Natural Gesture in Descriptive Monologues

Jacob Eisenstein and Randall Davis
MIT Artificial Intelligence Laboratory
200 Technology Square
Cambridge, MA 02139
{jacobe, davis}@ai.mit.edu

ABSTRACT

Gesture plays a prominent role in human-human interaction, and it offers promise as a new modality for human-computer interaction. However, our understanding of gesture is still at an early stage. This study explores gesture in natural interaction and describes how the presence of a display can affect the use of gesture. We identify common gesturing behaviors that, if accommodated, may improve the naturalness and usability of gestural user interfaces.

Keywords

Gesture, multimodal interaction, natural interaction

INTRODUCTION

Empirical evaluation of human-to-human gesture offers the potential to improve the usability of gestural user interfaces. However, existing research on gesture is not always applicable. Psychologists and linguists have studied gesture in natural human-to-human communication. A typical task in this type of study is to describe the events in a short movie [1]. Such research has provided a valuable starting point, but needs to be made more specifically relevant to human-computer interaction. In particular, while gestural user interfaces may involve a computer display for the user to gesture at, we have been unable to find studies from the psychology literature that describe how the presence of a display or diagram affects the use of the gesture.

On the other hand, researchers HCI community have investigated the use of gesture in command-driven multimodal user interfaces [2]. Through the use of Wizard-of-Oz techniques, these studies offer a fairly general picture of how multimodal language is used in command-driven interaction. However, commands comprise only a small fragment of human interaction.

We are interested in a combination of these two approaches: natural human-to-human interaction in the presence of a diagram. Although this type of discourse is very common – e.g., a teacher addressing students with the aid of a whiteboard – relatively little is known about it. The few studies that are relevant (e.g., [3]) are more specifically intended to improve multimodal disambiguation. In contrast, we are interested in establishing basic principles of gesture in the presence of a diagram with an eye toward arriving at a set of specifications for natural gestural user interfaces.

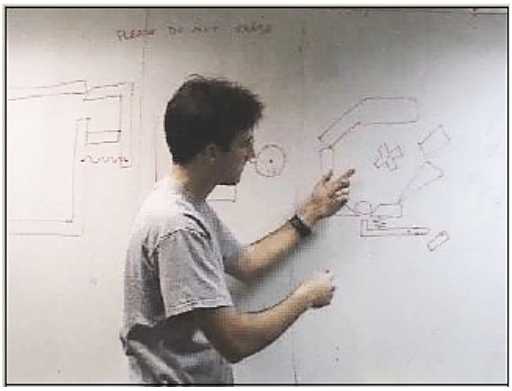

Figure 1: A two-handed gesture used in the description of a "pinball-machine" device

EXPERIMENT

Nine participants were shown simulations of three mechanical devices. After seeing each simulation, the participants were asked to describe that device with the aid of a predrawn diagram. Participants ranged in age from 22 to 28 and included four women and five men. Eight of the nine participants were right-handed, and one was a non-native English speaker. Explanations were videotaped and transcribed. They ranged from 15 to 90 seconds in length. 574 gesture phrases were transcribed; as many as 58 and as few as six gestures were used in a single explanation.

Gestures Refer to the Diagram

All of the speakers relied heavily on the diagram in their explanations. Of the 574 gesture phrases, 549 (95.6%) made reference to the diagram. Of the 25 gestures that did not refer to the diagram, 16 were accounted for by a single speaker. This suggests that the extent to which the diagram is used is

Speaker	Dom. Hand	Left	Right	Both
1	R	0	33	0
2	L	61	0	0
3	R	7	30	3
4	R	43	0	0
5	R	22	25	16
6	R	5	19	8
7	R	13	27	43
8	R	56	30	2
9	R	80	45	6
		287	209	78

Table 1: Distribution of gestures across hands

somewhat idiosyncratic. A Chi Squared test showed a significant difference across speakers ($\chi^2 = 59.1$, $p < 0.001$, df=8), and did not show a significant difference ($p > 0.05$) across tasks.

If it were found that a large proportion of gestures did not refer to the diagram, that would suggest that pens or touch-screens are inadequate as input modalities, since they constrain gesturing to the display surface. However, nearly all observed gestures were on the display surface, suggesting that this constraint does not substantially interfere with users' natural tendencies.

Speakers Often Use Both Hands

Table 1 indicates which hands were used to gesture. Although speakers 1, 2, and 4 gestured exclusively with a single hand, the other six speakers used both hands, sometimes simultaneously. Even though eight of the speakers were right-handed, left-handed gestures were more frequent. The diagrams were arranged from left to right on the board, and speakers almost always stood to the left of the diagram. From this position, gesturing with the left hand made it easier to face the camera, and might explain this preference.

The use of both hands by many of the participants suggests that glove-based user interfaces should offer two gloves, and vision-based systems should track both hands. Pen-based interfaces appear to be poorly suited for two-handed gestures, since few people are comfortable using a pen with their non-dominant hand.

Longer Gesture Units

A *gesture unit* is defined by McNeill [1] as "the period of time between successive rests of the limbs." In a study of gesture without a diagram, he found that more than 50% of gesture units contain only a single gesture phrase. In contrast, we observed much longer gesture units: a median of eight phrases per unit. Speakers frequently completed a gesture and then maintained their hand in a hold position on the diagram until initiating the next gesture. A comparison of these results is shown in Table 2. An area of future research is whether these longer gesture units correspond in any meaningful way with discourse structure.

	Number of Phrases in Unit						
	1	2	3	4	5	6	> 6
With diagram	6	8	11	11	8	2	55
No diagram (from [1])	56	14	8	8	4	2	8

Table 2: Distribution of Gesture Unit Lengths

Speaker	Deictic	Iconic			Total
		Trajectory	Tracing	Other	
1	8	18	3	3	24
2	18	37	5	1	43
3	17	17	3	3	23
4	19	23	1	0	24
5	26	30	2	4	36
6	11	19	1	1	21
7	34	31	5	6	42
8	43	40	3	1	44
9	46	54	22	5	81
	222	269	45	24	338

Table 3: The number of deictic and iconic gestures

Deixis is More Frequent

Iconic gestures are typically defined as the use of the hands to represent the semantic content of the accompanying speech. For example, a speaker may describe a path by tracing its outline, or use a closed fist to represent a rock. Deictic gestures are pointing motions that may refer to real objects or regions of space that were given referential value previously. Iconics have previously been found to occur roughly ten times as often as deictics [1].

As shown in Table 3, the presence of a diagram substantially increases the proportion of deictic gestures. Deictic references to parts of the diagram seem to have substituted for the role otherwise played by iconic gestures. Instead of describing the presence and static structure of objects, the majority of iconic gestures describe the trajectories that objects take. A very typical pattern was "[deictic] This lever moves over **here** [trajectory]." The increased use of deictics is encouraging from a gesture understanding perspective, since discerning the meaning of iconic gestures is thought to be very difficult in the general case [4].

REFERENCES

1. D. McNeill. *Hand and Mind.* The University of Chicago Press, 1992.
2. S. Oviatt. Ten myths of multimodal interaction. *Communications of the ACM*, 42(11):74–81, 1999.
3. I. Poddar, Y. Sethi, E. Ozyildiz, and R. Sharma. Toward natural gesture/speech HCI: A case study of weather narration. In *Proceedings of 1998 Workshop on Perceptual User Interfaces*, 1998.
4. C. J. Sparrell. Coverbal iconic gesture in human-computer interaction. Master's thesis, Massachusetts Institute of Technology, 1993.