Characterizing Deixis over Surfaces to Improve Remote Embodiments

Aaron Genest and Carl Gutwin
University of Saskatchewan
aaron.genest@usask.ca, gutwin@cs.usask.ca

Abstract. Deictic gestures are ubiquitous when people work over tables and whiteboards, but when collaboration occurs across distributed surfaces, the embodiments used to represent other members of the group often fail to convey the details of these gestures. Although both gestures and embodiments have been well studied, there is still little information available to groupware designers about what components and characteristics of deictic gesture are most important for conveying meaning through remote embodiments. To provide this information, we conducted three observational studies in which we recorded and analysed more than 450 deictic gestures. We considered four issues that are important for the design of embodiments on surfaces: what parts of the body are used to produce a deictic gesture, what atomic movements make up deixis, where gestures occur in the space above the surface, and what other characteristics deictic gestures exhibit in addition to pointing. Our observations provide a new design understanding of deictic gestures. We use our results to identify the limitations of current embodiment techniques in supporting deixis, and to propose new hybrid designs that can better represent the range of behavior seen in real-world settings.

Introduction

Distributed tabletop applications often provide user embodiments to facilitate communication and improve awareness among collaborators (Benford et al., 1995). These embodiments improve communication between distributed groups in part by showing gestures made by remote users. These gestures are often deictic – that is, indicatory gestures accompanied by speech patterns such as “this one” or “over there” (Bekker et al., 1995). However, embodiments on tables often fail to fully represent the range of expression that is evident in collocated settings.

Embodiments can take many different forms, but can be classified into *realistic* embodiments such as VideoArms (A. Tang et al., 2007) or DOVE (Ou et al., 2003) and *abstract* embodiments such as telepointers (Hayne et al., 1994). Abstract embodiments have been shown to permit the expression of deixis (e.g., Gutwin and Penner, 2002), but are generally limited to conveying simple points or paths. Realistic embodiments represent hands and arms on the remote table with much greater fidelity (e.g., with video), but only in two dimensions, since the embodiment must be projected onto the table surface. In addition, video-based solutions rely on imperfect image separation techniques, can be limited in resolution, and require greater network bandwidth.

The different limitations of both embodiment approaches mean that distributed surfaces often fail to represent the full richness of deixis that can be seen in co-located work – such as the height of a gesture over the surface, or the difference between lightly touching and pressing hard on the surface. Although there has been a great deal of work done on both embodiments and deictic gesture, there is still little information available about how an embodiment should be designed to maximize the information shown to remote users of tabletop groupware. Embodiments cannot currently be evaluated in terms of their expressiveness because there is no clear characterization of deixis over surfaces, 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, we carried out three observational studies in which we recorded and analysed more than 450 gestures. 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. Our observations provided 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 gesture occurs in the space above the table, and what other physical characteristics different types of gestures exhibit.

In this paper, we describe our investigations of these questions, and use our results to analyse the strengths and limitations of realistic and abstract embodiments. From this analysis, we present a set of recommendations for the design of tabletop embodiments, and propose a new set of hybrid visualizations that can better represent the full range of expression and subtlety that is evident in real-world deixis.

Previous Work

Gestures and deixis are a key component of collocated communication (McNeill, 1992), and representing gestures for remote collaborators is an important groupware design factor (Gutwin & Greenberg, 2000). There are two foundations to our work: understanding the role of deixis in communication, and developing systems that successfully represent deixis with an appropriate level of fidelity.

Understanding Deixis

Although there are several studies that examine the use of gestures in a variety of communication contexts (e.g., Clark, 2003; Hindmarsh et al., 2000), 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. (1995), has found that gestures are used as a communication medium in face-to-face meetings almost as frequently as words. Bekker identified four categories of gestures, one of which was deixis (described as a “point” gesture). Other studies have identified the presence of understandable deixis as critical to smooth collaboration (Hindmarsh & Heath, 2000). One important result of deixis research is an understanding of how to evaluate deictic communication: the success rate of conversational grounding (how quickly people come to a common understanding of the target of deixis), can be used as a success metric (Kirk et al., 2007).

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.* (2005) examined hand movements during mixed-ecology collaborative tasks and developed a coarse-grained analysis of their characteristics. A finer-grained analysis was performed by Kettebekov and Sharma (2000), 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.

