phone in the suggested remote communication scenario of the study. They were also prompted to suggest three positive and three negative aspects about using a phone in this scenario. Finally, we asked the participants to describe the features they would like a phone to have for this purpose, and we gave them the opportunity to provide any further comment. During the main study, we recorded audio and video footage. We placed one camera in the front of the center screen (Figure 2.b) and a

Figure 4. Example frames from the three panoramic videos used in the study (V1 to V3); playback happened in the immersive video environment (IVE).

secondary camera from the side (Figure 2.c). Extra footage from a webcam placed on the top of the center screen (Figure 2.d) was also recorded in order to have another reference view in case any gesture was not clearly visible from the other cameras. In addition, we recorded the skeleton data from a Kinect 2 device (Figure 2.e) placed in front of the participant.

ANALYSIS

In order to analyse the recorded video footage we devised a tagging scheme following an iterative process. First, we created a preliminary version of the scheme based on previous work in gestures classification [39, 46] and research on eliciting user-generated gesture interactions [50, 34, 15]. The resulting scheme was modified whenever we encountered new requirements in the form of different ways of gesture interactions. The final tagging scheme is shown in Table 3. We then analysed all the video footage using the previously defined tagging scheme with the goal to properly distinguish and classify every interaction.

Comparing participants gesture interactions, we identified three different main categories: *free-hand pointing*, *see-through pointing* and *device pointing*. We defined *free-hand pointing* as a natural pointing gesture in which the participant was using a hand to relay spatial information without explicitly using or positioning the phone. For a detailed analysis, we tagged each finger used for describing the requested spatial information (F1: thumb, F2: index, F3: middle, F4: ring, F5: little). The gesture could be performed using more than one finger. Hand gestures, of participants using all the fingers,

were also included in this category (Figure 8.b, g, h, i, k and l). If a participant performed the gesture by looking through the transparent screen of the mock-up device to describe the required targets, we classified it as a *see-through pointing* interaction. In this category we also tagged the orientation of the device as *portrait* or *landscape* (Figure 8.a, e, f, j). Finally, we define *device pointing* as the gesture interaction where people used the device as an extension of their arm for describing the targets. In this case, as we did not find any difference, all the gestures were tagged in the same way (Figure 8.c).

The tagging scheme include some further variables such as well: static (when participants remained in a static posture) or dynamic (when they purposefully moved body parts during the gesture) gestures. We tagged as well whether the gesture was performed to describe a single target (e.g. an individual building) or the same gesture was repeated for pointing out multiple targets (e.g. a number of bikes). We also recorded if the participants were using the dominant hand, the non-dominant hand or both hands. We further noted when the participant were using speech to support the gesture. Overall, the tagging scheme was very deterministic and easy to apply, which is why a single expert tagged all the videos.

RESULTS

The 20 participants performed 600 tasks in total resulting in three hours of video and audio material that we analyzed as described in the previous section. Figure 5 summarizes the distribution of interaction types we observed. 246 of the 600 task were performed using a *see-through pointing* gesture, and

Figure 5. Classification of the main interaction types

free-hand pointing occurred 245 times. Participants frequently used a single finger (183/245) or the whole hand (53/245) in the latter case. Multiple fingers were used less frequently (9/245). Finger pointing also occurred in 9/246 cases of *see-through pointing* (see Figure 8.f for an example). *Device pointing* was less common compared the other gestures; we observed it on 84 occasions. The remaining 25 tasks were completed either using another method or not at all. This includes 13 tasks, which were completed with non-gestural interaction such as imitating texting or verbal communication. Finally, 12 tasks were not completed by participants, which happened mainly when they had to answer open questions (see Figure 7.a), or for the first tasks of the experiment.

Users were very likely to stay with the same interaction model during the experiment. They stuck to the same interaction model independent of the targets or the tasks. The average interaction consistency was 0.93 with SD 0.1, with individual values ranging from 1.0 to 0.63. Figure 7.a and Figure 7.b provide a more detailed account of the gesture types participants used and how consistent they were with respect to their choices.

Dynamic gestures occurred much less frequently than static gestures (82/600). The majority of dynamic gestures occurred for the 'street' object, which stretched across all three screens of the IVE. 60 of the 600 tasks were related to showing this target, and in 40 cases participants performed a dynamic gesture to describe the street. Other targets, where participant performed dynamic gestures multiple times, included the plants (5 times out of 20 where it was the target object), the pharmacy (4/20), the white building (4/20), and the bikes (9/60). Dynamic gestures were also occasionally used for one of the open questions: "if you would like to open a business here, where would you place it?" (6/60).

