

Understanding Naturalness and Intuitiveness in Gesture Production: Insights for Touchless Gestural Interfaces

Sukeshini A. Grandhi, Gina Joue, Irene Mittelberg

Natural Media and Engineering, HumTec, RWTH Aachen University, Germany

{grandhi, joue, mittelberg}@humtec.rwth-aachen.de

ABSTRACT

This paper explores how interaction with systems using touchless gestures can be made intuitive and natural. Analysis of 912 video clips of gesture production from a user study of 16 subjects communicating transitive actions (manipulation of objects with or without external tools) indicated that 1) dynamic pantomimic gestures where imagined tool/object is explicitly held are performed more intuitively and easily than gestures where a body part is used to represent the tool/object or compared to static hand poses and 2) gesturing while communicating the transitive action as how the user habitually performs the action (pantomimic action) is perceived to be easier and more natural than gesturing while communicating it as an instruction. These findings provide guidelines for the characteristics of gestures and user mental models one must consciously be concerned with when designing and implementing gesture vocabularies of touchless interaction.

Author Keywords

Gestures, intuitiveness, naturalness, user centric interaction design, cognitive semiotic principles, metonymy

ACM Classification Keywords

H.5.2 Information Interfaces and Presentation: User Interfaces

General Terms

Design, Human Factors, Experimentation.

INTRODUCTION

Touchless gestural interfaces have many potential applications such as in sterile/clean room environments, collocated shared technology and robotics [13,14]. However, unlike touch gestures, touchless gestures remain largely a notion developed in science fiction (as depicted in the popular sci-fi movie *Minority Report*) and have only been implemented to a limited degree in a few proof-of-concept research applications (e.g. [1,2]) and video games (e.g. Sony Eye Toy, Microsoft Kinect). This limited implementation of touchless gestures is due to challenges in 1) achieving accurate and meaningful gesture recognition

[1,2] and 2) identifying natural, intuitive and meaningful gesture vocabularies appropriate for the tasks in question [10]. While computer vision researchers have long been working on gesture recognition, the challenge of generating natural and intuitive touchless gesture vocabulary while keeping user experience in mind, has received relatively very little research attention. The gesture vocabulary used has often been an ad-hoc choice made by the designer to trigger certain pre-assigned actions (e.g. the gesture of clapping one's hands to turn on the computer). Mostly chosen for their ease of implementation to facilitate distinctive recognition and segmentation, such gestures have arbitrary mappings that require users to learn a set of gestures and the associated commands they trigger. This often puts a strain on user memory and defeats the purpose of using gestures as a way to facilitate intuitive and natural interaction [5,10]. The aim of the work presented in this paper is to take a small step towards understanding what makes touchless gesture production natural and intuitive to provide design guidelines for such interfaces. We adopt a vernacular definition of “natural” as being marked by spontaneity and “intuitive” as coming naturally without excessive deliberation. While several classifications of gestures and their functions exist [5,11,12,16], we adopt a semiotic or communicative approach to gestures [12] for HCI, which corresponds to Caldoz's classification of *semiotic* hand movements [4], or movements that communicate information from shared common ground. Since 90% of semiotic gestures are accompanied by speech [11], we look at gestures in association with speech.

RESEARCH QUESTIONS

In particular, this paper focuses on the communication of transitive actions, or actions involving the manipulation of objects with or without external tools. For example, how do we communicate to the system using gestures that a screen be erased? Specifically we view gesture forms in terms of its relationship to what is being communicated or represented. This was motivated by neuropsychology, psychiatry and developmental psychology literature which all suggest that for transitive gestures (e.g. brushing teeth with a toothbrush), it is more normal for the hand itself to hold an imagined object (e.g. holding an imaginary toothbrush) rather than representing a body part as the object itself (e.g. finger representing the toothbrush itself). Using the hand to represent the object itself is termed as “body-part-as-object” in neuropsychology, and is an example of “object substitution” in developmental

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.

Copyright 2011 ACM 978-1-4503-0267-8/11/05...\$10.00.

psychology. In neurological examinations, this type of action is considered an error associated with movement planning disorders [7]. The ability to perform hand actions where the hand holds an imagined object, referred to as pantomimed action with an imagined object in neuropsychology, is considered the proper developmental trend in children [3]. Based on semiotic theory, “body-part-as-object” gestures are referred to as exhibiting “internal metonymy”, and “holding-imagined-object” gestures are referred to as “external metonymy” [12]. Therefore, we hypothesize that body-part-as-object gestures, i.e. reflecting internal metonymy, are generally not as intuitive as hand actions holding-imagined-objects, i.e. reflecting external metonymy. In order to develop design principles for effective communication of transitive actions to a system using gestures, we considered the following two broad research questions:

1. What aspects of a gesture (such as motion, hand shape and form) are natural and intuitive when communicating a transitive action? In particular, are gestures using hand actions holding-imagined-objects more intuitive and natural than using body-part-as-object?
2. Is it easier to gesture a transitive action when the user communicates as how s/he habitually performs the action or when the user communicates it as an instruction? That is, should one gesture as “this is how I do it” or “this is how you should do it”?

