Cross-Device Interaction via Micro-mobility and F-formations

Nicolai Marquardt[1,2], Ken Hinckley[1], Saul Greenberg[2]

[1]Microsoft Research
One Microsoft Way, Redmond WA 98052
kenh@microsoft.com

[2]University of Calgary, Dept. of
Computer Science, Calgary, AB, Canada
{nicolai.marquardt, saul}@ucalgary.ca

ABSTRACT

GroupTogether is a system that explores cross-device interaction using two sociological constructs. First, *F-formations* concern the distance and relative body orientation among multiple users, which indicate when and how people position themselves as a group. Second, *micro-mobility* describes how people orient and tilt devices towards one another to promote fine-grained sharing during co-present collaboration. We sense these constructs using: (a) a pair of overhead Kinect depth cameras to sense small groups of people, (b) low-power 8GHz band radio modules to establish the identity, presence, and coarse-grained relative locations of devices, and (c) accelerometers to detect tilting of slate devices. The resulting system supports fluid, minimally disruptive techniques for co-located collaboration by leveraging the *proxemics of people* as well as the *proxemics of devices*.

Author Keywords

Proxemics; F-formations; micro-mobility; tablets; sensors

ACM Classification Keywords

H.5.2 [Information Interfaces and Presentation]: Input

INTRODUCTION

Despite the ongoing proliferation of devices and form-factors such as slates and electronic whiteboards, technology often hinders (rather than helps) informal small-group interactions. Whereas natural human conversation is fluid and dynamic, discussions that rely on digital content—slides, documents, clippings—often remain stilted due to the awkwardness of manipulating, sharing, and displaying information on and across multiple devices.

To address these shortcomings, we leverage two sociological constructs that have been observed in human-human social activity: first, the interpersonal (*proxemic*) dynamics of small-group gatherings known as *F-formations*; and second, how people employ *micro-mobility* of physical artifacts to afford nuanced collaboration. Multiple sensing modalities enable us to capture and make use of these constructs in novel multi-device interaction techniques. Paired together, these approaches afford a lightweight federation of multiple devices and displays that leverage proxemic cues derived from: (a) the human-to-

human proxemic relationships resulting from the spacing and relative body orientation among people; and (b) the distance and relative orientation between the devices themselves. We believe that recognizing and respecting the *proxemics of devices* as a related yet nonetheless separate construct from the *proxemics of people* is an important conceptual distinction contributed by our work.
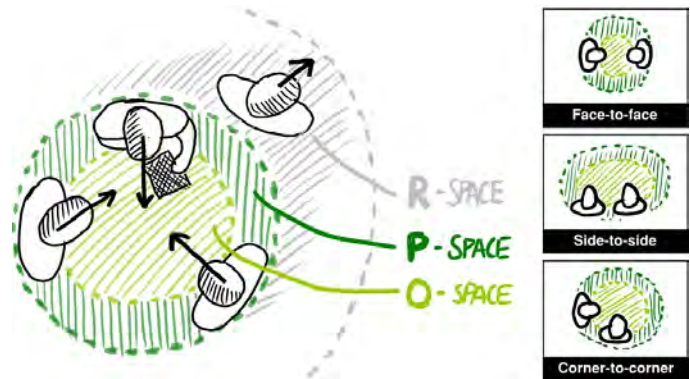

Fig. 1. An F-formation consists of two or more persons engaged in joint activity. Their bodies define three, roughly circular, regions: the inner o-space, the ring of p-space, and the surrounding r-space.

Our paper contributes the following. Using the theoretical lenses of F-formations and micro-mobility (detailed in depth in the following sections), and further informed by an observational study we conducted of people working together with foam-core mock-ups of mobile displays, we designed a number of cross-device interaction techniques that support nuanced gradations of sharing, from the subtle to the overt, with the goal of minimizing the transaction costs—and social disruption—of sharing information across a small-group ecology of digital devices and situated displays. The techniques are unique in that they consider device-to-device relationships as well as the context of the social group to mediate interaction. Our system realizes this through a hybrid of on-device sensors (motion sensors and 8GHz band radios) as well as extrinsic sensors in the environment (depth cameras and fixed-location radios).

The remaining sections of this paper first delve further into the theory of proxemics, F-formations, and micro-mobility that motivate our work, and discuss additional related work. Next, we present an observational study which informs the behaviors that serve as the building blocks of our interaction techniques. We then detail the interaction techniques and sensing system implementation. We close with an informal evaluation of the techniques as well as a discussion of salient issues raised by this research.

BACKGROUND THEORY AND MOTIVATION

Our research leverages insights from three sociological concepts: proxemics, F-formations, and micro-mobility.

Proxemics

The sociological field of proxemics [4,10] studies people's use of personal space to mediate social interactions. While proxemics has many aspects (e.g. nonverbal cues, body orientation, and cultural variation), the defining characteristic is that people equate interpersonal physical distance with social distance. Hall [10], for example, describes social zones based on physical distance between people, ranging from *intimate* (0-50cm), *personal* (1m), *social* (4m), and *public* (>4m) distances. In the context of ubiquitous computing, the implication is that proximity becomes, in part, an estimation of people's desire to communicate with one another via the devices they carry.

The Focused Encounter: F-formations

Distance in and of itself is not a complete description of the social connectedness of co-located persons. Whereas Hall's notion of proxemics primarily concerns the impact of *distance* on the perceptions available to the organism—and hence the types of communications afforded—the study of *F-formations* further considers the *physical arrangements* that people adopt when they engage in focused conversational encounters. Specifically, F-formations [5,17] are a macro-level theoretical lens through which one observes small-group interactions. F-formations consider the spatial relationships that occur as people arrange themselves during face-to-face interaction for optimal task-dependent communication and clarity of perception (Fig. 1). A typical arrangement is a roughly circular cluster that contains 2-5 persons who are actively part of the group. The inner space of that circle (called *o-space*) is reserved for the main activity the group is pursuing. The ring of space occupied by the people (*p-space*) determines group membership. The surrounding region (*r-space*) buffers the group from the outside world. Thus persons who are nearby but not in p-space are excluded from the fine-grained social circle that defines the F-formation. Still, the group monitors r-space to see others who may be trying to join. For example, an approaching person in r-space may be greeted via eye contact, while a person who is facing away, even if close to the group, is not treated as a potential member.