In case of the *free-hand pointing* interaction, it was interesting to observe that participants supported by verbal information in 93 of the 245 cases. 43 of those 93 supporting utterances used the demonstrative "over there" and 42/93 used just "there". Less frequently used demonstratives were "here", "over here",

Figure 6. Number of people agreeing to use the mobile device for remote deictic communication in the future on a seven-point Likert scale (1=strongly disagree, 7=strongly agree), answers colored by predominantly used gesture type.

"in front", "at", and "in". We observed verbal support of gesture interaction as well: 16 times in case of the *device pointing* interactions and two times in case of the *see-through pointing* interactions.

For the *see-through pointing* gestures, we also analyzed how the screen was positioned. For 127 of the 246 gestures in this category, participants used the phone in portrait position, while 119/246 used it in landscape position. The targets, where portrait mode position was most frequent, were: tree (9/9, i.e. for all 9 uses of see-through pointing with the tree as target object, participants held the phone in portrait mode), the bus stop information screen (8/8), and the streetlamp (8/8). Targets, where the landscape mode was more frequent, were the bikes (6/8), the building under construction (6/8).

After completing all pointing tasks, we asked the participants fill a questionnaire to gather information about whether they intend to use a mobile device for remote deictic communication. We used a seven-point Likert scale, where one corresponded to "strongly disagree (1)" and five to "strongly agree (7)". Figure 6 summarizes the responses we obtained: 1 (4×, 20%), 2 (1×, 5%), 3 (2×, 10%), 4 (3×, 15%), 5 (6×, 30%), 6 (4×, 20%), and 7 (0×, 0%). The bars are colored according to each participants predominantly used gesture type. As can be seen from the diagram, participants using see-through pointing were much more positive about using the device in for remote deictic communication than those using free-hand pointing. For those predominately using device pointing, responses were mixed. In addition to querying participants about their intention to use, the questionnaire also asked them for three positive and negative aspects of using a mobile device as in the experiment scenario. The most frequent comments are summarized in the Table 4.

DESIGN IMPLICATIONS

From our results we can infer three main implications for the design of future gestural interfaces that facilitate remote deictic communication, in particular pointing. These relate to gesture diversity, environmental aspects, and technical considerations. Below we discuss these three areas in more detail.

Figure 7. Consistency of gesture types: a) participants' interaction agreements [60] per target; b) interaction consistency per participants across tasks.

Positive Facts

Sending pictures is possible
It is possible to search for information on the Internet
Makes remote communication easy and/or fast

Negative Facts

I did not need/use it. It is not helpful.
It is distracting.
It is dangerous

Table 4. Participants' most common positive and negative feedback.

Gesture Diversity

Communicating deictic information through pointing gestures raises several challenges regarding gesture diversity that designers for mobile devices may want to consider. Current mobile devices primarily use touchscreen input for gestural interaction. Increasingly, alternative mobile gesture recognition methods are emerging that use the built-in or external sensors. The results of our study indicate that the recognition of pointing gestures should not only rely on a single touch input. Only a few users used the screen as input method when describing the position of the targets. We identified two dominating interaction methods – *free-hand pointing* and *see-through pointing* – which should instead be supported.

In the case of *free-hand pointing*, users ignored the mobile device and used their hand to perform natural pointing gestures. For this type of gesture, primarily the index finger was used, followed by multi-finger pointing and hand gestures. A small set of participants even used both hands for pointing (Figure 8 h). The observed variety in *free-hand pointing* indicates that interaction designers would be well-advised to consider more than just one type of pointing for remote deictic communication.

In the case of *see-through pointing*, users imagined a camera in different positions of the device, e.g. on the top or the back. They also used the mock-up prototype as a see-through display and pointed with the center of the device to the target. On some occasions, users performed a *free-hand pointing* gesture with their remaining free hand in order to support the interaction.

While we found three main interaction pattern (*free-hand pointing*, *see-through pointing* or *device pointing*), participants were highly consistent in terms of which one they used – mostly regardless of the different types of targets. They stuck to the same interaction method when describing spatial infor-

mation for most of the experiment. The patterns we observed can inform the design of future interfaces for remote deictic communication as well.

