Natural Group Binding and Cross-Display Object Movement Methods for Wearable Devices

Tero Jokela
Nokia Technologies
Tampere, Finland
tero.jokela@nokia.com

Parisa Pour Rezaei
Tampere Univ. of Technology
Tampere, Finland
parisa.pourrezaei@tut.fi

Kaisa Väänänen
Tampere Univ. of Technology
Tampere, Finland
kaisa.vaananen@tut.fi

ABSTRACT

As wearable devices become more popular, situations where there are multiple persons present with such devices will become commonplace. In these situations, wearable devices could support collaborative tasks and experiences between co-located persons through multi-user applications. We present an elicitation study that gathers from end users interaction methods for wearable devices for two common tasks in co-located interaction: group binding and cross-display object movement. We report a total of 154 methods collected from 30 participants. We categorize the methods based on the metaphor and modality of interaction, and discuss the strengths and weaknesses of each category based on qualitative and quantitative feedback given by the participants.

Author Keywords

Co-located interaction; multi-device user interfaces; wearable devices; smartwatches; smartglasses; device ecosystem binding; group association; pairing; cross-display object movement; guessability study; elicitation study.

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous

INTRODUCTION

We are seeing an increasing diversity of digital devices with the emergence of new wearable form factors, such as smartwatches and smartglasses. Currently, wearable devices are still relatively rare and the early applications focus on individual use. However, as more and more people acquire and start to use such devices, situations where there are multiple persons present with wearable devices will become commonplace. In such situations, wearable devices provide interesting opportunities to support collaborative tasks and experiences between co-located persons through multi-user applications [12]. For example, two families on a holiday trip together could collect all photos and videos that they capture with their smartglasses into a shared album; a group of friends biking together could use their smartwatches to track the locations and performances of each other; or a team of rescue workers arriving at an accident scene could use their wearable devices to communicate and to share situational information with each other.

But before a group of co-located persons can engage in spontaneous collaborative interactions with their devices, they must first connect their devices together into a group to make the devices aware of each other and to enable data communications between them. This problem is called ad-hoc group binding or group association [6] (also known as pairing or coupling). Another common challenge in co-located interaction is moving virtual objects (such as content items or application windows) between device displays. This problem is generally known as cross-display object movement [21]. Both tasks involve complex technical procedures, but still from a user's perspective, it should be possible to perform these tasks in a fast and easy way – if the process is too time-consuming or tedious, the users might lose their interest in multi-user applications in the first place. Both device binding and cross-display object movement have been extensively studied in prior research and a wide range of potential solutions have been proposed to both problems. However, most of the existing solutions have been designed for individual users rather than for groups and have been driven by technology and security considerations rather than by user experience. Most of the existing methods have also been designed for conventional devices and may not be suitable for or take full advantage of the features of wearable devices [12].

In this paper, we present an elicitation study which aims to collect methods for group binding and cross-display object movement tasks on wearable devices from groups of ordinary end-users. The question we address is what would a group of co-located users naturally do to connect their wearable devices or to move objects between them. We cover both smartglasses and smartwatches, currently the two most common types of wearable devices. We report a total of 154 methods collected from 30 participants, categorize the methods based on the metaphor and modality of interaction, and discuss the strengths and weaknesses of each category based on qualitative and quantitative feedback given by the participants. The results inform the design of future wearable device user interfaces and applications that involve group binding and cross-display object movement tasks.

RELATED WORK

Ad-Hoc Device Binding Methods

The problem of ad-hoc device binding has been extensively studied in human-computer interaction, ubicomp, and security research, and a wide range of different methods for device binding has been proposed [6, 29]. The most common device-binding method today, which is widely used in Bluetooth and WLAN networks, is called Scan & Select. It is based on scanning the environment for available devices and presenting a list of the found devices to the user who can then select the other devices they wish to bind with. The connections are authenticated with short strings that the user should copy or compare between devices.

Examples of alternative device binding methods proposed in prior literature include a range of methods based on synchronous user actions, for example, pressing buttons on both devices simultaneously [25] or shaking the devices together [10]. Methods based on spatial alignment of devices include pointing [19], touching [26], or placing the devices in close proximity of each other [17]. Devices can also be bound by using auxiliary devices, for example, by attaching tokens to the devices [1].

Binding methods are not only means for connecting devices – they have strong social and emotional aspects [12]. Many factors influence persons' preferences of binding methods, including the place, the social setting, the other people present, and the sensitivity of the data [11]. Device association in large groups is a highly collaborative group activity [13, 14, 16]. Therefore, the group binding methods should pay special attention to supporting groupwork and social interactions. The group creation task can be divided in different ways between the members of the group, including leader-driven and peer-based approaches [13, 16]. Group association can be seen as a one-step procedure of binding all devices with a single action, or as a sequence of pairwise associations [5, 29]. The group binding methods should also be flexible and robust, allowing people to adapt them to the changing needs of the situation and to recover from failures [14].

Cross-Display Object Movement Methods

Conventional solutions for moving content items between devices include connecting the devices directly using physical cables or wireless short-range radio technologies, such as Bluetooth or WLAN. Objects can also be moved through portable storage media, such as memory cards, USB memory sticks, and external hard drives. It is also a common practice to move objects between devices by sending them as e-mails or other kinds of messages. Further, objects can be saved to network folders or storage services and re-opened with other devices. Many cloud services support transferring specific content types (for example, browser tabs, photographs, video, or music) between devices.

In the field of human-computer interaction research, the problem of cross-display object movement has been studied in the context of multi-monitor computer setups, large composite displays, collaborative interaction spaces, tabletop displays, and ad-hoc ecosystems of mobile devices [21]. A wide range of methods for moving visual objects between device displays has been proposed in the prior literature. For example, in Pick-and-Drop [24] the user can pick up an object by touching it with a digital pen and then drop it by repeating the touch action on another display. Objects can also be moved from one display to another by making a throw touch screen gesture towards the target. Many variations of the basic throw gesture, for example, Drag-and-Throw and Push-and-Throw [9] have been suggested, aiming to provide better control over and to improve the accuracy of the throw. In ARIS [3], the user can manage objects across screens on a radar-like map view showing all displays in proximity. Further, objects can be manipulated across displays by using pointing gestures [23]. In MediaBlocks [30], objects can be bound to physical tokens that can be moved between displays.

