"It's Natural to Grab and Pull:" Retrieving Content from Large Displays Using Mid-Air Gestures

Ville Mäkelä, Jobin James, Tuuli Keskinen, Jaakko Hakulinen, Markku Turunen
University of Tampere, Finland

***Abstract*—Mid-air gestures have been largely overlooked for transferring content between large displays and personal mobile devices. To fully utilize the ubiquitous nature of mid-air gestures for this purpose, we developed SimSense, a smart space system which automatically pairs users with their mobile devices based on location data. Users can then interact with a gesture-controlled large display, and move content onto their handheld devices. We investigated two mid-air gestures for content transfer, *grab-and-pull* and *grab-and-drop*, in a user study. Our results show that i) mid-air gestures are well suited for content retrieval scenarios and offer an impressive user experience, ii) grab-and-pull is preferred for scenarios where content is transferred to the user, whereas grab-and-drop is presumably ideal when the recipient is another person or a device, and iii) distinct gestures can be successfully combined with common point-and-dwell mechanics prominent in many gesture-controlled applications.**

***Index Terms*—content transfer, large displays, mid-air gestures, mobile devices, smart spaces, ubiquitous computing.**

I. Introduction

EXCHANGING information between large displays and personal devices is a feature of growing importance in smart spaces. Past research has shown great interest in studying different interaction mechanics for such tasks.

Previous research has largely focused on proxemics-based devices, such as NFC (Near field communication) readers, to enable communication between displays and mobile devices [1][2][3]. Some novel solutions have also been proposed, such as one utilizing the camera of the mobile device to enable drag-and-drop interactions [4]. However, while these solutions may work well for a specific purpose, they also require users to spend time establishing a connection, or require users to specifically walk up to the device to carry out the task. Most notably, they require users to interact with their mobile device in order to successfully transfer content, which may, in many cases, require too much effort from the user. This drawback calls for more advanced and seamless approaches for content transfer.

With this need in mind, we envision ubiquitous spaces where a) ad-hoc connections between mobile devices and large displays are established automatically, and b) content can be moved from one device to another by utilizing mid-air gestures. This vision enables users to simply pass by a display and "grab" interesting content onto their mobile device with a simple gesture and without the need to directly manipulate the recipient device.

In this article, we focus on exploring gestural interaction for content transfer. We built a novel smart space system, SimSense, capable of automatically pairing users with their mobile devices, allowing exchange of content between mobiles and the large screen with simple gestures and without the need to manually set up a connection. We report the results of a user study in which participants experienced two different gestures for information retrieval, *grab-and-pull* and *grab-and-drop*.

II. Designing Mid-Air Gestures For Content Transfer

In our previous work, we investigated transferring content between mobile devices and a large display by utilizing mid-air gestures and an NFC reader [5]. Users could select items from the screen with *point-and-dwell*, after which the selected content could be moved to the mobile device by touching the nearby NFC device with the mobile. We discovered that the point-and-dwell mechanic was suboptimal for content transfer. Primarily, when users were doing something other than interacting with the screen, the screen would still interpret the user's movement as interaction. Then, due to point-and-dwell, actions were occasionally triggered on the screen without the user's intent.

Our approach in this article is inherently different, as we aim for a more novel and seamless solution by completely negating the need to interact with the handheld during content transfer. Regardless, we utilize our previous findings by moving away from point-and-dwell mechanics and instead designing distinct gestures for content retrieval. Additionally, this allows us to utilize point-and-dwell for other, more lightweight interactions, such as navigating the content on the large display.

Received June 2016; revised October 2016. This work was supported in part by Tekes (The Finnish Innovation Agency) for UXUS (User experience and usability in complex systems) research programme under grant 1170/10.

V. Mäkelä, J. James, T. Keskinen, J. Hakulinen, and M. Turunen are with Tampere Unit for Computer-Human Interaction (TAUCHI) at University of Tampere, Kalevantie 4, 33100, Tampere Finland. Contact the authors at ville.mi.makela@sis.uta.fi, jobinjamez@gmail.com, tuuli.keskinen@sis.uta.fi, jaakko.hakulinen@sis.uta.fi, markku.turunen@sis.uta.fi

Our literature review did not reveal existing works that involve mid-air gestures specifically in the context of transferring content between devices. However, studies for more general-purpose mid-air gestures are numerous, which in part affected the design of our content retrieval gestures as well as our user study.

