SHOWING THE POINT:

UNDERSTANDING AND REPRESENTING DEIXIS OVER DISTRIBUTED SURFACES

A Thesis Submitted to the College of

Graduate Studies and Research

In Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy

In the Department of Computer Science

University of Saskatchewan

Saskatoon

By

Aaron Myles Genest

Permission to Use

In presenting this thesis in partial fulfilment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Computer Science
University of Saskatchewan
Saskatoon, Saskatchewan, S7N 5C9

ABSTRACT

Deictic gestures, which often manifest as pointing, are an important part of interpersonal communication over shared artifacts on surfaces, such as a map on a table. However, in computer-supported distributed settings, deictic gestures can be difficult to see and understand. This problem can be solved through visualizing hands and arms above distributed surfaces, but current solutions are computationally and programmatically expensive, rely on a limited understanding of how gestures are executed and used, and remain largely unevaluated with regards to their effectiveness. This dissertation describes a solution to these problems in four parts:

1. Qualitative observational studies, both laboratory-based and in the wild, that lead to a greater understanding of how gestures are made over surfaces and what parts of a gesture are important to represent. In particular, these observations identified the height of a gesture as a characteristic not well-supported in distributed groupware.
2. A description of the design space available for representing gestures and candidate designs for showing the height of a gesture in distributed groupware.
3. Experimental evaluations of embodiments that include the representation of gesture height.
4. A toolkit for facilitating the capture and representation of gestures in distributed groupware.

This work is the first to describe how deictic gestures are made over surfaces and how to visualize these gestures in distributed settings. The KinectArms Toolkit is the first toolkit to allow developers to add rich arm and hand representations to groupware without undue cost or development effort. This work is important because it provides researchers, designers, and developers with new tools for understanding and supporting communication in distributed settings.

ACKNOWLEDGMENTS

My wife, Emily Jenkins, who went before me and helped me through. My wonderful kids, Ian and Oliver, who did neither.

My supervisor, Dr. Carl Gutwin, who taught me more than I could imagine.

Jeff Dyck, who first suggested that I meet Carl. I blame you.

Members of the HCI lab, including David Flatla, Dr. Scott Bateman, Andre Doucette, Adrian Reetz, Nelson Wong, Gregor McEwan, Jared Cechanowicz, Dr. Miguel Nacenta, Dr. Tad Stach (who caused this), and Dr. David Pinelle. The research assistants and developers at the University of Saskatchewan and the University of Calgary: Zenja Ivkovic, Stephen Damm, Mike Kalyn, and Robert Post. I look forward to continuing to work with you all.

My committee members, Dr. Nate Osgood, Dr. Regan Mandryk, Dr. Eric Neufeld, Dr. Scott Bell, Dr. Stacey Scott, and Dr. Mark Keil. Dr. Brett Elkin, Alasdair Veitch, Dr. Susan Kutz, and Dr. Ted Leighton, who provided access into the experiences of wildlife researchers.

The Department of Computer Science; the people of the Sahtu Region, NWT; and the class of 2008 WCVM Wildlife Rotation for participating (sometimes unwittingly) in my studies.

To the many participants and facilitators of the CSCW 2011 Doctoral Colloquium in Hangzhou, China and the ECSCW 2011 Doctoral Colloquium in Aarhus, Denmark.

TABLE OF CONTENTS

page

LIST OF TABLES

Table page

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

When people collaborate they often use gestures as an important part of their communication. Gestures can take many forms, but when people share the same space, they often use a form of gesture called deixis – an indicative gesture that commonly takes the form of pointing – to identify objects, locations, or directions in the space around them [1]. This behaviour is particularly critical for communication when people are collaborating over a shared artifact, such as a map on a table. Map-based collaboration, also called geocollaboration, is a frequent collaborative activity and occurs in a variety of domains, including urban planning, emergency response, foreign and domestic policy, and a variety of research and industrial settings. In most of these settings, the collaboration can be described as surface-based, (i.e., it occurs over a table or whiteboard).

Although collaborations are often collocated, they are increasingly distributed, taking place in multiple locations, at different times, or both. In distributed, surface-based geocollaborations, communication is usually improved through the use of embodiments, which are representations of remote users, their characteristics, and their activities. Embodiments are useful for showing deixis, which often manifests as indicative pointing gestures targeting artifacts on the shared surface. However, current embodiment designs are not effective in expressing deixis during distributed, surface-based collaborations.

Research Problem

Despite the importance of deixis in communication [2], current embodiments fail to adequately represent gestures in distributed settings. In particular, they are two dimensional representations that fail to show the height of gestures, a feature of gestures over surfaces; they frequently do not accommodate for the challenges of distributed settings, such as missed gestures, a problem because of the reduced salience of remote representations of people; and their effectiveness in representing gestures has not been extensively evaluated. In addition, embodiments are difficult to add to groupware (software that supports multiple collaborating users), since there are no tools for easily capturing and showing gestures. This means that distributed collaboration over shared artifacts requires more time, is less expressive, less successful, and more error-prone than collocated collaboration [1] [3] [4] [5]. In some domains, such as emergency management [6] and political negotiation, such problems can mean the difference between success or failure [7].

There are four challenges to creating embodiments that show deixis in distributed collaboration: deixis is complex and poorly understood, especially over surfaces; embodiments are designed without a systematic understanding of the design space for representing deixis; when embodiments represent deixis, they do so in only two dimensions; and there are no easy-to-use solutions for capturing and showing three dimensions of deixis over tables in groupware.

Deixis Over Surfaces is not Understood

There is no comprehensive characterization of deixis over surfaces. Existing characterizations of gestures are limited to restricted domains, such as gestures made by TV weather anchors [8], describe gestures made in other settings, such as storytelling [9] or workplaces [10], or are limited examinations of the gestural space [11]. This is a problem because with no clear characterization, there are no principles available to tell designers which

gestures are most common and which parts of a deictic gesture are most important. Such a characterization would allow designers to make informed decisions about what to support when creating embodiments. It could also inform other questions, such as how to best convey gesture information in environments with limited network, display, or computational resources; or, how to design command gestures so that they do not collide with natural communication gestures.

The Embodiment Design Space is not Described

The embodiment design space for showing deixis over surface-based collaboration has not been described. Design space descriptions allow designers to systematically explore disparate ways of representing information. Without such a description, embodiment designs are more likely to fail because they are not based on a complete understanding of what representations are appropriate for the characteristics of deictic gestures.

Embodiment Designs are in Two Dimensions

Most embodiments represent gestures in two dimensions, despite that gestures move above surfaces in three dimensions. There exists no evaluation of the effectiveness of digital embodiments in expressing gestures above the surface (e.g., they fail to represent gesture-height information). There are three aspects to this problem:

1. Can an embodiment that includes gesture-height information improve the accuracy of the interpretation of gestures made by remote collaborators?
2. Are embodiments with gesture-height visualizations interpreted with the same expressive qualities that the height of gestures conveys in collocated settings?
3. Are embodiments with gesture-height visualizations usable in distributed collaboration: does it permit normal gesturing and is it preferred over un-enhanced alternatives?

Without evaluating these aspects of embodiments, we have no understanding of whether or not embodiment designs are successful in representing three-dimensional gestures over surfaces in distributed settings.

Deixis is Hard to Include in Groupware

There are no light-weight solutions for capturing and showing three-dimensions of gesture movement above surfaces in groupware. Existing solutions are either difficult to set up, interfere with natural gestures with cumbersome sensors, or are extremely bulky to transport, limiting their use to specialized locations such as laboratories or custom-designed communication rooms. Even after these challenges are overcome, there are no toolkits that allow developers to easily incorporate the capture information into visualizations for distributed groupware.

For above-the-surface communication to be usable in distributed settings, three-dimensional gestures must be captured with a minimal setup and specialized equipment. In particular, the capture and representation of gestures should be easy for developers to add to software. Without such a solution for capturing deixis, groupware will continue to use embodiments that are ineffective for showing gestures, making distributed collaborations take more time and be less successful than collocated collaborations.

Motivation

This research is inspired by work underway in the Canadian Arctic by a team of biologists and veterinarians [12]. Their work in capturing and understanding the traditional knowledge and experiences of the aboriginal hunters and trappers is hampered by a lack of accurate logging and tracking during their interviews. These challenges have led to an interest in understanding how gestures are made over surfaces, how they might be tracked in distributed

contexts, and how gestures might be best communicated to remote collaborators – whether distributed geographically or temporally.

Eliciting Traditional Ecological Knowledge Over Maps

Every year since 1998, a team of Canadian biologists and veterinarians from Saskatchewan, Alberta, and the North-West Territories have undertaken a community tour of the Sahtu Region, an isolated part of the Canadian sub-arctic. Part of this tour involves the extraction of traditional ecological knowledge (TEK) from local hunters and trappers. This knowledge is used to inform policy, monitor wildlife and ecosystem health, and establish benchmark knowledge for use in later research.

The process of TEK extraction used by these researchers acknowledges the particular cultural preferences of the people they interview, namely a story-telling setting based on the strong oral tradition of the Dené and Inu people. The researchers ask questions about the land and the hunters and trappers respond by telling stories with reference to a map displayed on a table (see Figure 1). As a result, capturing the information in such a way that it can be later encoded in Geographic Information Systems (GIS) involves a convoluted process. When a community member points at a location on the map, a researcher makes a descriptive mark on an overlay. The marks are (days or weeks later and in another location) associated with video and audio recordings of the session and the resulting interpretation is encoded in a GIS.

The researchers have reported that capturing the gestures of community members and incorporating the gestures into GIS without intermediate steps would lead to better (more accurate and reliable) results and interfere less with the process than the current method. Such a technology, unfortunately, does not exist.

The researchers need a system that transparently (so as not to interfere with the storytelling process) captures deixis over maps. Current systems, however, require extensive set-

up, calibration, and often need highly specialized and expensive equipment. Even if such a capture technology available, we do not know how to represent deixis over surfaces in a way that shows the full, three-dimensional gesture. To do that, we must understand what are the important aspects of deixis over surfaces. This last problem became the starting point for my dissertation work.

Figure 1: Hunter and trapper interview in a Sahtu community. The primary researcher is occluded on the right side. The researcher marking the transparency is occluded by the woman in the red shirt on the right.

Solution

The solution to the problem of showing deixis over surfaces in distributed settings is to create user embodiments that represent the deictic gestures of remote users. There are four steps to this solution:

1. Understand what information should be included or excluded in the expression of deixis by observing and analyzing deictic gestures made during collaborative tasks.
2. Describe the design space for creating embodiments that represent gestures in three dimensions over surfaces.

3. Design and evaluate embodiments based on the knowledge gained in step 1 and 2.
4. Develop a practical, light-weight toolkit for capturing and representing deixis in three dimensions over distributed surfaces.

Understanding Deixis over Surfaces

Chapter 3 describes experimental observations performed to further understanding of how people perform deixis over surfaces. In particular, the observations focused on what parts of their body people use, how those parts are positioned, and how they move. The analysis from these observations provided an understanding of which body parts are important to show in embodiments, why those body parts should be shown, and what details about users and their actions are essential for understanding deixis.

Describing a Design Space for Representing Deixis

Chapter 4 describes how embodiment designs can be classified according to how they represent users. Abstract representations have been shown to be able to express deixis (e.g., [13]), but are generally limited to conveying simple points or paths. Realistic representations can represent hands and arms on the remote table with much greater detail (e.g., with video [14]–[16]), but only in two dimensions, since the embodiments are almost always projected onto the table surface. Hybrid approaches, which use both abstract and realistic visual components (e.g., [17]) may balance the drawbacks and advantages of both approaches, but there are few examples of hybrid embodiments and there has been no exploration of how this design space might best be used.

Since most embodiments are two-dimensional representations of three-dimensional gestures, the height of a gesture above the surface is usually not well-represented. Chapter 4 describes how different visual variables are more or less effective for showing this missing information. In addition, Chapter 4 explains how temporal information about where the gesture

has been in the past can be conveyed through temporal traces and introduces a set of candidate designs for showing three dimensions of gestures with temporal traces.

<u>Evaluate the Design of Embodiments for Representing Deixis</u>

Embodiment designs that are informed by the design space described in Chapter 4 must be evaluated to see if they effectively represent deixis over surfaces. In particular, there are three questions that need to be answered in order to better understand the design of height visualizations:

- *Accuracy*: do the new embodiment designs improve people's ability to determine the type or target of a gesture?
- *Expressiveness*: do the new embodiment designs reliably convey qualities of deixis visible in collocated settings?
- *Usability*: can people make use of the new embodiments in realistic work, and do they prefer these representations?

Chapter 5 describes an evaluation of several of embodiment designs created to improve a single aspect of deixis identified as important in the previous chapters: the height of a gesture above the surface.

<u>A Light-Weight Toolkit for Capturing and Representing Deixis</u>

Many state of the art digital embodiments use specialized and expensive capture technology (e.g., infrared-based motion capture), are able to produce only limited fidelity embodiments using desktop computers (e.g., VideoArms and an enhanced version of VideoArms [17]), and/or are difficult to use in groupware. These are major barriers to the adoption of embodiments that support gestural communication in groupware. Thus, there are three requirements for including embodiments capable of expressing gestures into distributed groupware:

1. high fidelity images using desktop computing power,
2. cheap and easy to use capture technologies, and
3. powerful but simple Application Programming Interfaces (API).

Chapter 6 describes the development and evaluation of a toolkit that meets these three requirements.

Contributions

The research described in this dissertation solves the problem of representing three-dimensional deictic gestures in distributed groupware applications. It does this by making five smaller contributions:

1. It provides the first extensive characterization of deixis over surfaces.
2. It identifies of the height of gestures above a surface as an important aspect of deixis.
3. It describes the design space for embodiments that show deixis over surfaces.
4. It introduces effective designs for showing height above the surface in embodiments.
5. It introduces a novel, light-weight toolkit for capturing and showing rich and understandable gestures over distributed surfaces.

Thesis Outline

This thesis is arranged as follows:

Chapter 2 discusses related work in the fields of geocollaboration and Computer Supported Cooperative Work (CSCW). In particular, the chapter reviews literature on maps, geovisualizations, geocollaboration, embodiments, deixis, representations of height in surface-based applications, and some models of collaboration.

Chapter 3 describes a set of observations about the use of deixis, both in the laboratory and in the wild, during geocollaboration. These observations are analyzed in terms of how body parts are used (their *morphology*), their atomic components of movement, and the way in which

height is a component of the gesture. Chapter 3 concludes with a characterization of deixis over surfaces for the purpose of identifying which features are most important to include in embodiment designs.

Chapter 4 builds on the characterization of deixis by describing a design space for embodiments. The chapter concludes with possible designs for showing the height of gestures.

Chapter 5 explains how the designs described in Chapter 4 were evaluated. The evaluation shows that enhancing embodiments with height and historical information is an effective solution for improving the interpretation of deixis in distributed settings.

Finally, Chapter 6 introduces KinectArms, a toolkit for capturing and representing deictic gestures, including their height. The toolkit contains capture and processing tools, and a set of effects for embodiments that show height, show gesture history, or manipulate embodiment visibility.

Chapter 7 includes a discussion of the work in the previous chapters. In particular, it discusses how the research described in the previous chapters addresses the challenges of understanding deixis over surfaces, describing the embodiment design space, extending embodiment designs into three dimensions, and developing more effective tools for capturing and representing deixis over surfaces.

Chapter 8 provides a summary of the previous chapters, ideas for future research, and a review of the contributions of this dissertation.

LITERATURE REVIEW

This chapter sets the stage for the work that follows: it reviews literature on geocollaboration, introduces current models of collaboration, and compares how the geocollaborative and CSCW fields address collaboration in the design of their systems. Map-based collaboration is the motivation for this research and remains a domain of interest, but is not the primary focus of this research. Beyond geocollaboration, this chapter examines gestural communication and in particular deixis in larger workspaces and in the context of surface computing. Finally, this chapter reviews current solutions for presenting gestures in remote collaboration and some methods of visualizing gestures in collocated collaborations.

Introduction to Collaboration with Maps

Ptolemy argued that maps are useful because they are scaled representations of physical systems [18]. Thus, maps serve as models for systems, facilitating communication and collaboration by providing a common ground [18] and as such, maps have been a critical part of civilization for thousands of years [19]. The recent development of Geographic Information Systems (GIS) permits new ways of creating, displaying, and editing maps [20], opening new avenues for map-based communication and collaboration. However, although GIS can be used to create complex and valuable geovisualizations (visualizations of geographical information), they have not matured as collaborative tools [21]. In particular, tools designed to support collaborative geovisualizations (also known as *geocollaborative* systems) fail to effectively represent the presence and movement of remote collaborators [22]. This is unsurprising, since the study of embodiments, the most common solution to representing remote users, is an ongoing topic of research in the field of CSCW (e.g., [23], [24]).

A more extensive review of the literature related to collaboration and geocollaboration can be found in Appendix A.

Supporting Gestural Communication in Remote Collaboration

One key element of geocollaboration, as in other forms of collaboration, is gestural communication. Gestural communication is involved in several of the mechanics of collaboration [3], but is only minimally supported in geocollaborative or CSCW systems (e.g., [25] and [14]). In particular, current systems fail to convey the height of gestures (or the distance of gestures from a vertical display), even though most map-based collaborative systems and many CSCW systems are surface-based. Part of this reason may be an underdeveloped understanding of exactly what information needs to be conveyed in gestural communication over distributed surfaces.

Systems that do not show gesture height can still be used, since people are capable of accommodating for a lack of bandwidth in communication [26]. However, previous research has also shown that more effective support for communication does improve collaboration, so a clear goal is the development of distributed collaboration systems that more fully support gestural communication.

Systems for the Capture of Gestures Over Surfaces

Capturing gestures made over surfaces remains a major challenge for providing support for gestural communication in distributed collaboration (and geocollaboration). Many technologies, such as the Microsoft Surface and DiamondTouch [27], can detect when users touch the surface of a table, but showing gestures above a surface requires more complex hardware and software (e.g., infra-red (IR) motion capture). A primary tradeoff is between capture quality and computational efficiency. Although camera-based techniques such as VideoArms [14] and C-Slate [15] provide high levels of detail for hands and arms, making gestures easy to see and understand, they are computationally expensive, have large network

demands, cannot be used for top-down projection systems, and can fail when lighting conditions change. On the other hand, systems that track users through other means (e.g., Vicon IR motion tracking, Polhemus magnetic tracking), have minimal network and computational requirements, provide three dimensions of information, are robust for changes in lighting or display; but often track only a few points on the hand and arm, can be complex to set up and calibrate, and are intrusive for users.

Collaboration

To support distributed collaborative activities, we must first understand how collaboration occurs. This section explores the requirements for collaboration, common models of collaboration and geocollaboration, and the low-level mechanics of collaboration and how they relate to geocollaboration.

Models of Collaboration

CSCW has no strongly defined taxonomy for cooperation, collaboration, group work, coordination, and the other semi-synonymous terms associated with research in the field [28]. Instead, much of the research cited by CSCW researchers that attempts to describe, model, or define collaboration comes from psychology [29], ethnography [30], education [31], or other associated fields [32]. Some CSCW researchers have worked to define elements of collaboration for a limited set of collaborative settings. For instance, Tuddenham and Robinson[33] compared collocated, mixed-presence, and remote collaborations on tabletops and identified territoriality, orientation, and consequential communication (through embodiments) as mechanisms critical to the success of collaborative systems. Tang *et al.* [17] found that identity, awareness, spatial metaphor, and corporeal embodiments are significant design issues in remote collaboration over shared surfaces.

There is extensive CSCW research on problems in collaboration that occur when collocated collaboration becomes mixed-presence (where some, but not all collaborators are not collocated) or remote (where no collaborators are collocated), and proposes solutions for those problems on an individual basis. Examples of this research includes Fraser's [34] work on display trajectories to solve the problem of locating pen-based cursors on large shared displays, Fussell's [35] work on supporting remote gestures, and Stach's [36] work on rich embodiments.

The most comprehensive and detailed work on developing a taxonomy of collaborative mechanics is the work by Gutwin and Greenberg [3] and, later, Pinelle and Gutwin [37]. They identify seven mechanics of collaboration, two of which frequently involve gestures [3]:

- Explicit communication: the intentional verbal and physical (most often gestural) communication between collaborators.
- Consequential communication: the unintentional communication between collaborators that results from their explicit actions and unconscious behaviours.

Their proposed mechanics encompass or align well with Tuddenham and Tang's high-level work as well as the individual results of Fraser, Fussell, Stach, and others. Taken together, their mechanics can form the basis for a model of collaboration from the perspective of CSCW research.

The following section examines the two mechanics of collaboration directly related to gestural communication as identified by Gutwin and Greenberg [3], explicit and consequential communication. Each mechanic is situated within related CSCW literature, and includes a discussion of how the mechanic works in geocollaborative settings.

Explicit Communication

Gutwin and Greenberg describe explicit communication during collaboration as intentional communication between participants [3]. Interpersonal communication is significantly more efficient when the participants have *common ground*, a shared model of the world or, at least, the collaborative environment [35]. During collaboration, collaborators automatically participate in the process of grounding: exchanging knowledge and building a common model to which they can refer [38]. Grounding most often occurs through explicit verbal communication and physical communication, which encompass both explicit gestural communication and elements of consequential communication (discussed below) [39]. Hindmarsh and Heath have documented the process of grounding in explicit communication, especially how it is comprised of both explicit verbal communication and deixis [1].

State of the art systems designed for CSCW research generally incorporate support for explicit verbal communication either as part of a video connection [40] or through dedicated audio channels [17]. However, some research has found that explicit communication has an impact on other ideas present in CSCW models of collaboration. For example, Gutwin and Greenberg found that the verbal component of explicit communication can be used to increase the workspace awareness of collaborators [41]. Verbal communication is also used when other channels are not available: For instance, Rodden *et al.* found that changing the seating around a computer screen changed the relationship among collaborators. If a screen faced only one collaborator, he/she was forced to increase the level of verbal communication in the collaboration, much as if the collaboration occurred over the telephone [42].

Geocollaborative research emphasizes explicit verbal communication as the single critical component of communication for collaboration [22]. However, some researchers have

identified the complexity or intrusiveness of existing geovisualization tools as a barrier to verbal communication [43]. Since expert users of geocollaborative tools prefer verbal to text-based communication, reducing the barriers to verbal communication is a design goal for geocollaborative tool makers [44].

Collaborators use explicit non-verbal communication, most commonly in the form of gestures, almost as frequently as words in face-to-face communication [4]. Bekker *et al.* divided gestures into four categories based on previous research and experimental observations [4]:

- Kinetic: movement that is all or part of a storytelling performance (e.g., showing the sequence of steps to fold a box correctly).
- Spatial: a gesture that indicates distance, location, or size, independently of any object or interface in the collaborator's environment (e.g., a gesture to indicate the size of a model not currently present).
- Point: the gesture component of deixis indicating "an attitude, attribute, affect, direction, or location" (p.159) [4].
- Other: purposeful gestures that do not fit in any of the above categories.

In experimental observations, Bekker found that gestures are brief, complex, include both 2D and 3D trajectories, assist with building persistent, imaginary structures in the collaborative environment, and occur in sequences, often interleaved with other activities. Pointing gestures have been found to be particularly critical to smoothly running workplaces [1] [3] [4]. The effectiveness of collaborative tools in supporting pointing gestures can be measured by the rate of conversational grounding; or how quickly all parties come to a common understanding of which object, abstract or real, is the target of the gesture [5]. Research in CSCW focuses on two

categories of systems that support explicit physical communication: collaborative sketch and design systems; and systems to support collaborative physical tasks [5].

Geocollaborative research agrees that gestural communication between participants is important during collaborative use of geovisualizations [44]. Since almost all geocollaborative work focuses on a central or shared map artifact, support for representing deictic references to the map or map elements has been emphasized by several researchers and prototypes [21] [45].

Other forms of explicit communication include writing messages or drawing diagrams, either in synchronous collaboration (e.g., chat messaging systems as in [46]) or asynchronous (e.g., email or fax). Annotation, or adding written information to existing artifacts, has been explored extensively in CSCW research (e.g., DOVE [47]) as a tool for both asynchronous and synchronous collaboration. Asynchronous annotation is supported in several geocollaborative platforms (e.g., Google Maps, Yahoo Maps, and Microsoft Live Maps), has been identified as a requirement for geocollaborative toolsets [44], and has been observed to take place during regular use of GISystems [48][44]. Related research has emphasized that annotated collaborative artifacts can be used to improve the exploration of shared datasets [49]. Despite this evidence, very little research has explored the sharing of map-based annotations [50].

A more extensive exploration of the use of diagrams and annotations as explicit communication has occurred in the specialized field of spatial decision support (a sub-category of geocollaboration). An example is the work by Rinner on argumentation maps [51]. Argumentation maps are used to assist with asynchronous conflict-based collaboration, in which individuals or sub-groups wish to make arguments about specific geographic locations. Such solutions, however, do not address the general use of diagrammatic and textual annotation in non-conflict-based geocollaboration.

Consequential communication

Consequential communication is unintentional information transfer that results from the actions of collaborators in the workspace [52][3], either through interactions with the workspace [157] or during conversations with collaborators. Consequential communication assists in developing an awareness of the interests and actions of fellow collaborators without any explicit, intentional effort on their behalf [41]. The capture and transfer of consequential communication between remote collaborators is a stated goal in CSCW research [52].

CSCW research has identified two aspects of consequential communication [52]:

- *Communication through interpersonal interaction*: Consequential communication occurs through the relative physical locations and orientations of collaborators and the interface(s). This information also provides clues about whether collaborators are present and about their identity [23]. Head orientation and eye gaze are also key aspects of consequential communication, assisting with conversational flow, conveying information about relationships and emotional content, and providing feedback [53].
- *Communication through the interface*: Collaborators' interactions with the user interface of a collaborative system provides consequential communication. Gaze direction can be important within this context [23]. Actions and intentions of collaborators can also be deduced from the movement of mouse pointers (or telepointers in remote collaboration); hand and arm movement when engaging the interface; and the responses of the system to user interaction, also known as feedthrough (e.g., an audible click when a button is pushed) [52][23][41].

Geocollaborative research has not specifically addressed consequential communication. However, studies of geocollaborative environments and design exercises for geocollaborative tools have identified aspects of consequential communication as critical to the success of geocollaboration [44][22]. MacEachren and Brewer designed, but did not evaluate, a tool intended to facilitate monitoring (see below for more discussion of monitoring), which also provides consequential communication about the social presence (how much, how recently, and with whom collaborators have interacted [54]) of collaborators [43]. To date, however, features that support consequential communication in geocollaborative tools have been poorly implemented, if at all [22]. This may be in part because geocollaborative solutions focus on interactions with the system rather than interactions between users. For example, when evaluating how to support multiple collaborators sharing a single large screen, rather than embodying the collaborators and representing their actions, researchers have focused on turn-taking and data protection [25].

Research in both CSCW and geocollaboration has identified historical interactions with the interface as a component of consequential communication. Much CSCW research has built on work by Hill *et al.* on read wear and edit wear -- the idea that digital artifacts can represent historical usage in the same way that physical artifacts deteriorate as they are used (e.g. books fall open naturally to frequently examined pages) [55].

Supporting Gestural Communication with CSCW Research

User embodiments of collaborators have been developed specifically to solve the problem of supporting communication between distributed users (e.g., [14], [16], [47], [56]). Embodiments assist with coordination and awareness, permitting users to focus on work or areas of interest to other collaborators. They also assist with activity coordination and can reduce inadvertent competition for resources between collaborators. Although embodiments assist with

awareness, there are also several tools and techniques for awareness that assist with attention-drawing and coordination, support shared control over interfaces, and promote private spaces for personal work in collaborative systems. The following sections review a range of embodiment and awareness tools and techniques.

Embodiment through Avatars

Embodiments in digital systems were studied in detail by Benford *et al.* [23] who stated, "the inhabitants of collaborative virtual environments (and other kinds of collaborative systems) ought to be directly visible to themselves and to others through a process of direct and sufficiently rich embodiment" (p. 243). Embodiment is often used interchangeably with 'avatar', a term popularized by Stephenson as a representation of a user within a multi-user digital environment [57]. However, avatars are generally more complete renderings of the user's body, often complete three-dimensional representations within a virtual environment (e.g., those described by Gerhard *et al.* [58]) (see Figure 2), whereas embodiments can encompass more abstract forms of user representation, such as telepointers [13], or partial representations of the user's body [14]. Any embodiment can be enhanced with extra information such as gender, experience, age, or domain-appropriate information to improve coordination and monitoring [36]. Such rich embodiments mitigate the tendency of participants to prefer interactions with collocated participants during mixed-presence collaborations [14].

Developing effective avatars as embodiments is difficult and many of the design issues are contentious or poorly understood. For example, Mori's [59] 'uncanny valley' of interaction with human-like robots has been extended to avatar interactions [60], but remains highly contested in empirical studies [61][62]. Additional problems abound in creating avatars that provide a sense of presence for the users [63], realistic interaction options [53], and reasonable controls [64]. Many of the problems relating to the use of avatars as virtual embodiments lie in

understanding what components of an avatar are most useful for representing users and their actions [53]. However, in the future, as input devices and displays develop, avatars will likely be an excellent solution to the problem of embodiment in digital systems [65][13].

Figure 2: Avatars currently have varying levels of sophistication, but at their best are capable of only limited non-verbal communication.

More abstract embodiments trade realism for lower-cost (in terms of processing power and network bandwidth) representations of critical information. There are two major abstract embodiment technologies of interest to geocollaboration: telepointers and video-based embodiments.

Telepointers

Telepointers are cursors that follow each collaborator's input device (e.g., mouse) pointer in a distributed system [13]. Telepointers are an exceedingly low-cost solution to embodiment, can be very responsive [66], and can appear to be immediate and natural representations of the actions of remote users [67]. They can be enhanced with temporal traces to become reasonably effective in representing gestures made by remote collaborators [13], can be aggregated to support large numbers of collaborators [68], and can contain awareness information about the system and the collaborator's identity and actions (see Figure 3) [69]. Additionally, telepointers have been shown quantitatively to be effective in improving perception and performance in learning tasks [70].

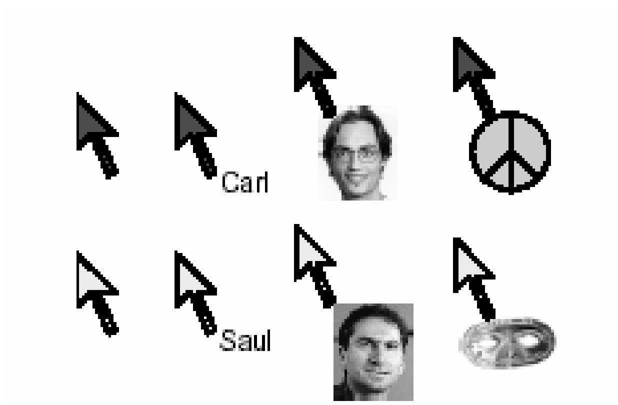


Figure 3: Examples of telepointers. Plain telepointers (leftmost) provide little information about collaborators, but can be enhanced (remaining images) to contain awareness information about collaborators' identify and actions (image from Greenberg *et al.* [69]).

There are also some limitations to telepointers in distributed collaboration. Pinelle et al. [24]found that location awareness is more difficult when users interacted with a system with an embodiment instead of direct touch, although their participants still preferred embodiments over direct touch in collocated collaboration. The design of successful telepointer-style embodiments on large surfaces is also highly dependent on the setting and tasks of the collaboration [72].

Where embodiments such as telepointers represent indirect input, they have been found to be unregulated by normal social protocols and can interfere with protection mechanics [71] although this can be mitigated by the introduction of automatic protection controls [73]. Finally, telepointer-based embodiments provide limited consequential communication bandwidth as compared to realistic avatars or video-based embodiments: facial expressions, body posture, physical location and other elements of consequential communication are absent.

Video-Based Embodiments

Video-based embodiments use digitized video, often filtered in some way, to create considerably more realistic embodiments than can be provided by telepointers. Video-based systems are often implemented using cameras perpendicular to the display surface, which capture any portions of collaborator's bodies that interpose themselves between the camera and the working surface. The captured video can be separated from the background and superimposed on collaborators' displays. Many systems, such as VideoWhiteboard [16], VideoDraw [56], VideoArms [14], and Apperley *et al.*'s work with video shadows [74], filter the captured video, reducing it to transparent, coloured shadows or outlines shown on remote displays (see Figure 4). Other systems such as DOVE [47] and Li *et al.*'s work with multiple cameras [75] transmit raw video images. Li *et al.* transmit facial information as well, as does Facespace [76], Facetop [77], and Clearboard [78].

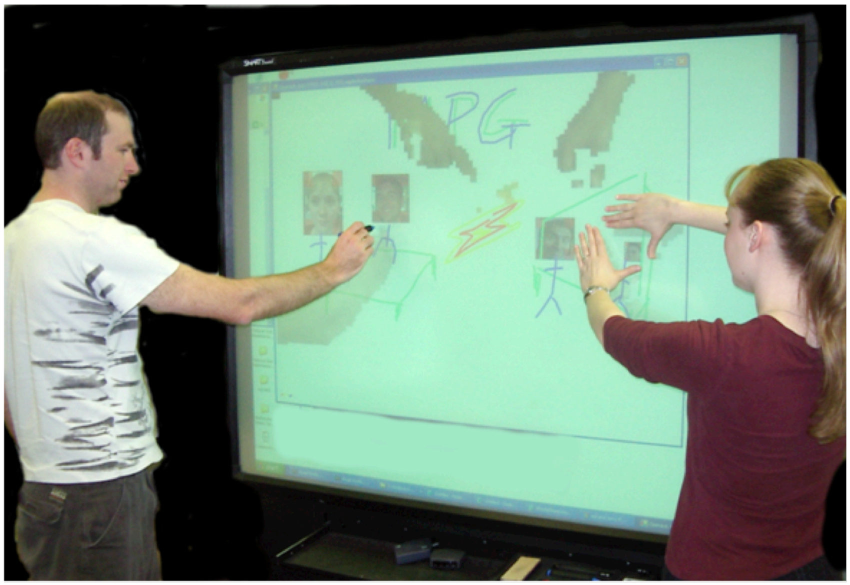


Figure 4: VideoArms in action. Remote participants are embodied by shadow-like images of their arms and hands, projected onto the workspace. Image from Tang *et al.* [14].

The above systems explore variations on three general techniques, each of which assists with presence and communication in mixed or distributed collaborative environments: Raw video, filtered video, and raw or filtered video with additional information (e.g., video of faces).

Raw video of hands and arms has been mostly used for collaborative physical or drawing tasks. In the DOVE system, the video embodiment was shown only to the teacher, allowing a teacher to view a learner [47]. An earlier prototype, VideoDraw, did provide a two-way collaborative interface [56]. User studies of both systems found that although the raw video feed was useful, the real benefit of the systems came from the versatility inherent in having both high fidelity video of hands and arms as well as a drawing tool (and, in the case of DOVE, a gesture recognition tool for assisting with drawing). These results have been supported in further studies of similar systems [35].

Filtered video of hands, arms, or entire bodies has been found to provide greater support for natural, explicit and consequential communication than telepointers [14]. Solutions of this type can also easily add components of telepointers, such as traces, to improve awareness [17].

As with raw video, however, not all interactions are well supported. For instance, there currently exists no solution that supports the representation of height or the distance from the surface of the table to gestures made with hands and arms.

Raw video of faces, possibly with heads and upper bodies, can be added to any of the above systems. Li *et al.* found that collaborators highly value the ability to see each other's faces (see Figure 5) [75], and other research suggests that video-mediated communication between collaborators is especially important during the negotiation stages of collaboration [79]. There is evidence, unfortunately, that the cognitive load for collaborators increases as their attention is split among more display windows [80]. An additional concern is that almost all of the research in this area has explored only very small groups of two or three collaborators. Whether any benefits from facial views through video during collaborative tasks exist for larger groups is unknown.

Under certain circumstances, however, incuring the additional cost associated with facial displays may be worthwhile: when a user is collaborating in a non-primary language, understanding is significantly improved (for that user) if the speaker's face is visible [79].

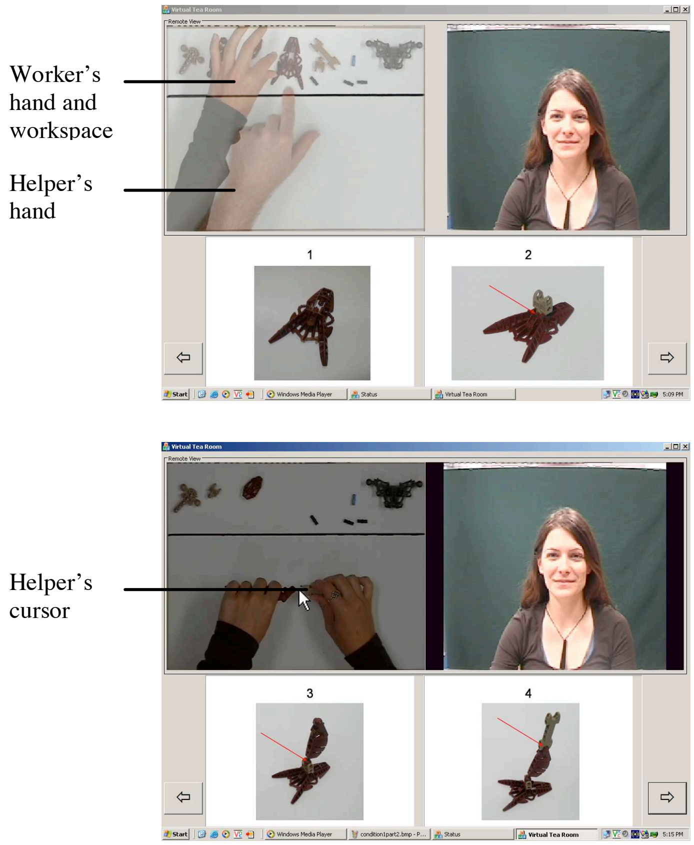


Figure 5: Li *et al.*'s system for supporting collaborative physical tasks. The addition of an extra video feed displaying the face and upper body of a remote collaborator is well liked, but may contribute to the cognitive load.

Despite the realism, the high level of unfiltered consequential communication and awareness, and the high fidelity and versatility of video-based embodiments, some results suggest that for certain kinds of tasks, such as collaborative sketching, telepointer-based embodiments are at least as effective as video-based embodiments, at a fraction of the bandwidth and processing costs [35][75].

<u>Awareness Tools and Techniques</u>

Techniques and tools that facilitate awareness in collaborative environments improve consequential communication, coordination of action, planning, monitoring, and protection. Embodiment techniques, such as those discussed above, can play a large role in awareness, but

there are several additional techniques, adjustments to techniques, or particular tools that can further improve awareness. Awareness is not limited to distributed or mixed-presence environments where embodiments are most commonly used, however, since maintaining an awareness of collaborators in a co-located system can also be difficult. The awareness literature reviewed earlier (Gutwin and Greenberg [41]; and Salvador *et al.* [81]) defines awareness as the ongoing ability to answer a set of questions about the environment and collaborators, whether collocated, mixed-presence, or distributed. This can be done through an awareness of the system state, the collaborators, or the collaborators' interactions with the system. Of the three methods, issues related to the awareness of collaborators and their interactions with the system are of the most interest in the design of geocollaborative systems and are discussed in greater depth in the following sections.

<u>Awareness of Collaborators</u>

Embodiments generally provide an awareness of both collaborators and their interactions with the system, but embodiments can be tuned specifically to improve awareness of collaborators. For example, Stach *et al.* found that a player's personal information, experience, activity, and information about their current and historical (and potential) interactions with the system could all be included effectively in an embodiment roughly the size of a telepointer [36]. Other research has found that very simple pieces of information about collaborators' relationships with the environment, such as their orientation, are easy to include in a telepointer, and are valuable to collaborators [69].

Awareness of social interactions between collaborators is often obscured during distributed collaboration, especially when some participants are collocated and others are distributed. Erickson and Kellogg, along with various collaborators, have developed systems that

are *socially translucent*: they transparently express the relationships among collaborators in the system [54][82][83][84]. For the most part, these systems have been developed to track chat messaging or other text-based interactions (other related examples are OpenMessenger [85] and the CommunityBar [86]). Other systems have been examined with respect to social translucence (e.g., Gibbs *et al.* [87]), but there is, as yet, no quantitative evidence that improving social translucence improves awareness.

Understanding and Representing Deixis

Deixis is part of non-verbal explicit communication, consequential communication, and awareness in collocated communication [2], [4], [39], [88] and representing gestures to remote collaborators is an important groupware (a sub-field of CSCW that encompases the design and use of software that facilitates collaborative work) design factor [3]. The field of CSCW has yet to identify a definitive solution for representing deixis in any environment. The following sections review the scientific literature on understanding the role of deixis in communication and the development of systems that successfully represent deixis with an appropriate level of fidelity.

<u>The Role of Deixis in Surface-Based Collaboration</u>

Although there are several studies that examine the use of gestures in a variety of communication contexts (e.g., [39], [88]), research that describes gestures in surface-based collaborations is less common. General research into deixis in communication contexts, such as the work of Bekker et al. [4], has found that gestures are used as a communication medium in face-to-face meetings almost as frequently as words. Other studies have identified the presence of understandable deixis as critical to smooth collaboration [1] and that certain kinds of tasks (e.g., description or identification tasks) tend to result in higher levels of deixis on larger displays (as opposed to smaller displays) [89].

Surface-based deictic gestures are different than the more general case because the target is often within reach and the target space is often limited. Kirk *et al.* [11] examined hand movements during mixed-ecology collaborative tasks (where some participants are colocated and others are distributed) and developed a coarse-grained analysis of their characteristics. One important finding was that participants worked hard to accommodate for 2D representations of 3D actions. A characteristic gesture was "the inhabited hand", where the instructing collaborator placed his/her hand over the remote hand and demonstrated the desired 3D motion, such as a wrist rotation.

A finer-grained analysis was performed by Kettebekov and Sharma [8], who developed a semantic classification for deixis based on observations of weather narrations. Because the goal of the research was automated gesture recognition, their conclusions focused on the effectiveness of the system in recognizing gestures rather than characterizing gestures within the context of collaboration.

Embodying Gesture Height

The height of a gesture above the surface of a shared workspace has been identified as a possibly valuable piece of information in distributed communication [34], [40]. Embodiments that represent height information have considered height in two ways: as one of many components of gestural communication, or as a way to provide feedthrough about others' actions on the surface.

Height in Existing Embodiment Techniques

One of the earliest video embodiments, VideoWhiteBoard, captured real shadows cast on the display surface [16]. This had the side effect of showing diffusion in the shadows as people moved away from the surface, a cue that could be used by remote participants to better

understand people's locations. Later implementations of the idea used digital rather than analogue shadows, and did not convey the same kind of distance information [14].

Distance information can also indicate intention to make a deictic gesture. Fraser *et al.* [34] found that explicitly visualizing the approach of participants' pens toward a wallboard significantly improved coordination in a video-annotation task, and reduced conversational latency. They found that lack of information about gesture height was a barrier to communication in collaborative video annotation: annotation events were occurring quickly and were often missed by remote users because it took too long to locate the remote user's writing tool icon on the video screen. Their solution involved subtle representations of the approach of the writing tool to the digital whiteboard surface, thus allowing remote users to anticipate the point of contact at the beginning of the annotation. Since they were using a sketch tool, rather than supporting deixis, they used a fuzzy dot as a representation of the pen location that increased in opacity as the stylus approached the surface – an abstract visualization. Both Fraser et al.'s work and VideoWhiteBoard are effective in showing heights or distances larger than 5cm, but neither of these projects explored a more complete design space for height.

Hilliges *et al.* [90] used the height from a surface as a method of providing more sophisticated interaction with 3D objects. They provided hand shadows as a feedback technique, inverted so as to appear smaller when hands were farther from the surface. As an interaction technique, distance from the surface can be mapped, as they suggest, to less engagement, and therefore the user occludes less space on the surface. This approach is the opposite of the natural feedback provided in VideoWhiteBoard's shadows, but Hilliges reports that users had no problems with interpretating this inversion. This decision is questionable for other reasons, however. Larger embodiments are not necessarily distracting (especially if translucent or shown

behind artifacts), but they do provide more subjective awareness than smaller embodiments [24]. It is likely that the system developed by Hilliges would provide less awareness in an environment with more competition (e.g., with more than two users).

Recently, Tang *et al.* enhanced the VideoArms technique to show 'touch pearls' [17] or contact traces when participants were touching the table surface, using an effect similar to telepointer traces [13]. They mentions that proximity information was not well modeled in their design, despite the improvement provided by the traces; however, they note that the traces did provide a level of awareness not available through the original VideoArms design. No evaluation of the contact traces was carried out.

Height as Feedthrough

Height or distance from the surface can also provide information about user activity, since surface touches are often used to invoke commands in the system. Visualizations of height are therefore used to give feedback to the user making the gesture, and can also provide feedthrough to other users [91]. The visualizations generally show the difference between hover and touching states on touch or pen interfaces (e.g., with the C-Slate interface [12], or with Wigdor *et al*'s ripple visualizations [22]). Although work on feedthrough that incorporates height information provides valuable insight into embodiment design, this research has not investigated how including height affects communication in distributed collaborations.