USER STUDY

Procedure

In order to address the above research questions we adopted a user-centric approach [13,15,16]. Participants were told that the study explored people’s intuitive preferences and natural tendencies in verbal and non-verbal expressions. Sixteen (8 female) native or near-native American English speakers of different U.S. regions were recruited from a university visitor pool. All except one (over 60 years old) were between 20-30 years old, and 68.8% were strongly right-handed based on the Edinburgh Handedness Survey.

Participants were presented pairs of pictures showing a “before” and “after” scenario. These scenarios (Table 1) reflected simple computer tasks (e.g. cut, erase), but were camouflaged as everyday non-computer scenarios to minimize the influence of conceptual models of performing these tasks on pre-existing input devices. Our definitions of intuitive and natural translate into the assumption that spontaneous and more frequently produced gestures in common tasks can indicate what is more “natural and intuitive,” and how well people can maintain a so-called unintuitive gesture, can shed further light on naturalness.

As soon as the picture pair disappeared, participants were asked to speak aloud while performing a gesture to “explain and show” the experimenter (seated in front of them in a position where the pictures could not be seen), what needed to be done to achieve what was in the after-picture from the

before-picture. Each picture pair was shown for 3 seconds on a 24” monitor at a resolution of 1024 x 768 (stimulus visual angle of 7.15°). A total of 19 pairs of before-after pictures were used. The scenarios were divided into three categories: Transitive actions usually done 1) with an external tool (e.g. slicing an apple with a knife), 2) without an external tool (e.g. laying plates on a table) and 3) with or without an external tool (e.g. wiping a surface with bare hands or a cloth). Participants performed the “explain and show” scenario for the above tasks under three conditions to distinguish between the styles of communication (Instructional vs. Habitual) and to understand naturalness and intuitiveness of performing gestures with internal metonymy (body-part-as-tool).

1. Instructional: Participants instructed the experimenter on what needed to be done and began their instructions with the sentence frame, “You need to...”
2. Habitual: Participants explained how they normally would perform what needed to be done, beginning their explanations with the sentence frame, “This is how I...”
3. Instructional using body-part-as-tool: Participants repeated the Instructional condition, but they additionally were required to use their hands to represent any tool they might need to use (internal metonymy).



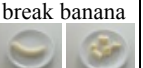


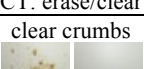
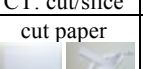
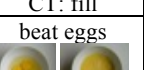
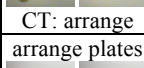
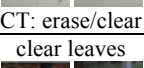
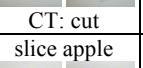
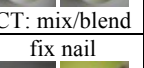
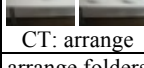
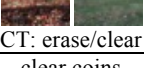
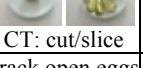

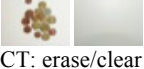


 CT: arrange	 CT: erase/clear	 CT: cut/slice	 CT: fill
 CT: arrange	 CT: erase/clear	 CT: cut	 CT: mix/blend
 CT: arrange	 CT: erase/clear	 CT: cut/slice	 CT: place/fix
 CT: arrange	 CT: erase/clear	 CT: open	 CT: place/fix
	 CT: erase/clear	 CT: close	 CT: write/sketch

Table 1 The 19 scenarios and corresponding computing task (CT)

The order of the three conditions as well as the 19 scenarios within each condition was randomized for each participant to minimize order effects. At the end of the study, participants were interviewed on their perspectives on the naturalness and intuitiveness of the communicating styles and tool/object representation in their gestures (internal or external metonymy). All sessions were recorded using two high-speed video cameras from two visual perspectives.

FINDINGS

Data from 912 video clips (16 participants x 19 scenarios x 3 conditions) were blindly annotated by a coder. Parts of

the annotations were verified for consistency by the authors, and a second annotator blindly coded 25% of the data. The inter-rater reliability for the annotators was good (Cohen's $\kappa = 0.64$, $p < 0.001$). Annotations consisted of 1) right- or left-hand use; 2) number of hands used; 3) use of body-part as tool/object vs. pantomimed action of imagined tool/object use; 4) dynamic (moving gesture) vs. static gesture poses; 5) distinctions between gestures involving a tool (device used to perform the dominant action, e.g. knife used for cutting) or object (entity upon which the dominant action is performed, e.g. the apple that is cut); and 6) distinctions between gestures that set up the context vs. execute the dominant action of the transitive action.