Nacenta et al. [21] present a taxonomy and classification of cross-display object movement methods based on three conceptual levels: on referential domain level, the methods can reference displays either spatially or non-spatially (for example, using textual names); on display configuration level, the mapping between spatial arrangement of the displays and the method's input model can be planar, perspective, or literal; and on control paradigm level, methods can be open-loop, closed loop, or intermittent. A few studies have compared different cross-display object movement methods against each other. In an early study comparing six different methods, Nacenta et al. [20] identified significant differences between methods, with Radar View and Pick-and-Drop as the fastest and most preferred methods. More recent studies by Jokela et al. [15] and Scott et al. [28] have found that many factors can influence the users' preferences of methods, including the numbers of objects and devices involved, device characteristics, physical and social environment, and personal working styles.

Elicitation Studies

To generate a set of user-defined input actions, Nielsen et al. [22] and Wobbrock et al. [32] suggest similar elicitation approaches: the participants are first presented with the end effect of an operation and are then asked to perform the action that caused it. Elicitation studies have been most commonly used to produce sets of touch screen gestures [for example, 31] but they have been adapted to other kinds of input actions as well, including hand gestures [27], kick gestures [2], and bend gestures [18]. Closest to our study, Chong and Gellersen report two elicitation studies that collected device binding methods from individual users [4] and groups of users [5]. While their studies covered a broad range of different device combinations, they did not address group binding methods for wearable devices.

OUR STUDY

Objectives

In our study, we were interested in collaborative multi-user applications for groups of co-located users with wearable devices. In particular, we were interested in interaction methods for two common tasks in co-located interaction: group binding and cross-display object movement. Regarding devices, we decided to focus on the two most common wearable device categories today: smartglasses and smartwatches.

Both device binding and cross-display object movement have been extensively studied in prior research and a wide range of potential solutions have been proposed to both problems. However, most of these solutions have been designed for a single user interacting with two devices. Scenarios involving multiple users and devices differ in many respects from single-user scenarios, making the single-user solutions not necessarily applicable to multi-user scenarios. Most of the prior research has also been driven by technology and security considerations. The existing solutions have often been invented by researchers and system designers, focusing on a particular technology solution, while the intuitiveness of the solution to non-technical users has been given little consideration. Finally, most of the existing methods have been designed for conventional devices such as computers, phones, and tablets, and are not necessarily applicable to wearable devices, which are far more personal and intimate. They also may not take advantage of the unique features of wearable devices that could enable more natural and innovative solutions [12].

To address these concerns, we wanted to approach the problem from a perspective of what pairs and small groups of people would naturally and intuitively do to connect their wearable devices into a group or to move virtual objects between their wearable devices. Therefore, we adopted an elicitation study approach and asked groups of participants to come up with their own techniques for these tasks. The participants took turns to suggest different methods and then immediately tested and evaluated the methods with mock-up devices. The participants could suggest any technique with only one restriction: as earlier studies suggest that people are unwilling to hand in their personal devices to strangers [8], the participants were not allowed give their devices to the other participants. The primary objective of our study was to collect and preliminarily evaluate a broad set of suggestions for natural group-binding and cross-display object movement methods for wearable devices, in order to inform the design of future multi-user applications for groups of co-located users with wearables.

Figure 1. Device surrogates used in the study: smartglasses (left) and smartwatch (right).

Our study was inspired by similar studies on device binding methods by Chong and Gellersen [4, 5]. Our study extends their work in three major directions: we address (1) groups of users binding their watches together (Chong and Gellersen only considered a single user binding a digital watch with a phone, a tablet, or a display); (2) binding of smartglasses (not considered by Chong and Gellersen); and (3) cross-display object movement (not considered by Chong and Gellersen).

Devices

We decided not to use any commercially available wearable devices in the study, as the devices' native user interface styles and technical features and limitations might have guided the participants' proposals and limited their creativity. Instead, we provided the participants with simple mock-up devices that acted as surrogates of real devices. Figure 1 illustrates the device surrogates used in the study. We simulated smartglasses with ordinary 3M safety spectacles. For simulating smartwatches, we built custom mock-ups by attaching a Casio watch band to a small block of polystyrene foam. The dimensions of the block were 48x38x12 mm (1.9x1.5x0.5 inches) which is comparable to the currently available commercial smartwatches.

Participants

We recruited eight groups of four participants for the study by posting an advertisement on local mailing lists and social media. However, two persons canceled their participation at the last moment, which resulted in two of the groups having only three participants and reduced the total number of participants to 30. We encouraged the participants to recruit also their friends, so that some participants would know each other to make the study situation more natural and comfortable. 15 of the participants knew one of the other participants in the same group and eight participants knew several of the other participants, while seven knew none of the other participants. 14 of the participants were female and 16 male. The ages of the participants varied between 15 and 56 years (M=32.7, SD=12.6). Two of the participants were left-handed and 28 right-handed. 14 participants had educational backgrounds related to information technology, while the other 16 participants represented a wide variety of different professions (for example, a school teacher, a nurse, and a bus driver). On a scale between 1 and 7 (1=novice, 7=expert), the participants rated their familiarity with information technology slightly above average (M=4.7, SD=1.9). Seven participants had tried smartglasses (for example, Google Glass) and 11 participants had tried a smartwatch or other smart wrist device. 11 participants were wearing eyeglasses continuously and six as needed, while 13 participants did not have eyeglasses. Eight participants were wearing a wristwatch every day and seven sometimes, while 15 participants did not have a wristwatch.

Procedure

We organized a total of eight study sessions. The sessions were arranged in our usability laboratory. In each session, there were three or four participants and a moderator present. Figure 2 shows the evaluation setup. The durations of the sessions varied between 75 and 120 minutes.

