How We Gesture Towards Machines: An Exploratory Study of User Perceptions of Gestural Interaction

Sukeshini A. Grandhi
Eastern Connecticut State University
Willimantic, CT USA
grandhis@easternct.edu

Chat Wacharamanotham
RWTH Aachen University
52056 Aachen, Germany
chat@cs.rwth-aachen.de

Gina Joue
RWTH Aachen University
52056 Aachen, Germany
joue@humtec.rwth-aachen.de

Jan Borchers
RWTH Aachen University
52056 Aachen, Germany
borchers@cs.rwth-aachen.de

Irene Mittelberg
RWTH Aachen University
52056 Aachen, Germany
mittelberg@humtec.rwth-aachen.de

CHI 2013 Extended Abstracts, April 27–May 2, 2013, Paris, France.
ACM 978-1-4503-1952-2/13/04.

Abstract
This paper explores if people perceive and perform touchless gestures differently when communicating with technology vs. with humans. Qualitative reports from a lab study of 10 participants revealed that people perceive differences in the speed of performing gestures, sense of enjoyment, feedback from the communication target. Preliminary analysis of 1200 gesture trials of motion capture data showed that hand shapes were less taut when communicating to technology. These differences provide implications for the design of gestural user interfaces that use symbolic gestures borrowed from human multimodal communication.

Author Keywords
Gesture interfaces, interpretation, naturalness, human centric, user experience, perception, interaction design, cognitive principles, models

ACM Classification Keywords
H5.2. User interfaces: User-centered design, Theory and Methods.

Introduction
The popularity of touchless gestures in video games (e.g. using Sony's Eye Toy, Microsoft's Kinect) and advances in gesture recognition algorithms have led to

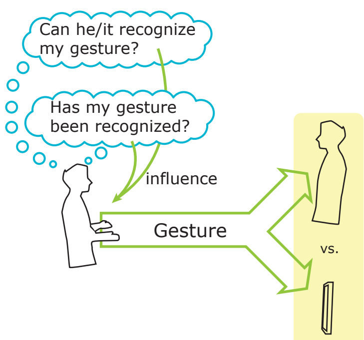

Figure 1: The communication target may influence gestural interaction and user experience. What aspects do users perceive differently when gesturing to technology?

applications in areas such as living room control, object search, and healthcare [2,3,7]. One goal of gestural interaction is to make gesturing less awkward and more natural [4,9]. To achieve this, researchers have often borrowed gesture vocabulary from symbolic gestures that occur naturally in human multimodal communication [1,4,7].

In this paper, we present an exploratory study as part of a larger research agenda that aims at understanding the nature of user perceptions and cognitive models of gestural interaction and how they influence user experience. We began by exploring if the user experience of gesturing in human–human multimodal communication translates to similar user experience when interacting with gestural interfaces even when using the same gesture vocabulary.

Background

The evaluation of gestural interaction based on user perceptions, even though an important factor in technology adoption and use, has not been extensively explored. Previous studies have been limited to pen-based gestures, touch gestures, or motion gestures involving handheld devices, typically based on agreement scores, ease of learning of and recall of gesture vocabulary [7,9]. Other studies have focused on perceived ease of use of gesture, its intuitiveness, its relevance to the task at hand, and social acceptability of gesture performance [6,8]. However, it is unclear what characteristics of gestural interaction users perceive differently when interacting with technology using touchless gestures, and how these differences influence user experience [Fig. 1]. Understanding the perception of gestural interaction can inform the design of gestural interfaces for an enhanced user experience.

A step towards such understanding is to compare interactions with humans and those with technology. The media equation [5] states that people frequently engage in human-computer interaction as they would human-human interactions. In computer-mediated communication, e.g., large or close-up faces on screen can be perceived as invading personal space, similar to human-human communication. However, do these perceptions hold when explicit actions such as gestures are required when communicating with computers? If differences in perceptions of communicating with humans vs. computers exist, they could potentially influence communication with technology using gestures, which in turn will influence how we design gestural interfaces.

Research Questions

This exploratory study is guided by two broad research questions pertaining to the influence of the communicating target on how one perceives and performs gestures. Specifically,

1. *Influence of communication target on user perception of gesture:* How do people perceive performing gestures to communicate with technology compared to humans? If people perceive and experience a difference, what are these differences?
2. *Influence of communication target on gesture production:* Does the gesture performance vary with the communication target (human vs. technology), and if so, what aspects of a gesture (such as speed, hand shape and gesture size) vary with the communication target?