Showing Gesture History

Gestures occur quickly and are easily missed in distributed collaborations when users may be attending to things other than remotely located collaborators when a gesture begins. Showing the historal location of embodiments has been found to be useful for helping users see just-missed actions but has also had some unexpected benefits. When Gutwin introduced

temporal traces on telepointers [92], he found that users also used them to create temporary annotations in addition to the benefits found in the ability to see and interpret gestures. A later expansion on this work [13] found that the traces also reduced challenges with losses in temporal fidelity (e.g., information received out-of-order by remote clients).

The benefits to communication can be profound: Yamashita *et al.* [93] found that a 'remote lag', essentially a short-term history of movement, reduced communication problems during mentoring tasks and Fussell *et al.* [35] found that pen drawing in addition to DOVE made distributed collaboration almost the same as side-by-side collaboration for instructional physical tasks. In Fussell's experiments, the pen permitted a kind of history (although it had to be manually erased), which might possibly have been achieved by permitting the same kind of fingertip drawing found in Tang *et al.* [17].

History also provides a mechanism for adding information in a distributed context that is not normally available in a collocated context: work by Erickson and Kellogg [82] has proposed displaying presence and activity between collaborators over time through social translucence graphs added to embodiments.

Capturing Deixis Over Surfaces

There has been considerable research on using height as a variable for input and capturing that height with cameras (often depth cameras). Most often, height has been used to identify touch interactions on variably distant surfaces, such as with OmniTouch [94][9], which used a depth camera similar to the Microsoft Kinect. Detecting touch on fixed-distance (non-touch-sensitive) surfaces has been shown to be possible using the Kinect camera (e.g., [28]). Marquardt *et al*. are one of the few to suggest that interactions should span the spaces between touch, hover, and above the surface [17]. More elaborate models of a workspace have been created with

moving cameras, such as in KinectFusion [13]. For most surface-based computing, however, a fixed camera and fixed surface is a more common scenario.

Although height detection in the layer above (or in front of) a surface has been used in a variety of gesture recognition and input techniques (e.g., [1] and [12]), there are few embodiments that represent height in this space. The exception is Fraser *et al.* [4], who showed that representing approaches to a surface can improve distributed interactions. They did not, however, capture detailed information about the hand and arm, only the position of the tip of a stylus.

Conclusion

A key aspect of supporting distributed collaboration, whether in geocollaboration or in other domains, is finding a way to allow the expression of communicative gestures. Although embodying users appears to be a promising solution, there is little related work that specifically examines the success of embodiment techniques in expressing various kinds of gestures, particularly deixis. One problem is that with the exception of Kettebekov and Sharma [8], there is little work that classifies and explicates surface-based natural deixis from the perspective of designing embodiments to convey those gestures. Without such a classification, designing and evaluating an embodiment to show deixis over surfaces becomes more difficult and time consuming. The next chapter describes research performed to develop an understanding of how people perform deixis over surfaces and how those gestures can be classified to assist with the design of remote embodiments.

CHARACTERIZING DEIXIS OVER SURFACES[1]

In Chapter Two, embodiments were described as one of the most common methods used by applications to support remote, synchronous collaboration. However, there is little information available about what to include or exclude from the design of embodiments for the purpose of representing deixis. The state of the art in embodiment design does improve communication between distributed groups; in part by showing deixis performed by remote users. However, embodiments cannot currently be evaluated in terms of their expressiveness because there is no clear characterization of deixis over surfaces (i.e., what movements comprise the gesture, the position of body parts during a gesture, and where above the surface the gesture occurs), and there are no principles available to tell designers which parts of a deictic gesture are most common, or most important. Such a characterization would help people make informed decisions about embodiment design, and could also inform other questions such as how to best convey gesture information in environments with limited network, display, or computational resources, or how to design command gestures so that they do not collide with natural communication gestures.

To address the lack of a surface-based deixis characterization scheme, three observational studies were performed in which more than 450 gestures were recorded and analysed. Two laboratory experiments investigated information-sharing tasks over projected maps, and a field study observed discussion and collaboration over maps between park wardens and science students. These observations provide new insights into four questions that have particular importance for the design and implementation of tabletop embodiments: what parts of the body are involved in deictic gestures, what atomic movements make up a deictic gesture, where the

---

[1] Material from this chapter was published as "Characterizing Deixis over Surfaces to Improve Remote Embodiments" in Proc. of ECSCW 2011 [95]

gesture occurs in the space above the table, and what other physical characteristics different types of gestures exhibit.

Examining Surface-Based Deixis

To further our understanding of what deixis looks like when performed over surfaces, three observational studies of deictic communication were performed: two laboratory-based studies and one field study. Ethics consent for all of these studies was granted and examples of the consent forms and questionnaires are included in Appendix B.

<u>Laboratory Study Methods</u>

In both laboratory studies, participants were asked to carry out a series of tasks using a top-projected tabletop that showed a Google Earth map of Saskatoon, Saskatchewan. Participants were initially seated, but were allowed to move freely around the table as needed. The map was a combination of street map and satellite image at a resolution sufficient for counting (but not identifying) houses in the image. Maps afford rich opportunities for deixis: for example, identification of individual or groups of artifacts, paths between or along artifacts, and areas that include multiple artifacts. In this way maps approximate many other cluttered workspaces used for planning or design tasks. Although maps do not provide a setting for all the collaborative tasks in which deixis is used, they do capture a large subset of this task space, particularly for two-dimensional displays, allowing the extension of these research findings beyond the domain of geocollaboration.

In the first study, four pairs of participants answered questions about the spaces represented by the displayed map. Questionnaires (included in Appendix B) were formulated using previous work by Kettebekov and Sharma [8] as inspiration. Although the questions were asked by the researcher, participants were instructed to direct their answers to each other. This meant that dialogues developed between participants rather than between the researcher and the

participants. The questions in the questionnaire were designed to elicit different kinds of deixis: some questions were designed to elicit path gestures, others to elicit indication of areas, and others to elicit pointing. These tasks simulate a variety of information-sharing collaborative activities seen in the real world. Participants (one female and seven male) were all staff or students at the University of Saskatchewan, and ranged in age from 22 to 56. Participants were from a variety of professional backgrounds, and were selected so that they did not work regularly with maps in their profession. All were familiar with Google Earth and had used it previous to the study.

Sessions were videotaped with a single camera at an angle oblique to the table's surface. The resulting recordings were reviewed several times by the investigator to identify episodes involving deictic gesture, and to determine a set of candidate classification categories for the observed deictic gestures. Deictic gestures were separated from other kinds of gestures, such as conversational gesticulation (e.g., shrugging or using descriptive gestures), which were discarded in the analysis.

Although the recordings contained a wide variety of deixis, subsequent analysis of the data from the first study identified limitations in the way that gesture data were captured: in particular, a single camera was insufficient for capturing all of the detail of a gesture. For example, the height of a gesture above the table was often difficult to determine, and when the gesture was performed above the table, the x and y-axis coordinates could be difficult to identify. With the single camera in study one, participants' arms, hands, and bodies also sometimes occluded their gestures from view. To resolve this problem, the second study used two cameras at 90 degrees to each other. One camera was located at the table's surface and aimed so that each

gesture's height was easily determined (see Figure 6). The second camera provided a top-down view of the table's surface (see Figure 7).

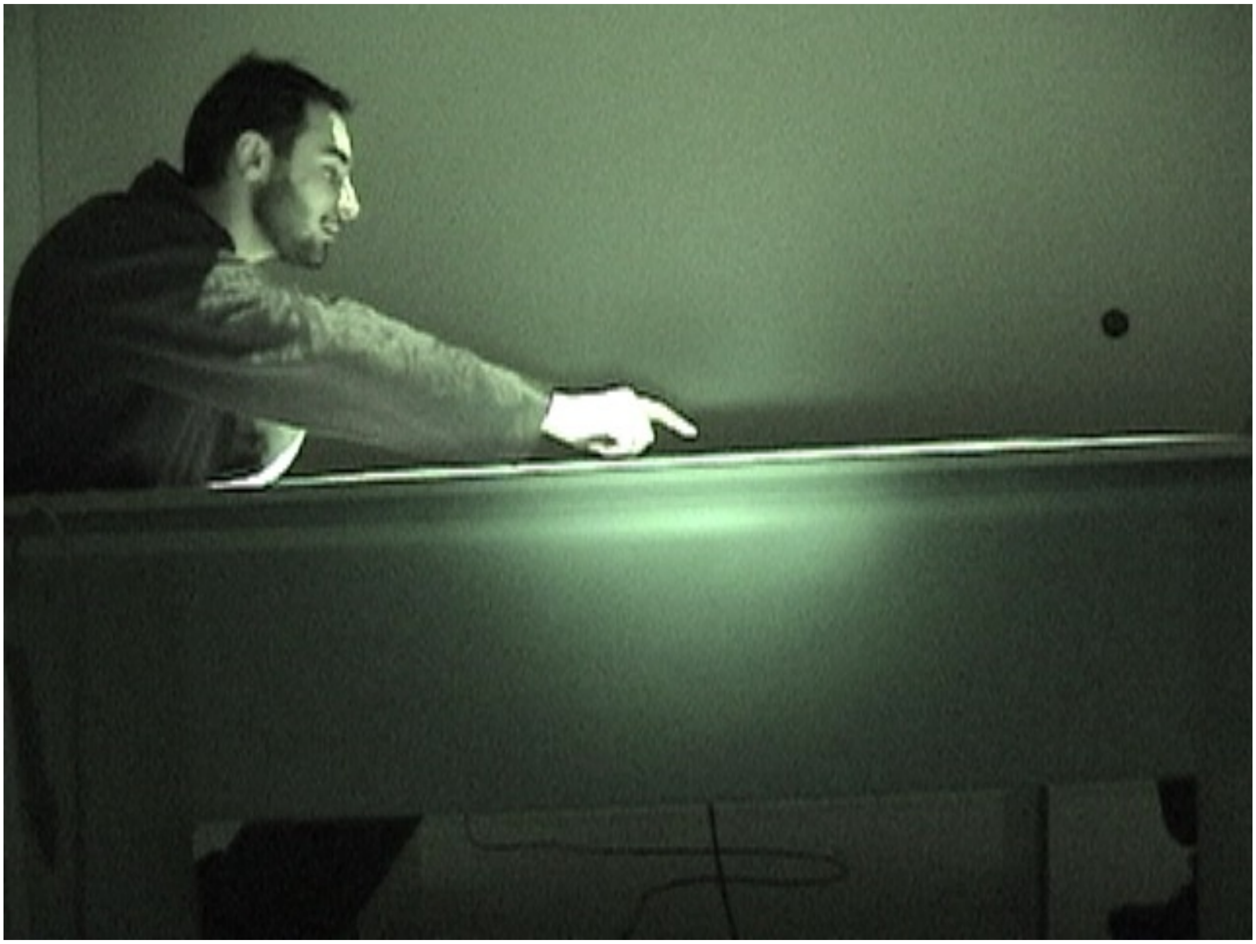


Figure 6: From from horizontal, table-top level camera in the laboratory experiment.

Figure 7: View from top-down camera in the experimental set-up.

The tasks in the second study were similar to that of the first, but with small alterations in order to explore the issue of people's confidence in the accuracy of their answers. Observations from the first study had suggested that people express these qualities by changing the height of their deictic gestures or hesitating during a gesture; the second study therefore asked participants to use new and unfamiliar information in some of their tasks. This study was performed by two pairs of male participants (from the University of Saskatchewan) who had not participated in the first study, aged 22 to 30. Both were graduate students in computer science, had previous experience with Google Earth, but did not use maps professionally.

The recordings of the second study were reviewed and analysed in a way similar to that of the first. The types of deixis seen, and the categories generated, from the second study were very similar those of the first study. No significantly new types of deixis were apparent, nor did

the association between atomic gestures and targets (i.e., paths, areas, or points) change in any substantive way. As expected, however, the two-camera setup of the second study did provide new information that allowed clearer delineation of gesture location.

Field Study Methods

A field study of a real-world collaboration was performed in order to gather additional observations and to compare the findings of the laboratory studies with those gathered from a more realistic setting with a larger group of collaborators. With this study, the goal was to observe a variety of collaborative tasks different from those created for the laboratory studies. A group of four veterinary science students, one graduate-level teaching assistant, and a park warden were observed and recorded with video and stills during a two-day workshop in which the students learned about how wardens effect the transfer of herds of wild ungulates between parks and preserves. The workshop took place in conference rooms, indoor and outdoor animal enclosures, and a variety of outdoor facilities in Elk Island National Park, Alberta, Canada. The students and wardens carried out numerous discussions over different types of maps including wall-mounted maps (Figure 8) and hand-held paper maps (Figure 9), in several indoor and outdoor settings. In addition, there were two cases of ad-hoc map use, one involving a sketched map on a blackboard, and one involving a map drawn in snow (Figure 10). Because of the nature of the workshop, most gestures came from one of the park wardens who provided a great deal of information to the other participants. Other participants, however, did perform gestures, usually in short bursts, and often for the purpose of achieving conversational grounding (see Figure 11).

Figure 8: A senior Parks Canada officer shares information about the geography of the park with students (out of frame on either side).

Figure 9: Students planning the afternoon’s route during the field study.

Figure 10: An ad hoc map drawn in snow by the senior wildlife management officer of their current location and surroundings. The officer and students took turns alternating between pointing at sections of the map and the landmarks in their environment.

Figure 11: Discussion around a wall-mounted map inside an elk management facility. Students took turns approaching the map to point.

Recordings and notes from the field study were analysed with methods similar to that used for the laboratory studies –all episodes where gestures took place were identified and each instance was categorized using the categories developed in the earlier studies. Overall, the types of deictic gestures seen in the field were similar to what was observed in the laboratory, and the existing categories were able to characterize all of the gestures of the field study. However,

many of the episodes observed in the field were in a presentation style, with a vertical surface and a seated audience. The effects of this difference are discussed in the analysis below.

Analysis: four basic questions about deictic gesture

The video tapes of the study sessions were analysed using four basic questions to help identify and characterize the ways in which deictic gestures can vary in real-world activity, and therefore imply the variations that remote embodiments should attempt to convey. The questions are:

- what parts of the body are used to produce a deictic gesture?;
- what atomic movements make up deictic gestures?;
- where does the gesture occur in the space above the table?;
- and what additional physical characteristics do gestures have in addition to pointing?

The analysis was performed by identifying each gesture in the video and coding it using the above questions. Coding was done by only the primary researcher. As with the laboratory experiments, non-deictic gestures were discarded from the analysis.

<u>What Parts of the Body are Used for Deictic Gesture?</u>

The position of the body parts used in the production of a gesture (i.e., its *morphology*) can provide insight into what information is needed to correctly interpret the gesture. This analysis is vital for the design of remote embodiments, as it tells us what should be tracked at the local site and visualized at the remote location, and how to optimize information about the gesture. It can also show what spaces are available for command gestures without risking mis-interpretation.

Variations in the morphology of deixis over maps were observed to come primarily from the fingers and hand. The lower and upper arm, the shoulder, and the rest of the body play less of a role in the meaning of a deictic gesture; the movement and orientation of these body parts is

most often the result of intended movement of the hand, not the result of a communicative intention. In some cases, the overall posture of the body and arm (e.g., an extended arm or a leaning-over body posture) provided valuable awareness information about the hand's location (e.g., that the speaker was reaching to point to something far away), but the idea of drawing attention to the gesture is a separate issue from the interpretation of the gesture itself. This means that the most important body parts for understanding and representing deixis are the hand and fingers, and their movement, posture, and orientation.

The parts of the hand available for use in the production of a gesture are the five fingers and the palm or the back of the hand. A part of the hand is considered to be engaged in the gesture if it is not de-emphasized (e.g. a finger curled into the palm) and it is integral to the interpretation of the gesture. Fingers can also be grouped or spread: for example, a gesture can be morphologically described as engaging the thumb by itself, first and second together, and third and fourth together, and either engaging the palm (see Figure 12, top) or not engaging the palm (see Figure 12, bottom) . The engagement of the palm (i.e., its importance in the gesture's interpretation) may not always be easily determined, but in the studies was often apparent in the larger context of the gesture.

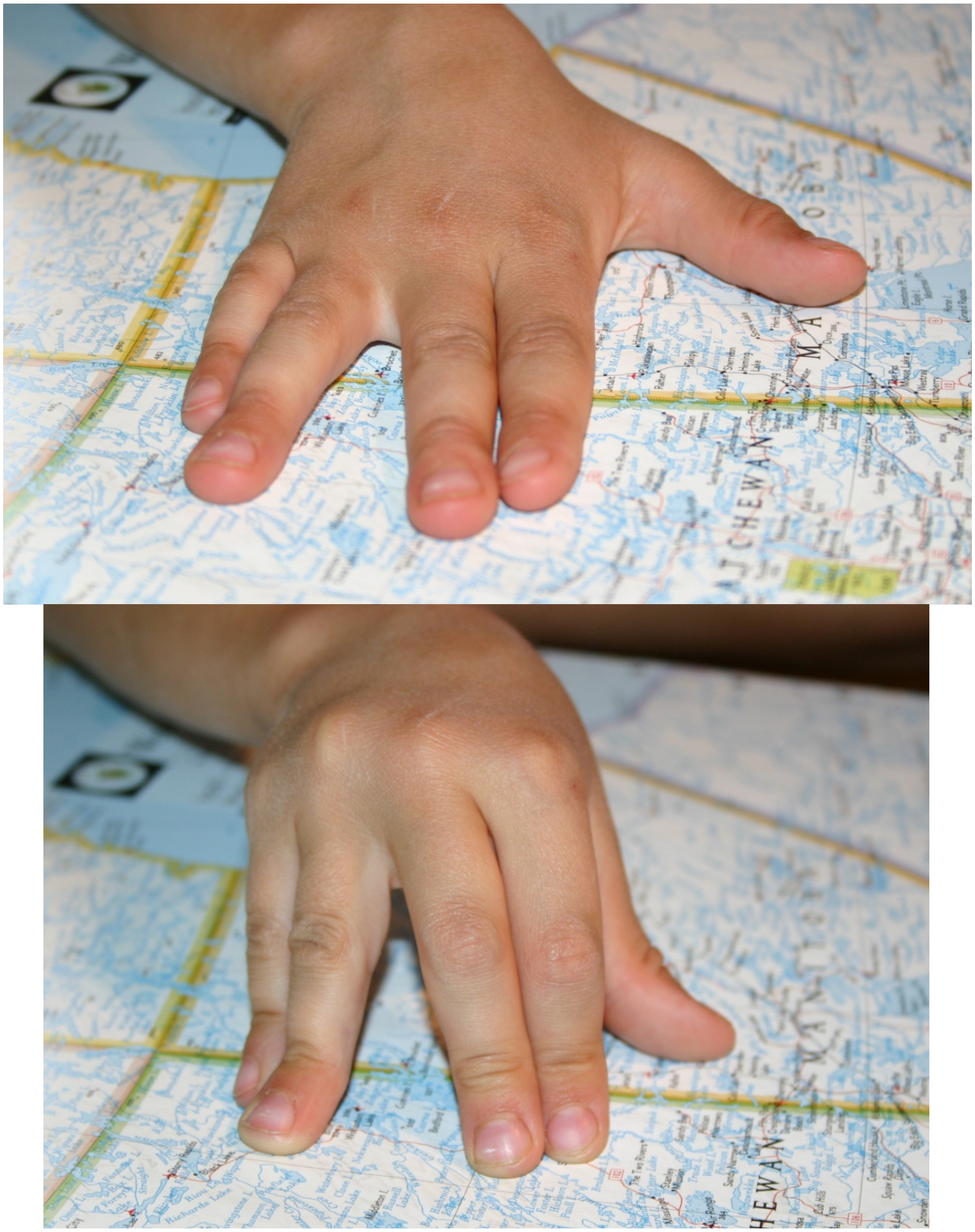


Figure 12: Examples of palm engagement/disengagement. Above, a gesture with the palm engaged, indicating an area of the map; below, the same morphology, except that the palm is not engaged, indicates several points or small areas.

Hand orientation describes the relative position of the palm with respect to the map surface. Gestures can be described as palm-down (palm faces the surface, Figure 13, top), palm-

up (palm faces away from the surface, Figure 13, middle), or sideways (see Figure 13, bottom). This category is independent of the morphology of finger and palm engagement.

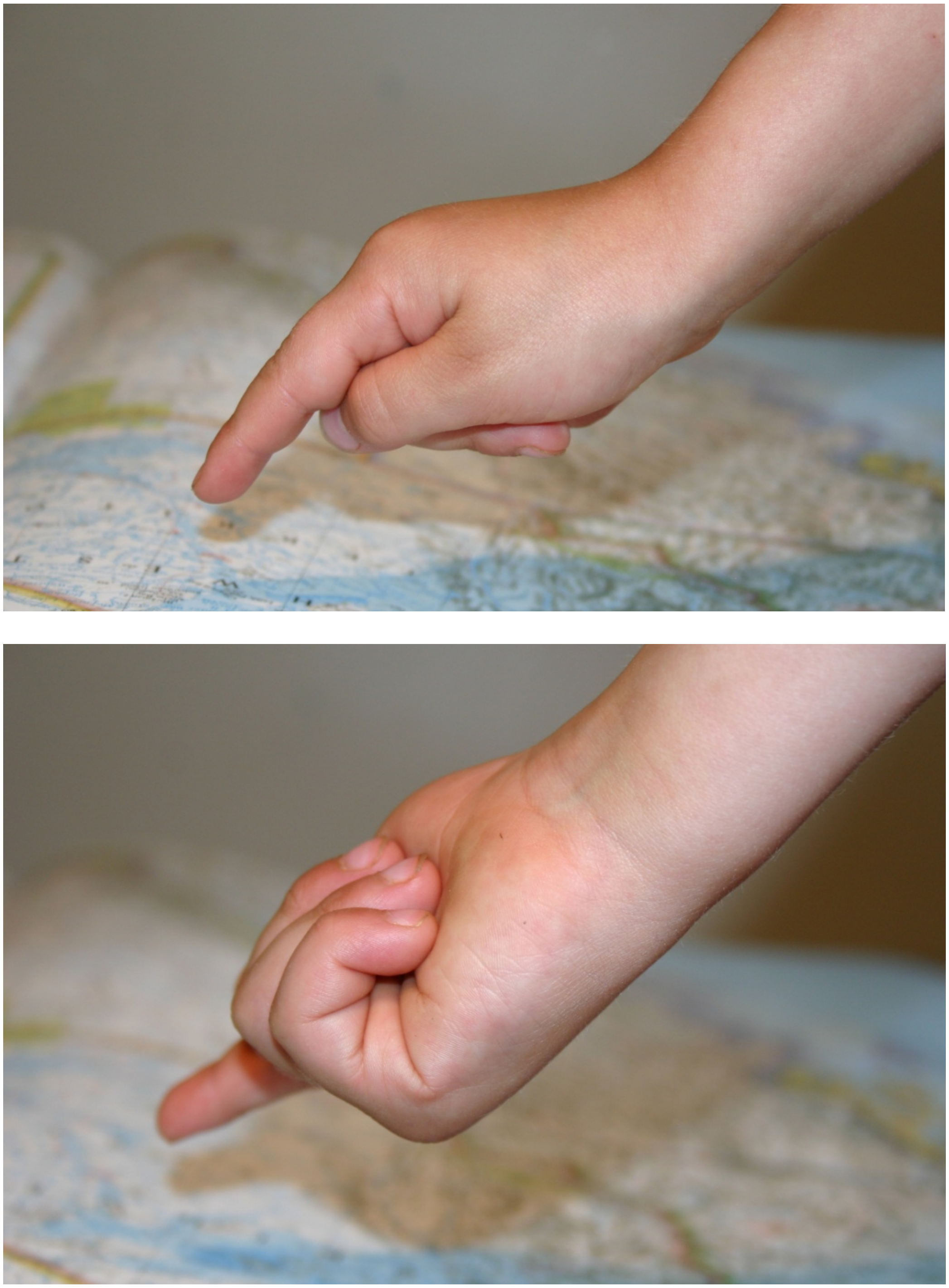


Figure 13: The same pointing gesture with different palm orientations. Top, palm down; middle, palm up; and bottom, palm sideways.

Of the parts of the hand actually used in deictic gestures, the index (1st) finger is of prime importance (see Figure 14). In the first study, only 15 (7%) of the 225 observed gestures did *not* engage the index finger. Of those, 11 used the palm of the hand to indicate an area on the map – and all of these were generated by two of the participants. The second study was very similar, with only 10 gestures of 146 that did not involve the index finger. Four of these engaged the palm, always in a sideways orientation (i.e., a ‘cutting’ or ‘separation’ gesture); the remainder engaged the middle (2nd) finger and had no palm engagement.

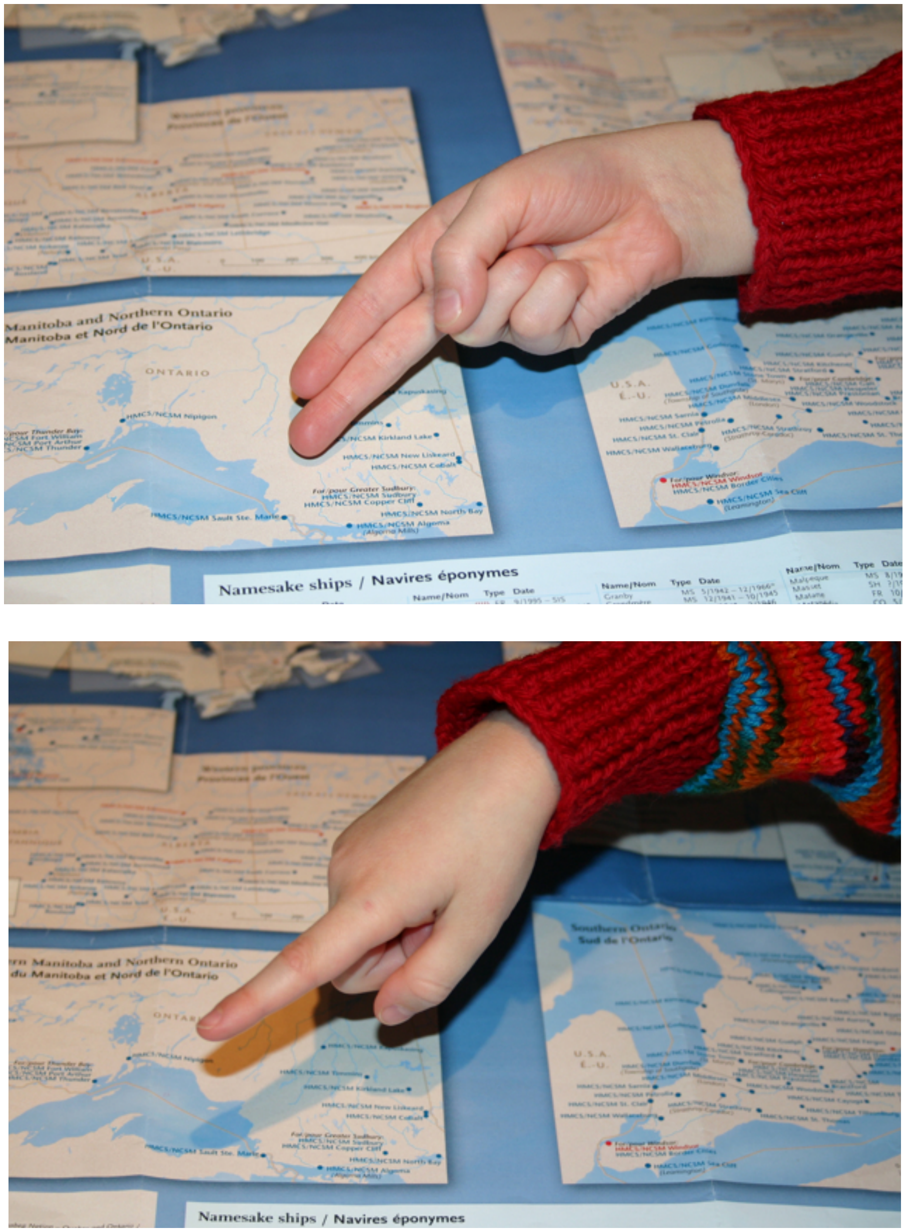


Figure 14: Examples of the two most common pointing gestures. Above, a double-finger point with the palm away from the surface. Below, a single-finger point with the palm down.

In contrast to the laboratory observations, during the field study 26 (32%) of the recorded gestures engaged the palm and only 41 (50%) of the gestures engaged the index finger. The engagement of the middle finger was used as an alternative to the index finger in 26 (32%) of the gestures. Additionally, 36 (44%) of the gestures were performed palm-up rather than palm-down to the surface, a position that was likely easier due to the positioning of the map (several maps in

the field study were tacked to the wall). All of these non-index-finger gestures in the field study data are from a single participant, however, and further work is needed to determine whether these morphologies are common.

Use of two hands for gestures was rare, with only one episode of deictic gesture involving both hands simultaneously for the gesture itself (a palm-engaged gesture with both hands to indicate a large area on the map). Two hands were used on other occasions, but with the second hand (always the non-dominant hand) used as a placeholder. For example, if the participant was tracing a large contour on the map, he/she might place the non-dominant hand at the start position and leave it there until the pointing hand returned to the start.

The observations of deixis morphology over maps can be summarized in two ways: first, a large majority of episodes used one of two pointing fingers in classic pointing gestures (Figure 1); second, the remaining smaller set of gestures were highly varied in their morphology. Between the two laboratory studies, 95% of deictic gestures engaged the index finger, although often in conjunction with additional fingers. Finger-based pointing with the first or second finger was also extremely frequent in the field study (82% of gestures).

<u>What Atomic Movements Create Deictic Gestures?</u>

Gestures have been previously characterized in terms of small, atomic blocks of movement, a scheme designed to assist with automated classification [8]. In this scheme, atomic gestures are strung together (often in long chains) to create complete gestures. Using this idea, seven distinct atomic blocks were identified from the observations, which uncovered several substantial problems for gesture embodiment as a result of this classification. The seven gesture ‘atoms’ observed were: preparations, strokes, points, contours, retractions, rests, and hesitations (see Table 1). These differ only slightly from those of Kettebekov and Sharma [8]. In particular, the stroke atom replaces a set of gesture atoms used by Kettebekov and Sharma, contour is a

close substitute for their circle atom, and the hesitation atom is a new atomic gesture, not described by Kettebekov and Sharma. These atomic gestures are discussed in greater detail, below.

Table 1: Gesture atoms and their short descriptions.

Gesture Atom	Description
Preparation	Hand moves into position, ready to gesture
Stroke	Movement along 2D or 3D curve
Point	Movement stops before continuing (intentional indication)
Contour	Movement in a closed shape (e.g., ellipse)
Retraction	Movement away from position, no longer ready to gesture
Rest	Not gesturing
Hesitation	Semi-aimless movements that accompany visual search or other conversational pause.

Gestures begin with *preparation*, a gesture atom with no explicit meaning, designed to move the hand and arm into a position where a meaningful gesture can occur. A preparation atom can, but does not always, serve to attract attention [96] to subsequent atoms.

The next three atom types involve meaningful movement, and form the core of the deictic gesture. *Stroke* atoms are movements along a line or path in two or three dimensions, *point* atoms are meaningful pauses in the gesture movement; and *contour* atoms are path-like gestures that curve and close, returning to the point of origin (or near the point of origin, depending on the precision of the gesture). Examples of stroke and contour atoms can be seen in Figure 15. All three of these atoms can be used to indicate any artifact in the workspace, but there is a natural mapping between stroke atoms for showing paths in the workspace, point atoms for showing point locations and directions; and contour atoms for showing areas. Strokes, contours, and

points are all indicative atoms, in that they can be used to indicate artifacts or locations on the working surface.

Retraction atoms occur at the end of a deictic gesture and before the start of another. Although not all gestures have retraction atoms, many do. Retraction atoms may lead to *rest* atoms, where the hand and arm are no longer engaged in deixis, but remain in the working space.

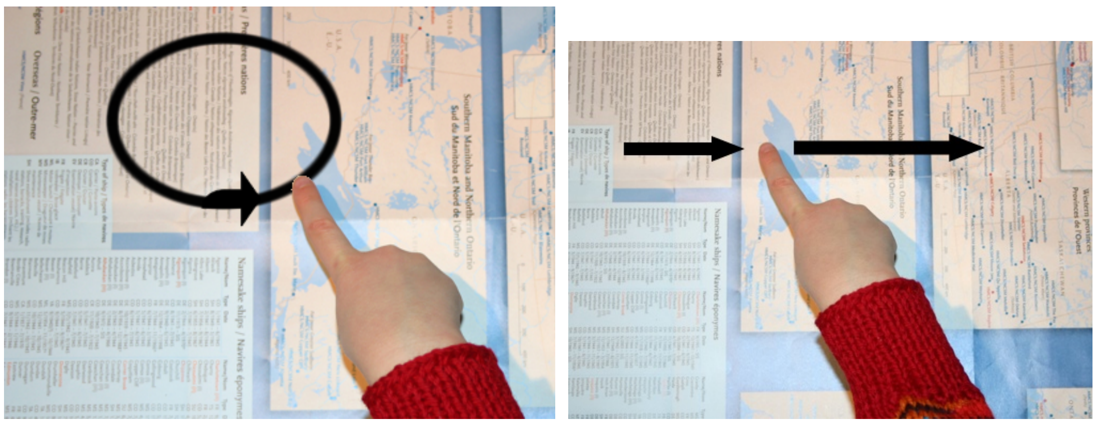


Figure 15: Contour atom (left) and stroke atom (right). Arrows indicate the movement direction of the finger.

The seventh and final atom is *hesitation*. In the time between a preparation atom and an indicative atom, people often hesitate in mid-gesture, performing a visual search of target locations or otherwise pausing in the conversational flow. During this time, people also pause the deictic gesture, but they rarely stop moving – instead, they often carry out a series of aimless movements over the potential target space. This movement is visually distinct from any of the other atoms, but is difficult to characterize, other than that the movements appear interrupted, hesitating often, and, in the case of visual searches, often loosely follow head orientation and gaze direction. A similar atom (although not described as such) was identified by Kirk et al. [11] as the “wavering hand.”

In all of the studies, the frequency of indicative atoms was counted. In the first study, there were 225 distinct indicative atoms, of which 91 (40%) were points, 104 (46%) were strokes and 30 (13%) were contours. In the second study, there were 146 indicative atoms, of which 56 (38%) were points, 61 (42%) were strokes, and 29 (20%) were contours. Deixis in the field study was less frequent than in the laboratory studies, but three segments of over three minutes of almost continuous deixis were observed. During these segments, there were 43 (52%) pointing atoms, 35 (43%) stroke atoms, and 4 (5%) contour atoms.

Table 2: Summary of gesture types by study.

Study	**Points**	**Strokes**	**Contours**	**Total Observed**
Laboratory 1	91 (40%)	104 (46%)	30 (13%)	225
Laboratory 2	56 (38%)	61 (42%	29 (20%)	146
Field Study	43 (52%)	35 (43%)	4 (5%)	82

Brief hesitation atoms were observed in almost every series of atomic gestures in every study, and when they were not present, the participant usually paused in a rest atom before answering the task's question (in the laboratory studies). Statements such as "somewhere over here", "I'm not sure where, exactly", or stalling vocalizations (e.g., "um") frequently accompanied hesitation atoms.

<u>Where Does the Gesture Occur in the Space Above the Table?</u>

The presence of a planar surface in a collaboration setting introduces the possibility of measuring the *height* of a gesture from the surface, a measurement not feasible in a more varied workspace. Based on the observations in these studies, height is an important characteristic of gestures over surfaces. It is clear that as the height of a gesture changes, the gesture can imply different meanings than if the height were to remain the same. Additionally, movements such as

tapping on a point or bouncing a finger along a path are difficult to express without considering variations in height.

In these studies, height of a gesture, and variation in that height, was highly meaningful. There were four main heights: gestures touching the surface (Figure 16, top), gestures moving back and forth between touching and not touching (e.g., tapping), gestures carried out entirely just above the surface (Figure 16, middle), and gestures performed above about 5cm from the surface (Figure 16, bottom). First, deictic gestures that touched the table surface were common, and almost always occurred when speakers were being more specific, more confident, and more precise. Second, gestures that moved between touching and just off the surface (e.g., tapping or bouncing actions) were also common, and were used for emphasis and to indicate a series of locations along a path. Third, gestures that hovered just above the surface, in a layer approximately 2-5cm above the surface, indicated less confidence or familiarity, or, occasionally, indicated areas rather than points or paths. Gestures above about 5cm were used to indicate reduced confidence, larger areas of the map, out-of-reach locations, or locations that were off the map. In a few cases, stroke gestures used height variations to represent variations in height in the real world. For example, one participant moved his finger in an arc while going "over the river" on a bridge.

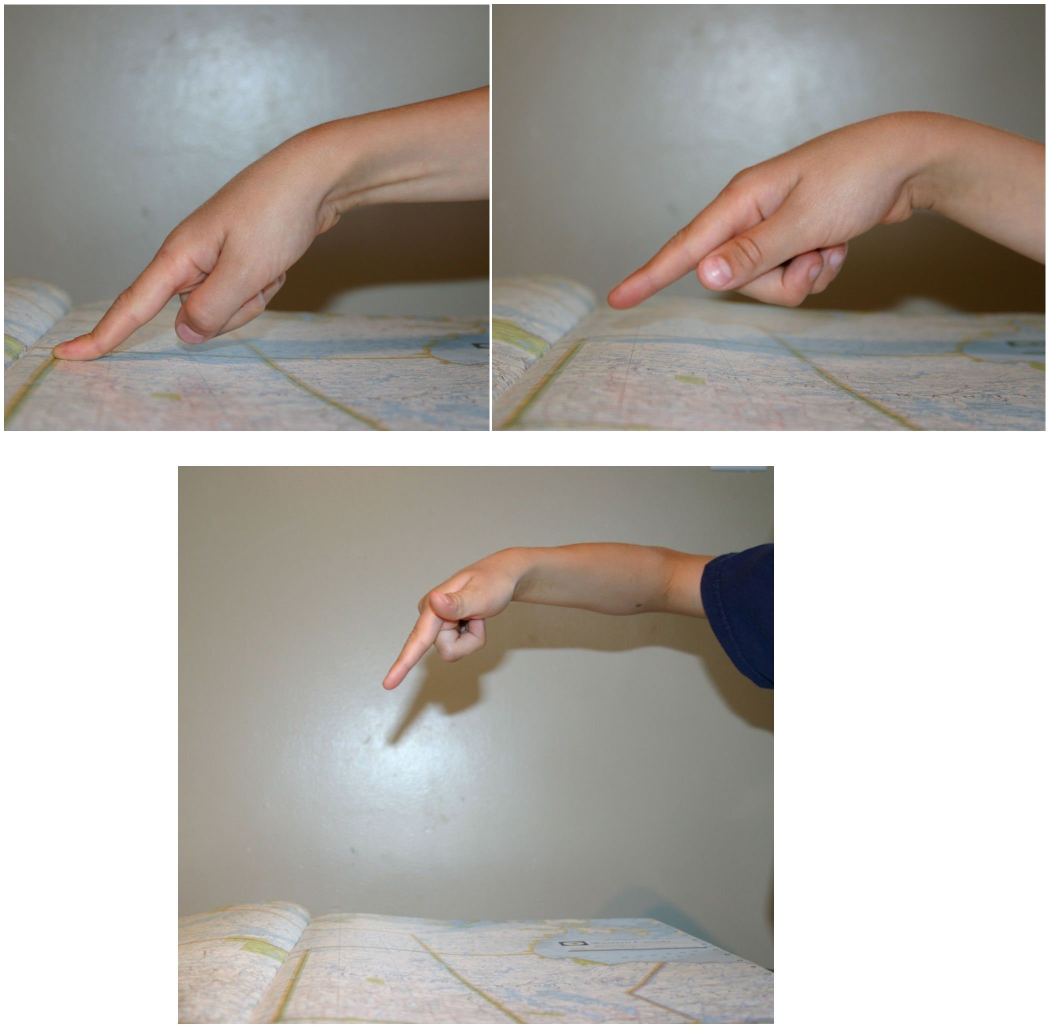


Figure 16: Three of the four behaviours in layers of height. Above, on the surface; middle, just above the surface; below, well above the surface.

The height of a deictic gesture is complex, however, and not always easily represented. As with other components of deixis, height has a wide range of context-sensitive semantics. For example, in the context of a hesitation atom, height above the surface means that the gesturer is uncertain and not ready to engage in specific deixis. However, a similar gesture presented as a series of strokes and points might be identifying paths, areas, and points on the surface. In this case, the height of the gesture above the table can indicate larger areas, wider paths, or large artifacts. Height is also used more frequently for secondary references, when the target of the deixis has already been referred to at least once in recent conversation. Secondary references

occur when the target has already been referenced once in a session or conversation. Secondary references often have low accuracy with respect to the target of the gesture, since the target is already assumed to be understood by the participants [97].

Although the absolute height of gestures varied depending on the size of the individual, whether they were seated or standing, their proximity to the surface, and other ergonimic considerations, the observed height layers were consistently present and highly similar between situations and participants. The exception to this was the top layer, where standing participants or participants using a vertical surface were able to make considerably higher gestures than others. The higher gestures did not have any different meaning, the participants simply used the space available to them to express the same qualities of communication.

Height was also used in a few other ways: as a component of 'ray-casting' gestures that pointed to out-of-reach objects; to mirror variations in height of the objects represented on the map; as another way of emphasizing a location; or to show variation in the precision of a location. Given the wide range of semantics for height, it seems clear that some representation of height is an important requirement for remote embodiment techniques – particularly to show whether a deictic gesture is touching the surface or not.

What Additional Physical Characteristics do Gestures Have?

In previous studies, a wide range of behaviours and subtleties have been observed in deixis (e.g., [2]). Much of the variation in behaviour occurs once the gesture has reached its target, rather than during the approach or retraction, which generally involve standard movements. Therefore, deixis on surfaces can also be characterized by the possible movements that can be made on the target. Understanding what kinds of gestures are available and when they are used (e.g., in conjunction with certain kinds of targets, or certain modes of speech) can assist in filtering or augmenting this component of gestures in distributed environments.

There were three additional characteristics that were seen several times during the studies. This is not an exhaustive set, but serves to indicate the range of additional possible behaviours that can be observed in a deictic gesture. In each example characteristic, a range of meaning is provided by small movements of the arm and hand, or changes in pressure on the surface after a deictic target is reached. These variations in movement or pressure do not change the target of the deixis, but rather provide emphasis or convey qualities that can only be determined through the verbal record. The three characteristics we observed are width variation in strokes, wiggle motions in pointing, and pressure in pointing.

- *Width variation* is variation in the movement along the plane of the surface that does not otherwise interfere with the target of the gesture. For example, width variation on a stroke atom could be sinusoidal movements that range along the axis perpendicular to the movement vector (i.e., a snake-like gesture, rather than a straight line).
- *Wiggles* are movements of the hand or arm that do not change the target pointed to by a finger or hand. A wiggle variation during a point atom that touches the surface would leave the pointing finger in place while moving the hand or arm.
- *Pressure variation* is a change in the pressure applied by a finger or hand on the table surface (presses occur only when the pointing hand is touching the surface). Pressure changes can be visually detected by the observer through subtle differences in the posture and appearance of the hand and finger (e.g., bending or colouration of the pointing finger).

The use of width variation and wiggling can occur on any of the indicative deixis atoms in the air or on the surface, but pressure variations are limited to deixis atoms located (at least briefly) on the surface. All three variations were found during both laboratory studies and in the

field. In general, wiggle and pressure variations, although present, were noticeably less common than width variations, with some participants using neither. However, visual indications of pressure are subtle, and the video recordings may have been insufficient to permit the correct identification of all instances of pressure variation. Where it was possible to view wiggle and pressure variations, they were used for emphasizing a gesture, or accompanied verbal attention-drawing. In a few cases, wiggles and pressure changes were used in the same way as hesitation atoms – that is, as a 'stalling tactic' while the next location was identified. In a few other situations, wiggle variations were performed when the pointing finger or hand partially occluded an area of the surface that the speaker needed to see. The occlusion avoidance that resulted involved a wiggle as the participant moved his/her hand from side to side while peering at the surface.

While existing embodiments already do a good job of representing width variations in atomic gestures, wiggle and pressure variations are more complex. Wiggle variations are expressed through changes in the angle and/or position of the hand without a major change to the primary point of contact with the surface; this means that representations showing more of the hand and arm will naturally represent these kinds of motions, but single-point representations such as telepointers will not.

Pressure shows few external signs other than a change in the posture or colour of the hand at the point of contact (and only if enough pressure is being applied). Even with sufficiently high video fidelity, realistic embodiments are unlikely to do well conveying information about pressure, and abstract embodiments do not represent pressure at all. However, the fact that pressure variations only occur on the surface suggest that sensing technology in the surface itself could be used to improve the expressiveness of an abstract embodiment. Some digital tables

naturally sense pressure and others detect secondary features of pressure (such as differences in the contact size for Frustrated Total Internal Reflection (FTIR) tables as people push harder), and a visualization of this sensed value could easily be added to the embodiment of a remote participant.

