User-Defined Gestures for Connecting Mobile Phones, Public Displays, and Tabletops

Christian Kray, Daniel Nesbitt, and John Dawson
School of Computing Science
Newcastle University
Newcastle upon Tyne, United Kingdom
c.kray@ncl.ac.uk*

Michael Rohs
Deutsche Telekom Laboratories
TU Berlin
Berlin, Germany
michael.rohs@telekom.de

ABSTRACT

Gestures can offer an intuitive way to interact with a computer. In this paper, we investigate the question whether gesturing with a mobile phone can help to perform complex tasks involving two devices. We present results from a user study, where we asked participants to spontaneously produce gestures with their phone to trigger a set of different activities. We investigated three conditions (device configurations): phone-to-phone, phone-to-tabletop, and phone to public display. We report on the kinds of gestures we observed as well as on feedback from the participants, and provide an initial assessment of which sensors might facilitate gesture recognition in a phone. The results suggest that phone gestures have the potential to be easily understood by end users and that certain device configurations and activities may be well suited for gesture control.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—*Interaction styles, evaluation/methodology, user-centered design*

General Terms

Design, Experimentation, Human Factors

Keywords

User-defined gesture, gesture, mobile phone, large display, tabletop, multi-device interaction, device pairing

1. MOTIVATION

Recent years have seen considerable progress in terms of how users can interact with computers. Mice, touchpads, keyboards and keypads might still be the predominant form of interaction both in traditional settings and in mobile scenarios but alternative means of interaction are proliferating rapidly. Touch-enabled devices are at the forefront of this trend but other techniques such as voice-control, tangible interface elements, or gestures are on the rise as well. At the same time the number of devices users routinely interact with is growing quickly. Many people now own desktop and laptop computers, media players and a mobile phone. In addition, we are surrounded by ever more devices, e. g. public displays, interactive kiosks and ATMs. Therefore, users often find themselves in a situation where they need to interact with two devices, for example to transfer files from one device to another or to establish a connection between the two. This process can be quite complicated, in particular when one of the devices belongs to another person or a third party, and it might involve configuring network settings and learning unfamiliar control mechanisms.

One driver behind the work presented in this paper is thus to investigate ways to simplify this process, i. e. by identifying gestures and activities that could be combined to make interactions more intuitive. Gestures hold some potential in a dual-device scenario for a number of reasons: they do neither require ownership of both devices nor physical reachability. In addition, gestures can make activities visible to another party, which particularly in the context of phone-to-phone interaction can be beneficial. They are also arguably more natural than current interfaces provided for device-to-device interaction, and potentially easier to learn as well.

A second motive for our work is to gain a better understanding what kind of gestures people naturally produce with their phone. Wobbrock et al. [21] recently presented a study on such gestures in conjunction with a tabletop system, which resulted in a vocabulary of gestures for use with tabletop systems. To the best of our knowledge, no such vocabulary exists for gestures with a mobile phone, and filling this gap is thus another driver for the work presented here.

Finally, mobile phones nowadays include a variety of sensors such as cameras, NFC readers, compasses and accelerometers, which in principle could be used to detect gestures. In order to assess, whether this is a realistic assumption, and also to understand which sensors might enable the recognition of common gestures, it is necessary to collect a set of naturally produced gestures, which can then be analysed in terms of how well they can be detected using various sensors.

The remainder of the paper is structured as follows. We first review previous work on interaction techniques to connect mobile phones and other devices, as well as on gesture-based interaction in general. The main part of the paper then describes the user study we conducted to investigate the research questions outlined above. After discussing the implications of our findings on interface design for mobile devices, the paper concludes by summarizing our key results.

*{daniel.nesbitt,john.dawson}@newcastle.ac.uk

Copyright is held by the author/owner(s).
MobileHCI'10, September 7–10, 2010, Lisbon, Portugal.
ACM 978-1-60558-835-3/10/09.

2. RELATED WORK

A key inspiration for the work presented here is Wobbrock et al.'s [21] study based on the idea to present users with the effects of gestures (*referents*) and then having them perform the actions that they think produced the given effects (*signs*). While [21] focused on gestures for surface computing, we are interested in gestures for coupling mobile phones to other devices, such as other mobile phones, large displays, and interactive tabletops. The gestures we focus on are thus not restricted to a 2D plane, but are performed in 3D space.

Augmented Surfaces [16] was an early project with the goal of exchanging information between mobile devices, interactive surfaces, and physical objects. It introduces the concept of "hyperdragging" virtual objects from a mobile device display onto a projected surface and thus moving information across the boundary of devices. *Hyper Palette* [1] presents the idea of using a PDA as an input device for an interactive table. The technique combines movement across the table with tilting the PDA relative to the movement direction. Tilting the front edge (in movement direction) down is called *scoop* and is used for transferring virtual items from the table to the PDA. This gesture implements the metaphor of using the PDA as a "scoop" for virtual content on the table. Tilting the front edge (in movement direction) up is called *spread* and implements the reverse operation, i.e., transferring content from the PDA to the table. The metaphor used here is that of "spreading" items stacked in the PDA onto the table. *Pick and Drop* [14] extends direct manipulation to interactions across devices. One application example is that of creating text and graphical items on a PDA and then copying them to a nearby whiteboard. The PDA is used like a painter's tool in this case. *Touch & Interact* [5] allows for connecting mobile phones to particular positions on a large display. The system is implemented with an NFC/RFID reader integrated into the phone and a matrix of RFID tags integrated into the large display. This allows to "pick" particular media items or to "drop" them somewhere on the screen. The selection accuracy is limited by the granularity of the RFID tag arrangement.