Environmental Implications

Spatial pointing gestures interact with the targets and their different properties such as size or position in the environment. For example, for specific types of targets users performed static or dynamic gestures. By static gestures, we mean gestures where the hand or device was moved towards the target and then its position did not change anymore. In a dynamic gesture the hand or the device was moved along the target one or more times, describing their shape as in what McNeill termed iconic gestures [38]. This dynamic behavior was found when participants targeted objects that covered a large area of the scene such as a street or a large building. For single or small objects such as people or bikes, static gestures were clearly preferred. However, we found no difference in the usage of dynamic or static gestures regarding the *free-hand pointing*, *see-through pointing* or *device pointing* interaction modalities. It is worth mentioning, though, that we did not present any moving targets, which might have resulted in more dynamic gestures. Overall, we did not find any systematic relationship between the used interaction methods and specific targets. These findings imply that future interaction methods for remote deictic communication should focus on the link between static/dynamic gestures and the small/extended size/shape of target objects.

Technical Implications

When designing mobile devices it is crucial to consider the right selection and placement of sensors and displays on the device. In addition, the shape of the device plays an important role in the context of supporting different interaction modalities. These factors might influence how people use the mobile device in general and when communicating deictic information in particular. In this context, the use of a 360 degrees camera (such as [32] or a dedicated hand tracking device (such as the leap motion ⁷) could be very beneficial to support all different types of gesture interactions. Particularly in the case of *free-hand pointing* interaction, this would be useful as those users mostly ignored the phone. They held it with one hand while performing the pointing gesture with their other hand. Generally, designers should also take into account the different orientations of the device we observed during the study.

⁷<https://www.leapmotion.com/>, accessed 9 February 2017

For example, users who performed the *see-through pointing* gestures were alternating between portrait and landscape orientation, probably depending on the shape and extend of the target object.

Finally, in our study we found participants frequently used verbal expressions such as *here*, *this* or *that* in support of pointing gestures. Given that this is also a very common phenomenon in face-to-face communication, it would make a lot of sense to support multi-modal communication (in particular: audio and gestures) for remote deixis.

DISCUSSION

When designing our experiment, we aimed to simulate a real life scenario while ensuring a high degree of control. For this purpose, we ran the study in an immersive video environment. There was some evidence that the study setup achieved a high degree of immersion. For example, one participant performed a reverse pointing gesture when asked for where they would set up a business. The participant referred to the part of the city that would be located at her back if she would be located at the location shown by video 2 (Figure 4)

It surprised us somewhat to see a substantial number of participants perform *free-hand pointing* gestures as their predominant interaction method. The experimenter had informed them up to three times prior to the start of the study that they should use the mobile device for their interactions but these instructions were more or less ignored by this group. One possible explanation would be that participants still assumed to be collocated with their communication partner (despite them not being visible). Alternatively, this might be an artifact of the exploratory nature of the study and the rather open questions. Finally, their behavior might also reflect the implicit assumption that the phone can somehow capture their natural pointing behavior and transmit it to the remote communication partner.

Another phenomenon worth highlighting is the legacy bias reflected, for example, in the negative feedback to some questions (4). It is possible that participants were influenced by the limitations of current technology, which could explain the lack of intention to use the device in the suggested remote communication scenario of the experiment. Another potential legacy bias is the popularity of the *free-hand pointing* method as it replicates what people currently do in face-to-face situations. It is also worth noting that the lowest intention of use results were predominately observed for participants who had *free-hand pointing* as their main interaction method (see Figure 6).

Our study was subject to a number of limitations that might have affected various aspects of the results we reported above. Besides a limited number of participants and targets, the use of a mock-up device may have affected the study outcome. By providing participants with an imaginary and omnipotent mobile phone, we did not only rely on their imagination but also might have biased them towards envisioning a smartphone-inspired interaction such as the one we observed in the *see-through pointing*. We took the decision of introducing the mock-up as a mobile device over not giving them any tool at all to avoid confusing participants. We felt it might have been

overly demanding to ask them to imagine any kind of portable device they could use to perform the gestures. (Even with the mock-up, some participants who mainly relied on *free-hand pointing* interaction did not use the mock-up device at all.) Another limitation relates to the use of the immersive video environment. Even though it has been described as resulting in behavior similar to what is observed in field studies [11], we cannot be certain that participants would produce the same gestures in the field. Finally, while we simulated parts of a remote communication, it did not constitute an actual conversation between human partners.

