http://dx.doi.org/10.14236/ewic/HCI2018.11

Deictic Communication across Distances: Visualising Remote Pointing Gestures on Mobile Devices

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

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

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

Deictic expressions, such as '*what is that*' while pointing at an object, play an important role in face-to-face communication, for example when describing locations and orientation or when identifying objects. If two parties are not collocated, e.g. when communicating via mobile phones, such deictic expressions cannot easily be exchanged between the remote parties. In this paper, we propose three ways to visualise deictic pointing gestures to a remote communication partner: 1) *fingerprint overlay, 2) natural hand overlay* and *3) map-with-viewshed (see Fig. 1)*. We evaluated these visualisations in a lab-based user study, where participants had to identify various realistic targets on a mobile phone. Overall, participants preferred and were most successful with *fingerprint overlay*. We also identified properties of the target objects that affected how well a pointing gesture could be transmitted. Our results can inform the design of future interfaces to transmit pointing gestures across distances.

Remote, pointing gesture, mobile phone, visualisation.

1. INTRODUCTION

When people communicate in a face-to-face setting, spoken language plays a key role, but various other elements are also essential to ensure successful communication. *Deictic expressions* are particularly ubiquitous and important among these elements. They include terms such as "this", "that" or "there" but pointing gestures as well. Frequently, spoken terms and pointing gestures are combined to convey meaning. Deictic expressions are used, for example, to establish joint attention, to express where people or objects are located, or to describe the spatial relation between entities (Diessel 2006). Given the prevalence of deictic communication in everyday interactions, it is not surprising that deictic expressions are among the first words children learn (Carpenter et al. 1998), and that deictic terms are present in almost any known language (Diessel 1999). A typical example of deictic communication in a face-to-face scenario would be one person uttering *'what is that?'* while pointing at an object. Without the deictic information contained in the message ("that" and pointing gestures), the other party will have difficulties to identify what the speaker is referring to. While deictic information plays an important role in face-to-face communication, it is often misunderstood or lost

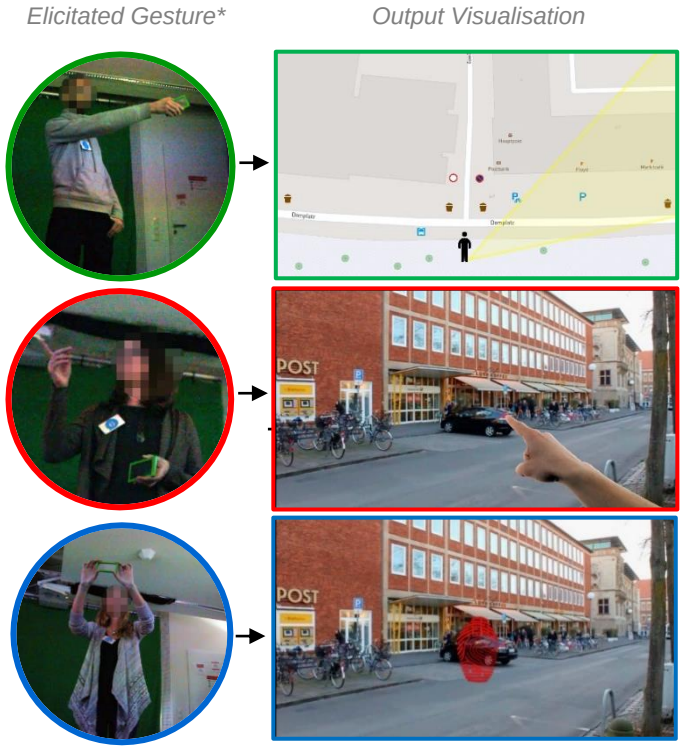


***Figure 1**: Device pointing + map-with-viewshed (green), free-hand pointing + natural hand overlay (red), and, see-through pointing + fingerprint overlay (blue)*
** Navas Medrano et. al (2017).*

Figure 2: *Top-down; Scenario 1 (S1), Scenario 2, and, Scenario 3 (S3) by Navas Medrano et.al (2017) used in the study.*

when communicating over a distance. When making a phone or video call both parties lack information about the other party's immediate environment and about their relative position and orientation to each other. Modern mobile devices facilitate strategies to somewhat compensate for these issues such as adding verbal descriptions, taking pictures or videos (and annotating them), or sending map locations. However, as these strategies differ substantially from natural behaviour, they may be insufficient and lead to misunderstandings. Furthermore, they require additional steps compared to face-to-face communication and they deviate from natural behaviour in collocated settings.

Our goal is thus to facilitate the use of intuitive deixis in remote communication. Specifically, we investigate deictic pointing gestures in a mobile setting. To realise real-time deictic communication between two non-collocated parties via mobile phones, we need to understand how people perform such gestures (Navas Medrano et al. 2017), to capture pointing gestures on a mobile device and then communicate these gestures effectively to the remote party.

In this paper, we focus on the latter aspect by proposing three different methods to visualise pointing gestures and evaluating them through a lab-based user study. Our key contributions are (1) three visualisation methods (*map-with-viewshed, natural hand overlay* and *fingerprint overlay*) to display deictic gestures to a remote party (see Figure 1); (2) the characterisation of these methods through a lab-based user study; and (3) the identification of how the size of the target object affects the success rate of different visualisation methods. Our results can benefit researchers and practitioners working on facilitating natural communication over distances.

2. DEICTIC POINTING

When discussing *deictic* pointing, it makes sense to look beyond the preconceptions commonly associated with pointing in the Human-Computer Interaction literature. Pointing in HCI has mainly been understood as a way of providing input to a system, i.e. to identify a target object. In face-to-face communication between two people, deictic pointing is more than *index finger* pointing to a specific target object. It is a complex communicative act that is subject to contextual and cognitive factors and often relays spatial information.

Different classifications of gestures have been proposed by various researchers before, as Wundt (1973), Efron (1972), Ekman and Friesen (1969) or McNeill (1992). All of them have categorised pointing or deictic gestures as a separated distinctive category. However, this provides limited detailed information about how the gesture has actually been performed. Later on literature, researchers realised there are many possibilities to perform deictic gestures, for example using the head instead of the index finger movement. Different realisations might even change the meaning of the gesture (Calbris 1990). Kita (2003) argued that due to the ubiquitous nature of pointing and the fact that humans interpret it quite easily, it might look as if pointing were a trivial phenomenon. However, the author states that '*the versatility and interpretability of pointing are based on complex underlying biological, psychological and semiotic processes.'*

As mentioned before, people perform pointing in different ways, for example, by just using one's gaze or rotating one's body into a certain direction. We can point with our hands or use various tools for that purpose, such as a simple stick or a mobile phone. Performing pointing gestures often involves more than just a single static gesture. Pointing gestures can be dynamic and iconic to represent shape or motion. They can also be abstract (McNeill 2008), if a person is pointing to something that does not physically exist.