Conclusion

When people perform deixis over surfaces, they create widely varied and complex gestures. However, most of the gestures can be classified as some form of index-finger pointing. The complex movements that occur during deixis can be divided into seven atomic movements, of which three are very important to show: point, path, and contour. There are also numerous small variations in the way that gestures are made, although the most frequent and noticeable are the wiggle, changes in pressure, and a width variation. All of these convey subtle information about the gesture meaning.

Finally, the height of a gesture above a surface is a critical piece of information for understanding many qualities about the gesture, such as the specificity, the emphasis, and the confidence level intended to be conveyed. Height is also used to disambiguate similar gestures, mirror real-world changes in the height of objects on a map, and to assist with the indication of distant objects. The following chapter will discuss how these findings, especially how the height of a gesture can affect its interpretation, can help describe a design space for embodiments intended to show deixis to remote users.

ARM EMBODIMENTS FOR DEIXIS: A DESIGN SPACE[2]

Groupware solves the problem of representing users and their actions in remote settings with embodiments: representations of users, their characteristics, and their actions in and around digital (or digitally-enhanced) spaces. Embodiments in digital systems were studied in detail by Benford *et al.* who argued that, “the inhabitants of collaborative virtual environments (and other kinds of collaborative systems) ought to be directly visible to themselves and to others through a process of direct and sufficiently rich embodiment” [23]. Benford introduced embodiments as potentially complex and nuanced representations but typical embodiments use simple representations, such as telepointers, which often show only a few aspects users or their actions (e.g., [17], [36], [47]).

Embodiments have been used to represent gestures in distributed settings. Hayne *et al.* showed that very simple embodiments are sufficient for allowing distributed users to express a variety of common gestures [99]. More complex embodiments, such as VideoArms, have also been shown to be effective for gesture representation [14]. However, there has been no comprehensive exploration of the design space of embodiments, particularly with respect to how well deictic gestures can be represented.

Designing an embodiment requires answering two questions:

1. What information needs to be conveyed?
2. How can this information be conveyed in an embodiment?

The following sections examine these questions in more depth, suggest how they might be answered, and explore how these questions can be modified to inform the design of embodiments that show deixis over surfaces.

---

[2] Material from this chapter was published as “Evaluating the Effectiveness of Height Visualization for Improving Gestural Communication at Distributed Tables” at CSCW 2012 [98].

An embodiment shows specific information about a user and her actions to other users in distributed environments. In some cases this information is simple, such as the location of a computer cursor on a screen. In other cases, the information is more complex, such as the location and shape of a user's body. Conveying all available information is not always possible or desirable, since resources for transmitting, expressing, and understanding the information are often limited. The limits may be technological, such as bandwidth or computational power, or biological, such as cognitive and perceptual limits. Sometimes, representing all of the available information may be undesirable, such as when privacy is important or when attention needs to be drawn to particular aspects of interactions. Other times, presenting one kind of information may obscure another kind of information. For example, showing large embodiments may obscure the workspace or cause embodiments to overlap more often.

The information to be contained in an embodiment is selected on the basis of its importance to the domain and activity supported by the distributed application. This selection is highly dependent on the circumstances and requirements for the collaboration. For instance, showing user's faces might be very important in diplomatic collaboration, but it is less important during mixed-ecology (where digital and physical elements are both part of the workspace) instructional tasks (as in [75]). For such instructional tasks, showing high detail representations of the three-dimensional movement of hands and arms is much more important than in other tasks (e.g., a collaborative sketching task). Discovering what information is important to include in an embodiment can be done through careful observation and characterization of the movements and behaviours of users when they perform the task in collocated settings.

Information can be endcoded visually using techniques developed by Bertin [100] and others (e.g., [101]), who described how designs can use visual patterns to show different types of

information. (The representations discussed in this chapter are visual, although there are other methods, such as auditory or tactile, of embodying users.) For example, continuous data is less effectively encoded through changes in the shape of an object than it is by changes in the size of an object. In this example, both size and shape are *visual variables*. Embodiments represent information about people and their activities by using a variety of visual variables in the components of the embodiment. Given a particular piece of information to encode (e.g., the height of a hand above a table), the embodiment designer must select one or more visual variables appropriate to the data type (e.g., for the continuous variable of height, changes in opacity and size may be appropriate).

The movement of a gesture, or position of hands and arms over time, is difficult to encode with many of Bertin's visual variables because time is continuous and can have a wide range. Showing movement using the visual variable of 'position' is the most common solution. Position on an axis almost universally represents time in time-dependent visualizations, from time series graphs to the Lund School aquariums (as in [102]), spiral graphs [103], time wheels [104], calendar views [105], ThemeRiver [106], and many others (see [107]). However, gestures take place in three axes of position, so encoding movement must be done without colliding with representations of the current position.

In many cases there may be constraints to representing information as an embodiment. These constraints can be domain-dependent, such as visually cluttered and complex workspaces; situation-dependent, such as users who are colour-blind; or global, such as cognitive or technological limits. Understanding how these constraints mediate the range of what information can be conveyed and how provides clues to how an embodiment might be designed.

Although information may be encoded with visual variables, the encodings must still be transformed into a design. For example, if opacity and size are used to encode the height of a gesture, a design that incorporates those visual variables might represent them as a shadow (which decreases in opacity and increases in size as the gesture gets higher). There might be many possible designs that make use of the same visual variables, however, some will be more successfully interpreted by users. The success of the design (as defined by how closely the intended embodiment interpretation matches people's actual interpretation) can be measured by watching how people use the embodiments in distributed settings and by running controlled user tests.

The following sections answer these questions for deictic gestures. Section 0 discusses what information about deixis needs to be conveyed in embodiments for supporting distributed collaboration. Section 0 explores how gesture characteristics can be conveyed using abstract and/or real representations. Section 0 examines the representation of gesture height and how some visual variables are better than others for showing different kinds of height. Next, Section 0 shows how temporal information about where the gesture has been in the past can be conveyed through temporal traces. Finally, Section 0 introduces a set of candidate designs for showing three dimensions of a pointing gesture and describes how these designs were developed through pilot experiments.

What Information About Deixis Needs to be Conveyed?

Chapter 3 introduced several important components of deictic gestures: the morphology of the gesture, the movement of the gesture, the height of the gesture, and subtle variations in the way gestures are executed (such as wiggles, or pressure). For embodiment design, these components can be the foundation of three questions:

1. Where is the gesture?

2. How did the gesture move?
3. What does the gesture look like?

The location of the gesture is critical for understanding deixis because, unlike other forms of gesturing, deictic gesture always has one or more targets: without a clear understanding the location of a pointing gesture, it is impossible to know the target. The location of a gesture is complicated: it can be expressed in relation to the user, the workspace, the environment, or even other users. It can also be expressed in terms of a point, an area, or a volume. In particular, location of gestures has not previously included much height information, but Chapter 3 identified the height of a gesture as a critical component of how deixis is communicated.

The motion of a gesture through space (or the absence of motion) is a critical aspect of a deictic gesture. Movement in the x,y plane helps differentiate between different kinds of targets (e.g., a point versus a path) and provides tools for conversational mediation, such as attention-getting. Movement in the z plane can also help differentiate between gestures (e.g., a path versus points on a path) and can also show qualities such as emphasis (e.g., in a tapping gesture). In a distributed context, because gestures happen quickly and distributed users are often not as aware of the remote environment, some historical information about the gesture – where the gesture was in recent history -- may also be important.

The appearance of a gesture -- that is how it looks to an observer -- includes the morphology of the gesture, but also its context: who is performing it, where the gesturer is standing in relation to the workspace and other users, and perhaps for whom the gesture is made. Much of the meaning of the gesture, both in explicit communication (e.g., identifying the target of the gesture) and consequential communication (e.g., the emotional state of the gesturer), is contained in the appearance.

Location

Conveying the x,y Location of a Gesture. The location of a deictic gesture can be described as a point or multiple points in two or three dimensions, or as an area or volume (summarized in Table 3). One way of describing the location of a gesture is by using the location of the tip of the primary pointing finger in two or three dimensions relative to the workspace (as described in Chapter 3). Whole hand or arm gesturing, however, is more difficult to show, since a single point is insufficient for expressing the position of the hand and arm. For deixis, there is an added complication: the target of deixis is less certain and less precise with only a single point of information, especially if expressed in only two dimensions.

In this context, designers must consider the position of the hand relative to the workspace. Multiple points, such as fingertips, palm centre, and elbow, can provide some of this information, suggesting the location of a gesturer's body. Multiple points can also provide a more accurate indication of the target of deixis if the target is large, an area, or a more than one artifact, rather than a single, small point.

Conveying the Height of a Gesture. There are two important characteristics of height in deictic gestures: the difference between touch and hover, and the relative height of gestures above the surface. As noted in Chapter 3, the small difference between touching the surface and hovering can indicate large variations in specificity, confidence, and emphasis. The relative height of gestures above the surface can also convey specificity, confidence, and emphasis, but also permits the identification of out-of-reach targets, differentiating between different kinds of targets (e.g., a point and an area), and the identification of targets off the edge of the surface or workspace artifacts (e.g., just beyond the extent of a map).

Conveying Location as Areas or Volumes of Hands and Arms. Showing the boundary of a user's hand and arm (the area or volume of the workspace it occupies) provides the most

location information. In some cases, such as when part of the hand may occlude someone's view of a pointing finger, this may be less desirable. However, a volume or area representation of a hand and arm allows the expression of more complex deixis (such as gestures that use the palm), and may convey subtle changes in posture and position relative to the location of the workspace that help understand deictic gestures. For example, during the studies described in Chapter 3, participants were observed moving their hand and arm without displacing the tip of their finger. They did this to avoid occluding a target or potential target with body parts not actively engaged in pointing. If a remote user could see only a few points representing the position of the hand, the same hand shape and location might appear to be indicating other targets with the finger or wrist. Only with a volumetric representation of the whole arm and hand does the purpose of the gesture become apparent. In this situation, it might also be valuable to see some portion of the body and head.

As mentioned above, representing the height of important body parts plays a major role in facilitating communication between distributed users through deictic gestures. However, the difference between representing the arm and hand as a two dimensional area or representing it as a three-dimensional volume can be computationally expensive and increase bandwidth requirements in distributed settings. These problems will be discussed in detail both in this chapter and in Chapters 6 and 7.

Table 3: Information provided by different kinds of location information for hands and arms.

Location Information	**Information Provided**
2D Point (x,y position)	2D position of pointing finger; very low precision for possible targets; no information about body position
3D Point (x,y,z position)	3D position of pointing finger; low precision for possible targets; no information about body position
Multiple 2D Points	2D position of hand and arm; low precision for possible targets; minimal information about body position

Multiple 3D Points	3D position of hand and arm; medium precision of possible targets; minimal information about body position
Area	2D position of hand and arm; low precision of possible targets; some information about body position
Volume	3D position of hand and arm; high precision of possible targets; some information about body position

Movement

Chapter 3 characterized a gesture's movement through space using small, atomic blocks of movement. This work helped identify what kinds of movement convey information during deixis. There are five key characteristics of movement that should be identifiable by remote users, suggesting a minimum level of embodiment detail to faithfully represent deixis (see Table 4):

1. The x,y movement of the gesture through the workspace. This conveys critical information about the meaning of the gesture, in particular, the intended target and type of target (e.g., area or path).
2. Representing the movement before the deixis reaches its intended target, including small preparatory movements. Fraser *et al.* [34] found that showing the approach of a user's hand to a surface was critical for focusing the attention of distributed users. Preparatory gestures also play a role in conversational mediation and attention-getting [96], [108].
3. Changes in height, such as those that occur during tapping, are of particular importance in conveying information about the specificity, confidence, and emphasis of the gesture. Changes in height also help differentiate between common gestures, such as a paths or points along a path.
4. Small movements, such as wiggles, can contain valuable information, such as lower levels of certainty, or higher levels of emphasis. These movements can be very subtle,

and require a high level of detail in both capture and reproduction if they are to be usefully conveyed to users.

5. Although not, strictly speaking, movement, pressure applied to a surface does convey information during deixis, such as emphasis, or aspects of emotional state of the gesturer.

As with the other characteristics of a gesture, low levels of detail are sufficient for conveying enough information about deixis that the gesture will likely be understood (i.e., the target correctly identified), but additional detail will provide valuable and meaningful communicative information.

Table 4: Gesture movement characteristics and the information they convey.

Movement Characteristics	**Information**
Movement in the x/y plane during main part of deixis	Core communication; the target of the gesture.
Preparation atom	Attention-getting; conversational mediation
Changes in height	Differentiation between similar gestures; varying levels of emphasis, confidence, specificity
Small movements (e.g., wiggles)	Varying levels of certainty or emphasis
Pressure	Emphasis or emotional state

Morphology

Chapter 3 describes how most critical information contained in deictic gestures is conveyed through the shape of the hand, in particular through one or two-finger pointing. Palm orientation and the posture of unengaged fingers was also shown to convey deictic information, although this information is more nuanced and less critical to the interpretation of the reference. More extensive variations in morphology have been found in other studies [11][8], especially when the gestures include narrative gesticulations.

Low levels of detail in replicating the morphology of a gesture have been shown to cause some problems in interpretation [14], but the success of telepointers in conveying gestural information [13] suggests that choosing which morphological element to represent is more important than high levels of detail. Deixis can be interpreted with relatively low morphological detail, so long as the tip of the index and/or middle fingers are visible – thus the success of telepointers in conveying deixis. However, people vary in their choice of primary pointing finger, potentially causing problems for systems that track only one finger for deixis (see Chapter 3).

Narrative gestures rely on more complex, whole-hand morphologies [9] and therefore a higher corresponding level of detail. Higher levels of detail will also provide more nuanced information about the communicative intent of the gestures, but the work in Chapter 3 shows that the baseline morphological detail for conveying most occurrences of deixis is very low. Table 4 summarizes how varying levels of morphological detail can convey gestures and what challenges there are in using particular detail levels.

Table 5: Changing the level of morphological detail also changes what kinds of gestures can be represented.

Morphological Detail	Effectively Conveys	Challenge
Low (e.g., tip of finger)	Single-finger pointing	There are individual variations in pointing technique
Medium (e.g., major points of hand and arm)	More complex deixis, simple gesticulation	Fails to show palm orientation and other subtle characteristics
High (e.g., full colour images)	Narrative gesticulation, subtle qualities of deixis	Costly to track, reproduce, and transmit over networks

How can Deictic Gestures be Conveyed in an Embodiment?

Each characteristic of a gesture (i.e., location, movement, and morphology) can be represented using realistic, abstract, or hybrid representations. These different styles of

representing information can vary in effectiveness for representing gesture characteristics. An in-depth examination of how realistic, abstract, and hybrid representations have been used in previously developed embodiments can be found in Appendix C.

Embodiments may use both abstract representations, such as symbols, and more realistic representations, such as images, to convey the characteristics of users. Embodiments are not limited to showing hands and arms: they can also express actions and behaviours in or around a shared workspace by enhancing the visual representations of users; provide information about the state of users, such as relationships between users, awareness, or emotional state; and the characteristics of users, such as their domain expertise, demographic information, or geographical location [36]. For instance, if a designer wishes to show that a user has been touching the surface for several seconds or that a user is angry, she must include extra information not available in a completely realistic representation of hands and arms – this necessarily means that the embodiment is no longer strictly realistic.

Realistic representations of hands and arms serve as effective tools for conveying many of the characteristics of a gesture. Where realistic representations fail, more abstract representations can accommodate, providing a range of options for conveying gesture characteristics. Each gesture characteristic can be expressed with either abstract or realistic representations. Embodiments that use both abstract and realistic styles are hybrid, a portion of the design space that is largely unexplored.

Candidate Designs

The following sections describe designs based on the above techniques that were evaluated in pilot studies. Many designs incorporated one or more features because the same feature could not apply to height, touching, and traces. During pilot studies, designs were refined by adding additional elements to dual-encode information or to capture an aspect of movement

above the surface in deixis that was not able to be expressed using a single technique. For example, although both size and transparency were successful in showing absolute height, participants much preferred designs where the two techniques were combined.

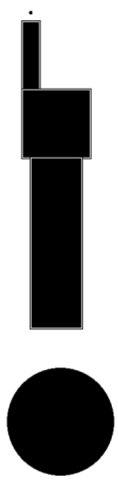
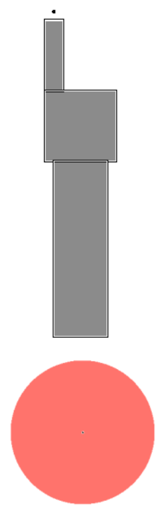
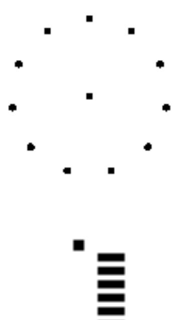
Telepointers are a simple and well-researched benchmark against which to evaluate new abstract representation-based embodiments. As a result, arrows, crosshairs, and other, similarly specific graphics were used for showing touch instead of wider, blob-like representations, like those extracted from FTIR tables. The goal of a touch representation is to show, as accurately as possible, the point of contact. As a result, the designs were optimized for a high level of accuracy in representing touch locations.

Although Chapter 3 describes how people use different layers in the above-the-table space, these differences are not uniform, so these designs represent height mostly with a continuous visualization, but occasionally with some added elements that can allow people to make use of different layers.

There were three temporal traces designs available (ripples, telepointer traces, and none), six touching designs (four shape changes, a colour change, and none), and 14 designs for showing absolute height. These can be seen on

Table 6: one shape design; three colour designs; one transparency design; two size designs; two size designs with a transparency secondary encoding; one size design with a displacement secondary encoding; two designs where graphical elements were added; and two designs where graphical elements were added that used colour as a secondary encoding; and a plain telepointer arrow as a benchmark.

Table 6: Above the surface encoding designs. Not all secondary codings were explored.

Primary Height Encoding	Primary Encoding Version	Colour Design	Transparency Design	Displacement Design
Shape				
Colour				
Transparency				
Size				
Adding Graphical Elements				

Pilot Studies of Candidate Embodiment Designs

Design candidates were pruned and refined using informal pilot studies. Six unpaid participants were recruited from the Department of Computer Science graduate student body at the University of Saskatchewan. Participants were in graduate studies with an emphasis on Human-Computer Interaction or Software Engineering, and thus had some experience with design concepts and collaborative work.

Methods. Participants performed informal evaluations of each candidate design over a 30-minute session. The evaluations took the form of researcher-guided open ended interviews (initial questions can be found in Appendix A). Each participant examined each design for an unlimited period of time and then was asked to comment positively and negatively on the design between viewings. Participants were allowed to revisit previously viewed designs at any time. Participants sat or stood at a large table as they wished. Each candidate design was shown on the table's surface with a white background, initially, that could be changed to high contrast satellite map images at their request.

Rather than create static candidate designs, software was written to support alteration of design features while using the embodiments. With this software, pilot participants could change the touch visualization, absolute height visualization, or temporal traces independently while continuing to experiment with their gestures. This allowed pilot participants to rapidly move between designs while performing similar gestures, permitting a more structured exploration.

Participants used keyboard keys to change between backgrounds and prototypes. They manipulated the prototype's location and appearance using a Polhemus Liberty stylus. The prototypes represented the location of the tip of the stylus in three dimensions. Participants were

asked to imagine that the prototype was representing the tip of their finger to someone in a remote location.

Results. Although there were some individual differences, most participants had clear preferences for using cursor images as historical traces, a shape or colour change for indicating touching, and size and transparency for showing the height above the surface. Shape changes were difficult to differentiate for absolute height for two reasons: increases in the number of sides when the polygon had few sides appeared to represent larger movements than was the case; and changes when the polygon had many sides were difficult to identify, especially when the entire gesture occurred well above the surface. Reversing the visualization, with fewer sides representing higher gestures, created unacceptable interpretation challenges for all of the participants.

Participants experienced similar problems with colour changes for absolute height representation: it was challenging to find a natural mapping (even with the grey scale design) and identify small changes in height.

Adding graphical elements was more preferred by participants, although they found that the designs felt abstract and that there was a fairly large cognitive load in trying to both identify the location of the intended target and interpret the height information. In this respect, variations in size worked particularly well, allowing people to interpret the location of the target as an area rather than a point when the gesture was higher. Occlusion was identified as a potential problem for size alone, but the combined transparency and size design worked particularly well, as did the combined transparency and displacement design. These designs were further refined based on participant comments and combined into a single embodiment with size, displacement, and transparency encodings for changes in height (see Section 0).

One unanticipated factor that emerged was visibility: participants could vary the background between a plain white background with abstract shapes and several satellite maps. When using the map backgrounds, participants complained about the visibility of many of the embodiments and recommended that they be altered to include vibrant colours with high-contrast stroking (see Figure 17).

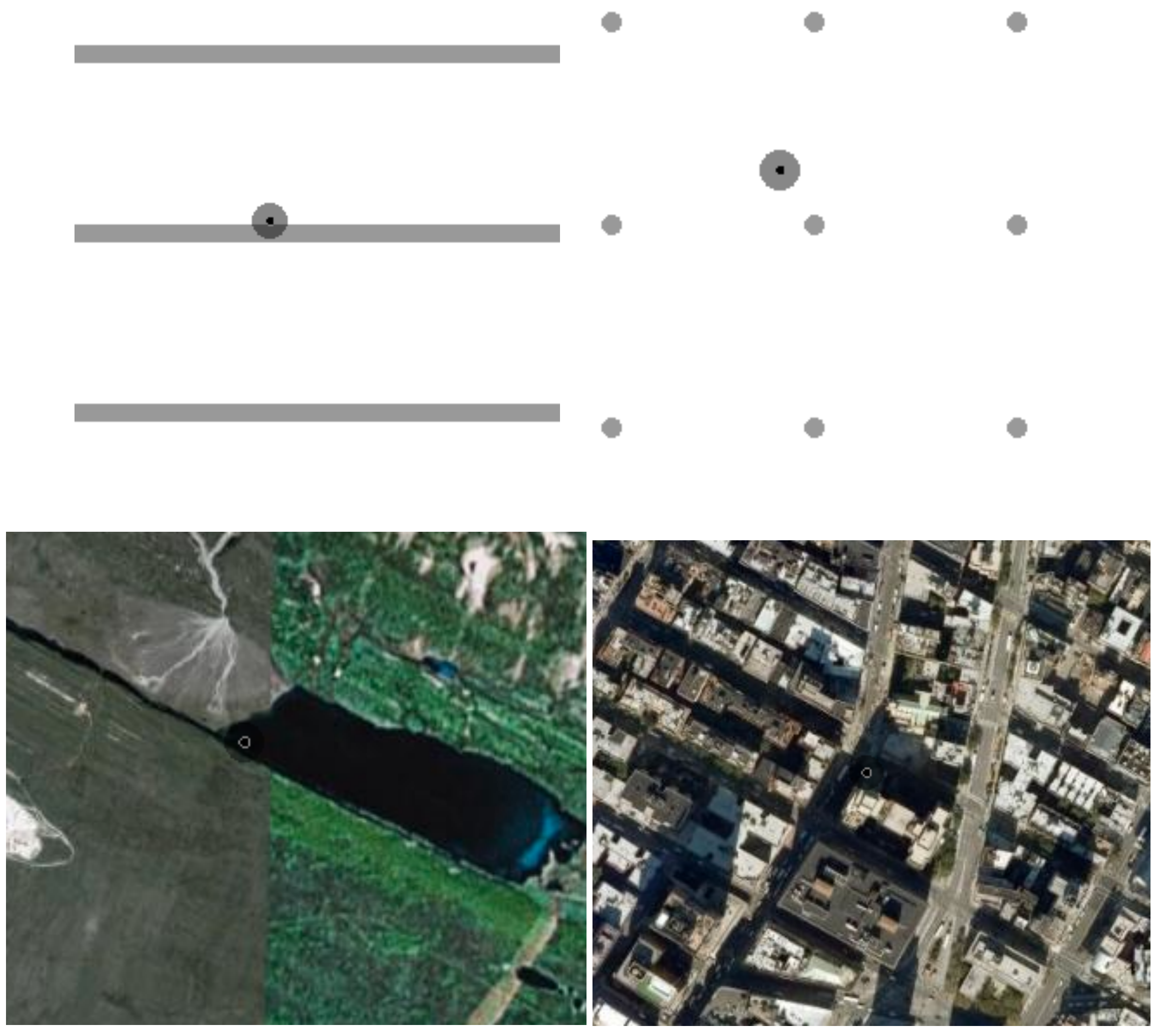


Figure 17: Although many embodiments were visible on some backgrounds (above), their visibility was limited on more complex backgrounds (below). Note that the white stroking around the black dot means that the location of the embodiment is visible, even if the transparency effect designed to show gesture height is uninterpretable.

Refined Embodiment Designs for Showing Gesture Height

As with the candidate designs, the refined designs for representing gesture height in embodiments are separated into three components: representing the absolute height of the gesture, representing touches, and representing the immediate past of gestures. Each component is designed to allow the other components to be experimentally varied independently.

Representing Absolute Height. The above-the-surface embodiment design uses shape, size, transparency, and displacement to encode height. Colour is used to improve visibility. Height above the surface is represented as an ellipse that increases in size and becomes more transparent as the gesture moves upwards (see Figure 18). These visual cues are compatible with the idea that higher gestures are less specific and that higher gestures might need to appear less engaged (to the workspace) than lower gestures [15]. The ellipse is multi-coloured, increasing its visibility on a variety of backgrounds (see Figure 19). The representation also includes a temporal trace (in order to emphasize high-level path gestures) by fading out each drawn ellipse after one second; the effect is one of a series of shapes left behind in the trail of the cursor.

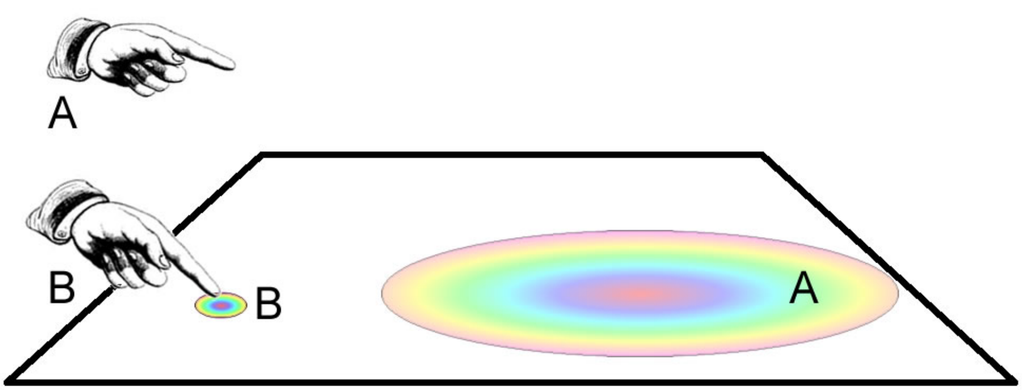


Figure 18: Absolute height visualization. A high gesture (A) with the large, transparent, displaced ellipsoid and a hovering gesture (B) with the small, solid, sphere.

The change between hovering gestures (just above the surface) and higher gestures is subtly reflected by changing from a circle (the hovering representation) to an increasingly narrow and long ellipse (for gestures further above the surface) (see Figure 18). As the gesture increases in height, the ellipse becomes narrower and longer, and the centroid of the ellipse moves away from the user (i.e., as a shadow would when cast from a light behind the user; see Figure 19). These effects are shown only when the gesture rose more than 5cm above the table's surface. The design assumed that users would be seated at pre-determined, fixed locations, allowing the displacement and shape change to be fixed with respect to a given side of the table.

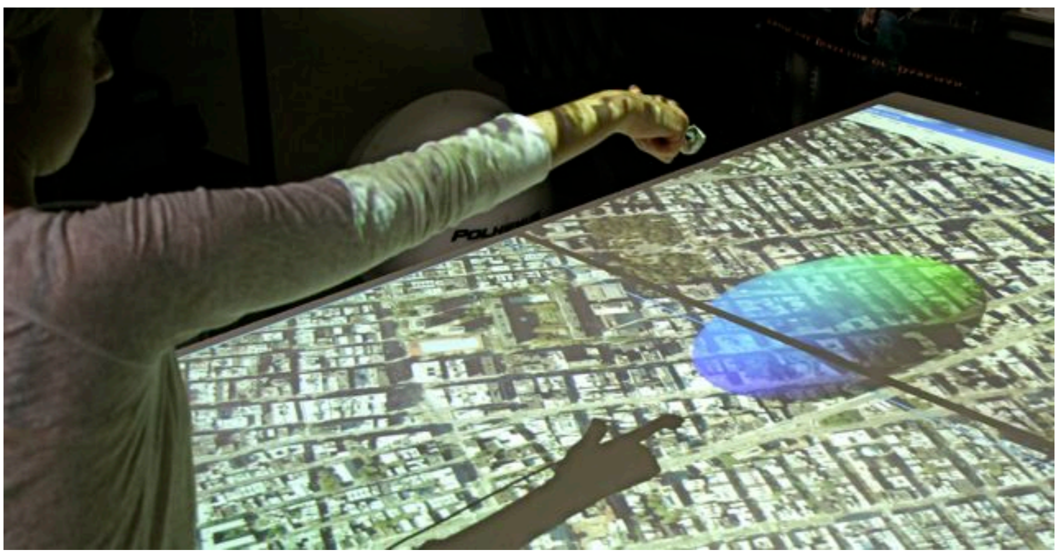


Figure 19: Simulated ray-casting with the ellipsoid embodiment. N.B., the shadow is an artifact of the overhead projection and is not part of the embodiment.

Representing Touches. To emphasize the difference between hovering and on the surface, the embodiment dual-encoded the touch state with shape and colour. Gestures above the surface were represented with an arrow (red and black for improved visibility on a variety of backgrounds). When a touch is detected, a crosshair shape, highlighted in green, is added under the arrow (see Figure 20).

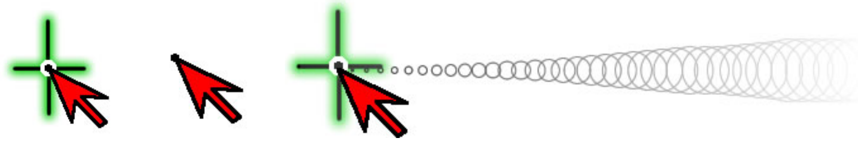


Figure 20: The table-touching designs, on the surface colour and shape change (left); above the surface, plain (centre); and combined with contact traces as ripples (right).

Contact Traces. Since the design for representing absolute height includes temporal traces, all that remains to represent to show a complete history of movement is a representation of historical contact with the surface. Rather than temporal traces, these are contact traces.

The refined designs had two variants of contact traces. The first variant used successive, expanding and increasingly transparent ripples, as in the pilot design, to show historical movement. Each ripple was actually three, concentric circles: a white circle, a black circle, and a white circle. This design noticably improved contrast against a variety of backgrounds and addressed a major concern expressed by participants during the pilot study (see Figure 21, top).

There were, however, some problems with this design (which were revealed during the first experimental evaluation, see Chapter 5). In particular, people can mistake successive ripples for individual surface touches. Therefore, the second variation shows simple sketch lines along the path of the gesture. Ripples are still shown at the point of initial and final contact with the table, which permits tapping gestures or gestures that touch the surface at a single point to retain some form of contact trace (see Figure 21, middle).

The trace lines also showed the speed of the gesture: slow touching gestures resulted in straight and parallel sketch lines, and faster gestures showed lines that were slightly rotated and slightly further apart (see Figure 21, bottom). This visualization was added to facilitate previously observed behaviour where people ‘colour in’ space with quick path gestures to indicate an area (as noted in Chapter 3).

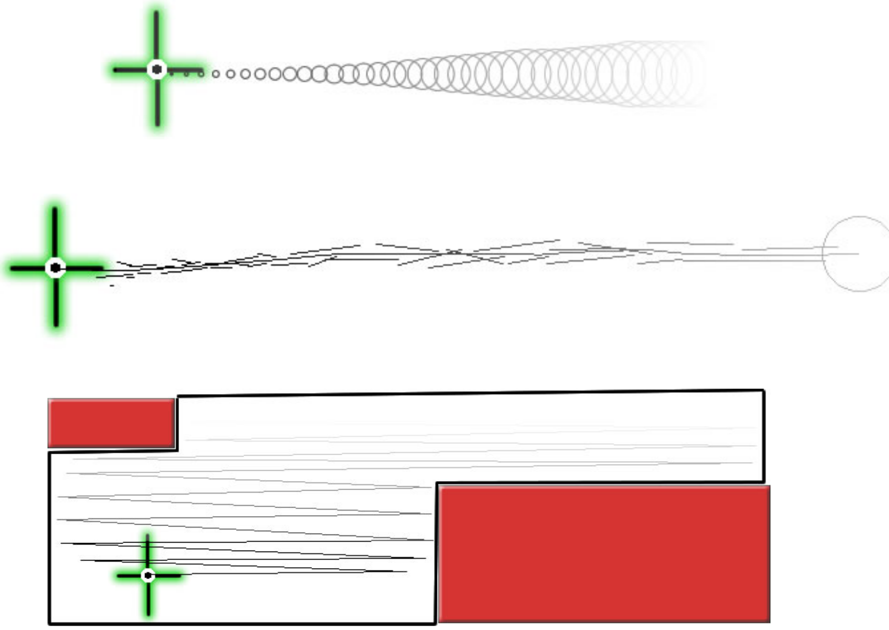


Figure 21: Top: ripple contact traces. Middle: line contact traces (slow movement with initial ripple at touch point). Bottom: line contact trace 'colouring in' an area.

Conclusion

Effective representations of deixis must show the characteristics of a gesture: where it occurs and what it looks like. This can be done by showing gesture characteristics through realistic representation, abstract representation, or with a combination of the two techniques. Both realistic and abstract representations are incapable of showing all of the aspects of a 3D gesture on a 2D display, a problem that can be solved by combining realistic and abstract representations into hybrid designs. This chapter has reviewed the many of the existing techniques for showing gestures along three axes: how well does the embodiment technique show where the gesture is, how the gesture moves, and what it looks like.

This review showed that the height of a gesture is an aspect that is not well represented by the state-of-the-art in embodiment design. The remainder of the chapter describes an

extensive prototyping process that developed a set of abstract visualizations that show both gesture height and changes in gesture height (and the location of a gesture) over time.

The next chapter describes how the designs were tested to see if they accurately represent height information to distributed users, are interpreted in the same way as gesture height is interpreted in collocated settings, and enhance distributed collaboration without interfering with natural gestural communication.

EVALUATING HEIGHT-ENHANCED EMBODIMENTS[3]

This chapter describes the experimental evaluation of the embodiment designs introduced in Chapter Four.

Introduction

Chapter Four introduced a three-part abstract represention for showing the height of gestures above collaborative surfaces in distributed settings. This chapter describes three studies performed to evaluate the effectiveness of those designs. Effectiveness is measured by answering three questions about how height visualizations are interpreted relative to the way that height is interpreted in collocated settings:

1. *Accuracy*: does height information improve people's ability to determine the type or target of a gesture?
2. *Expressiveness*: can height visualizations reliably convey qualities such as specificity, confidence, and emphasis?
3. *Usability*: can people make use of height visualizations in realistic work, and do they prefer these representations?

These design questions were answered with three studies. The first study showed that representing touch and hover significantly improve people's ability to determine both the target of the gesture and the type of gesture. The second study showed that people use height visualizations to interpret a gesture's specificity, confidence, and emphasis, and showed that these interpretations are consistent with the ways that people see real-world gestures. The third study looked at the ways in which people use height representations in realistic collaboration, and showed that people quickly make use of the additional height information in their deictic

[3] Material from this chapter was published as "Evaluating the Effectiveness of Height Visualization for Improving Gestural Communication at Distributed Tables" at CSCW 2012 [98].

gestures, that the height-enhanceded embodiments caused no new usability problems, and that height-enhanced embodiments were strongly preferred by users.

Study 1: Interpretation Accuracy

Research Questions

The first study examined how accuracy was affected by the addition of gesture height enhancements to telepointer embodiments. In particular, the study asked two questions:

1. Do touch and hover visualizations allow observers to better identify the target of pointing gestures?
2. Do touch and hover visualizations improve people's ability to identify the type of pointing gesture (i.e., whether the gesturer's intent was to point to an object, a path, an area, or none of these)?

Experimental Conditions

This study examined four embodiment conditions:

- *Arrow*: a red-outlined, standard arrow pointer (Figure 22, left);
- *Touch*: a red-outlined, standard arrow pointer that was enhanced to change shape- and colour when a gesture touched the surface (Figure 22, centre);
- *Trace*: a red-outlined, standard arrow pointer (Figure 22, centre) enhanced to add ripples as a contact trace (as in Figure 22, right);
- *Touch+Trace*: a red-outlined, standard arrow pointer enhanced with a combination of Touch and Trace when a gesture touched the surface (Figure 22, right) .

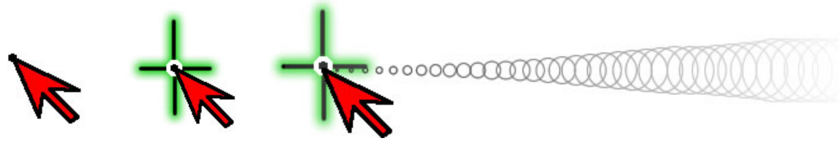


Figure 22: Left, the Arrow visualization; centre, the Touch visualization; and, right, the Touch+Trace visualization.

The Arrow visualization was used as a control condition because this representation provides the highest gesture specificity (e.g., compared to an FTIR blob), and because arrows are a standard embodiment in distributed groupware.

Tasks

The visualizations were tested on two tasks, both of which used aerial photographs as background images (see Figure 23).

Accuracy Task. In the accuracy task, participants were required to identify the targets of pre-recorded gestures displayed using one of the four visualization techniques. Each pre-recorded gesture touched the table's surface between 3 and 5 times. Touches occurred at realistic targets (such as street intersections) and the appropriate point of each visualization was used for the touch (the tip of the arrow representation, and the center of the crosshair when touching). Three to five points were used to increase the difficulty level of the task without exceeding the natural limitations of short term memory [109]. Gestures containing several targets are also realistic: they were observed frequently during the observational studies described in Chapter 3.

Gesture Type Identification Task. In the gesture type identification task, participants differentiated between three kinds of pre-recorded gestures that can look similar. The three types (see Figure 23) were *paths*, which indicated a straight or zigzagging path on the map; *single points*, which indicated single targets on the map; and *multiple points*, which indicated several points scattered across the map [6]. Gestures were carried out at a variety of heights: paths could

be on or above the table, and point gestures either touched the table or paused at each target in the gesture.

Figure 23: Examples of gesture types used in Study 1: left (blue), a scattered gesture; middle (red), a path gesture; right (green), a point gesture. An example of the background used in the studies.

<u>Proceedure</u>

Participants were seated at a large table, on which was projected a satellite photograph of New York (see Figure 23). The image was selected on the basis of its high number of closely situated potential targets. For both tasks, participants were asked to first watch a gesture's representation as it played on the surface of the table in front of them and then answer one or more questions. Participants were informed that the gesture would begin at the edge of the table immediately in front of them, move around on the table, before returning back to the approximate point of origin.

Participants completed a demographic survey and an orientation session that included previews of each visual condition prior to each task. They were also provided the opportunity to

ask questions of the researcher. The gestures played without accompanying speech. Participants could pause between trials for as long as they wished to ask questions or rest.

For the gesture-type identification task, participants viewed 28 gestures (eight single-point, eight multiple point, and twelve paths) in random order. After each gesture, they were asked a multiple choice question that asked them to identify whether the gesture they had just seen had indicated a point, a path, or multiple points. They were then asked to rate their confidence in their answer on a 5-point Likert scale, with 1 as "not at all confident" and 5 as "completely confident".

For the accuracy task, after each of the 48 gestures (12 for each condition), participants used the mouse to click on each location where the gesture touched the surface. Participants were asked to be as accurate as possible (but were also aware that the task was timed). Participants could select the points in any order (the analysis (see below) used a best-fit technique that calculated the best possible accuracy value). After selecting the points, participants rated their confidence in their selections on a 5-point Likert scale, with 1 as "not at all confident" and 5 as "completely confident".

<u>Participants</u>

Sixteen participants (12 male, 4 female, 21-45 years old, mean 27.6) were recruited from the University of Saskatchewan. Participants were paid CND$10 for partipating. All participants reported using a computer every day; 13 most commonly used a mouse as input device (two used a trackpad and one a touchscreen); and on average, the participants played video games at least once a week (range 1 (never) to 4 (several times a week) on a five point scale, mean 2.9).

<u>Apparatus</u>

Gestures were recorded by using a Polhemus Liberty sensor to track a researcher's gestures over a projected map (the same map used during the experiment). The sensor was taped

to the tip of the index finger of the researcher and the position of the sensor was logged using custom software.

Custom software was used to replay the gestures, show the embodiments, and guide the study. Gestures were recorded and played back on a large, 124cm by 185cm table with a top-down projection using two projectors showing a stitched 1024x1534 pixel image (two 1024x768 images stitched top to bottom). The resulting display had an 8.3 dots per cm (14 dpi) resolution.

Setting

Participants were seated at the long end of the table (see Figure 24) in a darkened room. Although not all of the table was within reach of the participant, gestures were designed to remain in the bottom half of the display, ensuring that the participant could easily see the gesture representation (and click on any targets in the accuracy task). The researcher was in the room and available for the duration of the study.

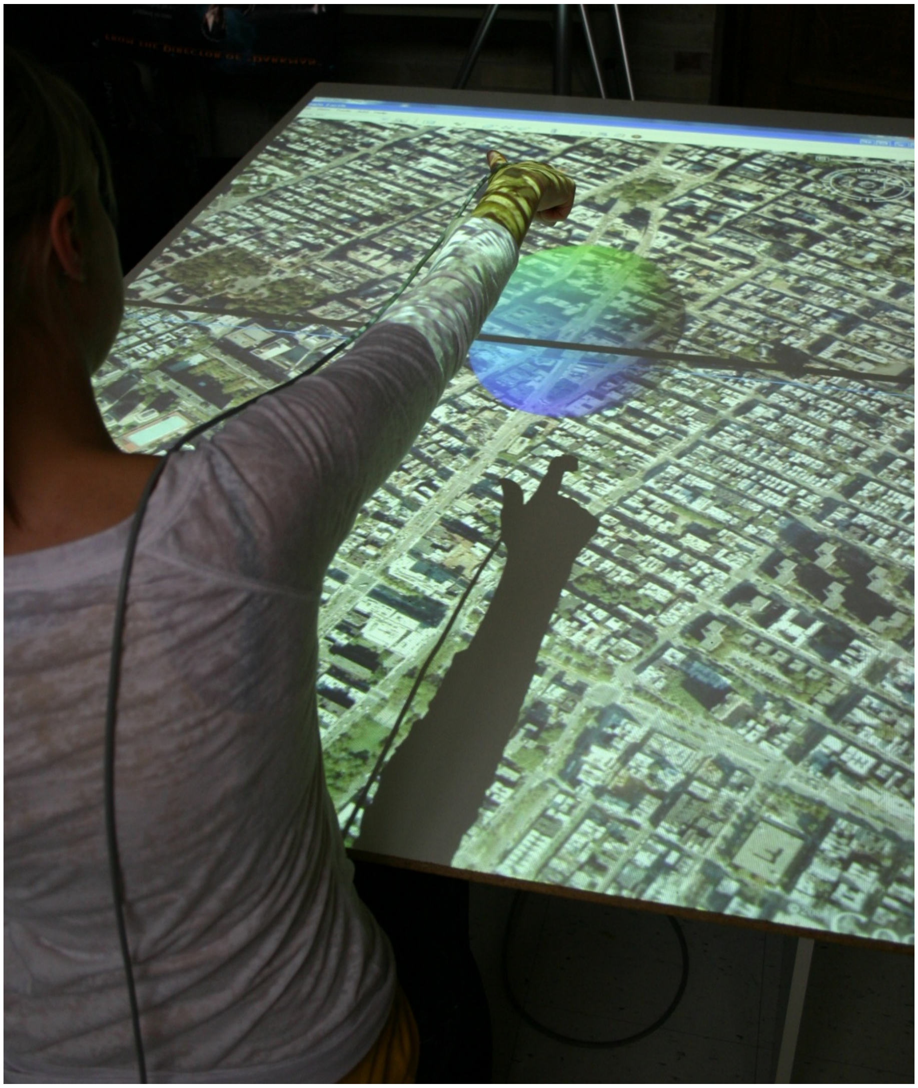


Figure 24: The top-down projected table used for study 1, study 2, and by the local participant in study 3. In this image, the user is seated at the foot of the table, as were participants in all of the studies. The black line in the middle of the table is where the stitch between the two projector images.

