Touch and Gesture: Mediating Content Display, Inscriptions, and Gestures across Multiple Devices

Gerard Oleksik
Instrata Ltd.
12 Warkworth Street, Cambridge
United Kingdom
+44 (0)1223 301 101
gerard@interactables.com

Natasa Milic-Frayling
Microsoft Research Ltd.
7 J J Thomson Ave., Cambridge
United Kingdom
+44 (0)1223 479 700
natasamf@microsoft.com

Rachel Jones
Instrata Ltd.
12 Warkworth Street, Cambridge
United Kingdom
+44 (0)1223 301 101
racheljones@instrata.co.uk

ABSTRACT

Recent advances in design and technology have broadened the range of devices that enable human-computer interaction through multi-touch and increased their adoption in collaborative work settings. Since most of the research has focussed on optimal use of individual devices, we now need to expand our understanding of how these devices are used in concert and what is required to support user interactions across multiple devices. We conducted in-situ observations of team meetings that involve the use of a tabletop computer, tablet PCs, and a vertical display. The study shows how inscriptions and gestures naturally emerge around the content and how important it is to maintain their spatial congruence. Furthermore, the combination of the tablet PCs and the tabletop encourages the use of gestures and touch across devices. The users often apply sequential and synchronous gestures to bind the content and inscriptions across the devices in support of sense-making. The observed binding gestures extend the notion of multi-touch beyond the individual devices and require a unified approach to the touch and gesture support.

Categories and Subject Descriptors

H.5 [**Information Interfaces and Presentation**]: H.5.2. User Interfaces.

General Terms

Design, Human Factors.

Keywords

Gesture, inscription, touch, tabletop, tablets, vertical display, meeting place, deictic gestures, binding gestures.

1. INTRODUCTION

Touch enabled displays and user interfaces have long been researched as the means of facilitating natural interaction with computing devices. Recently, the commercialization and take-up of multi-touch mobile phones, slate computers, and tabletop computers have increased the use of touch interactions and opened up opportunities for studying emerging user practices and experiences.

Particularly interesting are scenarios where multiple devices are used to facilitate collaborative work, each device contributing its specific interaction facilities. Generally, multi-device settings have been studied ([10],[20]) but only few have looked at the real usage scenarios [2]. In order to deepen our understanding of the issues that arise in such environments, we conducted in-situ observations of team meetings that involved the use of a tabletop computer, tablet personal computers (tablet PCs), and a PC with a large vertical display. By observing a real work setting, we study interactions that naturally occur as individuals displayed the content across the devices and used gesture and inscriptions to facilitate their discussions.

We conducted the meeting observations over a period of six months, starting with a setup that first included only tablet PCs and a large vertical display, and then was extended with a tabletop computer. Based on the previous studies of the tabletop use we anticipated increased use of gestures ([15],[18]). However, from our data we gained new insights and studied the emergence and purpose of gestures across the devices. In particular, we noted the importance of sequential and synchronized deictic gestures that were used to indicate connections among related resources on the same device and across devices. We refer to them as *binding* gestures. The binding gestures essentially extend the notion of multi-touch across devices and highlight the need for a unified approach in supporting touch, gesture, and inscription. Furthermore, we noted a critical importance of the spatial configuration of devices and participants in the meetings. It became apparent that multiple devices cause bifurcation of the user's attention across devices and thus require an effort to maintain the congruence of the content, inscriptions and gestures that are used in communication.

In the following sections we describe the study findings in detail and put forward recommendations for designs and methodologies to improve multi-device meeting environments.

2. RELATED RESEARCH

From previous studies of meeting places, we expected meeting conversations to be strongly supported by gestures and inscriptions ([14],[4],[1]). Thus, in our study we pay a particular attention to how they are manifested in the interaction with multiple touch-enabled devices. In the following section we briefly review selected literature related to gesture, inscription, and multi-device environments.

2.1 Gestures

2.1.1 Meaning and Role of Gesture

Gestures refer to physical movements of hands, head, and other parts of the body used in information exchange and interaction. The breadth of human gesture is broad, and they have been studied in conjunction with inscriptions, touch, and speech.

Kendon [11] proposes a 5 point continuum for describing the degree of formalism underpinning gesture. That continuum ranges

Setup 1. *Setup 2.* *Setup 3.*

Figure 1. Three different device configurations used in the meetings.

from free-form *gesticulation* that accompanies speech to the sign language, complete with vocabulary and grammar.