Furthermore, pointing can be ambiguous in many ways (Kita 2003). Ambiguity does not have to be a negative aspect that needs to be avoided at all cost. Sometimes, ambiguity is intended by a person, e.g. when they are not sure or explicitly want to indicate multiple options without picking a specific one. In such a scenario, approaches for remote communication which focus solely on identifying the pointing target would falsely map the pointing gesture to one object rather than conveying what the person actually meant. This is one reason why it makes sense to transmit the deictic pointing gesture to a remote recipient in a natural way as it enables that person to interpret the gesture in its context and deal with the ambiguity that comes with it. For example, a person might perform a pointing gesture to describe an abstract spatial element (e.g. describing the house she will build in an empty lot). Since the abstract target (the house) does not exist, it cannot be easily identified and transmitted across distances. Transmitting the gesture together with its context may enable the receiver to recreate the target in their imagination.

3. RELATED WORK

In the following, we shortly discuss relevant work in related areas, in particular on deictic information and on remote communication via mobile phones.

3.1 Deictic Communication

Though deictic communication is an active research field across multiple disciplines, there is no universally agreed upon definition that covers all possible facets or uses. Deictic expressions are not limited to language alone as gestures play an important role as well. Frequently, spoken and gestural deictic expressions are combined. Previous research has shown, for example, that deictic demonstratives are more related to gestures than other words (Diessel 2006). There is evidence that people are more likely to use gestures when communicating spatial information (Alibali et al. 2001). Additionally, there is evidence on being more challenging not only to understand but also to express spatial information without the support of gestures (Rauscher et al. 1996). Occasions where deictic gestures are used include, for example, describing the layout of a room or building (Seyfeddinipur et al. 2001), an irregular shape (Graham and Argyle 1975), or motion in space (Kita and Özyürek 2003). Giving directions (Allen 2003) is another prominent example where deictic gestures are key to the successfully transmission of spatial information. This example highlights why it is important to communicate deictic information to a remote party: when giving directions to a remote person or when answering spatial questions via a mobile phone connection, not being able to easily relay deictic information can lead to misunderstanding or lengthy explanations.

3.2. Remote Communication

Mobile technology enables remote human interaction and eliminates the requirement to be co-present in order to interact (Giddens 1984; Kern 1983; Harvey 1989). Remote communication has had a profound impact on how people live and work and has changed how people understand time and space (Cairncross 2001; Bauman 2000). With the arrival of high-speed mobile network access, video traffic has been on the rise (Liu et al. 2008). Using video conferencing is one possible option to convey deictic information, but previous work has identified a number of limitations with this kind of system (Eisert 2003; Egido 1990, 1988), including eye contact (D'Angelo and Gergle 2016) and shared understanding (Isaacs and Tang 1994). Various approaches have been proposed to improve video conferencing in order to overcome these limitations. Fussell et al. (2004) and Kirk et al. (2007), for example, investigated tools and approaches that enable gesture interaction for remote collaboration in video systems. Gauglitz et al. (2012) proposed a framework to enable mobile remote collaboration on tasks that involve the physical environment. Wong and Gutwin (2010) looked at how people produce and understand pointing gestures when interacting with collaborative virtual environments. Avellino et al. (2015) investigated how accurately users can understand remote pointing from video footage on a large wall-sized display. Systems such as HyperMirror try to tackle issues such as deictic communication by using a mirror metaphor that places local and remote partners in a joint 'mirror' space (cf. e.g. Morikawa and Maesako 1998; Hirata et al. 2008). While these approaches address some of the shortcomings of video conferencing systems, they frequently require an extensive infrastructure (e.g. public displays and/or sensor arrays) or detailed knowledge about the environment of the communication party. Therefore, these approaches are not well suited for real-time use in an unstructured real-world setting.

3.3. Remote Deictic Communication

Since the seminal paper describing the Put that there system (Bolt 1980), various technologies and systems have been developed to enable gestural interaction, such as, accelerometer-based (e.g. WiiRemote) or camera-based approaches (e.g. Kinect). Mitra and Acharya (2007) and Rautaray and Agrawal (2015) provide comprehensive surveys of technological approaches for hand gestures recognition. Several different types of gestures can be realised with these technologies, including deictic ones. Mayer et al. (2015), for example, proposed pointing interaction using an allocentric perspective. The authors investigated hand pointing as well as gaze pointing (Akkil and Isokoski 2016). However, such approaches frequently require an instrumented environment to work properly, which limits their use to those augmented areas.

Other approaches have looked at enabling pointing gestures in a more mobile context. The ShowMe system (Amores et al. 2015) is based on a head-mounted display and facilitates pointing (amongst other gestures) in a mobile setting by recording arm movements and transmitting them to a communication partner to support remote collaboration. Myopoint (Haque et al. 2015) captures deictic gestures using electromyographic and inertial measurements that are taken by an arm-mounted commercial device (but are not transmitted to a remote party). While these approaches are more mobile than those discussed in the context of video conferencing, both still require additional hardware for capturing gestures. In addition to capturing deictic gestures, remote communication also entails visualising these to the remote communication partner. Balakrishnan et al. (2008) investigated how different visualisations of information in general improved remote collaboration performance. Other

work has looked at map-based visualisations and how to communicate actions by remote collaboration partners (Fechner et al. 2015). Biocca et al. (2006) proposed attention funnels as a form of visualisation for mobile augmented reality platforms intended to support collaboration. Gauglitz et al. (2012) introduced a framework which facilitates textual annotations by remote collaboration partners. While these forms of visualisations can thus remotely communicate certain aspects, they are not specifically targeted at relaying deictic gestures. In addition, they are not grounded in how people actually perform such gestures in a mobile setting.

Navas Medrano et al. (2017) carried out an elicitation study to shed light on the latter aspect. The authors placed participants in an immersive video environment showing a number of real-world scenes and then tasked them with pointing to a series of objects using a mobile device. They identified three main methods that people spontaneously use to perform deictic gestures with a mobile phone: see-through pointing, free-hand pointing and device pointing. Participants who used see-through pointing held up the smartphone so that the imagined camera captured the scene and then touched the screen at the location where the target object was shown. Those using free-hand pointing just used their hand to point at an object while holding the mobile phone in the other hand. In the case of device-pointing, participants held the phone in the pointing hand and aimed it at the target as if it were a tool or an extension of their body. In the following section, we propose three visualisation techniques for remote deictic communication that are based on these findings about how people point with mobile phones (Navas Medrano et al. 2017).