Similar to the findings in our previous work, Yoo and colleagues [6] recently found that users preferred gestures such as *pushing* and *grabbing-and-pulling* over the popular point-and-dwell method for interactions with large information displays. Other techniques for interacting with on-screen item from a distance have also been proposed, such as AirTap and Thumb Trigger [7], and the Gunslinger [8]. These techniques are comparable to usual target selection interaction techniques such as the point-and-dwell. With SimSense, we extend *beyond* the selection of on-screen items, and enable transferring those items to one's personal device. We believe such interactions would benefit from gestures that suggest a link between the large display and the recipient device.

Several works present guidelines for designing mid-air gestures. Nielsen and colleagues [9] provide a set of guidelines for user-centered gesture design. They list that gestures should be 1) easy to perform and remember, 2) intuitive, 3) metaphorically and iconically logical towards functionality, and 4) ergonomic. In one of our earlier studies [10], we also listed several guidelines for designing mid-air gestures, which include e.g. designing for one-handed interaction, and allowing changing the gesturing hand on the fly. Ackad and colleagues [11] studied the design of learnable mid-air gestures, and also presented a set of relevant guidelines, which include e.g. that gesture sets should be small, and on-screen feedback for gestures is important. In their user study, Ackad and colleagues found that an on-screen skeleton helps users understand how the system interprets their movement.

III. SimSense

We developed SimSense[1], a smart space system that allows seamless transferring of content from a large information display to mobile devices by utilizing mid-air gestures and automatic connections between the display and mobile devices [Fig. 1A]. SimSense consists of a large display application, a Kinect sensor, a mobile application, and a set of Bluetooth beacons (the system is presented in detail in [12]).

SimSense is permanently deployed within the premises of the School of Information Sciences at the University of Tampere. The deployment space is occasionally used for lectures and meetings, but is primarily a relaxed, open space filled with tools and displays for students and staff to utilize freely.

The large display application of SimSense runs on a full HD projector screen with physical dimensions of roughly 3.0 meters in width and 1.5 meters in height. The Kinect sensor is positioned at the horizontal midpoint below the screen. Five Bluetooth beacons are positioned on the ceiling of the space above this area, one on each of the four corners and one in the center.

Fig. 1. A) A user is transferring content from the large information display to his mobile device by performing the grab-and-pull gesture. B) An overview of the elements in the large display application. C) "New feed(s) available" notification on the mobile device about received content. D) The SimSense mobile application displaying content that was transferred from the large display.

A. Large Display Interaction

The SimSense large display application serves as a gesture-controlled information display. The primary elements of the display are presented in [Fig. 1B]. The user's hand movements are visualized with an on-screen hand cursor. Interactive elements are triggered with the point-and-dwell mechanic, i.e., by pointing at an element and waiting for a short period of time. During dwelling, a progress bar is overlaid on the element to visualize that the element is being selected. Dwell time was set to be relatively long, 2.5 seconds, to prevent accidental target selections while performing content retrieval

[1] https://youtu.be/RkpjCsNBu3U

gestures.

Interactive elements include content elements and navigational elements [Fig. *1*B]. Triggering a content element will display the content of said element in full in a new page. The page can be closed via a close button displayed on the bottom of the page. Navigational elements can be triggered to cycle between different pages of content.

The user is visualized on the screen in two ways [Fig. *1*B]. An avatar is displayed on the bottom of the screen to reflect the user's horizontal position relative to the screen. Furthermore, the same figure is utilized for feedback. Secondly, we visualize the user's joint movements with an animated stick figure on the top-right corner of the screen.

The large information display can be easily set to present any kind of content. For the study, we chose to include news from Yle, a Finnish news portal, which are fetched via a public API.

B. Connecting Mobile Devices and Users with a Large Display

When a user enters the space with a mobile device that has the SimSense mobile application installed, the device is automatically paired with its owner based on location data. In order to determine the location of the mobile device, the area in front of the large display application is divided into 8 blocks of 50cm X 50 cm dimension, and 100 RSSI (Received Signal Strength Indicator) values from five Bluetooth beacons attached to the ceiling of the space were measured at each of these blocks and stored in a database. The location of the device is predicted by comparing the RSSI values to these previously stored set of values. The location of the user is received via a Kinect sensor. If the two locations match, the entities are paired. Consequently, all content that the user selects on the large screen to be transferred, will be sent to the correct device.