Stitching [7] allows inputting pen gestures that span two or more devices. A user starts pen input on one screen and then moves across the bezels of the devices to the second screen. The gesture is then interpreted to perform different application-defined actions, such as copying images from one device to the other. Hinckley [6] uses "bumping" devices together as an interaction method for connecting two tablet PCs. Accelerometers attached to the devices pick up the event as simultaneous sensor readings and associate the devices accordingly. Hinckley describes various applications of such synchronous gestures, such as sharing information or dynamically tiling together displays to show an image across multiple devices. SyncTap [15] allows to connect two devices by simultaneously pressing and then releasing buttons on both devices. The co-occurrence of these events is used to establish the pairing of the two devices.

Swindels et al. [18] use infrared pens to detect pointing actions and transfer device identities between multiple devices. *Gesture Connect* [13] is a system for combining devices. It identifies selection targets using NFC and enables command input by detecting gestures with an accelerometer. Both steps are combined in a single physical action. The *ConnecTables* [19] system also allows the connection of multiple pen-operated devices to shared workspaces as well as the exchange of items between them. Built-in RFID tags and readers allow for detecting nearby devices and establishing ad-hoc connections. BlueTable [20] is a vision-based system which enables the association of a mobile device with an interactive surface. A camera detects objects placed on the table as connected components of a certain size and shape. To check whether the connected component is a mobile device the system sends a request over Bluetooth to each device in range and waits for the device to blink its IRDA port.

RELATE [4] uses special hardware to exchange radio and ultrasound signals between nearby devices to infer their relative spatial relationship with high accuracy, i. e. the relative location and orientation of a set of collocated devices. *BeepBeep* [11] is an acoustic range sensing system that sends and receives sound signals between two devices in order to infer the distance between them. This allows for ad-hoc connections between devices without additional hardware beyond speaker and microphone. In order to avoid inaccuracies in measuring the signal travel time, each device sends a signal and simultaneously listens for the arrival of its own signal at its microphone. The authors report accuracies around 1-2cm within a range of more than 10m [11]. *Point & Connect* [12] elaborates that idea and implements pointing gestures of moving one device towards another in order to enable spontaneous device pairing. These examples show that current off-the-shelf hardware can support ad-hoc device connections based on intuitive spatial gestures.

In our user study we observed several instances of impulsive gestures with a relatively large movement amplitude, such as moving the whole arm. An example of an investigation of more coarse-grained gestures is [3], which focuses on "throw" gestures to both transfer media items from a mobile device onto a large display as well as to fetch them back. *Toss-It* [23] uses similar gestures that consist of swinging actions to "throw" a media item from one PDA to another. The system tries to estimate the trajectory of the swing action and thereby the target location.

Woo and Lim [22] investigate various design options for pairing devices by touching them. They look for gestures that are based on intuitive metaphors, such as lighting a candle with another one, and also make the pairing process visible, for example through light "flowing" from one device to the other. Kela et al. [9] describe an accelerometer-based gesture system for a smart environment. Their study results suggest that gestures were especially natural when the commands had a spatial association.

There are multiple schemes for classifying human gesture. A brief overview of work on gesture classification is given in [21]. Incidental human gesture is often performed without the executing person being consciously aware of this. The types of gestures we study in this work are deliberate and conscious ways of reaching a goal in an interaction task. Karam and Schraefel [8] provide a taxonomy of gesture-based interaction in HCI.

While the majority of the research described in this section is concerned with facilitating and simplifying the interaction between two devices using some form of physical gestures, very little work (except [21]) so far has looked into which gestures users produce naturally. One goal of the work presented here is thus to fill this gap for a specific usage scenario, namely the use of a mobile phone in conjunction with another device.

It is worth mentioning that there is a broad range of techniques currently available to connect a mobile phone to another device, which does not involve any gestures. Most commonly, this is achieved either via attaching a cable to both devices or by a wireless connection (e. g. via Bluetooth, GSM, WLAN).

3. TOWARDS USER-DEFINED GESTURES FOR INTERACTION ACROSS DEVICES

As stated at the beginning of this paper, gesture-based interaction has the potential to be beneficial in terms of making activities involving two devices easier to use. A logical first step to assessing this potential is to probe users about which gestures they would produce naturally and to gather some initial feedback about these gestures in the context of a number of activities. More specifically, we were interested in finding answers to the following research questions:

- Which gestures do users produce naturally to trigger various activities involving a mobile device and another device?
- Which of these activities do lend themselves well to being triggered by gestures, and which ones do not?
- What is the impact of different types of content and devices on the gestures being generated?

In the following, we describe the study we ran to find initial answers to those questions and the results we obtained in the course of the experiment.

3.1 Study Design

As our goal was to gain a better understanding of which gestures people would produce naturally, we decided to look at a number of device combinations as well as activities that are commonly performed using those devices. In order to keep the study manageable, we initially focussed on three device combinations (each consisting of two devices only) that people might encounter in public: phone to phone, phone to public display, and phone to tabletop computer. A secondary motive for investigating those combinations was the fact that other (privately owned) devices such as laptops, desktops or media players already have established means for triggering certain actions (e.g. plugging in a cable to start synchronization), whereas this is much less the case for the scenarios we chose.

In addition to the three device configurations, we also had to select a number of activities that could be triggered by phone gestures. In order to generate a list of such activities, we first conducted an informal survey of activities that occur between two devices of any type. We used these gestures as a basis for a brainstorming session and then selected twenty activities, which could be performed between a phone and any of the three devices we had chosen previously. As we were interested in seeing whether the type of content would have an impact on the gestures being generated (e. g. would people perform different gestures depending on whether they were downloading an image or an application to their phone), we included a set of activities that only varied according to their content. Table 1 contains a complete set of the questions we used in the study.

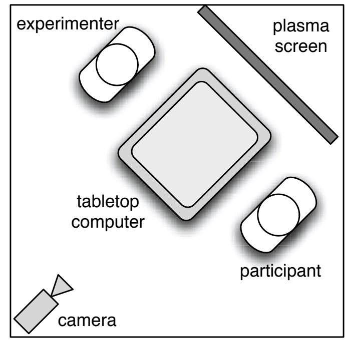

Figure 1: Top view of study setup: the experimenter was facing the participant. The tabletop systems was located between them and the public display was placed to the right of the participant. A video camera was recording each trial.