CONCLUSION

In this paper, we reported on an exploratory study aimed at eliciting smartphone gestures for remote deictic communication. Our analysis of the collected gestures revealed two dominant and one secondary category of gestures: *free-hand and see-through pointing as well as device pointing*. Most participants either used the (assumed) camera of the device or their fingers for pointing. A substantially smaller group used the body of the phone for pointing. Our results also indicate that the type or shape of the target does not affect which type of pointing gestures people use, though there are systematic variations within the same type of gestures (e.g. the tendency to use dynamic gestures for elongated targets such as a street). Generally, we observed a very high degree of consistency regarding individual pointing behavior: most people relied on one specific type of pointing gesture throughout the majority of the study. Based on these and other results we obtained during the study, we were able to derive a number of guidelines for the design of user interfaces aimed at supporting remote deictic communication.

Our main contributions are thus initial insights into natural user behavior for remote deictic communication and the identification of three categories of pointing gestures. Further contributions include the tagging scheme we developed and initial guidelines for designing devices and interfaces that facilitate remote deictic communication in the form of pointing gestures. These contributions can benefit designers and researchers developing and investigating future interfaces and tools that support pointing across distances. Based on our results, we identified several promising directions for future research. Investigating technological realizations of the gesture types we discovered is a logical next step. In order to realize deictic communication across distances, we also want to look into how to relay pointing gestures to the remote communication partner. Finally, we did not yet consider the case where a person wants to point at targets in the environment of the remote communication partner.

ACKNOWLEDGMENTS

The authors gratefully acknowledge funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 676063 (DCOMM, "Deictic Communication" <http://www.dcomm.eu/>). We would also like to thank the participants who took part in our study.