As the participants arrived in the laboratory, the moderator asked them to fill in the background information and consent forms. The moderator then introduced the participants to wearable devices and the idea of several co-located persons using their wearables together for collaborative applications. The moderator explained that the objective of the study was to invent and evaluate interaction methods for two common tasks in multi-user applications: group binding and cross-display object movement. The moderator then continued by describing the detailed study procedure.

To begin the actual study, the moderator gave each participant the first mock-up device. If the participant was wearing eyeglasses, they could wear the protective glasses used to simulate smartglasses over the eyeglasses if needed. However, if the participant was wearing a wristwatch, they were asked to remove the wristwatch when wearing the smartwatch mock-up. The moderator then described the device in more detail and gave several examples of different visual, audio, and haptic interaction methods that the device could support. However, the moderator emphasized that these were just examples and that the devices could have any capabilities the participants wanted them to have. After that, the moderator gave the participants the first task and explained it in detail. The moderator asked one of the participants to suggest a method how the task could be achieved with the devices. The moderator encouraged the participants to suggest the first intuitive ideas that spontaneously occurred in their minds and reminded that they could do anything but to give their devices to another participant.

When the first participant had described a method, the moderator asked all participants to stand up and try it out with the device mockups, first in pairs and then as a single group of three or four persons. The moderator portrayed an external person not to be included in the group or to receive the object. Trying the method with the mockup devices

Figure 2. Participants suggesting and evaluating methods during the study.

clarified the details of how it would work and gave the participants a better understanding of its strengths and weaknesses in practice. After the participants had tried the method, the moderator asked them to fill in a paper form to evaluate the method in terms of practicality (that is, how easy, effortless, efficient, and error-free it was to use) and pleasantness (that is, how human, connective, inspiring, and inventive it was). After filling in the forms, the moderator asked the participants to provide brief immediate free-form verbal comments about the method. The moderator then asked the next person to suggest another method, which was similarly tested and evaluated. When every participant had suggested a method, the moderator offered an opportunity to any participant to suggest further ideas. Overall, the groups tested three to seven methods for each device-task combination.

The same procedure was then repeated for the other task with the same device, and after that for both tasks with the other device. The order of the devices and methods was systematically varied between study sessions to counter-balance any learning effects. At the end of the study session, the moderator briefly interviewed the participants for general comments about the wearable devices and tasks used in the study. To close the session, the moderator thanked the participants and gave each participant a small reward to compensate for their time.

The study sessions were recorded with a video camera. All the proposed methods were documented and categorized by three researchers. All participant comments about the methods as well as interview responses were transcribed and thematically analyzed. A quantitative analysis of the evaluation responses was done separately.

RESULTS

General

The participants proposed a total of 154 different methods during the study. Of these methods, 73 were intended to be used with smartglasses, including 38 group binding methods and 35 cross-display object movement methods. The remaining 81 methods were intended to be used with smartwatches and consisted of 40 group binding methods and 41 cross-display object movement methods.

Both group binding and cross-display object movement tasks could be divided into three main phases: preparation, target identification, and confirmation. The *preparation phase* consisted of various activities that were needed to initiate the task. It could include discussing with the other participants and agreeing on a common target and strategy to achieve it. It could also include initiating the necessary technical functions on the devices, such as activating a certain device mode, or creating a new group for others to join, or browsing and selecting the object to be shared with the others. In the *target identification phase*, links between the devices were created, that is, it was indicated which devices were intended to participate in the group or to receive the object. The *confirmation phase* consisted of making sure that the intended action was successfully completed. The person who created the group or shared the object wanted to make sure that all the intended persons had been included in the group or received the object, and that there were no external persons included. On the other hand, the persons who joined in the group or received the object wanted to make sure that they were only in those groups that they wanted to be part of, or received only those objects that they wanted to have. It is important to note that all phases consisted of both technical actions involving interactions with the devices and social actions involving interactions with the other participants. Many actions had both technical and social dimensions – for example, if one person was interacting with their device by tapping the screen with their finger, this could be observed by the other persons and was intertwined with the social interaction within the group.

After an initial analysis of the collected methods, we decided to categorize the methods based on the techniques used in the target identification phase, as that appeared to be the most characteristic phase of the interaction where the actual connections between the participants were formed. The same categorization was used for both group binding and cross-display object movement methods. The categorization was primarily based on the *metaphor of interaction* and secondarily on the *modality of interaction*. Three researchers first independently analyzed and categorized each method. The individual categorizations were then compared and the methods where the researchers disagreed were commonly discussed and resolved. Table 1 presents the resulting categorization. We will next explain and discuss each main category in detail.

Spatial

The methods in the Spatial category were based on the relative or absolute positions of the participants. Spatial methods were the most popular category for smartglasses, but they were also commonly suggested for smartwatches. Spatial methods were equally proposed for both group binding and object movement tasks. The methods in this category could be divided into three subcategories: pointing-based, proximity-based, and area-based methods.

The participants suggested several different approaches for selecting the target person by pointing. For smartglasses, by far the most popular approach was to point at the target with gaze. *"[P25] Eye contact. I look directly at my target person and then the [object] will be transferred to her."* In some variants, it was enough just to look in the direction of the target person, while in other variants it was required to establish an eye contact between the persons, providing the target with a possibility to reject the interaction. While some participants considered looking at another person a natural element of human communication, others felt that it was disturbing to have to stare at the eyes of another, possibly unfamiliar, person. Gaze pointing, like all pointing-based methods, was also thought to be slow for larger groups of people as the person had to point at every target person in sequence, one at a time. As a solution, some participants suggested that all the persons in the field of view could be selected with a single operation. The person could then select their viewpoint so that only the target persons they wished to select were visible, or alternatively they could hide the unwanted persons, for example, with their hands. For smartwatches, the most popular approach was to point at the target with the device, aiming and shooting like with a camera or a laser ray. These methods were considered as fun and playful. *"[P21] There is the Star Trek fun factor. Pew, pew, pew… You could add a lazer sound."* A third possible approach, which was suggested for both smartglasses and smartwatches, was to point at the target with a finger or with a hand. While generally considered easy and natural, some participants were concerned that pointing with a finger could appear as rude or ridiculous in some social situations.