3.2 Participants

In order to recruit participants, we advertised the study through flyers, posters and various mailing lists at the university. Over the three days that the study was running, 23 people participated, of which eight were female. Participants were between 18 and 50 years old, with the majority (20) being less than 35 years.

3.3 Apparatus and Material

The study took place in a corner of a media lab, where we installed a video camera that was recording the study. Only the experimenter and the subject were present during the study. In the lab space, we had set up a tabletop system (Microsoft Surface) and large plasma screen (42"). Both devices were turned off in order to prevent any screen content from influencing the gestures people were performing. Participants were free to use their own phone during the study but if they did not want to or had not brought a phone, we provided them with a Nokia N95 smartphone. In the phone-to-phone condition, the experimenter held out a phone at about arms length so that participants would have a point of reference when performing their gestures. The distances were set up so that a participant could easily reach either device without having to move about in the room.

The study followed a within-subject design. We counterbalanced the order of exposure to the three conditions. For each condition, there were twenty questions; the order in which they were asked was randomized across conditions and participants.

3.4 Procedure

Upon arrival, participants received an information sheet, which provided them with a short overview over the study and its aims, and an initial questionnaire that contained a small number of questions about their background (age, gender, experience in using the devices used in the study). After subjects had read the information sheet and filled in the questionnaire, the experimenter provided them with more detailed information about the experiment and what would be expected of them in each trial. Participants were also informed that there would be three sections. After finishing the explanations, the experimenter provided subjects with the opportunity to ask any questions they might have. We

Table 1: Activities subjects were asked to generate gestures for (order of exposure randomised in study)

	What gesture would you perform with your phone ...
1	... to send an item that is visible on your phone screen to the other device?
2	... to send a phone number or contact from your phone to the other device?
3	... to send an application (e.g. a game) from your phone to the other device?
4	... to send a media file (e.g. photo, video, music) from your phone to the other device?
5	... to download an item that is visible on the other device's screen to your phone?
6	... to download a phone number or contact from the other device to your phone?
7	... to download an application (e.g. a game) from the other device to your phone?
8	... to download a media file (e.g. photo, video, music) from the other device to your phone?
9	... to stream a media file (e.g. video, music) from your phone to the other device?
10	... to stream a media file (e.g. video, music) from the other device to your phone?
11	... to synchronise your phone with the other device (e.g. time, calendar, contacts)?
12	... to select the other device (e.g. among a number of devices)?
13	... to authenticate your phone with the other device (e.g. to make a payment)?
14	... to scroll the screen content of the other device to the left?
15	... to abort the current interaction between your phone and the other device?
16	... to pause the current interaction between your phone and the other device?
17	... to rewind the current interaction between your phone and the other device?
18	... to move the current interaction between your phone and the other device forward?
19	... to mirror the content of your phone's screen on the other device?
20	... to vote for the content being displayed on the other device?

explicitly discouraged questions during the study to avoid situations where a participant would engage in a dialogue about their gestures. The main study began after all questions of the participant were answered.

For each condition, the experimenter would first read out a brief explanation to the subject, which highlighted the device configuration that pertained to the condition. Additionally, we instructed subjects to clearly indicate once they felt they had finished a gesture (e. g. by saying "done"). In case of the tabletop computer, this explanation also included a brief description of what a tabletop system is, since we assumed not everyone would be familiar with it. The order of the three conditions was counter-balanced. For each of the conditions, subjects had to answer twenty questions. The order of these questions was randomized across conditions and subjects.

For each question, the experimenter first read out the question to the participant. The subject then performed a gesture or indicated that they were unable to do so. Once they indicated that they had finished, the experimenter asked them to rate how well their gesture matched the device configuration and the activity described in the question. After participants had responded, the experimenter moved on to the next question. At the end of the third condition, subjects were given a final questionnaire, which contained questions about their general attitude towards using gestures with a phone to trigger activities as well as some further questions. After they filled in this questionnaire, they received a small payment to compensate them for their time. We then discharged participants from the study.

4. RESULTS

All but one participant owned a mobile phone, and all had used a public display. Eleven out of the 23 subjects had not used a tabletop computer before the study. (We were surprised by this finding as we had expected few people to even have heard about this type of system. One possible explanation for this unusually high number of people with prior exposure to tabletops is that our research group recently provided a tabletop system for an exhibition, which marked the re-opening of the most popular museum in the city, and which thus attracted a large number of visitors.) If participants had used a device beforehand, we asked them to rate their expertise on a five-point Likert scale, where one corresponded to "very inexperienced" and five to "very experienced." The average rating for expertise with a device was highest in the case of mobile phones (3.91) followed by public displays (3.59). The tabletop computer received the lowest average rating (3.00).

Each subject performed twenty gestures per condition and experienced all three conditions. In total, we thus recorded more than 1300 individual gestures. On average, it took participants about 27.5 minutes to complete the study, with the longest time taken being 45.5 minutes and the fastest being a little less than 20 minutes. Subjects took on average about nine minutes per condition, which includes the time taken to read out the initial instructions and the questions preceding and following each gesture. In total, we collected about eleven hours of video material during the study.

4.1 Gesture Analysis Method

A detailed annotation and analysis of all the gestures would require a thorough annotation according to a well-defined set of criteria to accurately capture aspects such as the 3D trajectory of the gesture or changes in velocity during the gesture. In order to assess whether this would be a worthwhile endeavor, we decided to first screen the videos for some basic gestural properties (see Table 2). Two members of the research team split up the task of annotating the entire set of gestures we had recorded. Since the properties were very basic and of factual nature (rather than of interpretative nature such as labeling a gesture as 'pointing' or 'flicking'), we did not deem it necessary to have several annotators work through the entire set in order to be able to perform a cross-validation between their individual judgements afterwards.