3.4. Diversity and Deictic Pointing

Humans can perform pointing gestures in widely diverse ways. Even if the most common (human) way of performing pointing gesture is using the hand, it is sometimes performed using the head, the gaze, by protruding the lips (Sherzer 1973), or by a movement of the elbow. Hand pointing can also be done with a tool as an extension of your arm.

Researchers have considered pointing as being critical in the evolution of language (Hewes 1981, 1996; Rolfe 1996) and as one way that distinguishes humans from primates (Povinelli and Davis 1994). Some researchers even argue it is an innate part of human communication (Bates et al. 1987). Previous research has investigated if humans are biologically programmed to point with an extended index finger making index-finger pointing universal across cultures. Povinelli and Davis (1994) claimed that index finger pointing is a universal human behaviour found in every known culture. However, Wilkins' (2003) preliminary findings suggested that there might be cultures where adults do not perform or rarely perform index pointing. This would suggest that pointing might not be universal in sociocultural and semiotic terms. Wilkins (2003) argues that forms of pointing (e.g. flat hand pointing) differ from culture to culture. Kendon and Versante (2003) found six different kinds of manual pointing in Neapolitan population, which is well known for its rich gesture culture. Two forms include the use of the index finger, one only the thumb and in the three remaining ones, the hand has all the fingers extended. The authors concluded the differences in the way of pointing might have an impact on the message intended to be transmitted. The different ways of hand pointing are thus not used arbitrarily.

In the case of lip pointing, Hewes (1996) and Morris (1978) argue that deliberate lip pointing happens only in cultures where finger pointing is considered rude or taboo. However, it has been discovered that lip pointing is more predominant than index pointing in some cultures (e.g. Awtuw speakers in Papua New Guinea (Feldman 1986)). In others, finger pointing is very rare or even non-existent (e.g. Kuna indigenous in Panama (Sherzer 1973, 1993) and Barai of Papua New Guinea (Wilkins 2003).

In summary, we can thus conclude that the act of deictic pointing is universal in humans, and forefinger pointing undoubtedly is the most common form and a canonical case of pointing (Franco and Butterworth 1996). However, we need to acknowledge and consider the existence of different variations in pointing across cultures and individuals. Therefore, it is also relevant for technology facilitating remote deictic communication to support different ways of pointing both, on the producer's and the recipient's side. Consequently, there is a need to provide technology that can sense different types of deictic pointing gestures, and that can then relay these gestures to the remote party in an appropriate way.

4. VISUALISING DEICTIC GESTURES ON MOBILE DEVICES

In order to enable remote deictic communication on mobile devices, there are several contextual aspects to consider. Mobile users are free to move in space and can thus be located in many different environments. While it is safe to assume some basic information (e.g. map data) is available for most locations, we cannot presume that detailed 3D information or a sophisticated external infrastructure (beyond network connectivity) exists. In addition, it is highly desirable to convey deictic information without delay to not interrupt the flow of a conversation and to account for the high mobility of users. Consequently, approaches relying on smart environments, image recognition, and detailed world knowledge are not well suited for conveying deictic gestures in mobile settings.

An alternative approach to solving there issues is to find out how people perform pointing gestures with a mobile phone (Navas Medrano et al. 2017). Then, it is critical to visualise information about how such a deictic gesture was performed to the remote communication partner, based on what can easily be measured by built-in sensors of a smartphone. Rather than using image recognition or other AI-based approaches to identify the target object of a pointing gesture and then presenting this directly to the receiver, the presented alternative approach relies on the receiver to identify the object from a faithful visual representation of the remote deictic pointing gesture. The key advantages of this approach are: it does not require a detailed digitalised model of the world, it can work with built-in sensors alone and, facilitates real-time communication. In previous work, Navas Medrano et al. (2017) investigated how people envision remote pointing when using mobile devices. They reported on an elicitation study where they asked participants to perform a series of remote pointing tasks. Their work resulted in initial insight into three different natural user behaviours specific to this context: *see-through, free-hand,* and *device pointing* (Figure 1, left). Based on these considerations, we designed a visualisation method to depict each of the pointing gestures that we had elicited in our previous study (Figure 1, right).

4.1. Fingerprint overlay

One of the two most frequently observed pointing interactions with a mobile phone as reported by Navas Medrano et al. (2017) was *see-through pointing*. Here, people used the built-in camera to capture a scene and then tapped the display of their mobile to point at objects. Based on Gauglitz video annotations (Gauglitz et al. 2014), we propose the visualisation shown in Figure 1 (bottom) to convey deictic information: *fingerprint overlay*. We overlay the scene captured by the camera with a fingerprint that corresponds to the area touched by the fingertip of the person performing the deictic gesture. In order to realise this visualisation, no external infrastructure or knowledge is needed as the scene itself is captured by the camera of the smartphone and the touch events are registered by the device as well.

Other approaches, such as annotations or highlights, would not be as natural for reflecting *see-through pointing* behaviour. Visualisations identifying the target of the pointing gestures and highlighting the pixels of that target, or its contour would require more intelligent algorithms that are always able to identify the target. In some cases (e.g. abstract deixis) it is not feasible. An annotation such as drawing a complex shape around the target rather than covering it with a fingerprint overlay would prevent the partial occlusion of the target but it would be inconsistent to how the user behaved in Navas Medrano et al. (2017). In a real-time communication scenario, it is much faster to just tap on the screen.

4.2. Natural hand overlay

The second most observed method for mobile phone pointing reported in Navas Medrano et al. (2017) was *free-hand pointing*. Participants performing this interaction pointed directly at objects with one hand while holding the mobile phone in the other, in the ‘put-that-there’ (Bolt 1980) spirit. To communicate essential information about this pointing method, we designed the natural hand overlay visualisation method shown in Figure 1 (middle). It overlays the scene captured by the camera of the smartphone held in the user’s one hand with their other arm pointing at the target object. This visualisation can also be realised easily without any external infrastructure or knowledge as the smartphone camera can simultaneously capture the scene and the user’s arm pointing at the target.