a) See-Through Pointing

b) Free-Hand Pointing

c) Device Pointing

d) Pokémon Go Swipe Pointing

e) Portrait See-Through Pointing

f) See-Through Pointing + F2 Free-Hand Pointing

g) F2 + F5 Free-Hand Pointing

h) Two hands F2 Free-Hand Pointing

i) Reverse F2 Free-Hand Pointing

j) See-Through pointing + Free-Hand Pointing

k) Free-Hand Pointing while holding the device

l) Hand (F12345) Free-Hand Pointing

Figure 8. Main gesture interactions and gesture variations

REFERENCES

1. Deepak Akkil and Poika Isokoski. 2016. Accuracy of interpreting pointing gestures in egocentric view. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 262–273.
2. Martha W Alibali, Dana C Heath, and Heather J Myers. 2001. Effects of visibility between speaker and listener on gesture production: Some gestures are meant to be seen. *Journal of Memory and Language* 44, 2 (2001), 169–188.
3. Gary L Allen. 2003. Gestures accompanying verbal route directions: Do they point to a new avenue for examining spatial representations? *Spatial cognition and computation* 3, 4 (2003), 259–268.
4. Judith Amores, Xavier Benavides, and Pattie Maes. 2015. Showme: A remote collaboration system that supports immersive gestural communication. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 1343–1348.
5. Zygmunt Bauman. 2000. Liquid modernity. 2000. *Polity, Cambridge* (2000).
6. Richard A. Bolt. “Put-that-there”: Voice and Gesture at the Graphics Interface. In *Proceedings of the 7th Annual Conference on Computer Graphics and Interactive Techniques* (1980) (*SIGGRAPH '80*). ACM, 262–270.
7. Karl Bühler. 1934. *Sprachtheorie: die Darstellungsfunktion der Sprache*. Gustav Fischer.
8. Frances Cairncross. 2001. *The death of distance: How the communications revolution is changing our lives*. Harvard Business Press.
9. Malinda Carpenter, Katherine Nagell, Michael Tomasello, George Butterworth, and Chris Moore. 1998. Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the society for research in child development* (1998), i–174.
10. Eve V Clark. 2009. *First language acquisition*. Cambridge University Press.
11. Ioannis Delikostidis, Holger Fritze, Thore Fechner, and Christian Kray. 2015. Bridging the gap between field-and lab-based user studies for location-based services. In *Progress in Location-Based Services 2014*. Springer, 257–271.
12. Holger Diessel. 1999. *Demonstratives: Form, function and grammaticalization*. Vol. 42. John Benjamins Publishing.
13. Holger Diessel. 2006. Demonstratives, joint attention, and the emergence of grammar. *Cognitive linguistics* 17, 4 (2006), 463–489.
14. Robert MW Dixon. 2003. Demonstratives: A cross-linguistic typology. *Studies in Language. International Journal sponsored by the Foundation “Foundations of Language”* 27, 1 (2003), 61–112.
15. Tanja Döring, Dagmar Kern, Paul Marshall, Max Pfeiffer, Johannes Schöning, Volker Gruhn, and Albrecht Schmidt. 2011. Gestural interaction on the steering wheel. In *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11 (CHI '11)*. ACM Press, New York, New York, USA, 483.
16. David Efron. 1941. *Gesture and Environment: A tentative study of some of the spatio-temporal and “linguistic” aspects of the gestural behavior of eastern Jews and southern Italians in New York city, living under similar as well as different environmental conditions*. King’s crown Press.
17. Paul Ekman and Wallace V Friesen. 1969. The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *semiotica* 1, 1 (1969), 49–98.
18. Leopoldina Fortunati. 2000. The mobile phone: new social categories and relations. *University of Trieste* (2000).
19. Norbert Freedman. 1972. The analysis of movement behavior during the clinical interview. *Studies in dyadic communication* (1972), 153–175.
20. Susan R Fussell, Leslie D Setlock, Jie Yang, Jiazhi Ou, Elizabeth Mauer, and Adam DI Kramer. 2004. Gestures over video streams to support remote collaboration on physical tasks. *Human-Computer Interaction* 19, 3 (2004), 273–309.
21. Steffen Gauglitz, Cha Lee, Matthew Turk, and Tobias Höllerer. 2012. Integrating the physical environment into mobile remote collaboration. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services*. ACM, 241–250.
22. Hans Geser. 2004. Towards a sociological theory of the mobile phone. *E-Merging media: communication and the media economy of the future* (2004), 235–260.
23. Anthony Giddens. 1984. *The constitution of society: Outline of the theory of structuration*. Univ of California Press.
24. Jean Ann Graham and Michael Argyle. 1975. A cross-cultural study of the communication of extra-verbal meaning by gestures. *International Journal of Psychology* 10, 1 (1975), 57–67.
25. Jean Ann Graham and Simon Heywood. 1975. The effects of elimination of hand gestures and of verbal codability on speech performance. *European Journal of Social Psychology* 5, 2 (1975), 189–195.
26. David Harvey. 1989. The condition of postmodernity: An enquiry into the origins of social change. *Malden, MA: Blackwell* (1989).
27. Jana M Iverson and Susan Goldin-Meadow. 2005. Gesture paves the way for language development. *Psychological science* 16, 5 (2005), 367–371.