We leave the thorough measurement of the accuracy of the pairing process for another paper; however the process has proven reliable during our evaluations. It is possible that the process fails if two unpaired users enter the space at the same time while being very close to each other and without separating before starting interaction, which is unlikely. We have planned a fallback method for such situations, wherein the system would ask users to step away from each other to complete the pairing reliably. Interaction-wise, two simultaneous users are supported, as the Kinect sensor can accurately track two users at the same time.

C. Content Retrieval

Based on the findings of our previous work as well as existing gesture design guidelines, we designed two content retrieval gestures, *grab-and-pull* and *grab-and-drop*.

Grab-and-pull and grab-and-drop gestures are both initiated by first pointing at a transferable element on the screen and performing a grab gesture (forming a fist with the pointing hand). We note that pointing at an element will also begin the dwell time based selection on said element. To prevent accidental selections, dwell time progression is stopped when the element is grabbed, i.e., when a content retrieval gesture is started.

With grab-and-pull, the user then moves the hand back while keeping the grab gesture – as if physically pulling an object from the screen towards the user. The target element gets larger as the pull progresses to visualize that it is being pulled from the screen. After pulling a certain distance, the grab-and-pull is completed and the content is transferred [Fig. *2*A]. The pull distance was tested via pilot studies and was set to 150 mm.

With grab-and-drop, a drop area is generated on top of the user's avatar when an element is grabbed. The user then drags the grabbed element into the drop area and releases the grab, after which the content is transferred [Fig. *2*B].

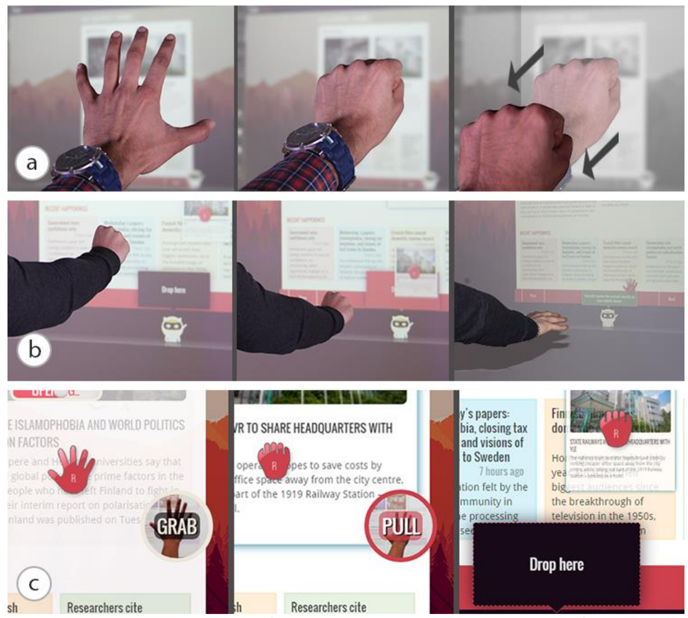

Fig. 2. Interaction techniques for content retrieval and examples of contextual feedback. A) Grab-and-pull: 'grab' and 'pull' an element from the screen. B) Grab-and-drop: 'grab' content and drag and drop it onto a drop area. C) Left: an animated icon visualizing a 'grab' gesture when a transferable element is being hovered. Middle: an animated icon visualizing the 'pull' phase of grab-and-pull after the user has grabbed the corresponding element. Right: with grab-and-drop, a drop area is generated on top of the user's avatar when an element has been grabbed and can be moved around the screen.

Based on the characteristics of both gestures, we produced two minor hypotheses. First, we assumed grab-and-pull to be faster. This was because grab-and-drop requires users to position a dragged element in a specific area, whereas grab-and-pull only requires a one-dimensional pull. Second, we assumed grab-and-drop to appear more familiar and hence result in e.g. increased clarity and confidence. This was due to grab-and-drop's similarity to traditional drag-and-drop mechanics, common in e.g. mouse and touch based interactions.