4.3. Map-with-viewshed

Navas Medrano et al. (2017) reported *device pointing* as the third most frequently observed method for mobile phone pointing. It was used by a substantial number of users though less frequently than the other two methods. In *device pointing*, participants used their smartphone as a laser pointer: They held it in their hand while extending their arm towards the target so that the device would point at it. The visualisation we designed for this case (*map-with-viewshed*, see Figure 1, top) consists of a map that is augmented by the user’s location and a viewshed extending from the user’s location in the direction the device is pointing. The basemap was extracted from Open Street Map Mapnik (OSM) and is rotated and zoomed to include nearby objects. Additionally, we added icons for potential targets that were not included in the basemap such as parked cars or trash bins. We prioritised the use of standard OSM icons to maintain consistency with the original style but used additional symbols when needed (e.g. for street lights). With respect to traffic signs, we used the ones from the 1968 Vienna Convention on Road Signs and Signals 623 (Economic Commission for Europe, Transport Division, 2007).

We used maps as an abstract visualisation, as they are common interfaces that can be realised without very detailed work knowledge (e.g. a 3D model) and without an external infrastructure. Maps are feasible to implement with current technology. The built-in sensors of a smartphone can provide location and orientation of the device. Such information can be used to automatically select the map area that is depicted and to overlay the viewshed. The viewshed representation was chosen since the orientation information returned by current smartphone sensors

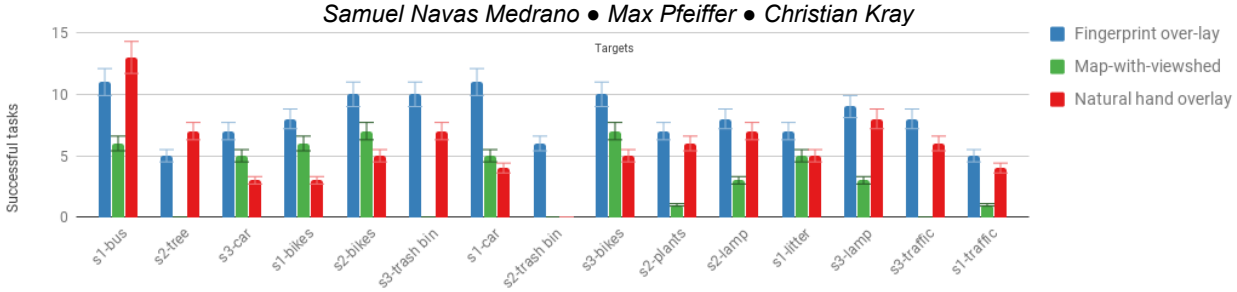


Figure 3: *Small targets that are successful identified for the three visualizations of the three scenarios (S1-S3).*

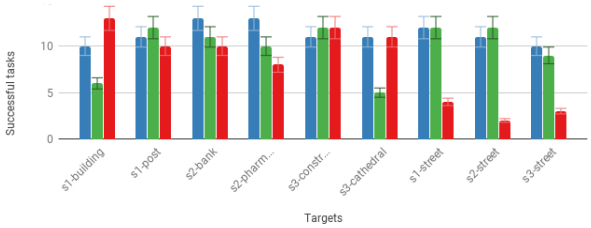


Figure 4*: Large targets that are successful identified for the three visualizations of the three scenarios (S1-S3).*

frequently is not very accurate. This representation is also widely used by other systems (e.g. Google maps) and provides a realistic representation of location ambiguity. However, it is highly desirable for the basemap being used to be augmented with some additional information. The *map-with-viewshed* visualisations would represent the physical action of pointing performed by the sender.

5. STUDY

In order to evaluate how *fingerprint overlay*, *natural hand overlay* and *map-with-viewshed* visualisations convey deictic gestures to remote parties, we carried out a user study contrasting the three types of visualisations. Our overall aim was to gain an initial understanding of their (relative) effectiveness, their properties and dependencies, and to gather feedback on how pointing gestures information is (best) communicated to a remote partner's mobile phone. The study was approved by the institutional ethics review board.

5.1. Study Design

We conducted a lab-based experiment to ensure a high degree of control. To avoid confounding factors resulting from including a second communication partner, we used a Wizard-of-Oz approach to simulate the remote communication process. Instead of speaking to another person, the remote partner was simulated by an automated voice, who was asking 'Which element of the does scene Sarah mean?' The experimenter played back this message before each task. We selected 24 targets from three urban scenarios (eight targets per scenario) based on the literature (Navas Medrano et al. 2017, Figure 2). To cover a wide and balanced range of different spatial properties (e.g. small/large, wide/narrow, individual objects/groups of objects). The targets were typical objects commonly found in urban environments such as buildings, streets, traffic signs or bikes. They were shown to the participants on a mobile device using the three different types of visualisations introduced (Figure 1, right). The participants then had to identify what object the (simulated) remote partner had pointed to. To avoid any order or selection effects, we performed a within-subjects study and randomized the order of exposure to stimuli. All participant completed the 72 task in random order (8 targets x 3 scenarios x 3 visualisations) in an average time of 19.84 minutes (SD: 3.48, Range: 14.41 to 25.43 minutes). We recorded the entire experiment (video and audio) for later analysis and used the Santa Barbara Sense of Direction Scale (SBSDS) test (Hegarty et al. 2002) to measure the participants' environmental spatial ability, which might influence their performance. We also composed a post-test questionnaire to gather qualitative feedback about the visualisation methods and the transmitted deictic information.

5.2. Participants

We recruited participants through flyers and online channels such as international students groups from the local university community. Thirteen people, two males and eleven females, with an average age of 25 years (SD: 5.17, Range: 21-35 years), participated in the experiment. Twelve participants were right-handed and one was left-handed. All participants owned a smartphone and had a university-level education. They all stated that they were fluent speakers of English, and that they were residents of the city where the pictures, which we used in the study, were taken. Participants received a monetary compensation of 10 EUR for their time.

5.3 Apparatus and Material

The experiment was performed in a lab environment with only the experimenter and the participant present, who sat at the table facing each other. The experimenter used a laptop to run the study, and the participant was asked to perform the tasks on a common smartphone (Samsung Galaxy S6). The verbal cue (which was consistent across all tasks) was played back on the mobile device. It consisted of a single sentence ('Which element of the scene does Sarah mean?') which was produced by a standard voice synthesiser. In order to record the interaction, a Canon EOS 55D camera was placed behind and above the participant facing towards the mobile device.