Study Design

The study used a within-participants design and each visualization was shown as a block with order counterbalanced using a Latin square. The identification task was presented first, then the accuracy task. The dependent variables in the accuracy task were the completion time and the accuracy of target identification. Accuracy was calculated by using an error value: the average number of pixels between the participants targets and the real targets for the trial. Since there

were always several targets in a gesture and since participants were not required to specify the points in order, error was calculated using a best-fit technique that used the lowest overall error value (in pixels) for any possible mapping of participant selections to actual targets.

In the gesture type identification task, the dependent variable was the number of correct gesture-type identifications. This variable was calculated as an error rate, the number of failed identifications divided by the total number of identifications in the condition.

For both tasks, the independent variable was the kind of visualization: System, Touch, Trace, or Touch+Trace.

Data Collection

Computer logs were used to track the participant's completion time, selected points for the accuracy task, and answers in the gesture-type task. The logs also collected participant's answers to the confidence question for both tasks. Demographics were gathered using the aforementioned paper-based demographic survey (see Appendix 1).

Results

RM-ANOVAs were used to to look for effects of the visualization on accuracy of target identification and type identification. All tests of effects used α=.05; post-hoc pairwise comparisons were carried out with t-tests using the Bonferroni correction. There were no statistically significant differences among conditions for completion time.

Accuracy in Identifying Target Locations. RM-ANOVA showed a main effect of visualization on location accuracy ($F_{3,45}$=15.76, $p<0.001$). Figure 25 displays the average error amount per target, and shows that participants were approximately 12 and 16 pixels (1.5 cm and 2 cm) closer per target, on average, to the targets with the two trace visualizations (Trace and Touch+Trace) than with the other techniques. Pairwise comparisons (see summary in Table 7) showed that both trace visualizations had significantly lower error amounts than Arrow and

Touch (all p<0.01, no other differences found). In addition, RM-ANOVA of the completion time data found no effects from different visualizations ($F_{3,45}$=0.252, p<0.859, see Figure 25).

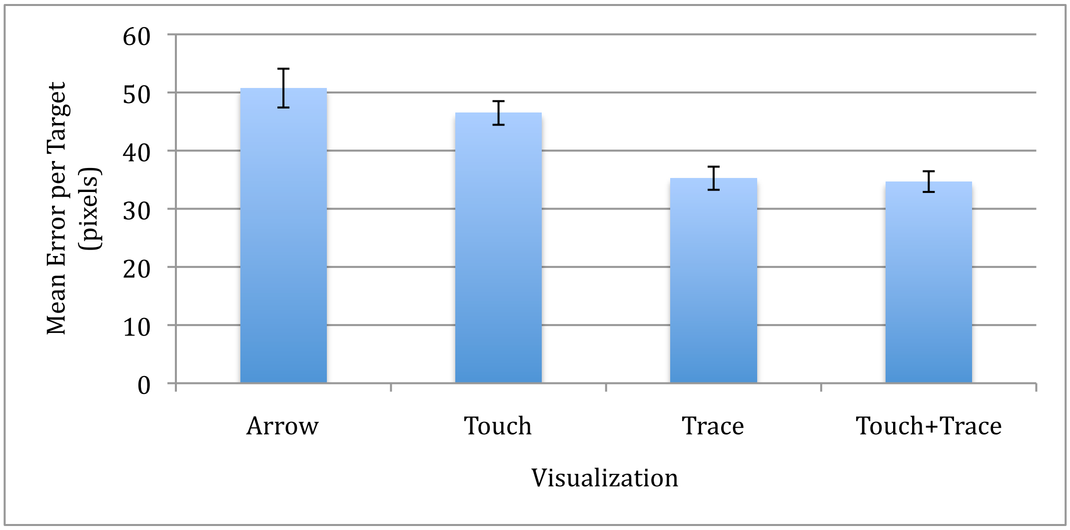


Figure 25: Mean error in pixels by visualization condition in the accuracy task.

Table 7: Pairwise T-test results among all accuracy conditions for error in pixels. Significant results were found between Arrow and Trace; Arrow and Touch + Trace; Touch and Trace; and Touch and Touch + Trace.

	Touch	**Trace**	**Touch + Trace**
Arrow	T_{30}=1.094 p=0.283	T_{30}=4.005 p<0.001	T_{30}=4.056 p<0.001
Touch		T_{30}=3.957 p<0.001	T_{30}=4.151 p<0.001
Trace			T_{30}=0.068 p=0.946

RM-ANOVA showed a main effect of visualization on confidence ($F_{3,45}$=38.38, p<0.001; Figure 26). Post hoc, pairwise T-tests found that participants were significantly (p<0.001) less confident about their answers in the Arrow condition when compared to all other conditions. There were no other significant results (see Table 8). These findings are interesting because participants were unduely confident about the correctness of their answers in the Touch condition, despite that they, on average, performed significantly worse in that condition than in the Trace and Touch+Trace conditions.

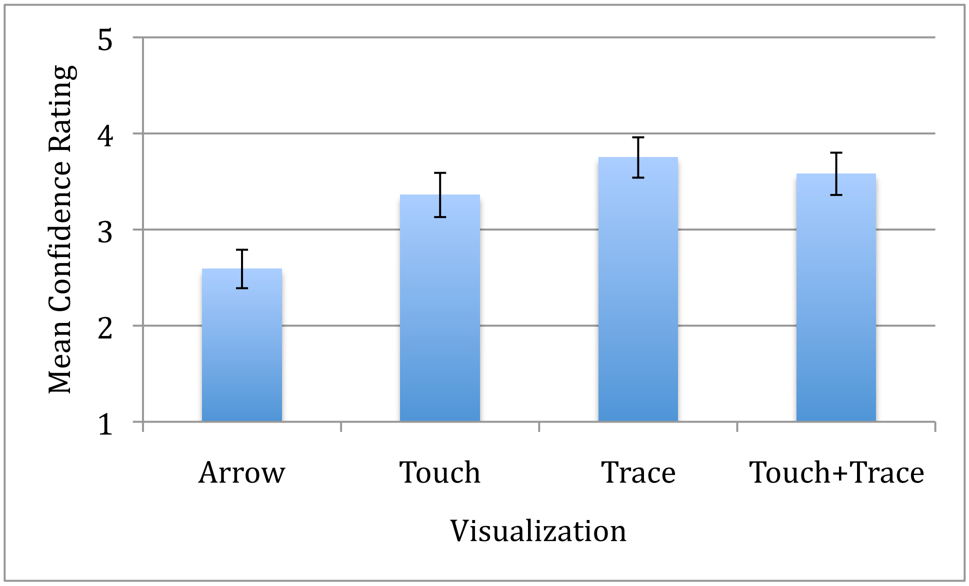


Figure 26: Mean participant rating of confidence in answers by condition in the accuracy task.

Table 8: Pairwise T-test results among the conditions for participant's reported confidence levels in their target identification accuracy. Significant results were found for Arrow and Touch; Arrow and Trace; and Arrow and Touch+Trace.

	Touch	**Trace**	**Touch + Trace**
Arrow	T_{30}=2.261 p=0.031	T_{30}=3.567 p=0.001	T_{30}=2.539 p=0.017
Touch		T_{30}=1.350 p=0.187	T_{30}=1.201 p=0.239
Trace			T_{30}=0.192 p=0.849

Accuracy of Gesture Type Identification. RM-ANOVA showed a main effect of visualization on identification accuracy ($F_{3,45}$=154.85, p<0.001). As Figure 27 shows, all of the visualizations that represented touch (either with a visual state change or with a contact trace) had significantly lower error rates than the standard arrow (p<0.001 for all post-hoc pairwise comparisons involving the arrow; no other differences found; see Table 9). The differences were substantial: the nearly 50% error rate of the arrow cursor was reduced to less than 5% for all of the touch visualizations. This finding suggests that gesture identification is improved with any touch information, which helps to disambiguate between simple movement above the table (e.g., moving to the start position of the gesture) and the gesture itself.

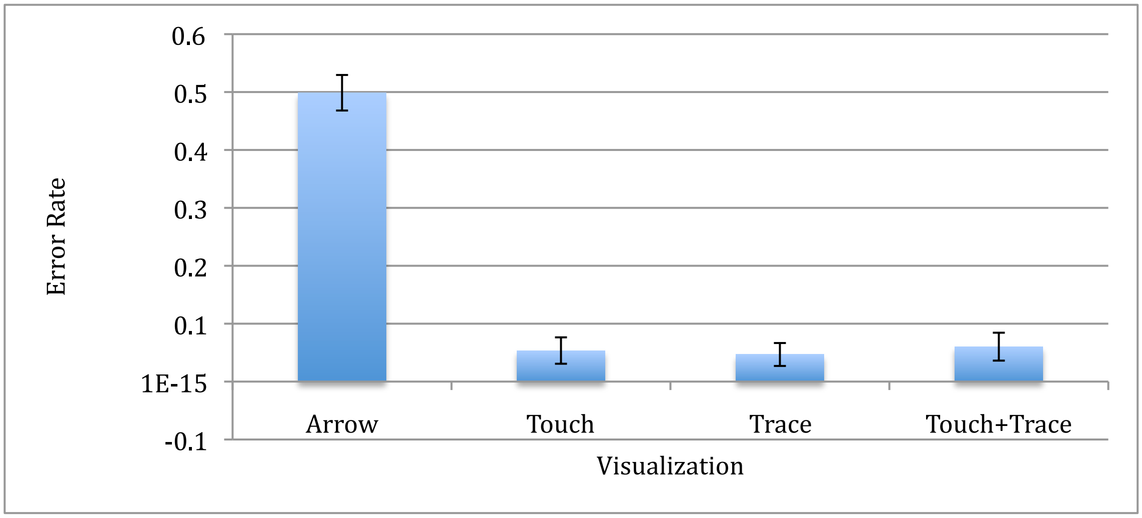


Figure 27: Error rate by visualization condition on the identification task.

Table 9: Pairwise T-test results on all conditions for the identification task. Significant results were found between Arrow and Touch; Arrow and Trace; and Arrow and Touch + Trace.

	Touch	**Trace**	**Touch + Trace**
Arrow	T_{30}=4.392 p<0.001	T_{30}=3.416 p=0.002	T_{30}=3.619 p<.001
Touch		T_{30}=0.690 p=0.495	T_{30}=0.518 p=0.608
Trace			T_{30}=0.158 p=0.876

RM-ANOVA showed a main effect of visualization on confidence ($F_{3,45}$=46.43, p<0.001, see Figure 28). Post-hoc pairwise comparisons found that participants were significantly less (p<0.001) confident about their selections in the Arrow condition compared to any other condition. No other differences were found (see Table 10). These results mirror the error rate results: participants were rightly confident (or unconfident) about their answers.

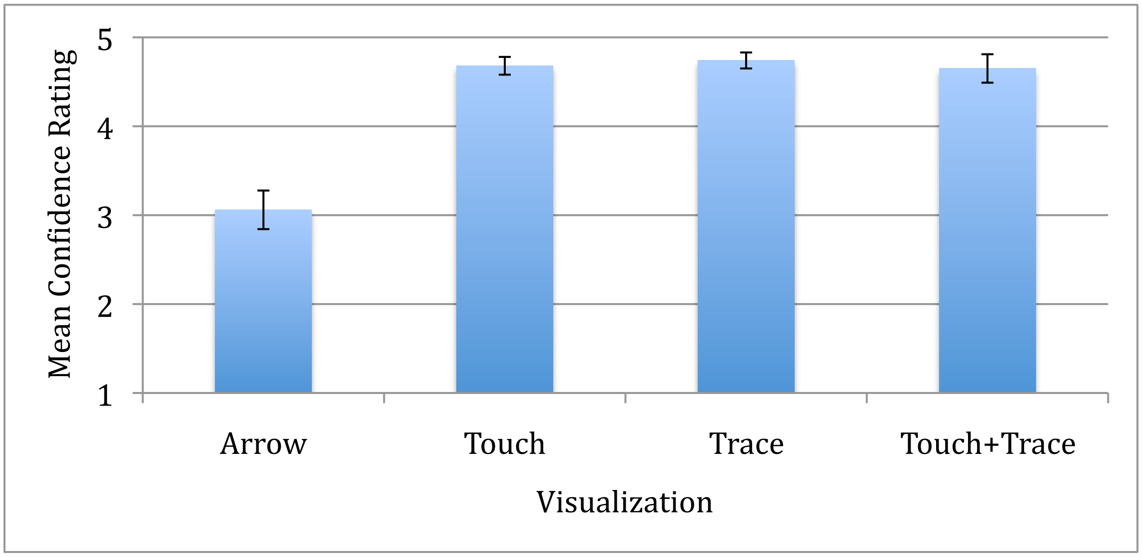


Figure 28: Participants' mean rating of their confidence in their answers by condition during the identification task.

Table 10: The results of pairwise T-tests on all conditions for participant's confidence in their answers during the identification task. Significant results were found between Arrow and Touch; Arrow and Trace; and Arrow and Touch+Trace.

	Touch	**Trace**	**Touch + Trace**
Arrow	$T_{30}=6.446$ $p<0.001$	$T_{30}=5.965$ $p<0.001$	$T_{30}=5.599$ $p<0.001$
Touch		$T_{30}=0.000$ $p=1.000$	$T_{30}=0.640$ $p=0.527$
Trace			$T_{30}=0.593$ $p=0.564$

Study 1 Discussion

Despite better accuracy rates for both trace conditions in both tasks, participants had some problems with the ripple-based contact trace visualization. Many participants mis-interpreted a series of ripples as multiple, individual points of contact rather than a constant line. As well, some of the pre-recorded gestures generated two or three ripples in close proximity when the surface was touched, even when the touch was intended to be a single point (see Figure 29). This was not an error, but a realistic representation of the length of a touch when indicating several locations in succession on the surface of a table. In these cases, participants sometimes interpreted it as multiple points and scored lower on accuracy measures.

Figure 29: Single contacts intended to identify points could sometimes result in several ripples, since the hand could brush the surface slightly during the gesture.

In response to these problems, the next studies use a line-based contact trace (similar to that in Tang's expanded VideoArms [17]) with an added ripple at the beginning and end of the contact. This visualization is described in further detail in Section 0, above.

Figure 30: Modified Touch + Trace visualization. Note the faded circle on the right at the point of contact. Over time, the circle fades and expands, simulating a ripple.

The Value of Contact Traces. Contact traces were valuable for improving the precision of target identification, but were not more effective than touch visualizations for differentiating between similar gestures. Previously, traces have been shown to improve gesture interpretation (Gutwin and Penner [13]) and the Study 1 results do not contradict this finding. Instead, Study 1 helps identify what aspects of a gesture need contact traces and for what aspects touching information is sufficient. Contact traces help accurately identify targets and this is likely why they worked well in the Gutwin and Penner study: the traces were used to identify a path-based target (e.g., showing a shape). However, many gestures do not have highly specific targets or are referring to targets for the second time (or later). For these gestures, contact traces are probably

less important, since the target is either easy to identify or already known, and touching visualizations are likely sufficient for conveying the gesture.

Confidence. Any visualization with more information than that contained in the arrow telepointer made participants more confident of their interpretation of the gesture. This happened even when they were incorrect, such as in the touching condition for the accuracy task. This shows that providing touching and historical information in an embodiment is a powerful tool for increasing participants' confidence during a collaboration.

<u>Study 1 Summary</u>

This first study clearly shows the benefit of touch and trace embodiment enhancements:

- touch information improves gesture identification;
- contact traces improve the accuracy of target identification;
- users are more confident in their answers when provided with touch information and/or contact traces than they are with a plain arrow telepointer.

These results present clear design choices for designers: If the correct expression of gestures is important, but not the accuracy of deictic target interpretations, simple touch visualizations are sufficient. For the most understandable embodiments, both touching information and contact traces should be included.

Study 2: Expressiveness of Height Visualizations

The second study investigated the question of whether or not added height visualization enhancements aid in conveying subjective qualities that are common in real-world gestures, such as the degree of confidence, specificity, and emphasis in the gesture. This study also examined whether or not the addition of above-the-surface detail in an embodiment would further improve gesture recognition.

Study One found that table-touching information in an embodiment allowed people to more accurately identify remote gestures. It also found that the intended targets of remote deixis were more accurately identified by people watching embodiments that contained contact traces than those that did not. Study One did not explore how either of these factors might change if an embodiment contained more explicit visualizations of how high the gesture was above the surface of the table, rather than a simpler on/off visualization.

Research Questions

Study Two examines three separate questions:

1. Do people interpret height information in a remote gesture as having an effect on the confidence, specificity, and emphasis of the gesture?
2. Do people associate visualizations of tapping with confidence, specificity, and emphasis?
3. Does additional, above-the-surface height information help people to interpret gesture type?

Given that improvements in the interpretation of intended targets in Study One was linked to the presence of contact traces (and not table-touching), it is unlikely that a more detailed view of the gestures above the surface of the table would improve the accuracy of target identification any further. There is a small possibility that any additional visualization would actually degrade performance on that metric due to occlusion and distraction, but careful design could likely minimize such an effect. As a result of these factors, this study did not examine whether there was a change in the accuracy of target identification with the additional height information.

Experimental Conditions

The experiment had three independent variables: the *global height*, indicating the overall height of the entire gesture above the table; the *local height variation*, which involved hand-bouncing or finger-tapping motions but did not change the global height of the gesture; and the *gesture type*, targeting a point, path, or area.

Gestures were recorded at three global heights: *high* (more than 20cm above the surface), *low* (between 5 and 20cm above the surface), and *surface* (in contact with the table). Half of the gestures were recorded with local height variation within the global height (i.e., a tapping or waving movement that remained within the bounds of a specific global height), and half where the hand and fingers maintained a constant, global height (i.e., no tapping). Each gesture set (high, high with local variation, low, low with local variation, surface, and surface with local variation) included two kinds of path gestures, two kinds of point gestures, and two kinds of area gestures:

- Pointing gestures indicating a single target;
- Pointing gestures indicating multiple scattered points;
- A simple path between two points;
- Paths between multiple points;
- An area delineated with a contour gesture;
- An area that is ‘coloured in’ with the gesture.

The dependent variables were the accuracy of gesture identification (as in Study 1) and the interpreted level of confidence, specificity, and emphasis.

Tasks

The study task was similar to that of the first study: participants were shown pre-recorded gestures and asked to respond to several questions. As in the first experiment, high contrast, cluttered aerial photographs were used as background images. Where areas, paths, or points were indicated in the pre-recorded gestures, the target of the gesture was believable (e.g., a group of similar buildings, an intersection, or a path along several roadways).

Participants answered a series of seven questions:

1. What was the gesture indicating (a point or points, a path, or an area);
2. How specific the gesture was in indicating a location;
3. How confident you feel the person was in making the gesture;
4. How much emphasis was intended to be conveyed by the gesture;
5. Whether or not the gesture was indicating something out of reach;
6. Whether or not the gesture was indicating something off of the edge of the map;
7. Whether or not this was the first time that the target of the gesture had been discussed during the conversation.

Questions 5, 6, and 7, were included to reduce the chance of participants guessing the hypothesized link between height and specificity, confidence, and emphasis. Question 1 was a multiple choice question; questions 2, 3, and 4 were answered on a 7-point Likert scale; and questions 5, 6, and 7 were true or false.

Proceedure

The procedure was almost identical to that in Study 1. Participants were seated at a large table, on which was projected the satellite photograph of New York (see Figure 23). Participants were asked to first watch a gesture's representation as it played on the surface of the table in front of them and then answer the seven questions listed above (in Section 0). Participants were

informed that the gesture would begin at the edge of the table immediately in front of them, move around on the table, before returning back to the approximate point of origin.

Participants completed a demographic survey and an orientation session that included previews of each visual condition prior to the study's start. They were also provided the opportunity to ask questions of the researcher. The gestures played without accompanying speech. Participants could pause between trials for as long as they wished to ask questions or rest.

Participants viewed 63 gestures in random order (some gestures were not replicated in all conditions, see below).

Participants

Sixteen participants (11 male, 4 female, 19-33 years old, mean 23.9) were recruited from the University of Saskatchewan. Participants were paid CND$10 to participate. Participants all reported using a computer every day; most used a mouse as primary input device (12 mouse, 3 trackpad, 1 touchscreen); and none reported having previously used or seen collaborative tabletop systems, large public touchscreen displays, collaborative GIS, distributed groupware, or screen sharing.

Apparatus

Gesture height was shown using a combination of two designs described above: height above the table was shown using the ellipse and sphere visualizations (Figure 18 and Figure 19), but the moment the gesture touched the table, the sphere disappeared and the line-based contact traces, state-change cursor and starting and ending ripples were used (Figure 30). As discussed above, line-based traces were used rather than ripples because lines are less likely to indicate multiple points to viewers.

Setting

As in Study 1, gestures were recorded and played back on a 124cm by 185cm table with top-down projection using two projectors. The displayed image was a stitched 1024x1534 pixel image (two 1024x768 images stitched top to bottom). The display had an 8.3 dots per cm (14 dpi) resolution.

Gestures took place in the area immediately in front of the participant. The room was darkened and the researcher available during the study for questions or concerns.

Design

The study used a within-participants design to examine the effects of two factors: the global height of the gesture, and the presence of local height variation in the gesture. These factors were fully crossed and contained either three or four trials for each of the gesture types (path, point, area). The difference in number results from the fact that some gestures would not realistically be used in some situations (e.g., hovering over an area works only above the surface). There were a total of 63 randomly-ordered trials.

The dependent variables were:

1. the level of interpreted emphasis of the gesture;
2. the level of interpreted confidence of the gesture;
3. the level of interpreted specificity of the gesture; and,
4. the accuracy of gesture type identification.

Data Collection

Demographic data were collected using a single-page pen and paper questionnaire. Experimental data were collected using custom software integrated with the gesture-display software. Participants entered answers to questions using single keystrokes on a keyboard (e.g., selecting a multiple choice answer).

Results

There were two groups of data: participant measures of gesture qualities, and an accuracy test. The participant measures data were examined to determine how each height variation (global height, and local height variation) affected the way that participants interpreted gesture qualities.

Interpretation of Gesture Qualities: Global Height. All participants interpreted the ellipse height visualization as we expected, inferring that lower height implied greater confidence, specificity, and emphasis. We carried out an ANOVA to determine whether the height visualizations at our three global heights led to differences in participant ratings of these three qualities. In all cases, the ANOVA showed a significant effect of height on the perceived level of the quality (for confidence, $F_{2,30}=24.84$; for specificity, $F_{2,30}=30.27$; for emphasis, $F_{2,30}=15.64$; all $p<0.001$), as shown in Figure 31.

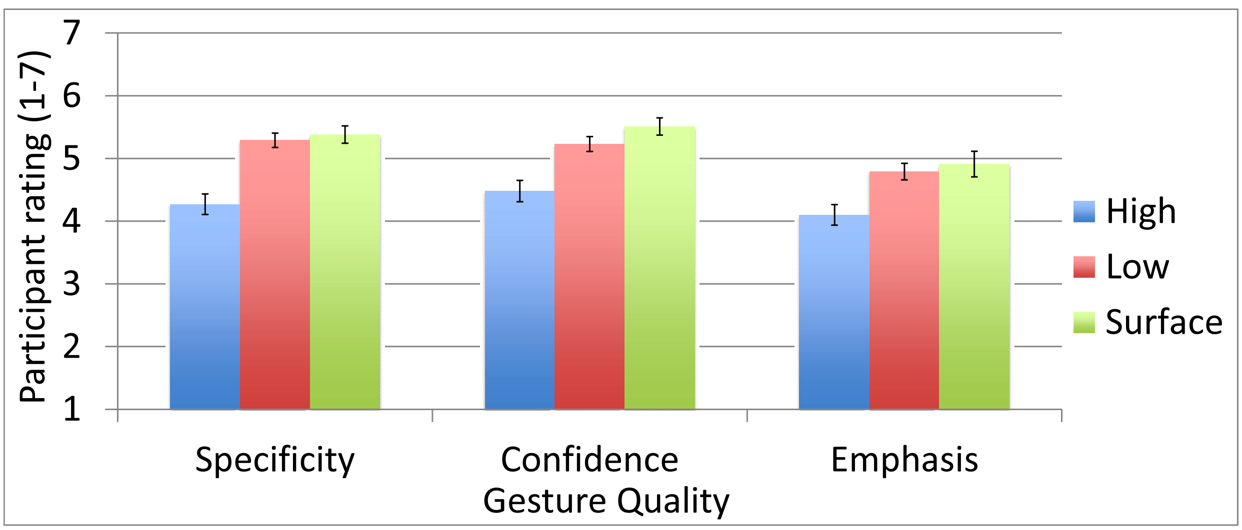


Figure 31: Interpretation of gesture qualities, by global height.

Interpretation of Gesture Qualities: Local Height Variation. RM-ANOVA also showed a main effect of local height variation on interpretation of qualities ($F_{1,15}=6.39$, $p<0.001$). As shown in Figure 32Figure 33 andFigure 34, people were much more likely to interpret a

gesture as being emphatic, specific, and confident if there was a bouncing or finger-tapping motion in the gesture. There was also an interaction between global height and local height variation ($F_{2,30}$=4.68, p<0.001). As the figures show, when people saw only the global height, they were much more likely to reduce their estimation of emphasis with High gestures, but when there was local height variation, this information was consistently interpreted as implying that the gesture was emphatic.

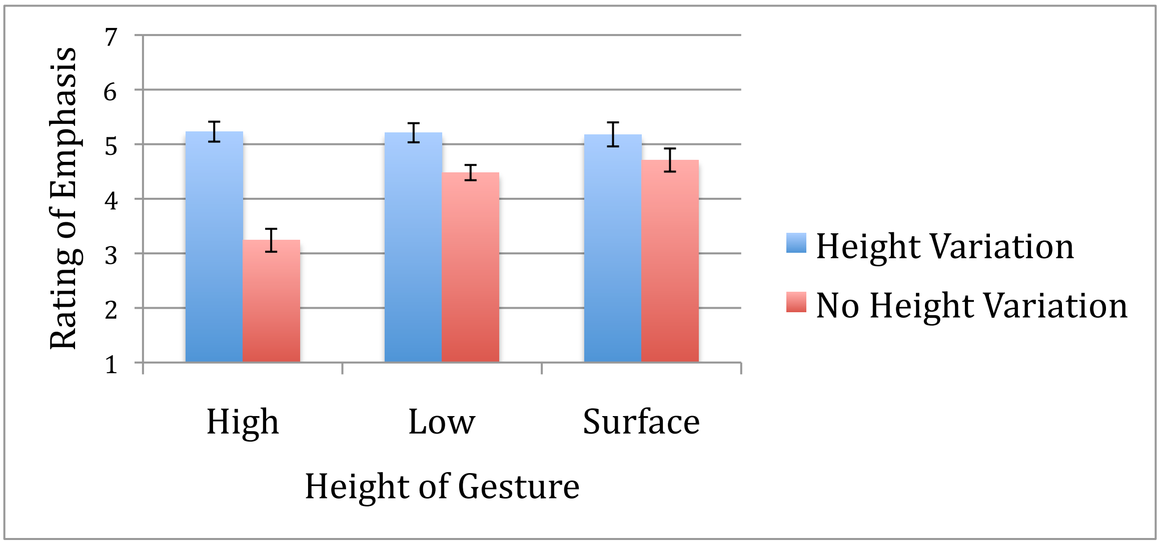


Figure 32: Interpretation of emphasis, by local height

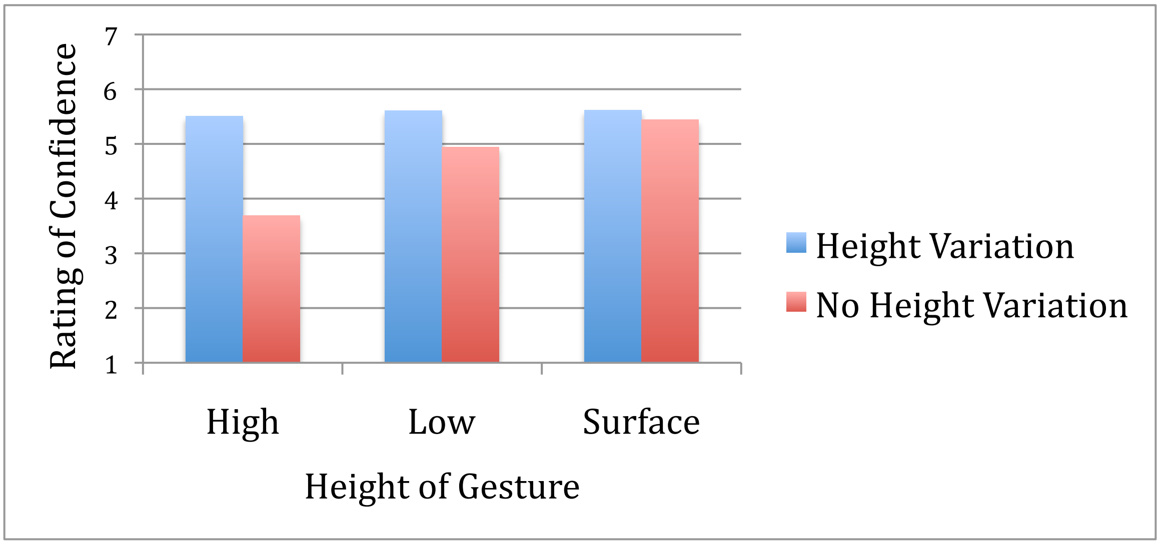


Figure 33: Interpretation of confidence, by local height

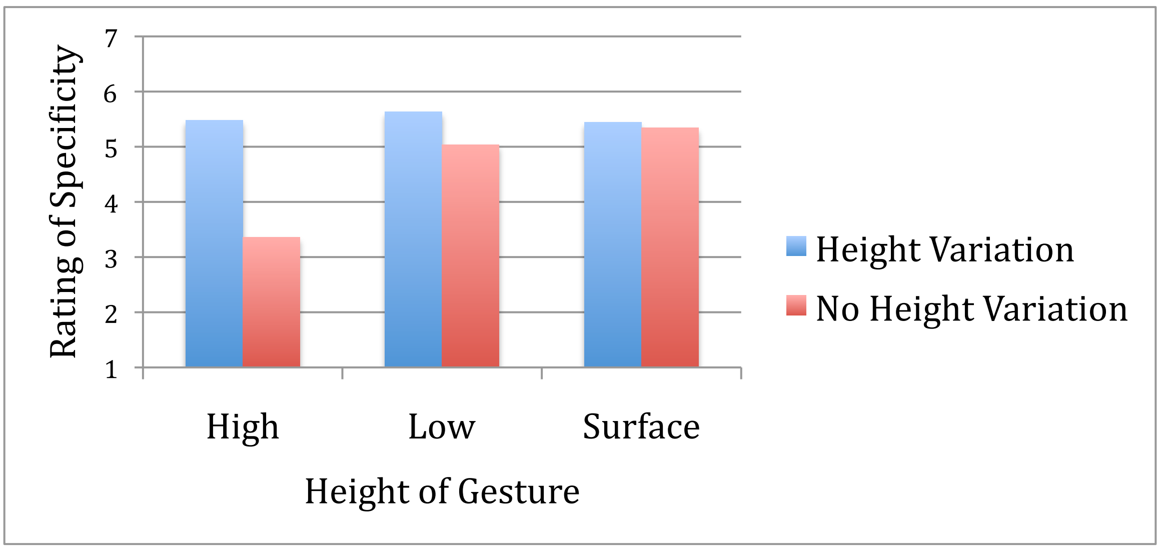


Figure 34: Interpretation of specificity, by local height

Accuracy of Identifying Gesture Type. RM-ANOVA showed a main effect of global height on interpretation accuracy ($F_{2,30}$=9.027, $p<0.001$). The high error rates observed in some conditions were investigated post hoc (Figure 35). The exceptional error rates were specific to particular gesture types: paths with points had high error rates above the surface and both multiple-point paths and area contours had high error rates on the surface. Participants largely

misclassified the multiple-point paths and area contours surface gestures as path gestures, and both of these gesture types do share considerable information with paths. Part of this result is attributable to the complicated nature of gestures. The points-on-a-path gesture could also have been classified as a pointing gesture – participants interpreted it as a path on the surface and as pointing when in the air (see Figure 36).

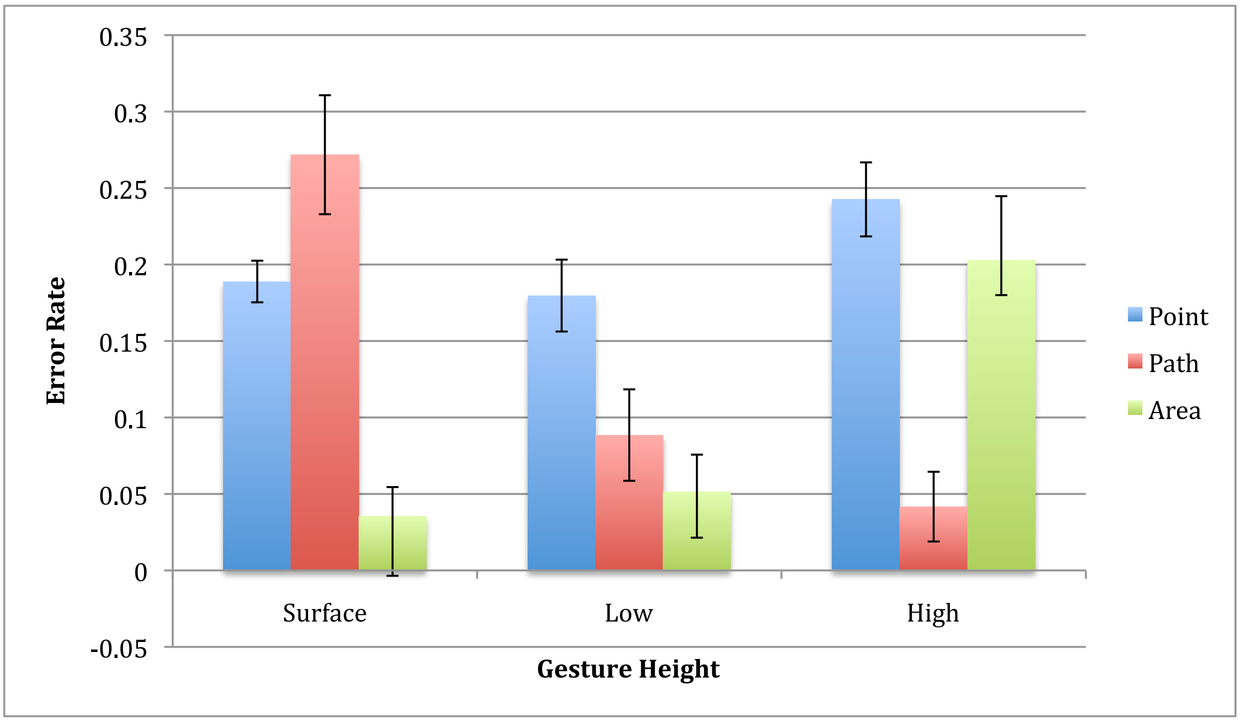


Figure 35: Interpretation errors, by gesture target and height

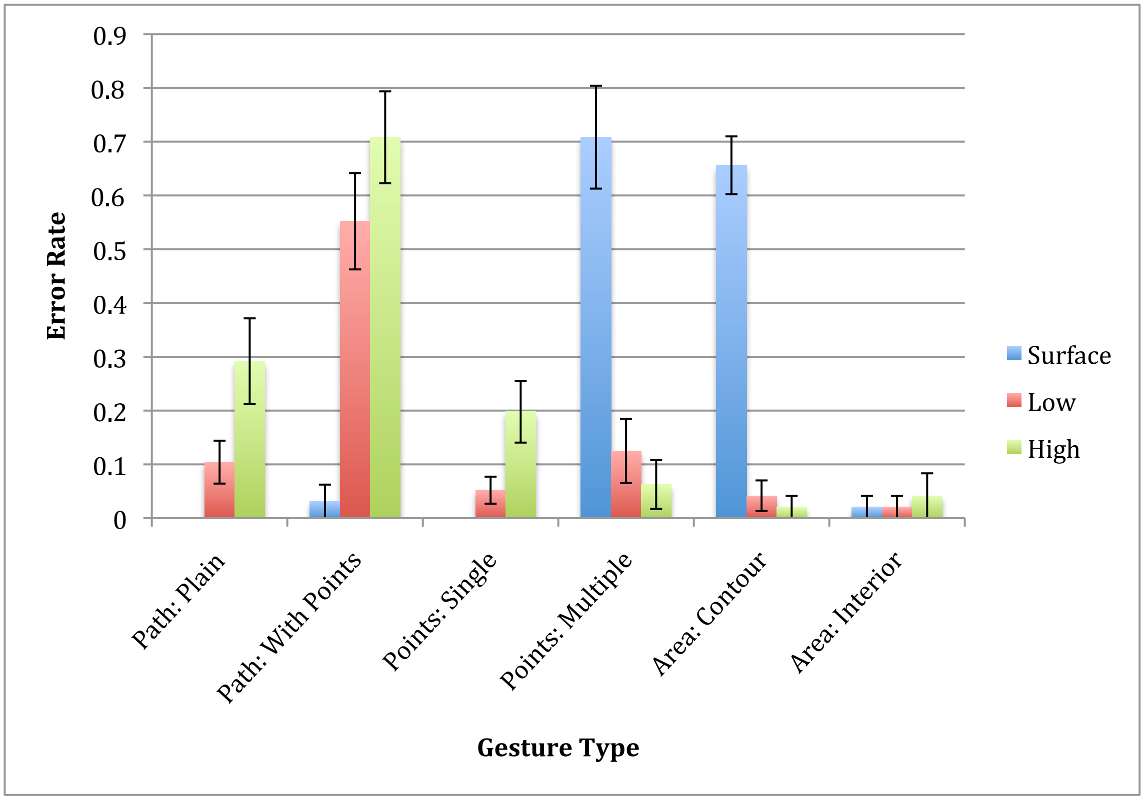


Figure 36: Interpretation errors, by gesture group and height

<u>Study 2 Discussion</u>

Height Variation and Global Height. Users interpreted gestures that contained local height variations as similarly emphatic, precise, and confident as gestures that were on the surface of the table. This result implies two things: where embodiments do not show variations in height beyond touching and hover, users can still convey higher or lower levels of emphasis, confidence, and specificity; and that embodiments that model the full range of height used in gestures can show a range of gesture qualities not available through touch and hover representations.

A tapping gesture is a series of rapid accelerations, decelerations, and changes in direction. In the way that a gesture's height can be shown, changes in a gesture's height can also be shown with embodiment techniques. One direction for future research might be discovering

whether adding visualizations of acceleration or changes in acceleration can improve the interpretation of remote gestures.

Errors in Gesture Identification. The high error rates in identifying particular kinds of gestures are undertstandable:

- Multiple Points on the Surface: the gesture only pauses at each point, but does not lift. This gesture could equally be performed when someone is uncertain about a path or area and pauses to take stock. Thus, many of these were incorrectly identified as paths.
- Area Contour on the Surface, Plain Path Low and High: there is little practical difference between a contour and a closed path (one that returns to its starting point). At higher levels, such gestures are interpreted as areas (because they are high), thus Low and High Plain Path gestures were incorrectly classified as Area Contour gestures and the Surface Area Contour gesture was incorrectly classified as a Plain Path.
- Path with Points, Low and High: points along a path can also be described as multiple points. The Low and High versions of Path with Points, were regularly misinterpreted as Multiple Points.

These understandable errors were made frequently enough that the mean error rates for each height condition were artificially raised. However, there were several kinds of gestures that were regularly correctly interpreted. The ‘colouring in’ method of showing an area was interpreted correctly almost without fail, as were the above the surface versions of Contour Areas and Multiple Points.

<u>Study 2 Summary</u>

Study Two shows that extra information about gesture height above the surface, shown as a continuous, ordinal visualization, is interpreted by people in much the same way as colocated height. The global height of a gesture is consistently interpreted as being inversely proportional

to the gesture's emphasis, confidence, and specificity, but local height variations, can change this interpretation. Local height variation is interpreted as having roughly the same level of emphasis, confidence, and specificity as a gesture with a global surface height. Tapping while already on the surface, oddly, is not interpreted (through the visualization) as being more emphatic, for instance, than a gesture that remains on the surface. This means that local variation essentially over-rides any interpretation that might be applied to global height variations: people interprete all locally-varying gestures in the same way that they do on-the-surface gestures with respect to confidence, emphasis, and specificity.

These results suggest that both parts of the above-the-surface height are important: the global height at which the gesture is made and any local changes in height within that space. They re-inforce the design decisions described in Chapter Four: despite that there are apparent levels in height (as described in Chapter 3), the height above the table is a continuous variable and should be represented as such, not a simple nominal switch between layers.

Whether above-the-surface visualizations assist in differentiating between similar gestures is a more difficult question. People confused some kinds of gestures but not others. Some of the problem may lie with the precision of the embodiment. The continuous representation that assisted with accurately representing gesture quality may complicate gesture differentiation. Natural gestures above the surface may unintentionally have small height variations. These variations, when made more explicit by a height-enhanced embodiment, may lead to interpretation errors. In sum, Study Two provides no clear evidence if above-the-surface visualizations assist with gesture differentiation or if they cause problems: there were cases in which each occurred.

Study 3: Usability in Realistic Collaboration

Research Questions

The third study investigated the use of height-augmented embodiments in a realistic collaborative situation, and looked at three main issues:

- *Usage*: do people make use of the height visualization when making gestures, and for what do they use them?
- *Interpretation*: are gestures interpreted as intended, and do the visuals of the augmented embodiments cause any communication difficulty or confusion?
  - *Preference*: do people prefer augmented embodiments over standard versions?

Experimental Conditions

The study compared two different embodiments in a realistic distributed setting: a height-augmented embodiment and a ‘standard’ embodiment. The height-augmented embodiment used in the study (and shown equally to both participants) were the same as those used for the second study (Figure 18, Figure 19, and Figure 30). The standard embodiment was a standard arrow telepointer (Figure 22, left).

Tasks

The study involved a realistic collaborative task between two distributed participants: a local participant whose gestures were tracked and displayed, and a remote participant who could see the local participant’s gestures on a display, but whose gestures were not tracked. The remote participant was given a list of 22 questions about locations in Saskatoon, the local city (similar to those used by Kettebekov and Sharma [13]; Appendix A includes an example of the questionnaire used for this study). The questions were selected to explore a wide range of open-ended interactions. Some questions involved areas that were familiar to one or both participants,

and some questions were designed to involve unfamiliar areas of the city (e.g., safe boating areas on the river, or optimal pathways through certain neighbourhoods).

The remote participant asked each question to the local participant, who responded by speaking and gesturing over the map. Participants were asked to continue discussing the question until both were satisfied that the answer was understood and correct.

Proceedure

Participants were recruited in groups of two and provided simultaneous orientations. After signing an ethics form, the participants were given an overview of the study task and introduced to the two different embodiments that were used during the study. The participants were encouraged to ask questions and experiment with using the embodiment prior to being separated. When the participants were ready, they were selected for the remote and local roles based on their knowledge of the city. Those with more knowledge were assigned the role of local participant, since they would be better able to provide more detailed answers to the questions. In this way, participants were more likely to spend longer discussing the questions and gesturing over the map, providing more data.

The local participant was seated at the same large table used in Studies 1 and 2 and the remote participant was escorted to an adjoining room with a second display. The two participants were provided with headsets for an audio connection and asked to ensure that they could hear each other easily. When ready, the remote participant began asking questions of the local participant, who replied to each answer using both gesture and speech. Participants were encouraged to discuss the question and answer as long as possible to ensure they understood each other.

Half of the groups began with the height-enabled embodiment, and the other half with the telepointer. After 11 questions, they switched to the other visualization. Before each condition,

participants were again given as much time as needed to familiarize themselves with the embodiments.

Between each question, participants filled out a questionnaire about the conversation they had just completed. (Example questions from the questionnaire are included in Appendix A.) The questionnaire asked the participants to identify different qualities of the gestures that had been used in the exchange, and also asked how quickly they had achieved understanding, how well that understanding was achieved, and whether any locations discussed had been out of reach or off the table. After each session, an open-ended interview was conducted to discuss the participants' experiences and to ask about preferences.

Participants

Eight pairs of participants were recruited from the local university (5 female, 11 male, ages 21-33, mean 24.4). All participants reported using a computer every day; 14 reported using a mouse most frequently as input device, and two reported using a touchpad most frequently; none reported having previously used or seen collaborative tabletop systems, large public touchscreen displays, collaborative GIS, distributed groupware; and, two participants had previously used screen sharing software. Participants were paid CND$10 for their participation.

Apparatus

The local participant had a Polhemus Liberty 240/80 sensor taped to his or her primary pointing finger and was seated at the same top-down projection table as was used in Studies 1 and 2. The remote participant was seated at a 40 inch plasma screen television placed on its back at coffee-table height. Both participants were seated at the foot of their tables with the same orientation to the display.

Participants were provided microphones and headphones with which to communicate between the rooms. Headphones were not noise-cancelling, but participants found it difficult to

hear researchers or extraneous noises without removing them. The audio connection was provided by initiating a Skype call between the two computers and the audio quality remained high for all participants.

Setting

The study took place in two darkened rooms. The local participant was in a larger space that had been used for Studies 1 and 2. The remote participant was in a small obervation room adjoining the local participant's room, seated with his/her back to the observation window. Participants were not able to see each other and the door was closed between the two rooms, limiting audio contact to that provided by the headsets.