With both content retrieval gestures, users need to first reach out with their hand (straighten the hand) before elements can be grabbed. This is because i) we wanted to avoid accidental content transfers and ii) the Kinect sensor occasionally triggers false positives and negatives with the grab - reaching out with the hand alleviates the issue somewhat. However, more trivial interactions with the screen, such as navigating the content and reading news, do not require the hand to be reached out.

D. Mobile Application

The SimSense mobile app is an Android application that needs to be installed on the user's mobile device to enable automatic pairing and content transfer. The application triggers a notification on the device whenever new content is received [Fig. 1C]. The application displays all received content in a scrollable list [Fig. 1D]. With the prototype, we concentrated on simple textual and image-based content, such as news. There are currently no other functionalities in the mobile application, although we will investigate extending it in the future.

E. Feedback

One of our emphasized themes in the design was multi-sensory feedback to the user, in a similar manner to what Vogel and Balakrishnan [7] utilized in their study of mid-air pointing and clicking techniques. First, we included traditional visual and auditory feedback on all interactions, such as when an element is hovered or grabbed. Second, and more importantly, we included contextually changing feedback [Fig. 2C]. Finally, the avatar on the screen reacts to the location of the user by moving horizontally on the bottom of the screen, while an animated stick figure reacts to the user's skeletal movements [Fig. 1B].

Additionally, feedback is provided on the mobile application. Whenever content is received, a simple audio tone is played and the mobile vibrates. Moreover, received content is immediately available on the mobile device to be explored further.

IV. User Study

We conducted a 12-participant user study (six males, six females) following a within-subjects design. The participants were 21–43 years old (M=27, SD=6.4). Most participants were students; however their educational background varied from, e.g., linguistics to business. Their prior experience with mid-air gestures was relatively low: three participants had no prior experience, six had used such systems once or twice, and three reported using such systems a few times a year.

We wrote a script for the user study within the SimSense system that would randomly highlight one news headline from the main page of the application. The participant should then select the corresponding headline and transfer it to the mobile device by utilizing one of the gestures (only one was available at a time). After a successful task, the system would wait for three seconds until highlighting the next item.

In the beginning of each study session, the participant filled in a background questionnaire and signed a consent form for video recording. Then, participants were introduced to the application. Next, a mobile phone with the SimSense application was given to the participant, and all participants put the phone in their pocket.

Each participant experienced both techniques (in a counter-balanced order) according to the following pattern:

- Watching a 5-second video demonstrating the technique once.
- Filling in an *expectations* questionnaire about the demonstrated technique.
- Performing 20 content retrieval tasks, with a short break after the first 10.
- Filling in an *experiences* questionnaire about the technique.

After all tasks, subjects filled in a questionnaire comparing the two techniques. Finally, participants were interviewed.

We used a simplified SUXES method [13] to measure participants' expectations and experiences of both interaction techniques. According to the SUXES method [13], *expectations* can show how significant the different factors are and by themselves already provide insight into how people perceive new types of interactive systems, therefore providing better understanding of the user experience. Moreover, we hypothesized that natural user interfaces seen in, e.g., various science fiction films may have resulted in raised expectations, which could not necessarily be met in reality.

We also included open-ended questions, and had participants choose their preferred technique based on varying categories (described in the following chapter). Finally, objective measures from the content retrieval tasks were logged.

V. Results

Participants' expectations and experiences for both techniques are presented in [Fig. 3A]. As we hypothesized, expectations were indeed very high in the majority of categories; however we were able to answer with an equally positive user experience.

Grab-and-pull exceeded expectations in five out of eight categories, and met the expectations in three. Grab-and-drop method, on the other hand, failed to meet the expectations for three of eight statements, although it also met the expectations for four conditions, and exceeded them in one.

Additionally, we presented three statements that were only answered after the experience [Fig. 3B]. Combining statements from both Fig. 3A and Fig. 3B, the user experience of grab-and-pull received higher median scores than grab-and-drop in 10 out of 14 categories. Grab-and-drop scored higher in one category (novelty), and for the remaining three, median scores were the same (naturalness, fun element, confidence that content retrieval was successful).