F-formations are nuanced. They need not be circular. Different relative body orientations— face-to-face, side-by-side (as in Fig. 1), or corner-to-corner (L-shaped)— afford different types of collaborative tasks: competitive, collaborative, or communicative, respectively [26]. Group size varies, but tends to be small. Freely forming groupings rarely surpass 5 persons; 95% are 4 persons or less, and more than half are dyadic [7]. Gestures made or objects held within the o-space become the focus of conversation, whereas objects held down (in p-space) or outside the circle (in r-space) are excluded [5]. Fixed or semi-fixed features of a room, such as walls and furniture, can also influence the structure of F-formations [5]. As we will see, these

observations—as further reinforced by our own behavioral study—inform the sensing requirements and desired transactional properties of our interaction techniques.

Micro-Mobility

The *micro-mobility* of physical artifacts is the fine-grained orientation and repositioning of objects so that they may be fully viewed, partially viewed, or concealed from other persons. For example, Luff & Heath [20] reported on patient-doctor consultations via paper records and observed how individuals rely on the micro-mobility of these physical artifacts to facilitate communication. The doctor might gesture at a record, orient it such that it invites the patient to view material, or pull it back to give the doctor time to read information. These subtle manipulations of the paper record afford the shifting demands of an activity.

Thus micro-mobility posits that orienting an artifact towards other persons, moving it closer to them, or even subtly tilting it towards or away from that person affords powerful and nuanced ways for individuals to share (or not share) information, as well as to fluidly manage the focus of conversation and make their intentions clear to others.

Using Theory to Motivate Group Interaction Techniques

Our belief is that ubiquitous computing environments can sense social proximity of people in the form of F-formation assemblages as well as the micro-mobility of the physical devices used by the group. In particular, we sense F-formations of both the people and the devices they carry to automatically federate devices and mediate cross-device interaction in a way that corresponds to real-time social groupings. Physically nearby people in the r-space, especially those facing away, are recognized as socially distant from the current assemblage of users and therefore can be excluded from the exchange. If people within the F-formation then orient their devices towards (say) the central o-space (i.e., using micro-mobility), the system can offer lightweight ways to share device contents across the group.

ADDITIONAL RELATED WORK

Previous efforts have explored techniques for mediating cross-device interaction, sometimes informed by proxemics [10] or a more general notion of proximity sensing, but past work rarely applies the multiple lenses of proxemics, F-formations, and micro-mobility in unison.

Sensing Presence, Proximity, and F-Formations

Various early systems sense and take action based on the distance between people and a device. However, what often goes unrecognized is that the fidelity of sensing (*coarse-grained* vs. *fine-grained* detection of physical distance) as well as the sensing approach itself (*intrinsic* sensors on the device or worn by the user, vs. *extrinsic* sensors mounted in the environment) entail many tradeoffs that speak to the nature and type of interactions that can be supported.

For example, sensing mere presence represents one very coarse-grained type of ubicomp proxemics, with correspondingly limited interactive capabilities, such as turning on the lights when a user enters a room [4]. If a

rough measure of distance is available, more fine-grained interaction is afforded by sensing proximity in several discrete zones (akin to Hall's proxemic zones [10]). For example, Vogel [28] describes an ambient wall display that exhibits responses ranging from public to semi-private interaction depending on what zone a person is in, a theme further explored by Ju [16]. Systems may also provide proximity-ordered lists of nearby resources [25,30].

Greenberg [9] argues that computing systems—which are carried about as well as embedded in the world around us—could be enhanced by sensing proxemic dimensions such as distance, orientation, movement, identity, and location. Ballendat [1], for example, describes a proxemic media player that responds to movement, proximity, and relative body orientation of multiple persons. Other examples include games, advertising displays, and presentation systems based on nuanced notions of proximity [9,21,29]. Such systems typically leverage environments augmented with sensing technologies [1,3,28,31], though approaches such as orienting proximate mobile devices using their back-facing cameras have also been explored [6].

In computing, F-formations have been used to analyze social interactions in crowded environments [14,22], but they have rarely been applied to interaction technique design. Models have been proposed for federating sets of cooperating devices, such as Rodden's awareness model for groupware applications [24], or the boundary principle [18] articulated by Kindberg & Fox. F-formations offer a lens grounded in sociological theory to enrich these approaches.

Micro-mobility for Tabletops and Other Interactions

Luff and Heath [20] introduced micro-mobility and its implications for the design of groupware systems mostly as a cautionary tale, i.e., that new technologies may disrupt the natural movement of artefacts necessary for effective communication. Micro-mobility has since been applied to other areas, particularly the spacing and orientation [8,19] of digital objects on tabletop systems to denote personal territoriality as well as groupings of objects. Other related techniques include adjusting the posture of dual-screen devices [13], bumping and pouring [12], stitching [23], and "chucking" content from one device to another [11].

Implications

Our work differs from this prior work in several ways. First, we apply the above concepts in unison. By explicitly incorporating sensing mechanisms for both interpersonal proxemics (via F-formations) as well as device-to-device orientation and identification (via micro-mobility), our system embraces these approaches in a hybrid design.

Second, our focus is on cross-device interaction methods, and in particular how to decide when devices should be federated (by sensing F-formations) and how that information should be shared (by sensing micro-mobility).

Third, the sociological study of proxemics and F-formations has primarily considered interpersonal distance.

As such, this literature rarely concerns itself with "object territoriality" – for our purposes, that is, the *proxemics of devices* such as slates or mobiles. Our work, including our formative study, seeks to explicitly address this gap.

DESIGN STUDY: PROXEMICS OF PEOPLE & DEVICES

We conducted an exploratory design study to gain insight into: (a) how people vary proxemic variables, such as distance and orientation, while interacting with one another; and (b) the less well-studied proxemic relationships of device to people, and device to device. Our goal was to inform the behaviors we should look for with our sensing system, as well as to provide fodder for naturally occurring gestures and device movements that we might leverage—or need to disregard—for our interaction techniques.