Representing Deixis

CSCW research has two primary approaches to embodying users who are interacting with remote physical spaces. Abstract techniques, such as telepointers, are designed with little or no regard to a true representation of the physical world, and instead represent relevant characteristics of users (Benford et al., 1995; Stach et al., 2007) or their actions (Greenberg et al., 1996). Conversely, realistic embodiments show users as accurate representations of their physical presence. Often, this takes the form of video superimposition of hands and arms into real workspaces. This technique takes a variety of forms, from the vertical surfaces and mirrored displays of VideoDraw and VideoWhiteBoard (J. Tang & Minneman, 1991a; 1991b) to DOVE's mixed-ecology system (Ou et al., 2003). Digital video of arms alone has also been used: in this approach, video of users' arms on the table surface is extracted and transmitted. VideoArms is an example of this technique (A. Tang et al., 2007), and was later enhanced to use temporal traces to show the previous positions of users' hands (A. Tang et al., 2010).

Research into abstract embodiments such as basic telepointers (Hayne et al., 1994) has examined which components of the users' movement and characteristics are important to visualize (Stach et al., 2007; Pinelle et al., 2008). Research has also demonstrated how telepointers can convey gestures, finding that the use of temporal traces can improve gesture recognition (Gutwin & Penner, 2002). Realistic embodiment research has concentrated less on the effectiveness of gesture representation than on the addition of additional channels of communication such as facial expression (Li et al., 2007). However, there is evidence that video-realistic embodiments still fail to express the full range of deixis, forcing users to adjust their gestures to accommodate the limitations of the representation (Ou et al., 2003).

With the exception of the work by Kettebekov and Sharma (2000), there is little work that classifies and explicates surface-based natural deixis from the perspective of designing embodiments to convey those gestures. Because of this lack, there are there are no systematic attempts to understand whether current embodiment techniques express the full richness of deixis to remote collaborators.

Examining Surface-Based Deixis

To further our understanding of how deixis is used over surfaces and whether embodiments capture this use, we performed three observational studies of deictic communication: two laboratory-based studies and one field study.

Laboratory Study Methods

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 the local city. Participants were allowed to move 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. We used maps for two reasons. First, map-based collaborations are common activities performed both by amateurs (e.g., tourism) and by professionals (e.g., emergency response, urban planning, land-use negotiation). Second, 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.

We recognize that maps do not provide a setting for all the collaborative tasks in which deixis is used, but maps do capture a large subset of this task space, particularly for two-dimensional displays.

In the first study, four pairs of participants answered questions about the spaces represented by the displayed map. Questionnaires were formulated using previous work by Kettebekov and Sharma (2000) as inspiration. Although the questions were asked by the researcher, participants were instructed to direct their answers to each other. 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 local university, and ranged in age from 22 to 56.

Sessions were videotaped with a single camera at an angle oblique to the table's surface. The resulting recordings were reviewed several times to identify episodes involving deictic gesture (deixis was separated from other kinds of actions such as conversational gesture), and to determine a set of candidate classification categories for the observed deictic gestures. Although the recordings contained a wide variety of deixis, our analysis identified limitations in our ability to capture gesture data: in particular, we determined that a single camera is 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 a single camera, participants' arms, hands, and bodies also sometimes occluded their gestures from our view. To resolve this problem, our 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. The second camera provided a top-down view of the table's surface.

The tasks in the second study were similar to that of the first, but we made small alterations in order to explore the issue of people's confidence in the accuracy of their answers. 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 local university community), aged 22 to 30.

We reviewed and analysed the recordings of the second study in a way similar to that of the first. We found that 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

We carried out a third study of a real-world collaboration, 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, we also hoped to observe a variety of collaborative tasks different from those we had created in the laboratory studies. We observed and recorded a group of four veterinary science students, one graduate-level teaching assistant, and a park warden 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 and hand-held paper maps, 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. 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.

Recordings and notes from the field study were analysed with methods similar to that used for the laboratory studies – we identified all episodes where gestures took place, and then categorized each instance, using the categories developed in the earlier studies. Overall, we found that the types of deictic gestures seen in the field were similar to what we observed in the laboratory, and that our existing categories were able to characterize all of the gestures of the field study. We did note, however, that many of the episodes we observed in the field were in a presentation style, with a vertical surface and a seated audience. We discuss the effects of this difference in our analysis below.