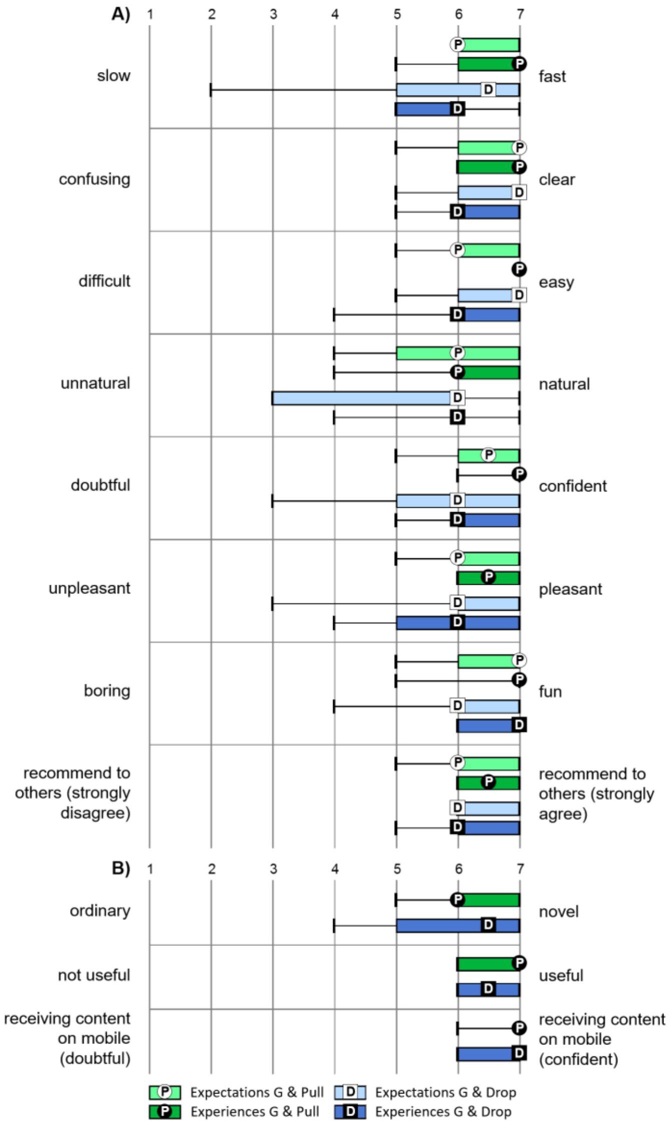

Fig. 3. Boxplots for expectations and experiences for grab-and-pull and grab-and-drop. P and D letters represent the median values.

We assumed that grab-and-pull would be faster than grab-and-drop. Indeed, in addition to users reporting it to be faster [Fig. 3A], a paired samples *t* test shows that the time to complete the grab-and-pull gesture was significantly shorter than the time for grab-and-drop; $t(11) = -10.002$, $p < 0.001$. Grab-and-pull took an average of 0.65 seconds (SD=0.20), whereas grab-and-drop took an average of 1.99 (SD=0.48) seconds.

In general, both techniques received high scores in all categories, as all median scores landed within the 6-7 spectrum. Presumably, one factor in high UX scores for both techniques was that they worked well and users were able to interact with the system successfully. This is supported by the fact that with grab-and-pull, there was only a single time when incorrect content was moved to the mobile, and with grab-and-drop, there were only eight such instances (from two different users). Moreover, instances where a content retrieval gesture was started but not completed were also rare. Users performed an average of 3.0 unfinished gestures (SD = 3.1) with grab-and-pull, and 2.75 unfinished gestures (SD = 2.59) with grab-and-drop. However, we attribute the majority of these occurrences to the Kinect sensor, as the sensor occasionally reported both false positives and false negatives regarding whether the user's hand was in the 'grab' position. A natural open hand stance, with fingers curved slightly inward, seemed to be in a relatively grey area which could be interpreted as either one by the sensor. Some users reported this by stating that sometimes the object they were dragging would "disappear" and they would need to start over. Grab-and-drop may have suffered more from this issue, as the disappearance of an object being dragged is easier to notice than that of an object being pulled (wherein it is more difficult to attribute the issue to a single factor).

We were also interested to see how content retrieval gestures would work in parallel with standard point-and-dwell selections. Positively, we did not encounter a single instance wherein a participant would have accidentally triggered an element via dwelling while aiming to transfer it to the handheld device. Moreover, no participants reported any issues in regards to dwell selections.