Participants performed joint activities, consisting of both cooperative and competitive tasks, using handheld displays of different sizes (slate, reader, and phone; 30.5 x 21 cm, 15 x 21cm, and 6 x 12cm, respectively). Tasks included *collaborative* (discuss and compare graphics divided across displays), *competitive* (find locations on a map as quickly as possible), and *co-present but individual* work: memorize a list of items, and then enumerate them while the other participant checks the list. We also included scenarios where participants read pre-printed "private" emails.
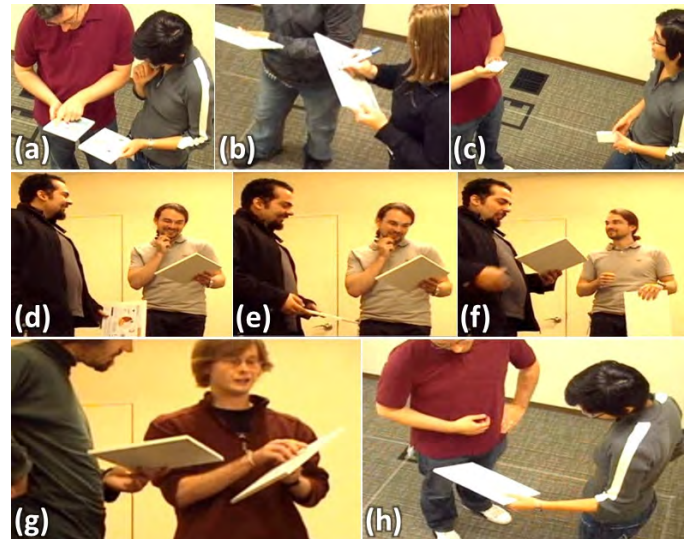
Fig. 2. Behaviors from study. Use of **(a)** side-by-side, **(b)** L-shaped, and **(c)** face-to-face formations depends on task. Devices shift in and out **(d,e,f)** of the shared o-space as users attend to them. Tilting **(g)** while pointing to an item on the display. Moving a display into o-space **(h)** and orienting to a compromise viewpoint.

We videotaped ten paired participants (2 pairs male-female, 3 pairs male-male) as they worked with 1cm thick foam-core mock-ups of slates, readers, and phones fitted with paper "displays." We used foam-core models so that we could focus on the human behaviors of micro-mobility and F-formations, rather than limitations induced by the affordances (weight, screen glare, etc.) of current devices.

From these sessions we observed the following behaviors (B1-B8) that illustrate interpersonal postures, movements, and gestures that people naturally exhibit in such situations:

B1. Device as extension of person. Participants treated the devices as part of their person (Fig. 2a). They were reluctant to bring displays into direct contact or to touch one another's displays (though they sometimes did so, albeit briefly). Users clearly exhibited a notion of "personal space" surrounding their devices, but the socially acceptable *device-space* was compressed as compared to normal interpersonal distances.

B2. Distance and shape of F-formations vary by task. In the presence of hand-held devices, the task influenced the choice of formation, reinforcing related findings from the proxemics literature [26]. For collaborative tasks that required synthesis of information across displays, users moved close together and adopted side-by-side formations (Fig. 2a). For parallel individual work, users moved apart, often in L-shaped formations (Fig. 2b). For competitive or private tasks, users often moved face-to-face (Fig. 2c), as well as further apart.

B3. Movement of devices in and out of focal zone. Users extended displays into the o-space defined by their joint gaze when discussing content. Displays not the topic of communication moved into p-space at the user's side, with the display facing inward (Fig. 2d,e,f).

B4. Incidental tilting. As long as a display was not oriented towards the other person, users seemed largely oblivious to the particular tilt angles of their device. That is, any particular tilt angle is not a definitive cue in and of itself, but rather has to be interpreted in the context of the F-formation (e.g. Fig. 2d, e).

B5. Pointing while tilting within the o-space. People often pointed to an item while tilting the display towards the other person (Fig. 2g). That is, people often discuss *specific pieces of content* rather than the entire display.

B6. Reorientation with gradation in response. As with micro-mobility [20], we observed subtle tilting and reorientation to nonverbally indicate when people wanted to say something about their display. Larger, more overt movement of the display to a compromise viewpoint was a powerful cue to redirect the other person's attention (Fig. 2h). That is, people employed a gradation in response, from the subtle to the overt, to suit the current communicative need.

B7. Avoid persistent spatial invasion. When users gestured at each other's display, such intrusions always remained brief, typically just a second or two. Users often want to indicate referents on the other's display, yet doing so may be uncomfortable; this suggests cross-device interactions should avoid direct interaction with another user's display [23].

B8. Matching pose while side-by-side. When actively working on a joint task, users tended to mirror the tilt angle of one another's slates (e.g. Fig. 2a), suggesting that equal tilt angles offer a context where it might be useful to reduce the barriers to sharing across devices.

Our intent is not to mimic B1-B8 in our design *per se*, but rather to use them as building-blocks, or points of departure. Furthermore, some of the above points—while perhaps obvious in retrospect—did not strike us as important observations until after we had completed our initial implementation and usability study. These we will return to later as we discuss the strengths, weaknesses, and possible extensions to our techniques.

THE GROUPTOGETHER SYSTEM

The GroupTogether system explores these concepts in action. We implemented a prototype "informal information workspace" where users can freely arrange collected content (photos, sketches, clippings, etc.) on an interactive canvas. We implemented the application for both slates and digital whiteboards, and each instance of the application may connect to others nearby, as mediated by our sensing of F-formations and device micro-mobility. We chose this application context because collecting and loosely organizing content is a common need of information workers, such as during active reading (e.g. [13,27]).

In the following sections we describe four main interaction techniques supported by GroupTogether, focusing initially on the micro-mobility aspects. We then describe how we extend these techniques to federated devices held by users in F-formations consisting of two or more persons. We also consider the case where the F-formation encompasses a digital whiteboard. We then describe our implementation.

INTERACTION TECHNIQUES

We developed four interaction techniques facilitating the sharing of digital information between the devices of people standing in an F-formation. These techniques are directly inspired by behaviors B1-B8 noted above, and as we discuss each technique we will refer back to these behaviors to reinforce our rationale, design considerations, and behavioral constraints. Note, however, that B1 and B2 are more meta-observations underlying all our techniques: they validate our preconception that there exists a proxemics of devices (B1), reinforce our design approach (devices must be proximal, but do not have to touch; again B1), and furthermore that a variety of F-formation structures must be properly sensed and supported in a consistent manner by the interaction techniques (B2). Furthermore, while in the following descriptions of specific techniques we primarily focus on micro-mobility gestures, keep in mind that these gestures are only active when the user is currently sensed as standing in an F-formation.

Our interaction techniques facilitate transient sharing, copying, transfer, and reference to digital information across federated devices. These actions suit our application context and allow us to explore various possibilities in order to generate design insights. In particular, the system offers multiple ways to support co-located collaborative activity, ranging from the subtle to the more overt, with various and nuanced semantics of what it means to "share" content. For clarity, in the following we refer to a two-user F-formation involving handheld devices: the user initiating the interaction is the *sender*; and the other person is the

recipient. We later describe how our techniques work with more than two people, and also with a large display.