Analysis: four basic questions about deictic gesture

We analysed the video of the study sessions using four basic questions that 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?

What Parts of the Body are Used for Deictic Gesture?

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.

Our observations suggest that variations in the morphology of deixis over maps come primarily from the fingers and hand. The lower and upper arm, the shoulder, and the rest of the body are unlikely to play 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) provides valuable awareness information about the hand's location (e.g., that the speaker is 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. In our observations, we considered a part of the hand to be engaged in the gesture if it is not de-emphasized in the gesture (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, but not engaging the palm. The engagement of the palm (i.e., its importance in the gesture's interpretation) may not always be easily determined, but in our studies was often apparent in the larger context of the gesture.

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), palm-up (palm faces away from the surface), or sideways. This category is independent of the morphology of finger and palm engagement.

Of the parts of the hand actually used in deictic gestures, the index (1st) finger is of prime importance. In our first study, only 15 (6.7%) 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.

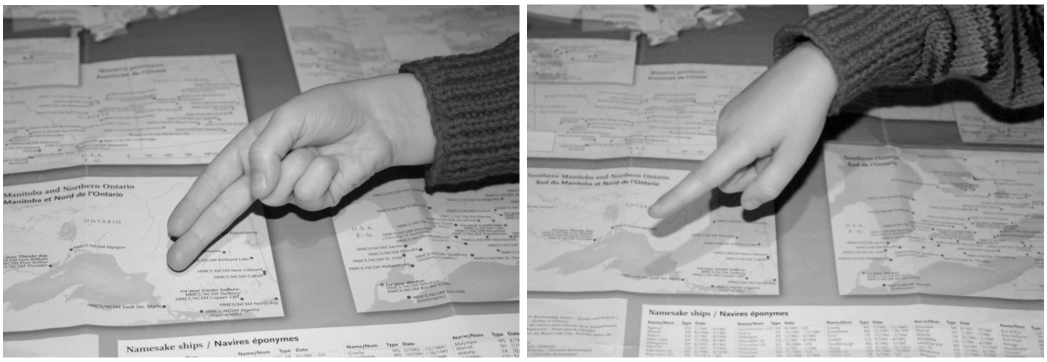

Figure 1. Examples of the two most common pointing gestures. Left, a double-finger point with the palm away from the surface. Right, a single-finger point with the palm down.

In contrast to the laboratory observations, during the field study 26 (31.7%) of the recorded gestures engaged the palm and only 41 (50.0%) of the gestures engaged the index finger. The engagement of the middle finger was used as an alternative to the index finger in 26 (31.7%) of the gestures. Additionally, 36 (43.9%) 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.

Our observations of deixis morphology over maps can be summarized in two ways: first, a large majority of episodes we observed 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, 94.8% of deictic gestures engaged the index finger, although often in conjunction with additional fingers. Finger-based pointing was also extremely frequent in the field study (81.7% of gestures).

The fact that pointing dominates deixis may not be surprising, but the degree to which we observed pointing-based gestures (even within other types of deixis) reinforces that this aspect must be considered as a primary component for remote embodiments. In current techniques, however, there are many situations where a remote embodiment may not make these gestures as clear as they need to be. For example, VideoArms suffers from inaccurate video-separation techniques and low-resolution video capture, which can create blocky and unclear boundaries in the arm visualizations, and can fail to correctly show an extended finger in the embodiment (Tang *et al.*, 2007). This guideline also helps explain why telepointer embodiments can be so effective – the telepointer clearly represents the tip of a pointing finger, and therefore expresses the most important morphology of deixis.

However, telepointers do not show any of the other more complex and varied morphologies we observed, and in these cases, higher-bandwidth representations of hands and arms (e.g., video or multiple-sensor setups) 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 (Li *et al.*, 2007) have a clear advantage over more abstract representations. The challenge in finding a middle ground between realistic and abstract embodiments is that identifying the important aspects of gesture morphology is difficult. We noted that two of our fourteen laboratory participants were responsible for 11 of the 15 observed non-single-finger pointing gestures. This suggests that between-subjects variability in deixis morphology could be a significant factor in the design and evaluation of gesture embodiments. In addition, we found that vertical displays, especially when used for presentation, encourage inverted hands (palm-up) and non-single-finger pointing. Therefore, abstract embodiment techniques that may work well for horizontal surfaces may not be as effective for vertical surfaces.