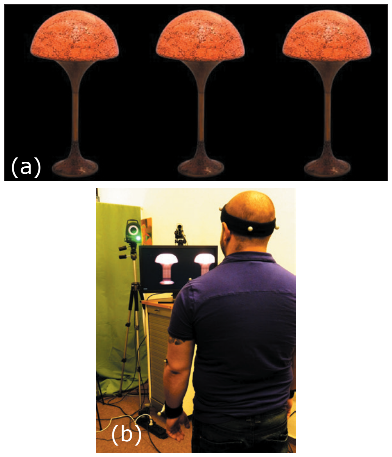

Figure 2: Study & stimuli. Participants monitored problems with stimuli (a) shown sequentially on a screen (b).

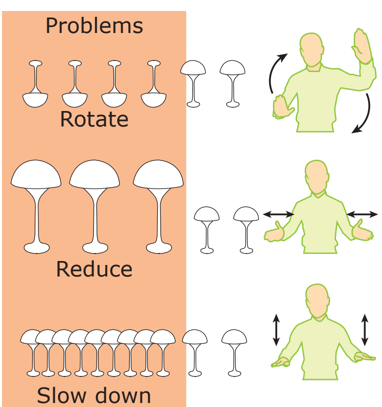

Figure 3: Three problems to be communicated and the corresponding gestures.

User Study

We conducted a lab study with 10 native German speakers (6 female; age 20–40 years old) recruited from various disciplines (medicine, waste management, physics, and computer science) on a local university campus. All participants were strongly right-handed (Edinburgh Handedness Survey). Three participants rated their skills in use of free-form gesture based technology (Nintendo Wii) to be above average, whereas most frequent Wii usage was once a week (*Mdn* = "Every few months").

The task was to monitor pictures of a lamp shown in an assembly line sequence [Fig. 2], and to communicate how to fix one of three problems with the pictures using a corresponding predefined gesture. The three problems were that the pictures were (1) too big, (2) upside down, and (3) moving too fast. We limited the number of problems/gestures to be communicated to minimize confounding variable. The three predefined gestures [Fig. 3], motivated by their iconicity or based on existing symbolic gestures as identified by consulting semiotic gesture researchers, were

> *Rotate*: Move the hands along the coronal/frontal plane in a clockwise circular motion with the palms flat and facing outward.
>
> *Reduce*: Move the hands in a horizontal motion towards each other with the palms flat and facing each other.
>
> *Slow Down*: Move the hands up and down in vertical motion with the palms flat and facing down.

At the beginning of the study, the experimenter explained and demonstrated the gestures to the participants. They were told they had to do the task under two separate conditions: (1) Gesturing to a human experimenter: they were told that the experimenter sitting across them would fix the problem when the appropriate gesture was performed. The experimenter was strictly instructed not to provide any feedback to the participants. (2) Gesturing to technology (Wizard of Oz): they were told that the computer screen—that displayed the pictures—was fitted with a gesture recognition technology, which would fix the problem when the appropriate gesture was performed. The experimenter was hidden behind the curtain to ensure that the participants believed that they were working with the technology and not a human Wizard of Oz [Fig. 4].

While we did not specify how often participants needed to perform each gesture, to ensure that gestures were initiated exactly at the time of communication in both conditions, participants were instructed to step on a foot pedal before gesturing. In the human condition, the pedal signaled the experimenter to look at them when they wanted to gesture towards him. In the technology condition, we informed participants that the pedal activated the gesture recognition system.

Each of the three problems was presented 10 times in each condition in two trials. The order of the two conditions as well as the problems/trials within each condition were randomized for each participant to minimize order effects. All sessions were recorded using two high-speed video cameras from two visual perspectives and 14 infrared motion capture cameras. Participants were fitted with 18 markers [Fig 5]. At the end of the study, participants were interviewed on their perceptions and experience of communicating using gestures to technology and to humans. We asked them

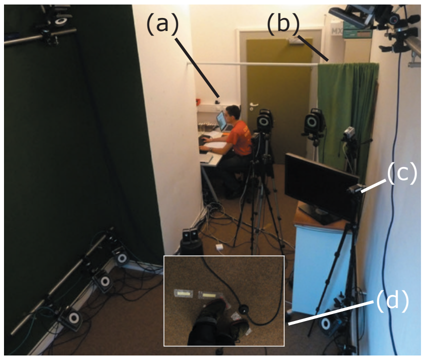

Figure 4: Experimental setup showing position of (a) experimenter, (b) curtain, (c) stimuli, and (d) participant.