Focusing on discursive human gestures, McNeil [14] identified (1) iconic gestures that relate to 'the semantic content of the speech' and provide a visual back up for what is being said; (2) metaphoric gestures that are pictorial but present an abstract idea rather than a concrete object of event; (3) beat gestures that are rhythmic accompaniments to speech and may emphasize the importance of particular words; (4) cohesive gestures which bind together what is being said, and (5) deictic or pointing gestures that direct listeners' attention to specific objects as they are mentioned. Bekker et al. [1] studied the use of gesture in face-to-face meetings among 10 design teams and found that team members often perform multiple gestures in sequence. Such sequenced gestures work in concert with each other.

2.1.2 Multi-device Environments

Most of the research concerning multi-device environments focused on interactions that support sharing and replication of data across devices. They involved prescribed interactions that users needed to follow in order to achieve a given objective.

For example, *Toss-It* facilitates transfer of data between PDAs and mobile devices through simple 'throwing' gestures between mobile devices [21]. *Point&Connect* enables users to pair mobile devices by moving the devices closer together [18]. With *Touch and Interact* the user can pass data from a mobile device to a large display by touching the screen with the phone [6]. Some include a pen to enable users to move data and pair devices. For example, *Pick Up and Drop* allows the user to use a pen to touch a digital object on a display and drop it onto another display or a different part of the same display [17]. The system by Lee et al. [13] enables users to connect mobile, large screen, and tabletop devices and share data between them through semaphore and pointing gestures. In order to share digital objects, the user touches the item and then points to a screen or a device where data needs to be transported. Hinckley [7] uses synchronous gestures to enable users to establish connections between tablet devices by bumping them together. Through a titling gesture, the user can then 'pour' data from one device to another.

2.2 Inscriptions

Inscription refers to persistent marks, sketches, or images made through the act of writing, drawing, printing, and engraving onto a surface. In the case of tablet PCs, many applications aim to record and recognize hand-written inscriptions.

Cortina et al. [3] report on the importance of inscriptions in support of mathematical learning and problem solving. They describe how the inscription of a mathematical problem in the classroom becomes a representation of the problem and a scaffold for collective reasoning and attention. The work by Goodwin [4] underlines the importance of placing inscriptions in the close proximity of their focal point, e.g., an archaeological artifact that cannot be physically moved. The interpretative function of the inscription is actuated within the same visual field as the content which inspired it. Furthermore, the research shows a fine interplay between inscriptions and gestures. Streeck & Kallmeyer [19] state that, because of their persistent nature, "inscriptions can become the targets or components of further symbolic actions", including physical gestures.

In our research, we aim to explore (1) the modes of interaction that naturally arise in multi-device environments and (2) the role and meaning that user gestures assume. We were aware that the existing multi-touch support may influence and limit our understanding of the natural gestures. Fortunately, most of the applications used by the study participants on the tabletop were not touch enabled and required the use of the mouse. Thus, the findings can be applied to extending the current multi-touch facilities with support for the identified gestures and inscriptions.

3. STUDY

We conducted in-situ observations of meeting sessions at a university research centre. We investigated how participants use multiple devices to display, manipulate, and create content during their discussions and what impact the computing technologies have on their actions.

Meetings are held in the research leader's office, and are attended by the research leader, post doctorate staff, and doctoral students from two closely related research groups. The computing infrastructure of the meeting setup involved several networked computing devices: a static PC with large display, tablet PCs, and a tabletop PC. Each researcher is equipped with their own tablet PC and skilled in touch based interaction and inscriptions using a stylus. They adapted MS OneNote software to serve as a Lab book. The tabletop computer provides multi-touch interaction with software applications that take advantage of the multi-touch capability. Otherwise, the content is accessed using the mouse. In fact, that was the case with the documents used in most meetings. The vertical display was primarily used to project content for group viewing. These multiple devices are used in concert to facilitate the meeting.