***Table 1**: Logistic regression model on a participant identifying a target successfully.*

Explanatory variable	Estimate	Exp (Estimate)
(Intercept)	*1.98****	*7.22*
method[target.map]	*-1.25****	*0.29*
method[target.hand]	*-1.06****	*0.34*
size[target.small]	*-1.48****	*0.23*
Deviance	*1132.8*	*(p=6.25e-6)*
AIC	*1140.8*	

****** Significant p < 0.001%***

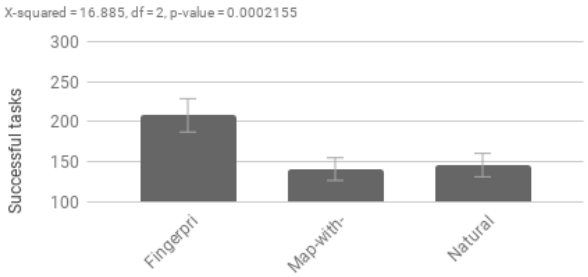


***Figure 5:** Overall successful selected targets per visualisation: map-with-viewshed overlay, natural hand overlay and fingerprint overlay.*

5.4. Procedure

After welcoming the participants, they received information about the experiment, a consent form, and a questionnaire with demographic questions (age, gender, dominant hand and expertise of using smartphones) together with the SBSDS test. Next, the experimenter briefly explained the experiment and the three visualisation methods and then handed out the smartphone to the participant. Any questions related to the experiment and the use of the device were answered. After this step, the main phase of the study began, and the recording device was started.

During the study's main phase, the 72 tasks were presented to the participants in a random order and in the following manner. First, a black screen with a 'repeat' button was shown while the audio cue was played back. If participants did not understand or wanted to hear the cue again, they could press the button to listen to the question again (which was the same for all tasks). The participants then signalled the experimenter that they were ready (verbally or via gestures). The experimenter then triggered the display of a visualisation on the mobile phone. The participants then had to identify the intended target object and verbally communicate this to the experimenter. Once they had done so (or had indicated that they were unable to identify it), again began with showing the black screen while playing back the audio cue.

After performing all tasks, all participants received the post-test questionnaire in which they were asked to answer a number of questions related to the three visualisation methods (Figure 9). Lastly,

***Figure 6**: Success rate per visualisation and target size (X^2 = 15, p=0.004).*

the experimenter debriefed them and they were paid their compensation.

5.5. Analysis

We classified each task as either a success or an error based on whether or not the intended target object was identified correctly. We disregarded any variations or difference with respect to how participants verbally described an object. For example, in the case of target building being the pharmacy, we classified the task as a success if participants referred to it by name, type or something more generic such as 'that building' as long as it was clear that they were referring to the pharmacy. The task was counted as an error if participants referred to a different target, e.g. the bike instead of the car.

During the study and the analysis, we noted a difference in user behaviour depending on the target type. We therefore decided to also analyse the results with regard to the size of target objects. For this purpose, we defined, a posteriori, two different categories: big and small targets. Big targets consisted of buildings and streets, while small targets were all other objects, which were neither streets nor buildings (e.g. bikes, cars, streetlamps, trash bins, etc.). Size in the real world also translated very well to size in pixels within the three visualisations. The only exception to this is the bus station, which is classified as a small target but overlaps in pixel size with the big targets in terms of its representation in the picture.

6. RESULTS

Together, the 13 participants performed 936 tasks in total. We analysed the resulting four hours of video footage as described in the previous section. 495 of the 936 tasks were classified as successful; on average users successfully completed 38/72 tasks (SD: 7.3), individually completing between 25 and 52 tasks correctly. Figure 5 summarises the success rates across the three visualisations. The fingerprint overlay visualisation was significantly more successful than both others ($\chi 2 = 16.885$,d f = 2, p = 0.0002), with a success rate of 208/495. The natural-hand overlay (146/495) and the map-with-viewshed visualisations (141/495) scored similarly in this respect. We did not find any significant differences regarding the three scenarios we used in

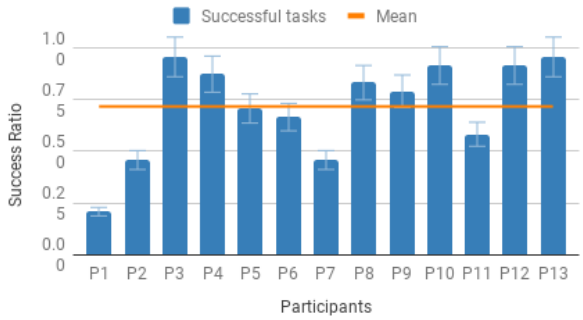


Figure 7: *Fingerprint overlay success ratio per participant.*

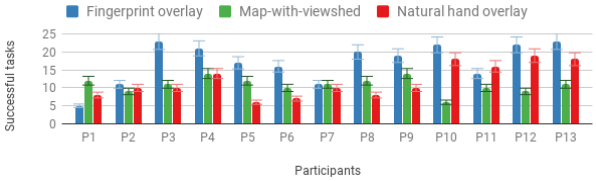
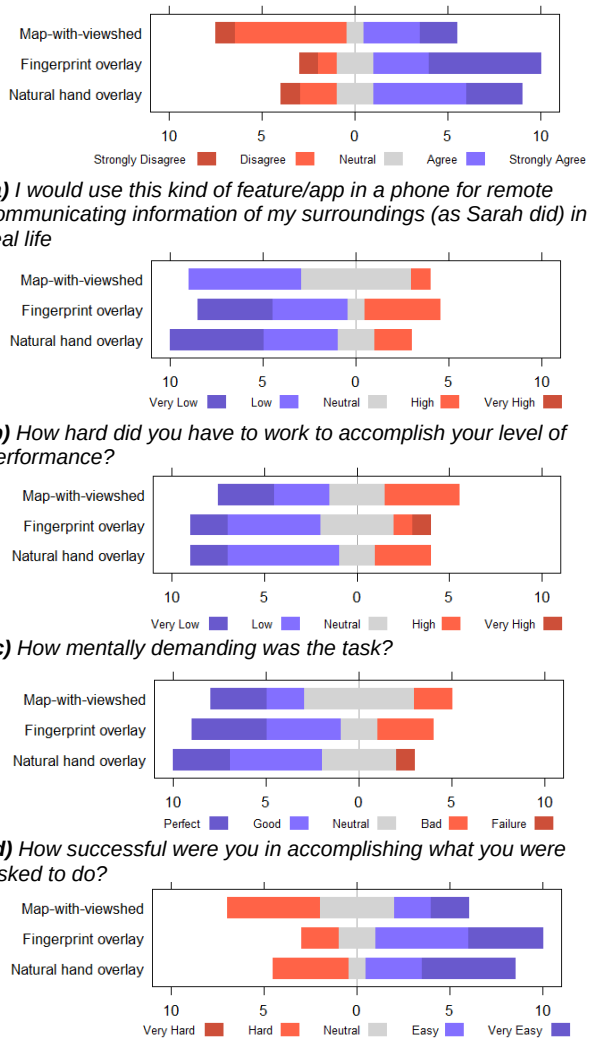


Figure 8*: Participants success trend per visualisation.*

the study. Scenario 1 accounted for 161 out of 495 successful tasks, Scenario 2 for 162 out of 495, and Scenario 3 for 172 out of 495.