Finally, our classification describes only what portions of the hand are engaged in the gesture, not how that engagement (or disengagement) was manifested. For example, for an index-finger pointing gesture, we could also characterize the posture of the remaining fingers (curled tightly into a fist, curled loosely, or in a rest position next to the index finger but not touching the surface). None of these differences are represented in the morphology characterization we have used, but there are clear differences in how much of this variation the different embodiment approaches would be able to convey. If these postures are important for a particular collaboration setting, designers must consider whether a realistic embodiment (e.g., DOVE) has enough video fidelity to accurately represent these aspects of the gesture, or whether an abstract representation can be augmented to increase its expressive power (e.g., with additional sensors).

What Atomic Movements Create Deictic Gestures?

Gestures have been previously characterized in terms of small, atomic blocks of movement, a scheme designed to assist with automated classification (Kettebekov & Sharma, 2000). In this scheme, atomic gestures are strung together (often in long chains) to create complete gestures. Using this idea, we identified seven distinct atomic blocks from our observations, and uncovered several substantial problems for gesture embodiment as a result of this classification. The seven gesture 'atoms' we observed are: preparations, strokes, points, contours, retractions, rests, and hesitations. These differ only slightly from those of

Kettebekov and Sharma: we identify strokes as a kind of primitive, rather than a set of gesture primitives, and as a substitute for their 'circle' gesture; we re-define 'contour' as a closed stroke (very similar to their 'circle' primitive; and we add the hesitation atom. We discuss these atomic gestures in greater detail, below.

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 (Gutwin & Greenberg, 1998) 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 2. 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.

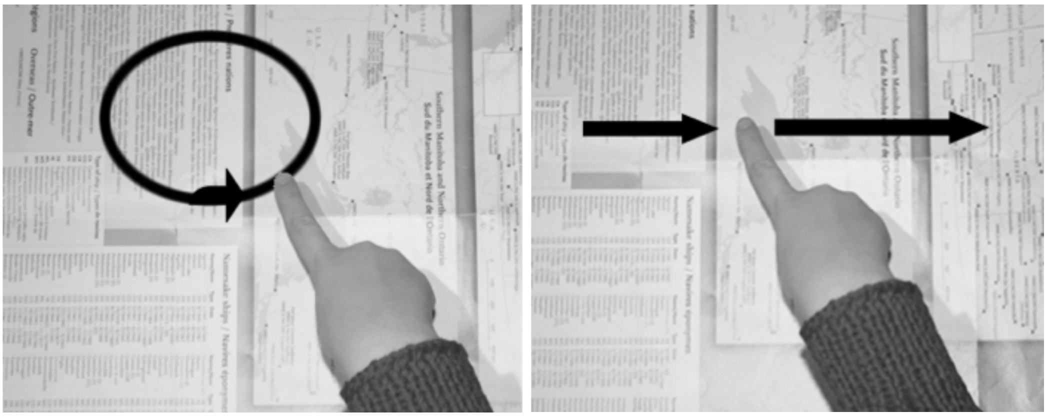

Figure 2. Contour atom (left) and stroke atom (right). Arrows indicate movement of the finger.

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. (2005) as the "wavering hand."

In all of the studies, we counted the frequency of indicative atoms. In the first study, there were 225 distinct indicative atoms, of which 91 (40.4%) were points, 104 (46.3%) were strokes and 30 (13.3%) were contours. In the second study, there were 146 indicative atoms, of which 56 (38.4%) were points, 61 (41.7%) were strokes, and 29 (19.9%) 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 (52.4%) pointing atoms, 35 (42.7%) stroke atoms, and 4 (4.9%) contour atoms.

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. Statements such as "somewhere over here", "I'm not sure where, exactly", or stalling vocalizations (e.g., "um") frequently accompanied hesitation atoms.

Our observations of gesture atoms suggest two potential problems with existing embodiment techniques, one for abstract embodiments and one for realistic embodiments. First, abstract embodiments like telepointers may have difficulty conveying the difference between different atoms. In some of our episodes, we observed 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 embodiments with higher fidelity and bandwidth (e.g., video-based representations) will be more effective in differentiating similar atoms.