The study was run by two researchers, one in each room. Participants congregated in the larger room for the initial instructions and the closing interview.

Study Design

The two embodiment conditions were crossed amoung the 8 groups: half started with standard embodiments for the first 11 questions before changing to the height-augmented embodiment and the other half began with the height-augmented embodiment and finished with the standard embodiment. The post-task interviews were semi-structured and open-ended. Each participant encouraged to answer, especially when one deferred regularly to the other.

Data Collection

The task and the post hoc interview were recorded on video in the large room and with audio in the small (remote) room.The questionnaire completed after each question and answer segment was paper-based. Performance data was not gathered, since the open-ended nature of the tasks and the differences in familiarity with the locations led to high variance in communication amounts and completion times.

Results and Discussion

The results of Study Three are organized into three main categories: how people used the height-augmented embodiments, how the embodiments were interpreted, and which embodiments were preferred by participants.

Data were analysed using a variation of Interaction Analysis [110]. In particular, participants were observed during the execution of the tasks at which point questions for post hoc interviews were formulated. Interview questions were re-used for subsequent participants when appropriate. Questions were based on apparent deviations from normal gesturing behaviour, gestures that were not understood by the researcher, and any experiences of note, such as moments where communication seemed more challenging than usual.

After the experiments were complete, video logs of the experiments and interviews were reviewed. An assessment of deviations from expected behaviour was undertaken and when examples identified, the video logs were searched for similar examples to identify problems or successes.

Usage of height-augmented embodiments. Observations during sessions and during later video analysis showed that the local participant (the person constructing gestures) adapted quickly to the additional capabilities provided by the enhanced embodiment. We observed that all participants clearly understood what the remote participant would see of the visualization, and all participants showed evidence that some of their gestures were constructed to make use of the height visualization.

Further evidence of how the embodiment was interpreted came from verbal interactions between participants. On several occasions, the gesturing participant asked the other participant, early in the session, if he/she was able to see that a particular gesture was higher or lower, confirming that the embodiment was being interpreted correctly.

Indicating areas. Local participants regularly used the ellipse to indicate areas by hovering or moving slightly at a height that made the ellipse both an appropriate size for the target under discussion (e.g., Figure 37, right), and to indicate emphasis. For example, the following exchange occurred during discussion of shopping areas (LP: Local Participant):

LP: It's on this road. [*leaves hand on surface while tracing along road, scanning ahead to find target*]

LP: [*raises hand slightly to create ellipse*] And here's the mall. [*taps finger in the air, varying the size of the ellipse*]

Improved gesture visibility. When local participants constructed a path or contour gesture with the augmented embodiment, it was clear that they knew the other person would see the path (because of the contact traces). As a result, there were no observed instances where the local participant repeated the gesture. When people constructed gestures with the plain telepointer, however, it was common for the participant to repeat the gesture several times, even without prompting from the remote participant.

Indicating out-of-reach targets. When tasks required the identification of far-away targets, local participants often used the augmented embodiment's simulated ray-casting ability (see Figure 37, left and centre). The following exchanges illustrate the difference between the two embodiment conditions; participants were able to come to a shared understanding with much less effort using the augmented embodiment.

Telepointer condition:

LP: Somewhere up there... I can't reach very far but... [*stretches out of chair to reach as far as possible*]
RP: Yeah…
LP: [*sits down and retracts hand*] Up along the river.

Enhanced embodiment condition:

LP: Up along the river, in the middle, there. [*at limit of reach, raises hand to use raycasting, stays seated*]
RP: OK

Figure 37: Left, reaching for a distant target in the plain condition; centre, reaching in the enhanced condition; right, hovering over an area in the enhanced condition.

Controlling specificity. Participants using the enhanced embodiment were aware of the way that height was shown to the other person and exercised much more control over gesture height than we observed in the telepointer condition. For example, in several sessions, it was observed (and the participant later confirmed during interviews) that the local participant avoided touching the table unless they intended a high level of specificity (as would be conveyed by the visualization). In the telepointer condition, however, participants touched the table indiscriminantly (e.g., at times where there was no high level of specificity or emphasis) and carried out their gestures above the table and heights that appeared to be based on comfort rather than communication. This likely occured because there was only one level of specificity possible in the embodiment. This did not cause a problem for the telepointer condition (although we wonder what would happen in mixed-presence settings), but indicates again that participants were aware of and made use of the additional expressive capabilities of the embodiment when these were available.

Interpretation of height-augmented embodiments. The video data were analyzed by looking for examples in which the additional visual information of the augmented embodiments caused clutter, confusion, or errors. In no cases did the extra information cause any noticeable problem, and nor did any participant report such problems in the post-session interview. Particular attention was paid to issues of distraction and occlusion for the ellipse visualization, but the transparency of this effect appeared to successfully remove any difficulties for the participants.

It was also clear that remote participants made use of the features of the augmented embodiment in their interpretations. Evidence for the value of the visualizations came out in the interviews, where it was clear that remote participants understood the visual effects and wanted the local participant to make use of them. For example, one remote participant remarked, "I was hoping that she [the local participant] would use the difference between the ball [ellipsoid] and the pointer more than she did."

Gestures in this study were accompanied with speech, which likely limited interpretation problems (for either condition), in that accompanying speech could often be used to disambiguate a gesture. This indicates that (as expected) the information conveyed by height visualization may not be critical to the successful completion of the task; rather, the augmentations add subtlety and expressiveness to the overall interaction. (The fact that people did notice and appreciate these changes is shown in the preference data described below).

Preferences. When asked about their preferences after the session, 14 of the 16 participants stated that they preferred the height-augmented embodiments. In addition, remote participants universally preferred the height-augmented version. In one typical response a

participant described the visualization as, "really nice, especially the trail left behind." The touch traces were appreciated by both participants: one remote participant said that the "pathways were a lot easier [to see]" with touch traces; a local user agreed, stating: "I liked the afterwards line. That was useful for generating paths." The two participants who preferred the plain telepointer were in the role of local participant; both of these participants indicated that the plain telepointer was more familiar and therefore more desirable.

Other analyses. Analysis of the post-conversation questionnaires (using RM-ANOVA) found no significant differences between telepointer conditions (all $p>0.05$). Major differences in these measures were not expected because of the high variance introduced by the unconstrained task.

Study 3 Summary

Study Three found that the height-enhanced embodiments described in Chapter 4 are usable in real-world contexts. Participants preferred the embodiments to telepointers, used many aspects of the design as intended, gestured more naturally than they did with telepointers, and observers were easily able to understand remote gestures. The study was designed with unconstrained tasks to promote natural interactions between collaborators during sessions. Unfortunately, this meant that significant associations between gesture height and the observed qualities of the gesture were not found. This does not mean that those qualities were not expressed – in fact, based on the naturalness of interactions between collaborators during the study, it is quite likely that these qualities were accurately conveyed. Study Three provides strong evidence that the height-enhanced embodiments are usable in real-world collaboration.

Conclusion

The studies described in this chapter show that adding abstract height information to embodiments can improve the use and understanding of gestures in distributed collaboration.

Embodiments can improve the accuracy of gesture and target identification by including table-touching visualizations and contact traces. Further enhancing an embodiment to include information about how high the gesture is above the table allows users to interpret the embodiment in much the same way they would interpret a collocated gesture. The evidence suggests that more information about the height of the gesture, both where it is and where it has been, is better for communication. These effects persisted in an ecologically valid, open-ended collaboration and users adapted quickly to the embodiments, pointed more naturally than they did with plain telepointers, and developed a new pointing language to take advantage of the opportunities afforded by the designs.

This chapter has provided evidence for the value of representing height in remote tabletop embodiments: height visualizations can improve interpretation accuracy, can improve the expressiveness of remote gestures, and can be quickly learned and used in realistic collaboration.

KINECTARMS: A TOOLKIT FOR CAPTURING AND DISPLAYING ARM EMBODIMENTS[4]

The previous chapters identified two main problems with capturing and displaying deictic gestures over surfaces. First, the complexity and subtlety of many gestures cannot be adequately conveyed using simple pointing mechanisms such as telepointers. Therefore, some CSCW systems have explored video-based embodiments (e.g., VideoArms [14]) that provide more information about the details of the gesture. However, video-based techniques often suffer from problems of poor video separation and low resolution.

Second, information about the height of the remote gesture is lost in distributed settings – but there are many situations where gesture height is a critical aspect of the communication [1]. Height information is difficult to capture, usually requiring expensive tracking technologies, such as the Polheums used in Study 3 of the previous chapter; in addition, since remote arms are displayed on the table surface, height information is difficult to convey through arm representations.

This chapter introduces a new toolkit, called KinectArms, which solves all three of the issues mentioned above. First, it uses a depth camera to quickly and efficiently extract arm images from the video stream – the system runs easily at 30 frames per second. Second, KinectArms uses the depth camera to identify fine-grained height information about the gesture, and can attach that information to extracted features of the images such as hands and fingers. Third, KinectArms provides several visualization techniques to show gesture height, to enhance the visibility of embodiments, and to highlight arm movements (e.g., abstract height indicators, color and transparency modifications, and motion lines and traces) (see Figure 38) . The toolkit also provides several other capabilities, including automatic calibration to the table surface,

---

[4] Material from this chapter was published as “KinectArms: A Toolkit for Capturing and Representing Gestures over Surfaces” at CSCW 2013.

tracking of arms in the scene, and wrappers to allow development in several programming languages.

Figure 38: Simulated use of KinectArms in a photo-sharing application, showing shadow and height indictor (circle).

This chapter describes the design, implementation, and use of KinectArms and makes three contributions.

1. KinectArms is the first toolkit to provide a simple solution to all the aspects of remote arm embodiments, including capture and representation of gesture height. It goes well beyond standard Kinect libraries, which are not set up to recognize arms over table surfaces.
2. This chapter describes how KinectArms can reproduce a wide variety of previous designs (e.g., DOVE, VideoArms, and C-Slate) and show that the platform can be used to produce new representations as well, providing an opportunity to compare different designs.

3. Finally, this chapter provides an evaluation of the toolkit in terms of performance, developer effort, expressiveness, and simplicity, and show that KinectArms provides a powerful yet easy to use solution to the problem of remote gestures.

With KinectArms, freely available (at https:// github.com/AaronGenest/KinectArms), designers can easily add support for remote gesture to tabletops, which can dramatically improve the usability and communicative capabilities of distributed tabletop groupware.

Requirements for an Arm-Embodiment Toolkit

A new toolkit for supporting the visualization of gestures in distributed systems would be valuable for researchers and developers. With the advent of cheap depth sensors, it has become important to understand how all three dimensions of gesture can be represented in a variety of distributed settings. This cannot be easily done with the state-of-the-art in software tools for capturing and representing gestures. A toolkit for arm embodiments has four main requirements:

1. Address the three problems described above – of easily separating arm images, capturing and displaying height information, and enhancing embodiment visibility;
2. Permit the rapid reproduction of existing embodiment techniques to allow experimentation and comparison;
3. Support easy creation of new embodiment types with both realistic and abstract representation techniques;
4. Provide simple setup and calibration.

The KinectArms Toolkit

KinectArms is a toolkit that simplifies the capture of remote tabletop gestures and the display of those gestures through arm embodiments. KinectArms has two parts: a capture module that recognizes hands and arms above the surface, performs video separation, and identifies the height of each pixel of the separated image; and a display module that provides built-in effects to

show height, to improve visibility, and to provide movement traces. Additional effects can be easily developed and added to the toolkit.

Overview and Setup

KinectTable uses an Xbox or PC Kinect sensor, fixed above the table (1.8m is optimal for the Xbox version) and pointed down. The camera need not be perfectly perpendicular to the display surface; the automatic table detection (see below) can handle small angular variation. The Kinect has a field of view of 57° horizontally and 43° vertically, so at 1.8 meters height, the camera can accommodate a 1.42m by 1.95m table surface. This area can be increased for larger tables (at the cost of resolution) by raising the Kinect.

Although 2D resolution will increase as the Kinect approaches the surface, the depth image has an optimal resolution at 1.8m, with an accuracy of approximately 2mm at that distance [111]. This level of accuracy means that user touches on the table surface can be approximately determined (i.e., without requiring a touch-sensitive surface); however, our tests described below used a PQLabs touch overlay on a 60-inch LCD television.

After the Kinect is suspended above the table surface, the KinectArms software can be started (usually by a client application). On startup, KinectArms automatically detects the table surface (detailed further below). Table detection takes less than a second, and no further calibration is required – KinectArms is then ready to capture and process arm images, and add visual effects to these images.

The general algorithm for using KinectArms in an arbitrary tabletop application is as follows:

1. Create an instance of the KinectArms client
2. Set up initial visual effects

3. While the application is running:
    a. Get image data from the KinectArms client
    b. Apply visual effects to the current frame
    c. Draw the current frame to the table

KinectTable: The Capture Module

The capture module of KinectArms uses a Microsoft Kinect as its primary sensor. The Kinect incorporates a full-colour camera and a depth camera (both 640x480), which can be aligned to produce a depth-mapped image. From this image, we accomplish both separation of arms and hands from the table background, and recognition of the structure of the arms in the scene. This recognition is *not* already provided by the Kinect – although the SDK provides sophisticated body recognition, it is not designed to capture isolated arms and hands over a table surface (it needs to see the entire body for accurate recognition).

The main features provided by the KinectTable module are:

1. Fast setup and calibration, with automatic detection of the table surface.
2. An API for accessing information about arms and table:
    a. Image masks for the table and each of the arms
    b. Fingertip, palm, and arm locations in three dimensions
    c. Basic user tracking
    d. Geometric properties for each of the arms.

<u>The KinectTable API</u>

The C++ version of the KinectTable API provides access to image data, table information, and arm information.

Image Data. The raw color and depth images from the Kinect are available in a custom structure; the application programmer can use the raw information, or can obtain a more standard image representation from the KinectViz module.

```
KinectArmsClient *client = KinectArmsGetClient();
KinectData data;
client->GetData(data);
DepthImage& depthImage = data.depthImage;
```

Table Information. KinectTable provides information about the table surface (from the automatic recognition step, as described below) including the height of the table, the table corners, and a bitmap mask where white pixels indicate the table.

```
BinaryImage& maskImage = data.tableMaskImage;
```

Arm and Finger Information. Information for each arm is stored in a C++ struct; these are provided in an array representing all arms above the table (stored in the order that they enter the space). Each struct contains the following arm information:

1. Geometric values:
    a. Mean depth of the arm (using all pixels);
    b. Total number of pixels corresponding to the arm;
    c. The pixel at the geometric center of the arm;
2. An array of points defining the arm boundary;
3. A bitmap mask corresponding to the arm;
4. An array of points corresponding to fingertips;
5. An array of points for between-finger locations;
6. A point representing the center of the hand;
7. A point representing the base of the arm (i.e., the intersection of the arm with the edge of the table);

8. A unique ID, retained between frames;

The height of any point in the arm image can be obtained from the `getHeight(Point p)` function. For example, the following code retrieves the depth of an arm's first finger:

```
client->GetData(data);
Arm& firstArm = data.arms[0];
int fingerHeight = client->getHeight(firstArm.fingers[0]);
```

<u>KinectTable System Architecture</u>

KinectTable is comprised of four components: a camera component, a table detector, an arm and hand detector, and an arm tracker (Figure 39). KinectTable makes use of the Kinect SDK or the OpenNI library to obtain the Kinect images, and the OpenCV library for image processing (opencv.willowgarage.com).

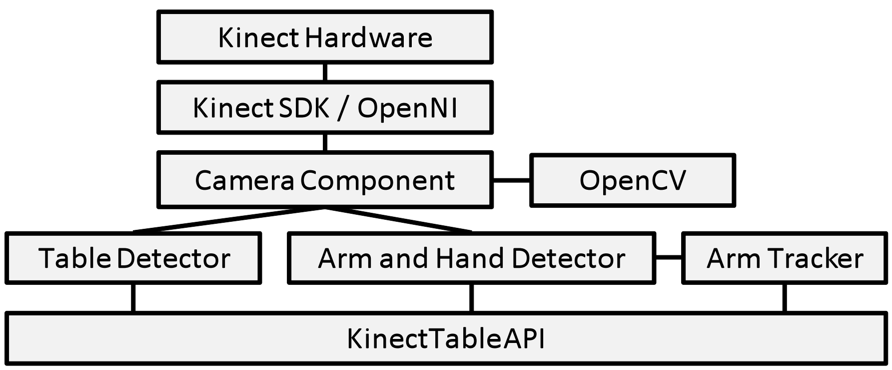


Figure 39: KinectTable system architecture.

Camera Component. The camera component fetches image data through the Kinect SDK, OpenNI, or the Kinect for Windows SDK, sets the SDK to align the colour and depth images, and performs low-level processing to prepare the data for table, arm, and hand detection. This processing involves converting the Kinect images to our own custom C++ structure, and passing a 5x5 median filter over the image to fill in error pixels (some pixels fail to update every frame).

Table Detection. The height and extents of the table are critical in the process of identifying arms, and most systems use an interactive calibration step or similar procedure to determine these values [14], [94]. KinectArms simplifies this step with an automatic table detector. Through the `recognizeTable()` function, KinectTable can be asked to identify the largest visible and roughly planar surface in the image as the table. After identification, any pixels in the depth map that are farther than the table, or outside its boundaries, are ignored.

Table detection is performed with the following steps:

1. A Laplacian filter and thresholding function are passed over the depth image to identify sharp changes in depth;
2. Depth edges are dilated to ensure continuity;
3. The center region enclosed by the continuous edge is filled – this region represents the table;
4. Table corners are found using k-curvature (as in [112]) on the region boundary (k-value of 30, threshold of 70°);
5. Table height is set as the lowest value of the table pixels.

Table identification is robust for slanted and curved tables, but any large non-linear changes in depth may cause incorrect table identification.

Arm and Finger Detection. Arms and fingers are identified using image processing techniques similar to those used for table detection (Figure 40):

1. Background subtraction is performed by removing pixels in the depth image that are not above the table region. The remaining pixels are arm candidates.
2. A Canny edge detector (max threshold 200, min 150, sobel filter order 5) finds the edges of the arms. Edges are dilated to make them continuous.

3. Contours of each arm candidate are found using OpenCV's Suzuki85 algorithm.
4. All small contours (< 40 pixels in length) are discarded as noise. The remaining contours are considered arms.
5. Geometric properties of each arm (mean depth, total pixels, center point) are calculated from the contours.
6. The base of each arm is determined by finding the intersection of the arm with the edge of the table.
7. Fingertip locations and the points between the fingers are found using k-curvature (k-value 30, threshold 85°).

The center of the hand is calculated using a Euclidean distance transform of finger and between-finger points.

Arm Tracking and User Identification. Arms are given unique ID numbers so that the application can track arms across multiple frames. Tracking takes advantage of the location where the arm intersects the edge of the table (a location that does not change rapidly in tabletop work). KinectTable matches arm bases in the current frame to known arm bases in the previous frame, by comparing distance and time values between the frames. If an arm in a previous frame cannot be matched, we assume it has left the table area. If an arm does not return at the same base position within five seconds, we re-use the ID for the next arm that enters the table area.

.NET Wrapper. Developers can access KinectTable with .NET languages through a wrapper created with Visual C++, which provides .NET versions of all data structures. The wrapper's API is similar to the C++ API but also takes advantage of extra work in an event-driven fashion whereby methods are called automatically when new data is available.

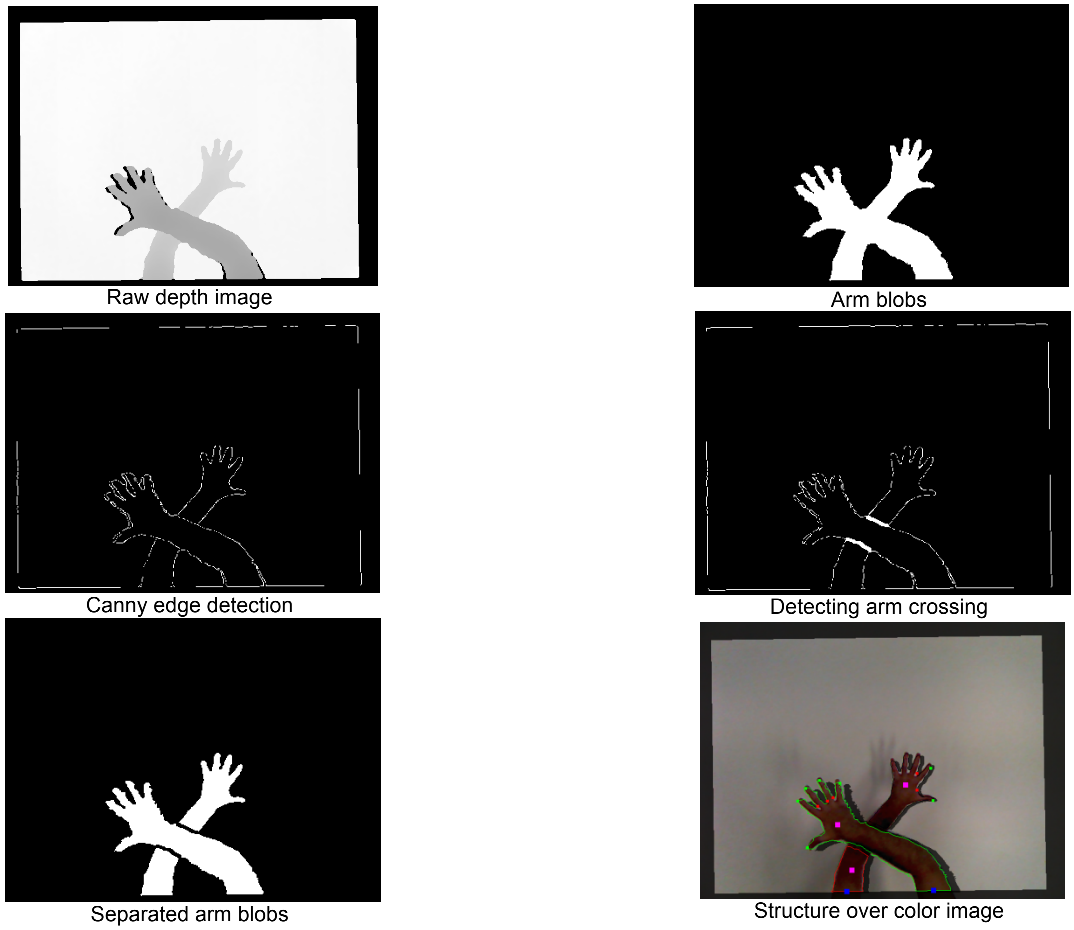


Figure 40: Steps in processing arm images

KinectViz: The Display Module

KinectViz provides a set of standard effects that can be added to arm images obtained from KinectTable. The library provides effects to visualize height, increase or decrease arm visibility, improve user identification, and provide motion traces. Using these effects, it is possible to replicate many previous embodiments; KinectViz also allows programmers to design new effects. Programmers can set up effects with single API calls, and effects can be changed dynamically and can be assigned to specific height layers. Effects can be applied to all arms or to specific arms (the examples below show the global versions).

<u>Basic Arm Representations</u>

VideoArms. KinectViz provides full-colour background-subtracted images of hands and arms that can be drawn on top of existing tabletop objects. The basic representation (similar to VideoArms [23]) is built in a few lines of code:

```
client->GetData(data);
ColorImage image;
applyVizEffects(data, image);
```

The `applyVizEffects()` method translates raw KinectTable data into a drawable image, and also alters the image to add any visual effects that have been chosen by the application programmer (no effects are needed for the standard arm).

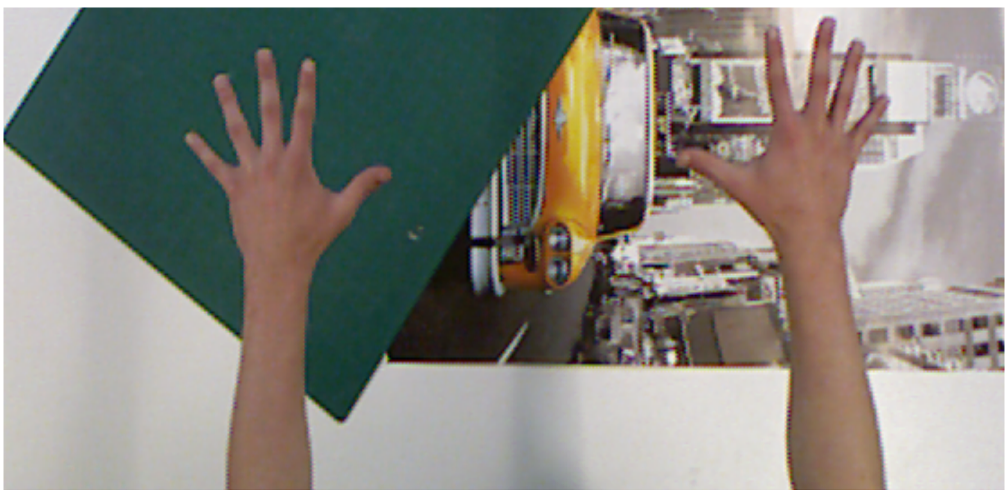


Figure 41: Standard arms drawn over table artifacts.

StructureArms. The extracted structure of the arm (base point, hand center, and finger points) can be used as the basis for an abstract representation of the arm. The main advantage over video-based embodiments is that structure information is much smaller than video, and can therefore be used even in poor network conditions. The structure points could also be used as anchors for artificial textures (e.g., cartoon arms or even images of people's real arms). The points on the structure can be accessed through KinectData (see above); as a demonstration, KinectVis includes a 'stick figure' effect in its API (Figure 42).

```
setSkeleton(bool enabled);
```

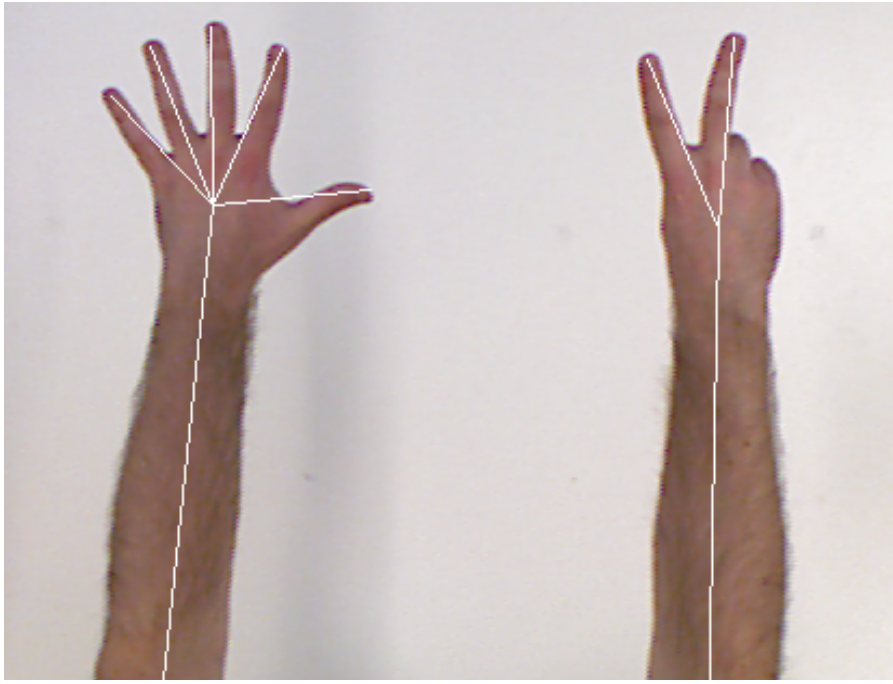


Figure 42: Stick-figure arms using structure points.

Visual Effects 1: Height Indicators

Showing the height of a gesture can greatly improve people's ability to interpret the meaning of that gesture. KinectViz provides three kinds of height indicator.

Circles. KinectViz can add abstract visualizations to the realistic embodiment. To show height, we add a circle that changes size and transparency based on the arm's height above the table (see Figure 43). KinectViz uses the lowest fingertip as the centre of the circle, which assumes that people are pointing downwards towards the surface.

```
setCircle(bool enabled);
```

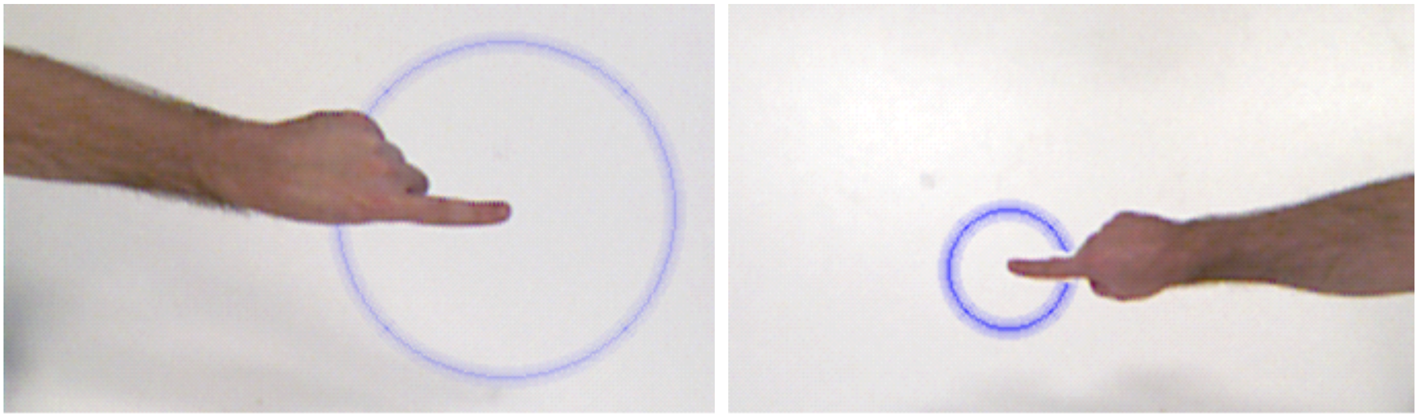


Figure 43: Circles -- abstract visualization of gesture height.

Shadows. Arm shadows have been part of analog video representations (e.g., [24]), but are not usually captured in digital embodiments. KinectViz provides a shadow effect in which the shadow is displaced to the side as the arm moves higher above the table (Figure 44).

```
setShadow(bool enabled);
```

Figure 44: Shadows -- simulated representation of height.

Gradients. Richer representations of arm height are also possible with KinectVis. One novel technique shows height as a false-colour gradient (see Figure 45): parts of the arm closer to the table are shown in cool colors (blues and greens); higher parts of the arm are shown in warm colors.

```
setGradient(bool enabled);
```

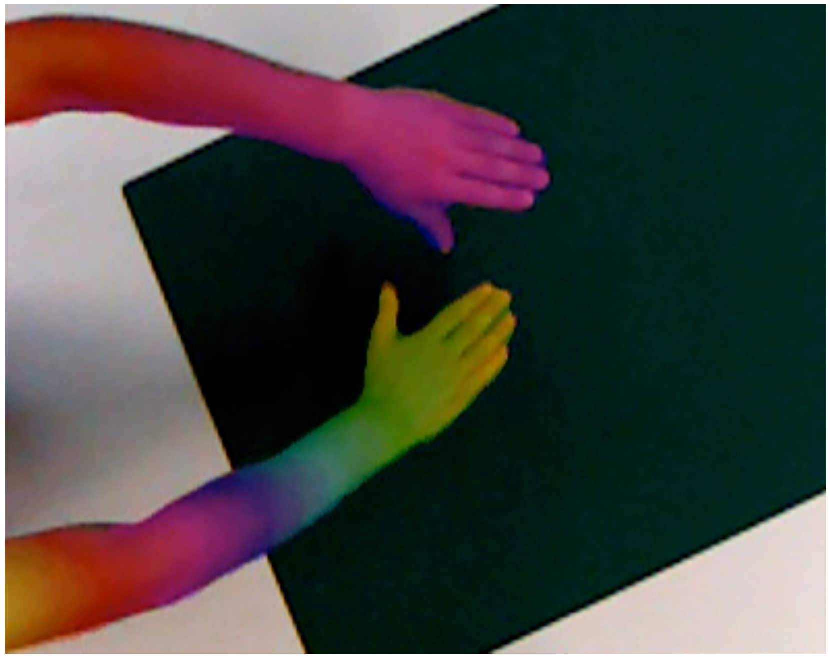


Figure 45: Gradients -- richer representation of arm height.

Visual Effects 2: Visibility and Identification Enhancements

Virtual embodiments can suffer from several kinds of visibility problems: they are harder to notice and harder to see than real arms [24, 41], they are often difficult to identify, and they can occlude objects on the table. To address these problems, KinectViz provides effects that manipulate the visibility of an embodiment and that assist identification.

Outlines. To increase noticeability, KinectViz includes an effect that draws a colored outline around the arm (Figure 46); the effect can be applied to one or all of the arms.

```
void setArmOutline(bool enabled, Color colour);
```

Figure 46: Outlines -- visibility on complex backgrounds.

Transparency. Embodiment visibility can be varied with a transparency level defined by the application programmer, or with a level that varies according to the height from the surface (Figure 47). Transparency can be used in several ways: as a basic way to avoid occlusion, or as a dynamic effect (e.g., to reduce salience of less-active participants).

```
setTransparency(bool enabled, int level);
```

Figure 47: Transparency – reducing visibility and occlusion.

Tattoos. KinectArms can apply virtual tattoos or markings to enhance identification (Figure 48). The tattoos move with the user's arm and are maintained even if the user removes her arm from the Kinect's field of view (see description of IDs above). Tattoos can be any image or text, and are defined by providing a directory for tattoo images. Tattoos are applied in the order that arms enter the workspace, and are set up similarly to other effects:

```
setTattoos(bool enabled);
```

Figure 48: Tattoos -- improving identification of embodiments

Tinting. A second effect for helping people differentiate between several embodiments applies a semi-transparent color to the arm image. The effect is similar to the gradient effect ofFigure 45, but with a single color.

```
setTint(bool enabled, Color color);
```

<u>Visual Effects 3: Motion Traces</u>

As discussed in previous chapters, gestures are easily missed in distributed settings because of the reduced salience of virtual representations, or network problems. KinectViz provides two effects that help solve this problem: motion lines and motion blur.

Motion lines. Fingers can leave motion-line traces as they move above the table. This effect is created by storing the finger point, connecting the points with lines, and fading older

lines to avoid cluttering the space (Figure 49). No previous embodiment technique enables these kinds of traces above the table surface.

```
setFingerTrace(bool enabled, int historySize, int maxAge);
```

Motion blur. A blur effect on the entire arm can help people notice and understand arm movement and gesture [29]. KinectViz produces this effect by overlaying previous frames on the current image (Figure 49), and increasing the transparency on older frames.

```
setMotionBlur(bool enabled, int historySize, int maxAge);
```

In both effects, the parameter `historySize` indicates how many previous samples to use in the effect, and `maxAge` controls how quickly traces fade away.

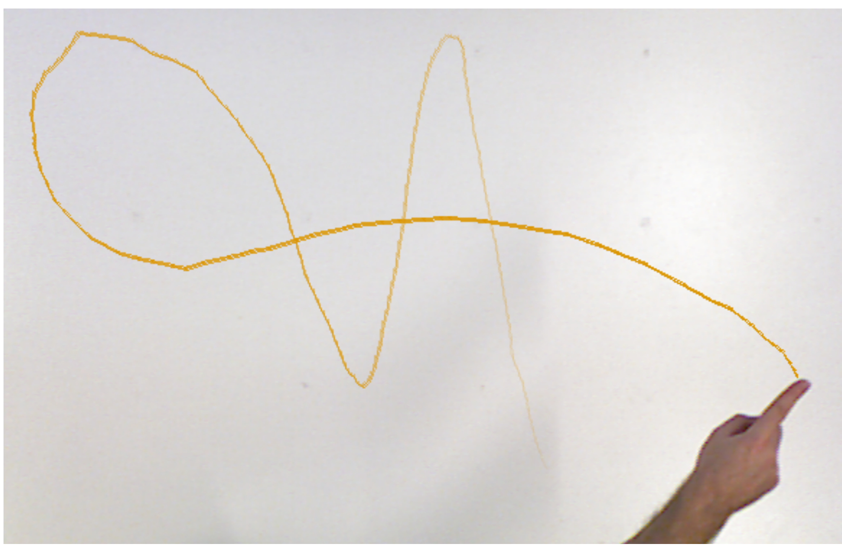


Figure 49: Motion lines (above) and motion blur (below).

<u>Assigning Effects to Specific Height Layers and Users</u>

People use different heights above the table surface to provide different information in a gesture [94] – for example, touching the surface, hovering, and medium and high levels above the table. Therefore, embodiments should be able to represent users differently depending on the height layer in which they are gesturing. All of the effects in KinectViz can be restricted to be visible only in certain layers above the table. For each of the API calls above, there are additional versions with two extra parameters – the lower and upper height of the layer in which the effect should be shown. This capability allows composite visualizations to be created without needing to add new code. For example:

```
	setCircle(bool enabled, float minHeight, float maxHeight);
	setMotionBlur(bool enabled, int historySize, int maxAge, float
minHeight, float maxHeight);
```

Similarly, additional versions of the API calls allow programmers to assign effects to single users – these calls add a parameter for the ID as determined by the library.

```
	setCircle(bool enabled, int armID);
```

Evaluation of KinectArms

Our experiences with KinectArms allow an analytical evaluation of several issues: the toolkit's performance, the complexity it adds to applications, and its extensibility, generality, and expressive power.

Performance

KinectArms is fast, providing visualized arm images at about 30 frames per second, even with multiple users. KinectTable uses the depth camera and a set of fast image-processing algorithms to carry out video separation, so a major computational expense seen in other systems is avoided. The Kinect hardware produces 30 frames per second, and KinectTable easily processes at the same rate.

KinectVis carries out additional image processing and other computation to add visual effects to the arm images. Depending on the effect, multiple or demanding effects in KinectViz, or a large number of arms that need to be processed, could reduce the frame rate depending on hardware. In our experience, with four users and simultaneous use of several of the effects described above, a standard PC (Windows 7, Core i5 processor) was able to maintain a rate of 30 frames per second. In distributed tables, networking constraints are more likely to impose the upper limit on frame rate (at remote sites) than KinectArms.

Complexity and Usability for Application Programmers

Using KinectArms with groupware applications is simple. API calls to KinectTable return arrays of hands and arms with unique identifiers that match users from frame to frame. Applications can choose how and where to display the hands and arms depending on their specific requirements. Similarly, KinectViz effects can be added with single API requests, and

the API allows developers to parameterize the effects (to different users and specific height layers) and change effects dynamically at runtime.

We carried out two informal tests of how easily KinectArms can be used for client application projects. First, we integrated KinectArms with a distributed photo-sharing application running on two interactive tables. With this application, users could manipulate shared images using the touch capabilities of the networked tables and see each other's gestures with KinectViz effects (e.g., Figure 38 and Figure 50). The integration of KinectArms and the photo-sharing application took less than one afternoon, suggesting that developers will be able to easily access KinectTable data and use KinectViz effects.

Second, we provided the KinectTable toolkit to another researcher in our lab (not an author) who needed a toolkit for exploring tabletop gesture recognition. Using calls to the KinectTable API, the researcher was able to begin developing gesture recognition systems within a day of starting development. The KinectTable hand structure provided enough information for the researchers to write software that recognizes hand orientations and postures (e.g., fingers up or fingers down), grasping gestures, and pointing gestures. This research is currently ongoing, and is continuing to use the KinectArms toolkit.

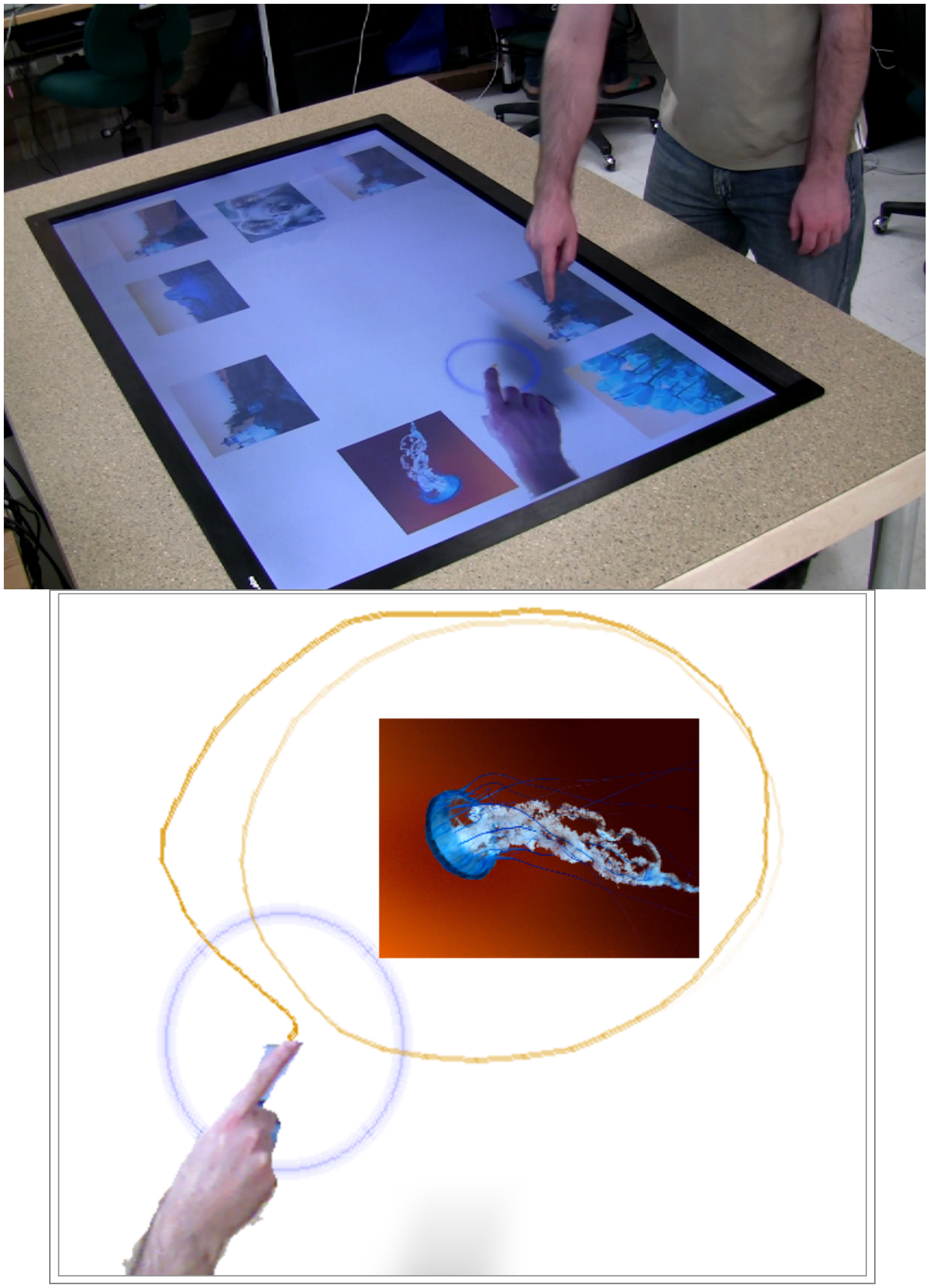


Figure 50: A remote user gesturing to indicate a shared image (below) in the photo-sharing application (above).

Expressiveness: Replicating Existing Techniques

A measure of KinectArms' breadth is its ability to duplicate other arm embodiments that have appeared in previous research. Using only the stock effects of KinectArms, we are able to

reproduce embodiments from a wide range of previous work, including VideoArms [14], DOVE [47], C-Slate [15], Fraser's approach visualization [34], and Yamashita's gesture replay [93].

For example, Tang's VideoArms shows hands and arms as two-dimensional silhouettes over a workspace. A later version that used a Microsoft Surface added 'trace pearls': small, fading lines where fingers came in contact with the surface [17]. KinectViz can replicate these two techniques using its basic video arm representation, and a motion-line effect that is limited to a small layer just above the surface (0-1cm). When a pointing finger (defined as the lowest finger on a hand) enters the low layer, traces are drawn. This embodiment can be built with the following calls:

```
setFingerPointer(true, 0, 10);
setTint(true, 0, 800, grey, 255, 0);
```

The monochrome image of users' hands and arms in VideoArms can also be changed to full colour at no cost (part of the basic video arm representation in KinectArms).

As a second example, both DOVE and C-Slate overlaid video embodiments on a workspace and used transparency either to reduce occlusion or to indicate that hands were further from the tablet surface. In addition, DOVE provide contact traces when the user touched the surface. These techniques can be replicated using transparency, shadows, and motion lines:

```
setTransparency(true, 50, 0, 800);
setShadow(true, 10, 800);
setFingerPointer(true, 0, 10);
```

In addition, KinectArms shows its expressive power in two other ways. First, by combining effects for different users and different height layers, KinectViz can create complex representations that go well beyond existing examples. Second, KinectVis provides several novel effects not seen elsewhere (e.g., outlines, color gradients, stick-figure arms, tints, and above-the-table motion lines); and these new effects were simple to build by using KinectArms' high-

resolution images, extracted arm structure, and detailed height information, capabilities which are not provided by any other current toolkit.