After performing a Spearman Correlation test, we observed that the overall success rate per participant was not correlated to their score in the SBSDS test (r=0.44, p=0.13). However, we found evidence that the score might be influenced by the target size (r=0.7, p=0.0001). We analysed the data using a logistic regression model to further explore the impact on the target size on the success rate (see Table 1). The results suggest that the chances of a user successfully identifying a target are significantly reduced when the target is small, when it is visualised by the *mapwith-viewshed*, or by the *natural hand* overlay. Figure 6 illustrates this effect: It highlights the difference in overall performance by the fingerprint overlay, which is less affected by small target sizes compared to *natural hand* and *map-withviewshed*.

In addition to comparing success rates across all three methods, we looked into the consistency within and between methods. Figure 8 shows that while some participants performed consistently across different methods (e.g. P2, P7), there also were many who were much more successful with one particular method (e.g. P3 P4) or much less successful with another one (e.g. P10, P12). Looking at the most successful visualisation (i.e. *fingerprint overlay*), we note significant individual differences (see Figure 7). The average success rate for this method was 0.71, with individual rates varying between 0.21 and 0.96. Overall, there is no clear trend regarding the consistency of success rates. However, we noted a consistent low success rate in the case where the target was the street itself, in all scenarios, when it was visualised by *natural hand overlay* (see Figure 4). Other targets (e.g. three, plants, trash bin, and traffic sign) also performed particularly bad when visualised by *map-with-viewshed* (see Figure 3).



(a) *I would use this kind of feature/app in a phone for remote communicating information of my surroundings (as Sarah did) in real life*

(b) *How hard did you have to work to accomplish your level of performance?*

(c) *How mentally demanding was the task?*

(d) *How successful were you in accomplishing what you were asked to do?*

(e) *How easy did you understand what Sarah was saying?*

Figure 9*: Qualitative questions: (a) Intention of use, (b) Effort, (c) Mental Demand, (d) Performance, and, (e) Understanding.*

The results from the post-study questionnaire, summarised in Figure 9 shed some light on how participants perceived the different visualisation methods. In general, participants were quite positive about the fingerprint overlay and natural hand overlay. While fingerprint overlay was rated a bit more highly in the categories 'intention of use' and 'easy to understand', participants felt to some degree more successful with the natural hand overlay. The perceived performance and actual performance for the natural hand overlay visualisation were thus to a certain extent perceived inconsistent. In contrast, participants perceived the map-with-viewshed visualization as mentally demanding, arduous, difficult to comprehend. In addition, they did not feel successful using it. Finally, the participants also had a chance to provide additional feedback to the experimenter. Four participants expressed than the fingerprint icon from the Fingerprint overlay visualisation was frequently totally or partially hiding the targets. Three participants relayed that they felt ambiguity from the proposed visualisations regarding some of the tasks. Two participants noted that they were missing

contextual information from the simulated remote party when trying to complete the tasks.

7. DISCUSSION

In this study, we aimed to take the pointing interactions in remote communication already described in previous literature and design a way of presenting them as visualisations to a hypothetical remote partner in order to evaluate their feasibility.

Compared to the other two visualisations, using *fingerprint overlay* visualization had the highest success rate in identifying targets. Particularly for small targets, its success rate was in many cases substantially higher than the other two visualizations. In addition, its success rate was less reduced with regard to large targets. Still, there were still some targets and participants for which the fingerprint overlay method did not result in the maximum success rate. Additionally, the relative differences between the three methods varied between individual participants (Figure 8). Some achieved a similar success rate for all three visualisations (P2, P7), while others had an outstanding performance with one particular visualisation (P1, P3, P4-6, P8, P9). The third group of individuals had a similar success rate for two visualisations (P10-P13).

Presenting deictic information is challenging when it is transmitted across distances. When people aim to transmit information about an object in the real world, its characteristics (such as its size, surrounding and position) as well as the sender's perspective are all important to successfully interpret the message. In remote transmission, this information becomes even more important if there are other objects near the target, if the target is occluded by other objects or if the chosen visualisation method does work well with the specific target type.

7.1. Design Implications

Based on our results we identified a series of implications that could help designers, researchers and developers to gain a better understanding of how deictic pointing gestures can be conveyed over distances using mobile devices.

7.1.1. Impact of target size
Our results indicate that large targets are more easily identified than small targets. Nonetheless, while the success rates of map-with-viewshed and natural hand overlay were substantially reduced when these methods were applied to small targets, the success rate of the fingerprint overlay was less affected. Designers would be well advised to consider the size of the target object when designing interfaces that transfer and visualise deictic gestures. In addition, we encourage to investigate visualisations that cover, overlap or include the target object in a similar way to the fingerprint overlay visualisation. Even though this visualisation partially occluded the target, it still performed better than the other two visualisation types.

7.1.2. Impact of visualisation method
Previous work (Navas Medrano et al. 2017) has demonstrated that people choose different methods to point at target objects in their environment when using a mobile device. Analogously, we observed variations between people (and targets) with respect to how successful different visualisations were. While overall the *fingerprint overlay* resulted in the highest success rates, this was not consistently true for individual users or specific targets: each visualisation method was the most successful one for at least on target (see Figure 3 and Figure 4). It makes sense to consider the inclusion of multiple visualisation methods in a system that facilitates the transmission of deictic gestures. It may be possible to use some contextual information such as how the gesture was performed or what inherent properties were present in the pointer's environment to automatically select a specific visualisation method. The results we obtained for combinations of different types of objects and visualisations could then be used to inform this process. Alternatively, it could be useful to let users switch between different visualisation methods. This would not only eliminate the need for deeper contextual analysis but also account for individual preferences thereby enabling adaptive support for diverse user groups.

7.1.3. Static vs dynamic gestures
With respect to elongated targets (such as streets), which are large enough to extend across the entire shown scene (Figure 3), the natural hand overlay performed quite poorly. A possible explanation for this phenomenon is that people frequently perform dynamic pointing gestures for this type of targets (Navas Medrano et al. 2017), e.g. when describing the shape of the target (McNeill 1992). This dynamic aspect is not represented by the static natural hand overlay method but might be expected by participants in order to identify, for example, a street as the target object. We advise to animate the gesture representation (hand overlay) when communicating gestures that represent elongated targets could result in higher success rates.