Second, realistic embodiments may have difficulty representing certain types of stroke atoms – but these problems could be reduced by combining abstract visualizations with the realistic base representation. In our observations, 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 performed gestures where multiple repeated strokes were used to emphasize points (similar to a rubbing gesture, as previously reported by (Kirk et al., 2005). 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.

This is a situation where abstract embodiments, which can be arbitrarily augmented to show different types information, can be superior. There is evidence that using temporal traces on remote embodiments can improve collaborative communication (Gutwin & Penner, 2002); (A. Tang et al., 2010). The prevalence of stroke atoms, combined with their versatility, suggest that traces and similar techniques that enhance stroke atoms could be particularly valuable in improving the noticeability and interpretation of deictic gesture. A few previous embodiments have implemented such enhancements, often of contour atoms, and usually in conjunction with automated gesture recognition or gestures as commands (e.g., Ou et al., 2003). This space is largely unexplored, however, and there may be great value in combining realistic and abstract embodiments to convey stroke atoms. This approach could provide the clear indications of a telepointer, the ability to visualize temporal movement with visual traces, and the richness and expressivity of a video representation.

Where Does the Gesture Occur in the Space Above the Table?

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. However, representing height (or distance, in the case of a vertical surface) is limited in prior work. For example, this quality was shown incidentally as the opacity of the shadow in VideoWhiteBoard (J. Tang & Minneman, 1991), but later digital systems did not display this effect (e.g., Apperley *et al.*, 2003). Other systems display a separation effect, such as a narrow space between touching and hovering, as in the C-Slate system (Izadi *et al.*, 2007). In a system for collaborative video review (Fraser et al. 2007), distance from the display was explicitly and intentionally represented, in order to draw the attention of remote collaborators to an upcoming pointing action; this visual information was shown to improve coordination and reduce conversational latency. In our terms, this system visualized preparation atoms in order to improve the interpretation of pointing atoms.

Our observations also suggest that 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 our studies, height of a gesture, and variation in that height, was highly meaningful. We observed four main heights: gestures touching the surface, gestures moving back and forth from just above the surface to touching; gestures carried out entirely just above the surface, and gestures performed above about 5cm from the surface. 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.

Although a few embodiment solutions express some component of gesture height (e.g., VideoArms with traces (A. Tang et al., 2010)), no current technique expresses the full range of height seen in our studies. Some evaluations of remote embodiments have also noted the limitations of current approaches – for example, an evaluation of DOVE and DOVE-like embodiments (Ou et al., 2003; Li et al., 2007) found that participants modify their behaviour to adapt to the lack of height information in the embodiment. Our results suggest that height is a significant component of the information conveyed by a deictic gesture. A failure to fully represent height means that modified gestures or speech patterns must be employed to make up for the lost information. At best, conversational grounding takes longer without height information; at worst, information is lost, requiring greater effort to maintain the communication.

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, which often have cursory accuracy (Krauss & Fussell, 1990).

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., McNeill, 1992). 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.

From our observations, we present three additional characteristics that were seen several times during the studies. This is by no means an exhaustive set, but will serve to indicate the range of additional possible behaviours that can be observed in a deictic gesture. In each of our example characteristics, 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.

(1) *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).

(2) *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.

(3) *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 that rest (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. We note however that visual indications of pressure are subtle, and our 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.

Width variations were less homogeneous in usage: they were used to emphasize, to cover vague boundaries, to suggest areas rather than paths, to express lower levels of confidence, or as secondary referents to already-described locations. Width variations were more frequently accompanied with verbal utterances, which often clarified the reason for the width variation.

While both realistic and abstract 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 (e.g., FTIR technologies), and a visualization of this sensed value could easily be added to the embodiment of a remote participant.

Design Recommendations and New Designs

Our observations indicate that neither realistic nor abstract embodiments are entirely capable of representing the complex space of deixis over surfaces. In the next section we summarize the assessment of how each embodiment approach supports the representation of deictic gesture, and state a set of design principles that can be identified from our studies. We then consider the question of designing the best possible embodiment for showing deixis. We present one possibility – a hybrid design that combines the fidelity of realistic embodiments with an abstract embodiment’s ability to selectively emphasize or de-emphasize components of the gesture. This approach can offer an effective solution to the problems seen in traditional approaches, and can substantially improve the expressiveness of embodiments on tables and surfaces.