Gestures exhibited motion than static poses

Gestures representing the dominant action were seldom shown as static hand forms or poses (e.g. palm held sideways with extended fingers to represent cutting), but were almost always (95.7%) dynamic motions demonstrating the “use of tool”.

Pantomimed actions of holding-imagined-tool were more intuitive and easier than body-part-as-tool

In conditions 1 and 2 where participants were not told how to represent tools in their gestures, participants gestured holding an imaginary tool significantly more often than representing a part of their hand as a tool ($\chi^2(4) = 54.5$, $p < 0.001$). Figure 1 and Video Clip 1 (video file uploaded with the paper) show a subject holding an imaginary knife in the right hand to cut an imaginary apple held in the left. Figure 2 and Video Clip 2 show the right palm to represent a knife. In gestures where no external tool was used and the dominant action was pantomimed, the object, by default, was manipulated (Video Clip 3). In condition 3 where participants were explicitly asked to use body-part-as-tool to represent any tool used in the action, they failed to do so 77.5% of the time when an external tool was necessary. Only 3.7% of these participants “corrected” themselves to the instructions of the condition without prompting (Video Clip 4). The intuitiveness and ease of pantomiming with an imaginary tool in hand is further corroborated by participants' comments such as the following that were typical: *“most difficult was using my hand as a tool. That really didn't come naturally”, “It felt weird to use my finger as the tool because I don't do that. It's just more usual to use the tool...it's kind of awkward.”*

Gestures represented objects explicitly rather implicitly

In all the three conditions, bimanual gestures were performed three times more often than one-handed gestures. Objects on which an action was performed, be it with an external tool or the hand itself, were depicted explicitly 45% of the time, as opposed to implicitly assuming its presence in one of three forms (illustrated in Figures 1-3 by the left hand for an apple being cut): pantomimed action of holding-imagined-object (52.5% in all conditions, Fig 1), using a hand to situate the space in which an object is placed (34.0% in all conditions, Fig 3) and part of hand as

object when not explicitly asked to do so (18.1% in conditions 1 and 2, Fig 2).



Figure 1

Figure 2

Figure 3

Communicating as “actor” easier than as commander”

Almost all participants reported that gesturing how they habitually perform an action was easier and more intuitive than gesturing an instruction. Comments such as the following were typical from the participants: *“it was easier to visualize and pantomime what I would normally do than relay that information to you...”*, *“I found showing how I do it easier because it was more natural because I know how to do it and instructing you I don't know if you will understand my hand gestures”*. Video Clip 5 exemplifies the struggle and awkwardness in terms of body orientation of a participant gesturing while instructing.

DISCUSSION AND IMPLICATIONS

Our findings suggest that communicating with a system through gestures may be easier when designers adopt an embodied approach [6]. Namely, user experience can be enhanced when the gesture vocabulary is developed based on the understanding that actions are embodied, i.e. situated in real-life social and physical scenarios. In particular, our findings provide the following key guiding principles for the design of touchless gestures involving transitive actions.

Firstly our findings have implications for how designers should communicate the presence and use of touchless gestural interactions in a system, be it in the form of text, symbols or illustrations and demonstrations [15]. Gestures triggering manipulation of objects should be dynamic iconic representations of the motion required for the manipulation, rather than static iconic hand poses. For example, a gesture to trigger “delete” should be a “wiping” hand movement rather than a static hand sign. Instructions for use of such gestures may also be helpful as illustrations of the pantomimed action. For example, rather than showing an image of a knife to indicate the gesture to use, an illustration of hand holding a knife performing the cutting act could be shown.

Secondly, gestures to trigger tasks that suggest the use of a tool should not impose body-part-as-tool compositions because such gestures do not seem to come naturally and intuitively. Instead designers should consider selecting gestures that pantomime the actual action with imagined tool in hand. For example, a gesture to slice could be a pantomiming motion of cutting with an imaginary knife in hand rather than the gesture, open palm moving vertically up and down. However, it must be noted that when the hand shape can represent the tool or object, or draw attention to an important aspect of the tool/object being communicated, in an unambiguous way due to anatomy [13], using body-part-as-tool may make more sense. This may be why certain

body-part-as-object gestures have become emblematic, e.g. index and middle fingers showing scissors for cutting.

Thirdly, gestures in space to trigger manipulation of objects should be two-handed, as the non-dominant hand often appears to provide a reference frame while the dominant hand performs the transitive gesture. This is consistent with principles of Guiard's kinematic chain model of asymmetric bimanual actions [8] and other research, which suggests that two hands together provide a better sense of virtual space than one hand for task execution [9].