In proximity-based methods, the devices were capable of determining the closest person or group of persons. The participants could then make selections by moving close to each other. While simple, the methods were considered error-prone in situations where there were a lot of people present. Some participants also considered moving very close to unfamiliar persons annoying. Finally, in area-based methods the participants defined an area to select all persons within that area. The area could be defined either by setting the radius of a circle around the user, or by explicitly drawing the area boundaries. Defining an area was seen as an efficient though imprecise way to select a large group of people. Selecting a small group people in a crowded area could force the participants to define a very small area and to move very close to each other to fit in that area.

Touching

The most commonly suggested method for smartwatches was device touch, that is, bringing the watches into physical contact or very close proximity of each other for a brief period of time. *"[P32] You select [the object to share] and*

	Smartglasses												Smartwatch											
	All				Group binding				Object movement				All				Group binding				Object movement			
	S	N	PRA	PLE	S	N	PRA	PLE	S	N	PRA	PLE	S	N	PRA	PLE	S	N	PRA	PLE	S	N	PRA	PLE
ALL	8	73	5,03	4,99	8	38	5,03	4,99	8	35	5,03	4,99	8	81	5,34	5,32	8	40	5,38	5,32	8	41	5,31	5,32
SPATIAL	7	30	4,92	4,99	7	14	5,02	5,07	7	16	4,84	4,91	7	13	4,89	5,02	6	7	5,17	5,13	4	6	4,53	4,88
Pointing	7	25	5,04	4,99	5	11	5,19	5,05	7	14	4,92	4,94	5	8	4,91	4,98	2	3	5,62	5,5	3	5	4,41	4,62
With gaze	7	21	5	4,97	5	9	5,21	5,09	7	12	4,84	4,88	1	1	4,12	4,25	0	0	N/A	N/A	1	1	4,12	4,25
With hand	4	4	5,23	5,1	2	2	5,06	4,88	2	2	5,43	5,36	1	2	4	3,67	0	0	N/A	N/A	1	2	4	3,67
With device	0	0	N/A	N/A	0	0	N/A	N/A	0	0	N/A	N/A	4	5	5,37	5,55	2	3	5,62	5,5	2	2	4,93	5,64
Proximity	3	3	3,64	4,27	2	2	3,57	4,29	1	1	3,75	4,25	1	2	5	5,67	1	1	4,83	5	1	1	5,17	6,33
Define area	2	2	5,31	5,94	1	1	5,75	6,75	1	1	4,88	5,12	3	3	4,79	4,79	3	3	4,79	4,79	0	0	N/A	N/A
TOUCHING	5	9	5,01	4,82	5	7	5,17	4,85	2	2	4,5	4,75	8	21	5,53	5,63	8	11	5,4	5,58	6	10	5,67	5,68
Device-device	4	6	5,27	4,95	4	5	5,31	4,94	1	1	5,12	5	8	19	5,64	5,72	8	10	5,54	5,71	6	9	5,75	5,74
Person-person	3	3	4,54	4,58	2	2	4,88	4,62	1	1	3,88	4,5	1	1	5	5,25	0	0	N/A	N/A	1	1	5	5,25
Person-device	0	0	N/A	N/A	0	0	N/A	N/A	0	0	N/A	N/A	1	1	4,12	4,38	1	1	4,12	4,38	0	0	N/A	N/A
COMMAND	6	15	4,88	4,62	4	6	5	4,89	5	9	4,79	4,46	6	16	5,06	4,75	4	5	4,95	4,61	6	11	5,11	4,82
GUI	3	7	4,25	3,77	3	3	4,27	3,95	3	4	4,23	3,63	4	7	5,26	4,91	3	3	5,14	4,77	3	4	5,34	5
Voice	3	4	4,96	5,25	2	2	5,5	5,64	2	2	4,43	4,86	6	7	4,96	4,87	1	1	4,38	4,75	6	6	5,07	4,89
Thought	3	4	5,81	5,47	1	1	6,12	6,12	3	3	5,71	5,25	1	2	4,69	3,88	1	1	5	4	1	1	4,38	3,75
SCAN & SELECT	3	3	4,41	3,86	3	3	4,41	3,86	0	0	N/A	N/A	8	13	5,62	5,48	7	7	5,44	5,19	4	6	5,81	5,79
List	3	3	4,41	3,86	3	3	4,41	3,86	0	0	N/A	N/A	6	10	5,97	5,68	5	5	5,81	5,42	4	5	6,12	5,92
Spatial map	0	0	N/A	N/A	0	0	N/A	N/A	0	0	N/A	N/A	3	3	4,5	4,83	2	2	4,62	4,69	1	1	4,25	5,12
SHARED ACTION	3	5	5,47	5,81	3	5	5,47	5,81	0	0	N/A	N/A	5	10	5,39	5,39	5	9	5,61	5,6	1	1	3,62	3,75
Hand gesture	1	1	4,5	5,38	1	1	4,5	5,38	0	0	N/A	N/A	4	6	5,23	5,55	3	5	5,58	5,94	1	1	3,62	3,75
Touch gesture	0	0	N/A	N/A	0	0	N/A	N/A	0	0	N/A	N/A	1	1	6,33	6,5	1	1	6,33	6,5	0	0	N/A	N/A
Eye gesture	1	1	6	6,67	1	1	6	6,67	0	0	N/A	N/A	0	0	N/A	N/A	0	0	N/A	N/A	0	0	N/A	N/A
Button press	2	2	5,57	6,14	2	2	5,57	6,14	0	0	N/A	N/A	1	1	5,17	3,83	1	1	5,17	3,83	0	0	N/A	N/A
Passkey	1	1	5,88	5	1	1	5,88	5	0	0	N/A	N/A	2	2	5,57	5,07	2	2	5,57	5,07	0	0	N/A	N/A
VIRTUAL OBJECT	4	7	5,67	5,65	1	1	4,88	5	4	6	5,82	5,77	3	6	5,73	6,02	0	0	N/A	N/A	3	6	5,73	6,02
Throwing	4	5	5,34	5,34	1	1	4,88	5	4	4	5,47	5,43	3	5	5,69	5,94	0	0	N/A	N/A	3	5	5,69	5,94
Touch screen	1	1	6,5	7	0	0	N/A	N/A	1	1	6,5	7	3	3	5,73	5,91	0	0	N/A	N/A	3	3	5,73	5,91
AR	3	4	5,12	5,03	1	1	4,88	5	3	3	5,21	5,04	0	0	N/A	N/A	0	0	N/A	N/A	0	0	N/A	N/A
Imaginary	0	0	N/A	N/A	0	0	N/A	N/A	0	0	N/A	N/A	2	2	5,64	6	0	0	N/A	N/A	2	2	5,64	6
Giving	2	2	6,57	6,5	0	0	N/A	N/A	2	2	6,57	6,5	1	1	5,88	6,38	0	0	N/A	N/A	1	1	5,88	6,38
Imaginary	1	1	6,83	6,83	0	0	N/A	N/A	1	1	6,83	6,83	1	1	5,88	6,38	0	0	N/A	N/A	1	1	5,88	6,38
AR	1	1	6,38	6,25	0	0	N/A	N/A	1	1	6,38	6,25	0	0	N/A	N/A	0	0	N/A	N/A	0	0	N/A	N/A
NATURAL BEHAVIOR	3	3	5,79	5,67	1	1	5,5	5,12	2	2	5,94	5,94	2	2	5,38	5,12	1	1	6,25	5,88	1	1	4,5	4,38
Natural discussion	3	3	5,79	5,67	1	1	5,5	5,12	2	2	5,94	5,94	1	1	4,5	4,38	0	0	N/A	N/A	1	1	4,5	4,38
Similar behavior	0	0	N/A	N/A	0	0	N/A	N/A	0	0	N/A	N/A	1	1	6,25	5,88	1	1	6,25	5,88	0	0	N/A	N/A
REAL OBJECT	1	1	4	4,62	1	1	4	4,62	0	0	N/A	N/A	0	0	N/A	N/A	0	0	N/A	N/A	0	0	N/A	N/A
Clothes	1	1	4	4,62	1	1	4	4,62	0	0	N/A	N/A	0	0	N/A	N/A	0	0	N/A	N/A	0	0	N/A	N/A