Extensibility: Defining New Effects

Although KinectViz has a wide variety of stock visualizations, developers may wish to create their own. The extensibility of the KinectArms toolkit is shown in the simplicity of the process of creating a new effect. KinectTable provides easy access to hand and arm position data and KinectViz automatically applies any enabled effect when applyVizEffects is called. Adding a new effect involves three steps:

1. Add a flag in the KinectViz class for the new effect;
2. Create a method that performs image manipulation using the KinectData structure as a basis; and
3. Add a call to the new method in applyVizEffects that executes conditionally on the flag being set to true.

Generality

KinectArms works with a wide variety of tabletop setups and table hardware. Tables with existing touch sensing can be used alongside KinectArms (and can be used to augment KinectViz effects by providing additional accuracy for touch events). Top-down projection, a problem for most video-separation techniques, works well with KinectArms, since separation is done with the depth camera, not the colour camera. KinectArms can also be used with non-interactive tables (similar to Ishii's TeamWorkStation [113]). One-way connections can be used to show remote users the contents of a normal table, thus supporting collaboration over paper-based artifacts.

Conclusion

Representing rich and subtle gestures in distributed tabletop groupware can be expensive, programmatically complex, and can require specialized hardware and lengthy calibration. KinectArms, a toolkit that helps groupware developers build arm user embodiments with a minimum of effort, is a solution to the problems of capturing and displaying arm embodiments for remote work. KinectArms handles video capture and separation, determines fine-grained height above the table, and provides visualizations that improve arm visibility, representation of gesture height, and movement. The KinectArms toolkit allows simple replication of most existing arm embodiment techniques, and enables the creation of many new types of representation.

DISCUSSION

In Chapter 1, four challenges were identified as barriers to creating remote embodiments that support deixis in distributed collaboration:

1. deixis over surfaces is poorly understood;
2. the embodiment design space is not described;
3. embodiment designs are in only two dimensions; and,
4. deixis is hard to capture and show over surfaces.

In this chapter, each of these challenges will be discussed in the context of the research described in Chapters 3 to 6. In particular, the next sections will explore the limitations in this work and discuss how the above research can be generalized to other domains.

Understanding Deixis

The observational experiments described in Chapter 3 significantly improve our understanding of deixis over surfaces. These lessons can be extended to provide some generalizations for design and for other kinds of gestures. However, the use of maps in the experiments may also limit the generalizability of the findings. These issues are discussed below.

How Height can be just Height

In the observations reported in Chapter 3, participants sometimes used the height of a gesture to represent the real-world height of an object represented on a map. Because the context of these gestures is often clear in other ways, such as the state of the conversation or accompanying verbal cues, participants were not confused. These same contextual cues remained in force in the distributed and simulated distributed experiments described in Chapter 6. Despite the possibility that the spatiality of the domain could be conflated with the spatiality of the gesture, even without the presence of some of the disambiguating cues (e.g., study 1 and 2 of Chapter 6) study participants were able to differentiate between the two kinds of gestures.

The ability of participants to easily differentiate between the different uses of height in a gesture, despite the lack of certain contextual cues, suggests that height representations geared to suggesting variations in confidence, specificity, and emphasis are equally able to be interpreted for other kinds of intent in the use of height. A more rigorous examination of these situations should be performed, but it seems likely that the embodiments proposed in this dissertation will not add confusion to communication in the face of alternate uses of height in gestures.

How Interactions with Maps can Generalize to Other Artifacts

The kinds of deixis over maps described in Chapter 3 were categorized into three gesture types: indicating points, paths, and areas. These gestures were used as the foundation for the design and evaluation of embodiments to represent those gestures in distributed settings, as described in Chapters 4, 5, and 6. This work assumed that these gestures correspond well to the majority of deixis that is performed over a wider variety of shared artifacts. This assumption is in two parts: first, the largest portion of deixis, regardless of the circumstances in which it takes place, is constrained to indicating points, paths, and areas; and second, these kinds of indicatory gestures are applicable to all artifacts.

Even with maps, describing deixis as gestures that indicate points, paths, and areas fails to completely describe the corpus of gestures. The remaining gestures, however, are quite rare. For example, gestures that divide the workspace (e.g., “this side of the map”) were seen but not included in the model because of their relative rarity in the both the laboratory and field studies. However, the frequency of these gestures may have been limited by the kind of tasks during these studies: largely information transfer tasks. Negotiation tasks (e.g., [7]), may generate more division gestures (an example of a non-pointing deictic gesture), for instance. Nevertheless, there is evidence that the assumption is correct: despite possible moderate increases in prevalence of

certain kinds of gestures for certain tasks or domains, the key indicatory gestures identified in much of the literature on deixis are points, paths, and areas (e.g., [1], [4], [11], [35]).

The second assumption, that points, paths, and areas can be described for all artifacts for which deixis is used, is a more challenging assumption to support. If artifacts on a shared display are considered to be either targets or distractors during deixis, then points, paths, and areas do successfully describe most of the tasks that can occur. People can indicate a target (point), move between targets (path), indicate part of a target (area) or multiple targets (area), avoid targets (path), or perform combinations of those gestures. The 'division' gesture suggested above is a possible case where points, paths, or areas do not describe the gesture, but even in that circumstance, it is possible to consider it as a gesture indicating an area (e.g., half of the workspace).

It seems likely that the model of deixis as indicating points, paths, and areas on surface workspaces holds for the majority of circumstances, allowing the research described above to be generally relevant for most kinds of collaboration. Exceptions will be special cases arising from unusual circumstances or certain kinds of workspaces. More likely, however, some kinds of workspaces will encourage non-pointing deixis, creating an environment where the assumptions about showing a pointing finger as an abstract representation will not hold. This possibility is discussed further, below.

How Maps Can Influence Deixis

Maps are frequent collaborative artifacts, used for everything from day-to-day navigation tasks to large-scale planning. However, when maps are used in a distributed collaboration, they may encourage different kinds of user behaviours than other kinds of artifacts, such as collections of photos, puzzle pieces, or text documents. These differences may limit the applicability of the work in this dissertation.

Maps can be boundary objects, or objects that occupy different kinds of roles in different environments [114] and, depending on the role, encourage different kinds of interactions between people. For instance, people might turn a paper map to orient themselves with respect to their environment or their direction of travel, an activity seen frequently in the use of tourist maps *in situ* [115], and people can turn maps to orient themselves to a collaborator's point of view. Map turning can lead to different kinds of deixis: in the field study described in Chapter 3, participants on opposing sides of maps sometimes rotated their hands so that pointing fingers were oriented as though they were side-by-side (a pointing gesture very similar to one that could be used to point at oneself). This form of deixis was not observed when maps could not be rotated. De facto rotation can be achieved by moving around maps, a technique observed frequently when participants worked around fixed maps. This kind of change in orientation is subtly different from map rotations: side-by-side collaboration can be interpreted as more cooperative [116].

Developing tools to change the orientation of digital objects and understanding how orientation influences collaboration are major reseach topics for groupware researchers (e.g., [116], [117]). Collaborations can be influenced by the kind of rotation and orientation possible with collaborative artifacts. The nature of maps (single, large artifact) and the lack of available rotation tools during the experiments described in Chapters 3 and 5 may limit the applicability of the results to collaborations that do not involve maps. In particular, pointing at out-of-reach targets may occur less frequently where collaborative artifacts can be rotated or where people are free to move around the surface.

Another way in which pointing may differ when people collaborate over maps is the use of above-the-shoulder gestures. Chapter 3 describes how participants rarely pointed above their

shoulder level in the laboratory study. However, the task was constrained: the map was relatively small (with effort, participants could reach across), and the room had no windows and therefore the map was not obviously in situ, despite that the map showed the local city. A map more obviously in situ, for example, near windows with a panoramic view of the spaces represented by the map, may elicit different kinds of pointing activities. Such pointing to indicate areas outside of the workspace occurs not just for maps, but for other boundary artifacts and for tasks where some collaborators or resources are outside the shared workspace.

The research described in the previous chapters fails to capture how gestures directed at artifacts outside of the workspace or gestures that occur above the shoulder may be different from gestures that occur below the shoulder and target artifacts within the workspace. Such gestures may need to be represented differently than with the representations introduced and evaluated in Chapters 4, 5, and 6.

Describing the Embodiment Design Space

Chapter 4 describes the embodiment design space in terms of abstract designs, realistic designs, and hybrid designs. It also examines how each of these design approaches succeeds and fails at representing the location, movement, and morphology of deictic gestures. The analysis shows that neither realistic nor abstract design approaches are sufficient by themselves for showing all aspects of deictic gestures.

The designs in Chapter 5 and the tools provided by KinectTable (in Chapter 6) are a small step in solving this problem by using hybrid embodiments that incorporate both abstract and realistic design elements. However, there remain some limitations to the work in Chapter 4. The following sections explore some of these limitations: how emphasizing or increasing the amount of information in an embodiment may interfere with communication; what limits there

might be to the information conveyed by an embodiment; and what problems there are for representing height information in distributed embodiments.

Chapter 5 found that including temporal traces in the embodiment designs had significant advantages in distributed settings. This finding suggests that temporal traces may be of use in other collaborative contexts, such as histories of activity over time. Representations of height, also found to be of value in Chapter 5, could be of value for collaborations performed over vertical surfaces (e.g., [34]). Both these possible extensions to embodiment design and the limitations to the work in Chapter 4 are discussed below.

How Information can Interfere with Communication

Although hybrid embodiments can combine realistic and abstract techniques to show a wide variety of information about deictic gestures, combining both kinds of representations could interfere with embodiment interpretation. For example, the abstract height representations introduced in Chapter 4 occlude fingertips if placed over a realistic representation of a finger, possibly reducing the expressiveness of the realistic representation. However, inverting the layers (displaying the hand and finger over top of the abstract touch representation) would cause occlusion for touch representations, reducing the precision with which touches can be shown. This problem might be addressed by increasing the transparency of one representation, but that presumes that transparency is not already used for showing other information about the user's gesture.

Too much information in an embodiment can also increase the chance that unneeded information distracts users, interfering with collaborative activites. As mentioned earlier, some information in an embodiment can be emphasized to draw attention to important aspects of user behaviour, such as the temporal traces and contact traces used in Chapters 4, 5, and 6. Attention, however, is limited: if attention is drawn to one aspect of an embodiment, other aspects receive

less attention. Designers should consider that each new visual element in an embodiment design will compete with the other elements for limited attention, and design accordingly.

Limited information about gesture characteristics (such as gesture height or temporal history) are not necessarily a problem for all collaborative activities, however, and designers may choose to limit the amount of gesture information in embodiments designed for specific tasks or where there are particular aesthetic goals. The work described in Chapters 4, 5, and 6 does not consider how communication can be impeded by too much information about gestures or how emphasized gesture characteristics may interfere with the interpretation of other less-visible but still important characteristics, some possible limitations in the utility of the hybrid design approach.

<u>Representing Information: How there may be Cognitive Limits</u>

There is limited research that has explored how much information can be successfully conveyed through user embodiments. Stach *et al.* [36] found that a large number of discrete pieces of information could be added to an abstract embodiment and users would still recall the purpose of each piece. They theorized, however, that their study participants were using only a select few pieces of information at any time and that some pieces of information were not used at all. Their findings suggest that not only is there a limit on what amount of information is usable while interacting with remote embodiments, but that there may be a limit on the amount of information that can be understood and used (although they did not manage to find such a limit, if it exists).

These limits, what those limits may be, and how the limits may vary among people are critical pieces of information for embodiment designers. In particular, if designers add many different pieces of information to embodiments, they may find that some or many are ignored or useless. This would compound the potential problem of distraction, discussed above, since not

only might a piece of information not be used, but it might distract from more important information as well.

The work in the previous chapters does not take into account these limits. It is possible that embodiment designs with hybrid approaches will be limited in their usefulness beyond representing height and one or two other gesture characteristics.

How Temporal Traces can be Used Differently

In Chapter 4, historical information about above-the-surface movement is represented very similarly to how the current state is displayed: the only difference is that the historical traces fade over time. With a short interaction history displayed, this technique works well. However, with a long interaction history (e.g., more than 60 seconds), the display would be cluttered with almost-identical copies of the embodiment and the current state would be difficult to detect.

The designs in Chapter 4 and 5 are limited in their ability to clearly represent longer term histories of gestural communication. For longer-term historical traces, the past may be shown with noticeably different symbols or visual characteristics than the representation of where the gesture in the present time. The contact traces used in the Chapter 5 experiments are an example of this technique: rather than leave copies of the arrow pointer as historical traces, lines and ripples were used. Such representations can be smaller than the original, thus reducing display clutter, yet still convey some important aspects of the gesture (such as its two-dimensional location).

How Temporal Traces can Generalize to other Collaborative Settings

Chapters 5 and 6 established that short term histories of gestures were valuable in distributed settings. This finding can be generalized to other distributed contexts where gestures are important. Gestures provide nuance to communication even when speech is present, but deixis is unique in that it provides a clear indication of areas of interest even without associated

speech. When this information is collected and visualized, it has the potential to quickly answer questions that would otherwise be very difficult or time consuming to answer.

For example, in many crisis management Emergency Operation Centres (EOC), there are a variety of workspaces, each dedicated to differing kinds of information such as logistics, planning, or contact information. When emergency managers are away from the EOC for extended absences, it can be difficult to understand what information has changed. Note that this is not simply a question of which information has changed; indeed, much information may have changed but much of it will be relatively unimportant. Some small changes in a model, however, may have been the subject of large amounts of attention. Tracking the deixis in the room over a period and showing long term traces of that deixis (e.g., as a heat map) can provide critical clues to members of a new shift, allowing them to focus on the important pieces of information that need to be tranferred between shifts. Such visualizations would allow people to gain an awareness of the attention over time paid to artifacts in the workspace.

In other scenarios, such as distributed urban planning (or other geographical resource-based planning), different kinds of questions can be important. After a collaborative session, stakeholders may be interested in reviewing parts of the collaboration to isolate decisions made about particular geographical features. Without historical gesture information, this process can be time-consuming, often requiring the replay of audio and video recordings of the entire session. However, a visual history of gestures can help identify at what time during the collaboration the geographical features were discussed, permitting stakeholders to jump to the correct location in the audio and video recordings.

In this case, long term historical traces showing the height of the gestures can be of help. There may have been several occasions during a collaborative session when participants pointed

at a location of interest. However, since lower gestures and tapping gestures are associated with greater specificity and emphasis, a historical trace with height information can more easily show when (or what) geographical features were of specific or emphatic interest.

Designing Embodiments for Three Dimensions

A key result from Chapter 3 was the identification of the height of a gesture as an important aspect of deixis. Chapter 4 introduced approaches to embodiment design that can show the height of a gesture and Chapter 5 described evaluations of designs based on those approaches. In particular, these chapters show that representing the height of gestures is valuable for interpreting deixis in distributed settings; that a visual representation of height is interpreted by people in a distributed setting in the same way that the height of a gesture is interpreted in a collocated setting; and that changes in height, such as tapping, are at least as important to represent as the global height of the gesture. The next sections extend this work by discussing how information can be dual-encoded or decomposed in embodiments, how the power of speech can overcome many of the limitations of embodiments, the power of traces in distributed settings, how showing deixis with embodiments can help with the representation of other kinds of gestural communication, and how the designs from Chapter 4 were adopted, adapted, and co-opted by the participants of the Chapter 5 studies.

<u>How Information can be Dual-Encoded and Decomposed</u>

There is value to dual-encoding important information (using two different visualization techniques to represent the same piece of information) and decomposing compound information (breaking information into its constituent parts, if possible). The pilot studies described in Chapter 4 found that high visibility was a key component of embodiment design when the embodiment was intended to be used in realtime collaboration over a variety of different backgrounds. This problem was solved by dual-encoding much of the information: touching was

encoded as both a shape change and a colour change; above-the-surface movement changed size, shape, and transparency. The results from Study One (Chapter 5) reinforced this finding with the only aspect not dual-encoded in the design, the contact traces. Participants struggled to identify the start and end of contact points and misinterpreted single sustained contacts (paths) as multiple points. When the design was adjusted in Study Two to encode contact traces in two ways, lines and ripples, those problems disappeared.

This experience also highlighed the importance of understanding the components of gestures at an atomic level: a sustained contact with the table can be broken into a touch, a slide, and a release. By representing each of these separately, or decomposing the information as much as possible, the gestures became more understandable to participants than when only one visualization was used to represent all three components.

<u>How Speech can Help</u>

Although speech is a very powerful communication tool, gestures and other non-verbal communication channels are important for understanding context and nuance in ways that are difficult to convey quickly and easily with language. This is not immediately apparent, since when study participants were able to converse through speech (in the third study of Chapter 5), they appeared to have little difficulty collaborating even with the plain telepointer. These results indicate the power of accompanying speech to clarify and disambiguate gestures, a finding in line with previous work (e.g., [5]).

However, although speech may be used to overcome the lack of visual cues during deixis (in the way that telephone communication remains effective despite the lack of visual communication channels), gestural communication adds subtlety to language, facilitating mutual understanding. Participants made use of all of the information that was available in the enhanced embodiment, altered their gestures to accommodate a richer and more efficient gestural language

based on the affordances of the embodiment design, gestured more naturally where those alterations did not take place, and also strongly preferred the enhanced embodiment to the plain version. Study participants also preferred the additional communication tools provided by height-enhanced embodiments, suggesting that people are aware of how gestures enhance communication. More profound effects may be found when height enhancements are incorporated with realistic, video-based embodiments (as can be done with the KinectArms toolkit).

The adoption of embodiment features despite the power of speech suggests that representations of gestures and other physical characteristics should be an important component of distributed groupware, regardless of the presence of alternative channels of communication.

How Traces are Valuable

Traces, and contact traces in particular, were found to be valuable for disambiguating similar gestures, but it is likely that they would be valuable for any situation where small changes in gesture characteristics are hard to see or where distractions are prevalent. Although other research has introduced the idea of contact traces or touch representation in various forms, the studies in Chapter 5 were the first to comprehensively examine the value of contact information as feedthrough to remote collaborators, rather than strictly as feedback. The first and second studies in Chapter 5 showed that contact traces (ripples in the first study, lines in the second) are powerful visual indicators to remote collaborators. In the first study, the contact trace made it much easier for people to recall and keep track of the multiple points in the gesture – and this is particularly valuable in remote collaboration, where the user's embodiment cannot come close to the visual salience of a real body leaning over the table to point. Others have noted that actions in large distributed tables are easy to miss [24], and the first study helps confirm the idea that traces can help groups overcome this problem.

The final study of Chapter 5 unambiguously found contact traces to be valuable to both remote and local users. Not only were users enthusiastic about the usefulness of contact traces, but contact traces encouraged people to touch the surface of the table only when precision was called for, a behaviour more in line with how people naturally use touch in deixis over surfaces. Contact traces also enabled the optimization of gestures: using the plain telepointer, people frequently repeated gestures on the surface, circumscribing an area several times or drawing the same line repeatedly along a path. In the enhanced condition, this behaviour did not manifest itself, since the trace served the same function, replacing the repetitive movement with the fading line. This second behaviour suggests that contact traces can substitute for the stylus sketching that participants used in Li *et al.* [75] to identify regions of the workspace.

Overall, touch traces for showing recent actions appears to be a worthwhile design feature to add to remote embodiments in most contexts. In particular, contact traces will be valuable in settings with large displays, many collaborators, or other awareness challenges that could result in momentary user distractions.

<u>How Showing Deixis May Improve Other Gestural Communication</u>

Although the height representations presented in Chapter 4 do little to show non-deictic gesturing, KinectTable can be used to capture and display most kinds of gesticulation. Any gesturing that occurs in the volume of space detectable by the Kinect camera can be replicated with reasonable fidelity. In other words, the technology and designs necessary for capturing and representing deixis in three dimensions are also sufficient for capturing and representing other kinds of gestures.

There are, however, some limitations. First, KinectTable remains a top-down capture technology and some gesture information can be occluded by arm and hand positions or by the presence of additional people in the workspace. Second, gestures out of view of the camera will

not be captured or represented. This is a particular problem when workspaces are situated in a larger ecology of interaction, such as a table in a room with whiteboards. In this situation, information in other, associated workspaces would likely be the topic of conversation, meaning that deixis will be missed when referencing these off-camera locations, and users may frequently turn away from the table to converse, even while the table remains the primary workspace.

It is possible to conceive of alternative designs where gestures are captured regardless of where they occur in a room, using, for example, motion capture technology; but two key design features of KinectTable are its low cost and rapid deployment. Solutions for capturing more of the workspace ecology would need to be re-imagined to retain those design qualities.

How People Adopt, Adapt, and Co-opt Designs

Any time people are provided with a new mechanism for communicating, they adapt their existing communication methods, adopt new ones, and co-opt aspects of the design for their own purposes. With the telephone, for example, people have adopted new communication methods to account for the lack of visual connection, adapted existing communication methods (e.g., the greeting and farewell are different on the phone than in person), and co-opted design elements of the phone, such as the keypad (such as with the development of interactive voice recordings like automated phone menu systems).

The amount this occurs can be a loose metric for the success of the design. Adoption, adaptation, and co-option are signals that people are investing time and effort in understanding and using the design. Adoption shows that people find the design useful, adaptation shows that people are find it sufficiently useful to change their regular practices, and co-option shows that the design is successful enough that people want to use it in additional ways that the designers did not originally envision. In the case of embodiment design for supporting gestures, a successful design should inspire new gestures, encourage people to change some existing

gestures to take advantage of new opportunities provided by the design, and suggest ways in which a user can co-opt design features to improve their interaction with the system or with other users.

Although the studies in Chapter 5 were intended to evaluate how successfully an embodiment can represent height to remote collaborators, it is almost impossible to fully duplicate natural gestures for a remote participant, and so gesture visualizations will always to some degree be new visual languages rather than simply a representation of the natural gesture itself. In the third study, although generally more natural pointing behaviour was observed with the enhanced telepointer, people also used new behaviours to take advantage of the facilities provided by the enhancements, in particular when participants used the displacement of the ellipse to identify out of reach targets or the expansion of the ellipse to indicate areas. This raises several questions: for example, whether people should be asked to learn a new visual language in order to gesture over tables, and whether producing gestures for either co-located collaborators or remote partners will produce a substandard representation for the other party.

Capturing Deixis over Surfaces

Chapter 6 describes a system, KinectTable, that captures three dimensions of gestures over surfaces cheaply, quickly, and easily. The next sections discuss how this system can be useful outside of table-based systems, how representing gestures with 3D technology is not the answer, how KinectTable can solve the problems identified in the motivation for this research (TEK elicitation in the Canadian Arctic), and some limitations due to the challenge of differentiating deictic gestures from other kinds of gestures.

<u>How KinectTable is Limited and can be Extended</u>

KinectTable is intended as a tool to facilitate the inclusion of embodiments in distributed groupware. Despite that the examples in the previous chapters are for tabletop surfaces, there is

no practical reason that KinectTable could not be used for vertical surfaces or open areas as well. Vertical surfaces have some additional occlusion challenges, especially for deictic gestures, but these might be resolved through alternate sensor placements. KinectTable is flexible enough that it can be used to extend proximity-based software (e.g., [118]) into distributed settings. The visual techniques provided in KinectViz can be applied to whole body movement as easily as hands and arms. Since the Microsoft Kinect SDK already supports whole-body recognition, this would be a relatively simple accomplishment.

KinectTable is currently limited by its relatively low capture resolution compared to modern displays, a problem that will be more apparent on larger and/or higher resolution displays than those used to demonstrate KinectViz in Chapter 6. With larger, higher resolution displays, KinectTable will become effectively less accurate, with one KinectTable pixel mapping to several display pixels. In addition, although it captures three-dimensions of a deictic gesture, it is unable to capture three dimensional models of people in space, only a single-sided topology. This works well for displaying user embodiments on screens, but does not work effectively for creating avatars in virtual environments. Thus, KinectTable cannot be easily extended to provide embodiments in the form of virtual avatars.

How 3D Technology is not the Answer

Recent advances in 3D display technology have allowed the distribution of 3D movies and television in homes and workplaces. Intuitively, since gestures are three dimensional movements and hands and arms are three dimensional objects, the use of 3D technology to show them should obviate the need for many of the embodiments available through KinectViz. However, current 3D displays require a perpendicular viewpoint not accessible to tabletop users, and the state-of-the-art in accessible 3D technology does not support true perspective-shifting movement (e.g., moving the point of view to look around an occluding object).

Until 3D technology improves to allow oblique angles of view and to permit users to move their heads to change their perspective, the 2D embodiments available with KinectViz are a valuable contribution to gesture visualization.

How this Research Solves the Motivating Problem: TEK in the Arctic

Chapter 1 introduced a motivating problem for the work described in the previous chapters: gathering traditional ecological knowledge (TEK) from aboriginal peoples in the Canadian arctic. The problem had four parts:

1. Researchers needed to be able to reliably and accurately capture participant gestures during interview sessions over maps.
2. The capture system needed to be easy to set up and require a minimum of calibration.
3. The capture system needed to be transparent to users: participants were uncomfortable with new technology and unwilling to use devices (such as a stylus).
4. The participants needed to be able to tell their stories without interruption from technology or the researchers present.

KinectTable is close to a solution to these problems. Combined with simple map-display software, it can reliably and accurately capture gestures made over maps; it is simple to set up and calibrate; it has no obvious sensors with which participants must interact; and would be largely transparent to eliciting TEK.

Some challenges remain. The task of integrating KinectTable and a GIS remains incomplete and may prove to be time consuming or difficult due to the complex nature of GIS software. KinectTable does not track users other than through their rough location around the table. Users who move around the table to gesture will be mistaken for new or different users. In addition, gesture visualizations have not previously been used as metadata in GIS and some

effort would have to go into deciding how to store these data and how to show them in conjunction with other kinds of GIS information.

Beyond the technical challenges related to GIS, there remain a host of problems related to the real-world deployment of KinectArms. The harsh transportation requirements (-40 C and exposed to weather), unreliable power, and inconsistent workspaces create engineering challenges that must be addressed. The TEK process may also extend to spaces not covered by the Kinect camera, and it may be that adjustments to how KinectArms collects data (and how the data are interpreted) may be necessary. There may be social issues as well: despite its lack of attachable sensors, KinectArms-based systems still used cameras and need wired connections with computers. This may create interaction barriers with some First Nations elders.

The most logical next step in the deployment of KinectArms as a solution for the motivating problem would be a brief field trial accompanied by a trained ethnographer who could identify the successes and failures of the system under realistic conditions.

Differentiating Gestures

One key limitation of KinectArms is the unsolved problem of identifying when a gesture is deictic and adding appropriate information (e.g., showing a height shadow), and when a gesture is non-deictic whereupon additional visual information is not added. In Chapter 3, 93% of the observed deictic gestures were pointing gestures that incorporated the index finger. This is a good clue for the identification of deixis, but the remaining 7% of gestures were widely different in their morphology. The long tail of gesture morphologies means two things: the gestures are difficult to classify meaningfully; and some gestures are likely to be similar (or even identical) to non-deictic gestures.

The long tail of gesture morphologies are difficult to classify because of a lack of common salient features (such as an extended index finger). Many gestures are very similar to

others except for relatively small details (a finger curled or not; a subtle change in palm orientation) to such an extent that the gesture corpus might be considered a continuum of types rather than distinct groups for the purpose of any extended classification. This kind of classification scheme, where many instances fall between archetypes, is difficult to use for gesture recognition, which is a key requirement for differentiating between deixis and other gestures for the purposes of supporting gesture-specific enhancements.

These problems are compounded by gestures that are context-dependent in their interpretation. For instance, the division gesture described earlier (dividing the workspace into pieces) can be equally used as a non-deictic gesture: e.g., while accompanying the verbal cue, "We divided the pie in half" [11]. This means that to provide accurate gesture differentiation, a system must understand natural language as well. Since natural language comprehension by computers is still an unsolved problem, a true gesture recognition system that differentiates between deictic and non-deictic gestures remains some distance in the future.

Despite these issues, designers can still rely on users to differentiate between gesture types. This means that systems can likely be generous in their application of enhancements to gestures, assuming that most of them are deictic and relying on the users to identify when gestures are not intended as deixis. Furthermore, when users have sufficient feedback about the appearance of their embodiments, they adapt their movements to accommodate for how they wish the embodiments to appear. This behaviour was seen in the evaluations described in Chapter 5 and can be relied on as a further filter of non-deictic gestures.

Conclusion

This chapter has discussed how the research in this dissertation can be generalized to other domains and problems and also how the research is limited. I reviewed how the use of maps in the qualitative studies of Chapter 3 may have meant that some kinds of gestures were

under-represented in the experimental studiesm but also discussed how map-focussed deixis can be generalized to other settings. I discussed how information representation is a critical design decision, identifying that too much information can interfere with communication and that there may be ccognitive limits to the information we can understand in a visualization. I also introduced some new ideas about showing long term historical traces and how traces can be extended to other collaborative settings.

In discussion of the findings of Chapter 4 and 5, I reviewed how three-dimensional embodiments benefit from the dual-coding and decomposition of information and how the power of speech can overcome limitations in designs. I also explored how the designs described in this dissertation may be sufficient for showing non-deictic gestures, how designs can be evaluated through their adoption, adaption, and co-option by users, and the enormous value of traces in distributed collaborations.

Finally, this chapter has explored the extensions and limitations of the KinectArms tool, how this tool is limited in its ability to differentiate between deictic and non-deictic gestures, but also how it might solve the motivating problem of this dissertation by assisting with the capture of TEK in the arctic.

CONCLUSION

This chapter provides a summary of the work, a review of the contributions of this thesis, and directions for future work.

Summary

Deixis is an important part of communication, yet the state of the art in embodiments for distributed collaboration fail to effectively show the full richness of gestures. This is a problem because without expressive deixis, communication can be limited, difficult, be misinterpreted, or fail completely. Surface-based, distributed collaborations are used in many critical settings, including international diplomacy, crisis management, and natural and urban resource management. In such settings, failures in communication can be catastrophic.

This dissertation has described research that solves the problem of showing gestures over surfaces during distributed collaborations. The problem was solved in several steps. First, I achieved a greater understanding of how people perform deixis over surfaces and what portions of deixis are important for people to see so that they can correctly interpret the gesture. Next, Iexplored how realistic and abstract designs can best represent those gesture characteristics I found to be important in my observations. I then created designs that showed the height of gestures and the history of movement over the surface, characteristics of gestures that I had found to be important but that have not been represented well by remote embodiments.

I tested the embodiments I had designed by comparing them to basic telepointers and each other. I found that embodiments that show height improve gesture differentiation, target identification, and allow people to gesture more naturally during distributed collaborations. With this knowledge, I designed and oversaw the creation of a toolkit, called KinectArms, that supports the capture of gestures over surfaces and the representation of those gestures in

distributed groupware. This toolkit has now been released as open source software and is in use at several universities and colleges in the Americas.

Based on the work described in this dissertation, groupware applications can now be built with a minimum of effort to include embodiments that show gesture height and historical traces. These embodiments are a significant step forward in representing deixis in distributed collaborations and represent a major step forward in undertstanding how deixis works, how it can be represented, and what representations are effective.

Contributions

The primary contribution of this dissertation is a solution to the problem of representing three-dimensional deictic gestures in distributed groupware applications. In providing this solution, this dissertation makes six smaller contributions:

1. It provides a characterization of deixis over surfaces grounded in laboratory and field studies. The characterization is on four major axes: morphology, atomic movements, gesture height, and common variations.
2. It describes the design space for embodiments that show deixis over surfaces, dividing it into abstract and realistic representations and introducing the concept of hybrid representations.
3. It identifies of the height of gestures above a surface as an important aspect of deixis, in that height conveys qualities such as emphasis, differentiates between similar gestures, conveys real-world changes in height, attracts attention, and allows the indication of distant artifacts. In particular, this dissertation identified three layers of height used over surfaces during deixis: touching, hovering, and a layer above hovering. It also identified variations in height during a gesture (e.g., tapping) as conveying similar levels of emphasis, confidence, and specificity as gestures that touch the surface.

4. It introduces set of effective designs for showing height above the surface in embodiments. The designs were built through an extensive exploration of prototypes and tuned through pilot studies, and evaluated with three experimental studies. The studies showed that including height information in embodiments improves the recognition and accuracy of gestures on remote tables, allows users to gesture more naturally in distributed settings, and that users prefer such embodiments and will even extend their gestural language to accommodate them. Generally, people interpret height representations in distributed settings in the same way as they do gesture height in collocated settings.
5. It introduces a novel toolkit for capturing and showing rich and understandable gestures over distributed surfaces. In particular, the toolkit provides several ways to show the height of gestures using abstract representations, similar to those designed in this work, in combination with realistic hand and arm representations.

Together, these contributions provide a significant improvement to our understanding of how deixis is used over surfaces and how groupware can support the use of deixis in distributed settings.

Future Work

The work described in this dissertation leads to several promising avenues for future work: showing rates of changes in height; emphasizing local variations in height changes over global variations, allowing tapping to be further emphasized; managing many embodiments over a surface, rather then the one or two used in most previous work (including the previous chapters); differentiating deictic gestures from other kinds of gestures and supporting each differently with embodiments; using KinectArms as tool for testing distributed collaboration

support technologies; improving KinectArms with better tracking, more visualizations, and better calibration tools; and using KinectArms-based groupware to support TEK elicitation in the wild.

Representing the Rate of Change in Height

Study Two found that local height variation can trump the global height of a gesture in conveying specificity, emphasis, and confidence. It is possible that there is another component of movement involved in the interpretation of these qualities: the rate of movement. Slow changes in the height of a gesture may mean something different than rapid changes. One can imagine that a rapid descent of a finger tip to hover just over the table might be interpreted with the same emphasis (or more) than a slow tapping gesture on the surface. Study Two did not examine this component of movement because it was not identified in either the work described in Chapter 3 or the subsequent design sessions.

The rapidity of change in height and in lateral movement is represented in the current embodiment design (as the shape, size, and transparency of the above-the-surface representation change over time), but this quality of the gesture is not emphasized. Chapter 4 discussed the reasons why such emphasis may be important. Future work should examine whether the rate of movement (in any direction) conveys qualities that should be expressed more explicitely in embodiments.

Understanding and Emphasizing Differences in Height Variation

During the evaluation of the effectiveness of height visualizations, significant differences were found in the way that gestures are interpreted depending on the presence or absence of local height variations. These findings suggest that users may benefit from the enhancement of local variations over global height information. Although local variations are discernable in the existing embodiment, they are not obvious and may, at times, be obfuscated by the temporal traces component of the above-the-surface visualization. A visual effect that emphasizes smaller

variations in height during a gesture may provide remote collaborators with a clearer image of the gesture in progress. Future work should examine whether this is the case or whether a visualization more sensitive to small variations would cause new problems for users.

Patterns of height variation may be associated with certain kinds of non-deictic gesturing or deixis that identifies specific classes of objects (e.g., perhaps points, as opposed to lines or areas, are always associated with lower heights of gestures). A closer examination of how height relates to particular classes of gestures needs to be performed in order to answer these questions and may ultimately help with automated gesture differentiation (discussed further, below).

Multiple Embodiments on a Surface

Introducing embodiments that have the potential to take up a larger virtual space than the hand and arm do in real space could create problems when multiple users are interacting in the same workspace. In the studies described above, only one user was embodied. It is possible that multiple embodiments could interfere with communication or obstruct a user interface. Although the seriousness of this potential problem is unknown, there are a few design considerations that may help resolve this potential problem.

Gestures in motion and gestures in contact with the surface were interpreted as having much higher levels of emphasis, confidence, and specificity. Thus, when a user remains still and above the surface, the embodiment might be designed to become less noticable – perhaps by increasing the transparency, regardless of the global height of the gesture.

Gestures, especially deictic gestures, are almost always accompanied by speech. Workspaces could be rendered less cluttered by embodying only speaking users (or users who have very recently spoken). De-emphasizing unmoving hand and arm positions and eliminating the embodiments of non-speakers would in turn emphasize those movements that are perhaps more critical to collaboration, attracting attention more easily to important gestures over and on

the workspace. This increased attentiveness might mediate the requirement for multiple visual encodings, allowing additional information to be encoded in the embodiment and creating a richer set of possibilities for communication.

Embodying Deixis and non-Deixis Differently

Deictic and non-deictic gestures have different communicative roles and groupware should be able to support those roles in different ways. For instance, during conversational gesticulation, users may not want to have their embodiments leaving traces or emphasizing touches. Unfortunately, the goal of distributed groupware that accurately and consistently differentiates between different kinds of gestures is currently unobtainable. However, there are several short-term goals that are logical next steps from this research.

1. Perform a similar set of observations for deixis with artifacts other than maps, maps *in situ*, and a more extensive set of observations for vertical (e.g., wall-mounted) surfaces. There is an imperfect understanding of how these factors affect the way that deixis manifests in collocated collaboration.
2. Use natural language comprehension to identify deixis. Although non-deictic gestures may be difficult to interpret without sophisticated natural language comprehension, the linguistic components of deixis are usually more clear and limited (e.g., including words such as “this” or “there”). It may be possible to leverage this simpler set of language patterns to more clearly identify when a deictic gesture is occurring during a collaboration and to respond appropriately with embodiment enhancements.
3. Use simple heuristics to identify deictic events and differentiate them from other kinds of gesturing. There may be a set of simple heuristics (e.g., the speaker and at least one other collaborator are facing the surface) that help identify when deixis is occurring. It is possible that deixis occurs in groupings, where several deictic gestures are made before non-deictic

gesturing begins to occur. A greater understanding of the environmental conditions and collaborative patterns that are present during deixis may help improve deixis differentiation.

A Test Bed for Distributed Collaboration

KinectArms provides the opportunity to cheaply add user embodiments to distributed groupware applications. Several Canadian university labs are currently using KinectArms in this respect. Because the hardware requirements are low, these labs can cheaply and easily integrate KinectArms into their custom groupware applications, share it between labs, and test in a true distributed setting.

These advantages allow the examination (or re-examination) of several research problems in distributed settings. The following are two examples of such research problems:

1. How robust is surface-based groupware to real-world network challenges? Much groupware is, by necessity, tested in one location over local networks. However, many of the settings envisioned for the deployment of groupware are less constrained, involving international distances. Similar systems distributed among multiple institutions in Canada will allow researchers to test their applications in a more realistic environment.
2. How effective are techniques for managing disconnections during real-world collaborations (e.g., [119])? The presence of several distributed tabletops will encourage the use of surfaces in real collaborations among connected labs. For the first time, many theorized problems related to constant use of distributed groupware, such as disconnections and reconnections during collaborative sessions, will arise in real-world settings. Such events will allow researchers to evaluate the success of their solutions to these problems more effectively.

Extending the KinectArms Project

KinectTable can be extended in several ways to improve its usability for developers, extend the number and variety of visualizations, and provide opportunities for using it to detect user input.

Currently, KinectArms provides no support for calibrating between multiple tables, and the examples in Chapter 6 (see Figure 51) were achieved by aligning two systems by setting both Kinect sensors at fixed heights above the (equal-size) surfaces. Based on this experience, the next step in improving KinectTable will be to add an interactive calibration tool to the library, allowing the Kinect to be hung at different heights above different tables.

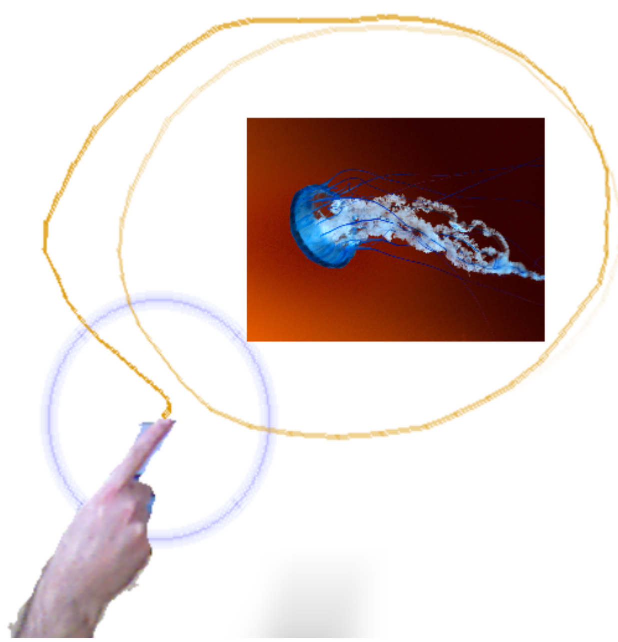


Figure 51: Remote user annotating shared image in the photo-sharing application.

The API for accessing KinectArms features is, at present, difficult to use. A set of wrappers around existing features would improve KinectArms' usability as a toolkit for groupware developers.

There is space for designing and implementing an almost unlimited number of visualizations for user embodiments. KinectViz provides a sampling of visualizations, but further development could provide a large set of visualizations suitable for a wide variety of tasks. One addition of value to the research proposed earlier in this section would be long-term history visualizations. Additive heat maps or other such visualizations of gestures over collaborative sessions would provide an easy tool for exploring the value of such techniques.

Finally, KinectTable currently detects only fingertips, the spaces between fingers, palm centres, the intersection of the arm and table edge, and the arm centroid. Further articulating the hand images with knuckle detection, wrist identification, and perhaps other body parts visible to the camera (such as the head and shoulders) would provide several clear benefits. First, a greater awareness of users' locations around the surface would improve user tracking even when they are not visible in the workspace. Second, a more articulated model of the hand and arm would set the stage for gesture recognition, allowing the exploration of how gesture recognition might integrate with above-the-surface communication. Finally, many embodiment techniques for supporting deixis rely on identifying the primary pointing finger. In the work in Chapter 6, this was done by using the lowest finger, however, there are circumstances where this method fails. A more robust arm and hand model could improve the recognition of the engaged portions of a hand during deixis, and thus improve embodiment support for deixis.

Evaluation in the Wild: The Sahtu Hunter Tour

This project was inspired by ongoing work in the Canadian arctic by biologists, veterinarians, and ecologists. If KinectTable is combined with a short-throw projector, it can be hung over tables in a variety of northern settings to provide a digital map surface over which gestures can be tracked. This combined technology has the potential to solve the problem that

was the original inspiration for this project: How can the gestures of interviewees be tracked during TEK elicitation in the Canadian arctic?

For future research, and as a clear next step, I would like to deploy this system during the Sahtu Hunter tour (or a similar TEK extraction research trip) and evaluate its success, both in terms of how well researchers are able to use the gestures to extract GIS data, and how easily it is used in the context of TEK extraction.

REFERENCES

[1] J. Hindmarsh and C. Heath, "Embodied reference: A study of deixis in workplace interaction* 1," *Journal of Pragmatics*, vol. 32, no. 12, pp. 1855–1878, 2000.

[2] D. McNeill, *Hand and mind*. University of Chicago Press, 1992.

[3] C. Gutwin and S. Greenberg, "The mechanics of collaboration: Developing low cost usability evaluation methods for shared workspaces," in *Proceedings of the 9th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises*, 2000, pp. 98–103.

[4] M. M. Bekker, J. S. Olson, and G. M. Olson, "Analysis of gestures in face-to-face design teams provides guidance for how to use groupware in design," in *Proceedings of the 1st conference on Designing interactive systems: processes, practices, methods, & techniques*, 1995, p. 166.

[5] D. Kirk, T. Rodden, and D. S. Fraser, "Turn it this way: grounding collaborative action with remote gestures," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2007, p. 1048.

[6] I. Rauschert, P. Agrawal, R. Sharma, S. Fuhrmann, I. Brewer, and A. MacEachren, "Designing a human-centered, multimodal GIS interface to support emergency management," in *Proceedings of the 10th ACM international symposium on Advances in geographic information systems*, 2002, p. 124.

[7] R. Johnson, "Negotiating the Dayton Peace Accords through digital maps," in *annual meeting of the International Studies Association, Washington, DC. Retrieved September*, 1999, vol. 11, p. 2000.

[8] S. Kettebekov and R. Sharma, "Understanding gestures in multimodal human computer interaction," *International Journal on Artificial Intelligence Tools*, vol. 9, no. 2, pp. 205–223, 2000.

[9] M. Cavazza, F. Charles, S. J. Mead, O. Martin, X. Marichal, and A. Nandi, "Multimodal acting in mixed reality interactive storytelling," *Multimedia, IEEE*, vol. 11, no. 3, pp. 30–39, 2004.

[10] C. Heath and P. Luff, "Collaborative activity and technological design: Task coordination in London Underground control rooms," in *Proceedings of the second conference on European Conference on Computer-Supported Cooperative Work*, 1991, pp. 65–80.

[11] D. Kirk, A. Crabtree, and T. Rodden, "Ways of the hands," in *ECSCW 2005*, 2005, pp. 1–21.

[12] R. K. Brook, S. J. Kutz, A. M. Veitch, R. A. Popko, B. T. Elkin, and G. Guthrie, "Fostering community-based wildlife health monitoring and research in the Canadian North," *EcoHealth*, vol. 6, no. 2, pp. 266–278, 2009.