7.1.4. Standardized visualisations
The *map-with-viewshed* overlay performed particularly poorly when the intended targets were visualised by non-standarised icons (e.g trash bin). Some targets even performed poorly despite being visualised by standard icons (e.g. traffic signs). One possible explanation for these observations could be that users generally perform less well with a map-based visualisation when faced with icons they do not commonly encounter on maps.

Consequently, it seems advisable to carefully test icon sets that are to be used with the *map-with-viewshed* method, in particular when icons are not commonly used on maps.

7.2. Feasibility of current technology

In this work, we focused on transmitting deictic information across distances using technology widely available. The presented visualisation methods would easily work on current existing mobile phones with no need for technical limitations or hardware expansions. The evaluation of the suggested visualisation methods proved their effectiveness for successfully transmitting deictic pointing gesture information to our participants. Our results also showed that the proposed visualisations performed well not only in ideal conditions but in challenging situations as well. They proved to have the potential to be implemented successfully in the field. Moreover, the qualitative feedback provided by our participants suggested that a technological solution implementing those visualisations would be usable and easily accepted by potential users.

7.3. Limitations

When designing the experiment, we aimed to ensure a high degree of control. As a consequence, we provided no linguistic context in the form of audio cues for each task in order to minimize any possible interference. As deictic communication is context dependent, this condition might have influenced the participant's results towards a lower performance due to the lack of verbal context, which would have made the identification of targets easier to the remote party. Furthermore, the number of participants we tasked and scenarios we depicted were limited. We chose to consider only urban scenery as giving directions is one of the de facto settings where deictic information plays a key role but also to maintain consistency with previous research. Additionally, opportunistic recruitment led to an imbalance in gender and handedness, which might have affected the outcome of the study.

We tested a limited number of possible representations for transmitting deictic information from pointing gestures. Such decision was taken in order to prove the feasibility of transmitting gestural deictic information across distance only considering existing mobile phones. Further developing in emergent technologies could expand the possibilities for communicating deictic information across distances. Further studies may be needed to assess the previously mentioned aspects.

8. CONCLUSION

In this paper, we investigated how deictic information can be transmitted and visualised in a remote communication process. Based on previous work in this research topic, we presented three visualisation methods for transmitting deictic gestures across distances and evaluated them in a user study. The study revealed that the *fingerprint overlay* visualisation performed better than *natural hand overlay* and *map-with-viewshed* visualisations. Additionally, the results indicated that the size of the target affects their probability of being successfully identified. The success rate of *fingerprint overlay* visualisation is less affected by target size. We also report participants preferences for the *fingerprint overlay* and *natural hand overlay* visualisations over the *map-with-viewshed* visualisation. From our results, we were able to derive a number of design implications for user interfaces aiming to support remote deictic communication. Our main contributions are thus initial insights into the visualisation of remote deictic gestures. Further contributions include the identification of the properties of target objects that affect how well a pointing gesture can be transmitted in mobile communication. These contributions can benefit designers and researchers developing and investigating future interfaces to communicate gestures successfully across distances. Such interfaces would be beneficial for everyday mobile communication.

This work constitutes a first step towards understanding the transmission and visualisation of deictic information in remote communication. We identified several promising directions for future research. As a next step, we plan to investigate the possible impact of contextual factors, in particular the linguistic context, in the remote pointing processes. We are also interested in the role of ambiguity when communicating deictic pointing gestures. Additional future research could explore different communication channels to present deictic pointing information to a remote party. This includes nonvisual methods such as using Electrical Muscle Stimulation to remotely control the recipient hand and arm (Kaul et al. 2016), or other types of haptic feedback (e.g. Tsukada and Yasumura 2004). From a methodological point of view, it would be interesting to further investigate the approach we used to design the visualisation (i.e. from elicited gestures) and to characterise this process in detail. Finally, it makes sense to test its feasibility by evaluating it in relevant use cases, such as providing directions to remote communication partners or remotely selecting one of many options.

9. 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).

REFERENCES

Akkil, D. and P. Isokoski (2016). Accuracy of interpreting pointing gestures in egocentric view. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 262–273. ACM.

Alibali, M. W., D. C. Heath, and H. 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), 169–188.

Allen, G. L. (2003). Gestures accompanying verbal route directions: Do they point to a new avenue for examining spatial representations? *Spatial cognition and computation 3*(4), 259–268.

Amores, J., X. Benavides, and P. 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*, pp. 1343–1348. ACM.

Avellino, I., C. Fleury, and M. Beaudouin-Lafon (2015). Accuracy of deictic gestures to support telepresence on wall-sized displays. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 2393–2396. ACM.

Balakrishnan, A. D., S. R. Fussell, and S. Kiesler (2008). Do visualizations improve synchronous remote collaboration? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1227–1236. ACM.

Bates, E., B. O'Connell, and C. Shore (1987). Language and communication in infancy.

Bauman, Z. (2000). Liquid modernity. 2000. *Polity, Cambridge*.

Biocca, F., A. Tang, C. Owen, and F. Xiao (2006). Attention funnel: omnidirectional 3d cursor for mobile augmented reality platforms. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pp. 1115–1122. ACM.

Bolt, R. A. (1980). Put-that-there: Voice and gesture at the graphics interface. In *Proceedings of the 7th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '80, New York, NY, USA, pp. 262–270. ACM.

Cairncross, F. (2001). *The death of distance: How the communications revolution is changing our lives*. Harvard Business Press.

Calbris, G. (1990). *The semiotics of French gestures*, Volume 1900. Indiana Univ Pr.

Carpenter, M., K. Nagell, M. Tomasello, G. Butterworth, and C. 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*, i–174.

D'Angelo, S. and D. Gergle (2016). Gazed and confused: Understanding and designing shared gaze for remote collaboration. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, New York, NY, USA, pp. 2492–2496. ACM.

Diessel, H. (1999). *Demonstratives: Form, function and grammaticalization*, Volume 42. John Benjamins Publishing.

Diessel, H. (2006). Demonstratives, joint attention, and the emergence of grammar. *Cognitive linguistics 17*(4), 463–489.

Economic Commission for Europe, Transport Division (2007). Convention on road signs and signals of 1968; *European agreement supplementing the convention and protocol on road markings, additional to the European agreement*. pp. ix, 276 p.

Efron, D. (1972). Gesture, race and culture. *JSTOR*.