Table 1. Categorization, numbers, and evaluation scores of the suggested methods (S = number of sessions where the method was proposed, N = number of times the method was proposed, PRA = practicality, PLE = pleasantness).

then you just bump your watch with the other person's watch." Device touch was equally suggested for group binding and object movement tasks. The participants considered device touch easy, fun, straightforward, and accurate to use. Some participants commented that bumping devices strengthened the team spirit and compared it to clinking glasses when making a toast. As an interaction method, device touch was familiar to many participants as they had experience of using NFC. However, a few participants were concerned that touching might physically damage their devices. In some variants of device touch, the target persons had to be selected one at a time, making the method very inefficient with large amounts of people. In most of the suggested variants, however, it was possible that several devices could touch each other simultaneously which improved the efficiency in large groups. Still, the method required the participants to be within touching distance of each other. In a few of the suggested methods, the target persons who had already been selected could also select further persons, making the group membership or the shared object to spread like a chain reaction or a virus within the group. Several participants also suggested device touch for smartglasses, especially for group binding. These methods required the participant to take off their smartglasses to touch the other person's glasses, which was considered impractical, especially if the person had poor eyesight and the smartglasses were assumed to have prescription lenses.

Another form of touching that was suggested by the participants was two persons touching each other. This typically took the form of a handshake. *"[P9] When you shake hands, both persons are added to the group."* While shaking hands was considered a familiar and natural gesture, it was primarily used in formal situations and felt comical in more relaxed situations such as when meeting with friends. People usually shook hands once when they met – shaking hands many times, first to greet and then to form groups or move objects felt inconvenient. A few participants were also worried that they could accidentally share files when shaking hands, for example, in a job interview. With smartglasses, the persons could shake hands naturally with their right hands. However, most participants wore the smartwatch on their left hands, requiring them to shake hands with the “wrong hand”. Sometimes the participants wore their watches in different hands making the handshake very cumbersome. Finally, one participant also suggested a method where the person creating a group could add other devices to the group by touching the devices with their hand, that is, by pressing the device screen with their thumb. The touching person was identified by the fingerprint which raised concerns about privacy and security.

Command

Yet another category of methods that was common for both smartglasses and smartwatches was methods that were based on giving direct commands to the device. Such methods were particularly popular for object movement, but were suggested also for group binding. In Command methods, the target persons were referred to using pre-defined identifiers such as names or contact cards. The most common subcategory of Command methods was methods based on traditional GUI interactions. *“[P7] You have a list of contacts on your watch. You can tap the persons you want to add to the group.”* Using pre-defined contacts enabled the participants to prepare groups in advance and include persons who were not currently present, but on the other hand it was limited to persons that the participants knew beforehand. On a smartwatch, the GUI was shown on the watch screen and the user could interact with standard touch screen techniques. This was seen as a familiar, reliable, and discreet way to interact but the screen was considered very small for handling large amounts of contacts. On smartglasses, the participant could see the GUI on a virtual screen floating in the air and interact with gaze, hand gestures, or some kind of a pointing device, for example, a touchpad, integrated into the frames of the smartglasses. The participants were worried that using a virtual display might be dangerous, for example, when walking, as it partly blocked their vision, and had doubts about how well the novel interaction methods would work in practice.