28. Tero Jokela, Parisa Pour Rezaei, and Kaisa Väänänen. Using Elicitation Studies to Generate Collocated Interaction Methods. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct* (2016) (*MobileHCI '16*). ACM, 1129–1133.
29. Stephen Kern. 1983. The culture of time and space. Cambridge, MA: Harvard UP (1983).
30. David Kirk, Tom Rodden, and Danaë Stanton Fraser. 2007. Turn it this way: grounding collaborative action with remote gestures. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. ACM, 1039–1048.
31. Sotaro Kita and Asli Özyürek. 2003. What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and language* 48, 1 (2003), 16–32.
32. Jonathan Knighten, Stephen McMillan, Tori Chambers, and Jamie Payton. 2015. Recognizing social gestures with a wrist-worn smartband. In *Pervasive Computing and Communication Workshops (PerCom Workshops), 2015 IEEE International Conference on*. IEEE, 544–549.
33. Christian Kray, Daniel Nesbitt, John Dawson, and Michael Rohs. 2010a. User-defined gestures for connecting mobile phones, public displays, and tabletops. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*. ACM, 239–248.
34. Christian Kray, Daniel Nesbitt, John Dawson, and Michael Rohs. 2010b. User-defined gestures for connecting mobile phones, public displays, and tabletops. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*. ACM, 239–248.
35. Joanne Lavergne and Doreen Kimura. 1987. Hand movement asymmetry during speech: No effect of speaking topic. *Neuropsychologia* 25, 4 (1987), 689–693.
36. Charlotte Magnusson, Kirsten Rassmus-Gröhn, and Delphine Szymczak. 2010. Scanning angles for directional pointing. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*. ACM, 399–400.
37. Sven Mayer, Katrin Wolf, Stefan Schneegass, and Niels Henze. 2015. Modeling distant pointing for compensating systematic displacements. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 4165–4168.
38. David McNeill. 1985. So you think gestures are nonverbal? *Psychological review* 92, 3 (1985), 350.
39. David McNeill. 1992. *Hand and mind: What gestures reveal about thought*. University of Chicago press.
40. Sushmita Mitra and Tinku Acharya. 2007. Gesture recognition: A survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 37, 3 (2007), 311–324.
41. Saied Moezzi, Arun Katkere, Don Y Kuramura, and Ramesh Jain. 1996. Immersive video. In *Virtual Reality Annual International Symposium, 1996., Proceedings of the IEEE 1996*. IEEE, 17–24.
42. Meredith Ringel Morris. Web on the Wall: Insights from a Multimodal Interaction Elicitation Study. In *Proceedings of the 2012 ACM International Conference on Interactive Tabletops and Surfaces* (2012) (*ITS '12*). ACM, 95–104.
43. Meredith Ringel Morris, Jacob O. Wobbrock, and Andrew D. Wilson. Understanding Users' Preferences for Surface Gestures. In *Proceedings of Graphics Interface 2010* (2010) (*GI '10*). Canadian Information Processing Society, 261–268.
44. Leysia Palen, Marilyn Salzman, and Ed Youngs. 2001. Discovery and integration of mobile communications in everyday life. *Personal and ubiquitous computing* 5, 2 (2001), 109–122.
45. A. Pance. 2016. Transparent electronic device. (2016). US Patent 9,367,093.
46. Isabella Poggi. 2008. Iconicity in different types of gestures. 8, 1 (2008), 45–61.
47. Frances H Rauscher, Robert M Krauss, and Yihsiu Chen. 1996. Gesture, speech, and lexical access: The role of lexical movements in speech production. *Psychological Science* 7, 4 (1996), 226–231.
48. Siddharth S Rautaray and Anupam Agrawal. 2015. Vision based hand gesture recognition for human computer interaction: a survey. *Artificial Intelligence Review* 43, 1 (2015), 1–54.
49. Ruth Rettie. 2005. Presence and embodiment in mobile phone communication. *PsychNology Journal* 3, 1 (2005), 16–34.
50. Jaime Ruiz, Yang Li, and Edward Lank. 2011. User-defined motion gestures for mobile interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 197–206.
51. Christopher Schmandt and Eric A Hulteen. 1982. The intelligent voice-interactive interface. In *Proceedings of the 1982 conference on Human factors in computing systems*. ACM, 363–366.
52. Douglas Schuler and Aki Namioka. *Participatory Design: Principles and Practices*. CRC Press.
53. Mandana Seyfeddinipur, Sotaro Kita, C Cave, I Guaitella, and S Santi. 2001. Gesture and dysfluency in speech. *Oralite et gesturalite: Interactions et comportements multimodaux dans la communication* (2001), 266–270.

54. Shaikh Shawon Arefin Shimon, Sarah Morrison-Smith, Noah John, Ghazal Fahimi, and Jaime Ruiz. Exploring User-Defined Back-Of-Device Gestures for Mobile Devices. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services* (2015) (*MobileHCI '15*). ACM, 227–232.
55. Michael Tomasello. 2009. *Constructing a language*. Harvard university press.
56. Radu-Daniel Vatavu and Ionut-Alexandru Zaiti. Leap Gestures for TV: Insights from an Elicitation Study. In *Proceedings of the 2014 ACM International Conference on Interactive Experiences for TV and Online Video* (2014) (*TVX '14*). ACM, 131–138.
57. N. Vats. 2014. Flip mobile phone with transparent display. (2014). US Patent D712,867.
58. Barry Wellman. 2001. Physical place and cyberplace: The rise of personalized networking. *International journal of urban and regional research* 25, 2 (2001), 227–252.
59. Morton Wiener, Shannon Devoe, Stuart Rubinow, and Jesse Geller. 1972. Nonverbal behavior and nonverbal communication. *Psychological Review* 79, 3 (1972), 185.
60. Jacob O. Wobbrock, Meredith Ringel Morris, and Andrew D. Wilson. User-defined Gestures for Surface Computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2009) (*CHI '09*). ACM, 1083–1092.