Assessment of Realistic and Abstract Embodiments

Both extremes of the abstract-to-realistic embodiment spectrum facilitate the visualization of some of the important components of deixis over a collaborative surface. However, as shown in Figure 3, neither abstract nor realistic embodiments adequately represent the full range of possible behaviours and subtleties in deictic gesture. Although telepointers and other abstract embodiments are able to convey point and stroke atoms extremely well, and can in some implementations represent variations such as height, they cannot represent the full range of gesture morphology we observed, nor are they effective in differentiating between similar gesture atoms, such as rest and point atoms. The main advantage to abstract representations is that additional information such as height, pressure, or past locations can be shown by enhancing the visualization.

Component of Deictic Gesture	**Abstract Embodiments**	**Realistic Embodiments**
Show differences between similar atoms	No	Possible
Represent stroke atoms	Yes	No
Express one or two finger pointing	Yes	Yes
Express full range of possible morphology	No	Possible
Represent height	Possible	No
Represent wiggle variation	Possible	Yes
Represent pressure variation	Possible	No
Represent width variation	Yes	Yes

Figure 3. Gesture visualization capabilities of embodiments

Realistic embodiments such as video-based solutions can express a greater range of gesture morphology than abstract representations, although with the substantial caveat that the source images must be effectively separated and transmitted at a high enough resolution for those morphologies to be identifiable. Realistic embodiments generally fail to represent height effectively (due to the 2D projection of the image), do not show 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.

Between these two extremes, however, lies a largely unexplored design space that combines the best parts of the two main approaches. The hybrid approach would use video-fidelity representations of hands and arms as the base representation for the embodiment, and enhance the image with abstract visualizations to represent components of the gesture that are not easily represented in two-dimensional video. An early example of this approach is the enhancement of VideoArms to show temporal traces whenever a user's fingers touched the surface (A. Tang *et al.*, 2010). This change provided some emphasis of stroke atoms (but only on the surface) and an indirect representation of height (by separating the touching layer from the other layers). This idea could be taken much further, however: a hybrid embodiment could show a definite focus point, temporal traces both above and on the surface, as well as visualizations of height, pressure, and other qualities. We present one possible hybrid visualization below.

Design Recommendations

Our observations and evaluation of existing approaches suggest five main recommendations for designers of embodiments for surface groupware:

(1) Embody the hand and some portion of the arm with a high-fidelity image. This representation provides a near-complete coverage of the different gesture morphologies, assists with the differentiation of similar atoms, and helps identify wiggle variations.

(2) Add an abstract visualization of the point of the index finger (i.e., a telepointer). This addition will clearly indicate point and stroke atoms, regardless of problems in the separation and resolution of the video image, and can serve as the basis for other abstract visualizations.

(3) Enhance the representation of movement with visual traces or other visual effects. This will assist in emphasizing stroke gestures and in representing width variations.

(4) Use an abstract representation for height. An abstract height representation as an addition to a realistic embodiment will assist in differentiating similar atoms and provide a clear mechanism for showing height variations. Particular attention should be paid to representing the difference between touching, just off the surface, and above the surface.

(5) Use an abstract representation for pressure variations (if these values can be gathered from other sensors). Since pressure is a subtle communication channel, the abstract representation should be equally subtle. Since we noted little use of pressure, compared to other variations, this is a low priority design recommendation.

Finally, we stress that both the task and the display/collaborative environment can have a profound impact on design. For example, we found that vertical displays tended to influence palm orientation, which could influence the selection of sensor suites and display techniques.

An Example Hybrid Embodiment

The five design recommendations above can be used to create a hybrid visualization that is able to represent all of the important components of deictic gesture. One example embodiment that includes some of this information (shown in Figure 4) uses the fidelity of realistic techniques and the flexibility of abstract techniques to create an expressive hybrid visualization. We assume that 3-D sensing techniques are available and that we can therefore gather the actual location of the hand and arm.

The embodiment contains the following components:

(1) High-fidelity top-down video image of the remote participant's hand and arm, reduced in opacity to avoid completely occluding the artifacts that are beneath the arm's representation.

(2) Telepointer spot that shows the position of the index finger.

(3) Temporal traces implemented as a motion blur to emphasize stroke atoms.

(4) Height representation as a shadow, offset slightly with respect to the seating position of the user to permit a parallax 'raycasting' effect that can improve interpretation of pointing direction for more distant targets.