For each gesture, we noted down six properties. Four of them were binary attributes, where we simply recorded whether or not a property was true, and two were time measures recorded in seconds. Among the binary attributes, we looked at whether or not the distance between the mobile phone held by the subject and the other device (either a mobile phone held by the experimenter, the tabletop system or the public display) changed while the subject was performing the gesture. We also noted whether the devices physically touched. In addition, we recorded whether the mobile phone was rotated along any axis, and whether its location in space remained constant. It is worth pointing out that the latter is different from the relative distance property. For example, if the phone was moved in front of the public display while maintaining the same distance to it, we would record a location change but no change in the relative distance.

The two time based measures were the delay before producing a gesture and the duration of the gesture itself. We defined the former as the interval between the experimenter finishing reading out a question, which asked the subject to perform a gesture, and the time when subjects started to physically move the mobile phone in any way. The latter point in time was also marked as the start of the gesture. The duration of a gesture was specified as the time in seconds between the start and the end of a gesture. As part of the instructions at the beginning of each condition, we asked participants to clearly indicate when they were done with performing a gesture, i. e. by saying "done" but they frequently forgot to do this during the study. We therefore defined the end of a gesture either being marked by the subject indicating it explicitly or by them providing a rating for the quality of the match (as quite a few subjects started doing this unprompted). If neither of these clues were present, we defined the end of a gesture as the point in time when the experimenter started reading out the question about the quality of the match. Consequently, the duration measure is not very precise and provides only an indication of the upper bound of how long gestures took.

As we had not instructed participants to keep quiet during the study, quite a few subjects were either explaining the gestures as they performed them or engaged in think-aloud activities. Occasionally, this led to ambiguous situations, where participants would say something along the lines of "I would perform the same gesture as I did for activity X" and then either would not do anything or even would perform a different gesture than the one indicated. When annotating the video, if we encountered such a case we focused on noting physical movements of the mobile phone held by the subjects.

4.2 Basic Properties of Gestures

Figure 2 summarizes the results we obtained from annotating the gestures according to the basic properties outlined above. More than 70% of all gestures in all three conditions incurred a change in relative distance, and more than 70% of them resulted in a change in the (absolute) location of the mobile phone. Both rotation and touch occurred considerably less frequently across all conditions, and there were also differences between the device configurations. Despite all target devices being roughly at the same distance to the subjects, only 20% of all gestures targeting the public display involved the two devices touching. Touch events occurred twice as frequently in both the phone-to-phone and the phone-to-tabletop conditions. In terms of frequency rotation falls between touch and the other two properties, with 47 to 59% of all gestures involving a rotation of the mobile phone. Interestingly, rotations occur considerably less frequently in the phone-to-phone condition compared to the other two conditions.

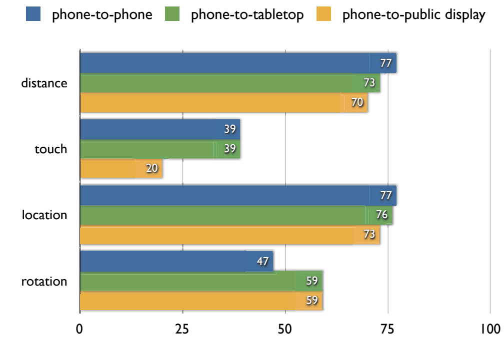

Figure 2: Basic properties of phone gestures in different device configurations (in percentage). Example: 77% of phone-to-phone gestures involved a clear change in distance between the phones.

We also measured the delay until subjects started to produce gestures as well as the overall duration of a gesture. In the phone-to-phone condition, the average delay across all subjects was approximately 4s and the average duration of a gesture was 5.0s. In case of gestures targeting the tabletop, subjects took approximately 5.0s until they started to perform a gesture, while the actual gesture took 6s to complete. In the phone-to-public display condition, the average delay was 3.5s and the average duration of a gesture 4.5s. Due to difficulties in exactly pinpointing the start and end of a gestures (see previous sections), these times should be considered as estimates of the upper boundary. Times also varied considerably both between subjects and within subjects, with some subjects taking up to 2.5 minutes before starting to produce a gesture.

4.3 Appropriateness of Gestures

After each gesture that participants performed we asked them to rate the quality of the match of their gesture and the current activity and device combination. They could choose amongst five different answers: they could respond by saying that they thought the gesture, activity and device combination matched well (well), that it was a neutral match (neutral), or that they thought that they matched poorly (poorly). Subjects could also indicate if they did not know (don't know) or if they thought that the combination did not make sense (inappropriate). Figure 3 summarizes the responses in this category. 62% of all gesture-device-activity combinations across all conditions were rated as being a good match, 16% as neutral, and 17% as being poor matches. In 4% of the cases, subjects did not know what to make of the match, and only 1% of all gesture-device-activity combinations were rated as being inappropriate.

Table 2: **Properties extracted from video footage for each recorded gesture**

property	values	how we measured it
distance	constant? (yes/no)	whether or not the relative distance between mobile phone held by the subject and target device changed during gesture
touch	yes/no	whether or not the two devices physically touched during gesture
location	constant? (yes/no)	whether or not the absolute position in space of the mobile phone held by the subject and target device changed during gesture
rotation	constant? (yes/no)	whether or not the mobile phone held by the subject was rotated along any axis during the gesture
delay	seconds	time between the experimenter reading the question and the subject starting to physically move the mobile phone
duration	seconds	time between start of the gesture (see above) and its end (see text for definition)

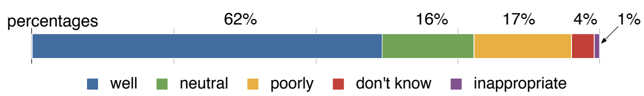

Figure 3: Quality of the match of various gesture-device-activity combinations across all conditions and subjects as rated by the participants ("How well did your gesture match the device and the activity?")