Participants named their preferred technique in four categories, and also named their overall preferred technique for content retrieval (Fig. 4). Nine out of 12 participants reported that grab-and-pull was more natural than grab-and-drop. More balanced results were received when asked which technique they were more confident performing (i.e., how sure they were that they were committing the gesture correctly), however the fun element and pleasantness categories again showed preference towards grab-and-pull. Many participants added that grab-and-pull felt more pleasant because it was slightly easier to perform. Finally, 10 out of 12 participants chose grab-and-pull as the overall preferred content retrieval method, while grab-and-drop was preferred by two subjects.

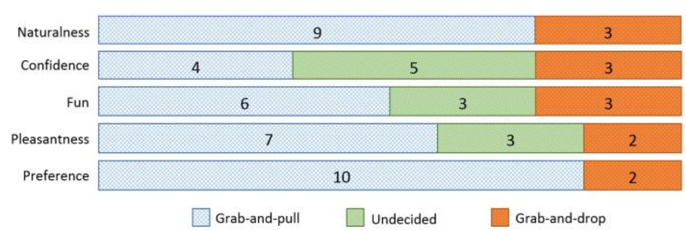

Fig. 4. Subjects' preferred techniques based on naturalness, confidence, fun element, pleasantness, and overall preference.

Many participants explained their preferences for naturalness based on who or what is receiving the content, e.g.:

> *If I want something for myself, it's simply natural to just 'grab and take', and not 'grab and move somewhere'.*
> - Participant 1

> *Grab-and-pull was more natural and more appropriate for this specific purpose. But if I'm thinking multiple devices, transferring from one device to another, grab-and-drop would be more natural. I could also want to transfer content to another person, and again grab-and-drop would be a better choice.*
> - Participant 3

Interestingly, four participants explicitly reported that they initially expected to prefer grab-and-drop, but changed their opinion after experiencing both techniques. Presumably, users were initially attracted by the perceived familiarity of grab-and-drop, but ultimately preferred the speed and ease of use of grab-and-pull.

Participants gave positive remarks of the feedback mechanisms. Generally, feedback was reported to be clear and helpful, and that the system was successfully guiding participants through the interaction process. One participant mentioned:

It was good to know that it [content] went to the phone immediately. I had a feeling that I always knew what was happening.
- Participant 5

Finally, participants were seemingly impressed and satisfied with the overall experience. Some commented:

I felt like I was in a movie, in the future.
- Participant 4

It [interacting] feels almost like I'm a superhero.
- Participant 5

[Interacting] is so much fun actually that you don't even care about the content, you just want to do it [interact].
- Participant 6

VI. Discussion

The multifaceted results of our user study suggest that *mid-air gestures are well-suited for content retrieval scenarios*. Both gestures, grab-and-drop and grab-and-pull, received high UX scores and were generally appreciated by participants. In addition, users reported a positive, novel and impressive experience with the system as a whole, and some users made remarks that interacting made them feel like they are superheroes or are a part of a futuristic movie. One participant reported having so much fun with the system that he did not even care about the content, but just wanted to keep interacting.

We successfully mixed point-and-dwell item selections with distinct (content retrieval) gestures. Participants did not commit accidental item selections, not even during content retrieval, nor did they report issues with mixing two different kinds of interactions. This is an encouraging finding and suggests that mid-air gestures can be utilized for a wider range of interactions than many existing works have proposed [9][10][11]. We believe the key factors in this success are that the two mechanics did not overlap beyond the initial act of pointing to an item, and that dwell time was stopped whenever an element was grabbed. Therefore, the dwell time of 2.5 seconds could likely be shortened if needed.

We were able to design two successful gestures with the help of existing guidelines. Based on recent findings, we made the initial decision to move away from dwelling-based mechanics [5][6]. In addition, we followed a number of existing gesture design guidelines [9][10][11] in designing the grab-and-drop and grab-and-pull gestures.

Interestingly, despite highly positive results for both gestures, *grab-and-pull was clearly preferred by users for content retrieval* (10 out of 12 participants). Based on our results, we identified three major factors for the outcome. First, grab-and-pull was reported to feel conceptually more natural, i.e., pulling content towards oneself to transfer it felt more appropriate than dropping it onto an area on the screen. Second, grab-and-pull was faster, and third, it was easier to perform as it required less accuracy. We note here that while the sample size of 12 was relatively limited, participants were recruited from various backgrounds, and the results were consistent across participants.