(5) Nominal and binary representation of the touching/hovering height state is represented by a small ripple effect (as in Wigdor *et al.*, 2009) on contact (note that this also helps to emphasize strokes on the surface by acting as an implicit temporal trace).

(6) Pressure variations represented as a heat map on contact points that change colour from blue for a light touch to red for heavy pressure.

This example embodiment addresses each of the design principles introduced above. The methods we used to address these design recommendations are arbitrary, although future work could determine whether, for instance, stroke atoms are best enhanced through trace-like effects (Gutwin and Penner, 2002), motion blur (as used here), or continuous fading lines as in the enhanced VideoArms technique (A. Tang et al., 2010).

Figure 4. A mock-up of elements of a hybrid visualization. Left, a coloured ripple effect enhances the contact point and provides pressure information. Right, temporal traces emphasize a stroke atom and a ray-cast shadow at the pointing target provides height information.

Conclusion

Deictic gestures are ubiquitous in tabletop collaboration, but the embodiments used in distributed tabletop groupware do not adequately express the range and subtlety of real-world deixis. To provide design information about what components and behaviours are important in deictic gesture, we carried out three observational studies. Our observations of collocated tasks provide new understanding of how deictic gesture works, and helps to explain why both abstract and realistic user embodiments fail to express the full range of deixis over surfaces. In particular, abstract embodiments fail to represent a small but significant portion of gesture morphology and can make it difficult to differentiate between similar gestures; realistic embodiments do not effectively represent the height or pressure of gestures, and have no mechanism for emphasizing movement. We identified five design recommendations to improve the expression of deixis through remote embodiments, and proposed a new embodiment approach that combines the best features of both realistic and abstract techniques.

Our studies suggest numerous avenues for further research. The most obvious next step is to implement the new embodiments for remote deictic gesture, and refine and test visualizations that combine the realistic and abstract approaches. We will then evaluate these designs to determine whether they can, as expected, improve the richness of gestural communication in tabletop groupware. Finally, we plan to expand and refine our characterization of deictic gestures through additional studies and further real-world observations.

Acknowledgements

Thank you to Dr. Ted Leighton, CCWHC; Parks Canada; and the WCVM Wildlife Health Rotation, class of 2009 for their tolerance and participation.

References

Apperly, M., McLeod, L., Masoodian, M., Paine, L., Phillips, M., Rogers, B., and Thomson, K (2003): ‘Use of video shadow for small group interaction awareness on a large interactive display surface’, *Proceedings of the 4th Australasian user interface conference,* vol. 18, 81-90.

Bekker, M. M., Olson, J. S., & Olson, G. M. (1995): ‘Analysis of gestures in face-to-face design teams provides guidance for how to use groupware in design’, *Proceedings of the 1st conference on Designing Interactive Systems,* 157-166.

Benford, S., Bowers, J., Fahlén, L.E., Greenhalgh, C., and Snowdon, D. (1995): ‘User embodiment in collaborative virtual environments’, *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* , 242-249.

Clark, H.H. (2003): ‘Pointing and placing’, in S. Kita (eds): *Pointing: Where language, culture, and cognition meet,* Erlbaum, Hillsdale, NJ, 2003, 243-268.

Fraser, M., McCarthy, M.R., Shaukat, M, and Smith, P. (2007): ‘Seconds Matter: Improving Distributed Coordination by Tracking and Visualizing Display Trajectories’, *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, 1303-1312.

Greenberg, S., Gutwin, C., and Roseman, M. (1996): ‘Semantic telepointers for groupware’, *Proceedings of the 6th Australian conference on Computer-Human Interaction*, 54-61.

Gutwin, C., & Greenberg, S. (1998): ‘Effects of awareness support on groupware usability’, *Proceedings of the SIGCHI conference on Human Factors in Computing Systems,* 511-518.

Gutwin, C., & Greenberg, S. (2000): 'The mechanics of collaboration: Developing low cost usability evaluation methods for shared workspaces', *Proceedings of the 9th IEEE workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises*, 98–103.

Gutwin, C., & Penner, R. (2002): 'Improving interpretation of remote gestures with telepointer traces', *Proceedings of the ACM conference on Computer Supported Cooperative Work,* 49-57.