Giving commands by voice was also commonly suggested. *“[P23] I say: ‘Create group: [first person’s name], [second person’s name], [third person’s name].’ And then it is done. I use the names I have stored on my watch.”* Voice commands were more often suggested for smartwatches, and some participants commented that using voice was more natural with a smartwatch, as there was a physical device to talk to that they and the other people in proximity could see. Many participants had experience of using voice commands with conventional devices, and they raised several common problems of voice commands, for example, that they could be annoying to use in some social situations or sensitive to background noise. The voice commands also required the participant to remember the names of the other persons, and listing many names could be slow, which would be problematic especially with larger groups of unfamiliar people.

Finally, some participants envisioned methods where they could give commands to their device by thought. Using thought commands was slightly more often suggested for smartglasses than smartwatches. While considered futuristic, the participants usually were initially excited about the possibility and perceived thought commands as a very easy, effortless, and discreet way to interact. However, when the participants tested the method in practice, the experience was described as strange and unsocial. *“[P2] I had a feeling that I was missing [the] speech. I think sharing is interaction between people. First, it felt like WOW, this was easy and nice, but when we tried it, it suddenly felt that we are all here silent and everybody is doing something alone in his own bubble with his own brain UI. You don’t even have to look [at the others].”* As there were no perceivable indications, the other participants found it very difficult to know whether a person had completed a task, or even started it, and had to query the status by asking verbal questions.

Scan & Select

Scan & Select is currently the most commonly used device binding method and it is widely implemented in commercially available phones, tablets, computers, and other devices. As such, it was familiar to most participants and many participants suggested to use it for the study tasks, particularly with smartwatches, which were perceived to be closer to conventional devices. *“[P23] My watch shows all the devices in proximity, like in Bluetooth. Then I select and mark the devices I want to share with from the list and press ‘Send’.”* Scan & Select methods were equally suggested for both group binding and object movement tasks. In most variants of the method, a participant scanned the environment for available devices and then selected the devices to join in the group or to receive the object. In other variants, a participant first created a group or selected an object and made it accessible to other devices in the proximity. The other participants could then scan the environment with their devices for available groups or objects and select the one they wanted to join or receive.

In general, Scan & Select was considered a practical but not a very exciting solution. Some participants were concerned that in a crowded location, such as in a sports stadium, with a lot of persons and devices, scanning might take a long time and result in a large number of possible targets, which could be difficult to manage on a small smartwatch display. In the vast majority of suggested methods, the discovered targets were presented as a simple list of textual names, posing an additional problem of mapping the names to real-world persons and devices [21]. A few participants suggested spatial representations, for example, showing the targets on a map, which could enable easier mappings between targets on the screen and in the real world.

Shared Action

For the group binding task, the participants also suggested a range of methods where the persons indicated that they wanted to join in the group by doing simultaneously a common action. *"[P17] I make some movement, for example, shake my hand like this. If you do the same, it detects a similar movement in proximity and connects us into the same group."* While a similar approach could also have been applied to object movement, only one such method was suggested, indicating that a shared action was primarily considered appropriate only for the group binding task. The shared action could be a gesture, for example, a hand gesture, a touch screen gesture (that is, drawing a picture on the screen), or an eye gesture. It could also be a button press or entering a textual passkey. The gestures could be exactly similar for every user, or they could form symmetric pairs, for example, with the first user pushing their hand forward and the second user pulling their hand backward. Typically, all participants could do the action in parallel, making the methods efficient in large groups.

In general, the methods based on shared actions were considered as practical and simple to use and remember as there was only a single action that was done by every person. The participants appeared to enjoy doing a physical action together and sometimes spontaneously accompanied the action by shouting a common phrase together, like in team gestures in sports. The actions could also be clearly observed, making it easy to follow the status within the group. On the other hand, some participants were concerned that the actions might appear as ridiculous to external observers. Several participants were worried that if the actions were common everyday actions, such as nodding or rotating your head, external persons could accidentally join in the group, especially in places with large numbers of people. A malicious external person could also observe the action and do it on their device, in order to join the group without permission. Pressing a button or entering a passkey was also considered somewhat old-fashioned and boring. Text entry in general was considered difficult with wearable devices.

Virtual Object

In the Virtual Object methods, the object to be shared was represented as a virtual object that could be manipulated like a real object. *"[P13] I can see the file [virtually floating in the air]. I can grab the file and throw it towards you."* With the exception of a single group binding method, the Virtual Object methods were suggested only for the cross-display object movement task. The virtual object could be a purely imaginary object, or it could be visualized as a graphical object on the smartwatch display or as an augmented reality object on the smartglasses. The participants suggested two different interactions for delivering the object: throwing and giving. In the throwing methods, the sender grabbed the object and threw it towards the recipient. In the giving methods, the sender took the object in their hand and offered it on their palm. The recipient could then grab the object and collect it from the sender's palm.

In general, the Virtual Object methods were well liked by the participants and they received high ratings in the questionnaire. The methods were considered practical, natural, fun, and intuitive – even magical. However, like with the Shared Action methods, some participants were worried that they could accidentally share their files by making unintentional gestures. Several participants were also concerned that they might miss the target when throwing and send the file to a wrong destination. The giving methods were considered more secure to use. The throwing methods were also considered inefficient with a large number of people as they required the sender to throw the object to each recipient separately. The giving methods were more efficient as several recipients could take the object from the sender's palm simultaneously.

Natural Behavior

In addition to methods where the participants did explicit actions to form groups or move objects, a few participants also suggested methods where the devices monitored the participants' natural behavior and pro-actively executed commands on their behalf. These methods could be further divided into two subcategories. In the first subcategory, the devices were monitoring the natural discussion within the group, identified the persons based on their voices, and automatically executed commands based on the discussion. *"[P29] If I wanted to share a photo with a lot of guys, I would just say: 'I have this cool photo. Do you want to see it?' Everyone who would like to have it just said: 'Yes." Then the photo would be magically shared."* In the second subcategory, the devices were monitoring the participants' behavior in general and clustered in the same group everybody who was behaving in a similar way (for example, moving together or doing similar actions at the same time).