[13] C. Gutwin and R. Penner, "Improving interpretation of remote gestures with telepointer traces," in *Proceedings of the 2002 ACM conference on Computer supported cooperative work*, 2002, p. 57.

[14] A. Tang, C. Neustaedter, and S. Greenberg, "Videoarms: embodiments for mixed presence groupware," *PEOPLE AND COMPUTERS*, vol. 20, p. 85, 2007.

[15] S. Izadi, A. Agarwal, A. Criminisi, J. Winn, A. Blake, and A. Fitzgibbon, "C-Slate: exploring remote collaboration on horizontal multi-touch surfaces," *IEEE Tabletop 2007*, pp. 3–10.

[16] J. C. Tang and S. Minneman, "VideoWhiteboard: video shadows to support remote collaboration," in *Proceedings of the SIGCHI conference on Human factors in computing systems: Reaching through technology*, 1991, pp. 315–322.

[17] A. Tang, M. Pahud, K. Inkpen, H. Benko, J. C. Tang, and B. Buxton, "Three's company: understanding communication channels in three-way distributed collaboration," in *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, 2010, pp. 271–280.

[18] M. J. Geller, "Mapping the universe," *Bulletin of the American Academy of Arts and Sciences*, vol. 45, no. 2, pp. 41–54, 1991.

[19] J. Pickles, *A history of spaces: cartographic reason, mapping, and the geo-coded world*. Routledge, 2004.

[20] J. L. Morrison, "The State of Government Cartography in 1998.," *Cartography and Geographic Information Science*, vol. 26, no. 3, 1999.

[21] A. M. MacEachren, "Cartography and GIS: facilitating collaboration," *Progress in Human Geography*, vol. 24, no. 3, pp. 445–456, 2000.

[22] A. M. MacEachren, "Moving geovisualization toward support for group work," *Exploring Geovisualization*, vol. 461, 2005.

[23] S. Benford, J. Bowers, L. E. Fahlén, C. Greenhalgh, and D. Snowdon, "User embodiment in collaborative virtual environments," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 1995, pp. 242–249.

[24] D. Pinelle, M. Nacenta, C. Gutwin, and T. Stach, "The effects of co-present embodiments on awareness and collaboration in tabletop groupware," in *Proceedings of graphics interface 2008*, 2008, pp. 1–8.

[25] A. MacEachren, G. Cai, I. Brewer, and J. Chen, "Supporting Map-Based geocollaboration through Natural Interfaces to Large-Screen Displays," *Cartographic Perspectives*, vol. 54, pp. 4–22, 2006.

[26] R. E. Kraut, S. R. Fussell, and J. Siegel, "Visual information as a conversational resource in collaborative physical tasks," *Human-computer interaction*, vol. 18, no. 1, pp. 13–49, 2003.
[27] P. Dietz and D. Leigh, "DiamondTouch: a multi-user touch technology," in *Proceedings of the 14th annual ACM symposium on User interface software and technology*, 2001, pp. 219–226.
[28] L. J. Bannon and K. Schmidt, "CSCW: Four characters in search of a context," 1991.
[29] D. J. Wood and B. Gray, "Toward a comprehensive theory of collaboration," *The Journal of Applied Behavioral Science*, vol. 27, no. 2, p. 139, 1991.
[30] J. Hughes, V. King, T. Rodden, and H. Andersen, "Moving out from the control room: ethnography in system design," in *Proceedings of the 1994 ACM conference on Computer supported cooperative work*, 1994, p. 439.
[31] P. Dillenbourg, "What do you mean by collaborative learning," *Collaborative learning: Cognitive and computational approaches*, pp. 1–19, 1999.
[32] G. M. Olson, J. S. Olson, M. R. Carter, and M. Storr\osten, "Small group design meetings: An analysis of collaboration," *Human-Computer Interaction*, vol. 7, no. 4, pp. 347–374, 1992.
[33] P. Tuddenham and P. Robinson, "Territorial coordination and workspace awareness in remote tabletop collaboration," in *Proceedings of the 27th international conference on Human factors in computing systems*, 2009, pp. 2139–2148.
[34] M. Fraser, M. R. McCarthy, M. Shaukat, and P. Smith, "Seconds Matter: Improving Distributed Coordination by Tracking and Visualizing Display Trajectories," 2007.
[35] S. R. Fussell, L. D. Setlock, J. Yang, J. Ou, E. Mauer, and A. D. . Kramer, "Gestures over video streams to support remote collaboration on physical tasks," *Human-Computer Interaction*, vol. 19, no. 3, pp. 273–309, 2004.
[36] T. Stach, C. Gutwin, D. Pinelle, and P. Irani, "Improving recognition and characterization in groupware with rich embodiments," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2007, p. 20.
[37] D. Pinelle, C. Gutwin, and S. Greenberg, "Task analysis for groupware usability evaluation: Modeling shared-workspace tasks with the mechanics of collaboration," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 10, no. 4, pp. 281–311, 2003.
[38] H. H. Clark and S. E. Brennan, "Grounding in communication," *Perspectives on socially shared cognition*, vol. 13, pp. 127–149, 1991.
[39] H. H. Clark and C. R. Marshall, "Definite reference and mutual knowledge," *Psycholinguistics: critical concepts in psychology*, p. 414, 2002.
[40] J. Ou, S. R. Fussell, X. Chen, L. D. Setlock, and J. Yang, "Gestural communication over video stream: supporting multimodal interaction for remote collaborative physical tasks," in *Proceedings of the 5th international conference on Multimodal interfaces*, 2003, pp. 242–249.
[41] C. Gutwin and S. Greenberg, "A descriptive framework of workspace awareness for real-time groupware," *Computer Supported Cooperative Work (CSCW)*, vol. 11, no. 3, pp. 411–446, 2002.
[42] T. Rodden, Y. Rogers, J. Halloran, and I. Taylor, "Designing novel interactional workspaces to support face to face consultations," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2003, p. 64.
[43] A. M. MacEachren and I. Brewer, "Developing a conceptual framework for visually-enabled geocollaboration," *International Journal of Geographical Information Science*, vol. 18, no. 1, pp. 1–34, 2004.
[44] I. Brewer, A. M. MacEachren, H. Abdo, J. Gundrum, and G. Otto, "Collaborative Geographic Visualization: Enabling shared understanding of environmental processes," in *IEEE Information Visualization Symposium*, 2000, pp. 137–141.
[45] M. J. Shiffer, "Augmenting geographic information with collaborative multimedia technologies," in *AUTOCARTO-CONFERENCE-*, 1993, pp. 367–367.
[46] D. G. Tatar, G. Foster, and D. G. Bobrow, "Design for conversation: lessons from Cognoter," *International Journal of Man-Machine Studies*, vol. 34, no. 2, pp. 185–209, 1991.
[47] J. Ou, X. Chen, S. R. Fussell, and J. Yang, "DOVE: Drawing over video environment," in *Proceedings of the eleventh ACM international conference on Multimedia*, 2003, pp. 100–101.
[48] W. J. Craig and S. A. Elwood, "How and why community groups use maps and geographic information," *Cartography and geographic information systems*, vol. 25, no. 2, pp. 95–104, 1998.
[49] K. L. Ma, "Image graphs—a novel approach to visual data exploration," in *Proceedings of the conference on Visualization'99: celebrating ten years*, 1999, p. 88.

[50] B. Congleton, J. Cerretani, M. W. Newman, and M. S. Ackerman, “Sharing Map Annotations in Small Groups: X Marks the Spot,” in *Proceedings of the 12th IFIP TC 13 International Conference on Human-Computer Interaction: Part II*, 2009, p. 110.
[51] C. Rinner, “Argumentation maps: GIS-based discussion support for on-line planning,” *Environment and Planning B: Planning and Design*, vol. 28, no. 6, pp. 847–863, 2001.
[52] K. Baker, S. Greenberg, and C. Gutwin, “Heuristic evaluation of groupware based on the mechanics of collaboration,” *Lecture Notes in Computer Science*, pp. 123–140, 2001.
[53] M. Garau, M. Slater, S. Bee, and M. A. Sasse, “The impact of eye gaze on communication using humanoid avatars,” in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2001, p. 316.
[54] T. Erickson and W. A. Kellogg, “Social translucence: an approach to designing systems that support social processes,” *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 7, no. 1, pp. 59–83, 2000.
[55] W. C. Hill, J. D. Hollan, D. Wroblewski, and T. McCandless, “Edit wear and read wear,” in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 1992, p. 9.
[56] J. C. Tang and S. L. Minneman, “VideoDraw: a video interface for collaborative drawing,” *ACM Transactions on Information Systems (TOIS)*, vol. 9, no. 2, pp. 170–184, 1991.
[57] N. Stephenson, *Snow crash*. Spectra, 2000.
[58] M. Gerhard, D. Moore, and D. Hobbs, “Embodiment and copresence in collaborative interfaces,” *International journal of human-computer studies*, vol. 61, no. 4, pp. 453–480, 2004.
[59] M. Mori, “Bukimi no tani [the uncanny valley],” *Energy*, vol. 7, no. 4, pp. 33–35, 1970.
[60] J. Seyama and R. S. Nagayama, “The uncanny valley: Effect of realism on the impression of artificial human faces,” *Presence: Teleoperators and Virtual Environments*, vol. 16, no. 4, pp. 337–351, 2007.
[61] C. Bartneck, T. Kanda, H. Ishiguro, and N. Hagita, “Is the uncanny valley an uncanny cliff?,” in *The 16th IEEE International Symposium on Robot and Human interactive Communication, 2007. RO-MAN 2007*, 2007, pp. 368–373.
[62] D. Hanson, A. Olney, S. Prilliman, E. Mathews, M. Zielke, D. Hammons, R. Fernandez, and H. Stephanou, “Upending the uncanny valley,” in *PROCEEDINGS OF THE NATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE*, 2005, vol. 20, p. 1728.
[63] F. Biocca, “The Cyborg’s dilemma1:: Progressive embodiment in virtual environments,” *Human Factors in Information Technology*, vol. 13, pp. 113–144, 1999.
[64] J. Lee, J. Chai, P. S. . Reitsma, J. K. Hodgins, and N. S. Pollard, “Interactive control of avatars animated with human motion data,” *ACM Transactions on Graphics*, vol. 21, no. 3, pp. 491–500, 2002.
[65] N. Ferrand, “Modelling and supporting multi-actor spatial planning using multi-agents systems,” in *Proceedings of the Third NCGIA Conference on Integrating GIS and Environmental Modelling, Santa Fe, New Mexico, USA, January*, 1996, pp. 21–25.
[66] J. Dyck, C. Gutwin, S. Subramanian, and C. Fedak, “High-performance telepointers,” in *Proceedings of the 2004 ACM conference on Computer supported cooperative work*, 2004, pp. 172–181.
[67] C. Gutwin, J. Dyck, and J. Burkitt, “Using cursor prediction to smooth telepointer jitter,” in *Proceedings of the 2003 international ACM SIGGROUP conference on Supporting group work*, 2003, p. 301.
[68] N. Osawa, “Aggregate pointers to support large group collaboration using telepointers,” in *CHI’06 extended abstracts on Human factors in computing systems*, 2006, p. 1174.
[69] S. Greenberg, C. Gutwin, and M. Roseman, “Semantic telepointers for groupware,” in *Proceedings of the 6th Australian Conference on Computer-Human Interaction (OZCHI’96)*, 1996, p. 54.
[70] J. Adams, B. Rogers, S. Hayne, G. Mark, J. Nash, and L. Leifer, “The effect of a telepointer on student performance and preference,” *Computers & Education*, vol. 44, no. 1, pp. 35–51, 2005.
[71] M. A. Nacenta, D. Pinelle, D. Stuckel, and C. Gutwin, “The effects of interaction technique on coordination in tabletop groupware,” in *Proceedings of Graphics interface 2007*, 2007, p. 198.
[72] D. Pinelle, M. Barjawi, M. Nacenta, and R. Mandryk, “An evaluation of coordination techniques for protecting objects and territories in tabletop groupware,” in *Proceedings of the 27th international conference on Human factors in computing systems*, Boston, MA, USA, 2009, pp. 2129–2138.
[73] D. Pinelle, M. Barjawi, M. Nacenta, and R. Mandryk, “An evaluation of coordination techniques for protecting objects and territories in tabletop groupware,” in *Proceedings of the 27th international conference on Human factors in computing systems*, 2009, pp. 2129–2138.
[74] M. Apperley, L. McLeod, M. Masoodian, L. Paine, M. Phillips, B. Rogers, and K. Thomson, “Use of video shadow for small group interaction awareness on a large interactive display surface,” in *Proceedings of the Fourth Australasian user interface conference on User interfaces 2003-Volume 18*, 2003, pp. 81–90.

[75] J. Li, A. Wessels, L. Alem, and C. Stitzlein, “Exploring interface with representation of gesture for remote collaboration,” in *Proceedings of the 19th Australasian conference on Computer-Human Interaction: Entertaining User Interfaces*, 2007, p. 182.
[76] D. Stotts, J. M. . Smith, and K. Gyllstrom, “Facespace: endo-and exo-spatial hypermedia in the transparent video facetop,” in *Proceedings of the fifteenth ACM conference on Hypertext and hypermedia*, 2004, pp. 48–57.
[77] D. Stotts, J. M. . Smith, and K. Gyllstrom, “Support for distributed pair programming in the transparent video facetop,” *Extreme Programming and Agile Methods-XP/Agile Universe 2004*, pp. 150–192, 2004.
[78] H. Ishii and M. Kobayashi, “ClearBoard: a seamless medium for shared drawing and conversation with eye contact,” in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 1992, pp. 525–532.
[79] E. S. Veinott, J. Olson, G. M. Olson, and X. Fu, “Video helps remote work: Speakers who need to negotiate common ground benefit from seeing each other,” in *Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit*, 1999, p. 309.
[80] P. A. Chalmers, “The role of cognitive theory in human-computer interface,” *Computers in Human Behavior*, vol. 19, no. 5, pp. 593–607, 2003.
[81] T. Salvador, J. Scholtz, and J. Larson, “The Denver model for groupware design,” *ACM SIGCHI Bulletin*, vol. 28, no. 1, pp. 52–58, 1996.
[82] T. Erickson and W. A. Kellogg, “Social translucence: Using minimalist visualizations of social activity to support collective interaction,” *Designing information spaces: The social navigation approach*, pp. 17–42, 2003.
[83] T. Erickson, D. N. Smith, W. A. Kellogg, M. Laff, J. T. Richards, and E. Bradner, “Socially translucent systems: social proxies, persistent conversation, and the design of ‘babble’,” in *Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit*, 1999, p. 79.
[84] T. Erickson, C. Halverson, W. A. Kellogg, M. Laff, and T. Wolf, “Social translucence: designing social infrastructures that make collective activity visible,” *Communications of the ACM*, vol. 45, no. 4, p. 44, 2002.
[85] J. P. Birnholtz, C. Gutwin, G. Ramos, and M. Watson, “OpenMessenger: gradual initiation of interaction for distributed workgroups,” in *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, 2008, pp. 1661–1664.
[86] G. McEwan and S. Greenberg, “Supporting social worlds with the community bar,” in *Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work*, 2005, pp. 21–30.
[87] M. R. Gibbs, K. Hew, and G. Wadley, “Social translucence of the Xbox live voice channel,” *Entertainment Computing–ICEC 2004*, pp. 407–499, 2004.
[88] J. Hindmarsh and C. Heath, “Sharing the tools of the trade: The interactional constitution of workplace objects,” *Journal of Contemporary Ethnography*, vol. 29, no. 5, p. 523, 2000.
[89] P. Bao and D. Gergle, “What’s this you say?: the use of local references on distant displays,” in *Proceedings of the 27th international conference on Human factors in computing systems*, 2009, pp. 1029–1032.
[90] O. Hilliges, S. Izadi, A. D. Wilson, S. Hodges, A. Garcia-Mendoza, and A. Butz, “Interactions in the air,” in *Proceedings of the 22nd annual ACM symposium on User interface software and technology - UIST ’09*, Victoria, BC, Canada, 2009, p. 139.
[91] G. Shoemaker, A. Tang, and K. S. Booth, “Shadow reaching: a new perspective on interaction for large displays,” in *Proceedings of the 20th annual ACM symposium on User interface software and technology*, 2007, p. 56.
[92] C. Gutwin, “Traces: Visualizing the immediate past to support group interaction,” *Graphics Interface 2002: proceedings: Calgary, Alberta, 27-29 May, 2002*, p. 43, 2002.
[93] N. Yamashita, K. Kaji, H. Kuzuoka, and K. Hirata, “Improving visibility of remote gestures in distributed tabletop collaboration,” in *Proceedings of the ACM 2011 conference on Computer supported cooperative work*, New York, NY, USA, 2011, pp. 95–104.
[94] C. Harrison, H. Benko, and A. D. Wilson, “OmniTouch: wearable multitouch interaction everywhere,” in *Proceedings of the 24th annual ACM symposium on User interface software and technology*, 2011, pp. 441–450.
[95] A. Genest and C. Gutwin, “Characterizing Deixis over Surfaces to Improve Remote Embodiments,” in *ECSCW 2011: Proceedings of the 12th European Conference on Computer Supported Cooperative Work, 24-28 September 2011, Aarhus Denmark*, 2011, pp. 253–272.

[96] C. Gutwin and S. Greenberg, “Effects of awareness support on groupware usability,” in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 1998, p. 518.
[97] R. M. Krauss and S. R. Fussell, “Mutual knowledge and communicative effectiveness.,” 1990.
[98] A. Genest and C. Gutwin, “Evaluating the effectiveness of height visualizations for improving gestural communication at distributed tabletops,” in *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, 2012, pp. 519–528.
[99] S. Hayne, M. Pendergast, and S. Greenberg, “Implementing gesturing with cursors in group support systems,” *Journal of Management Information Systems*, vol. 10, no. 3, pp. 43–61, 1993.
[100] J. Bertin, *Semiology of graphics*. University of Wisconsin Press, 1983.
[101] J. Mackinlay, “Automating the design of graphical presentations of relational information,” *ACM Transactions on Graphics (TOG)*, vol. 5, no. 2, p. 141, 1986.
[102] M. P. Kwan, “Gender, the Home-Work Link, and Space-Time Patterns of Nonemployment Activities*,” *Economic geography*, vol. 75, no. 4, pp. 370–394, 1999.
[103] M. Weber, M. Alexa, and W. Müller, “Visualizing time-series on spirals,” in *proceedings of the IEEE Symposium on Information Visualization*, 2001, p. 7.
[104] C. Tominski, J. Abello, and H. Schumann, “Axes-based visualizations with radial layouts,” in *Proceedings of the 2004 ACM symposium on Applied computing*, 2004, pp. 1242–1247.
[105] J. J. Van Wijk and E. R. Van Selow, “Cluster and calendar based visualization of time series data,” in *Information Visualization, 1999.(Info Vis’ 99) Proceedings. 1999 IEEE Symposium on*, 1999, pp. 4–9.
[106] S. Havre, B. Hetzler, and L. Nowell, “ThemeRiver: Visualizing theme changes over time,” in *Information Visualization, 2000. InfoVis 2000. IEEE Symposium on*, 2000, pp. 115–123.
[107] W. Muller and H. Schumann, “Visualization methods for time-dependent data-an overview,” in *Simulation Conference, 2003. Proceedings of the 2003 Winter*, 2003, vol. 1, pp. 737–745.
[108] M. Svensson, C. Heath, and P. Luff, “Instrumental action: the timely exchange of implements during surgical operations,” *ECSCW 2007*, pp. 41–60, 2007.
[109] G. A. Miller, “The magical number seven, plus or minus two: some limits on our capacity for processing information.,” *Psychological review*, vol. 63, no. 2, p. 81, 1956.
[110] B. Jordan and A. Henderson, “Interaction analysis: Foundations and practice,” *The journal of the learning sciences*, vol. 4, no. 1, pp. 39–103, 1995.
[111] A. D. Wilson, “Using a depth camera as a touch sensor,” in *ACM international conference on interactive tabletops and surfaces*, 2010, pp. 69–72.
[112] J. Segen and S. Kumar, “Human-computer interaction using gesture recognition and 3D hand tracking,” in *Image Processing, 1998. ICIP 98. Proceedings. 1998 International Conference on*, 1998, pp. 188–192.
[113] H. Ishii, “TeamWorkStation: towards a seamless shared workspace,” in *Proceedings of the 1990 ACM conference on Computer-supported cooperative work*, 1990, pp. 13–26.
[114] A. M. MacEachren, *How maps work: representation, visualization, and design*. The Guilford Press, 2004.
[115] W. Seager and D. S. Fraser, “Comparing physical, automatic and manual map rotation for pedestrian navigation,” in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2007, p. 776.
[116] R. Kruger, S. Carpendale, S. D. Scott, and S. Greenberg, “Roles of orientation in tabletop collaboration: Comprehension, coordination and communication,” *Computer Supported Cooperative Work (CSCW)*, vol. 13, no. 5, pp. 501–537, 2004.
[117] D. Wigdor and R. Balakrishnan, “Empirical investigation into the effect of orientation on text readability in tabletop displays,” in *ECSCW 2005*, 2005, pp. 205–224.
[118] T. Ballendat, N. Marquardt, and S. Greenberg, “Proxemic interaction: designing for a proximity and orientation-aware environment,” in *ACM International Conference on Interactive Tabletops and Surfaces*, New York, NY, USA, 2010, pp. 121–130.
[119] B. Roy, N. Graham, and C. Gutwin, “DiscoTech: a plug-in toolkit to improve handling of disconnection and reconnection in real-time groupware,” in *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, 2012, pp. 1287–1296.
[120] A. M. MacEachren, “Cartography and GIS: extending collaborative tools to support virtual teams,” *Progress in Human Geography*, vol. 25, no. 3, pp. 431–444, 2001.
[121] T. Nyerges, P. Jankowski, D. Tuthill, and K. Ramsey, “Collaborative water resource decision support: Results of a field experiment,” *Annals of the Association of American Geographers*, vol. 96, no. 4, pp. 699–725, 2006.

[122] B. Faber, W. Wallace, and J. Cuthbertson, "Advances in collaborative GIS for land resource negotiation," in *Proceedings GIS'95*, 1995, pp. 183–189.
[123] R. Reitsma, I. Zigurs, C. Lewis, V. Wilson, and A. Sloane, "Water-Resources Negotiations," *Journal of water resources planning and management*, 1996.
[124] I. Casas, R. Garrity, D. Mandloi, M. Sunm, J. Weaver, R. Nagi, and R. Batta, "A Spatial Decision Support System Combining GIS and OR Tools to Optimize District Boundaries and Bus Routes for a Suburban School District," *OR Insight*, vol. 21, no. 2, pp. 3–16, 2008.
[125] K. Brodlie, "Models of Collaborative Visualization," in *Exploring Geovisualization*, J. Dykes, A. M. MacEachren, and M. J. Kraak, Eds. Elsevier, 2005, p. 463.
[126] P. Antunes, G. Zurita, and N. Baloian, "A Model for Designing Geocollaborative Artifacts and Applications," in *Groupware: Design, Implementation, and Use: 15th International Workshop, Peso Da Régua, Douro, Portugal, September 13-17, 2009, Proceedings*, 2009, p. 278.
[127] J. Wood, H. Wright, and K. Brodlie, "Collaborative visualization," in *Proceedings of the 8th conference on Visualization'97*, 1997.
[128] M. P. Armstrong and P. Densham, "Cartographic support for collaborative spatial decision-making," in *AUTOCARTO-CONFERENCE-*, 1995, pp. 49–58.
[129] G. Mark, K. Carpenter, and A. Kobsa, "A model of synchronous collaborative information visualization," in *To appear in Proceedings of IV03 7th International Conference on Information Visualization. July*, 2003, vol. 16, p. 18.
[130] R. Johansen, *Groupware: Computer support for business teams*. The Free Press New York, NY, USA, 1988.
[131] D. Wigdor, S. Williams, M. Cronin, R. Levy, K. White, M. Mazeev, and H. Benko, "Ripples: utilizing per-contact visualizations to improve user interaction with touch displays," in *Proceedings of the 22nd annual ACM symposium on User interface software and technology*, 2009, pp. 3–12.
[132] J. Bertin and others, *Sémiologie graphique*. Mouton Paris, 1973.

APPENDIX A:
GEOCOLLABORATION AND COLLABORATION

Geocollaboration was the inspiration for this dissertation, but did not figure significantly in the research. However, there is value in understanding how research on geocollaboration has addressed supporting collaboration.

Geographical Information Systems to support digital and dynamic map use were created for individuals rather than collaborative groups [120] and remain complex applications requiring extensive training or expert “chauffeurs” [121], despite efforts to improve accessibility for non-expert users [122]. There are dynamic or digital map applications designed for naive users (such as Google Earth, MapQuest, and Google Maps), which can be used for common map-based tasks such as wayfinding. However, large and complex tasks (e.g., resource management [123], urban planning [124], or emergency response [6]), commonly involve the collaborative use of GIS [21]. As a result, there are few collaborative GIS tools, despite the importance of collaborative, map-based work [125].

The field of Computer Supported Cooperative Work focuses on several issues related to collaborative work, including articulating the process of cooperative work, understanding how information spaces are shared, and (within the context of the sub-field of groupware) adapting single-user systems to support collaboration [28]. Thus, CSCW research is applicable to the design and development of geocollaborative tools. In particular, CSCW research has identified specific mechanics of collaboration that can be used to evaluate collaborative systems [3], such as explicit communication, consequential communication, and protection of spaces or resources.

Research into geocollaboration has attempted to describe the collaborative aspects of geocollaboration (e.g., [43][126]) by creating frameworks for application development and research. However, geocollaborative frameworks are more domain-focused than CSCW frameworks because they are created for describing collaboration with respect to geography. They also tend to be more more abstract than the mechanics identified in CSCW. For this reason, CSCW frameworks are of greater use when approaching the design of remote embodiments to support gestural communication in distributed geocollaborations.

Geocollaboration is not new: people have been collaborating using maps for as long as there have been maps [43]. However, geocollaboration with geovisualizations is performed using computer software as an intermediary for both map-based interactions and, sometimes, interpersonal interactions. Computer mediation increases the opportunities for novel methods of collaboration and map interaction, but it can also create barriers to communication, both with the map and with collaborators.

The field of geocollaboration is relatively new and researchers have recognized that the process of merging geovisualization and collaboration is far from complete. Where GIS supports geocollaboration, collaborative tools have usually been added on to existing GIS with little regard to a cohesive design or state of the art collaborative research [21], [44], [120], [125], [127], [128]. In fact, GIS developed primarily as single-user, desktop-based systems have been described as impeding collaboration [25]. This problem is compounded in the many GIS that are complex and difficult to use [129].

The historical origin of GIS as single-user applications is not the only barrier to geocollaboration. Another is the wide variety of collaboration possible: collaboration can occur between peers or stakeholders with similar levels of knowledge [31] or collaboration can be

instructional, involving a worker and a helper, where the levels of knowledge are necessarily different in the field of instruction [35]. Additionally, collaboration can be geographically collocated or remote as well as either asynchronous or synchronous [130]. Accommodating the requirements of any of these kinds of collaboration is a significant design challenge.

Designing a collaborative system, whether geocollaborative or not, requires a useful model of collaboration. In particular, an understanding of the individual mechanics of the collaborative process can provide designers with specific targets for creating user embodiments for use in distributed settings.

APPENDIX B
EXPERIMENT CONSENT FORMS, DEMOGRAPHIC DATA FORMS, AND POST-STUDY QUESTIONNAIRES

For all of the experiments described in this dissertation, the experiments and the handling of participants was carried out according to the University of Saskatchewan's guidelines for experiments with human subjects, under approved ethics applications (investigator, Prof. Carl Gutwin).

Contents

DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF SASKATCHEWAN
INFORMED CONSENT FORM

Research Project: **Observing Deictic Gestures over Tabletop Maps**

Investigators: Dr. Carl Gutwin, Department of Computer Science (966-8646)

Aaron Genest, Department of Computer Science (966-2327)

This consent form, a copy of which has been given to you, is only part of the process of informed consent. It should give you the basic idea of what the research is about and what your participation will involve. If you would like more detail about something mentioned here, or information not included here, please ask. Please take the time to read this form carefully and to understand any accompanying information.

This study is part of a larger set of studies to determine how pointing gestures work when they are used over tabletop maps. During this study you will be asked a series of questions which you may answer using the map in front of you. We are interested in seeing how people point depending on what kind of questions they are asked.

At the end of the session, you will be given more information about the purpose and goals of the study, and there will be time for you to ask questions about the research.

The data collected from this study will be used in articles for publication in journals and conference proceedings.

As one way of thanking you for your time, we will be pleased to make available to you a summary of the results of this study once they have been compiled (usually within two months). This summary will outline the research and discuss our findings and recommendations. If you would like to receive a copy of this summary, please write down your email address here.

Contact email
address:____

All personal and identifying data will be kept confidential. If explicit consent has been given, textual excerpts, photographs, or video recordings may be used in the dissemination of research results in scholarly journals or at scholarly conferences. Anonymity will be preserved by using pseudonyms in any presentation of textual data in journals or at conferences. The informed consent form and all research data will be kept in a secure location under confidentiality in accordance with University policy for 5 years post publication. Do you have any questions about this aspect of the study?

You are free to withdraw from the study at any time without penalty and without losing any advertised benefits. Withdrawal from the study will not affect your academic status or your access to services at the university. If you withdraw, your data will be deleted from the study and destroyed.

Your continued participation should be as informed as your initial consent, so you should feel free to ask for clarification or new information throughout your participation. If you have further questions concerning matters related to this research, please contact:

- Dr. Carl Gutwin, Assistant Professor, Dept. of Computer Science, (306) 966-8646, gutwin@cs.usask.ca

Your signature on this form indicates that you have understood to your satisfaction the information regarding participation in the research project and agree to participate as a participant. In no way does this waive your legal rights nor release the investigators, sponsors, or involved institutions from their legal and professional responsibilities. If you have further questions about this study or your rights as a participant, please contact:

- Dr. Carl Gutwin, Assistant Professor, Dept. of Computer Science, (306) 966-8646, gutwin@cs.usask.ca
- Office of Research Services, University of Saskatchewan, (306) 966-4053

Participant's signature:____

Date:______________________

Investigator's signature:____

Date:______________________

A copy of this consent form has been given to you to keep for your records and reference. This research has the ethical approval of the Office of Research Services at the University of Saskatchewan.

Study Script, Researcher, Gesture Observations, Lab Study 1

You will be asked to describe various things to each other related to the map of Saskatoon in front of you. Whenever a question involves some component of geography that is off the map, please answer as well as you can with what you can see. Each of you will be asked to answer some of the questions, while only one person will be asked to answer others. The selection is random.

When you answer the questions, please answer to each other as if you need to describe the answer to the other person at the table. Feel free to use your hands and arms to illustrate your answer on the map. If you get stuck, feel free to collaborate to answer the question.

<Locator – point>
(Both) Where do you live?

<Navigation -- familiar>
(Both) What route do you use to get to work each day?

<Locator – area>
(U1) What land area is owned by the University of Saskatchewan?

(U2) What land area is considered “downtown” Saskatoon?

(Both) What are the best places to live in Saskatoon?

<Locator – point>
(U1) Where is Midtown Plaza?

(U2) Where is Market Mall?
(Alternate questions if they can’t locate the specific mall: Confederation Park Mall, Centre at Circle and 8th, etc.)

<Proximity>
(Both) Where is the shopping area closest to your house?

<Locator – point>
(Both) Where does one of your friends live?

<Navigation -- unfamiliar>
(Both) How would you get from [the other participant’s] house to your friend’s house?

(Both) How would you get from the University to [the other participant’s friend’s] house?

Participant A: Instructions

Please ask these questions of the other study participant. You will begin by asking a question, then he/she will ask a question before you continue with the next question.
Each question requires information which is partly contained on the map in front of you.
Whenever a question involves some component of geography that is off the map, please answer as well as you can with what you can see.

Feel free to use your hands and arms to illustrate your answer on the map. If you get stuck, feel free to collaborate together to answer the question.

1) Where do you live?

2) What route do you use to get to work each day?

3) What land area is considered “downtown” Saskatoon?

4) What do you consider the best places to live in Saskatoon?

5) Where is Market Mall?

6) Where is the shopping area closest to your house?

7) Where does one of your friends live?

8) How would you get from my house to your friend’s house?

9) How would you get from the University to the my friend’s house?

10) How do you get from Market Mall to Winnipeg?

11) From where you are right now, where is the nearest hotel?

Participant B: Instructions

Please ask these questions of the other study participant. You will begin by asking a question, then he/she will ask a question before you continue with the next question.
Each question requires information which is partly contained on the map in front of you. Whenever a question involves some component of geography that is off the map, please answer as well as you can with what you can see.

Feel free to use your hands and arms to illustrate your answer on the map. If you get stuck, feel free to collaborate together to answer the question.

1) Where do you live?
(Wait until the other participant has asked you question number 1 from their sheet.)

2) What route do you use to get to work each day?
(Wait until the other participant has asked you question number 2 from their sheet.)

3) What land area is owned by the University of Saskatchewan?

4) What do you consider the best places to live in Saskatoon?

5) Where is Midtown Plaza?

6) Where is the shopping area closest to your house?

7) Where does one of your friends live?

8) How would you get from my house to your friend's house?

9) How would you get from the University to my friend's house?

10) How do you get from Midtown Plaza to Edmonton?

11) From where you are right now, where is the nearest restaurant?

Open-Ended Questionnaire Used as Basis for Prototype Evaluation

For each prototype:

1. Did you like this embodiment? Why or why not?
2. What aspects of this embodiment work? Which ones do not?
3. Would you combine aspects of this embodiment with any other?

After all embodiments have been tried at least once:

1. Which embodiment prototype was your favourite?
2. Which embodiment prototype was your least favourite?
3. Which visual representation was the most useful for representing height?
4. Which one do you think would be the most intuitive for users?
5. Which one do you think will be the easiest to understand?
6. Do you have any other comments or questions?

DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF SASKATCHEWAN
INFORMED CONSENT FORM

Research Project: **Observing Deictic Gestures in the Wild**

Investigators: Dr. Carl Gutwin, Department of Computer Science (966-8646)

Aaron Genest, Department of Computer Science (966-2327)

This consent form, a copy of which has been given to you, is only part of the process of informed consent. It should give you the basic idea of what the research is about and what your participation will involve. If you would like more detail about something mentioned here, or information not included here, please ask. Please take the time to read this form carefully and to understand any accompanying information.

This study will help us understand how people gesture when they are collaboratively using maps. During this study a researcher will follow you for one or more days, observing you during your collaborations over maps. During that time the researcher may take photographs, shoot video, or ask you questions. You are not being tested – this is done to better understand how people behave outside of laboratory settings.

At the end of the session, you will be given more information about the purpose and goals of the study, and there will be time for you to ask questions about the research.

The data collected from this study will be used in articles for publication in journals and conference proceedings.

As one way of thanking you for your time, we will be pleased to make available to you a summary of the results of this study once they have been compiled (usually within two months). This summary will outline the research and discuss our findings and recommendations. If you would like to receive a copy of this summary, please write down your email address here.

Contact email address:____

All personal and identifying data will be kept confidential. If explicit consent has been given, textual excerpts, photographs, or video recordings may be used in the dissemination of research results in scholarly journals or at scholarly conferences. Anonymity will be preserved by using pseudonyms in any presentation of textual data in journals or at conferences. The informed consent form and all research data will be kept in a secure location under confidentiality in accordance with University policy for 5 years post publication. Do you have any questions about this aspect of the study?

You are free to withdraw from the study at any time without penalty and without losing any advertised benefits. Withdrawal from the study will not affect your academic status or your access to services at the university. If you withdraw, your data will be deleted from the study and destroyed.

Your continued participation should be as informed as your initial consent, so you should feel free to ask for clarification or new information throughout your participation. If you have further questions concerning matters related to this research, please contact:

- Dr. Carl Gutwin, Assistant Professor, Dept. of Computer Science, (306) 966-8646, gutwin@cs.usask.ca

Your signature on this form indicates that you have understood to your satisfaction the information regarding participation in the research project and agree to participate as a participant. In no way does this waive your legal rights nor release the investigators, sponsors, or involved institutions from their legal and professional responsibilities. If you have further questions about this study or your rights as a participant, please contact:

- Dr. Carl Gutwin, Assistant Professor, Dept. of Computer Science, (306) 966-8646, gutwin@cs.usask.ca
- Office of Research Services, University of Saskatchewan, (306) 966-4053

Participant's signature:____

Date:____________________

Investigator's signature:____

Date:____________________

A copy of this consent form has been given to you to keep for your records and reference. This research has the ethical approval of the Office of Research Services at the University of Saskatchewan.

DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF SASKATCHEWAN
INFORMED CONSENT FORM

Research Project: **Determining Effective Height Representations for the Height of Deictic Gestures**

Investigators: Dr. Carl Gutwin, Department of Computer Science (966-8646)

Aaron Genest, Department of Computer Science (966-2327)

This consent form, a copy of which has been given to you, is only part of the process of informed consent. It should give you the basic idea of what the research is about and what your participation will involve. If you would like more detail about something mentioned here, or information not included here, please ask. Please take the time to read this form carefully and to understand any accompanying information.

This study is part of a larger set of studies to determine how pointing gestures made in 3-Dimensional space can be visualized on a 2-Dimensional surface. In this study, we are examining what kinds of visualizations are most effective at indicating whether a pointing gesture touched the surface of a table. During this study you will be asked to view several pre-recorded sets of visualizations, similar to very short movies, and then answer some short questions about each visualization. You will use both the keyboard and mouse to answer these questions. Before the study starts, you will be given a short, pen and paper questionnaire with non-identifying demographic questions.

At the end of the session, you will be given more information about the purpose and goals of the study, and there will be time for you to ask questions about the research.

The data collected from this study will be used in articles for publication in journals and conference proceedings.

As one way of thanking you for your time, we will be pleased to make available to you a summary of the results of this study once they have been compiled (usually within two months). This summary will outline the research and discuss our findings and recommendations. If you would like to receive a copy of this summary, please write down your email address here.

Contact email address:____

All personal and identifying data will be kept confidential. If explicit consent has been given, textual excerpts, photographs, or video recordings may be used in the dissemination of research results in scholarly journals or at scholarly conferences. Anonymity will be preserved by using pseudonyms in any presentation of textual data in journals or at conferences. The informed consent form and all research data will be kept in a secure location under confidentiality in accordance with University policy for 5 years post publication. Do you have any questions about this aspect of the study?

You are free to withdraw from the study at any time without penalty and without losing any advertised benefits. Withdrawal from the study will not affect your academic status or your access to services at the university. If you withdraw, your data will be deleted from the study and destroyed.

Your continued participation should be as informed as your initial consent, so you should feel free to ask for clarification or new information throughout your participation. If you have further questions concerning matters related to this research, please contact:

- Dr. Carl Gutwin, Assistant Professor, Dept. of Computer Science, (306) 966-8646, gutwin@cs.usask.ca

Your signature on this form indicates that you have understood to your satisfaction the information regarding participation in the research project and agree to participate as a participant. In no way does this waive your legal rights nor release the investigators, sponsors, or involved institutions from their legal and professional responsibilities. If you have further questions about this study or your rights as a participant, please contact:

- Dr. Carl Gutwin, Assistant Professor, Dept. of Computer Science, (306) 966-8646, gutwin@cs.usask.ca
- Office of Research Services, University of Saskatchewan, (306) 966-4053

Participant's signature:____

Date:____________________

Investigator's signature:____

Date:____________________

A copy of this consent form has been given to you to keep for your records and reference. This research has the ethical approval of the Office of Research Services at the University of Saskatchewan.

Demographic Questionnaire for
Gesture Visualization Study (Study 1)

Date: **Participant Number:**

How old are you? _______________

Please circle: Male Female

How often do you use a computer?

1) Never
2) Rarely
3) Sometimes
4) Often
5) Every Day

Which input device do you use most frequently?

1) Mouse
2) Trackpad
3) Touchscreen
4) Stylus
5) Other (Please specify)

DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF SASKATCHEWAN
INFORMED CONSENT FORM

Research Project: **Determining Effective Height Representations for the Height of Deictic Gestures**

Investigators: Dr. Carl Gutwin, Department of Computer Science (966-8646)

Aaron Genest, Department of Computer Science (966-2327)

This consent form, a copy of which has been given to you, is only part of the process of informed consent. It should give you the basic idea of what the research is about and what your participation will involve. If you would like more detail about something mentioned here, or information not included here, please ask. Please take the time to read this form carefully and to understand any accompanying information.

This study is part of a larger set of studies to determine how pointing gestures made in 3-Dimensional space can be visualized on a 2-Dimensional surface. In this study, we are examining how people interpret a visualization we have designed that represents, in part, the height of a gesture. During this study you will be asked to view several pre-recorded sets of visualizations, similar to very short movies, and then answer some short questions about each visualization. You will use the keyboard to answer these questions. Before the study starts, you will be given a short, pen and paper questionnaire with non-identifying demographic questions.

At the end of the session, you will be given more information about the purpose and goals of the study, and there will be time for you to ask questions about the research.

The data collected from this study will be used in articles for publication in journals and conference proceedings.

As one way of thanking you for your time, we will be pleased to make available to you a summary of the results of this study once they have been compiled (usually within two months). This summary will outline the research and discuss our findings and recommendations. If you would like to receive a copy of this summary, please write down your email address here.

Contact email address:____

All personal and identifying data will be kept confidential. If explicit consent has been given, textual excerpts, photographs, or video recordings may be used in the dissemination of research results in scholarly journals or at scholarly conferences. Anonymity will be preserved by using pseudonyms in any presentation of textual data in journals or at conferences. The informed consent form and all research data will be kept in a secure location under confidentiality in accordance with University policy for 5 years post publication. Do you have any questions about this aspect of the study?

You are free to withdraw from the study at any time without penalty and without losing any advertised benefits. Withdrawal from the study will not affect your academic status or your access to services at the university. If you withdraw, your data will be deleted from the study and destroyed.

Your continued participation should be as informed as your initial consent, so you should feel free to ask for clarification or new information throughout your participation. If you have further questions concerning matters related to this research, please contact:

- Dr. Carl Gutwin, Full Professor, Dept. of Computer Science, (306) 966-8646, gutwin@cs.usask.ca

Your signature on this form indicates that you have understood to your satisfaction the information regarding participation in the research project and agree to participate as a participant. In no way does this waive your legal rights nor release the investigators, sponsors, or involved institutions from their legal and professional responsibilities. If you have further questions about this study or your rights as a participant, please contact:

- Dr. Carl Gutwin, Full Professor, Dept. of Computer Science, (306) 966-8646, gutwin@cs.usask.ca
- Office of Research Services, University of Saskatchewan, (306) 966-4053

Participant's signature:____

Date:____________________

Investigator's signature:____

Date:____________________

A copy of this consent form has been given to you to keep for your records and reference. This research has the ethical approval of the Office of Research Services at the University of Saskatchewan.

Demographic Questionnaire for Gesture Visualization (Study 2)

Date: **Participant Number:**

How old are you? _______________

Please circle: Male Female

How often do you use a computer?

6) Never
7) Rarely
8) Sometimes
9) Often
10) Every Day

Which input device do you use most frequently?

6) Mouse
7) Trackpad
8) Touchscreen
9) Stylus
10) Other (Please specify)

Have you ever used or seen any of the following technologies (please circle those that apply)?

- Collaborative tabletop systems
- Large public touchscreen displays
- Geocollaborative systems (also known as collaborative GIS)
- Distributed groupware (e.g. multiplayer painting games)
- VNC or similar screen-sharing systems (e.g. iChat screen sharing)

DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF SASKATCHEWAN
INFORMED CONSENT FORM

Research Project: **Determining Effective Height Representations for the Height of Deictic Gestures**

Investigators: Dr. Carl Gutwin, Department of Computer Science (966-8646)

Aaron Genest, Department of Computer Science (966-2327)

This consent form, a copy of which has been given to you, is only part of the process of informed consent. It should give you the basic idea of what the research is about and what your participation will involve. If you would like more detail about something mentioned here, or information not included here, please ask. Please take the time to read this form carefully and to understand any accompanying information.

This study is part of a larger set of studies to determine how pointing gestures made in 3-Dimensional space can be visualized on a 2-Dimensional surface. In this study, we are examining how people communicate, using visualizations we have designed, some of which represent, in part, the height of a gesture. During this study you will be ask or answer several short questions about the city in which you live, communicating about these questions with another study participant. After answering (or asking) each of these questions, you will answer several pen and paper questions. Before the study starts, you will be given a short, pen and paper questionnaire with non-identifying demographic questions.

At the end of the session, you will be given more information about the purpose and goals of the study, and there will be time for you to ask questions about the research.

The data collected from this study will be used in articles for publication in journals and conference proceedings.

As one way of thanking you for your time, we will be pleased to make available to you a summary of the results of this study once they have been compiled (usually within two months). This summary will outline the research and discuss our findings and recommendations. If you would like to receive a copy of this summary, please write down your email address here.