The reasons for preferring grab-and-pull are logical, however we were surprised by the relatively clear difference, especially considering the positive feedback and experiences for both techniques. We initially expected one of the strong points of grab-and-drop to be its familiarity from other modalities. Considering this, it is interesting to note that users may have had the same expectation: four users explicitly stated that they initially expected to prefer grab-and-drop, but switched to grab-and-pull after experiencing both techniques.

We note that despite grab-and-pull being preferred, said preference may only apply to a specific scenario of transferring content to *oneself*. Many users noted that it simply makes sense that when they want something for themselves, they pull it towards them. On the other hand, users also suggested other interesting scenarios, such as transferring content from one display to another, where they reported that grab-and-drop might work better than a pulling gesture, wherein a drop area on the screen [Fig. 2C] would represent the recipient display instead of the user. Another suggested scenario for grab-and-drop was transferring content to someone else, such as a friend. Therefore, we argue that grab-and-pull is favorable for transferring content to the user, and grab-and-drop is a valid option for scenarios where the recipient is someone or something other than the user. We believe that these techniques can even be mixed into the same system, both for their own purposes.

We note some limitations in our study. We did not measure the intuitiveness of the gestures due to showing participants short demonstrations before the study tasks. This was because we aimed to measure their expectations as opposed to experiences. Regardless, our results support our claim that the designed content retrieval gestures are intuitive as both were regarded as natural and logical. Moreover, Yoo et al. [6] investigated a relatively similar grab-and-pull gesture in a different scenario, and they also make a case for intuitive gestures.

Arm fatigue was also not measured, which is a known issue with mid-air gestures [14]. However, we did account for fatigue by including a break after each set of 10 content retrieval tasks, and no participants mentioned nor seemed to be getting fatigued. Moreover, interactions with large displays are generally short, as are realistic content retrieval scenarios: users would pass by the system, grab a few interesting items with them, and continue on their way.

It is possible that the participants' overall impressed attitudes would somewhat diminish over time when the novelty effect of the system wears of. However, the majority of our results are not directly related to the novelty of the interaction. For instance, the deciding factors in the preference of grab-and-pull were *naturalness*, *ease of use*, and *speed*.

One interesting question is how people feel about their mobile devices automatically connecting to a public service. No one reported privacy-related issues during the study, and the participants' overall positive attitude towards the system provides encouragement. However, the relaxed and calm deployment setting may play a role in this. Hectic and less familiar environments, such as city centers, would presumably propose privacy and other concerns. Another noteworthy addition to this discussion is the requirement of a separate mobile application. We argue this is a positive, as installing an application serves as a permission for the device to provide its location to the service. Generally, the aspect of user acceptance, privacy, and the effects of the deployment location should be investigated in the future.

VII. Conclusion

Mid-air gestures can offer an impressive user experience in content retrieval scenarios between large displays and mobile devices. We discovered an ideal gesture for personal content transfer interactions, *grab-and-pull*, wherein users "grab" an item from the screen and pull it towards themselves. We also successfully mixed content retrieval gestures with common point-and-dwell mechanics, which we believe encourages designing more diverse gesture-controlled applications.

Our work makes way for designing ubiquitous spaces wherein interaction between public and personal devices is made more seamless and accessible by enabling interaction from anywhere in the space. Possible continuations to our work include transferring content from the mobile device back to the large screen, between several large screens, and between people.

Acknowledgment

Authors gratefully acknowledge the support by Tekes (The Finnish Innovation Agency) for UXUS (User experience and usability in complex systems) research programme.

Biographies

All authors are affiliated with the Tampere Unit for Computer-Human Interaction (TAUCHI) at University of Tampere, Finland.

Ville Mäkelä is a Ph.D. candidate at TAUCHI. His research focuses on mid-air gestures and content transfer between public and personal devices. He received his M.Sc. in interactive technology from the University of Tampere in 2013. Contact him at ville.mi.makela@staff.uta.fi.

Jobin James was a researcher at TAUCHI. His research interests include design of interaction techniques for large public displays, and design of ubiquitous environments with such displays. He received his M.Sc. in human technology interaction from the University of Tampere in 2015. Contact him at jobinjamez@gmail.com.