Overall, the methods based on monitoring natural behavior were considered very easy to use and intuitive. *"[P3] This would probably be the most pleasant method. It feels natural."* The methods also did not require any advance setup effort but could be used at any time. They received high scores in both practicality and pleasantness in the questionnaire. However, the participants were skeptical about whether such methods could be implemented in practice and were concerned that they might trigger actions unintentionally, for example, by saying 'yes' accidentally at the wrong moment. A few participants also raised privacy issues related to continuous monitoring of users.

Real Object

One participant suggested a method where the group membership was indicated by wearing a specific piece of clothing, for example, a certain kind of a t-shirt or a baseball cap, that was recognized by the smartglasses. This method can be seen as an example of a more generic category of methods [1, 30] where the group membership or the object to be moved is represented by a real physical object (a token) that is attached to the target user or device to indicate that the user should join the group or receive the object.

The participants considered that the method was fast and effective but required that somebody prepared the tokens in advance. The method was seen suitable for organized events where every participant was handed out a piece a clothing, for example, an event t-shirt. Different subgroups, for example, competing teams, could have different kinds of t-shirts.

DISCUSSION

Compared to conventional devices such as smartphones, tablets, and computers, the wearable devices were seen to provide many new possibilities for interaction. As the wearables were seen as extensions of the user's body, the interaction methods were expected to be more instant, physical, and simple. Of the two wearable device types included in the study, smartwatches were considered to be closer to conventional devices that the participants had plenty of experience with. *"[P7] Smartwatch is like a small smartphone on your wrist."* The smartwatch was also considered to be more discreet and unnoticeable than the smartglasses – this might change in the longer term future, however, as the form and appearance of smartglasses approaches ordinary eyeglasses. The similarity of smartwatches to conventional devices made it easier to invent methods for them but on the other hand it resulted in a large share of traditional methods such as Scan & Select. Smartglasses, on the other hand, were considered futuristic devices with a lot of novel opportunities. *"[P21] Watches and glasses offer different magnitudes of potential. With glasses, you can do unbelievable things if we just invent them."* As the smartglasses had features like near-eye displays and augmented reality that most participants had little experience of using, it was more difficult to generate ideas for them but on the other hand the generated ideas were more versatile and novel.

As reported earlier by Chong and Gellersen [5], our study further confirms that the perceived affordances of the devices significantly influence the suggested methods. The perceived affordances were defined primarily by two factors: first, the natural physiological capabilities and social functions of the body part that the device was attached to (that is, the participant's head vs. the wrist), and second, the technical features that were perceived possible for a device attached to that body part (for example, what kind of body signals could be measured from that location). As a practical example, gaze pointing and thought commands were primarily suggested for smartglasses, as both the eyes and the brain are located in the user's head near the position of the glasses. On the other hand, the methods suggested for smartwatches emphasized actions made with hands, for example, touching and shared hand gestures, while it would have been equally possible to detect such actions with cameras mounted on the smartglasses.

The participants commented that the group binding and cross-display object movement tasks were similar and felt that in principle, similar solutions could suit to both tasks. However, while some method categories were equally suggested for both tasks, in many categories there was a bias towards one of the tasks. For example, Shared Action and Scan & Select methods were more commonly proposed for the group binding task, while Virtual Object and Command methods were more popular for the object movement task.

The level of how well the participants knew the other participants did not seem to have a major influence on the perceived goodness of the methods. Rather, as there were unknown persons present in the same study group, the participants suggested only methods that were appropriate both with friends and strangers (for example, while handshakes were suggested, hugging was not).

Future Work

While our study has explored a wide range of methods for group binding and cross-display object movement tasks for both smartglasses and smartwatches, a single study can only cover a limited set of potential scenarios of use. There exist many other possible wearable device configurations that should be addressed in future studies. In addition to smartglasses and smartwatches, there are other wearable device types such as smart shoes, belts, and clothes. Users might wear different types of devices, for example, some wearing smartglasses and others smartwatches, requiring methods which can work across device types. Some of the users might have conventional devices, such as smartphones and tablets. A single user might be wearing multiple devices, such as smartglasses and a smartwatch, opening up a design space of methods that combine several devices [7].

Another factor that may have a strong impact on the group binding and cross-display object movement methods is the group size [29]. In our study, we have addressed pairs of users and small groups of three or four users. As the size of the group increases, a much wider variety of different approaches and strategies becomes possible, and the overall process may become more parallel [14]. Future studies could explore group binding and cross-display object movement with wearable devices in larger groups. Regarding group binding, our study covered only the initial group setup – it did not consider managing the changing group membership over time (that is, adding or removing participants later).

In addition to the methods themselves, we aimed to collect as much feedback as possible about the methods from the participants. However, the participants only evaluated the methods that were suggested in their own groups, implying that different groups evaluated different variants of the same method and some of the less common methods were not suggested at all in every group. Therefore, it is not possible to reliably prioritize the method categories based on our data – rather our study focused on collecting a broad range of ideas and initial feedback about those. To reliably evaluate the methods, follow-up studies that systematically present the same methods to every participant would be needed. Preferably these studies should use real prototypes as while many methods can work with mockups, in real life many technical aspects need to be considered [14]. The studies should also address different contexts of use [15, 13]. Ideally, the methods should be tested over extended periods of time in real-life environments [6].

CONCLUSION

We have presented an elicitation study collecting interaction methods for group binding and cross-display object movement tasks on wearable devices from ordinary end users. We have reported a total of 154 methods collected from 30 participants. We have categorized the methods based on the metaphor and modality of interaction, and discussed the strengths and weakness of each category based on qualitative and quantitative feedback given by the participants. The results of our study inform the design of future multi-user applications for wearable devices.

ACKNOWLEDGMENTS

We thank Jarno Ojala for helping with the study arrangements.