Contact email address:____

All personal and identifying data will be kept confidential. If explicit consent has been given, textual excerpts, photographs, or video recordings may be used in the dissemination of research results in scholarly journals or at scholarly conferences. Anonymity will be preserved by using pseudonyms in any presentation of textual data in journals or at conferences. The informed consent form and all research data will be kept in a secure location under confidentiality in accordance with University policy for 5 years post publication. Do you have any questions about this aspect of the study?

You are free to withdraw from the study at any time without penalty and without losing any advertised benefits. Withdrawal from the study will not affect your academic status or your access to services at the university. If you withdraw, your data will be deleted from the study and destroyed.

Your continued participation should be as informed as your initial consent, so you should feel free to ask for clarification or new information throughout your participation. If you have further questions concerning matters related to this research, please contact:

- Dr. Carl Gutwin, Full Professor, Dept. of Computer Science, (306) 966-8646, gutwin@cs.usask.ca

Your signature on this form indicates that you have understood to your satisfaction the information regarding participation in the research project and agree to participate as a participant. In no way does this waive your legal rights nor release the investigators, sponsors, or involved institutions from their legal and professional responsibilities. If you have further questions about this study or your rights as a participant, please contact:

- Dr. Carl Gutwin, Full Professor, Dept. of Computer Science, (306) 966-8646, gutwin@cs.usask.ca
- Office of Research Services, University of Saskatchewan, (306) 966-4053

Participant's signature:____

Date:____________________

Investigator's signature:____

Date:____________________

A copy of this consent form has been given to you to keep for your records and reference. This research has the ethical approval of the Office of Research Services at the University of Saskatchewan.

Demographic Questionnaire for
Gesture Visualization (Study 3)

Date: **Participant Number:**

How old are you? _______________

Please circle: Male Female

How often do you use a computer?

11) Never
12) Rarely
13) Sometimes
14) Often
15) Every Day

Which input device do you use most frequently?

11) Mouse
12) Trackpad
13) Touchscreen
14) Stylus
15) Other (Please specify)

Have you ever used or seen any of the following technologies (please circle those that apply)?

- Collaborative tabletop systems
- Large public touchscreen displays
- Geocollaborative systems (also known as collaborative GIS)
- Distributed groupware (e.g. multiplayer painting games)
- VNC or similar screen-sharing systems (e.g. iChat screen sharing)

Gesture Visualization (GVexp3)
Instructions and Questionnaire
Local Participant (creates visualizations)
Enhanced Telepointer Condition

Instructions

In this study, you will have to answer a series of questions, using your Skype connection, from your collaborator, the person in the next room. When you are asked a question, answer by talking over Skype and pointing (using the finger attached to the Polhemus sensor) at various locations on the map displayed in front of you.

Sometimes, as you point, you will see a special cursor (called an embodiment) that shows you where you are pointing. The embodiment has several qualities that reflect what you are doing with your finger.

1. The embodiment will look like a bubble. Its **size** and **transparency** represents the height of your finger above the table.

a. Larger and fainter means your finger is **higher**, or farther from the table.
b. Smaller and darker means your finger is **lower**, or closer to the table.

The embodiment leaves behind traces that **fade** as time goes by, so you know where you have recently been pointing. This happens quite quickly when your finger is above the table.

The embodiment changes **shape** when you **touch** the table, becoming a large, green plus sign (+).

The embodiment leaves a **fading line** when you move your finger along the surface of the table. In other words, when you remain touching the table while moving your finger, you'll see a fading line.

The embodiment leaves **ripples**, like in a pond, when you start or finish touching the table.

Take a minute now, with your collaborator, and get used to the way the embodiment works. If you have any questions, please ask the researcher.

If you are ready, you may continue with the study by turning the page. Wait for your collaborator to ask a question, then answer as well as you are able. Your collaborator may ask for clarification if he/she doesn't understand at what you are pointing. Feel free to clarify as much as needed.

After you have answered the question, write down your answers to the additional questions on the page and continue on the the next page. Take your time and be sure you're happy with your answers.

After eleven questions, there will be a short break. Be sure to carefully read the instructions for the second part of this experiment as well. Thank you and you may begin.
ID 1

Please wait for your collaborator to ask you a question. Once you have answered to his/her satisfaction, continue with the questions below.

Questions for you

1. Did you believe your collaborator understood which location(s) about which you were speaking?

   1 2 3 4 5 6 7
   Not at all understood **Completely**
   understood

2. How confident were you in indicating the location(s)?

   1 2 3 4 5 6 7
   Not at all confident **Completely**
   confident

3. How quickly do you believe your collaborator was able to understand what you were indicating?

   1 2 3 4 5 6 7
   Not at all quickly **Very quickly**

4. How much effort did it take you to convey the location(s) to your collaborator?

   1 2 3 4 5 6 7
   No effort **Lots of effort**

5. How much emphasis did you use while indicating the location(s) on the map?

   1 2 3 4 5 6 7
   No emphasis **Lots of emphasis**

6. How specific were you in indicating the location(s) on the map?

   1 2 3 4 5 6 7
   Not at all specific **Extremely specific**

7. Were you able reach all of the locations identified during this question?

   Yes **No**

8. Were any of the locations you indicated off the edges of the map?

   Yes **No**

[NOTE: The remaining questions were the same as the first eight, repeated for each trial. As a result, they are not reproduced.]

Gesture Visualization (GVexp3)

Instructions and Questionnaire

Remote Participant (sees visualizations)

Enhanced Telepointer Condition

Instructions

In this study, you will have to ask a series of questions, using your Skype connection, of your collaborator, the person in the next room. After you ask a question, the person in the next room will answer by talking and pointing at various locations on the map displayed in front of you. As the person points, you will see a special cursor (called an embodiment) that shows you where he/she is pointing. The embodiment has several qualities that reflect what the person in the other room is actually doing with their fingers, hand, and arm.

1. The embodiment will look like a bubble. Its **size** and **transparency** represents the height of your collaborator's finger above the table.
    a. Larger and fainter means their finger is **higher**, or farther from the table.
    b. Smaller and darker means their finger is **lower**, or closer to the table.

The embodiment leaves behind traces that **fade** as time goes by, so you know where your collaborator has recently been pointing. This happens quite quickly when your collaborator's finger is above the table.

The embodiment changes **shape** when your collaborator **touches** the table, becoming a large, green plus sign (+).

The embodiment leaves a **fading line** when your collaborator moves their finger along the surface of the table. In other words, when they stay touching the table while moving their finger, you'll see a fading line.

The embodiment leaves **ripples**, like in a pond, when your collaborator starts or finishes touching the table.

Take a minute now, with your collaborator, and get used to the way the embodiment works. If you have any questions, please ask the researcher.

If you are ready, you may continue with the study by turning the page. Ask a question, then listen to the answer. If you are uncertain about what your collaborator is asking, clarify as much as you need.

After you feel you understand what your collaborator was trying to tell you, write down your answers to the additional questions on the page and continue on the the next page. Take your time and be sure you're happy with your answers.

After eleven questions, there will be a short break. Be sure to carefully read the instructions for the second part of this experiment as well. Thank you and you may begin.

Questions for you

1. Did you understand which location(s) about which your collaborator was speaking?

   1 2 3 4 5 6 7
   Not at all understood **Completely understood**

2. How confident do you think your collaborator was in indicating the location(s)?

   1 2 3 4 5 6 7
   Not at all confident **Completely confident**

3. How quickly do you believe you were able to understand what your collaborator was showing you?

   1 2 3 4 5 6 7
   Not at all quickly **Very quickly**

4. How much effort did it take you to understand your collaborator?

   1 2 3 4 5 6 7
   No effort **Lots of effort**

5. How much emphasis do you believe your collaborator used while indicating the location(s) on the map?

   1 2 3 4 5 6 7
   No emphasis **Lots of emphasis**

6. How specific was your collaborator in indicating the location(s) on the map?

   1 2 3 4 5 6 7
   Not at all specific **Extremely specific**

7. In your opinion, was your collaborator able to reach all of the locations identified during this question?

   Yes **No**

8. In your opinion, were any of the locations indicated by your collaborator off the edges of the map?

   Yes **No**

   [NOTE: The remaining questions were the same as the first eight, repeated for each trial. As a result, they are not reproduced.]

APPENDIX C
REPRESENTING GESTURES

Abstract Representation

Abstract representations show information about users in non-realistic ways by using symbols and icons. Telepointers are an excellent example of a successful abstract design and highly effective in many contexts [13], [69], [70]. Abstract representations are successful for two reasons: they are relatively easy to implement, and they often have lower network bandwidth requirements than realistic representations.

There are also problems with abstract representations: they may require more cognitive effort to learn and use, and they may not be able to express all of the nuance of realistic representations. Whereas it takes little cognitive effort to interpret people's movements and body positions in a collocated context, people who see abstract representations must make an effort to ground the symbols or icons [114] (although certain kinds of abstractions require less effort to ground than others). This means that abstract representations must be learned in order to be usable and may require more cognitive effort than realistic embodiments. For example, although users were able to understand and recall most of the pieces of abstract information in Stach *et al.*'s rich embodiments [36], but did not use all of the pieces during collaborative activities (see Figure 52).

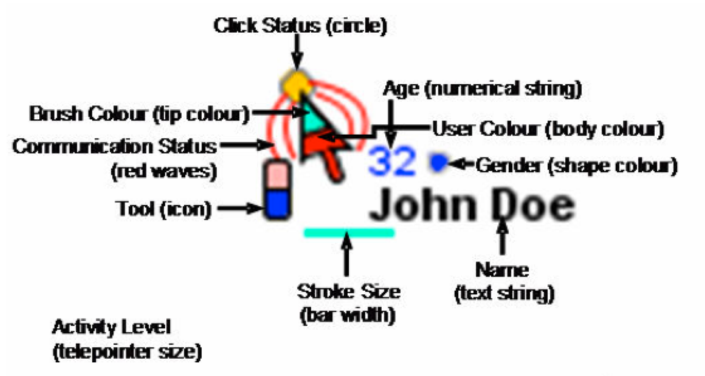


Figure 52: Abstract representations can be complex, but users may find using those pieces more cognitively challenging than if the information were conveyed through a realistic embodiment (image is from [36]).

The second potential problem for abstract representations is the range of nuance they convey. The human body is capable of complex physical communication through gesture. Abstract representations naturally limit communication because the process of abstraction is one of reducing complexity. Although abstract representations can be designed to be more or less complex, there may be an upper limit on the amount of information about gestures that can be encoded symbolically. There is also a tension between adding information and keeping the representation simple: abstraction reduces the complexity of a representation to those elements that are most important for interpretation; adding too many pieces of information may defeat the purpose of the abstraction.

Abstract representations (with a few exceptions, discussed below) are unable to show many of the more complex and varied morphologies exhibited during deixis and may have difficulty conveying the difference between different atoms of movement. However, they can be used to represent basic forms of deixis, such as single-finger pointing.

In the observational work described in Chapter 3, most of the deictic gestures used a single finger to point. This suggests that most of the time, abstract representations of location and

morphology are sufficient for representing the core meaning of a deictic gesture: indicating a target. However, the observational study also showed that vertical displays used in presentation-style discussions seemed to encourage non-single-finger pointing. Therefore, abstract techniques that may work well for horizontal surfaces may not be as effective for vertical surfaces.

The morphological classification in Chapter 3 describes only what portions of the hand were engaged in the gesture, not the shape of the parts of the body not engaged in the gesture. For example, for an index-finger pointing gesture, a more complete characterization would include the posture of the remaining fingers (i.e., curled tightly into a fist, curled loosely, or extended but not engaged; see Figure 53). None of these differences are represented in the morphology characterization in Chapter 3, but there are clear differences in how much of this variation abstract representations would be able to convey. If these postures are important for a particular collaboration setting, designers must consider whether an abstract representation can be designed in such a way that these aspects of the characteristic will be apparent to remote users.

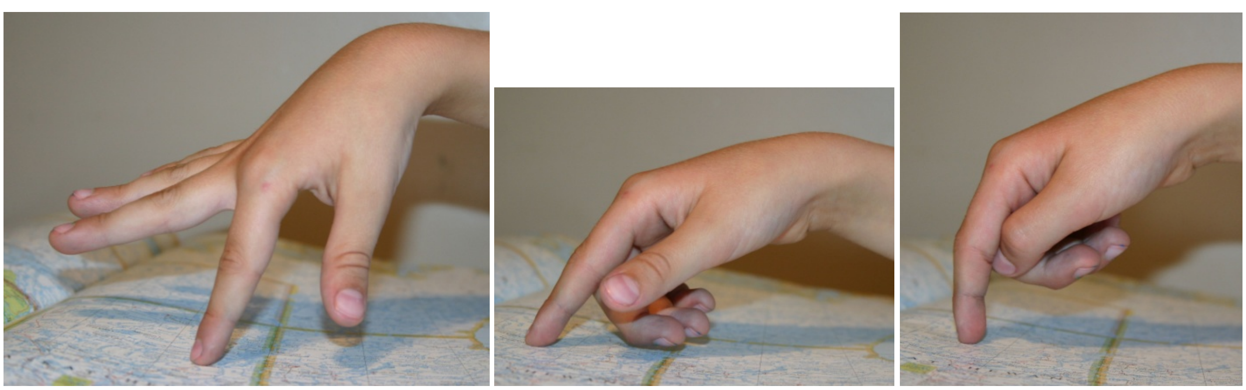


Figure 53: All three of the above gestures can be described morphologically (as per Chapter 3) as index-finger pointing (with no other engaged parts of the hand or arm). However, they look quite different and could possibly mean subtly different things.

One form of abstract representation can be effective in showing complex gesture morphologies: abstract 'skeletons' that show multiple linked points on user's hands and arms

(such as fingertips, palm centres, and elbows) can show complicated shapes and complex gesture variations. However, capturing this information is difficult, requiring either computationally expensive video separation or complex capture technologies (e.g., Vicon motion capture). These requirements create trade-offs: improved network performance comes at the cost of increased computational requirements. Abstract skeletons and similar representations also approach more realistic representations in their volumetric location information. The location of a gesture is properly the volume of space occupied by the hand and arm, a space that can be interpreted with a sufficiently detailed group of points..

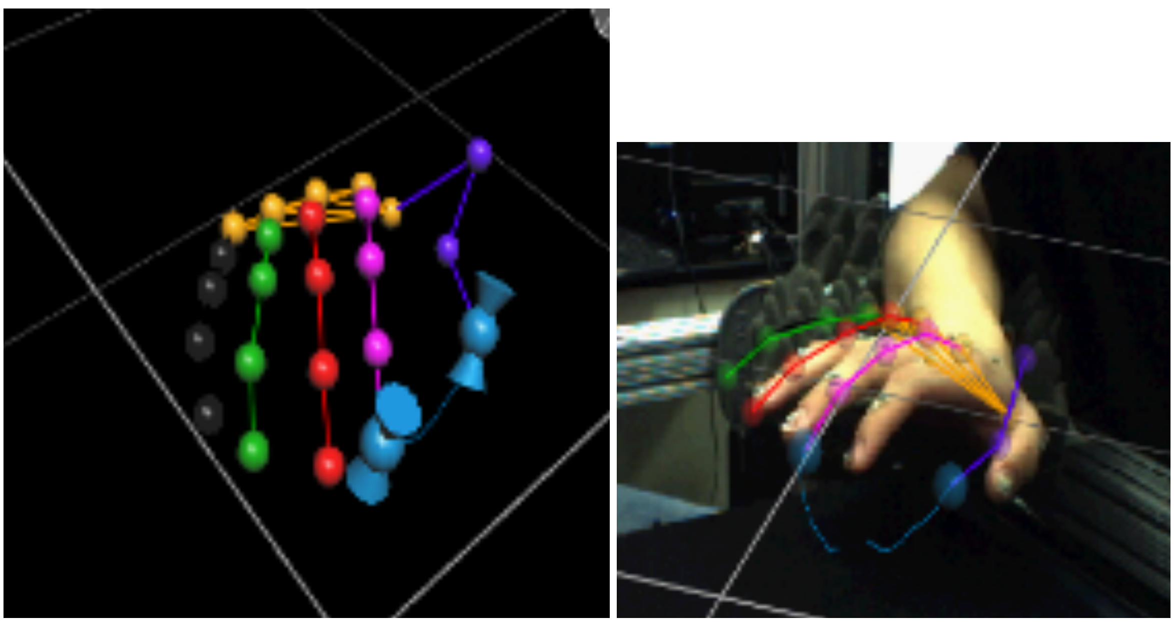


Figure 54: Vicon motion capture can identify enough points on the hand and arm (right) to allow people to interpret the resulting wireframe (left) as a volume of space. (Photo credit: Johns Hopkins Institute)

Representing the movement characteristics of deictic gestures is similarly complex for abstract representations. In some observed movements (reported in Chapter 3), there was very little visual difference between stroke and preparation atoms, between hesitation atoms and combinations of stroke and point atoms, and between rest atoms and point atoms. Although coincident conversation sometimes disambiguates these categories, differences between them

can also often be seen in the shape of the hand. For instance, a pointing finger is usually completely extended in a point atom, but may be partially curled under during a rest atom. This means that changes in the morphology of the hand can give clues to the kind of movement that is occurring during a gesture. Again, simple abstract representations of morphology may complicate interpretation in these cases.

Many abstract representations of the location and movement characteristics of gestures are limited to showing individual points or groups of points rather than volumetric representations of the hand and arm – often, an icon (such as a mouse pointer) represents the intended target of attention (e.g., [34], [36], [71]). The work in Chapter 3 suggests that this can be an effective solution for expressing most of the important communicative information in deixis, such as the target of a gesture. However, the state of the art in abstract representations (e.g., [36]) fails to represent much of the subtlety of gestures, including the height of a gesture above the work surface.

Realistic Representation

Realistic representations of users and their actions usually show the shape, visual texture, and location of a user's body or body parts. This is commonly done through the use of video (e.g., [47]), mediated video (e.g., [14]), or reconstruction based on video capture (many examples of this exist in the film industry). The realism provides a naturally high amount of information, but comes at the cost of higher bandwidth and computational costs. Like abstract representations, but for slightly different reasons, realistic representations also struggle with capture problems: capturing users' bodies with enough fidelity to usefully represent their movements for remote users can be computationally expensive (e.g., video separation), involve

complicated and expensive hardware (e.g., Vicon motion capture), and interfere with natural movement in a collaborative space (e.g., wired sensors).

In general, realistic representations excel at providing high levels of nuanced gesture information. Indeed, higher-bandwidth representations of hands and arms are likely the only solutions that can fully convey the subtleties of a gesture to remote collaborators. In terms of existing solutions, DOVE-like embodiments [75] have a clear advantage over more abstract representations in this regard (see, for example, Figure 55).

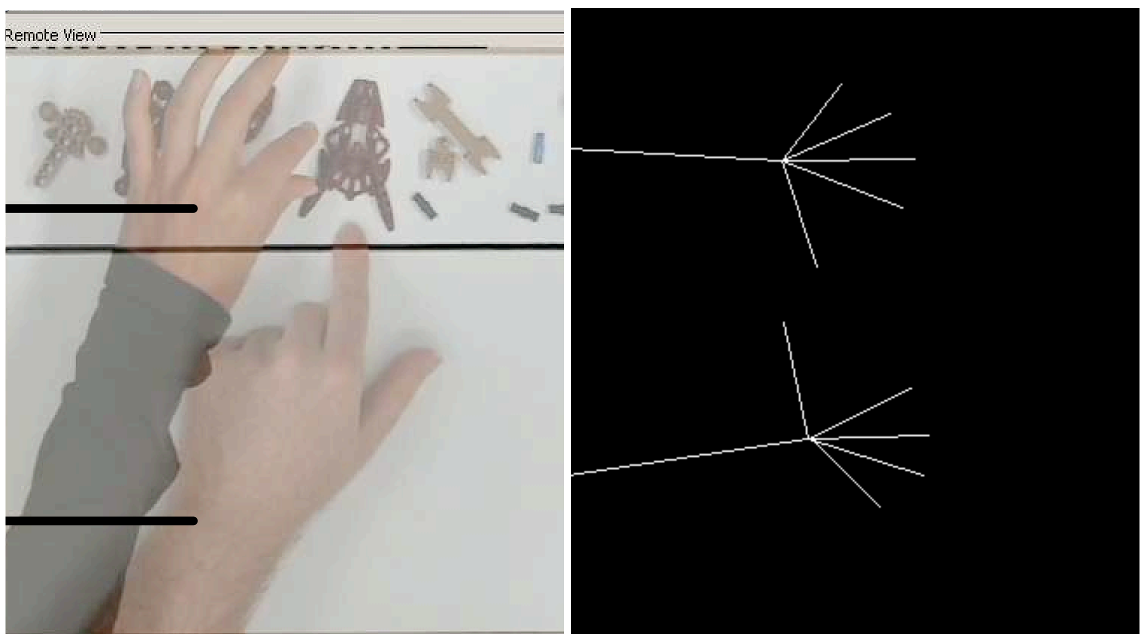


Figure 55: DOVE's realistic representations of hands and arms (left) provide far more subtlety and complexity than skeleton-like abstract representations based on individually tracked points on the user's hands (right).

However, designing an embodiment with realistic representations of gesture characteristics may be insufficient for effectively expressing a gesture. As previously mentioned, single-finger pointing is extremely common; however, some embodiments that use realistic hand and arm representations fail to properly express this behaviour. For example, VideoArms [14] suffers from inaccurate video-separation techniques and low-resolution video capture, which can

create blocky and unclear boundaries in the arm visualizations and thus fail to correctly show an extended finger in the embodiment [14] (see Figure 56). As with abstract representations, realistic representations must be designed with attention to the parts of a gesture that are important to convey. This may involve trade-offs in quality, depending on technological restrictions.

Figure 56: VideoArms in use. The embodiments are blocky and have interrupted boundaries, possibly making it difficult to interpret a gesture.

Realistic representations may also fail to express certain types of stroke movement atoms. In the observations detailed in Chapter 3, although point and stroke atoms were seen with almost equal frequency, stroke atoms were the most flexible in their application. Strokes were used to identify paths, as might be expected, but were also used for points and for areas. Almost every participant in the studies described in Chapter 3 performed gestures where multiple repeated strokes were used to emphasize points (similar to a rubbing gesture, as previously reported by Kirk et al. [11]). Strokes were also used to identify areas by performing a series of strokes similar to “shading in” the area. However, the limited visual representation of a remote embodiment can fail to adequately convey the salience of these gestures – for example, a ‘rubbing’ emphasis gesture with a 2D video arm may be less obvious than the corresponding

real-world original, due to lower sampling and display frequencies in remote capture and display technologies.

Realistic representations can express a greater range of gesture morphology than abstract representations, although with the substantial caveat that the quality must be high enough for those morphologies to be identifiable. Realistic representations generally fail to represent height effectively (due to the 2D projection of the image); do not show some gesture characteristics, such as pressure; and are less noticeable than real arms in co-located settings, so may not adequately convey temporal aspects of stroke atoms.

Table 11: Embodiment techniques and their effectiveness in conveying deixis.

Embodiment Technique	**Advantages for Conveying Deixis**	**Disadvantages of Conveying Deixis**
Abstract	Effective core deixis, such as identifying the target of pointing; low network cost; often low computational cost	Fail to represent complex morphologies; currently fails to represent gesture height
Realistic	Highly effective in conveying both complex and simple deixis	High network cost; often high computational cost; fails to represent gesture height.

Hybrid Representation

A single characteristic of a gesture, such as the location of the gesture, might be represented using either a realistic representations of the arm and hand or an abstract representations of those body parts. However, combining the two can achieve the best of both worlds (see Table 12 for how characteristics are better represented by either abstract or realistic representation). For instance, a realistic representation of the hand and arm combined with an abstract representation of its height (e.g., an artificial shadow cast on the workspace surface), provides the expressiveness of a realistic approach by showing the complex location information of an arm and hand while providing extra information about height.

Table 12: Capabilities of abstract and realistic representations in showing selected gesture characteristics.

Characteristic of Deictic Gesture	**Abstract Representation**	**Realistic Representation**
Differences between similar atoms	No	Possible
Stroke atoms	Yes	No
One or two finger pointing	Yes	Yes
Morphology	No	Possible
Height	Possible	No
Wiggle variation	Possible	Yes
Pressure variation	Possible	No
Width variation	Yes	Yes

In the real world, there is only one way of perceiving a gesture: the visual feedback received by watching it. In a digitally-enhanced environment, however, abstract and realistic representations can be combined to produce embodiments tuned to the specific requirements of the distributed work. Key aspects of an embodiment can be manipulated to emphasize or de-emphasize information, and information not available in one kind of embodiment style can be added by using a different design approach. These three methods of manipulating embodiment designs to provide more expressive designs in distributed settings are discussed below (see Table 13 for an example of how each of these methods can be applied to deixis characteristics).

In collocated settings, the intended targets of a gesture can be interpreted based on the direction of pointing. Although this characteristic is available in a collocated setting, it is not always perfectly clear and may be more difficult to identify in a distributed setting. Adding this characteristic of deixis to an embodiment by highlighting the target, artificially extending the pointing finger to the target, or performing other visual manipulations may be useful for

improving target identification in distributed settings. It may also be useful to provide this representation as feedback to the local user, not just through a remote embodiment.

With a hybrid approach to developing embodiments, some gesture characteristics can be emphasized by applying visual effects. For example, pressure applied to the surface of the table during a gesture, difficult to see in distributed settings, could be emphasized by changing the colour of parts of the hand in contact with the surface. Touch, difficult to differentiate from hover, could be emphasized by creating a ripple effect under the parts of the hand touching the surface, something already done by Wigdor *et al.* [131]. Movement history, something that can help increase the salience of remote gestures and make them easier to interpret, can be enhanced with short-term, iconic representations of history, such as telepointer traces [13], interaction ripples [131], trace pearls on VideoArms [17], and a motion replay technique [93].

Gesture characteristics can be diminished, or reduced in salience, as well. For instance, the presence of an arm high above the surface might be less important than one on the surface (and with top-down video capture, the higher arm may actually appear larger, despite being less important). Varying the opacity of the embodiment (i.e., making it more transparent as the height of the arm increases) can convey this intended quality by diminishing the presence of the arm as it gets higher (e.g., as in [15]).

Figure 57: Left, phantom presence in C-Slate diminishes the presence of arms and hands above the surface [15]; right, interaction ripples enhance surface touches [131].

Even with the best representations, information can be missing in remote settings. Showing 3D objects on a 2D display means that location and movement on the z-axis is usually missing. Network problems can mean that captured frames of hands and arms can arrive out of order, too late to be shown, or not at all. Pressure on the surface of a table can be easily missed if either capture or display technologies are not perfect. In many cases, however, embodiment designs can accommodate for some of this lost information. For example, height can be shown with abstract circles, movement can be simulated between missing frames, and pressure can be shown with floating text.

Table 13: Deixis characteristics and examples of how embodiments can add appropriate user characteristics, augment or diminish already existing information, or add missing information.

Deixis Characteristic	**Adding Characteristics**	**Augmenting/Diminishing Existing Information**	**Adding Missing Information**
Height	History of height movement shown as 3D abstract images	Increase transparency as gestures increase in height	Abstract representation of height, e.g., an expanding circle under finger represenation/reproduction
Touch	Frequency of touch events shown as heatmap on fingertips	Change in embodiment colour saturation when touching	Floating icon that representing touching or hover state
Pressure	Show nominal	Simulate cracked surface	Floating text showing

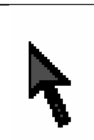
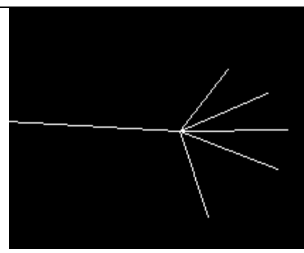
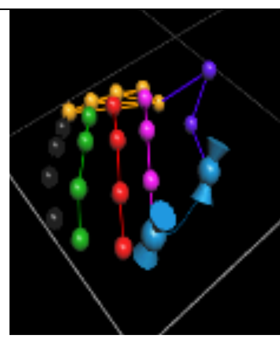
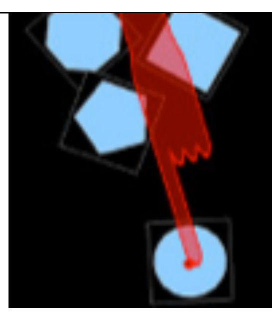
	levels of pressure (e.g., light, medium, heavy)	when pressure is high	kg/cm^2
Movement	Long-term history of movement over a session	Slowly fading motion traces	Movement between captured frames is extrapolated and displayed
Target	Frequently identified targets are highlighed	Pointing fingers emit laser pointers	Extension of hand and arm is extrapolated to potential targets, which are illuminated
Morphology	The skeletal structure of the hand is added.	When hand is in common positions for deixis, it receives a blue outline.	Hidden fingers are drawn as phantom fingers
Location	Body is represented even if not captured	Hands at the edges of the table are not shown	x and y positions of pointing finger relative to surface are shown as numeral values

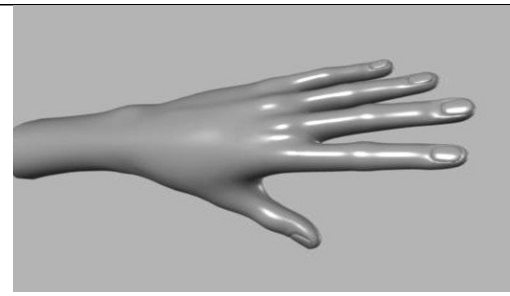
Representing Information about Deictic Gestures

Location

Location can be shown with both realistic and abstract representations, although (currently) abstract representations tend to show less detailed location information. Table 14 shows examples of how abstract and realistic representations have been used to show the location of pointing. There is almost no overlap between the two styles of representation; embodiments that represent users as points are abstract and almost all of those that represent them as areas and volumes are realistic. Of note is work by Fraser *et al.* (listed in Table 14 at the intersection of 3D Point and Abstract Representation), who encoded three dimensions of information into a single embodiment, but because of the nature of the visualization (higher gestures increased the size of the embodiment and blurred the boundaries more) sacrificed the precision present in a standard teletpointer.

Table 14: Examples of the way that embodiments represent location using abstract or realistic styles.

Location Information	Abstract Representation	Realistic Representation
2D Point (x,y position)	 [23], [69]	
3D Point (x,y,z position)	 [34]	
Multiple 2D Points		
Multiple 3D Points		
Area	 [24]	 [17]

Volume		

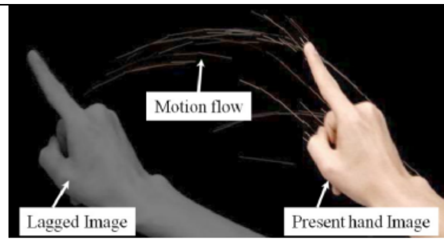
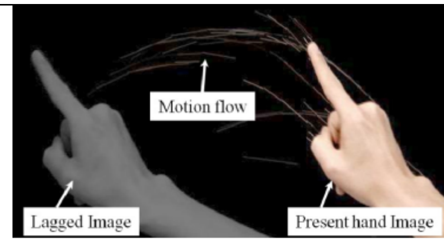
Movement

The movement characteristics of deictic gesture can be represented with either abstract or realistic techniques, although emphasizing certain characteristics works better with hybrid representations (see Table 15). One example of this is the work by Yamashita *et al.* on gesture replays [93]: Two realistic hand images, a the current location and a historical location, are joined by abstract lines showing the vector of movement. Pressure has been captured, but not yet represented in any remote embodiment technique. However, an abstract technique such as Ripples [131] could be adapted to show changes in pressure.

There are currently no embodiment techniques, either realistic or abstract, specifically designed to do a good job of showing small movements and emphasizing preparation atoms, although some of the more realistic techniques may convey this information without emphasis. There are some techniques for representing changes in height, and these are discussed in detail (along with some novel techniques) in Section 0, below.

Table 15: Examples of how movement characteristics can be represented using abstract and realistic representations.

Movement Characteristics	Abstract Representation	Realistic Representation

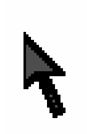
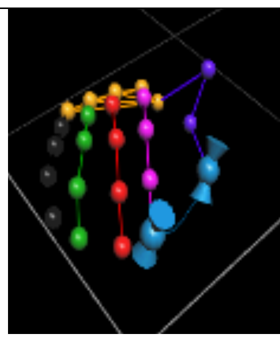
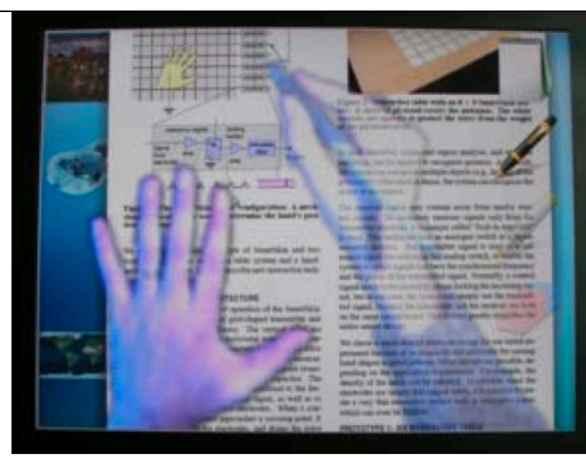
Movement in the x/y plane during main part of deixis	 [93] **Also, see Section 0, Representing Movement with Traces**	 [93]
Preparation atom	 [34]	
Changes in height	**See Section 0, Representing Height**	
Small movements (e.g., wiggles)		
Pressure		

Morphology

Abstract representations can show low or medium levels of morphological detail (see Table 16). Low levels are conveyed through techniques such as telepointers [23], [69] and Fraser's approach embodiment [34]. Medium levels of morphology are conveyed by techniques that use multiple points, perhaps in three dimensions. Realistic representations usually show medium and high levels of morphological detail. Two dimensional, shadow-like representations of hands and arms (e.g., [17]) show considerable morphology, but can make it difficult to

identify individual fingers or determine the relative height of hand components. More detailed representations (e.g., [15]) can provide these details, even if they are two dimensional images.

Table 16: Examples of how abstract and realistic representations can show varying levels of morphological detail.

Morphological Detail	**Abstract Representation**	**Realistic Representation**
Low (e.g., tip of finger)	 [23], [69]  [34]	
Medium (e.g., major points of hand and arm)		 [17]
High (e.g., full colour images)		 [15]

Representing Height

Despite the importance and richness of communication contained in gesture height, few embodiment designs represent gesture height. Some reasons for this limitation may be that height is difficult to capture with many common capture technologies and is difficult to reproduce on a 2D display. Capture technologies often gather touch information (e.g., FTIR tables) but technologies that can gather above-the-surface information (e.g., magnetic sensors or IR-based motion capture) are difficult to calibrate, cumbersome, or very expensive. Representations of height have commonly been limited to showing the difference between touching and not touching (e.g., [15] [17]).

Although some systems that show distance from a surface for the purposes of communication use a very basic and low fidelity representation (e.g., [34]), other techniques for using distance as feedback or interaction provide some possible design ideas. For instance, Shadow Reaching [91] used arm shadows that grow larger as users back away from the surface to interact with out-of-reach artifacts, a visualization that equally can be used as an embodiment. A similar shadow-like representation was used in the C-Slate system as feedback for 3D manipulations [15]. The original use of shadow in an embodiment, however, was by Tang and Minneman [16] in their VideoWhiteboard system, however the shadow was a side-effect of the analogue capture technique, not an explicit design decision. All of these systems use changes in size (and transparency, in most cases) as an indication of a user's distance from a workspace surface. Although the designs below explore a wider range of representations, the above work considerably influenced the design direction, therefore there are a greater variety and number of size-based designs.

There are general guidelines for visualizing data and specific examples of a few embodiment designs that do incorporate height information. Bertin established a set of guidelines

for visualizing information that includes a set of manipulations, or visual variable, that can be reasonably applied to visualizations (e.g., changes in position, colour, shape, or size) [132]. This helps narrow the field of possible visualizations and assists with identifying which visual variables are more appropriate for the ordinal above-the-surface height information, the nominal table touching information, and the geographic time series information for history. (Designs for showing historical movements are discussed further, below.) Mackinlay built on Bertin's work to rank the effectiveness of individual visual variables for different data types. In particular, position, colour hue and shape are effective for nominal categories; and density and colour saturation for ordinal information.

The next sections discuss how techniques for representing height based on the work referenced above can be used to show height, and in particular, how easy it is with each method to show touches or the absolute height of a gesture above the surface.

Shape

Changes in shape can be used to show touches and the absolute height of gestures. Touches can be represented by showing one shape above the surface and another on the surfaces (see Figure 58).

Figure 58: Shape representations of touch. Star and dot (left), mouse pointer and dot (centre), and bullseye (right).

Absolute height can be shown in two different ways with shapes. First, the height can be divided into an arbitrary number of layers and represented through changes in the number of sides, as in Figure 59.

Figure 59: Shape representations of height. More sides could either be higher or lower, depending on the direction of encoding.

Second, absolute height can be shown by gradually transforming a shape, such as a circle, into a different shape. Figure 60 shows an example of this kind of representation.

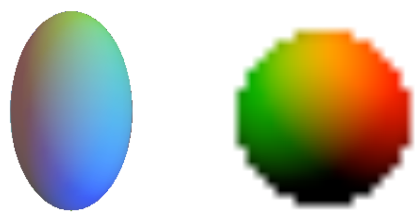


Figure 60: Changing shapes to represent changes in the absolute height of a gesture. In this case, low gestures (right) are represented with a circle. As gestures get higher, the circle elongates.

Colour

Colour is most effective for showing sudden changes, such as touches. A clear colour change (e.g., green to purple) can show touching/not touching or, when combined with similar distinct changes, can show layers of height (as with shape, above). Changes in absolute height can be represented with a black to white grey-scale spectrum where white is the maximum height and black the minimum; on a colour spectrum from red to blue with blue as maximum height and red as minimum height, and green to orange with orange as minimum height and green as maximum height (see Figure 61). However, variations through a spectrum of colour for absolute height are less effective than other uses of colour: many people are colourblind and would be

unable to see some of the colour changes and small changes in colour may not be easily visible when representing small changes in height.

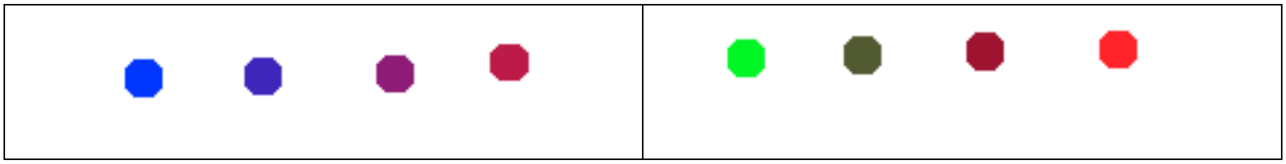


Figure 61: Examples of colour variations for absolute height representations. Left, from blue to red, right, from green to orange.

Size

Size variations are effective for showing touch and absolute height. For size to show touch, however, the embodiment must have a visible change in size between representing just above the surface and representing touching the surface. Without a visible change, touch would be very difficult to detect (see Figure 62). For example, if a 3-pixel diameter circle represents a touch and a 4-pixel diameter circle represents just above the surface, most users would be hard pressed to identify a touch as different from a hover.

Figure 62: Above, which one represents a touch? Small changes in size can be difficult to see.

Transparency

Although, like size and colour, changes in transparency can show touches, it most naturally maps to changes in absolute height (Figure 63), since small changes in the level of transparency can be hard to see. Transparency can be added to size changes to help avoid occlusion problems for very high gestures with correspondingly large embodiments (see Figure

64, top). Using both size and transparency may also increase the salience of changes in height (see Figure 64, bottom row).

Figure 63: Black circle that increases in transparency as gestures get higher. The dot in the middle is the centre of the circle and the only reference visible when the circle becomes completely transparent.

Figure 64: Using transparency as a secondary encoding helps avoid occlusion problems for size-based variations. Above, complex backgrounds are visible, despite the size of the embodiment. Bottom, the visual salience of changes in height may be increased by dual encoding height with both size and transparency.

Adding Graphical Elements

Graphical elements, such as symbols or shapes, can be added to an embodiment to show absolute height and touches. This method is least effective when representing touches for the same reason that transparency and size are poor methods: there is only a small change between

hovering (an embodiment with a single added graphic) and touching (an embodiment with no added graphics). For absolute changes in height, however, there are several approaches. One example of this is a method called satellites, which adds small, equidistantly spaced, graphics around a central graphic. The number of dots can increase with the height of the gesture, eventually forming an almost complete circle around the central circle when gestures are as high as a user could reach (see Figure 65).

Another option, called speedometer, uses progressive placement of graphics. They can be placed in a descending line (see Figure 66), or in a circle (Figure 67). The graphics can also be placed close together to appear as a line, to better represent the continuous nature of absolute height data (Figure 68). These can be combined with colour to take advantage of the strong, enculturated familiarity with speedometer gauges on vehicles (Figure 67 and Figure 68).

Figure 65: Satellites added small, equally spaced graphics around a central graphic. The graphics can increase in number as height increases (left to right).

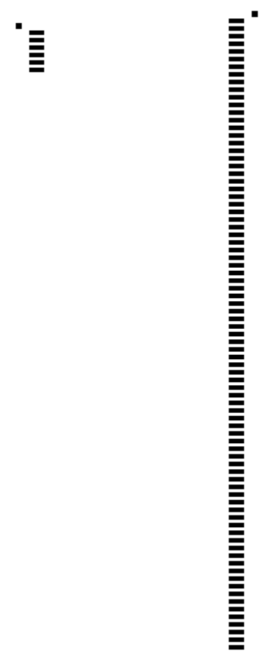


Figure 66: Speedometer uses an increasing number of graphics to show increasing values of gesture height (the left image represents a lower gesture than the right).

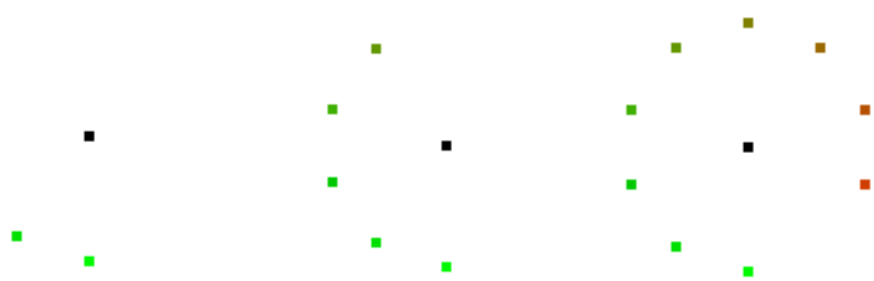


Figure 67: Speedometer can also use progressive, circular representations. In this example, the graphics appear at the bottom and increasing numbers are added clockwise as gestures increase in height (left to right). Colour can be used to increase the association with vehicle speedometers.

Figure 68: Graphics can be placed close together to appear as a line, to better represent the continuous nature of absolute height data. In this example, the technique eventually creates a circle for either a very high gesture or a gesture that touches the surface. Colour can be used to increase the association with vehicle speedometers.

Representing Movement with Traces

In distributed collaboration, gestures (or portions of gestures) can be missed by remote participants. Since many gestures are fast and incorporate subtle information, missed gestures can cause problems during distributed communication. In addition, distributed groupware can experience network jitter or low frame rates, which make it difficult to understand remote interactions and communication. Previous research has improved interpretation of and attention to gestures by including temporal traces as part of the embodiment [13], [93].

Contact traces, temporal traces, or other variations on visualizing past interactions (with others or with a system) were implemented by Tang *et al.* [17], Yamashita *et al*. [93], and Gutwin and Penner [13] in the context of remote embodiments. In the context of interface feedback, Wigdor *et al.* [131] experimented with using ripples to provide a very short term history of touches on a surface.

Temporal traces can be applied to both representations of gestures that touch the surface and gestures above the surface. On the surface, representations can either be sample-based, in that only periodic samples of the embodiment are shown (Figure 69 and Figure 70), or continuous. Sample-based representation allows the graphic that represents the touch to remain

as the temporal trace, whereas continuous representations have to find alternative trace representations.

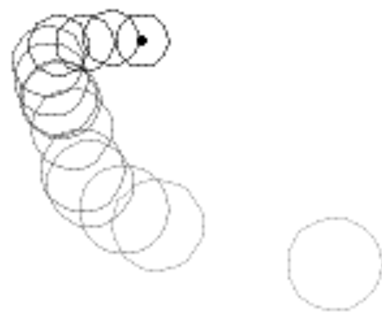


Figure 69: Ripples as temporal traces. Individual ripples expand and fade over time.

Figure 70: Temporal traces using a mouse pointer as a graphical icon. The current location of the gesture is the embodiment with the small square at the end of the pointer (bottom, right).

Figure 71: An example of continuous temporal traces. The ripple (right side) emphasizes the beginning of a touch.

Above the surface, embodiments can use temporal traces as well. Sample-based representations are likely better than continuous representations because of the complexity of showing the third dimension of movement (see Figure 72).

Figure 72: Temporal traces above the surface. The left-most, and highest, embodiment is the present location of the gesture. The location of earlier embodiments are represented by the gradually fading embodiments trailing down and to the right.