4.4 Observations

In addition to measuring various aspects and requesting direct feedback from participants, we also noted down some qualitative observations during and after the study. One such observation relates to the ease with which the majority of the participants took to producing gestures with their mobile phones. There was little confusion about what was asked of them and they were able to quickly perform gestures. Our overall impression was that the concept of phone gestures was very easy to understand and to put in practice.

For the two larger devices (tabletop and the public display) participants frequently talked about different regions that could be displayed on their screen and that could be associated with certain functions or activities. This was not the case for the phone-to-phone condition. In general, the set of gestures seemed to vary between different device configurations (see also section 5) but further analysis of the video data is needed to quantify this aspect exactly.

In terms of the gestures participants performed throughout the study, we observed a very large range of different types. These variations occurred both within and between subjects, and we observed far too many to provide a full account of them in this paper. Some common gestures included pointing the mobile phone at another device, pulling gestures (where the phone was pulled away from the other device) and flicking gestures (where the phone was moved along a short trajectory with considerable acceleration/deceleration component). Gestures people would perform less frequently were pouring gestures (where the phone was held at an angle with respect to the target device and the moved as if pouring liquid from it onto the other device) as well as directional touching (where the phone touched the other device at non-obvious angles or from the side). Uncommon gestures included, for example, placing the mobile phone on the top edge of the public display and then performing hand gestures over the display, or using the second hand in a scooping motion to move data from the target device to the mobile phone, which was held at a constant distance from the target device. The following section provides a more detailed account of a number of example gestures.

4.5 Example Gestures

Figure 4 shows four example gestures, which we recorded during the study. The photos shown in the figure are taken from the video footage we recorded, and show three steps involved in performing these gestures. On the figure, time advances from left to right, so (a) occurs before (b) and (b) occurs before (c). The topmost row depicts a typical 'pull away' gesture, where the phone is initially held in close proximity of the target device (here: another mobile phone) and is then pulled back towards the user, thereby increasing the distance between the two devices. This gesture was performed in response to question number 15 in Table 1.

The second row from the top (d-f) shows a 'pointing' gesture. The mobile phone is initially held in a default position and is then moved slightly and rotated so that a particular side of it (here: the top) is pointing in the direction of the target device (here: the tabletop). The phone is then held in this position for a certain amount of time (without touching the other device), and after some time moved back to the original (default) position. This gesture was performed in response to question 1 in Table 1. The third row from the top (g-i) shows a 'touch' gesture. The mobile phone is moved from the start position towards the target device (here: the public display) until physical contact is established. This gesture was a response to question 11 in Table 1.

The bottom row shows a more unusual 'facing gesture'. The participant moves his phone from the start location towards the target device (here: another mobile phone) while rotating it so that the phone screen is pointing downwards. He then brings his phone very close to the other phone so that their screens are facing but without establishing physical contact. This gesture was performed in response to question 13 in Table 1.

4.6 General Feedback

The final questionnaire contained a number of questions probing whether or not subjects could imagine using phone gestures to interact with another device, and which activities and devices they thought would work well in this context. The first section of the questionnaire asked participants to rate their degree of agreement with a number of statements