Tilt-to-Preview Selected Content

The *Tilt-to-Preview* technique provides an extremely lightweight way to *transiently share* selected digital content across devices. The receiving user can then elect to *grab a copy* of the transiently shared information.

Following the example of behavior B5, *Pointing while tilting within the o-space*, the sender shares a selected piece of content by holding his finger on said content while simultaneously tipping the slate slightly (by a minimum of 10 degrees, Fig. 3). This gesture is only active when the tablet is held up within o-space; when triggered it causes a transient semi-transparent representation of the selected item to appear on the display of all devices in the current F-formation. To make it easy for recipients to identify who is offering an item, an animation slides-in the shared item from the side of the screen where the sender is standing.
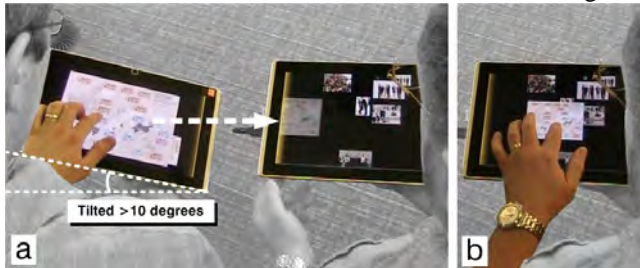

Fig. 3. Tilting tablet and touching content to transfer temporary copy (a), touch copy on second tablet to keep permanent copy (b).

We employ a tilt threshold of 10 degrees: during pilot testing we found this angle well beyond the *incidental tilting* (B4) that users typically exhibit while holding a slate. Tipping a slate beyond this threshold serves both as a gesture to trigger the behavior as well as a social cue observable to the recipient (and any other people nearby) that the sender wishes to share something. The gesture therefore also leverages observation B6, *Reorientation with gradation in response*, with a fairly subtle overture. Note also that the recipient can ignore such an overture merely by leaving his tablet down, in p-space (as in B3, *Movement in/out of focal zone*), or accept it by holding the tablet up.

When the sender lets go of the transiently shared content, it disappears from the recipient's screen. However, the recipient can choose to keep a *copy* of the transiently shared content by touching a finger down and grabbing it while it remains visible. This precludes any need for either user to reach onto the other's display, in accordance with observation B7 (*Avoid persistent spatial invasion*).

In dyadic F-formations the sender just tips the slate towards the other user. Alternatively, or if there are more than two people in the F-formation, the user can also tip the slate towards the o-space (i.e., the center of the formation).

Face-to-Mirror the Full Screen

As noted in B6, *Reorientation with gradation in response*, we observed both subtle and overt tipping gestures in the course of our design study. As we observed there, a user can employ a larger tilt as a more demanding nonverbal request to interrupt the current thread of conversation and introduce something else. We also noticed that users often employed larger tilts to show content to their more distant partner in face-to-face formations (B2, but not pictured).

Face-to-Mirror explores how we might provide digital affordances based on these observed behaviors. Using this technique, a person can share the full screen view of the primary digital content displayed on their screen to the other tablets in the social group of an F-formation.
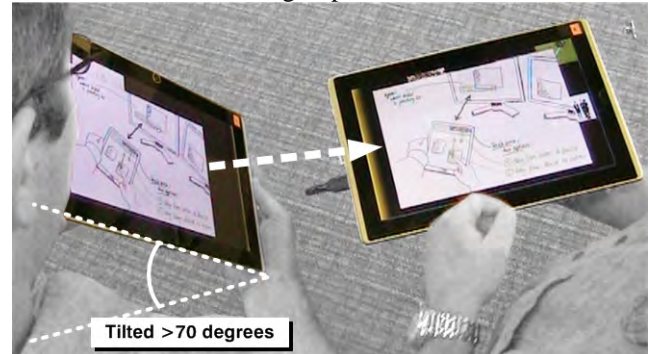

Fig. 4. Holding tablet vertically (angle larger than 70 degrees) shows a full screen copy on the other tablet.

When a person holds their tablet vertically (at an angle larger than 70 degrees), the interactive canvas is mirrored, at full-screen scale, to the display of all other tablets of the group (Fig. 4). Note that unlike Tilt-to-Preview, this is a pure tilting gesture; the user does not have to touch the screen to explicitly select content. Thus, while the tilting motion is larger, the transaction cost of sharing is potentially lower because the required action is simply "show your screen to the others." Again, the tilting motion is large enough that incidental tilting (B4) is not an issue with this technique, and furthermore we only observed the *Tilting while pointing* behavior (B5) in the context of more subtle tilting motions, so requiring pointing during Face-to-Mirror would run counter to our design study observations.

As with Tilt-To-Preview, Face-to-Mirror begins as a transient sharing technique that ends when the sender moves his slate away from the vertical posture, but where recipients can retain a copy by grabbing the mirrored item.

Portals

The above two techniques both share either a transient representation of an item, or a permanent copy if the recipient touches down and grabs it.

To explore an alternative semantic of *transferring* content from one device to another (that is, moving rather than copying content), we implemented the Portals technique, which builds on the notion of dragging items across federated devices (as explored in previous systems such as ARIS [2] and PointRight[15]). The difference here is the sensing approach to federate the devices in the first place, plus tilting as a micro-mobility gesture to establish a Portal.

When tilting a tablet (more than 10 degrees) towards the device of any other group member, a tinted edge appears along the shared screen edge of the two slates (Fig. 5a). By dragging an item through this edge and releasing the touch, the item is (permanently) transferred to the other device (Fig. 5b). A continuous cross-display animation reinforces the metaphor of the gesture: the content slides off the sender's screen, and slides into the recipient's screen. The recipient can then drag, resize, and otherwise manipulate the content that was transferred to their tablet (Fig. 5c). As with Tilt-to-Preview, the recipient will only receive items sent through a Portal if his tablet is held up in o-space (as opposed to moving it down to p-space), as observed in B3.
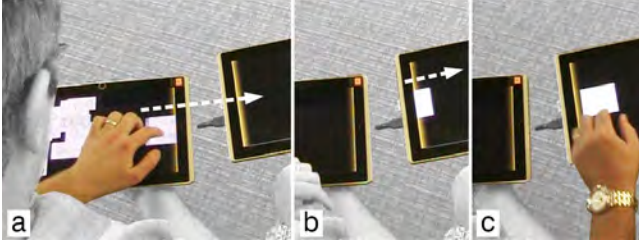

Fig. 5. Moving content from one slate to another: **(a)** dragging content through the tinted edge of screen,**(b)** item moves onto the other slate, and **(c)** the recipient repositions the item.