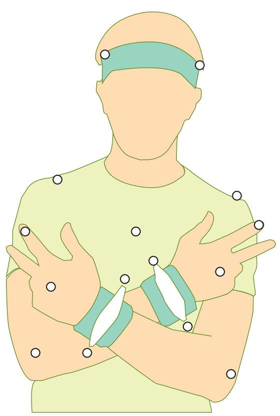

Figure 5: Positions of markers fitted to participants. (3 markers are not visible from this perspective)

to compare and contrast their experience in the two conditions.

Qualitative Findings

In expressing how they perceived the experience of gesturing to a human vs. technology, participants reported a range of preferences and factors.

Communication target preference:
Five participants reported a preference for performing the gestures when communicating with technology, three favored the human communicator and the remaining two participants reported no preference of gesturing to either technology or human.

Perceptions of gesture characteristics, communicator feedback, and communicator comprehension:
Participants who favored technology spoke in terms of the speed of performing gestures, sense of enjoyment, and feedback from the addressee. They perceived that gesturing "*felt faster with the system"* and that there was *"no need for feedback [a signal, indicating the readiness to read the gestures] with [the] monitor"*. When gesturing to humans, these participants felt compelled to *"wait for the person"* and to *"prepare the gesture for the experimenter".* One participant perceived a sense of fun and enjoyment with technology: *"it was like a game with the camera".* One participant perceived that gesturing to technology was "*focused inward*" but gesturing to a human was "*outward…I felt I had to convince someone*".

Participants who favored the human communicator spoke of naturalness and confidence in gesture recognition. They expressed their perception in terms of *"I knew that [the] experimenter would see me but not the monitor"*, *"I felt like [the] experimenter recognized the gesture earlier…had to be more careful with the screen"*, *"not sure if it was going to work with the screen…the monitor does not think like a person"*; *"I knew it would be OK with the experimenter…system didn't give feedback"*; *"easy to do with a person…not sure if it was going to work with the monitor"*.

Preliminary Quantitative Findings

Given the perceived differences in how confident certain users were in the gesture recognition technology compared to interacting with a person, we were interested in seeing if any differences existed in how well-formed the hand shapes in gestures were in gesturing to technology vs. humans. The hand shape for all gestures was designed to be flat and was considered to have *a tautness of hand shape* close to 180° [Fig. 6]. Preliminary analysis of motion capture data showed consistent differences of the tautness of hand shape. There was a significant main effect of gesturing target ($F_{1,1107.07}$ = 22.90, p <.0001, mixed-effects analysis of variance. Details in [Fig. 7, 8]). The hand shape was significantly less taut (smaller angle between palm and finger) when communicating with technology (M = 166.46° 95% CI = [165.70, 167.23]) than when communicating with humans (168.34° [167.69, 168.99]) for all three gestures, i.e., the hand shape was less taut when communicating with technology.

Implications

In this study, we focused on whether people perceived any differences in how they gestured to technology vs. humans. Eighty percent of the participants perceived differences in the act of gesturing itself in terms of factors such as ease, speed, naturalness of gesturing as

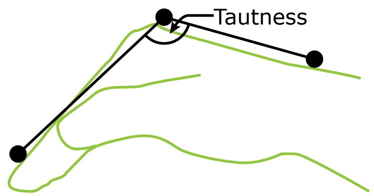

Figure 6: Tautness is the angle between palm and finger

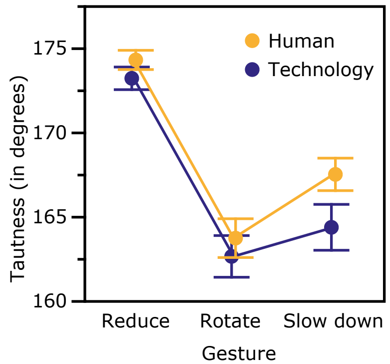

Figure 7: The differences in tautness of hand shape between gesturing to human vs. to technology (Mean and 95%CI)

Gesture	*t (df)*	*p*	*d*
Reduce	2.65 (169)	.009	0.20
Rotate	3.00 (188)	.003	0.22
Slow down	5.55 (186)	<.001	0.41

Figure 8: Result from the simple main effect of tautness of hand shape of each gesture. (Paired t-tests with Bonferroni correction: α = .017)

well as in their desire and expectation of feedback from addressee. This suggests that when using symbolic gestures occurring in human communication, designers cannot assume the same perceptions of interaction with technology. The different perceptions have to be specially catered for in the design of gestural interfaces. For example, a system that allows users to gesture without having to wait for a signal will increase user confidence in the gesture recognition. Also, providing feedback to acknowledge the gesture itself instead of just executing the operation will reduce the perceived uncertainty, cf.[10].