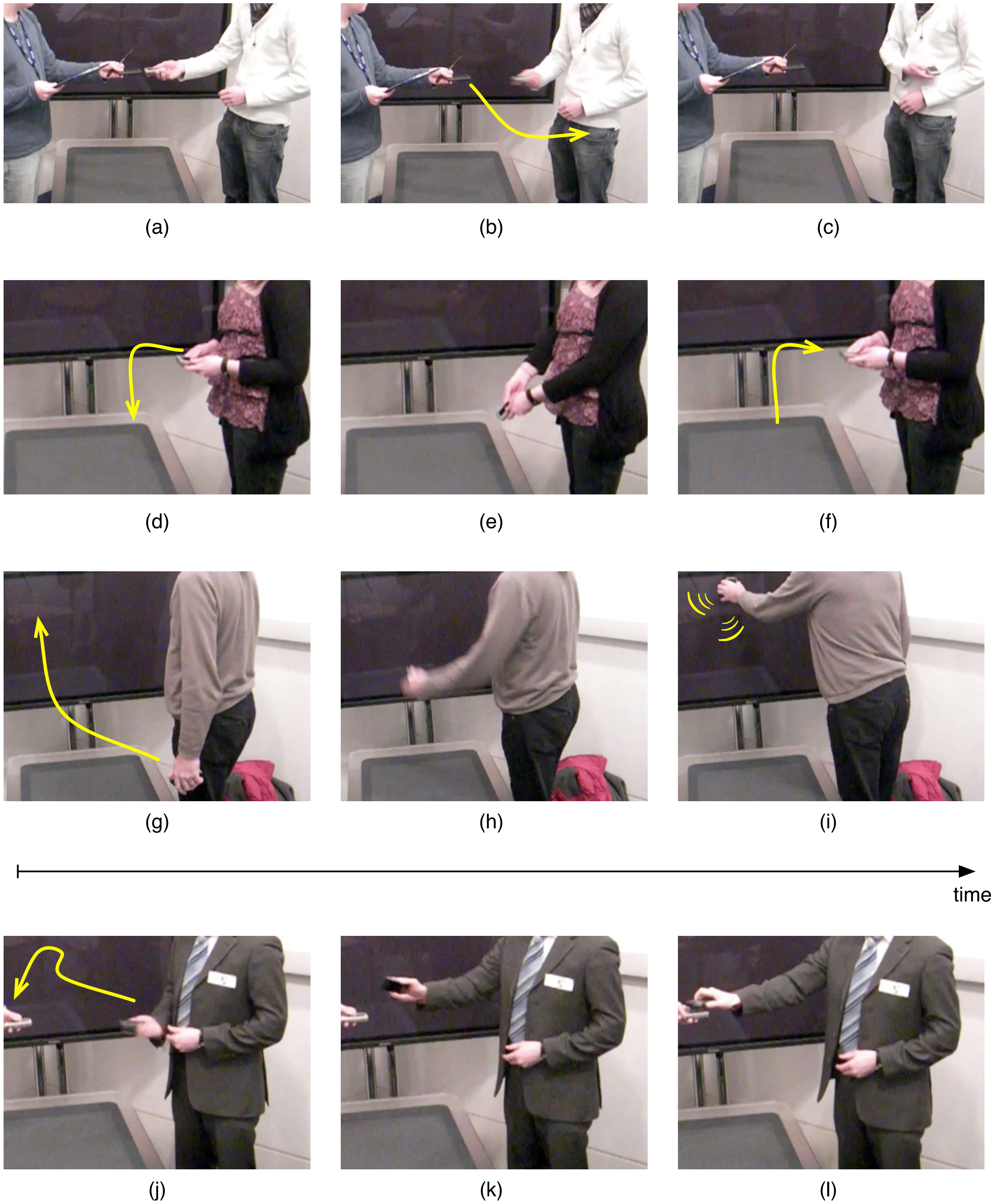

Figure 4: Four example gestures we observed during the study – images are taken from the video footage (time progresses from left to right). The top three rows depict gestures we observed frequently, whereas the bottom row shows an example for a less frequent gesture. Top row (a-c): 'pull away' gesture performed in the phone-to-phone condition to abort running activity; second row (d-f): 'pointing gesture' performed in the phone-to-tabletop condition to send an item to the tabletop; third row (g-i): 'touch gesture' performed in the phone-to-public display condition to trigger synchronization; bottom row (j-l): 'facing gesture' performed in the phone-to-phone condition to trigger authentication. (See text for more detail).

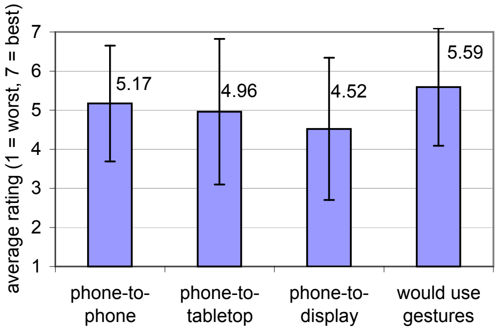

Figure 5: Subjective ratings to acceptability of phone gestures. Left three bars: suitability of phone gestures for each device combination. Rightmost bar: intent to use phone gestures overall. Error bars show standard deviation.

(on a seven-point Likert scale, where one corresponded to strong disagreement and seven to strong agreement). Figure 5 summarises their responses. When asked whether they would agree that phone gestures work well with a particular target device, they responded positively in all three cases. The phone-to-phone configuration received the highest average score (5.17, SD=1.48), closely followed by the phone-to-tabletop scenario (4.96, SD=1.86). The phone to public display configuration attracted the lowest overall rating (4.52, SD=1.82). When asked whether they would use phone gestures if they were available for their phone, the majority agreed (average rating of 5.59, SD=1.5).

In addition to the rating tasks, we also asked participants for qualitative feedback. More specifically, we solicited their feedback about the three most negative and positive aspects of using phone gestures. We also wanted to know which three activities subjects thought would work well, and which three they thought would not work well. Table 3 summarizes the most common responses.

The three most commonly mentioned negative aspects of using phone gestures were misinterpretation of gestures (mentioned eight times), feeling awkward performing phone gestures in public [17] (mentioned five times) and gestures being difficult to learn (also mentioned five times). Example comments that we classified as belonging to those categories include "Phone might mistake certain gestures for others.", "It could bring up the wrong interaction."; "Look a bit a prat", "You feel a bit silly performing gestures in public."; "Learning gestures takes time.", and "Learning and using a new way of interacting." There were a number of further comments that were not mentioned by many people, including negative impact on battery life, the danger of hurting bystanders while performing gestures, and concerns about costs incurred by adding gesture support to mobile phones.

On the positive side, the most frequently mentioned aspects included speed (mentioned seven times), ease of use (also mentioned seven times) and fun to use (mentioned five times). Example comments that we included in this category include "Quick", "Feels quicker than normal methods eg. Bluetooth"; "Easy", "Easier to perform simple tasks"; "Fun and 'modern"', and "Fun." Further, less frequently mentioned comments about positive aspects of using phone gestures related to aspects such as increased hygiene (due to not having to touch anything but one's own phone), not having to flick through menus on the phone and the benefit of physical activity resulting from performing gestures.

In terms of suitable activities, participants most commonly mentioned activity was sending information to the other device (mentioned eight times in a generic way; a further four people mentioned 'sending pictures'). Receiving information was the second most frequently mentioned activity (six times), while synchronizing devices was mentioned five times and rewinding/forwarding media was mentioned four times. Example comments that we classified as belonging to these categories include "Sending Info", "Sending data to other device"; "File download", "Downloading"; "Sync calendar from device to phone", "Sync your phone with someone else's"; "Fast forward/rewind" and "Fast forwarding media files". Playing games, watching television and remotely controlling another device were examples for activities, which subjects thought would be suitable for use with phone gestures but which were mentioned less often.

Amongst the activities listed by participants as being unsuitable for use with phone gestures only two were mentioned frequently: making payments (mentioned seven times) and authentication (mentioned six times). The other activities listed in Table 3 were each only mentioned three times. Example comments that we classified as belonging to the two most commonly mentioned categories include "Payment", "Any kind of payment"; "Authentication" and "Trying to make first time connection to other unknown devices - security issues." Voting, browsing the Internet and entering phone numbers were mentioned less frequently by participants as being activities less suitable for use with phone gestures.