Note that in one sense the gesture for Portals is a hybrid of Tilt-to-Preview and Face-to-Mirror: the user performs a fairly subtle (>10 degree) tilting motion (like Tilt-to-Preview) to create the portal, but does not have to touch the screen while doing so. On the one hand this means that Portals may be more prone to incidental tilting (B4). On the other hand, the feedback for Portals (a visually unobtrusive tinting along the matching edge of the devices) as well as the semantics of using the Portal (a transfer only occurs if the user explicitly passes an item through the shared edge of the Portal) means that there is very little impact if accidental activation of a Portal does occur.

In our evaluation (see below) users liked the Portals technique but found our distinction between *copy* and *transfer* confusing. We are therefore currently considering alternate mappings that treat all sharing as a copy, and thus free up tilting (any amount) for full-screen sharing.

Cross-Device Pinch-to-Zoom
Cross-Device Pinch-to-Zoom was motivated in part by B8, *Matching pose while side-by-side*. Here, we explore ways that users can explicitly share items when the slates are *not* tilted (e.g. as shown in Fig. 2a), but just held together side-by-side in o-space (B3 and B7) and at the same relatively flat angle (B8).

This technique allows viewing content across multiple tablet devices when using a pinch-to-zoom gesture. As typical of freeform canvas interfaces, a person can use a two finger pinch gesture to enlarge any content on the screen (Fig. 6a). But since our technique leverages the GroupTogether system's knowledge of F-formations and the pose of nearby devices, when the sender enlarges the zoomed content beyond the visible area of the slate's

display, the remaining content expands onto the surrounding tablets in o-space (Fig. 6b). That is, while the person zooms in, the content is displayed on the combined screen area of the tablets that form a particular f-formation (i.e., a larger cross-device virtual canvas is created).
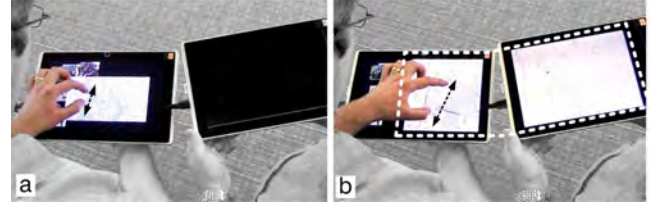

Fig. 6. Cross-Device Pinch-to-Zoom: **(a)** the sender begins pinch-to-zoom on the first slate; **(b)** the zoomed content is displayed on surrounding slates that are part of the F-formation.

Propagation through F-formations
While the above interactions illustrate how our ideas apply to two-person F-formations, the techniques also apply to larger groups (Fig. 7a). For *Tilt-to-Preview* and *Face-to-Mirror*, for example, a person can share content with the entire group by tilting their tablet towards the center of the formation (i.e., towards o-space) rather than just tilting towards a single person.

Furthermore, we implement the techniques described above for all three types of F-formations (side-by-side, face-to-face, and corner-to-corner). While it would be possible to support assemblage-specific gestures, we felt that this might not be intuitive to users. However, we do adapt the feedback on the screen (e.g. placement of the tinting indicating an active Portal) to match the spatial arrangement of users.

Likewise, users who are sensed as external to the F-formation cannot participate in group interactions, unless of course they move to stand within the group. While it might be interesting to explore various techniques for beyond-arms-reach interaction [3]—for example, to allow distant persons to send items to an F-formation, or to the digital whiteboard— by the same token in the context of GroupTogether this would go against the simplicity of the natural human behaviors that we observed in our design study, so we chose not to pursue such mechanisms here.

While we have fully implemented F-formation tracking and propagations, and performed pilot tests with groups of 3-4 persons, we did not include groups of this size in our informal user studies (see below). As such, while we believe propagation micro-mobility gestures through F-formations of more than two persons represents an important concept, as of this writing this aspect of our system has not yet been fully explored and evaluated.

A Digital Whiteboard as Part of an F-formation
Because F-formations can implicitly encompass fixed or semi-fixed features of the environment [5], we included a large screen digital whiteboard in our system as an exemplar. Users within a sensed F-formation can share content with the digital whiteboard in a manner analogous to sharing content to slates held by other participants. For

example, consider the *Hold-to-Mirror* technique. A person can hold their tablet vertically towards the large display, and a temporary copy of the tablet's content appears on the large screen (Fig. 7b). Similarly, a person standing next to the whiteboard can use the Portals technique to move content to the large display by dragging content onto the edge of the slate facing towards the whiteboard.
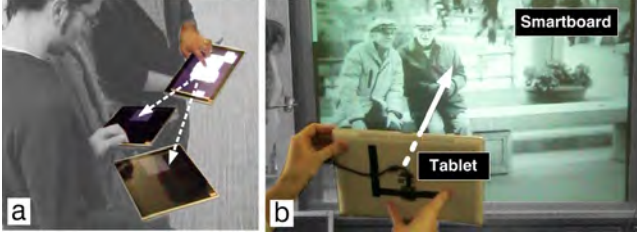

Fig. 7. (a) Shared content propagates to devices of all members of the current F-formation. **(b)** Using the Face-to-Mirror technique with a large digital whiteboard display sensed as part of an F-formation.

Our implementation considers the digital whiteboard in a manner similar to the human participants in an F-formation; that is, it has a particular location and a facing vector. When it falls along the perimeter of p-space it is treated as part of the F-formation, but if it falls outside the huddle, in r-space, it is not. For example, if a circle of users stands in front of the digital whiteboard, with one user's back to the display, and another user performs a Face-to-Mirror gesture, the content will be mirrored to the F-formation but not to the whiteboard. But if the same circle expands to encompass the whiteboard, then the content would be sent to the whiteboard as well.

IMPLEMENTATION
We now describe the major system components of GroupTogether, including the hardware and sensors, how we employ 8GHz band radios to associate devices to people, and how we process the raw Kinect depth data to sense F-formations.

Hardware System Components
We implemented all techniques on Asus EP121 slates (1.16 kg, 312 x 207 x 17 mm), with two-point capacitive multitouch screens running Windows 7. To the back of each slate, we attached a Phidgets Spatial 3/3/3 motion sensor and a Qualcomm Short Range Communication Technology ("QSRCT") 8GHz band radio module.

We mounted a pair of Kinect depth cameras in the ceiling above our prototyping environment. This enabled us to sense users in a roughly L-shaped active area (Fig. 8). The wall display was a fifty inch diagonal SmartBoard with single-point optically-sensed touch. We mounted fixed QSRCT radio modules on either side of the SmartBoard to enable triangulation of the slate-mounted radio modules. This enabled identification of specific devices (by their radio ID) as well as coarser-grained proximity sensing of devices that were nearby, but not held by a user within the viewing area of the Kinect cameras.