REFERENCES

1. Yuji Ayatsuka and Jun Rekimoto. 2005. tranSticks: physically manipulatable virtual connections. In *Proc. CHI '05*, 251-260.
2. Jason Alexander, Teng Han, William Judd, Pourang Irani, and Sriram Subramanian. 2012. Putting your best foot forward: investigating real-world mappings for foot-based gestures. In *Proc. CHI '12*, 1229-1238.
3. Jacob T. Biehl and Brian P. Bailey. 2004. ARIS: an interface for application relocation in an interactive space. In *Proc. GI '04*, 107-116.
4. Ming Ki Chong and Hans Gellersen. 2011. How users associate wireless devices. In *Proc. CHI '11*, 1909-1918.
5. Ming Ki Chong and Hans Gellersen. 2013. How groups of users associate wireless devices. In *Proc. CHI '13*, 1559-1568.
6. Ming Ki Chong, Rene Mayrhofer, and Hans Gellersen. 2014. A survey of user interaction for spontaneous device association. *ACM Computing Surveys.* 47, 1, article 8, 40 pages.
7. Jens Grubert, Matthias Heinisch, Aaron Quigley, and Dieter Schmalstieg. 2015. MultiFi: multi fidelity interaction with displays on and around the body. In *Proc. CHI '15*, 3933-3942.
8. Alina Hang, Emanuel von Zezschwitz, Alexander De Luca, and Heinrich Hussman. 2012. Too much information!: user attitudes towards smartphone sharing. In *Proc. NordiCHI '12,* 284-287.
9. Mountaz Hascoët. 2003. Throwing models for large displays. In *Proc. HCI '03*, 73-77.
10. Lars Erik Holmquist, Friedemann Mattern, Bernt Schiele, Petteri Alahuhta, Michael Beigl, and Hans Gellersen. 2001. Smart-Its Friends: A technique for users to easily establish connections between smart artefacts. In *Proc. Ubicomp '01*, 116-122.
11. Iulia Ion, Marc Langheinrich, Ponnurangam Kumaraguru, and Srdjan Čapkun. 2010. Influence of user perception, security needs, and social factors on device pairing method choices. In *Proc. SOUPS '10*, article 6.
12. Tero Jokela, Ming Ki Chong, Andrés Lucero, and Hans Gellersen. 2015. Connecting devices for collaborative interactions. *interactions* 22, 4, 39-43.
13. Tero Jokela and Andrés Lucero. 2013. A comparative evaluation of touch-based methods to bind mobile devices for collaborative interactions. In *Proc. CHI '13*, 3355-3364.
14. Tero Jokela and Andrés Lucero. 2014. FlexiGroups: binding mobile devices for collaborative interactions in medium-sized groups with device touch. In *Proc. MobileHCI '14*, 369-378.
15. Tero Jokela, Jarno Ojala, Guido Grassel, Petri Piippo, and Thomas Olsson. 2015. A comparison of methods to move visual objects between personal mobile devices in different contexts of use. In *Proc. MobileHCI '15*, 172-181.
16. Ronald Kainda, Ivan Flechais, and A. W. Roscoe. 2010. Two heads are better than one: security and usability of device associations in group scenarios. In *Proc. SOUPS '10*, article 5.
17. Christian Kray, Michael Rohs, Jonathan Hook, and Sven Kratz. 2008. Group coordination and negotiation through spatial proximity regions around mobile devices on augmented tabletops. In Proc. *TABLETOP '08*, 1-8.
18. Byron Lahey, Audrey Girouard, Winslow Burleson, and Roel Vertegaal. 2011. PaperPhone: understanding the use of bend gestures in mobile devices with flexible electronic paper displays. In *Proc. CHI '11*, 1303-1312.
19. Rene Mayrhofer and Martyn Welch. 2007. A human-verifiable authentication protocol using visible laser light. In *Proc. ARES '07*, 1143-1148.
20. Miguel A. Nacenta, Dzmitry Aliakseyeu, Sriram Subramanian, and Carl Gutwin. 2005. A comparison of techniques for multi-display reaching. In *Proc. CHI '05*, 371-380.

21. Miguel Nacenta, Carl Gutwin, Dzmitry Aliakseyeu, and Sriram Subramanian. 2009. There and back again: cross-display object movement in multi-display environments. *Human-Computer Interaction* 24, 1-2, 170-229.
22. Michael Nielsen, Moritz Störring, Thomas B. Moeslund, and Erik Granum. 2004. A procedure for developing intuitive and ergonomic gesture interfaces for HCI. In *Gesture-Based Communication in Human-Computer Interaction*, 409-220.
23. Dan Olsen, Jr. and Travis Nielsen. 2001. Laser pointer interaction. In *Proc. CHI '01*, 17-22.
24. Jun Rekimoto. 1997. Pick-and-drop: a direct manipulation technique for multiple computer environments. In *Proc. UIST '97*, 31-39.
25. Jun Rekimoto. 2004. SyncTap: synchronous user operation for spontaneous network connection. *Personal and Ubiquitous Computing* 8, 2 (May 2004), 126-134.
26. Jun Rekimoto, Yuji Ayatsuka, Michimune Kohno, and Hauro Oba. 2003. Proximal interactions: a direct manipulation technique for wireless networking. In *Proc. INTERACT '03*, 511-518.
27. Jaime Ruiz, Yang Li, and Edward Lank. 2011. User-defined motion gestures for mobile interaction. In *Proc. CHI '11*, 197-206.
28. Scott, Besacier, and McClelland. 2014. Cross-device transfer in a collaborative multi-surface environment without user identification. In *Proc. CTS '14*, 219-226.
29. Lucia Terrenghi, Aaron Quigley, and Alan Dix. 2009. A taxonomy for and analysis of multi-person-display ecosystems. *Personal and Ubiquitous Computing* 13, 8, 583-598.
30. Brygg Ullmer, Hiroshi Ishii, and Dylan Glas. 1998. mediaBlocks: physical containers, transports, and controls for online media. In *Proc. SIGGRAPH '98*, 379-386.
31. Jacob O. Wobbrock, Meredith Ringel Morris, and Andrew D. Wilson. 2009. User-defined gestures for surface computing. In *Proc. CHI '09*, 1083-1092.
32. Jacob O. Wobbrock, Htet Htet Aung, Brandon Rothrock, and Brad A. Myers. 2005. Maximizing the guessability of symbolic input. In *CHI EA '05*, 1869-1872.