3.1 Study Method

We adopted ethnographic approach of in-situ observations using video-recording as an aid to collect data and conduct post-observation analysis. The analysis was based on the total of 10 hour in-situ observations involving 7 separate meetings of 13 researchers from two research teams. All the meetings were held in the same physical location but in 3 different meeting setups, as the meeting space evolved over time to include the tabletop PC and assumed different spatial configurations of devices (Figure 1):

Setup 1 - Attendees sat around the research leader's desk, with their tablet PCs. A 26 inch vertical monitor was used to display content for group viewing.

Setup 2 - Attendees sat around the tabletop computer located next to the leader's desk, with their tablet PCs. The tabletop computer and the vertical monitor on the leader's desk were used for the group viewing of the meeting material.

Setup 3 - Attendees sat around the research leader's desk with the integrated tabletop computer, with their tablet PCs. The vertical monitor and the tabletop computer were used for group viewing of the meeting material. The applications used on the tabletop were not touch enabled.

Thus, the users' gestures did not interfere with the content display and therefore unfolded fully, without causing unintentional movement of objects, activation of software, and similar. This enabled us to detect emerging gestures that connect disparate content and inscriptions across devices.

3.2 Study Findings

As expected from previous research by Bekker et al. [1], deictic gestures were used extensively across all meeting setups: indirect, gestures to indicate a part of the screen with the mouse cursor or a finger pointing to a distant display, and direct, by touching the surface to point to a displayed artifact [5].

The highest use of indirect deictic gestures was observed in the meeting setup 1 where the vertical display provided the shared view of the content. Participants sat relatively far from the shared display. Thus, a high proportion of deictic gestures were made over distance. The participants situated closer to the display were able to directly point to parts of the screen to indicate what they were referring to.

In the meeting setups 2 and 3, the shared content was displayed on the tabletop computer. Gesturing to the content was markedly different, with high incidence of direct deictic gesture from both the meeting leader and other participants. The gestural language increased in complexity to include one-handed and two-handed gestural walkthroughs and finger tracing over content to support verbal explanations. We describe observed gestures in more details but, first, we reflect on the issues that arise due to the spatial configuration of the multi-device environment.

3.2.1 Content Management Across Devices

During the meetings, the participants used content and inscriptions across devices. The research leader, referred to as John, made annotations and sketches related to the discussed content and took notes on behalf of the group using MS OneNote application on his tablet. In order to make the content and the inscriptions visible to the group, he would display them on the shared monitor. While this action increased the visibility of the content, it had a knock off effect on certain interpretative gestures.

Indeed, in creating a sketch to explain the content, John would use his tablet as an inscription device. His sketches would then become a resource for interpretation and action. John periodically gestured to the parts of the inscriptions in support of the explanations that he gave during their creation. Critically, when John would start gesturing to the inscription on the tablet, a split between content and interpretation was created. The interpretative gestures made by John were not accessible to those looking at the vertical display. As a result, students would switch attention to the tablet to mediate the bifurcation of attention caused by two displays. They choose the one which unifies inscription and gestures.

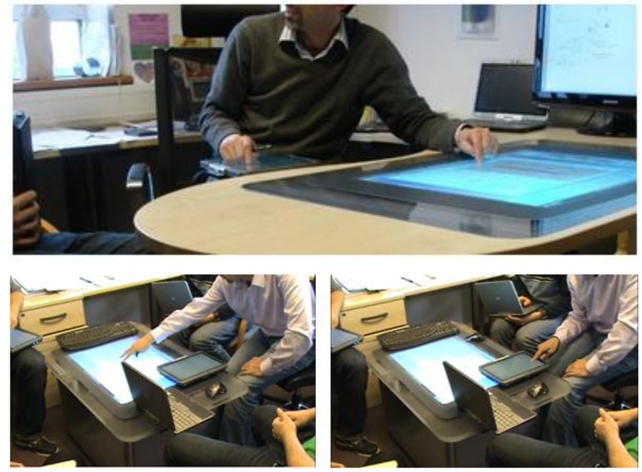

Figure 2. Direct synchronous gesture, binding the inscription on the tablet and the content displayed on the tablet (upper). Sequential binding between tablet and tabletop (lower image).

In another example, in the meeting setup 3, John repositioned the tablet, placing it on the tabletop in such a way that all meeting members can view the content (Figure 2, lower). This movement of the tablet to a more central and accessible location had an immediate effect on the meeting—more participants gestured to the content.