System Architecture
All the GroupTogether system components were connected via TCP over wireless Ethernet. We implemented a message protocol using Windows Communication Framework (WCF) to transmit all application and sensor state. A central server maintained a global model of all spatial relationships, and notified clients of state changes.
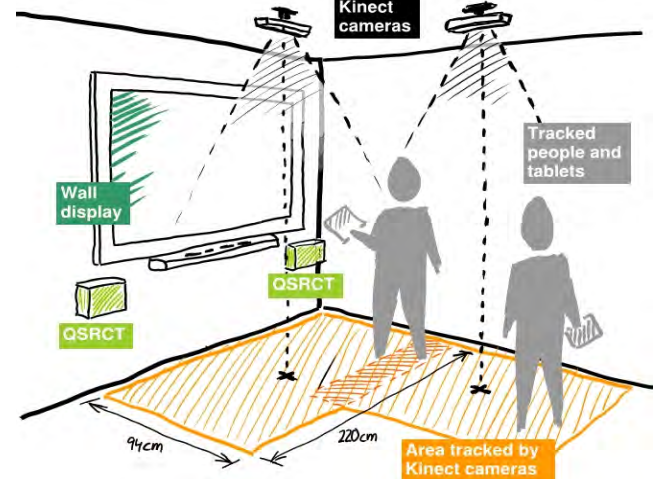

Fig. 8. Schematic of prototype environment with two overlapping Kinect cameras, and wall display with 8GHz band QSRCT radios (a radio was also attached to the back of the tablets, not shown).

Associating Devices to People via 8GHz Band Radios
The overhead Kinect cameras allow us to track moving blobs that we recognize as people, but this tells us nothing about the *devices* that they carry. We employ the QSRCT radios via *wireless radio signal trilateration* to associate devices to sensed persons.

The Qualcomm Short Range Communication Technology (QSRCT) radio modules transmit signals in the 8 GHz band, which avoids interference with the popular 2.4 GHz WLAN band, but also restricts QSRCT to line-of-sight communication. An advantage of this property is that the signals tend to stay within the social boundaries of meeting spaces delineated by walls, partitions, furniture, etc. The QSRCT radios have a maximum range of 15m and also *sense approximate distance* between radio modules (10 cm accuracy at 90% confidence). We employ *three point location trilateration* where we put three QSRCT radio nodes at fixed locations in the space around the edges of the area tracked by the depth cameras. The mobile radio sends range-finding requests to each fixed node and we Kalman-filter the resulting measurements to reduce noise. The device location is then the intersection of the three circles.

The server matches the device to a person by assigning the device to the person who is standing closest (as sensed by the Kinect cameras) to the triangulated location. If the distance between the device and all currently tracked people is above a tolerance threshold (currently 1m), the device remains unassigned. To assign all other devices, the process is repeated with the remaining QSRCT radios attached to the tablets.

Trilateration with the QSRCT radios—and matching these radio ID's to the devices held by people tracked via the Kinect depth cameras—allows us to not only associate each device to a particular person, but also to identify a person based on the device they carry (under the assumption that each person carries their own device). Furthermore, when people come and go from the field-of-view of the Kinect camera, the QSRCT radios enable us to keep track of *who* is entering or leaving, as well as to maintain some awareness of whether the person has walked away, or still remains nearby but out of depth-camera view.

KINECT TRACKING OF F-FORMATIONS

We use overhead Kinect depth cameras, mounted in the ceiling and looking downwards, to track spatial relationships between people, including the distance, viewing direction, and relative body orientation between persons. Our system then aggregates these sensed persons into F-formations and classifies the type of assemblage (that is, face-to-face, side-by-side, corner-to-corner, or as singletons not a member of any formation). We note that our implementation assumes users are *standing*, and does not yet handle seated participants.

The Kinect field-of-view of 58.5 degrees, with the cameras mounted 200cm above the floor, produces a 220x94cm tracking area per camera. We arrange the Kinects in an L-shaped sensing region (Fig. 8) and, with a one-time calibration step, compose the data into a combined depth image. For each depth image we calculate the orthographic projection of the camera view, and then use filters and linear interpolation to remove noise resulting from the partial overlap of the Kinects' structured light patterns.
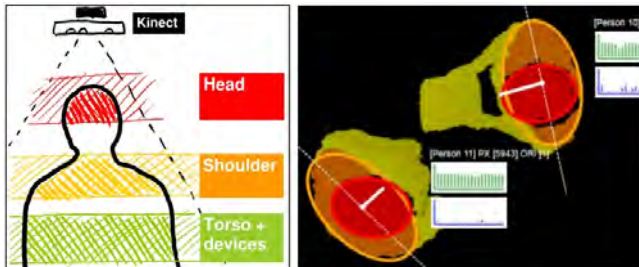

Fig. 9. Standing person as perceived by overhead Kinect depth cameras, with schematic (left) and raw depth image (right) labeled with recognized head, shoulders, and torso/arms. Note the facing vector (white line segments) and principal body axis (dotted lines).

Processing Pipeline for F-Formation Detection

We process the combined depth image in multiple steps to identify the location, orientation, and formations of people.

1. Filter. We first filter out all connected components too small or too large to be an actual person, leaving just those components most likely to represent a person.

2. Normalize height. Next, the height of all detected people is normalized to match. The algorithm finds the highest point of all detected people identified in the previous step, and shifts the depth values of all remaining outlines of people by the difference in height.

3. Detect heads. We assume the topmost depth band represents people's heads (Fig. 9). Our algorithm identifies all connected components in this separated 2D image. The results are ellipsoidal outlines of people's heads. We take the major axis of the ellipse as the orientation of the head.

4. Calculate body orientation. The second depth band (Fig. 9) includes all regions belonging to people's shoulders. We assign each detected shoulder region to the person it is closest to. We then take the convex hull (since the shoulder is not necessarily a single connected component) to get an ellipsoidal outline of a person's shoulders. The major axis of that ellipse gives us the orientation of the person's body.

5. Determine which way people face. The orientations calculated in steps 3-4 still have a 180-degree ambiguity. To determine which way the user is facing, we take a third depth band that corresponds to the user's torso, including their arms and hands (Fig. 9). We then identify which side of the major body axis a person's arms and hands appear on, as well as which side the head is shifted towards, and take this as the front. We further apply hysteresis to prevent momentary flips in the facing vector due to noise.