Hayne, S. , Pendergast, M., and Greenberg, S. (1994): 'Implementing Gesturing with Cursors in Group Support Systems', *Journal of Management Information Systems*, 10, 3, 1994 , 43-62.

Hindmarsh, J., Fraser, M., Heath, C., Benford, S., and Greenhalgh, C. (2000): 'Object-focused interaction in collaborative virtual environments', *ACM Transactions on Computer-Human Interaction*, vol. 7, no. 4, 477-509.

Hindmarsh, J., & Heath, C. (2000): 'Embodied reference: A study of deixis in workplace interaction', *Journal of Pragmatics*, vol. 32, no. 12, 1855–1878.

Izadi, S., Agarwal, A., Criminisi, A., Winn, J., Blake, A., Fitzgibbon, A. (2007): 'C-Slate: A Multi-Touch and Object Recognition System for Remote Collaboration using Horizontal Surfaces', *2nd Workshop on Horizontal Interactive Human-Computer Systems*, 3-10.

Kettebekov, S., & Sharma, R. (2000): 'Understanding gestures in multimodal human computer interaction', *International Journal on Artificial Intelligence Tools*, vol. 9, no. 2, 205–223.

Kirk, D., Crabtree, A., & Rodden, T. (2005): 'Ways of the hands', *Proceedings of the European conference on Computer Supported Cooperative Work*, 1–21.

Kirk, D., Rodden, T., & Fraser, D. S. (2007): 'Turn it this way: grounding collaborative action with remote gestures', *Proceedings of the SIGCHI conference on Human Factors in Computing Systems,* 1039-1048.

Krauss, R.M. and Fussell, S. R. (1990): 'Mutual Knowledge and Communicative Effectiveness', in J. Galegher, R. E. Kraut, and C. Egido (eds.) *Intellectual Teamwork: Social and Technological Foundations of Cooperative Work*, Erlbaum, Hillsdale, NJ, 111-145.

Li, J., Wessels, A., Alem, L., & Stitzlein, C. (2007): 'Exploring interface with representation of gesture for remote collaboration', *Proceedings of the 19th Australasian conference on Computer-Human Interaction: Entertaining User Interfaces,* 179-182.

McNeill, D. (1992): *Hand and mind*, University of Chicago Press, Chicago.

Ou, J., Chen, X., Fussell, S. R., & Yang, J. (2003): 'DOVE: Drawing over video environment', *Proceedings of the 11th ACM international conference on Multimedia,* 100–101.

Ou, J., Fussell, S. R., Chen, X., Setlock, L. D., & Yang, J. (2003): 'Gestural communication over video stream: supporting multimodal interaction for remote collaborative physical tasks', *Proc. of the 5th conference on Multimodal Interfaces,* 242–249.

Pinelle, D., Nacenta, M., Gutwin, C., & Stach, T. (2008): 'The effects of co-present embodiments on awareness and collaboration in tabletop groupware', *Proceedings of Graphics Interface,* 1–8.

Stach, T., Gutwin, C., Pinelle, D., & Irani, P. (2007): 'Improving recognition and characterization in groupware with rich embodiments', *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, 11-20.

Tang, A., Neustaedter, C., & Greenberg, S. (2007): 'Videoarms: embodiments for mixed presence groupware', in N. Bryan-Kinns, A. Blanford, P. Curzon and L. Niggy (eds.): *People and Computers XX*, Springer, London, 2007, 85-102.

Tang, A., Pahud, M., Inkpen, K., Benko, H., Tang, J. C., & Buxton, B. (2010): 'Three's company: understanding communication channels in three-way distributed collaboration', *Proceedings of ACM conference on Computer Supported Cooperative Work*, 271–280.

Tang, J. C., & Minneman, S. (1991a): 'VideoWhiteboard: video shadows to support remote collaboration', *Proceedings of the SIGCHI conference on Human Factors in Computer Systems*, 315–322.

Tang, J., and Minneman, S. (1991b): Videodraw: a video interface for collaborative drawing. *Transansactions on Information Systems* 9, 2 (April 1991), 170-184.

Wigdor, D., Williams, S., Cronin, M., Levy, R., White, K. (2009): 'Ripples: utilizing per-contact visualizations to improve user interaction with touch displays,' *Proceedings of the ACM symposium on User Interface Software and Technology*, 3-12.