5. DISCUSSION

The overall response of the participants with respect to using phone gestures to trigger activities involving another device was quite positive. This positive feedback does not only include the data reported in Fig. 5 and the positive aspects noted in Table 3 but also extended to informal feedback during and after the study. In addition, we found that subjects were able to produce gestures quickly and effortlessly, and that it did not take long for them to understand the concept and put it to practice. We feel that this type of interaction is thus worthwhile pursuing further.

We were also impressed with the creativity of the participants – they generated a large number of different gestures in a short amount of time. While some of these gestures did re-occur frequently both within subjects and between subjects, there were also some gestures, which were quite innovative such as writing one's signature in the air while holding the phone to authenticate. In our opinion, this indicates not only that there is much to explore in terms of possible gestures but also highlights the need to involve users in the design process (if only to not miss out on highly creative gestures.)

5.1 Implications for Gesture Recognition

As shown in Figure 2, the majority of gestures involves changing the relative distance between the mobile phone held by the user and the target device. Recognizing such gestures requires technologies that can estimate relative distances between devices. This can be realized by technologies measuring signal strengths or runtime differences between signals. Examples include sound and ultrasound sens-

Table 3: Most common qualitative feedback from participants (ordered according to the number of times they were mentioned) on negative/positive aspects of using phone gestures and well-suited/unsuitable activities

	negative aspects	positive aspects	well-suited activities	unsuitable activities
1	misinterpretation	speed	sending data	payments
2	feeling awkward in public	ease of use	receiving data	authentication
3	difficult to learn	fun to use	synchronizing devices	rewinding/forwarding
4	privacy	simplicity	rewinding/forwarding	placing a call
5	having to remember gestures	universal use	scrolling	entering text
6	damaging phone	looking good/cool		

ing (*BeepBeep* [11], *Point & Connect* [12], *RELATE* [4]), infrared light sensing (*SideSight* [2], HoverFlow [10]), and depth cams. Another type of gestures we observed relied on location changes, which for example can be detected with accelerometers [3, 9, 23]. Rotation changes were also quite common, which can be sensed by accelerometers, magnetometers, and gyroscopes. Moreover, absolute location in space was a component of some gestures. Except for pointing towards a target [18], this property is difficult to implement with high precision and low latency, but can be achieved, for example, by camera-based techniques using either natural features or special markers. The Wii controller, for example, uses an embedded infrared camera that detects external infrared LEDs as special markers. Surprisingly, direct touch, in which two devices physically touch during the gesture have been observed least often. Such gestures can relatively easily be detected with NFC/RFID tags [5, 13, 19] or via simultaneously occurring events [6, 15].

5.2 Implications for User Interfaces

Table 3 summarizes qualitative feedback from the participants, which can have some implications for the design of interfaces based on phone gestures. The most frequently listed negative aspect was the fear of gestures being misinterpreted. From a design perspective, this can be interpreted as the need to make sure recognition works reliably, or the need to provide means to users to easily abort such accidental interactions. The second most frequently named negative aspect, feeling awkward in public, somehow contrast with the positive aspects 'fun to use' and 'looking good/cool'. It may well be a matter of context, i. e. performing gestures in public is 'cool' in one location and awkward in another one.

Payments and authentication stood out as highly unsuitable activities for use with phone gestures. We attribute this to high visibility of gestures, which may be an undesirable side-effect when performing activities with security/privacy implications. Sending and receiving data were listed as being well-suited for phone gestures. It may well be that in this case, the visibility of gestures is a desirable feature as it can make it obvious to others what is going, in particular to the recipient or sender of the information.

5.3 Limitations and Next Steps

The current study only investigates interactions that span two devices, although their role and functionality can be quite diverse. It would be interesting to explore group interactions in which media items are exachanged from one sender to multiple receivers at once or in which input is collected from multiple sources. Moreover, we deliberately limited the study in not showing any content on the device displays. Consequently, an important future area of research would be to establish more clearly in how far properties of graphical content afford or constrain certain gestural interactions. Such a future study still needs to be conducted, however its design can be informed by the present investigation. We also think it would be beneficial to perform an "inverse" study, in which video footage from our current study is shown to other subjects and ask them to infer the intended action of a gesture. Moreover, in order to validate the approach of using gestures for multi-device interaction a comparison study with other technologies would be helpful.

6. CONCLUSION

In this paper, we presented a first investigation into eliciting gestures from users for combining mobile phones with other devices, i. e. other mobile phones, interactive tabletops, and large displays. The results are encouraging in that users generally liked gestures as a way to use their mobile phones to intuitively interact with other devices. Among the device combinations we tested, using phone gestures to connect to other phones received the highest rating, followed by the phone-to-tabletop, and phone-to-large-display scenarios. Our observations provide some initial insights that will inform the design of phone gestures in the future.

We were surprised by the degree of novelty and diversity of gestures that users invented and by their ability to spontaneously produce meaningful gestures. While there were a number of gestures that occurred frequently (such as pulling the phone back or flicking it in a particular direction) there were several gestures which were quite novel, e. g. placing a phone on another phone so that their screens were facing, or placing a phone on top of the frame of the public display. Cataloguing and categorizing these gestures in more detail will be a logical next step in this line of research.

Another contribution of the work described in this paper is an analysis of basic properties of phone gestures that can inform the design of future mobile phones in terms of which sensors to include in order to enable the recognition of common gestures. We found, for example, that a large number of gestures involve a change in the location of the phone and/or a change in the relative distance to the target device (which is difficult to measure given the sensors built into today's phones), while physical contact occurred less often.

A next step in this line of research will be to consolidate the observed gestures into a coherent gesture set based on the gestures that resulted from the study. This requires to look for gestures with a high degree of agreement between participants. We were already able to derive basic properties of user-defined phone gestures. These gestures and the user feedback will lay the foundation for future research in this area. Another logical extension to the work presented here is to conduct a series of studies, which asks subjects to associate activities to gestures they observe, i. e. the ones participants produced in the study reported here.