6. Detect F-Formations. Finally, our algorithm identifies F-formations. Two people can be in an F-formation if: (a) they are not standing behind each other, (b) the angle between their orientation vectors is smaller than 180 degrees (otherwise they would face away from each other), and (c) the distance between individuals is small enough so they can comfortably communicate and their o-spaces (Fig. 1) overlap. Our algorithm iterates over all pairs of people, calculates the distance and angle between them, and assigns an F-formation type (i.e., side-by-side, L-shaped, face-to-face, or none) based on tolerance thresholds. Hysteresis prevents frequent switching of detected formations if measurements lie close to a threshold. We also detect and tag singletons and persons in R-space (outside a formation).

We have not formally evaluated the error rates of our F-formation sensing approach; the current system is intended only as a proof-of-concept to illustrate that small groups can be sensed, as well as to suggest some directions as to how such sensing can be leveraged for small-group interaction. We have pilot tested the sensing with various arrangements of 2-4 person groups, including groups with nearby persons outside the formation (in R-space).

Note that in our system, sensed formations lead only to federated groups where the barriers for sharing are lowered: we do not necessarily assume that people in a group want to share. Indeed, this is precisely why we pursued a hybrid approach (of sensed F-formations + device micro-mobility) to initiate the actual cross-device sharing of content. We believe this approach appropriate for contexts such as small-group meeting spaces, where people have freedom of movement, but it might not be suitable for crowded environments such as a conference hall or subway station.

INFORMAL EVALUATION & REACTIONS FROM USERS

We conducted an informal evaluation to gather feedback of people using our techniques in practice. We recruited 6 participants for our experiments (all male, age 29-50, part of the same organization but not members of our group) paired into groups of two. Each was given a slate running our software. The experimenter explained the four main cross-device interaction techniques. We asked participants to use the technique for basic information-sharing and viewing tasks. They then responded to several 7-point Likert scale questions, discussed what the best and worst thing was about each technique. and ranked them for overall preference.

Results

We observed that participants were quickly able to perform each of the four techniques, and that they generally found them intuitive, quick, and easy to perform. Likert responses confirmed their overall positive feedback regarding ease-of-use, where they particularly liked the notion of lightweight sharing across devices. They ranked *Portals* as their favorite technique (and also gave it the highest average rating on a "comfortable to perform" Likert scale question), followed by *Tilt-to-Preview*, *Cross-Device Pinch-to-Zoom*, and *Face-to-Mirror*. However, a number concerns were raised by our participants or from our own observations, as summarized below.

Physical effort and tilting. Participants felt comfortable performing the smaller tilting gestures, such as that used for *Tilt-to-Preview*. They usually quickly moved the tablet back to the horizontal position after performing one of the sharing interactions. Yet the size, weight, and fragility of the sensor-augmented slates were an issue, as some participants found it difficult to hold the tablet in one hand while touching an item with the other hand. However, this points more to the limitations of these particular tablets and attached sensors than any particular failing of the micro-mobility idea: recall that people had no problem with tis using the lighter foam-core models during the design study.

Participants commented on the "intuitive" use of *Face-to-Mirror*: "this is like... I show it to you" or that there was "even less work to do" than the other techniques (i.e., just hold the slate vertically, with no touch required). Yet participants did raise concerns about the large tilting gesture (>70 degrees) required for *Face-to-Mirror*, particularly for side-by-side interaction. One user stated "you have to tilt too much" and another reported that the technique "required extreme tilt." Observations of people using the technique also revealed difficulties repeatedly holding the slate in the vertical position; users often changed grips to avoid fatigue and maintain a secure grip on the slate with both hands. This issue can be mitigated simply by relaxing the required angle.

Copy vs. move. The semantics of *copy* vs. *transfer* actions adopted by our techniques sometimes caused confusion for participants ("not sure when it copies vs. moves the image"). Our feedback did not make it sufficiently clear which type of "sharing" a particular gesture enabled.

Ownership, control, and handover. For all but the *Cross-Device Pinch-to-Zoom* technique (for the reasons described above), participants felt in control of what content was shared, transferred and kept between devices and the minimal effort required to do so. Participants commented positively: "it felt natural to either 'take it' or 'leave it'," or that the "receiving user decides to keep [it]." However, a few difficulties arose when users attempted to simultaneously share items (such as by dragging through Portals): "when both of us were trying to transfer images at the same time, it became confusing / difficult."

Structure of F-formations. Most participants agreed that the physical distance between themselves and the second person felt appropriate to enable sharing of data. Participants mostly stood in a side-by-side formation during the experiment, usually faced slightly inwards towards the o-space. Users also clearly expected the techniques to extend to more than two users ("Will this work in groups?"). As noted above, groups of more than two people were implemented in our system and tested in pilot sessions, but not were not included in this evaluation. Therefore the suitability of our techniques for larger groups of people remains to be fully demonstrated.

Overall preference. Participants ranked *Portals* as their favorite technique (and also gave it the highest average rating on a "comfortable to perform" Likert scale question), followed by *Tilt-to-Preview*, *Cross-Device Pinch-to-Zoom*, and *Face-to-Mirror*.

In future studies, we would like to evaluate our techniques more systematically for a variety of F-formation types as well as for larger groups of users (3-5 people). It would also be interesting to explore further variations of the techniques, such as *Face-to-Mirror* with a less extreme requirement for the tilt angle.

DISCUSSION

While our system explored a number of novel ideas, it is clear that much more could be done. For example, we explored a pair of techniques—*Tilt-to-Preview* and *Face-to-Mirror*—that looked at two extremes in gradation of response (B6). But from our observational studies, as well as from the commentary of Luff & Heath [20], it is clear that people's use of micro-mobility is more nuanced still. Therefore, it would be interesting to explore similarly nuanced techniques that explore this continuum further. These might include both implicit ways of using device tilt and motion for context sensing, as well as explicit gestures or posturing of devices (as explored in this paper) to support finely delineated notions of sharing content.

The observations in our design study, as well as those resulting from our implementation and informal evaluation, strongly suggest that people treat handheld objects as extensions of their person (B1). Yet by the same token the

sociological rules governing "object territoriality" are clearly not the same as those governing our physical bodies. To our knowledge the notion of object territoriality has never been systematically studied, which suggests the need for further experimental and observational studies to better understand behavioral principles that might inform interaction design for micro-mobility and F-formations.