These examples illustrate the spatial separation of the content and interpretative gestures that can occur with multiple displays. Participants refocus their attention to the area where content, inscriptions, and gestures are unified and provide a higher value than the content display alone.

Content Binding

Observed meeting discussions frequently involved resources displayed on separate devices: sketches on the tablets, notes on the vertical display, slides on the tabletop, and similar. Explanations often required referencing of distinct resources and directing the user attention to specific content through gestures.

Our analysis revealed a central role of the specific gesture patterns that we refer to as *binding gestures*. They serve to indicate associations and make the connections among displayed items explicit. Binding gestures manifest themselves differently across the meeting setups. We discuss examples of different types of binding gestures.

Hybrid sequential binding. In a one-to-one meeting in setup 1, John has produced a sketch to describe a process to the student, Peter, for the next experiment. On completing the sketch, John turns to Peter's slides on the shared vertical display and uses the mouse to gesture to a graph on a slide. Immediately upon doing so, John points directly to a part of his sketch on the tablet and then back to the monitor, emphasizing the connection between the two. This form of binding is sequential and hybrid as it involves indirect gesture via mouse and direct touch on the tablet.

Direct sequential binding. In setup 2, meeting participants are discussing slides prepared by a student. The slides are displayed on the tabletop computer and John is resting his tablet on the tabletop surface. He and post doc Zak are tracing their fingers over a graph displayed on the tabletop as they talk. John begins to describe a solution and starts to sketch it on his tablet as he talks. He then makes a number of binding gestures by pointing first to the slide on the tabletop and then to the sketch on the tablet. Thus, unlike the previous example, this binding is achieved through a sequence of two direct deictic gestures (Figure 2).

Direct synchronous binding across the tablet and the tabletop. We observed John holding the stylus on a part of the sketch on his tablet PC and simultaneously placing and holding his finger on the tabletop image, while verbally explaining the connection between the two. This gesture helped John elaborate on the relationship between content through direct deictic gestures that occur synchronously across the two devices (Figure 2, upper).

4. DESIGN CONSIDERATIONS

Our study revealed that multi-device collaborative environments present a significant challenge to preserving the congruence between the display of the content and the visibility of the gestures and inscriptions. The study individuals chose to trade the convenience of the large content display for the unified view of the gestures, inscriptions, and content. This suggests techniques to project or simulate gestures on the shared display. The touch gestures could be easily captured and overlaid over the content. C-slate by Izadi et al. [8] demonstrates that gesture tracking can be achieved in real time. However, the gestures above the display surfaces can be detected and represented digitally only through 3D gesture detection and tracking technologies. Generally, the existing techniques for projecting gestures into remote collaboration spaces could be adopted for that purposes [12]. We would first need to characterize the type and the objective of the gesture. For example, a kinetic gesture used to underscore a formula in the paragraph. Once the characteristics are known, we can apply appropriate display strategies to highlight the elements that are the focus of the gestures, touch, and inscription.

We observed that the gestures above and in-front of the display surfaces are essential for communication. There have been attempts to use that space for additional display functionalities [9]. However, that has to be done with care, particularly in the multi-device environments. Our observations of sequential and synchronized *deictic gestures* show how gestures form as part of the sense making process. In contrast to the touch gestures associated with specific commands, the binding gestures introduce multi-touch across devices to convey association among content pieces.

This raises important requirements for the design of touch support. First, it calls for the consistency and coordination of the touch commands across devices. Second, it requires that standard touch commands do not overlap with the direct deictic gestures that evolve during sense-making and, thus, may be confused for touch commands. Finally, with the advances of the real-time 3D gesture tracking and recognition, we anticipate closer integration of touch and 3D gestures as input techniques. Since 3D gestures naturally emerge in communication, similar gestures may have different meaning in the command mode verse the gesticulation mode. Thus, the methods for modality detection and enforcement would be of utmost importance to support intended user actions.

5. REFERENCES

[1] Bekker, M. M., Olson, J. S., and Olson, G. M. 1995. Analysis of gestures in face-to-face design teams provides guidance for how to use groupware in design. In Proc. of the conference on Designing interactive systems. 157–166

[2] Biehl, J. T., Baker, W. T., Bailey, B. P., Tan, D. S., Inkpen, K. M., and Czerwinski, M. 2008. Impromptu: a new interaction framework for supporting collaboration in multiple display environments and its field evaluation for co-located software development. In Proceedings of CHI'08, 939–948.