Finally, since gestures performed in habitual conditions were reported to be easier and more intuitive, adopting an egocentric rather than an allocentric (viewer's) frame of reference to gesture may ease the awkwardness of interacting with a system for the user, which is often a major concern in human robotic communication [14]. That is, instilling the mental model of an actor egocentrically pantomiming the action, i.e. with a communicative perspective of "I would like to do Task X", reduces the propensity of users to re-orient their reference frames to position gestures in space when communicating the perspective of "You do Task X for me".

LIMITATIONS AND FUTURE WORK

In this study we concentrated on hand forms, motion and communication style. However, many other characteristics such as the intensity of hold and other gesture kinematics may also afford insights on what comes naturally and intuitively. Consequently, promising directions for future work include using motion capture data to identify other characteristics of gestures which may help define intuitive and natural interaction, such as distance, speed, trajectory, image schemas and gestural space.

Although learned behavior can become common and even "natural", we wanted to avoid the preconceived assumptions of how one is familiar with interacting with existing system interfaces and disguised our tasks as everyday tasks. Of course, this also gives rise to limitations as to how translatable these tasks are to specific computer applications. We acknowledge that a good gesture vocabulary is typically task/context-specific, and it is a challenge to develop generic principles that can be applied across different contexts [13]. However, this challenge can be met if researchers collectively bring together their findings from diverse tasks/contexts. Thus, work presented here makes a small yet significant contribution towards the understanding of the core design principles to create intuitive and natural user experience for touchless gestural interfaces.

ACKNOWLEDGMENTS

This work has been funded by the Excellence Initiative of the German Research Foundation.

REFERENCES

1. Baudel, T. and Beaudouin-Lafon, M. Charade: Remote Control of Objects Using Free-Hand Gestures. *Communications of the ACM* 36, 7 (1993), 28-35.
2. Bolt, R. "Put-that-there": Voice and gesture at the graphics interface, Proceedings of the 7th annual conference on Computer graphics and interactive techniques, ACM, New York, NY (1980), 262-270.
3. Boyatzis, C.J. and Watson, M.W. Preschool children's symbolic representation of objects through gestures. *Child Development* 64 (1993), 729-735.
4. Cadoz, C. (1994). Les realites virtuelles. Dominos, Flammarion.
5. Cassell, J. A Framework For Gesture Generation and Interpretation. In R. Cipolla and A. Pentland (eds.), *Computer Vision in Human-Machine Interaction*. Cambridge University Press, New York (1998), 191-215.
6. Dourish, P. *Where the Action Is: The Foundations of Embodied Interaction*. MIT Press, Cambridge, 2001.
7. Goodglass, H., and Kaplan, E. Disturbance of gesture and pantomime in aphasia. *Brain* 86 (1963), 703-720.
8. Guiard, Y., "Asymmetric Division of Labor in Human Skilled Bimanual Action: The Kinematic Chain as a Model," *The Journal of Motor Behavior*, 19 (4), 1987, pp. 486-517.
9. Hinckley, K., Pausch, R., and Proffitt, D. Attention and visual feedback: the bimanual frame of reference. *Proc. Symposium on Interactive 3D Graphics*, ACM Press (1997), 121-126.
10. Lee, J.C. In search of a natural gesture. *XRDS: Crossroads, The ACM Magazine for Students* 16, 4 (2010), 9-12.
11. McNeill, D. *Hand and mind: What gestures reveal about thought*. University of Chicago Press, Chicago, 1992.
12. Mittelberg, I. and Waugh, L.R. Metonymy first, metaphor second: A cognitive-semiotic approach to multimodal figures of speech in co-speech gesture. In C. Forceville and E. Urios-Aparisi (eds.), *Multimodal Metaphor*. Mouton de Gruyter, Berlin (2009), 329-358.
13. Nielsen, M., Moeslund, T., Storrang, M. and Granum, E. A procedure for developing intuitive and ergonomic gesture interfaces for HCI. In A. Camurri and G. Volpe (eds.), *Gesture-Based Communication in Human-Computer Interaction: 5th International Gesture Workshop, GW 2003*, LNCS 2915, Springer, Berlin (2004), 409-420.
14. Powers, A. What Robotics Can Learn from HCI. *ACM Interactions Magazine* 15, 2 (2008), 67-69.
15. Saffer, D. *Designing Gestural Interfaces*. O'Reilly Media, 2008.
16. Wobbrock, J.O., Morris, M.R. and Wilson, A.D. User-defined gestures for surface computing. *Proc. of the 27th Int. Conf. on Human Factors in Computing Systems*, ACM, New York, NY (2009), 1083-1092.