Several sensing platforms currently exist that we could exploit for sensing the presence and location of people and devices (such as high-fidelity Vicon-based tracking [1,28], the proximity toolkit [21], etc.). In the present research, we deliberately wanted to design a practical, low-cost sensing system to illustrate that such systems could realistically be deployed in the near–rather than the far–future. While the implementation we pursued still requires equipment installation within a room, it could be practically and affordably done within (say) a small dedicated meeting room context such as a breakout room.

More generally, further investigation of lower-cost, coarser-grained, or more easily deployable sensing modalities to detect micro-mobility and F-formations remains important. Exploring the tradeoffs inherent in this liminal zone of imperfect sensing—rather than assuming high-fidelity optical sensing of fully tagged persons and devices, for example—is a worthy area of continuing research.

CONCLUSION AND FUTURE WORK

We have described GroupTogether, a sensing system that jointly brings into play the sociological constructs of micro-mobility and F-formations. We contributed a design study that illustrates how the *proxemics of people* as well as the *proxemics of devices* surface during joint activity. We then demonstrated how the attributes sensed by our system leverage these behaviors to enable a number of novel interactions that employ both the micro-mobility of devices and the social structure of F-formations. We described how the system senses these interactions using a combination of motion sensors, radio modules with coarse-grained range finding capability, and overhead Kinect-based depth camera tracking. And finally, we have presented an informal study illustrating user reactions to the techniques.

We believe we have only scratched the surface of a potentially rich space of such techniques. As devices become thinner and lighter, the opportunities for socially-situated interactions continue to expand. Likewise, systems and techniques that leverage ubiquitous proxemics should benefit from ongoing developments in spatially-aware technologies such as the Kinect depth cameras and QSRCT radio modules utilized in this research. These should be welcome developments for the many contexts where users need to share and discuss digital artifacts while remaining fully engaged in the seamless flow of social exchange with friends and colleagues.

REFERENCES

1. Ballendat, T., et al., Proxemic Interaction: Designing for a Proximity and Orientation-Aware Environment. *ITS '10.*

2. Biehl, J.T., Bailey, B.P. ARIS: An Interface for Application Relocation in an Interactive Space. *Proc. Graphics Interface (GI '04).*

3. Bragdon, A., et al., Code Space: Combining Touch, Devices, and Skeletal Tracking to Support Developer Meetings. *ITS '11.*

4. Buxton, W., Living in augmented reality: Ubiquitous Media and Reactive Environments, in *Video Mediated Communication*, ed. by K. Finn, Sellen, A., & Wilber, S. 1997, Erlbaum. p. 363-384.

5. Ciolek, T.M., Kendon, A., Environment and the Spatial Arrangement of Conversational Encounters. Sociological Inquiry 1980. **50**(3-4): p. 237-271.

6. Dearman, D., Guy, R., Truong, K. Determining the Orientation of Proximate Mobile Devices using their Back Facing Camera. *CHI'12.*

7. Dunbar, R., Duncan, N., Nettle, D., Size and Structure of Freely Forming Conversational Groups. Human Nature, 1995. **6**(1): p. 67-78.

8. Everitt, K., Shen, C., Ryall, K., Forlines, C. MultiSpace: Enabling Electronic Document Micro-mobility in Table-Centric, Multi-Device Environments. *Proc. IEEE TABLETOP '06.*

9. Greenberg, S., et al., Proxemic Interactions: The New Ubicomp? Interactions, 2011. **18**(1): p. 42-50.

10. Hall, E.T., *The Hidden Dimension*. 1966, New York: Doubleday.

11. Hassan, N., Rahman, M., Irani, P., Graham, P. Chucking: A One-Handed Document Sharing Technique. *INTERACT'09.*

12. Hinckley, K. Synchronous Gestures for Multiple Users and Computers. *UIST'03.*

13. Hinckley, K., Dixon, M., Sarin, R., Guimbretiere, F., Balakrishnan, R. Codex: A Dual-Screen Tablet Computer. *CHI'09.*

14. Hung, H., Kröse, B.J.A. Detecting F-formations as dominant sets. *Proc ICMI '11 Conf. on Multimodal Interfaces.*

15. Johanson, B., et al., PointRight: experience with flexible input redirection in interactive workspaces. *UIST'02.*

16. Ju, W., Lee, B.A., Klemmer, S.R. Range: Exploring Implicit Interaction through Electronic Whiteboard Design. *CSCW '08.*

17. Kendon, A., Spacing and orientation in co-present interaction, in *COST'09 Proceedings 2nd Int'l Conf on Development of Multimodal Interfaces: Active Listening and Synchrony*, Springer-Verlag Berlin: Heidelberg. p. 1-15.

18. Kindberg, T., Fox, A., System Software for Ubiquitous Computing. IEEE Pervasive Computing, 2002. **1**(1): p. 70-81.

19. Kruger, R., Carpendale, M., Scott, S.D., Greenberg, S. How People Use Orientation on Tables: Comprehension, Coordination and Communication. *Proc. ACM Group 2003 Conference.*

20. Luff, P., Heath, C. Mobility in collaboration. *CSCW '98.*

21. Marquardt, N., Diaz-Marino, R., Boring, S., Greenberg, S. The Proximity Toolkit: Prototyping Proxemic Interactions in Ubiquitous Computing Ecologies. *UIST '11.*

22. Marshall, P., Rogers, Y., Pantidi, N. Using F-formations to analyse spatial patterns of interaction in physical environments. *CSCW '11.*

23. Ramos, G., et al., Synchronous Gestures in Multi-Display Environments. Human-Comp. Interaction, 2009. **24** (1-2): p. 117-169.

24. Rodden, T. Populating the application: a model of awareness for cooperative applications. *CSCW '96.*

25. Schilit, B., Adams, N., Want, R. Context-Aware Computing Applications. *Proc. IEEE Workshop on Mobile Computing Systems and Applications*. 1994.

26. Sommer, R., *Personal space*. 1969: Prentice-Hall.

27. Tashman, C., Edwards, W.K. LiquidText: A Flexible, Multitouch Environment to Support Active Reading. *CHI'11.*

28. Vogel, D., Balakrishnan, R. Interactive public ambient displays: transitioning from implicit to explicit, public to personal, interaction with multiple users. *UIST'04.*

29. Wang, M., Boring, S., Greenberg, S., Proxemic Peddler: A Public Advertising Display that Captures and Preserves the Attention of a Passerby. *ACM 2012 Int'l Symp. on Pervasive Displays.*

30. Want, R., et al, The active bage location system. ACM Trans. Information Systems, 1992. **10**(1): p. 91-102.

31. Wilson, A.D., Benko, H. Combining multiple depth cameras and projectors for interactions on, above and between surfaces. *UIST'10.*