7. REFERENCES

[1] Y. Ayatsuka, N. Matsushita, and J. Rekimoto. Hyperpalette: a hybrid computing environment for small computing devices. In *CHI '00: CHI '00 extended abstracts on Human factors in computing systems*, pages 133–134, New York, NY, USA, 2000. ACM.

[2] A. Butler, S. Izadi, and S. Hodges. Sidesight: multi-"touch" interaction around small devices. In *UIST '08: Proceedings of the 21st annual ACM symposium on User interface software and technology*, pages 201–204, New York, NY, USA, 2008. ACM.

[3] R. Dachselt and R. Buchholz. Natural throw and tilt interaction between mobile phones and distant displays. In *CHI EA '09: Proceedings of the 27th international conference extended abstracts on Human factors in computing systems*, pages 3253–3258, New York, NY, USA, 2009. ACM.

[4] H. Gellersen, C. Fischer, D. Guinard, R. Gostner, G. Kortuem, C. Kray, E. Rukzio, and S. Streng. Supporting device discovery and spontaneous interaction with spatial references. *Personal Ubiquitous Comput.*, 13(4):255–264, 2009.

[5] R. Hardy and E. Rukzio. Touch & interact: touch-based interaction of mobile phones with displays. In *MobileHCI '08: Proceedings of the 10th international conference on Human computer interaction with mobile devices and services*, pages 245–254, New York, NY, USA, 2008. ACM.

[6] K. Hinckley. Synchronous gestures for multiple persons and computers. In *UIST '03: Proceedings of the 16th annual ACM symposium on User interface software and technology*, pages 149–158, New York, NY, USA, 2003. ACM.

[7] K. Hinckley, G. Ramos, F. Guimbretiere, P. Baudisch, and M. Smith. Stitching: pen gestures that span multiple displays. In *AVI '04: Proceedings of the working conference on Advanced visual interfaces*, pages 23–31, New York, NY, USA, 2004. ACM.

[8] M. Karam and M. C. Schraefel. A taxonomy of gesture in human computer interactions. Technical report, Technical Report ECSTR-IAM05-009, Electronics and Computer Science, University of Southampton, 2005.

[9] J. Kela, P. Korpipää, J. Mäntyjärvi, S. Kallio, G. Savino, L. Jozzo, and D. Marca. Accelerometer-based gesture control for a design environment. *Personal Ubiquitous Comput.*, 10(5):285–299, 2006.

[10] S. Kratz and M. Rohs. Hoverflow: expanding the design space of around-device interaction. In *Proc. of MobileHCI '09*, pages 1–8, 2009.

[11] C. Peng, G. Shen, Y. Zhang, Y. Li, and K. Tan. Beepbeep: a high accuracy acoustic ranging system using cots mobile devices. In *SenSys '07: Proceedings of the 5th international conference on Embedded networked sensor systems*, pages 1–14, New York, NY, USA, 2007. ACM.

[12] C. Peng, G. Shen, Y. Zhang, and S. Lu. Point & connect: intention-based device pairing for mobile phone users. In *MobiSys '09: Proceedings of the 7th international conference on Mobile systems, applications, and services*, pages 137–150, New York, NY, USA, 2009. ACM.

[13] T. Pering, Y. Anokwa, and R. Want. Gesture connect: facilitating tangible interaction with a flick of the wrist. In *TEI '07: Proceedings of the 1st international conference on Tangible and embedded interaction*, pages 259–262, New York, NY, USA, 2007. ACM.

[14] J. Rekimoto. Pick-and-drop: a direct manipulation technique for multiple computer environments. In *UIST '97: Proceedings of the 10th annual ACM symposium on User interface software and technology*, pages 31–39, New York, NY, USA, 1997. ACM.

[15] J. Rekimoto. Synctap: synchronous user operation for spontaneous network connection. *Personal Ubiquitous Comput.*, 8(2):126–134, 2004.

[16] J. Rekimoto and M. Saitoh. Augmented surfaces: a spatially continuous work space for hybrid computing environments. In *CHI '99: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 378–385, New York, NY, USA, 1999. ACM.

[17] J. Rico and S. Brewster. Usable gestures for mobile interfaces: evaluating social acceptability. In *Proc. of CHI '10*, pages 887–896, 2010.

[18] C. Swindells, K. M. Inkpen, J. C. Dill, and M. Tory. That one there! pointing to establish device identity. In *UIST '02: Proceedings of the 15th annual ACM symposium on User interface software and technology*, pages 151–160, New York, NY, USA, 2002. ACM.

[19] P. Tandler, T. Prante, C. Müller-Tomfelde, N. Streitz, and R. Steinmetz. Connectables: dynamic coupling of displays for the flexible creation of shared workspaces. In *UIST '01: Proceedings of the 14th annual ACM symposium on User interface software and technology*, pages 11–20, New York, NY, USA, 2001. ACM.

[20] A. D. Wilson and R. Sarin. Bluetable: connecting wireless mobile devices on interactive surfaces using vision-based handshaking. In *GI '07: Proceedings of Graphics Interface 2007*, pages 119–125, New York, NY, USA, 2007. ACM.

[21] J. O. Wobbrock, M. R. Morris, and A. D. Wilson. User-defined gestures for surface computing. In *CHI '09: Proceedings of the 27th international conference on Human factors in computing systems*, pages 1083–1092, New York, NY, USA, 2009. ACM.

[22] J.-b. Woo and Y.-k. Lim. Contact-and-connect: designing new pairing interface for short distance wireless devices. In *CHI EA '09: Proceedings of the 27th international conference extended abstracts on Human factors in computing systems*, pages 3655–3660, New York, NY, USA, 2009. ACM.

[23] K. Yatani, K. Tamura, K. Hiroki, M. Sugimoto, and H. Hashizume. Toss-it: intuitive information transfer techniques for mobile devices. In *CHI '05: CHI '05 extended abstracts on Human factors in computing systems*, pages 1881–1884, New York, NY, USA, 2005.