[3] Cortina, L. J., Zhao, Q., Cobb, P., McClain, K. 2003. Supporting students' reasoning with inscriptions. In Proceedings of 6th International Conference of Learning Science, 124–149.

[4] Goodwin, C. 2003. Pointing as Situated Practice. In S. Kito, Pointing: Where Language, Culture, and Cognition Meet. Mahwah, NJ: Lawrence Erlbaum Associates, Inc., 2003, 217–242.

[5] Ha, V., Inkpen, K. M., Whalen, T., and Mandryk, R. L. 2006. Direct Intentions: The Effects of Input Devices on Collaboration around a Tabletop Display. In Proc. of the First IEEE International Workshop on Horizontal Interactive Human-Computer Systems, 177–184.

[6] Hardy, R. and Rukzio, E. 2008. Touch & interact: touch-based interaction of mobile phones with displays. In Proc. of the 10th Int. Conf. on HCI with Mobile Devices and Services, 2008, 245–254.

[7] Hinckley, K. 2003. Synchronous gestures for multiple persons and computers. In Proc. of the 16th annual ACM symposium on User interface software and technology, 149–158.

[8] Izadi, S., Agarwal, A., Criminisi, A., Winn, J. Blake, A. & Fitzgibbon, A. 2007. C-Slate: Exploring Remote Collaboration on Horizontal Multi-touch Surfaces. Proc. IEEE Tabletop 2007, 3–10.

[9] Izadi, S., Hodges, S., Taylor, S., Rosenfeld, D., Villar, N., Butler, A., & Westhues, J. 2008. Going Beyond the Display: A Surface Technology with an Electronically Switchable Diffuser. In Proc. of ACM UIST'08, 269–278.

[10] Johanson, B., Hutchines, G., Winogradm T. 2002. The Interactive Workspaces Project: Experiences with Ubiquitous Computing Rooms. IEEE Pervasive Computing, 1 (2): 67–74.

[11] Kendon, A. 1988. How gestures can become like words. In Potyatos, F. (Ed.), Crosscultural perspectives in nonverbal communication, Toronto, Canada, 131–141.

[12] Kirk, D., Crabtree, A., Rodden, T. 2005. Ways of the hands. In Proceedings of the Ninth Conference on European Conference on CSCW, September 18-22, 2005, Paris, France, 1–21.

[13] Lee, H., Jeong, H., Lee, J., Yeom, K., and Park, J. 2009. Gesture-Based Interface for Connection and Control of Multi-device in a Tabletop Display Environment. In Proc. HCI 2009, 216–225.

[14] McNeill, D. 1992. Hand and mind: What gestures reveal about thought. Chicago: University of Chicago Press.

[15] Piper, A. M., Hollan, J, D. 2008. Supporting medical conversations between deaf and hearing individuals with tabletop displays. In Proc. of CSCW 2008, November 08-12, 2008, San Diego, USA, 147–156.

[16] Peng, C., Shen, G., Zhang, Y., Lu, S. 2009. Point&Connect: Intention-based Device Pairing for Mobile Phone Users. ACM/USENIX MobiSys 2009, 137–150.

[17] Rekimoto, J. 1997. Pick-and-Drop.: A Direct Manipulation Technique for Multiple Computer Environments. In Proc. UIST 1999, 31–39.

[18] Rogers, Y., Lindley, S.E. 2004. Collaborating around vertical and horizontal large interactive displays: Which way is best? Interacting with Computers, 16, 6, 1133–1152.

[19] Streeck, J. & Kallmeyer, W. 2004. Interaction by inscription. Journal of Pragmatics, 33(4).

[20] Streitz, N. A., Geißler, J., Holmer, T., Konomi, S., Müll+er-Tomfelde, C., Reischl, W., Rexroth, P., Seitz, P., Steinmetz, R. 1999. i-LAND: an interactive landscape for creativity and innovation. In Proc. SIGCHI, May 15–20, 1999, 120-127.

[21] Yatani, K., Tamura, K., Hiroki, K., Sugimoto, M., and Hashizume, H. 2006. Toss-It: Intuitive Information Transfer Techniques for Mobile Devices Using Toss and Swing Actions. IEICE-Trans. Inf. Syst., 89 (1): 150–157.