Egido, C. (1988). Video conferencing as a technology to support group work: a review of its failures. In *Proceedings of the 1988 ACM conference on Computersupported cooperative work*, pp. 13–24. ACM.

Egido, C. (1990). Teleconferencing as a technology to support cooperative work: Its possibilities and limitations.

Eisert, P. (2003). Immersive 3d video conferencing: challenges, concepts, and implementation. In *VCIP*, pp. 69–79.

Ekman, P. and W. V. Friesen (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *semiotica 1*(1), 49–98.

Fechner, T., D. Wilhelm, and C. Kray (2015). Ethermap: real-time collaborative map editing. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 3583–3592. ACM.

Feldman, H. (1986). A grammar of awtuw.(= pacific linguistics series b 94). *Canberra: Australian National University*.

Franco, F. and G. Butterworth (1996). Pointing and social awareness: Declaring and requesting in the second year. *Journal of child language 23*(2), 307–336.

Fussell, S. R., L. D. Setlock, J. Yang, J. Ou, E. Mauer, and A. D. Kramer (2004). Gestures over video streams to support remote collaboration on physical tasks. *HumanComputer Interaction 19*(3), 273–309.

Gauglitz, S., C. Lee, M. Turk, and T. 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*, pp. 241–250. ACM.

Gauglitz, S., B. Nuernberger, M. Turk, and T. Höllerer (2014). World-stabilized annotations and virtual scene navigation for remote collaboration. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*, pp. 449–459. ACM.

Giddens, A. (1984). *The constitution of society: Outline of the theory of structuration*. Univ of California Press.

Graham, J. A. and M. Argyle (1975). A cross-cultural study of the communication of extra-verbal meaning by gestures. *International Journal of Psychology 10*(1), 57–67.

Haque, F., M. Nancel, and D. Vogel (2015). Myopoint: Pointing and clicking using forearm mounted electromyography and inertial motion sensors. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 3653–3656. ACM.

Harvey, D. (1989). The condition of postmodernity: An enquiry into the origins of social change. *Malden, MA: Blackwell*.

Hegarty, M., A. E. Richardson, D. R. Montello, K. Lovelace, and I. Subbiah (2002). Development of a self-report measure of environmental spatial ability. *Intelligence 30*(5), 425–447.

Hewes, G. W. (1981). Pointing and language. *Advances in Psychology 7*, 263–269.

Hewes, G. W. (1996). A history of the study of language origins and the gestural primacy hypothesis. *Handbook of human symbolic evolution*, 571–595.

Hirata, K., Y. Harada, T. Takada, N. Yamashita, S. Aoyagi, Y. Shirai, K. Kaji, J. Yamato, and K. Nakazawa (2008). Video communication system supporting spatial cues of mobile users. *Proc. CollabTech*, 122–127.

Isaacs, E. A. and J. C. Tang (1994). What video can and cannot do for collaboration: a case study. *Multimedia systems 2*(2), 63–73.

Kaul, O. B., M. Pfeiffer, and M. Rohs (2016). Follow the force: Steering the index finger towards targets using ems. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '16, New York, NY, USA, pp. 2526– 2532. ACM.

Kendon, A. and L. Versante (2003). Pointing by hand in neapolitan. *Pointing: where language, culture, and cognition meet*, 109–137.

Kern, S. (1983). The culture of time and space. *Cambridge, MA: Harvard UP*.

Kirk, D., T. Rodden, and D. S. Fraser (2007). Turn it this way: grounding collaborative action with remote gestures. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pp. 1039–1048. ACM.

Kita, S. (2003). *Pointing: Where language, culture, and cognition meet*. Psychology Press.

Kita, S. and A. 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), 16–32.

Liu, Y., Y. Guo, and C. Liang (2008). A survey on peer-topeer video streaming systems. *Peer-to-peer Networking and Applications 1*(1), 18–28.

Mayer, S., K. Wolf, S. Schneegass, and N. Henze (2015). Modeling distant pointing for compensating systematic displacements. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 4165–4168. ACM.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago press.

McNeill, D. (2008). *Gesture and thought*. University of Chicago press.

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

Morikawa, O. and T. Maesako (1998). Hypermirror: toward pleasant-to-use video mediated communication system. In *Proceedings of the 1998 ACM conference on Computer supported cooperative work*, pp. 149–158. ACM.

Morris, D. (1978). Manwatching: A fold guide to human behaviour.

Navas Medrano, S., M. Pfeiffer, and C. Kray (2017). Enabling remote deictic communication with mobile devices: An elicitation study. In *Proceedings of the 19th International Conference on HumanComputer Interaction with Mobile Devices and Services*, MobileHCI '17, New York, NY, USA, pp. 19:1–19:13. ACM.

Povinelli, D. J. and D. R. Davis (1994). Differences between chimpanzees (pan troglodytes) and humans (homo sapiens) in the resting state of the index finger: Implications for pointing. *Journal of Comparative Psychology 108*(2), 134.

Rauscher, F. H., R. M. Krauss, and Y. Chen (1996). Gesture, speech, and lexical access: The role of

lexical movements in speech production. *Psychological Science 7*(4), 226–231.

Rautaray, S. S. and A. Agrawal (2015). Vision based hand gesture recognition for human computer interaction: a survey. *Artificial Intelligence Review 43*(1), 1–54.

Rolfe, L. (1996). Theoretical stages in the prehistory of grammar. *Handbook of human symbolic evolution*, 776– 792.

Seyfeddinipur, M., S. Kita, C. Cave, I. Guaitella, and S. Santi (2001). Gesture and dysfluency in speech. *Oralite et gesturalite: Interactions et comportemetns multimodaux dans la communication*, 266–270.

Sherzer, J. (1973). Verbal and nonverbal deixis: The pointed lip gesture among the san blas cuna. *Language in Society 2*(1), 117–131.

Sherzer, J. (1993). Pointed lips, thumbs up, and cheek puffs: Some emblematic gestures in social interactional and ethnographic context. In *Texas linguistic forum*, Number 33, pp. 196–211. University of Texas, Department of Linguistics.

Tsukada, K. and M. Yasumura (2004). Activebelt: Belttype wearable tactile display for directional navigation. In *UbiComp*, Volume 3205, pp. 384–399. Springer.

Wilkins, D. (2003). Why pointing with the index finger is not a universal (in sociocultural and semiotic terms). *Pointing: Where language, culture, and cognition meet*, 171–215.

Wong, N. and C. Gutwin (2010). Where are you pointing?: the accuracy of deictic pointing in cves. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1029–1038. ACM.

Wundt, W. (1973). *The language of gestures*, Volume 6. Walter de Gruyter.