Although the reported differences were not consistent across participants, they provide a richly structured set of factors that can be accessed to understand and enhance the user experience. Some participants perceived gesturing to technology as being unnatural and attributed it to a perceived lack of recognition feedback from the technology. Although the experimenter followed a strict protocol of not providing any feedback other than turning back to the system and fixing the error if the participant correctly performed the gesture, some participants perceived tacit feedback. In contrast, they expressed uncertainty of the technology's ability to recognize their gestures, despite (1) the assurance that pressing the pedal activated the gesture recognition and (2) the fact that the objects on the screen were visibly corrected. This highlights the following (1) the experimenter was providing subtle feedback knowingly which warrants new investigation with a different experimental set up where the human does not look at the participant and/or (2) some users were bringing in a social interaction mental model when gesturing to technology. Designers should consider bootstrapping users' tendencies to extrapolate social interaction models when using such gestures.

When communicating with technology, participants on average used less taut hand shapes. This decreased tautness may be the result of the egocentric manner (inward focus) of gesturing as reported by some participants. However, the decreased tautness is counter-intuitive to the qualitative responses of certain participants who reported the lack of confidence in gesture recognition technology. One would expect these participants to be more careful with gesture hand shape formation if they had less confidence in technology recognition. This contradiction calls for further investigation in understanding human perceptions of technology and how that influences gesture production. Our results also contradict the media equation, which suggests an underlying assumption of perceived similarity in the two target situations. Further research is needed to see if gestural input influences or mediates these varying perceptions.

Factors perceived to be favorable with technology were a sense of enjoyment, a sense of speed due to the diminished need for social interaction with human (no need to wait for human attention, feeling inward focused). This suggests that interaction designs can highlight fun and egocentric characteristics of gestural interaction to enhance the user experience. Thus, designers should pay attention to identifying scenarios for which gestures can be appropriate, and focus on aligning the task goals (fun, efficiency, etc.) to the gestural interaction type used.

Future Work

In this exploratory study, we used a small sample of iconic and metaphoric gestures. Our preliminary findings suggest the need to look at how these perceptions vary with a wider sample of gesture types and perhaps with user-defined or personalized gestures compared to gestures imposed by designers, as this is essential when designing an appropriate gesture vocabulary [1]. The differences in hand shape formation suggest a difference in gesturing to humans vs. technology. A deeper analysis of the motion capture data is required to answer whether there are differences in other measures such as gesture volume, distances, and speed, and how these gesture performance characteristics correlate with a user's perception of gesturing to different targets.

While our findings indicate differences in how users perceive gestural interaction with humans vs. technology, the lack of consistent user perceptions raises more interesting research questions that point to future research directions. We propose to explore further what the role of gestural interaction in the media equation is, what constitutes convincing gesture recognition feedback from technology, and how we can provide this without compromising seamless interaction. Collectively the initial findings of this study provide the basis for a research agenda that aims to draw on user perceptions and user experience to design effective gestural interfaces.

Acknowledgements

This work has been funded in part by the Excellence Initiative of the German Federal and State Governments and the German B-IT Foundation.

References

[1] Grandhi, S.A., Joue, G., Mittelberg, I. Understanding Naturalness and Intuitiveness in Gesture Production: Insights for Touchless Gestural Interfaces. In *Proc. CHI '11*, 821–824.

[2] Holz, C. and Wilson, A.D. Data Miming: Inferring Spatial Object Descriptions from Human Gesture. In *Proc. CHI '11*, 811–820.

[3] Johnson, R., O'Hara, K., Sellen, A., Cousins, C., Criminisi, C. Exploring the Potential for Touchless Interaction in Image-Guided Interventional Radiology. In *Proc. CHI '11*, 3323–3332.

[4] Nielsen, M., Moeslund, T., Storring, 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 (2004), 409–420.

[5] Reeves, B., & Nass, C. *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. Cambridge University Press, New York, NY, USA, 1998.

[6] Rico, J. and Brewster, S. Usable Gestures for Mobile Interfaces: evaluating social acceptability. In *Proc. CHI '10*, 887–896.

[7] Vatavu, R.D. User-Defined Gestures for Free-Hand TV Control, In *Proc. EuroITV '12*, 45–48.

[8] Vatavu, R. D., Vogel, D., Casiez G., Grisoni L., Estimating the Perceived Difficulty of Pen Gestures, In *INTERACT '11*, LNCS 6947. Springer, 89–106.

[9] Wobbrock, J.O., Morris, M.R., Wilson, A.D. User-Defined Gestures for Surface Computing. In *Proc. CHI '09*, 1083–1092.

[10] Norman, D.A. Natural User Interfaces Are not Natural. *Interactions,* 17, 3 (2010), 6–10.