Tuuli Keskinen is a postdoctoral researcher at TAUCHI. Her research focuses on user experience evaluation. In 2010, she received her M.Sc. degree and in 2015 her Ph.D. degree – both in interactive technology from the University of Tampere, Finland. Contact her at tuuli.keskinen@sis.uta.fi.

Jaakko Hakulinen is a university researcher at TAUCHI. His current research interests include multimodal interaction utilizing large scale of technology, e.g., gesture-based interaction, speech-based interaction, and intelligent environments. He received the M.Sc. degree in computer science from the University of Joensuu, Finland, in 1998, and Ph.D. degrees in interactive technology from University of Tampere, Finland in 2006. Contact him at jaakko.hakulinen@staff.uta.fi.

Markku Turunen is a professor of Interactive Technology at TAUCHI. His research interests include e.g. multimodal systems, novel user interfaces, and evaluation methods for mobile and pervasive computing applications. He received his Ph.D. in computer science from the University of Tampere in 2004. Contact him at markku.turunen@sis.uta.fi.

References

[1] G. Broll, W. Reithmeier, P. Holleis, and M. Wagner, "Design and evaluation of techniques for mobile interaction with dynamic NFC-displays", in *Proc. TEI'11*, 2011, pp. 205-212.

[2] R. Hardy, and E. Rukzio, "Touch & Interact: touch-based interaction of mobile phones with displays", in *Proc. MobileHCI'08*, 2008, pp. 245-254.

[3] K. Seewoonauth, E. Rukzio, R. Hardy, and P. Holleis, "Touch & connect and touch & select: interacting with a computer by touching it with a mobile phone", in *Proc. MobileHCI'09*, 2009, 36.

[4] M. Baldauf, K. Lasinger, and P. Fröhlich, "Real-world drag'n'drop – bidirectional camera-based media transfer between smartphones and large displays", in *Adj. Proceedings of the 10th International Conference on Pervasive Computing*, 2012.

[5] V. Mäkelä, H. Korhonen, J. Ojala, A. Järvi, K. Väänänen, R. Raisamo, and M. Turunen, "Investigating mid-air gestures and handhelds in motion tracked environments", in *Proc. PerDis'16*, 2016, pp. 45-51.

[6] S. Yoo, C. Parker, J. Kay, and M. Tomitsch, "To dwell or not to dwell: an evaluation of mid-air gestures for large information displays", in *Proc. OzCHI'15*, 2015, pp. 187-191.

[7] D. Vogel and R. Balakrishnan, “Distant freehand pointing and clicking on very large, high resolution displays”, in *Proc. UIST’05*, 2005, pp. 33-42.

[8] M. Liu, M. Nancel, and D. Vogel, “Gunslinger: subtle arms-down mid-air interaction”, in *Proc. UIST’15*, 2015, pp. 63-71.

[9] M. Nielsen, M. Störring, T.B. Moeslund, and E. Granum, “A procedure for developing intuitive and ergonomic gesture interfaces for HCI”, in *Gesture-Based Communication in Human-Computer Interaction*, vol. 2915, Springer Berlin Heidelberg, 2003, pp. 409-420.

[10] V. Mäkelä, T. Heimonen, M. Luhtala, and M. Turunen, ”Information wall: evaluation of a gesture-controlled public display”, in *Proc. MUM’14*, 2014, pp. 228-231.

[11] C. Ackad, A. Clayphan, M. Tomitsch, and J. Kay, “An in-the-wild study of learning mid-air gestures to browse hierarchical information at a large interactive public display”, in *Proc. UbiComp’15*, 2015, pp. 1227-1238.

[12] J.M. James, “SimSense - Gestural Interaction Design for Information Exchange between Large Public Displays and Personal Mobile Devices”, M.S. thesis, School of Information Sciences, University of Tampere, Tampere, Finland, 2015.

[13] M. Turunen, J. Hakulinen, A. Melto, T. Heimonen, T. Laivo, and J. Hella, ”SUXES – user experience evaluation method for spoken and multimodal interaction”, in *Proc. ISCA*’09, 2009, pp. 109-112.

[14] J.D. Hincapié-Ramos, X. Guo, P. Moghadasian, and P. Irani, “Consumed endurance: a metric to quantify arm fatigue of mid-air interactions”, in *Proc. CHI’14*, 2014, pp. 1063-1072.